

# Process Monitoring with Sparse Bayesian Model for Industrial Methanol Distillation

Lin Luo\* Lei Xie\*\* Hongye Su\*\* Jiusun Zeng\*\*\*

\* School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, Liaoning, PR China (e-mail: luolin@lnpu.edu.cn).

\*\* State Key Laboratory of Industrial Control Technology, Institute of Cyber-systems and Control, Zhejiang University, Hangzhou 310027, Zhejiang, PR China (e-mail: leix@iipc.zju.edu.cn)

\*\*\* College of Metrology & Measurement Engineering, China Jiliang University, Hangzhou 310018, Zhejiang, PR China.

---

**Abstract:** Following the intuition that not all latent variables in probabilistic principal component analysis method shifts simultaneously, this paper proposes a spike-and-slab regularization technique for nonlinear fault detection and isolation. Different from the existing probabilistic latent variable models, a spike-and-slab prior is introduced to downweight the irrelevant information of latent variables for the discriminative model. The resulting latent subspace supported by regularization parameters is not only sensitive to the informative variables, but it also eliminates the influence of the non-informative ones. The feasibility and efficiency of the proposed approach will be tested on an industrial methanol distillation dataset. Moreover, the performance will be compared with conventional probabilistic latent variables methods.

*Keywords:* Process monitoring, fault detection and isolation, Bayesian variable selection, spike-and-slab prior, industrial methanol distillation.

---

## 1. INTRODUCTION

Timely detection of faults and subsequent isolation of fault mode are becoming an indispensable component in fulfilling safety requirements, process stability and product quality demands. Technological development of machining equipments makes industrial records from numerous measured process variables available, calling for the development of multivariate techniques (Peres and Fogliatto (2018); Woodall and Montgomery (2014); Wang and Jiang (2009)). Multivariate statistical process control (MSPC) methods are commonly used to determine whether at least one property has deviated from the acceptable operating conditions and locate root causes of the detected shift.

Various multivariate techniques based on latent variable (LV) methods have been used in process monitoring, including principal component analysis (PCA), partial least squares (PLS), independent component analysis (ICA) and canonical correlation analysis (CCA) among others (Li et al. (2010); Lee et al. (2011); Chen et al. (2016); Qin (2012)). In general, a latent space in the LV model is explored to reveal the underlying low-dimensional structure of original measured variables and its complementary residual space aims at locating noises and outliers. Once the model has been determined, the MSPC control charts corresponding to the two aforementioned spaces

are required. Such charts are referred to as  $T^2$  and SPE control charts, respectively. After an out-of-control alarm is triggered, the contribution charts are used to determine which variables are deviating from their acceptable range. However, it is very common that process variables are contaminated by uncertainties, such as missing values and random noises, which could be non-Gaussian and characterized by outliers. As a result, traditional LV-based MSPC techniques are not effective and accurate anymore for such cases. In order to minimize the influence of such uncertainties, the probabilistic counterparts of LV models is usually necessary.

Recently, the probabilistic derivation of LV models has verified to be powerful tools for effectively addressing the uncertain effects, such as probabilistic PCA (Tipping and Bishop (1999))-, probabilistic ICA (Zhu et al. (2017))-, probabilistic CCA (Liu et al. (2018))-based methods. The probabilistic counterparts of LV models in their original form have no implementation of variable selection due to the extensive work involved in the construction of the latent space. In order to gain insight into high-dimensional processes, the information collected from the most of the relevant LVs are of utmost importance. However, the downside of the probabilistic LV model is the risk of including irrelevant variables in the models since multiple streams of variables is treated equally (Peres and Fogliatto (2018); Wang and Jiang (2009); Mehmood et al. (2012)). In such scenarios, the integration of variable selection methods to the probabilistic LV models are of significant interest.

---

\* This work is supported by National Natural Science Foundation of China under Grants 61703191, the Foundation of Liaoning Educational Committee under Grants L2017LQN028, the Scientific Research Foundation of Liaoning Shihua University under Grants 2017XJJ-012.

## 2. PRELIMINARIES

Research on variable selection in the probabilistic LV-based process monitoring can be divided into two categories (Ghosh et al. (2014)): filter-based and wrapper-based methods. Feature ranking in filter-based methods are generally independent of subsequent model or decision support, utilizing the intrinsic properties of the data. Since they have no implementation of the additional model learning and evaluation, filter-based approaches are more computationally efficient for the development of process monitoring (Srinivasan and Qian (2007); Ge et al. (2011)). Wrapper-based methods allow simultaneous feature selection and model construction to produce an appropriate subset of the latent variables for process monitoring. The most frequently used methods are k-fold cross validation and Bayesian regularization (Ge and Song (2010); Zeng et al. (2015); Kodamana et al. (2019)). Although monitoring performance has been improved using the former technique, the unbounded nature of the log-likelihood function makes maximum-likelihood estimator prone to favoring models of ever increasing complexity. Inspired by the potential of model combination, Lee et al. (2010) exploited a spike-and-slab prior to implement the inclusion and the exclusion of variables in the Bayesian method. The process of Bayesian regularization-based variable selection is inducing an indicator coefficients under the sparsity assumption. A sparse representation is desirable in situations where: (1) prior information is expected to be relevant or irrelevant for decision support, (2) only a subset of latent variables explains the data.

