A Unified Probabilistic Monitoring Framework for Multimode Processes Based on Probabilistic Linear Discriminant Analysis

Author:
Liu, Y; Zeng, J; Bao, J; Xie, L

Publication details:
IEEE Transactions on Industrial Informatics
v. 16
Chapter No. 10
pp. 6291 - 6300
1551-3203 (ISSN); 1941-0050 (ISSN)

Publication Date:
2020-10-01

Publisher DOI:
https://doi.org/10.1109/TII.2020.2966707

A unified probabilistic monitoring framework for multimode processes based on probabilistic linear discriminant analysis

***Abstract*—This work develops a novel probabilistic monitoring framework for industrial processes with multiple operational conditions. The proposed method is based on the probabilistic linear discriminant analysis (PLDA), which relies on two sets of latent variables, *i.e.*, the between-class and within-class latent variables. In order to deal with the large within-class variations in multi-mode industrial processes, this approach modifies the original PLDA by introducing a separate within-class loading matrix for each operational mode and designs an expectation maximization (EM) algorithm to estimate the model parameters from the training samples. Mode identification for test samples is achieved by investigating the cosine similarity in the between-class latent variables and two monitoring statistics corresponding to within-class latent variables and the residuals are considered for fault detection. To diagnose the process fault, this paper further develops a sparse probabilistic generative model based on PLDA for fault isolation. The enhanced performance of the proposed method is illustrated by applications to numerical examples and industrial processes.**

***Index Terms*—Multi-mode process monitoring, probabilistic linear discriminant analysis, latent variables, fault isolation.**

I. Introduction

To achieve improved system reliability and operational safety, more and more attention has been paid on the development of process monitoring methods. Thanks to the advancement of information technology, process monitoring methods, including data-driven techniques and model-based approaches, have been extensively studied over the past few decades [1]. In contrast to model-based approaches, data-driven techniques do not require much priori knowledge and have shown to be conceptually simple and with low implementation cost.

One type of data-driven techniques is based on the multivariate statistical process control (MSPC) [2]. The basic idea of MSPC methods is to obtain a set of latent variables by mapping the process data onto a lower dimensional subspace based on certain criteria. Once the latent variables are obtained, monitoring statistics like Hotelling's T^2 and squared prediction error (SPE) can be constructed for fault detection and isolation. It should be noted that the training datasets for these methods are usually assumed to admit a unimodal distribution. This can be problematic as practical industrial processes often undergo frequent production shifts due to changes in product quality specifications, manufacturing strategies, and working environments [3]. Thus, process monitoring models trained under the unimodal assumption may fail and produce significant number of false alarms. In order to deal with the monitoring problem of multi-mode processes, various methods have been developed, including adaptive methods [4], [5], similarity analysis based methods [6], [7], subspace separation [8], [9], and mixture models [10], [11]. More recently, monitoring of nonlinear processes using the kernel trick has also been considered and good results have been achieved [12], [13].

Among multi-mode process monitoring techniques, probabilistic approaches have received considerable attention [14]. Comparing to conventional methods, probabilistic approaches are advantageous in its capability of (i) incorporating prior knowledge using appropriate distributions; (ii) better handling of process uncertainty; (iii) easy accommodation of missing data and outliers [15], [16]. Various probabilistic methods has been reported in the literature, such as Gaussian mixture models [17], probabilistic principal component analysis [18], hidden Markov models [15], and conditional random fields [19]. They have been successful in detecting process fault and diagnosing operational modes in a number of applications. For fault isolation and diagnosis, however, possible solutions include formulating the fault isolation and diagnosis problem in a way similar to contribution analysis [20], or performing an additional root cause diagnosis using causal models like Bayesian networks [21], or transfer entropy [22]. Several probabilistic methods like probabilistic contribution analysis [23] and probabilistic reconstruction [24] have also been proposed. Despite the research progress on probabilistic monitoring methods, it still lacks a unified probabilistic framework which combines the fault detection and isolation for multi-mode processes.

In this paper, a novel probabilistic monitoring framework is proposed based on the probabilistic linear discriminant analysis [25]. However, the standard PLDA is not suitable for handling datasets with significantly large within-class variance [26], which is common in multi-mode industrial processes. In order to accommodate multi-mode industrial processes, this paper introduces a separate within-class loading matrix for each operational mode and designs two monitoring statistics, leading to better fault detection results. For mode identification, a cosine similarity is developed based on the between-class latent variables. In addition, a sparse probabilistic generative model based on PLDA is developed to isolate faulty variables. The contribution of this paper can be summarized as follows: (i) A unified probabilistic monitoring framework is developed for fault detection and isolation for multi-mode processes; (ii) The standard PLDA is extended by introducing a separate within-class loading matrix for each mode to accommodate the dataset with large within-class variance, leading to better fault detection results; (iii) A sparse probabilistic generative model based on PLDA is proposed to isolate faulty variables.

