Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Novel Dynamic Baysian Canonical Correlation Analysis Method for Fault Detection

Jiaxin Yu*, Zeyu Yang*, Le Zhou, Lingjian Ye***, Zhihuan Song***

**College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China (e-mail: JiaxinYu@zju.edu.cn; yangzeyu@zju.edu.cn; songzhihuan@zju.edu.cn).*
*** School of Automation and Electrical Engineering, Zhejiang University of Science & Technology, Hangzhou, 310023, China(e-mail: zhoule@zust.edu.cn;).*
**** Ningbo Institute of Technology, Zhejiang University, Ningbo, 315100, China (e-mail: lingjian.ye@gmail.com)*

Abstract: In the field of Multivariate Statistical Process Monitoring (MSPM), process dynamics has always been the focus. Besides, considering the uncertainty in chemical processes, latent variable models are extended to the probabilistic framework, in which maximum likelihood estimation with expectation maximization (EM) algorithm is adopted for parameter learning. However, the modelling performance is restricted owing to the reason that these models either neglect the static characteristics reflecting process structure or suffer from over fitting and local optimum. To tackle these issues, a dynamic Baysian canonical correlation analysis (DBCCA) model is developed through combining the consideration of process dynamics with the variational CCA and utilized for fault detection. More specifically, both static structural characteristics and process dynamics can be simultaneously captured in DBCCA model. In essence, the variational Bayesian approach renders effects of regularization, alleviating the dilemma in traditional maximum likelihood estimation methods by nature. The effectiveness of proposed method is testified on the well-known Tennessee Eastman (TE) benchmark, where improvements are attained.

Keywords: Dynamic Process Modeling, Dynamic Bayesian Canonical Correlation Analysis, Fault Detection, Variational Inference.

1. INTRODUCTION

Recent years have witnessed the rapid upgrading of automation production in chemical processes, which poses arduous challenges for the monitoring of process operation. Consequently, fault detection techniques play an indispensable role in automatic control systems (Qin, 2012). Multivariate Statistical Process Monitoring (MSPM) methods such as principal component analysis (PCA) and partial least squares (PLS), in particular, are receiving much attention due to their capability of dimension reduction of massive process data (Yin et al., 2012). In this setting, as a result of process units inertia and the adjustment of closed-loop control systems, process measurements sampled by distributed control systems (DCS) are auto-correlated and thus reflects process dynamics, adding difficulties to the process modelling task (Ge et al., 2013).

For dynamic process modelling, several researches have been made, among which dynamic PCA (Ku et al., 1995) is the most widely used technique. In this method, time lagged samples are stacked to form an augmented matrix, which is then dealt with ordinary PCA. After this pioneer work, many dynamic modelling methods based on augmented matrix were proposed, such as DPLS (Liu et al., 2019), DICA (Lee et al., 2004), etc. Besides, time series method is also proposed for dynamic process modelling (Negiz et al., 1994), where residuals are calculated for monitoring. However, time series method is more suitable for the processes with small static characteristics between variables, especially for univariate process (Li et al., 2011).

As an alternative, subspace-based approach is also proposed, in which canonical variate analysis (CVA) is the representative (Odiowei and Cao, 2009). CVA extracts the relationship between the past information and the current values of the measurements and its state-space model is identified for process monitoring (Russell et al., 2000). In addition, by projecting the measured data onto reduced subspaces, total projection to latent structures (T-PLS) is proposed for quality-related monitoring (Li et al., 2010).

Above mentioned subspace models own good description of the dynamics of process data, nevertheless, they leave the static characteristics un-modelled, which reflects the constraints between process variables (Li et al., 2014). To cope with this problem, process monitoring methods based on linear dynamic system (LDS) are conducted (Wen et al., 2010). For instance, linear Gaussian state-space model (LGSSM) depicts first-order Markov property by state variables in low dimensional latent space. Compared with the previous researches, the description of process dynamics is more compact. After that, autoregressive dynamic latent variable (ARDLV) model is proposed (Zhou et al., 2017, Zhou et al., 2018). In this model, both dynamic and static

This work was supported by the National Natural Science Foundation of China under Grant 61603342 and 61673349, Zhejiang Provincial Natural Science Foundation of China (LY19F030003). Corresponding Author: Le Zhou

characteristics of process data can be modelled. Moreover, high-order dynamic information of the process can be extracted in the latent space.

These early works based on latent space structure is derived in a probabilistic framework, which can describe process uncertainty (Ge, 2018). And these methods are commonly solved by the maximum likelihood estimation with expectation maximization (EM) algorithm (Rauch et al., 1965), which is, however, susceptible to local optimum or over fitting (Zhu et al., 2016). In view of such situation, probabilistic CCA (Bach and Jordan, 2005) is extended to the form of variational inference (Wang, 2007). By means of variational Bayesian approach (Attias, 1999, Bishop, 2006), the parameter estimation in variational CCA model has been proven more effective, because parameters are regarded as random variables, given prior distributions, and then updated in posterior distributions. Even so there still exits flaws in the variational CCA model as process dynamics are not taken into consideration, which remains a crucial issue.