In this paper, a novel process monitoring technique based on a sparse Bayesian model is proposed. The model uses spike-and-slab distribution as a prior over the latent variables and resolve the problem of the LVs selection in probability PCA, which we called SS-PPCA. In a spike-and-slab distribution, a binary indicator function selects either a point-mass at exactly zero (the 'spike') or a mixture of a broad Gaussian distribution (the 'slab'). The posterior distribution of these indicator parameters are utilized to select the more sensitive variables to faults. An expectation-maximization (EM) algorithm is presented for the maximum likelihood estimation (MLE) of the model parameters. In the proposed spike-and-slab regularization diagram, faults are usually detected by the resulting latent space, which is supported by the indicator variable. With the optimal subset of LVs, the corresponding statistics are sensitive to the most informative variables and more accurate to numerous local process behaviors. Moreover, a modified contribution plot is capable of an effective diagnostic tool due to the sparsity property of SS-PPCA.

The remainder of the paper is organized as follows. In Section 2, a brief review of probabilistic PCA-based process monitoring method is presented. Section 3 details the proposed SS-PPCA approach, in which model selection and novel monitoring statistics associated with SS-PPCA are proposed. In addition, model parameters are directly inferred by EM method. In Section 4, an industrial methanol distillation process is performed to illustrate the rationality and superiority of the proposed approach in contrast with traditional PPCA scheme. Finally, concluding remarks are drawn in Section 5.

### 2.1 Probabilistic principal component analysis

The PPCA is a generative probabilistic model for the measurement  $\mathbf{X} = \{\mathbf{x}_n | \mathbf{x}_n \in \mathbb{R}^D, n = 1, \dots, N\}$  where  $N$  is the number of observation and  $D$  is the dimension of the measured variables. The formulation of PPCA model is given by (Tipping and Bishop (1999))

$$\mathbf{x}_n = \mathbf{P}\mathbf{t}_n + \mathbf{e}_n \quad (1)$$

where  $\mathbf{P} \in \mathbb{R}^{D \times M}$  is the corresponding loading matrix;  $\mathbf{t}_n \in \mathbb{R}^M$  is the latent variable which is assumed to follow a Gaussian distribution;  $M$  is the number of latent factors;  $\mathbf{e}_n \in \mathbb{R}^D$  is the noise of the process which is also defined as a  $D$ -dimensional Gaussian distribution, that is,  $\mathbf{e}_n \sim \mathcal{N}(\mathbf{0}, \tau_D \mathbf{I})$ .

A common choice for a prior distribution over the latent variable  $\mathbf{t}_n$  is to use a unit covariance, zero mean Gaussian distribution,

$$\mathbf{t}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (2)$$

where  $\mathbf{I}$  is the identity matrix. In order to derive the likelihood function, the distribution  $p(\mathbf{X})$  can be obtained by the marginalization over a continuous latent space  $\mathbf{t}_n$ ,

$$p(\mathbf{X} | \mathbf{P}, \tau_D \mathbf{I}) = \int p(\mathbf{X} | \mathbf{t}_n, \mathbf{P}, \tau_D \mathbf{I}) p(\mathbf{t}_n) d\mathbf{t}_n \quad (3)$$

Given a dataset  $\mathbf{X} = \{\mathbf{x}_n\}$ , the corresponding log-likelihood function of complete data can be then formulated as

$$L = \ln \prod_{n=1}^N p(\mathbf{x}_n | \mathbf{P}, \tau_D \mathbf{I}). \quad (4)$$

The model parameters  $\Theta = \{\mathbf{P}, \tau_D \mathbf{I}\}$  can be determined by maximizing the log-likelihood function. The well-known expectation - maximization (EM) algorithm can be employed to accomplish such task.

### 2.2 Monitoring method

Let  $\mathbf{x}_{\text{new}}$  be the vector of a new process sample,  $T^2$  statistic associated with  $\mathbf{x}_{\text{new}}$  is defined as

$$T_{\text{new}}^2 = \mathbf{t}_{\text{new}}^T (\text{var}(\mathbf{t}_{\text{new}}))^{-1} \mathbf{t}_{\text{new}}, \quad (5)$$

where  $\mathbf{t}_{\text{new}}$  is the posterior mean depends on  $\mathbf{x}_{\text{new}}$ , which may be defined as,

$$\mathbf{t}_{\text{new}} = \mathbf{P}^T (\mathbf{P}\mathbf{P}^T + \tau_D \mathbf{I})^{-1} \mathbf{x}_{\text{new}}. \quad (6)$$

Similarly, SPE statistic associated with  $\mathbf{x}_{\text{new}}$  is defined as

$$SPE_{\text{new}} = \mathbf{e}_{\text{new}}^T (\tau_D \mathbf{I})^{-1} \mathbf{e}_{\text{new}} \quad (7)$$

where  $\mathbf{e}_{\text{new}} = \mathbf{x}_{\text{new}} - \mathbf{P}\mathbf{t}_{\text{new}}$  is the reconstruction error of the new data. In this paper, the upper control limits for  $T^2$  and SPE statistics are determined by performing kernel density estimation (KDE) on the corresponding statistics associated with training samples.