II. Probabilistic Linear Discriminant Analysis

In this section, the standard PLDA model [25], a probabilistic version of linear discriminant analysis (LDA), is briefly introduced. In PLDA, the j-th sample in the i-th class, $\mathbf{x}_{ij} \in \mathscr{R}^m$, can be described by two sets of latent variables with Gaussian priors. The corresponding probabilistic generative model can be given as

$$\begin{aligned} \mathbf{x}_{ij} &= \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \boldsymbol{\varepsilon}_{ij} \\ \mathbf{h}_i &\sim \mathscr{N}(\mathbf{0}, \mathbf{I}) \\ \mathbf{w}_{ij} &\sim \mathscr{N}(\mathbf{0}, \mathbf{I}) \\ \boldsymbol{\varepsilon}_{ij} &\sim \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned} \tag{1}$$

where $\mathbf{0}$ and $\mathbf{I}$ represent the zero vector and the identity matrix with appropriate dimensions, $\boldsymbol{\mu} = \frac{1}{N}\sum_{ij}\mathbf{x}_{ij}$ is the global mean value vector, with N being the size of the whole dataset. $\mathscr{N}(\cdot\,,\,\cdot)$ denotes the Gaussian distribution with the selected parameters. The assumption of Gaussian distribution is based on the fact that the data in a specific mode of multi-mode processes can be regarded as Gaussian or approximately Gaussian. Such assumption has also been used in a series of research work based on probabilistic models [15], [18]. $\mathbf{h}_i \in \mathscr{R}^{D_F}$ is the between-class latent variable shared by all the samples generated from the i-th class and $\mathbf{w}_{ij} \in \mathscr{R}^{D_G}$ is the within-class latent variable which explains the sample variation. $\mathbf{F} \in \mathscr{R}^{m\times D_F}$ and $\mathbf{G} \in \mathscr{R}^{m\times D_G}$ are two low-rank loading matrices. $\boldsymbol{\varepsilon}_{ij}$ is the stochastic Gaussian noise with zero mean and a diagonal heteroscedastic covariance matrix $\boldsymbol{\Sigma}$. The model parameters of PLDA can be grouped as $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}, \boldsymbol{\Sigma}\}$. An expectation maximization (EM) iteration procedure was introduced to estimate the model parameters in Ref. [25]. Comparing to PPCA, PLDA has better flexibility and is able to make probabilistic inferences about the class/mode identity.

III. Improved Probabilistic Linear Discriminant Analysis

As is discussed in Section I, the standard PLDA is not suitable for dealing with data with large within-class variations, which is common in industrial multi-mode processes. To cope with this problem, in this section, an improved PLDA (I-PLDA) model is considered.

A. Model Structure

To better deal with multi-mode processes, the standard PLDA is modified by replacing the within-class loading matrix $\mathbf{G}$ by $\mathbf{G}_i$ for $i = 1, \cdots, I$, with each $\mathbf{G}_i$ corresponding to a specific process mode/class. Hence, the generative model of I-PLDA can now be written as

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}_i\mathbf{w}_{ij} + \boldsymbol{\varepsilon}_{ij} \tag{2}$$

Note that in Eq.(2), by integrating out the latent variable $\mathbf{w}_{ij}$, the within-class variances are now obtained as $\mathbf{S}_i = \mathbf{G}_i\mathbf{G}_i + \boldsymbol{\Sigma}$, in contrast to $\mathbf{S}_i = \mathbf{G}\mathbf{G} + \boldsymbol{\Sigma}$ of PLDA, which is the same for all classes. This renders I-PLDA better flexibility, as it introduces a specific variance for data in each mode. Now the parameter set of I-PLDA is extended as $\tilde{\boldsymbol{\theta}} = \{\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}_i, \boldsymbol{\Sigma}, i = 1, \cdots, I\}$. The above treatment introduces significant flexibility to I-PLDA and leads to improved effectiveness in handling industrial multi-mode processes.

B. Model Estimation

Assume a dataset $\mathbf{X} = \mathbf{X}_1 \bigcup \mathbf{X}_2 \bigcup \cdots \bigcup \mathbf{X}_I$ has been collected, with $\mathbf{X}_i = \{\mathbf{x}_{ij}\}_{j=1}^{J}$ storing all J samples belonging to the i-th class, so that the size of $\mathbf{X}$ becomes $N = IJ$. In order to learn the parameter set $\tilde{\boldsymbol{\theta}}$ for I-PLDA, an iterative EM algorithm is developed. The EM algorithm is an optimization method particularly suitable for problems with some variables unobservable but whose probability distributions are known. It is a powerful and widely used tool for estimation of latent variable models [14]. The EM algorithm mainly consists of an E-step and a M-step as follows [27].

Part I: E-step. To facilitate the joint inference of the latent variables, the probabilistic generative model for the dataset $\mathbf{X}_i$ can be written as

$$\underbrace{\begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{iJ} \end{bmatrix}}_{\tilde{\mathbf{X}}_i} = \underbrace{\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}}_{\tilde{\boldsymbol{\mu}}_i} + \underbrace{\begin{bmatrix} \mathbf{F} & \mathbf{G}_i & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G}_i & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G}_i \end{bmatrix}}_{\tilde{\mathbf{A}}_i} \underbrace{\begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{i1} \\ \mathbf{w}_{i2} \\ \vdots \\ \mathbf{w}_{iJ} \end{bmatrix}}_{\mathbf{y}_i} + \underbrace{\begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \\ \boldsymbol{\varepsilon}_{i2} \\ \vdots \\ \boldsymbol{\varepsilon}_{iJ} \end{bmatrix}}_{\boldsymbol{\varepsilon}_i} \tag{3}$$