Hence, in this paper, a novel dynamic Bayesian CCA (DBCCA) model is proposed through combining the consideration of process dynamics with the variational CCA and then utilized for dynamic process monitoring. In the new method, the process dynamics and its static counterpart can be captured simultaneously. As a result of the latent variables in probabilistic generative model, a bond is built up between the current process information and the subsequent data. Intrinsically, since variational inference brings about effects of regularization, the dilemma in traditional maximum likelihood estimation methods is alleviated. Making use of the above mentioned merits, better performance on dynamic process monitoring can be obtained, which is demonstrated by the Tennessee Eastman (TE) benchmark.

The remainder of this paper is organized as follows. In Section 2, deterministic CCA model is revisited. Then, the proposed DBCCA for dynamic process modelling and its corresponding monitoring scheme are introduced in section 3. In the next section the proposed method is verified through the TE benchmark. Finally, conclusions are provided in Section 5.

2. DETERMINISTIC CANONICAL CORRELATION ANALYSIS REVISIT

Canonical correlation analysis (CCA) is a widely-used dimension reduction technique in multivariate statistical analysis. The target of CCA is to maximize the correlation between two sets of data, so as to obtain the corresponding lower-dimensional latent variables. Given two set of N samples datasets $\mathbf{X}\in\mathbb{R}^{M_1\times N}$ and $\mathbf{Y}\in\mathbb{R}^{M_2\times N}$ of M_1 and M_2 variables respectively, CCA seeks the largest correlation coefficients between them, which is for formulated by

$$\begin{aligned}&\max_{\mathbf{u},\mathbf{v}}\ \mathbf{u}^T\mathbf{X}^T\mathbf{Y}\mathbf{v}\\ &s.t.\quad \mathbf{u}^T\mathbf{X}^T\mathbf{X}\mathbf{u}=1\\ &\qquad\ \mathbf{v}^T\mathbf{Y}^T\mathbf{Y}\mathbf{v}=1\end{aligned}\tag{1}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the corresponding projection vectors for $\mathbf{X}$ and $\mathbf{Y}$.

After dimension reduction, the data is projected into the low-dimensional latent space where latent variables $\mathbf{T}=\left[\mathbf{t}_1,\mathbf{t}_2,...,\mathbf{t}_A\right]^T$ for $\mathbf{X}$ are obtained with the number of latent variables A.

Then the data can be constructed in the following form:

$$\begin{cases}\mathbf{X}=\mathbf{U}^T\mathbf{T}+\mathbf{E}_X\\ \mathbf{Y}=\mathbf{V}^T\mathbf{T}+\mathbf{E}_Y\end{cases}\tag{2}$$

where $\mathbf{E}_X$ and $\mathbf{E}_Y$ are the corresponding residuals for each group of data.

3. DYNAMIC BAYESIAN CANONICAL CORRELATION ANALYSIS MODEL AND PROCESS MONITORING APPLICATION

3.1 Dynamic Bayesian Canonical Correlation Analysis

CCA is usually used for quality related process monitoring purpose. In this work, distinctively, CCA is utilized under the probabilistic framework for dynamic process modelling, which is named as dynamic Bayesian canonical correlation analysis (DBCCA). Given observed data $\left[\mathbf{y}_1\quad \mathbf{y}_2\quad \cdots\quad \mathbf{y}_T\right]$, $\mathbf{y}_i\in\mathbb{R}^M$, consequent measurements at time instant t are induced as $\mathbf{p}_t=\mathbf{y}_{t-1}$ and $\mathbf{c}_t=\mathbf{y}_t$, which represents past and current information with respect to time t, respectively. Then, the probabilistic model structure of DBCCA can be expressed by the following formulas:

$$\begin{cases}\mathbf{p}\ =\mathbf{W}\mathbf{z}+\boldsymbol{\mu}+\boldsymbol{\varepsilon}\\ \mathbf{c}\ =\mathbf{H}\mathbf{z}+\boldsymbol{\eta}+\boldsymbol{\delta}\end{cases}\tag{3}$$

where $\mathbf{W}\in\mathbb{R}^{M\times D}$ and $\mathbf{H}\in\mathbb{R}^{M\times D}$ are the loading matrices. The latent variable that represents the essence of the process is characterized by $\mathbf{z}\in\mathbb{R}^D$. The corresponding mean vectors are given as $\boldsymbol{\mu}\in\mathbb{R}^M$ and $\boldsymbol{\eta}\in\mathbb{R}^M$. Process residuals are respectively indicated by $\boldsymbol{\varepsilon}\in\mathbb{R}^M$ and $\boldsymbol{\delta}\in\mathbb{R}^M$.