It should be noted that the number of latent variables  $M$  is an unknown prior and seems natural to seek a Bayesian approach to model selection. A simple way is to keep the features that have the largest log-likelihood on a validation data set. However, it is difficult to determine the value of dimensionality particularly if samples are not abundant. A more refined approach is to project the data onto a sparse principal subspace while retaining as much variance

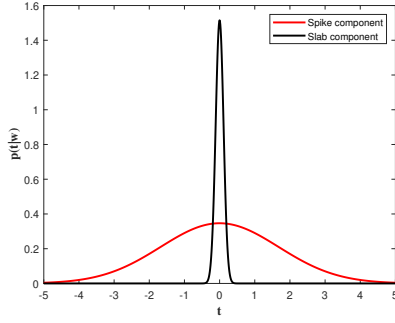


Fig. 1. Probability density of spike-and-slab priors. Spike-and-slab priors are a mixture of a broad Gaussian distribution (the slab) and a point mass (the spike).

as possible. The sparse representation is desirable in the situation where only a subset of underlying variables are actually relevant to explain the data. It means that the latent variables whose additional indicators include exact zeroes do not contribute to the decisions made by the model.

### 3. ONLINE MONITORING AND FAULT DIAGNOSIS

#### 3.1 Sparse subspace selection with spike-and-slab prior

To obtain a principal subspace with strong sparsity, a spike-and-slab prior has to be placed on the latent variables. The basic spike-and-slab prior is a mixture of a broad Gaussian distribution known as the 'slab' and a point mass at zero referred to as the 'spike'. Formally, the spike and slab prior has the form,

$$p(\mathbf{t}_n|\mathbf{w}) = \mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_M \mathbf{I})^{\mathbf{w}} \mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_{\mathbf{w}} \mathbf{I})^{1-\mathbf{w}} \quad (8)$$

where  $\mathbf{w} \in \{0, 1\}^M$  is a binary indicator variable,  $\mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_M \mathbf{I})$  is the slab component,  $\mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_{\mathbf{w}} \mathbf{I})$  is the spike component, and  $\tau_M \gg \tau_{\mathbf{w}}$  is the coefficients associated with the relevant variables. To complete the specification of the prior for  $\mathbf{t}_n$ , the indicator variable  $\mathbf{w}$  is assumed to be a Bernoulli distribution, which is expressed as,

$$p(\mathbf{w}|\mathbf{z}) = z^{\mathbf{w}} (1-z)^{1-\mathbf{w}} \quad (9)$$

When  $\mathbf{w} = \mathbf{0}$ ,  $p(\mathbf{t}_n|\mathbf{w})$  in Eq. (8) becomes a Dirac delta function centered at 0, indicating the spike component being chosen instead of the slab.

An illustration of a spike-and-slab prior is depicted in Fig. 1. A mixture of two densities with different scales allow to discriminate between the slab components, which are left almost unchanged, and the spike components, whose posterior is peaked around zero. Moreover, the posterior expected value of the binary latent variable  $\mathbf{w}$  yields an estimate of the probabilities which can be useful for identifying relevant features.

#### 3.2 Parameter estimation using the EM algorithm

For the spike-and-slab prior PPCA model (SS-PPCA) to be developed, the unknown information  $\Theta = \{\mathbf{P}, \mathbf{z}, \tau_D, \tau_M, \tau_{\mathbf{w}}\}$  should be estimated. Given a set of observation data  $\mathbf{X}$ , the EM algorithm finds the optimal estimation of the model parameters iteratively maximizing the expectation of log-likelihood (a.k.a. Q-function) through expectation step and maximization step.

**E-step.** The log-likelihood of the complete data can be written as

$$Q(\Theta|\Theta_{\text{old}}) = \sum_{n=1}^N [\mathbb{E}\{\ln p(\mathbf{x}_n|\mathbf{w}, \mathbf{t}_n, \Theta)\} + \mathbb{E}\{\ln p(\mathbf{t}_n|\mathbf{w}, \Theta)\} + \mathbb{E}\{\ln p(\mathbf{w}|\Theta)\}], \quad (10)$$

where the symbol  $\mathbb{E}$  presents the expectation under old parameters, that is  $\mathbb{E}_{\mathbf{w}, \mathbf{t}_n|\mathbf{x}_n, \Theta_{\text{old}}}$ .

The expectation expressions on the binary latent variable  $\mathbf{w}$  given the current parameters  $\Theta_{\text{old}}$  is obtained by the marginalization over continuous and discrete latent space,

$$\mathbb{E}\{\mathbf{w}\} = \sum_{\mathbf{w}} \int_{\mathbf{t}_n} p(\mathbf{w}, \mathbf{t}_n|\mathbf{x}_n, \Theta_{\text{old}}) \mathbf{w} d\mathbf{t}_n = \frac{z \mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_M \mathbf{I})}{z \mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_M \mathbf{I}) + (1-z) \mathcal{N}(\mathbf{t}_n|\mathbf{0}, \tau_{\mathbf{w}} \mathbf{I})} \quad (11)$$

The conditional expectation,  $\mathbb{E}\{\frac{1}{\mathbf{w}\tau_M + (1-\mathbf{w})\tau_{\mathbf{w}}}\}$ , is expressed by a weighted average of the precisions in Eq. (8),

$$\mathbb{E}\left\{\frac{1}{\mathbf{w}\tau_M + (1-\mathbf{w})\tau_{\mathbf{w}}}\right\} = \frac{\mathbb{E}\{\mathbf{w}\}}{\tau_M} + \frac{1-\mathbb{E}\{\mathbf{w}\}}{\tau_{\mathbf{w}}} \quad (12)$$