Here $\tilde{\mathbf{X}}_i$ is the data matrix storing all samples belonging to class i, $\tilde{\boldsymbol{\mu}}_i$ and $\tilde{\mathbf{A}}_i$ are the concatenated mean and parameter matrices. According to Eq.(1), both the concatenated hidden variable matrix $\mathbf{y}_i$ and the concatenated noise matrix $\boldsymbol{\varepsilon}_i$ follow a Gaussian distribution, with $\mathbf{y}_i \sim \mathscr{N}(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\varepsilon}_i \sim \mathscr{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}})$. $\tilde{\boldsymbol{\Sigma}}$ is a block diagonal matrix with each diagonal block being $\boldsymbol{\Sigma}$. With the distributions of hidden variable and noise terms determined, the parameters in Eq.(3) can be estimated using the EM algorithm. According to Ref. [25], the posterior expectation of $\mathbf{y}_i = \begin{bmatrix} \mathbf{h}_i^T & \mathbf{w}_{i1}^T & \cdots & \mathbf{w}_{iJ_i} \end{bmatrix}$ is obtained by

$$\begin{aligned} \langle \mathbf{y}_i \rangle &= \boldsymbol{\Xi}_i \tilde{\mathbf{A}}_i^T \tilde{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{X}}_i - \tilde{\boldsymbol{\mu}} \right) \\ \boldsymbol{\Xi}_i &= \left(\tilde{\mathbf{A}}_i^T \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{A}}_i + \mathbf{I} \right)^{-1} \end{aligned} \tag{4}$$

where $\langle \cdot \rangle$ denotes the expectation of a latent variable. $\boldsymbol{\Xi}_i$ is the posterior covariance of $\tilde{\boldsymbol{\Sigma}}$. Note that the matrix inversion in Eq.(4) is usually intractable due to the high dimension. Fortunately, this problem can be efficiently solved by the *partitioned inverse formula* [28] as

$$\begin{aligned} \boldsymbol{\Xi}_i &= \begin{bmatrix} \mathbf{A}_i & \mathbf{B}_i \\ \mathbf{C}_i & \mathbf{D}_i \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{U}_i & -\mathbf{U}_i\mathbf{B}_i\mathbf{D}_i^{-1} \\ -\mathbf{D}_i^{-1}\mathbf{C}_i\mathbf{U}_i & \mathbf{D}_i^{-1}\mathbf{C}_i\mathbf{U}_i\mathbf{B}_i\mathbf{D}_i^{-1} + \mathbf{D}_i^{-1} \end{bmatrix} \end{aligned} \tag{5}$$

where $\mathbf{U}_i$, $\mathbf{B}_i$, $\mathbf{C}_i$, and $\mathbf{D}_i$ are obtained from Eq.(3) and Eq.(4) as follows.

$$\begin{aligned} \mathbf{U}_i &= \left(\mathbf{I} + J\mathbf{F}^T\mathbf{Q}_i\mathbf{F} \right)^{-1} \\ \mathbf{B}_i &= \begin{bmatrix} \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}_i & \cdots & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}_i \end{bmatrix} \\ \mathbf{D}_i &= \mathrm{diag}\left(\begin{bmatrix} \mathbf{V}_i & \cdots & \mathbf{V}_i \end{bmatrix} \right) \end{aligned} \tag{6}$$

with $\mathbf{V}_i=\left(\mathbf{I}+\mathbf{G}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\right)^{-1}$, $\mathbf{Q}_i=\left(\boldsymbol{\Sigma}+\mathbf{G}_i\mathbf{G}_i^T\right)^{-1}$, and $\mathbf{C}_i=\mathbf{B}_i^T$. Based on the formulas of Eqs.(4∼6), the posterior expectations of the between- and within-class latent variables can be updated by

$$\langle\mathbf{h}_i\rangle=\mathbf{U}_i\mathbf{F}^T\mathbf{Q}_i\sum_{j=1}^{J}(\mathbf{x}_{ij}-\boldsymbol{\mu}) \tag{7}$$

$$\langle\mathbf{w}_{ij}\rangle=\mathbf{V}_i\mathbf{G}_i^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{F}\langle\mathbf{h}_i\rangle) \tag{8}$$

The second order moments of these two latent variables are estimated by

$$\langle\mathbf{h}_i\mathbf{h}_i^T\rangle=\langle\mathbf{h}_i\rangle\langle\mathbf{h}_i^T\rangle+\operatorname{cov}(\mathbf{h}_i,\mathbf{h}_i) \tag{9}$$

$$\langle\mathbf{w}_{ij}\mathbf{w}_{ij}^T\rangle=\langle\mathbf{w}_{ij}\rangle\langle\mathbf{w}_{ij}^T\rangle+\operatorname{cov}(\mathbf{w}_{ij},\mathbf{w}_{ij}) \tag{10}$$

$$\langle\mathbf{h}_i\mathbf{w}_{ij}^T\rangle=\langle\mathbf{h}_i\rangle\langle\mathbf{w}_{ij}^T\rangle+\operatorname{cov}(\mathbf{h}_i,\mathbf{w}_{ij}) \tag{11}$$

where “cov” is the covariance function and

$$\begin{aligned}&\operatorname{cov}(\mathbf{h}_i,\mathbf{h}_i)=\mathbf{U}_i,\\&\operatorname{cov}(\mathbf{h}_i,\mathbf{w}_{ij})=-\mathbf{U}_i\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\mathbf{V}_i,\text{and}\\&\operatorname{cov}(\mathbf{w}_{ij},\mathbf{w}_{ij})=\mathbf{V}_i\mathbf{G}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\mathbf{U}_i\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\mathbf{V}_i+\mathbf{V}_i.\end{aligned}$$