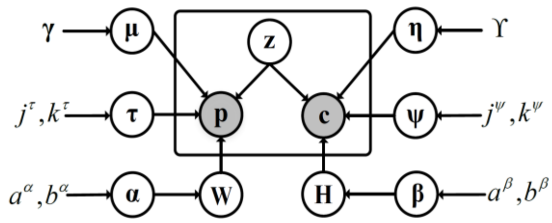

Fig. 1 Probabilistic graphic model representation of DBCCA

In Bayesian framework, these quantities are treated as random variables and updated in the form of a posterior distribution. The prior distributions are defined as follows.

$$P(\mathbf{z})=\prod_{n=1}^{T}\prod_{d=1}^{D}N\left(\mathbf{z}_{dn}\middle|0,1\right)$$

$$P(\boldsymbol{\mu})=\prod_{m=1}^{M}N\left(\boldsymbol{\mu}_m\middle|\mathbf{0},\boldsymbol{\gamma}^{-1}\right)$$

$$P(\boldsymbol{\eta})=\prod_{m=1}^{M}N\left(\boldsymbol{\eta}_m\middle|\mathbf{0},\Upsilon^{-1}\right)$$

$$P(\mathbf{W}|\boldsymbol{\alpha})=\prod_{m=1}^{M}\prod_{d=1}^{D}N\left(\mathbf{W}_{dm}\middle|0,\boldsymbol{\alpha}_d^{-1}\right)$$

$$
\begin{aligned}
P(\boldsymbol{\alpha}) &= \prod_{d=1}^{D} Ga\left(\boldsymbol{\alpha}_d \middle| a_d^{\alpha}, b_d^{\alpha}\right) \\
P(\mathbf{H}|\boldsymbol{\beta}) &= \prod_{m=1}^{M}\prod_{d=1}^{D} N\left(\mathbf{H}_{dm} \middle| 0, \boldsymbol{\beta}_d^{-1}\right) \\
P(\boldsymbol{\beta}) &= \prod_{d=1}^{D} Ga\left(\boldsymbol{\beta}_d \middle| a_d^{\beta}, b_d^{\beta}\right) \\
P(\boldsymbol{\tau}) &= \prod_{m=1}^{M} Ga\left(\boldsymbol{\tau}_m \middle| j_m^{\tau}, k_m^{\tau}\right) \\
P(\boldsymbol{\psi}) &= \prod_{m=1}^{M} Ga\left(\boldsymbol{\psi}_m \middle| j_m^{\psi}, k_m^{\psi}\right)
\end{aligned}
$$

where $N(\cdot)$ stands for Gaussian distribution and $Ga(\cdot)$ represents gamma distribution. In addition, $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are hyper-parameters of the loading matrices and $\boldsymbol{\tau}, \boldsymbol{\psi}$ are the precision matrices with their corresponding parameters given above. For expression convenience, $\boldsymbol{\theta} = \{\mathbf{z}, \mathbf{W}, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\psi}\}$ is defined. In DBCCA, both $\mathbf{p}$ and $\mathbf{c}$ are generated by the essential latent variable $\mathbf{z}$. The probabilistic graph model of DBCCA is shown in Fig. 1.

Given the parameters distributions, the likelihood of observation is derived as:

$$
\begin{cases}
P(\mathbf{p} \; |\mathbf{z}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\tau}) = \prod_{t=1}^{T}\prod_{m=1}^{M} N\left(\boldsymbol{\mu}_m \middle| \mathbf{w}_m^T \mathbf{z}_t + \boldsymbol{\mu}_m, \boldsymbol{\tau}_m^{-1}\right) \\
P(\mathbf{c} \; |\mathbf{z}, \mathbf{H}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\psi}) = \prod_{t=1}^{T}\prod_{m=1}^{M} N\left(\boldsymbol{\eta}_m \middle| \mathbf{h}_m^T \mathbf{z}_t + \boldsymbol{\eta}_m, \boldsymbol{\psi}_m^{-1}\right)
\end{cases} \tag{4}
$$

The joint probability density function is consequently given as:

$$
\begin{aligned}
P(\mathbf{p}, \mathbf{c}, \boldsymbol{\theta}) &= p(\mathbf{p}, \mathbf{c}, \mathbf{z}, \mathbf{W}, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\psi}) \\
&= P(\mathbf{p}|\mathbf{z}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\tau}) P(\mathbf{c}|\mathbf{z}, \mathbf{H}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\psi}) \times \\
&\quad P(\mathbf{W}|\boldsymbol{\alpha}) P(\mathbf{H}|\boldsymbol{\beta}) P(\mathbf{z}) P(\boldsymbol{\mu}) P(\boldsymbol{\eta}) P(\boldsymbol{\tau}) P(\boldsymbol{\psi})
\end{aligned} \tag{5}
$$

Then the posterior distribution can be obtained in the following form:

$$
P(\boldsymbol{\theta}|\mathbf{p}, \mathbf{c}) = P(\mathbf{p}, \mathbf{c}, \boldsymbol{\theta}) / P(\mathbf{p}, \mathbf{c}) \tag{6}
$$

However, such posterior distribution is intractable in complex situations and thus cannot be calculated analytically. Based on variational Bayesian approach, a restricted trial distribution

$$
Q(\boldsymbol{\theta}) = Q(\mathbf{z}) Q(\mathbf{W}) Q(\boldsymbol{\alpha}) Q(\mathbf{H}) Q(\boldsymbol{\beta}) Q(\boldsymbol{\mu}) Q(\boldsymbol{\eta}) Q(\boldsymbol{\tau}) Q(\boldsymbol{\psi}) \tag{7}
$$

is introduced to simplify and convert this problem. As the following equation holds:

$$
\int \log P(\mathbf{p}, \mathbf{c}) \, d\boldsymbol{\theta} = \underbrace{\int Q(\boldsymbol{\theta}) \log \frac{P(\mathbf{p}, \mathbf{c}, \boldsymbol{\theta})}{Q(\boldsymbol{\theta})} d\boldsymbol{\theta}}_{ELBO(Q)} - \underbrace{\int Q(\boldsymbol{\theta}) \log \frac{P(\boldsymbol{\theta}|\mathbf{p}, \mathbf{c})}{Q(\boldsymbol{\theta})} d\boldsymbol{\theta}}_{KL(Q\|P)} \tag{8}
$$