The posterior over the latent variable  $\mathbf{t}_n$  given the parameters  $\Theta_{\text{old}}$  observed in the previous M-step can be derived using Bayes' theorem,

$$p(\mathbf{t}_n|\mathbf{x}_n, \mathbf{w}, \Theta_{\text{old}}) \propto p(\mathbf{x}_n|\mathbf{t}_n, \mathbf{w}, \Theta_{\text{old}}) p(\mathbf{t}_n|\mathbf{w}, \Theta_{\text{old}}) \quad (13)$$

Since  $\mathbf{x}_n$  and  $\mathbf{t}_n$  are joint Gaussian distributed, the score probability  $p(\mathbf{t}_n|\mathbf{w}, \Theta_{\text{old}})$  also follows Gaussian distribution with its mean  $\mathbf{m}$  and variance  $\mathbf{V}$  given as follows,

$$\mathbf{m} = \mathbf{V} \mathbf{P}^T (\tau_D^{-1} \mathbf{I}) \mathbf{x}_n \quad (14)$$

$$\mathbf{V} = (\tau_M^{-1} \mathbf{I} + \mathbf{P}^T (\tau_D^{-1} \mathbf{I}) \mathbf{P})^{-1} \quad (15)$$

The derivation of expectation expressions required for the M-step are straight-forward through the integrations over the continuous part, which is given as follows,

$$\mathbb{E}\{\mathbf{t}_n\} = \sum_{\mathbf{w}} \int_{\mathbf{t}_n} p(\mathbf{w}, \mathbf{t}_n|\mathbf{x}_n, \Theta_{\text{old}}) \mathbf{t}_n d\mathbf{t}_n = \mathbf{m} \quad (16)$$

$$\mathbb{E}\{\mathbf{t}_n \mathbf{t}_n^T\} = \sum_{\mathbf{w}} \int_{\mathbf{t}_n} p(\mathbf{w}, \mathbf{t}_n|\mathbf{x}_n, \Theta_{\text{old}}) \mathbf{t}_n \mathbf{t}_n^T d\mathbf{t}_n = \mathbf{V} + \mathbf{m} \mathbf{m}^T \quad (17)$$

**M-step.** The M-step parameter updates,  $\{\mathbf{P}, z, \tau_M, \tau_{\mathbf{w}}, \tau_D\}$  are obtained by computing the derivatives of the Q function with respect to the corresponding parameter, and taking it to zero. The following results hold,

$$\mathbf{P} = \frac{\sum_{n=1}^N \mathbf{x}_n \mathbb{E}\{\mathbf{t}_n\}^T}{\sum_{n=1}^N \mathbb{E}\{\mathbf{t}_n \mathbf{t}_n^T\}} \quad (18)$$

$$z = \frac{1}{N} \sum_{n=1}^N \mathbb{E}\{\mathbf{w}\} \quad (19)$$

$$\tau_M = \sum_{n=1}^N \left[ \mathbb{E}\{\mathbf{t}_n \mathbf{t}_n^T\} \mathbb{E}\left\{\frac{1}{\mathbf{w}\tau_M + (1-\mathbf{w})\tau_{\mathbf{w}}}\right\} \times \left(\sum_{n=1}^N \mathbb{E}\{\mathbf{w}\}\right)^{-1} \right] \quad (20)$$

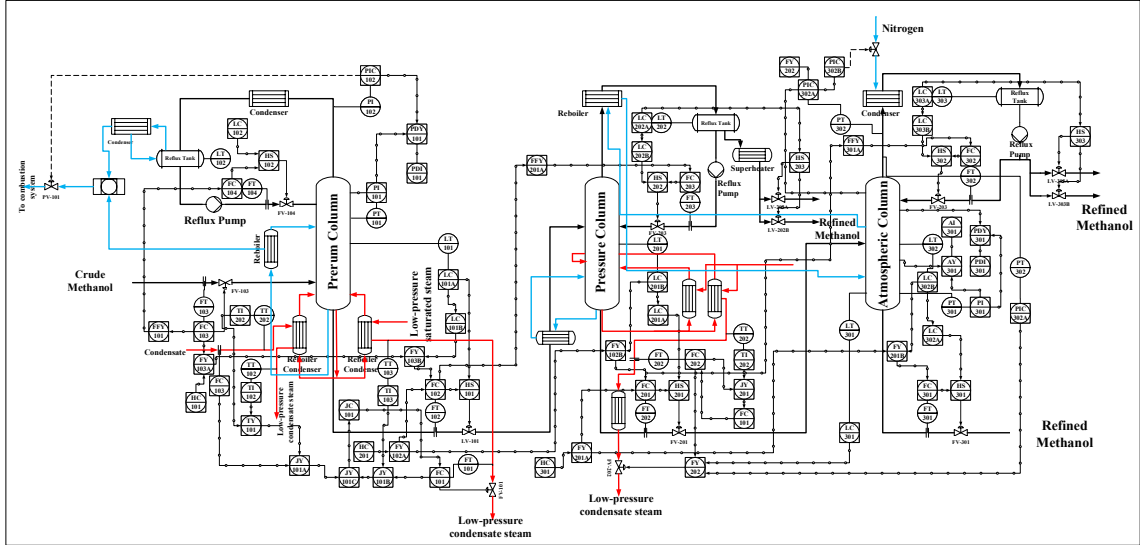


Fig. 2. P&ID diagram of methanol refining unit with a 3 column configuration. The red line and blue line represents temperature and pressure control scheme, respectively. The line with dot is feed control scheme.