Part II: M-step. For the purpose of parameter updating given the observed dataset and the posterior expectations, the EM auxiliary function from the complete log-likelihood is firstly extracted as

$$\begin{aligned}&Q(\boldsymbol{\Theta},\boldsymbol{\Theta}_{old})=\langle\ln p(\mathbf{X}|\boldsymbol{\Theta}_{old},\mathbf{H})\rangle_{q(\mathbf{H}|\mathbf{X},\boldsymbol{\Theta})}\\&\equiv-\frac{1}{2}\sum_{i=1}^{I}\left\{J\left(\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_i\right)+\ln|\boldsymbol{\Sigma}|\right)+\sum_{j=1}^{J}\langle\tilde{\mathbf{x}}_{ij}^T\rangle\boldsymbol{\Sigma}^{-1}\langle\tilde{\mathbf{x}}_{ij}\rangle\right\}\end{aligned} \tag{12}$$

where $q(\cdot)$ denotes the posterior distribution, $\equiv$ indicates equality up to an additive constant, and $\mathbf{H}$ the collection of all latent variables. The other terms can be expanded as $\langle\tilde{\mathbf{x}}_{ij}\rangle=\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{F}\langle\mathbf{h}_i\rangle-\mathbf{G}_i\langle\mathbf{w}_{ij}\rangle$ and $\boldsymbol{\Gamma}_i=\mathbf{F}\mathbf{U}_i\mathbf{F}^T+\mathbf{G}_i\operatorname{cov}(\mathbf{w}_{ij},\mathbf{w}_{ij})\mathbf{G}_i{}^T+2\mathbf{F}\operatorname{cov}(\mathbf{h}_i,\mathbf{w}_{ij})\mathbf{G}_i{}^T$.

The parameter set $\tilde{\boldsymbol{\theta}}$ for I-PLDA is then updated by taking the corresponding derivatives. For example, by setting the derivative with respect to $\mathbf{F}$ as zero, parameter matrix $\mathbf{F}$ can be updated as

$$\mathbf{F}=\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\bar{\mathbf{x}}_{ij}\langle\mathbf{h}_i^T\rangle-\mathbf{G}_i\langle\mathbf{w}_{ij}\mathbf{h}_i^T\rangle\right]\times\left[\sum_{i=1}^{I}J\langle\mathbf{h}_i\mathbf{h}_i^T\rangle\right]^{-1} \tag{13}$$

where $\bar{\mathbf{x}}_{ij}=\mathbf{x}_{ij}-\boldsymbol{\mu}$ and $\langle\mathbf{w}_{ij}\mathbf{h}_i^T\rangle=\langle\mathbf{h}_i\mathbf{w}_{ij}^T\rangle^T$. Similarly, other parameters are updated as follows.

$$\mathbf{G}_i=\sum_{j=1}^{J}\left[\bar{\mathbf{x}}_{ij}\langle\mathbf{w}_{ij}^T\rangle-\mathbf{F}\langle\mathbf{h}_i\mathbf{w}_{ij}^T\rangle\right]\times\left[\sum_{j=1}^{J}\langle\mathbf{w}_{ij}\mathbf{w}_{ij}^T\rangle\right]^{-1} \tag{14}$$

$$\boldsymbol{\Sigma}=\operatorname{diag}\left[\frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J}\langle\boldsymbol{\varepsilon}_{ij}\rangle\langle\boldsymbol{\varepsilon}_{ij}^T\rangle+\operatorname{cov}(\boldsymbol{\varepsilon}_{ij},\boldsymbol{\varepsilon}_{ij})\right] \tag{15}$$

where $\langle\boldsymbol{\varepsilon}_{ij}\rangle=\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{F}\langle\mathbf{h}_i\rangle-\mathbf{G}_i\langle\mathbf{w}_{ij}\rangle$ and $\operatorname{cov}(\boldsymbol{\varepsilon}_{ij},\boldsymbol{\varepsilon}_{ij})=\mathbf{F}\operatorname{cov}(\mathbf{h}_i,\mathbf{h}_i)\mathbf{F}^T+2\mathbf{F}\operatorname{cov}(\mathbf{h}_i,\mathbf{w}_{ij})\mathbf{G}_i^T+\mathbf{G}_i\operatorname{cov}(\mathbf{w}_{ij},\mathbf{w}_{ij})\mathbf{G}_i^T$.

Let k denote the iteration number, the full training scheme based on the joint updating EM algorithm can be summarized here.