The reason is that the left side term of the equation is a constant, and that the right side term is composed of evidence lower bound (ELBO) and KL divergence of $Q(\boldsymbol{\theta})$ to $P(\boldsymbol{\theta}|\mathbf{p}, \mathbf{c})$, which is noted in the formula. It's noticed that once ELBO is maximized after iterations, the KL divergence is approximately 0, making $Q(\boldsymbol{\theta})$ the closest to the $P(\boldsymbol{\theta}|\mathbf{p}, \mathbf{c})$ of interest. In this way, the intractable posterior distribution is indirectly transformed.

In the next procedure, substitute $Q(\boldsymbol{\theta})$ into the ELBO:

$$
\begin{aligned}
ELBO(Q) &= \int Q(\mathbf{z}) Q(\mathbf{W}) Q(\boldsymbol{\alpha}) Q(\mathbf{H}) Q(\boldsymbol{\beta}) Q(\boldsymbol{\mu}) Q(\boldsymbol{\eta}) Q(\boldsymbol{\tau}) \times \\
& Q(\boldsymbol{\psi}) \log \frac{P(\mathbf{p}, \mathbf{c}, \mathbf{z}, \mathbf{W}, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\psi})}{Q(\mathbf{z}) Q(\mathbf{W}) Q(\boldsymbol{\alpha}) Q(\mathbf{H}) Q(\boldsymbol{\beta}) Q(\boldsymbol{\mu}) Q(\boldsymbol{\eta}) Q(\boldsymbol{\tau}) Q(\boldsymbol{\psi})} d\boldsymbol{\theta}
\end{aligned} \tag{9}
$$

where the ELBO maximizes each parameter distribution in an iterative manner, which is known as the VBEM method. According to the theory of variational inference, the ELBO will reach convergence through the following update of distribution Q given in a general form:

$$
\log Q_i(\boldsymbol{\theta}_i) = \frac{\mathbf{E}\left[P(\mathbf{p}, \mathbf{c}, \boldsymbol{\theta})\right]_{k \neq i}}{\int \mathbf{E}\left[P(\mathbf{p}, \mathbf{c}, \boldsymbol{\theta})\right]_{k \neq j} d\boldsymbol{\theta}_j} \tag{10}
$$

where $\mathbf{E}_{k \neq i}[\cdot]$ stands for expectation with respect to the distribution $Q_k(\boldsymbol{\theta}_k)$ for all $k \neq i$. Under this pattern, the posterior distributions can be calculated alternatively and the intractable problem is solved.

The specific updating procedure is as follows. According to (10), the posterior distribution of latent variable $\mathbf{z}$ is given by

$$
Q(\mathbf{z}) = \prod_{t=1}^{T} N\left(\mathbf{z}_t \middle| \boldsymbol{\lambda}_t^{\mathbf{z}}, \boldsymbol{\upsilon}_{\mathbf{z}}^{-1}\right) \tag{11}
$$

and its parameters are

$$
\begin{cases}
\boldsymbol{\lambda}_t^{\mathbf{z}} = \boldsymbol{\upsilon}_{\mathbf{z}} \mathrm{E}_{Q(\mathbf{z})}\left[\sum_{m=1}^{M} \boldsymbol{\tau}_m \mathbf{W}_m (\mathbf{p}_{mt} - \boldsymbol{\mu}_m) + \sum_{m=1}^{M} \boldsymbol{\psi}_m \mathbf{H}_m (\mathbf{c}_{mt} - \boldsymbol{\eta}_m)\right] \\
\boldsymbol{\upsilon}_{\mathbf{z}}^{-1} = \mathbf{I} + \mathrm{E}_{Q(\mathbf{z})}\left[\sum_{m=1}^{M} \boldsymbol{\tau}_m \mathbf{W}_m \mathbf{W}_m^{\mathrm{T}} + \sum_{m=1}^{M} \boldsymbol{\psi}_m \mathbf{H}_m \mathbf{H}_m^{\mathrm{T}}\right]
\end{cases}
$$

Then the loading matrices, in a similar way, can be updated using the following formulas:

$$
Q(\mathbf{W}|\boldsymbol{\alpha}) = \prod_{t=1}^{T}\prod_{m=1}^{M} N\left(\mathbf{W}_{mt} \middle| \boldsymbol{\lambda}_{mt}^{\mathbf{W}}, \boldsymbol{\upsilon}_{\mathbf{W}}^{-1}\right) \tag{12}
$$

$$
Q(\mathbf{H}|\boldsymbol{\beta}) = \prod_{t=1}^{T}\prod_{m=1}^{M} N\left(\mathbf{H}_{mt} \middle| \boldsymbol{\lambda}_{mt}^{\mathbf{H}}, \boldsymbol{\upsilon}_{\mathbf{H}}^{-1}\right) \tag{13}
$$