$$\tau_{\mathbf{w}} = \sum_{n=1}^N \left[ \mathbb{E}\{\mathbf{t}_n \mathbf{t}_n^T\} \left( 1 - \mathbb{E} \left\{ \frac{1}{\mathbf{w} \tau_{\mathbf{M}} + (1 - \mathbf{w}) \tau_{\mathbf{w}}} \right\} \right) \right. \\ \left. \times \left( \sum_{n=1}^N [1 - \mathbb{E}\{\mathbf{w}\}] \right)^{-1} \right] \quad (21)$$

$$\tau_D = \frac{1}{N} \sum_{n=1}^N [\mathbf{x}_n \mathbf{x}_n^T - \mathbf{P} \mathbb{E}\{\mathbf{t}_n\} \mathbb{E}\{\mathbf{t}_n^T\} \mathbf{P}^T]. \quad (22)$$

The detailed derivations of Eqs. (18) - (22) are given in Appendix A. The expectation terms in Eqs. (18) - (22) are calculated using Eqs. (11), (12), (16) and (17).

### 3.3 Monitoring control charts based on SS-PPCA

To monitor the process based on the SS-PPCA developed from normal operations, various monitoring statistics are proposed, including monitoring chart based on (1) the latent variables, (2) the reconstruction residuals, and (3) measurement variables. The first two statistics are performed for fault detection. Once the fault has been detected, the latter statistic is to determine the location of the fault.

**Monitoring chart of latent variables.** The  $T^2$  statistic is used to measure the variation inside the sparse latent subspace. Since the posterior over the latent variable is distributed as a  $\mathcal{N}(\mathbf{m}, \mathbf{V})$ , the Euclidian norm of new latent variables  $\mathbf{t}_{\text{new}}$  follows a chi-square distribution. The  $T^2$  statistic associated with new samples  $\mathbf{x}_{\text{new}}$  can be computed as,

$$T^2 = \mathbb{E}\{\mathbf{t}_{\text{new}} | \mathbf{x}_{\text{new}}\} \mathbb{E}\{\mathbf{t}_{\text{new}} | \mathbf{x}_{\text{new}}\}^T \leq \chi_{1-\alpha, M}^{-2} \quad (23)$$

where  $\mathbb{E}\{\mathbf{t}_{\text{new}} | \mathbf{x}_{\text{new}}\}$  is identical to the posterior mean  $\mathbf{m}$  in Eq. (14),  $\chi_{1-\alpha, M}^{-2}$  is  $(1 - \alpha)$  - fractile of the inverse of the chi-square distribution with degrees of freedom  $M$  and  $\alpha$  is the confidence limit.

**Monitoring chart of reconstruction residuals.** The squared prediction error (SPE) statistic is used to measure the variation in the residual space. The SPE statistic

using the estimate of reconstruction residuals  $\mathbf{e}_{\text{new}}$  can be computed as

$$\text{SPE} = \mathbf{e}_{\text{new}}^T (\tau_D \mathbf{I}) \mathbf{e}_{\text{new}} \\ = (\mathbf{x}_{\text{new}} - \mathbf{P} \mathbf{t}_{\text{new}})^T (\tau_D \mathbf{I}) (\mathbf{x}_{\text{new}} - \mathbf{P} \mathbf{t}_{\text{new}}) \leq \chi_{1-\alpha, D}^{-2} \quad (24)$$

where

$$\mathbf{e}_{\text{new}} = \mathbf{x}_{\text{new}} - \mathbf{P} \mathbf{t}_{\text{new}} = (\mathbf{I} - \mathbf{V} \mathbf{P}^T \tau_D^{-1} \mathbf{I}) \mathbf{x}_{\text{new}}. \quad (25)$$

**Monitoring chart of measurement variables.** The contribution analysis of process variables or latent variables are widely used as fault isolation method in industrial systems. The contribution chart exploits correlations between variables to identify which variables contribute the most to the process fault. The contribution of the variable, subsequently, can be quantified by the posterior covariance. The data likelihood  $p(\mathbf{x}_n | \Theta_{\text{old}})$  can be derived by marginalizing the joint  $p(\mathbf{x}_n, \mathbf{w}, \mathbf{t}_n | \Theta_{\text{old}})$  over the latent variables  $\mathbf{t}_n$  and cumulating the result over the binary latent variables  $\mathbf{w}$ . The posterior distribution for the indicator variable  $\mathbf{w}$  given the current parameters  $\Theta_{\text{old}}$  is inferred as,

$$p(\mathbf{w} | \mathbf{x}_n, \Theta_{\text{old}}) = \frac{p(\mathbf{x}_n, \mathbf{w} | \Theta_{\text{old}})}{p(\mathbf{x}_n | \Theta_{\text{old}})} \\ \propto \mathcal{N}(\mathbf{x}_n | \mathbf{0}, \mathbf{P} (\tau_M^{-1} \mathbf{I}) \mathbf{P}^T + \tau_D \mathbf{I}) \quad (26)$$

The measurement on the correlations between variables, referred as  $M^2$  statistic, can be expressed as,

$$M^2 = \left\| (\mathbf{P} (\tau_M^{-1} \mathbf{I}) \mathbf{P}^T + \tau_D \mathbf{I})^{-0.5} \mathbf{x}_{\text{new}} \right\|^2 \\ = \mathbf{x}_{\text{new}}^T (\mathbf{P} (\tau_M^{-1} \mathbf{I}) \mathbf{P}^T + \tau_D \mathbf{I})^{-1} \mathbf{x}_{\text{new}} \leq \chi_{1-\alpha, D}^{-2} \quad (27)$$

Note that the confidence bounds based on the chi-square distribution in Eqs. (23), (24) and (27) are approximate since the mean and covariance are once estimated. In this work, the kernel density estimation (KDE) method can be utilized to determine the confidence bound.