Algorithm 1. EM for I-PLDA
Input: training data set $\mathbf{X}$, dimension sizes D_F and D_G.
1. Randomly initialize parameters $\mathbf{F}$, $\mathbf{G}_i$, and $\boldsymbol{\Sigma}$;
2. Compute the parameters $\mathbf{U}_i$, $\mathbf{V}_i$, and $\mathbf{Q}_i$ using Eq.(6);
3. Update the posterior expectations of $\{\mathbf{h}_i,\mathbf{w}_{ij}\}$ and the related second order moments using Eqs.(7∼11);
4. Update parameters $\mathbf{F}$, $\mathbf{G}_i$, and $\boldsymbol{\Sigma}$ using Eqs.(13∼15);
5. Calculate the likelihood of the total dataset as $L^{(k+1)}=\sum_{ij}\ln\mathcal{N}\left(\mathbf{x}_{ij}\middle|\boldsymbol{\mu}+\mathbf{F}^{(k+1)}\langle\mathbf{h}_i\rangle^{(k+1)},\mathbf{S}_i^{(k+1)}\right)$;
6. If $\left|\frac{L^{(k+1)}-L^{(t)}}{L^{(k)}}\right|>\varepsilon$, set $k=k+1$ and go back to Step 2, otherwise terminate the iteration and go to Step 7;
7. Output the model parameters and posterior expectations of all latent variables.

C. Parameter Selection

Note that the dimension sizes D_F and D_G for latent variables should be determined first in Algorithm 1. Usually, the natural constrains for them are $1\le D_F\le m$ and $1\le D_G\le m$, where m is the number of process variables. Although it has been suggested that these two parameters should be consistent [29], $1\le D_F\le m\le D_G\le 2m$ is adopted in this paper, as a smaller value of D_G will lead to a lower dimensional multivariate Laplace prior in the subsequent fault isolation, which tends to produce biased estimates and can be easily affected by outliers for high dimensional data [30]. The optimal values of D_F and D_G can be experimentally determined using trial and error.

IV. Process Monitoring Strategy

Until now, the estimations of model parameters and latent variables have been obtained. In order to construct a monitoring model for industrial processes with multiple conditions, this section develops a process monitoring strategy consisting of a mode identification step, a statistics construction step, and a fault isolation step.

A. Process Mode Identification

As shown in Eq.(1), the between-class latent variable $\mathbf{h}_i$ is used to indicate the class label. In order to identify the mode of a test sample $\mathbf{x}_t$, it is important to see how the test sample interacts with $\mathbf{h}_i$. Note that the estimation of $\mathbf{h}_i$ in Eq.(7) can be obtained by summing up the contribution of J samples in mode/class i, *i.e.*, $\langle\mathbf{h}_i\rangle=\sum_{j=1}^{J}\langle\mathbf{h}_{ij}\rangle$, with $\langle\mathbf{h}_{ij}\rangle$ being expressed as

$$\langle\mathbf{h}_{ij}\rangle=\mathbf{U}_i\mathbf{F}^T\mathbf{Q}_i(\mathbf{x}_{ij}-\boldsymbol{\mu}) \tag{16}$$

Thus, if a test sample $\mathbf{x}_t$ belongs to class/mode i, the corresponding contribution can be obtained by replacing $\mathbf{x}_{ij}$ with $\mathbf{x}_t$ in Eq.(16), resulting in a temporary between-sample latent variable $\mathbf{h}_{t|i}$, with the expectation of

$$\langle\mathbf{h}_{t|i}\rangle=\mathbf{U}_i\mathbf{F}^T\mathbf{Q}_i(\mathbf{x}_t-\boldsymbol{\mu}) \tag{17}$$

On the other hand, the average contribution to $\langle\mathbf{h}_i\rangle$ from all the samples in class/mode i can be defined as $\tilde{\mathbf{h}}_i$, with the expectation of

$$\langle\tilde{\mathbf{h}}_i\rangle=\frac{1}{J}\sum_{j=1}^{J}\mathbf{U}_i\mathbf{F}^T\mathbf{Q}_i(\mathbf{x}_{ij}-\boldsymbol{\mu}) \tag{18}$$

Identification of the mode label of a test sample $\mathbf{x}_t$ can be achieved by investigating the similarity between $\tilde{\mathbf{h}}_i$ and $\mathbf{h}_{i|t}$. The closer $\mathbf{h}_{i|t}$ is to $\tilde{\mathbf{h}}_i$, the more probable $\mathbf{x}_t$ can be assigned into mode/class i. To capture the similarity between $\tilde{\mathbf{h}}_i$ and $\mathbf{h}_{i|t}$, the cosine similarity is adopted here as

$$\cos\left(\tilde{\mathbf{h}}_i, \mathbf{h}_{i|t}\right) = \frac{\langle \tilde{\mathbf{h}}_i^T \rangle \langle \mathbf{h}_{i|t} \rangle}{\left\| \langle \tilde{\mathbf{h}}_i \rangle \right\|_2 \left\| \langle \mathbf{h}_{i|t} \rangle \right\|_2} \tag{19}$$

Consequently, the mode of $\mathbf{x}_t$ can be identified as

$$s = \arg\max_i \cos\left(\tilde{\mathbf{h}}_i, \mathbf{h}_{i|t}\right) \tag{20}$$

B. Process Monitoring Statistics

Once the mode label of a test sample is determined, it is required to construct monitoring statistics for fault detection. Here, conventional T^2 and *SPE* statistics are used. For I-PLDA, these two statistics are constructed as follows.

$$T^2 = \left\langle \mathbf{w}_{t|s}^T \right\rangle \left\langle \mathbf{w}_{t|s} \right\rangle \tag{21}$$

$$SPE = \left\langle \boldsymbol{\varepsilon}_{t|s}^T \right\rangle \left\langle \boldsymbol{\varepsilon}_{t|s} \right\rangle \tag{22}$$

where $\left\langle \mathbf{w}_{t|s} \right\rangle$ can be obtained using Eq.(8). $\left\langle \boldsymbol{\varepsilon}_{t|s} \right\rangle$ can be derived by the difference between the test sample $\mathbf{x}_t$ and the related estimation $\boldsymbol{\mu} + \mathbf{F}\langle \mathbf{h}_s \rangle + \mathbf{G}\left\langle \mathbf{w}_{t|s} \right\rangle$. The control limits of the two statistics can be obtained using kernel density estimation (KDE).