And their parameters are given as

$$
\begin{cases}
\boldsymbol{\lambda}_{mt}^{\mathbf{W}} = \boldsymbol{\upsilon}_{\mathbf{W}} \boldsymbol{\tau}_m \mathrm{E}_{Q(\mathbf{W})}\left[\sum_{t=1}^{T} \mathbf{z}_t (\mathbf{p}_{mt} - \boldsymbol{\mu}_m)\right] \\
\boldsymbol{\upsilon}_{\mathbf{W}}^{-1} = \langle diag \; \boldsymbol{\alpha} \rangle + \boldsymbol{\tau}_m \mathrm{E}_{Q(\mathbf{W})}\left[\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t^{\mathrm{T}}\right]
\end{cases}
$$

$$
\begin{cases}
\boldsymbol{\lambda}_{mt}^{\mathbf{H}} = \boldsymbol{\upsilon}_{\mathbf{H}} \boldsymbol{\psi}_m \mathrm{E}_{Q(\mathbf{H})}\left[\sum_{t=1}^{T} \mathbf{z}_t (\mathbf{c}_{mt} - \boldsymbol{\eta}_m)\right] \\
\boldsymbol{\upsilon}_{\mathbf{H}}^{-1} = \langle diag \; \boldsymbol{\beta} \rangle + \boldsymbol{\psi}_m \mathrm{E}_{Q(\mathbf{H})}\left[\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t^{\mathrm{T}}\right]
\end{cases}
$$

The hyper-parameters of the loading matrices are given prior gamma distributions and then updated in the same distributions which is given by the following posterior distributions:

$$
Q(\boldsymbol{\alpha}) = \prod_{d=1}^{D} Ga\left(\boldsymbol{\alpha}_d \middle| \lambda_d^{\alpha}, \nu_d^{\alpha}\right) \tag{14}
$$

$$Q(\boldsymbol{\beta}) = \prod_{d=1}^{D} Ga\left(\boldsymbol{\beta}_d \middle| \lambda_d^{\beta}, \nu_d^{\beta}\right) \quad (15)$$

with the parameters listed as

$$\begin{cases} \lambda^{\alpha} = a^{\alpha} + \frac{1}{2}M \\ \nu^{\alpha} = b^{\alpha} + \frac{1}{2}\mathrm{E}_{Q(\boldsymbol{\alpha})}\left[\sum_{m=1}^{M}\mathbf{W}_m^{\mathrm{T}}\mathbf{W}_m\right] \end{cases}$$

$$\begin{cases} \lambda^{\beta} = a^{\beta} + \frac{1}{2}M \\ \nu^{\beta} = b^{\beta} + \frac{1}{2}\mathrm{E}_{Q(\boldsymbol{\beta})}\left[\sum_{m=1}^{M}\mathbf{H}_m^{\mathrm{T}}\mathbf{H}_m\right] \end{cases}$$

The next step is to perform the posterior estimation of mean vectors, which are updated in the following manner respectively:

$$Q(\boldsymbol{\mu}) = \prod_{m=1}^{M} N\left(\boldsymbol{\mu}_m \middle| \boldsymbol{\lambda}^{\mu}, \boldsymbol{\upsilon}_{\mu}^{-1}\right) \quad (16)$$

$$Q(\boldsymbol{\eta}) = \prod_{m=1}^{M} N\left(\boldsymbol{\eta}_m \middle| \boldsymbol{\lambda}^{\eta}, \boldsymbol{\upsilon}_{\eta}^{-1}\right) \quad (17)$$

and the parameters of mean vectors are

$$\begin{cases} \boldsymbol{\lambda}_m^{\mu} = \boldsymbol{\upsilon}_{\mu}\boldsymbol{\tau}_m\mathrm{E}_{Q(\boldsymbol{\mu})}\left[\sum_{t=1}^{T}\left(\mathbf{p}_{mt} - \mathbf{W}_m\mathbf{z}_t\right)\right] \\ \boldsymbol{\upsilon}_{\mu}^{-1} = \boldsymbol{\gamma} + \mathrm{E}_{Q(\boldsymbol{\mu})}\left[\sum_{t=1}^{T}\boldsymbol{\tau}_m\right] \end{cases}$$

$$\begin{cases} \boldsymbol{\lambda}_m^{\eta} = \boldsymbol{\upsilon}_{\eta}\boldsymbol{\tau}_m\mathrm{E}_{Q(\boldsymbol{\eta})}\left[\sum_{t=1}^{T}\left(\mathbf{c}_{mt} - \mathbf{H}_m\mathbf{z}_t\right)\right] \\ \boldsymbol{\upsilon}_{\eta}^{-1} = \Upsilon + \mathrm{E}_{Q(\boldsymbol{\eta})}\left[\sum_{t=1}^{T}\boldsymbol{\psi}_m\right] \end{cases}$$