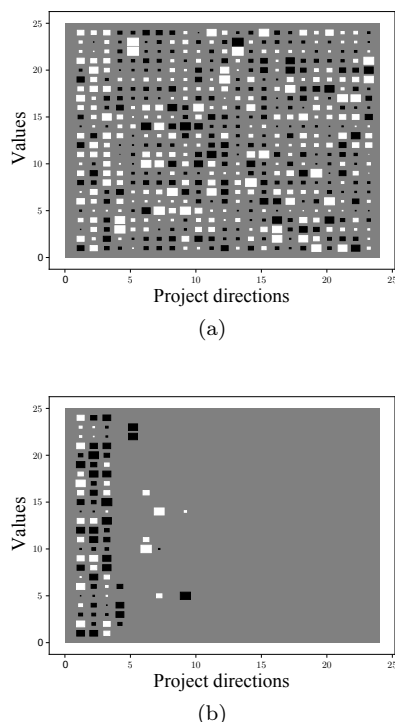


Fig. 3. Hinton diagrams of the loading matrix identified by (a) PPCA and (b) the proposed approach. Positive and negative values are represented by white and black squares, respectively, and the magnitude of value are represented by the size of square.

#### 4. APPLICATION: INDUSTRIAL METHANOL DISTILLATION

The application of this work is to detect process upsets in order to protect parts of an industrial methanol plant owned by China National Petroleum Corporation (CNPC). As one of the major feedstock for the olefins production, the methanol section consists of a standard arrangement of three distillation columns. A process diagram of the industrial methanol distillation is shown in Fig. 2.

To monitor and analyze the operation of the distillation section of a methanol plant, 24 variables based on engineering requirements have been continuously followed for more than 2 years. A total of 735 process samples used for the training datasets are collected from January 2017 to February 2018, while another 760 samples collected from February 2018 to May 2019 are required as the testing datasets to demonstrate the effectiveness of the proposed method. The variable time series trends after missing measurements imputations, outliers removing and z-score normalization. From February 2018 to May 2019, there were three different operating conditions causing by the process changing, which was consulted by the experts and engineers. They were located at 150th-165th, 305th-307th and 756th-760th samples in the test datasets.

The SS-PPCA and PPCA monitoring models were constructed based on the training datasets. The loading matrices initialized by these models are shown in Fig. 3. It has been seen that a total of 8 sparse components are pre-

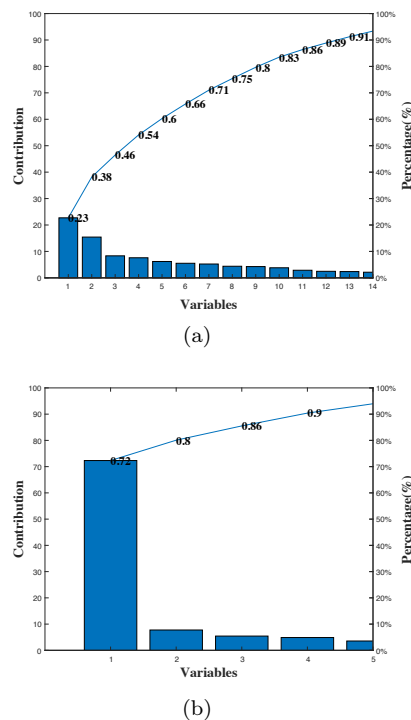


Fig. 4. Scree plot and cumulative variance explained plot for methanol refining plant, obtained through (a) PPCA and (b) the proposed approach.

served, while the other 16 are eliminated, as shown in Fig. 3(b). SS-PPCA could automatically capture the number of the local components and the sparsity of the loading matrix. It means that the weight associated with the useless components would be approximately zero, while that corresponding to the useful components should be one. For fair comparison, the numbers of latent variables accounted for more than 90 % variance were selected as 13 and 4 for the original PPCA and SS-PPCA, respectively. This can be illustrated as Fig. 4(a) and Fig. 4(b).

As shown in Fig. 5, an abnormal condition between the 150th and the 165th instances is indicated by the  $T^2$  charts of PPCA and SS-PPCA. PPCA based on a single Gaussian distribution, however, failed to detect the local mode between the 305th and the 307th instances, as shown in Fig. 5(a). Inspecting the  $T^2$  and SPE plots in Fig. 5(b) one sees that not only the  $T^2$ , but also the SPE statistics of SS-PPCA are moving simultaneously beyond confidence regions shortly after the samples 150th, 305th and 756th as a result of the different operating conditions being introduced. It can be judged that SS-PPCA shows the lowest missing detection rates for most conditions. This is due to the fact that the selection of the latent variables for each subgroup by SS-PPCA are discriminative and could be projected onto the latent space that contains the important information for abnormal data. Note that the tiny value of the  $T^2$  charts are found in Fig. 5(b) since the variation inside the sparse latent subspace is relative small.