C. Fault Isolation Method

Once a fault is detected, it becomes important to timely and accurately localize the faulty variables. In this subsection, a sparse probabilistic generative model based on I-PLDA is proposed to isolate faulty variables.

1) Sparse Probabilistic Generative Model: Assume a fault has been detected by the T^2 statistic, the following sparse probabilistic generative model (SPGM) is considered as

$$\begin{aligned} \mathbf{x}_{t|s} &= \underbrace{\boldsymbol{\mu} + \mathbf{F}\mathbf{h}_s + \mathbf{G}_s \mathbf{w}_{t|s}^*}_{\mathbf{x}_{t|s}^*} + \Delta\mathbf{G}\mathbf{w}_{t|s}^* + \boldsymbol{\varepsilon}_{t|s} \\ &= \mathbf{x}_{t|s}^* + \Delta\mathbf{G}\mathbf{w}_{t|s}^* + \boldsymbol{\varepsilon}_{t|s} \end{aligned} \tag{23}$$

Here $\mathbf{x}_{t|s}$ indicates that $\mathbf{x}_t$ has been assigned to mode s and $\mathbf{x}_{t|s}^*$ denotes the expected normal replica of $\mathbf{x}_{t|s}$. Note that $\mathbf{x}_{t|s}$ is equal to $\mathbf{x}_{t|s}^*$ and $\Delta\mathbf{G} = 0$ if $\mathbf{x}_t$ is normal. Hence the faulty components are totally incorporated into the term $\Delta\mathbf{G}\mathbf{w}_{t|s}^*$. The reason for not using $\mathbf{F}$ to isolate the fault is that the term $\mathbf{F}\mathbf{h}_i$ represents the between-class variation. Once a test sample is assigned into a specific mode and determined to be faulty, the dominant variations will be focused on within-class variation.

In order to isolate faulty variables from a faulty sample $\mathbf{x}_t$, it is desired to obtain the estimation of matrix $\Delta\mathbf{G}$. An appropriate assumption on $\Delta\mathbf{G}$ is that it is a matrix with row-wise sparsity. That is to say, the nonzero rows of $\Delta\mathbf{G}$ correspond to faulty variables while the zero rows correspond to normal variables [31]. Let $\Delta\mathbf{G} = \left[\Delta\mathbf{g}_1^T \ \Delta\mathbf{g}_2^T \ \cdots \ \Delta\mathbf{g}_k^T \ \cdots \ \Delta\mathbf{g}_m^T\right]^T$, with $\Delta\mathbf{g}_k$ being the k-th row of $\Delta\mathbf{G}$. The row-sparsity assumption can be achieved by imposing a multivariate Laplace prior [30] on $\Delta\mathbf{g}_k$

$$\Delta\mathbf{g}_k \,| a_k, c_k \sim c_k \exp\left(-\sqrt{a_k}\|\Delta\mathbf{g}_k\|_2\right) \tag{24}$$

where a_k and c_k are unknown parameters. As pointed out by Babacan *et al.* [30], this prior is sharply peaked at zero vector, indicating that a row corresponding to normal variable will be shrunk to zero, whilst the rows corresponding to faulty variables will be nonzero. It is a common practice to decompose the multivariate Laplace prior into a hierarchical prior, consisting of a scaled Gaussian and two Gamma distributions [30]. Eqs.(25∼27) show how to decompose the multivariate Laplace prior.

$$\Delta\mathbf{g}_k \,| z_k \sim \mathcal{N}\left(\mathbf{0}, z_k\mathbf{I}\right) \tag{25}$$

$$z_k \,| a_k \sim \frac{a_k^{(D_G+1)/4} z_k^{(D_G-1)/2} \exp\left(-a_k z_k/2\right)}{2^{(D_G+1)/2} Ga\left((D_G+1)/2\right)} \tag{26}$$

$$a_k \,| \lambda, \eta \sim Gamma\left(\lambda, \eta\right) \tag{27}$$

where $Gamma\left(a_k \,| \lambda, \eta\right) = \eta^\lambda a_k^{\lambda-1} \exp\left(-\eta a_k\right) \Big/ Ga\left(\lambda\right)$ with the Gamma function, $Ga\left(x\right) = \int_0^\infty y^{x-1} \exp\left(-\,y\right) dy$ and $\{\lambda, \eta\}$ are pre-determined hyperparameters.