For the precision matrices, which are again given gamma priors and therefore can be calculated in the following form:

$$Q(\boldsymbol{\tau}) = \prod_{m=1}^{M(l+1)} Ga\left(\boldsymbol{\tau}_m \middle| \lambda^{\tau}, \nu_{\tau}\right) \quad (18)$$

$$Q(\boldsymbol{\psi}) = \prod_{m=1}^{Ms} Ga\left(\boldsymbol{\psi}_m \middle| \lambda^{\psi}, \nu_{\psi}\right) \quad (19)$$

with the parameters of gamma distribution as

$$\begin{cases} \lambda^{\tau} = j^{\tau} + \frac{1}{2}T \\ \nu_{\tau} = k^{\tau} + \frac{1}{2}\mathrm{E}_{Q(\boldsymbol{\tau})}\left[\sum_{t=1}^{T}\left(\mathbf{p}_{mt} - \mathbf{W}_m\mathbf{z}_t - \boldsymbol{\mu}_m\right)^2\right] \end{cases}$$

$$\begin{cases} \lambda^{\psi} = j^{\psi} + \frac{1}{2}T \\ \nu_{\psi} = k^{\psi} + \frac{1}{2}\mathrm{E}_{Q(\boldsymbol{\psi})}\left[\sum_{t=1}^{T}\left(\mathbf{c}_{mt} - \mathbf{H}_m\mathbf{z}_t - \boldsymbol{\eta}_m\right)^2\right] \end{cases}$$

In this way, the required posterior distributions are updated iteratively until the ELBO reaches the convergence, after which the training procedure of DBCCA has been completed.

3.2 Dynamic Bayesian Canonical Correlation Analysis for Online Monitoring

In traditional process monitoring methods, T^2 and SPE statistics are calculated to detect process anomalies. Similarly, DBCCA also utilizes these two statistics, but with slight difference, as T^2 contains the comprehensive information of $\mathbf{p}$ and $\mathbf{c}$ while SPE only focus on the present data. Specifically, the monitoring statistics are given as follows:

$$T^2 = \left(\boldsymbol{\lambda}^{z}\right)^T \boldsymbol{\lambda}^{z} \quad (20)$$

For residual monitoring, the SPE statistics can be formulated as:

$$SPE = \boldsymbol{\varepsilon}^T\boldsymbol{\tau}\boldsymbol{\varepsilon} \quad (21)$$

where the residual of present information is calculated by:

$$\boldsymbol{\varepsilon} = \mathbf{p} - \mathbf{W}\mathbf{z} - \boldsymbol{\mu} \quad (22)$$

For online monitoring stage, newly coming data samples $\mathbf{p}_{new}$ and $\mathbf{c}_{new}$ can be given in the same way as that during the training stage. The statistics for online monitoring is derived as:

$$T_{new}^2 = \mathbf{z}_{new}^T\boldsymbol{\upsilon}_{\mathbf{z}}\mathbf{z}_{new} \quad (23)$$

using the testing latent variable

$$\mathbf{z}_{new} = \boldsymbol{\upsilon}_{\mathbf{z}}^{-1}\left(\mathbf{z}_{new}^{\mathbf{p}} + \mathbf{z}_{new}^{\mathbf{c}}\right) \quad (24)$$

where

$$\begin{cases} \mathbf{z}_{new}^{\mathbf{p}} = \mathbf{W}^T\boldsymbol{\tau}\left(\mathbf{p}_{new} - \boldsymbol{\mu}\right) \\ \mathbf{z}_{new}^{\mathbf{c}} = \mathbf{H}^T\boldsymbol{\psi}\left(\mathbf{c}_{new} - \boldsymbol{\eta}\right) \end{cases} \quad (25)$$

The testing residual statistics can also be obtained by:

$$SPE_{new} = \boldsymbol{\varepsilon}_{new}^T\boldsymbol{\tau}\boldsymbol{\varepsilon}_{new} \quad (26)$$

with

$$\boldsymbol{\varepsilon}_{new} = \mathbf{p}_{new} - \mathbf{W}\mathbf{z}_{new}^{\mathbf{p}} - \boldsymbol{\mu} \quad (27)$$

Considering the above monitoring statistics, the required control limits can be calculated by the following formulas:

$$T_{new}^2 \le T_{\lim}^2 = \chi_{\xi}^2(D) \quad (28)$$

$$SPE_{new} \le SPE_{\lim} = \chi_{\xi}^2(M) \quad (29)$$

where ξ is the significance level, D is the number of latent variable, and M is the number of process variables. Under this scheme, the dynamic process can be monitored in real time.

4. CASE STUDY ON TENNESSEE EASTMAN PROCESS

In this part, TE benchmark process is used to verify the effectiveness of the proposed DBCCA model and its feasibility for fault detection. The TE process is a real industrial process simulation platform, which is widely utilized to evaluate the performance of process monitoring methods. It consists of 5 process units, including 12 manipulated variables and 41 measurement variables, and detailed description of the process can be found in the original paper (Downs et al., 1993). The data under normal operation condition is used as training data, while the testing counterparts are obtained from 21 abnormal conditions of the process. When the monitoring statistics of the testing data exceed the corresponding control limit for several consecutive samples, it is considered that the process abnormally is successfully detected.