Fig. 6 depicts the  $M^2$  contribution plots of PPCA and SS-PPCA at the 158th instance in the methanol refining plant (faults are effectively detected by  $T^2$  chart). Contribution

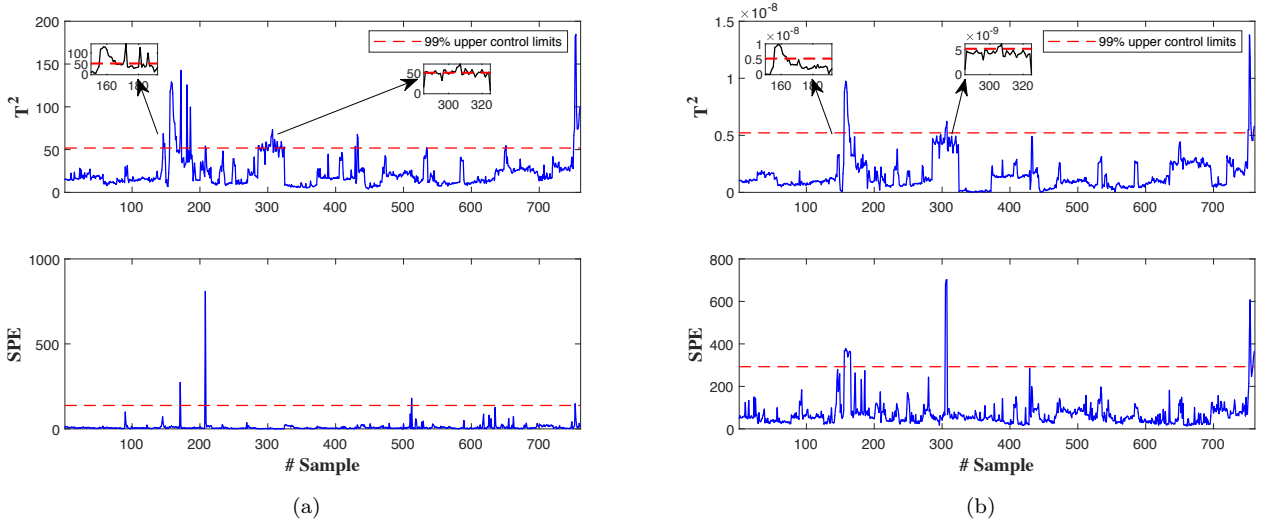


Fig. 5. Monitoring charts of (a) PPCA (the reduced dimension  $M = 13$ ), (b) the proposed method ( $M = 4$ ) for the methanol refining plant. Dashed horizontal lines are 99 % confidence limits computed by KDE.

plots provide important information for the diagnosis of the source of faults. It can be seen that the variable with precision degradation may contribute the more to the occurrence of abnormal condition. This can be attributed to the fact that 'smearing' effect on the contributing and non-contributing variable in the presence of control is different. Note that the variable 18 in Fig. 6(b) is the middle temperature of the atmospheric column, which has the long adjustment lags for the process load.

## 5. CONCLUSION

In this paper, a probabilistic regularization method SSP-PCA has been proposed for fault detection and isolation. SS-PPCA has the following notable features: (1) it explicitly considers that the selection of LVs has a great impact on the discriminative model. (2) SS-PPCA produces a sparse mapping from the input space that may be mixed with irrelevant variables. (3) The monitoring chart of measurement variables is an effective diagnostic tool since the sparsity pattern is common to all LVs. Simulation results show the superiority of SS-PPCA-based process monitoring to popular probabilistic LV model. Furthermore, the case study on industrial methanol distillation illustrates the proposed approach keeps desirable monitoring performance for industrial scenarios.

### Appendix A. DERIVATION OF THE M-STEP

**Updating the loading matrix  $\mathbf{P}$ :** Parameter  $\mathbf{P}$  is updated by setting the derivatives of the Q function with respect to  $\mathbf{P}$  and taking it to zero,

$$\sum_{n=1}^N \left\{ -\frac{1}{2} \frac{\partial}{\partial \mathbf{P}} \left[ (\mathbf{x}_n - \mathbf{P}\mathbb{E}\{\mathbf{t}_n\})^T \tau_D^{-1} (\mathbf{x}_n - \mathbf{P}\mathbb{E}\{\mathbf{t}_n\}) \right] \right\} = 0, \quad (\text{A.1})$$

the Eq. (18) holds.

**Updating the weight  $z$ :** Derivation of Eq. (19) is given as follows,

$$\begin{aligned} \frac{\partial Q(\Theta|\Theta_{\text{old}})}{\partial z} &= \sum_{n=1}^N \left[ \mathbb{E} \left\{ \frac{\mathbf{w}}{z} - \frac{1-\mathbf{w}}{1-z} \right\} \right] = 0 \\ \implies z &= \frac{1}{N} \sum_{n=1}^N \mathbb{E}\{\mathbf{w}\}. \end{aligned} \quad (\text{A.2})$$