To get a full Bayesian inference framework for the SPGM, the priors for other parameters are given by

$$\mathbf{x}_{t|s}^* \sim \mathcal{N}\left(\boldsymbol{\mu} + \mathbf{F}\mathbf{h}_s + \mathbf{G}\mathbf{w}_{t|s}^*, \nu\mathbf{I}\right) \tag{28}$$

$$\nu \,| \rho, \kappa \sim Gamma\left(\rho, \kappa\right) \tag{29}$$

$$\mathbf{w}_{t|s}^* \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right) \tag{30}$$

$$\boldsymbol{\varepsilon}_{t|s} \,| \gamma \sim \mathcal{N}\left(\mathbf{0}, \gamma\mathbf{I}\right) \tag{31}$$

$$\gamma | \alpha, \beta \sim Gamma\left(\alpha, \beta\right) \tag{32}$$

where ν and γ are the variances of $\mathbf{x}_{t|s}^*$ and $\boldsymbol{\varepsilon}_{t|s}$ respectively. As suggested by Yang *et al.* [32], these noise variances can be estimated by placing different Gamma hyperpriors over ν and γ. For these Gamma hyperpriors, $\rho, \kappa, \alpha, \beta$ are pre-determined hyperparameters.

If a faulty sample causes violation in the *SPE* statistic, a similar probabilistic generative model can be considered as

$$\mathbf{x}_{t|s} = \mathbf{x}_{t|s}^* + \mathbf{r}_{t|s} + \boldsymbol{\varepsilon}_{t|s} \tag{33}$$

$$\mathbf{r}_{t|s} \,| \tau \sim \mathcal{N}\left(\mathbf{0}, \tau\mathbf{I}\right) \tag{34}$$

$$\tau \,| b \sim \frac{\sqrt{b}}{2} \exp\left(-b\tau/2\right) \tag{35}$$

In this case, the faulty components are incorporated in the vector $\mathbf{r}_{t|s}$. For the purpose of fault isolation, this time, a univariate Laplace prior can be imposed on each element of $\mathbf{r}_{t|s}$. Similarly, the Laplace prior can be decomposed hierarchically into a Gaussian distribution in Eq.(34) and a Gamma distribution in Eq.(35), with b being a pre-determined parameter. Faulty variables now correspond to the nonzero elements of $\mathbf{r}_{t|s}$ and normal variables correspond to zero elements.

2) Moving Window Technique for Online Application: For a robust fault isolation as well as to alleviate the effect of noise/disturbance, a moving window approach is commonly applied in fault isolation. Assume a window of l faulty samples causing violations in the monitoring statistics have been collected as $\tilde{\mathbf{x}}_{t|\cdot} = [\mathbf{x}_{t-l+1|\cdot}, \cdots, \mathbf{x}_{t|\cdot}]$, with "$\cdot$" representing the identified mode for a sample. Let $\tilde{\mathbf{x}}^*_{t|\cdot} = [\mathbf{x}^*_{t-l+1|\cdot}, \cdots, \mathbf{x}^*_{t|\cdot}]$ denote the expected normal replicas of $\tilde{\mathbf{x}}_{t|\cdot}$. The two probabilistic generative models described in Eq.(23) and Eq.(33) can be rewritten as follows.

$$\tilde{\mathbf{x}}_{t|\cdot} = \tilde{\mathbf{x}}^*_{t|\cdot} + \Delta\mathbf{G}\tilde{\mathbf{w}}^*_{t|\cdot} + \tilde{\boldsymbol{\varepsilon}}_{t|\cdot} \tag{36}$$

$$\tilde{\mathbf{x}}_{t|\cdot} = \tilde{\mathbf{x}}^*_{t|\cdot} + \mathbf{R} + \tilde{\boldsymbol{\varepsilon}}_{t|\cdot} \tag{37}$$

where $\tilde{\mathbf{w}}^*_{t|\cdot} = [\mathbf{w}^*_{t-l+1|\cdot}, \cdots, \mathbf{w}^*_{t|\cdot}]$, $\mathbf{R} = [\mathbf{r}_{t-l+1|\cdot}, \cdots, \mathbf{r}_{t|\cdot}]$, and $\tilde{\boldsymbol{\varepsilon}}_{t|\cdot} = [\boldsymbol{\varepsilon}_{t-l+1|\cdot}, \cdots, \boldsymbol{\varepsilon}_{t|\cdot}]$. Again, the rows of $\Delta\mathbf{G}$ and the elements of $\mathbf{R}$ admit multivariate and univariate Laplace priors , respectively. By including the latest and discarding the oldest sample, faulty variables can be isolated online by identifying the non-zero rows of $\Delta\mathbf{G}$ and $\mathbf{R}$ sequentially.