Table 1 Fault Detection Rate of TE Benchmark (%)

	DPCA		LDS		ARDLV		DBCCA	
IDV	T2	SPE	T2	SPE	T2	SPE	T2	SPE
1	99.25	99.63	99.75	99.88	99.94	100	99.87	99.75
2	98.75	98.50	98.63	98.38	98.63	98.75	98.62	97.75
3	0.50	2.88	3.75	3.13	8.50	6.00	0.50	3.25
4	2.50	100	1.13	90.75	10.31	100	6.38	100
5	23.25	28.88	99.88	100	100	100	100	100
6	98.88	100	100	100	100	100	100	100
7	39.50	100	46.88	100	42.38	100	41.93	100
8	96.63	97.63	97.75	98.00	98.50	98.31	98.12	93.12
9	1.38	2.38	3.00	1.75	8.06	5.06	0.63	4.01
10	31.38	26.00	88.38	90.38	92.63	90.94	89.61	85.23
11	16.25	77.00	6.50	68.38	39.38	82.56	42.93	83.35
12	98.25	97.63	99.75	99.75	100	99.88	100	99.87
13	94.00	95.38	95.50	95.13	96.38	95.50	94.87	95.49
14	89.50	100	1.25	100	99.44	100	100	99.87
15	1.50	2.63	7.88	2.38	43.69	22.88	5.38	6.63
16	12.75	21.75	81.13	91.00	91.56	92.75	91.36	82.60
17	73.88	93.63	68.13	95.13	90.50	97.13	90.36	96.75
18	88.50	90.25	89.63	90.63	91.50	91.19	90.61	90.49
19	0.75	25.88	3.50	76.38	59.94	94.94	53.69	83.98
20	27.25	49.88	80.25	87.13	73.38	91.06	87.36	81.48
21	32.75	44.13	60.00	40.13	37.25	45.19	45.68	38.67

In this work, process monitoring methods based on DPCA, LGSSM, and ARDLV are used to compare with the proposed DBCCA model. The fault detection results of these methods are listed in the Table 1, which is evaluated by fault detection rate (FDR). The online monitoring results are carried out with confidence level of 99%. It can be observed that the performance of ARDLV and DBCCA significantly outperform the other methods. It's worth mentioning that although the detection rate of ARDLV on several faults is higher than that of DBCCA, it is at the cost of higher false alarm rate, which are not given here due to page limitation.

For some faults, take IDV 11 for instance, DBCCA has advantages over alternative methods both in false alarm and detection rate, which is displayed by Fig. 2, 3, 4, and 5 of online monitoring results of IDV 11 based on DPCA, LGSSM, ARDLV, and DBCCA respectively. The reason is that the proposed method owns good description of process dynamics, and static structural characteristics that are not considered in the subspace-based methods are compensated by CCA modelling. Furthermore, the introduction of variational inference on such basic model brings the effects of regularization and alleviates the problem of local optimum, which accounts for the satisfying results.