**Updating the precise in the latent variable and the noise:** Derivation of Eq. (19) is as follows,

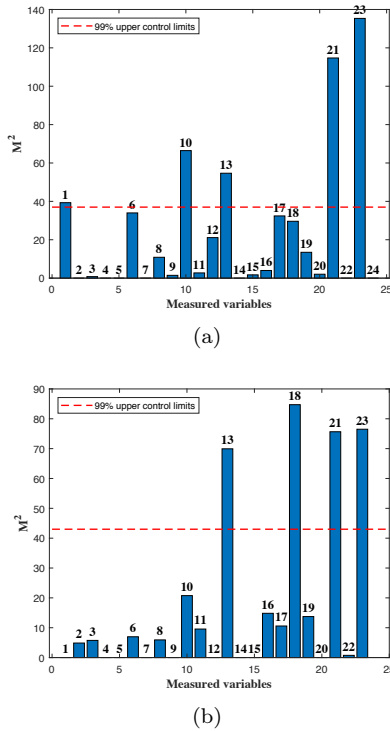


Fig. 6. The averaged contribution plots of (a) PPCA and (b) the proposed method after the detection of faults. Dashed horizontal lines are 99 % confidence limits computed by KDE.



$$\begin{aligned} \frac{\partial Q(\Theta|\Theta_{\text{old}})}{\partial \mathbf{P}} &= 0 \\ \Rightarrow \sum_{n=1}^N \left[ -\frac{1}{2\tau_M} \mathbb{E}\{\mathbf{w}\} + \frac{1}{2} \mathbb{E}\{\mathbf{t}_n \mathbf{t}_n^T\} \right. \\ &\quad \left. \times \mathbb{E}\left\{ \frac{1}{\mathbf{w}\tau_M + (1-\mathbf{w})\tau_{\mathbf{w}}} \right\} \right] = 0 \\ \Rightarrow \tau_M &= \sum_{n=1}^N \left[ \mathbb{E}\{\mathbf{t}_n \mathbf{t}_n^T\} \mathbb{E}\left\{ \frac{1}{\mathbf{w}\tau_M + (1-\mathbf{w})\tau_{\mathbf{w}}} \right\} \right. \\ &\quad \left. \times \left( \sum_{n=1}^N \mathbb{E}\{\mathbf{w}\} \right)^{-1} \right]. \end{aligned} \quad (\text{A.3})$$

Following a similar derivation, we can get the derivation of Eqs. (21) - (22).

#### ACKNOWLEDGEMENTS

The authors acknowledge the financial support by National Natural Science Foundation of China (Grants No. 61703191), the Foundation of Liaoning Educational Committee (No. L2017LQN028) and the Scientific Research Foundation of Liaoning Shihua University (No. 2017XJJ-012) for this work.

#### REFERENCES

Chen, Z., Ding, S.X., Zhang, K., Li, Z., and Hu, Z. (2016). Canonical correlation analysis-based fault detection methods with application to alumina evaporation process. *Control Engineering Practice*, 46, 51 – 58.

Ge, Z., Gao, F., and Song, Z. (2011). Two-dimensional bayesian monitoring method for nonlinear multimode processes. *Chemical Engineering Science*, 66(21), 5173 – 5183.

Ge, Z. and Song, Z. (2010). Mixture bayesian regularization method of ppca for multimode process monitoring. *Aiche Journal*, 56, 2838 – 2849.

Ghosh, K., Ramteke, M., and Srinivasan, R. (2014). Optimal variable selection for effective statistical process monitoring. *Computers & Chemical Engineering*, 60, 260 – 276.

Kodamana, H., Raveendran, R., and Huang, B. (2019). Mixtures of probabilistic pca with common structure latent bases for process monitoring. *IEEE Transactions on Control Systems Technology*, 27(2), 838–846.

Lee, J., Kang, B., and Kang, S.H. (2011). Integrating independent component analysis and local outlier factor for plant-wide process monitoring. *Journal of Process Control*, 21(7), 1011 – 1021.

Lee, Y.K., Mavris, D.N., Volovoi, V.V., Yuan, M., and Fisher, T. (2010). A fault diagnosis method for industrial gas turbines using Bayesian data analysis. *Journal of Engineering for Gas Turbines and Power*, 132(4), 041602.

Li, G., Qin, S.J., and Zhou, D. (2010). Geometric properties of partial least squares for process monitoring. *Automatica*, 46(1), 204 – 210.

Liu, Y., Liu, B., Zhao, X., and Xie, M. (2018). A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring. *IEEE Transactions on Industrial Electronics*, 65(8), 6478–6486.

Mehmood, T., Liland, K.H., Snipen, L., and Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62 – 69.

Peres, F.A.P. and Fogliatto, F.S. (2018). Variable selection methods in multivariate statistical process control: A systematic literature review. *Computers & Industrial Engineering*, 115, 603 – 619.

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

Srinivasan, R. and Qian, M. (2007). State-specific key variables for monitoring multi-state processes. *Chemical Engineering Research and Design*, 85(12), 1630 – 1644.

Tipping, M.E. and Bishop, C.M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(3), 611 – 622.

Wang, K. and Jiang, W. (2009). High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*, 41(3), 247–258.

Woodall, W.H. and Montgomery, D.C. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology*, 46(1), 78–94.

Zeng, J., Huang, B., and Xie, L. (2015). A bayesian sparse reconstruction method for fault detection and isolation. *Journal of Chemometrics*, 29(6), 349–360.

Zhu, J., Ge, Z., and Song, Z. (2017). Non-gaussian industrial process monitoring with probabilistic independent component analysis. *IEEE Transactions on Automation Science and Engineering*, 14(2), 1309–1319.