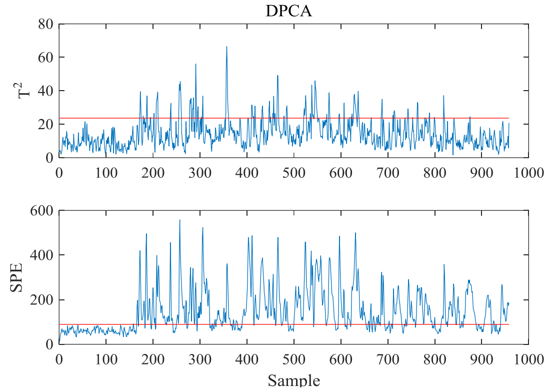

Fig. 2. Online monitoring of TE IDV11 based on DPCA

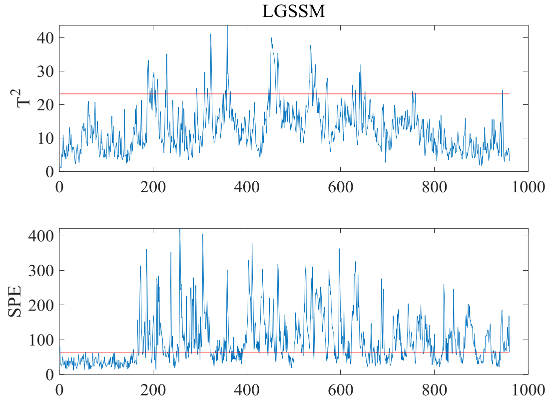

Fig. 3. Online monitoring of TE IDV11 based on LGSSM

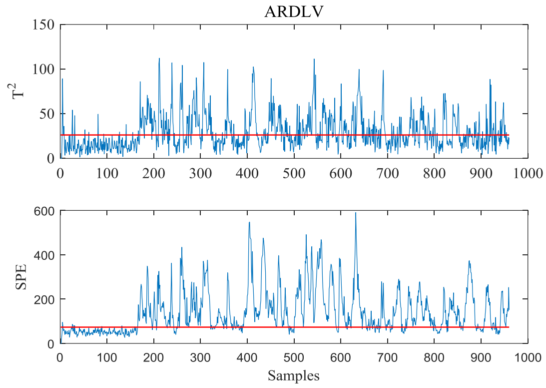

Fig. 4. Online monitoring of TE IDV11 based on ARDLV

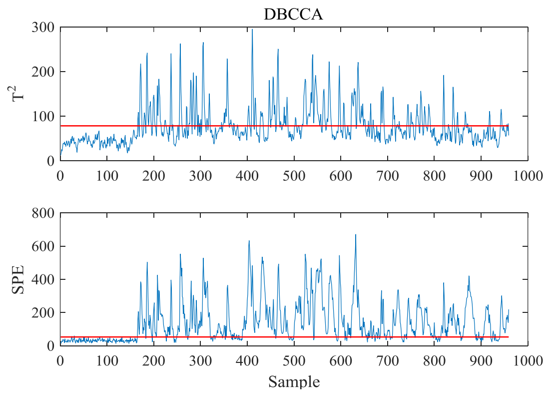

Fig. 5. Online monitoring of TE IDV11 based on DBCCA

5. CONCLUSIONS

Dynamic process monitoring is of great significance as it is related to process safety. However, methods considering process dynamics ignore the static structural characteristics. Although both characteristics are taken into account by probabilistic methods, they still face the problems of over fitting or local optimal in maximum likelihood estimation. In this work, a novel DBCCA model is proposed for dynamic process modelling and utilized for fault detection. Compared with former process monitoring methods, the proposed method not only simultaneously considers the dynamic and static characteristics of the process, but also alleviates the dilemma in traditional latent variable models based on maximum likelihood estimation. Through the case study on TE benchmark, it is proven that the proposed DBCCA model is superior to alternatives both in dynamic process modelling and fault detection task.

With the advent of intelligent manufacturing, massive amount of data can be collected and requirement on process safety surges. On one hand, since the proposed model is trained by approach of coordinate ascending (Bishop, 2006), the computational efficiency is limited. As for the future work, the current training mode maybe modified by stochastic gradient descent, which can be appropriately extended to the distributed parallel pattern (Zhang and Ge, 2019) for further training efficiency. On the other hand, subsequent process monitoring procedure, i.e. fault diagnosis ought to be conducted in time.

REFERENCES

Attias, H. (1999). Inferring parameters and structure of latent variable models by variational Bayes. *In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (pp. 21-30). Morgan Kaufmann Publishers Inc.

Bach, F. R., & Jordan, M. I. (2005). A probabilistic interpretation of canonical correlation analysis.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer New York.

Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & chemical engineering*, 17(3), 245-255.

Ge, Z., Song, Z., & Gao, F. (2013). Review of recent research on data-based process monitoring. *Industrial & Engineering Chemistry Research*, 52(10), 3543-3562.

Ge, Z. (2018). Process data analytics via probabilistic latent variable models: A tutorial review. *Industrial & Engineering Chemistry Research*, 57(38), 12646-12661.

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

Lee, J. M., Yoo, C., & Lee, I. B. (2004). Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical engineering science*, 59(14), 2995-3006.

Li, G., Joe Qin, S., & Zhou, D. (2010). Output relevant fault reconstruction and fault subspace extraction in total projection to latent structures models. *Industrial & Engineering Chemistry Research*, 49(19), 9175-9183.

Li, G., Liu, B., Qin, S. J., & Zhou, D. (2011). Dynamic latent variable modeling for statistical process monitoring. *IFAC Proceedings Volumes*, 44(1), 12886-12891.

Li, G., Qin, S. J., & Zhou, D. (2014). A new method of dynamic latent-variable modeling for process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), 6438-6445.

Liu, H., Yang, C., Carlsson, B., Qin, S. J., & Yoo, C. (2019). Dynamic Nonlinear Partial Least Squares Modeling Using Gaussian Process Regression. *Industrial & Engineering Chemistry Research*, 58(36), 16676-16686.

Negiz, A., Lagergren, E. S., & Cinar, A. (1994). Statistical quality control of multivariable continuous processes. *In Proceedings of 1994 American Control Conference-ACC'94* (Vol. 2, pp. 1289-1293). IEEE.

Odiowei, P. E. P., & Cao, Y. (2009). Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE Transactions on Industrial Informatics*, 6(1), 36-45.

Qin, S. J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual reviews in control*, 36(2), 220-234.

Rauch, H. E., Tung, F., & Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8), 1445-1450.

Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 51(1), 81-93.

Wang, C. (2007). Variational Bayesian approach to canonical correlation analysis. *IEEE Transactions on Neural Networks*, 18(3), 905-910.

Wen, Q., Ge, Z., & Song, Z. (2012). Data-based linear Gaussian state-space model for dynamic process monitoring. *AIChE journal*, 58(12), 3763-3776.

Yin, S., DING, S. X., Haghani, A., Hao, H., & Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of process control*, 22(9), 1567-1581.

Zhang, X., & Ge, Z. (2019). Automatic Deep Extraction of Robust Dynamic Features for Industrial Big Data Modeling and Soft Sensor application. *IEEE Transactions on Industrial Informatics*.

Zhou, L., Li, G., Song, Z., & Qin, S. J. (2017). Autoregressive dynamic latent variable models for process monitoring. *IEEE Transactions on Control Systems Technology*, 25(1), 366-373.

Zhou, L., Zheng, J., Ge, Z., Song, Z., & Shan, S. (2018). Multimode process monitoring based on switching autoregressive dynamic latent variable model. *IEEE Transactions on Industrial Electronics*, 65(10), 8184-8194.

Zhu, J., Ge, Z., & Song, Z. (2016). Bayesian robust linear dynamic system approach for dynamic process monitoring. *Journal of Process Control*, 40, 62-77.