To estimate the model parameters, a variational Bayesian (VB) inference method is applied, which considers the complete log-likelihood of Eq.(36)

$$\begin{aligned}
&\ln p\left(\tilde{\mathbf{x}}_{t|\cdot}, \tilde{\mathbf{x}}^*_{t|\cdot}, \Delta\mathbf{G}, \tilde{\mathbf{w}}^*_{t|\cdot}, \gamma, \nu, a_k, z_k, k=1,\cdots,m\right) \equiv \\
&-\frac{1}{2}\sum_{j=t-l+1}^{t} \frac{1}{\gamma}\left\|\Delta\mathbf{x}_{j|\cdot} - \Delta\mathbf{G}\mathbf{w}^*_{j|\cdot}\right\|_2^2 + \frac{1}{\nu}\left\|\bar{\mathbf{x}}^*_{j|\cdot} - \mathbf{G}_\cdot\mathbf{w}^*_{j|\cdot}\right\|_2^2 \\
&+\left\|\mathbf{w}^*_{j|\cdot}\right\|_2^2 + \sum_{j=t-l+1}^{t}\sum_{k=1}^{m} -\frac{1}{2}\left\|\Delta\mathbf{g}_k\right\|_2^2 \Big/ z_k + \phi(z_k, a_k) + \varphi(\gamma, \nu)
\end{aligned} \tag{38}$$

where $\Delta\mathbf{x}_{j|\cdot} = \mathbf{x}_{j|\cdot} - \mathbf{x}^*_{j|\cdot}$, $\bar{\mathbf{x}}^*_{j|\cdot} = \mathbf{x}^*_{j|\cdot} - \boldsymbol{\mu} - \mathbf{F}\mathbf{h}_\cdot$, $\phi(z_k, a_k) = (\lambda + D_G/4 - 3/4)\ln a_k - (z_k/2 + \eta)a_k + (D_G/2 - 1/2)\ln z_k$, and $\varphi(\gamma, \nu) = (\alpha - l/2 - 1)\ln\gamma + (\rho - l/2 - 1)\ln\nu - \beta\gamma - \kappa\nu$. The posterior expectation of each latent variable can be obtained by taking the expectation of Eq.(38) on an approximate joint posterior distribution as

$$\begin{aligned}
&p\left(\tilde{\mathbf{x}}^*_{t|\cdot}, \tilde{\mathbf{w}}^*_{t|\cdot}, \Delta\mathbf{G}, \gamma, \nu, a_k, z_k, k=1,\cdots,m \,\middle|\, \tilde{\mathbf{x}}_{t|\cdot}\right) \\
&\approx q(\gamma)q(\nu)\prod_{j=t-l+1}^{t} q\left(\mathbf{x}^*_{j|\cdot}\right) q\left(\mathbf{w}^*_{j|\cdot}\right)\prod_{k=1}^{m} q(\Delta\mathbf{g}_k)\, q(a_k)\, q(z_k).
\end{aligned}$$

Using Bayesian rule, Eq.(39) updates the posterior expectation of $\Delta\mathbf{g}_k^T$ as

$$\begin{aligned}
\left\langle \Delta\mathbf{g}_k^T \right\rangle &= \boldsymbol{\Phi}_k \sum_{j=t-l+1}^{t} \left\langle \gamma^{-1} \right\rangle \left\langle \Delta x_{jk|\cdot} \right\rangle \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle \\
\boldsymbol{\Phi}_k &= \langle z_k \rangle \left[\mathbf{I} + \sum_{j=t-l+1}^{t} \left\langle \gamma^{-1} \right\rangle \langle z_k \rangle \left\langle \mathbf{w}^*_{j|\cdot} \mathbf{w}^{*T}_{j|\cdot} \right\rangle\right]^{-1}
\end{aligned} \tag{39}$$

where $\left\langle \Delta x_{jk|\cdot} \right\rangle$ is the k-th element of $\left\langle \Delta\mathbf{x}_{j|\cdot} \right\rangle = \mathbf{x}_{j|\cdot} - \left\langle \mathbf{x}^*_{j|\cdot} \right\rangle$. Similarly, the posterior expectations of $\mathbf{x}^*_{j|\cdot}$ and $\mathbf{w}^*_{j|\cdot}$ can be estimated by:

$$\left\langle \mathbf{x}^*_{j|\cdot} \right\rangle = \frac{1}{\vartheta}\left[\left\langle \gamma^{-1} \right\rangle \left\langle \hat{\mathbf{x}}_{j|\cdot} \right\rangle + \left\langle \nu^{-1} \right\rangle \left(\boldsymbol{\mu}_\cdot + \mathbf{G}_\cdot \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle\right)\right] \tag{40}$$

$$\left\langle \mathbf{w}^*_{j|\cdot} \right\rangle = \mathbf{P}_\cdot \left[\left\langle \gamma^{-1} \right\rangle \left\langle \Delta\mathbf{G}^T \right\rangle \left\langle \Delta\mathbf{x}_{j|\cdot} \right\rangle + \left\langle \nu^{-1} \right\rangle \mathbf{G}_\cdot^T \left\langle \bar{\mathbf{x}}^*_{j|\cdot} \right\rangle\right] \tag{41}$$

$$\left\langle \mathbf{w}^*_{j|\cdot} \mathbf{w}^{*T}_{j|\cdot} \right\rangle = \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle \left\langle \mathbf{w}^{*T}_{j|\cdot} \right\rangle + \mathbf{P}_\cdot \tag{42}$$

$$\left\langle \mathbf{x}^*_{j|\cdot} \mathbf{x}^{*T}_{j|\cdot} \right\rangle = \left\langle \mathbf{x}^*_{j|\cdot} \right\rangle \left\langle \mathbf{x}^{*T}_{j|\cdot} \right\rangle + \vartheta\mathbf{I} \tag{43}$$

where $\vartheta = \left\langle \gamma^{-1} \right\rangle + \left\langle \nu^{-1} \right\rangle$, $\boldsymbol{\mu}_\cdot = \boldsymbol{\mu} + \mathbf{F}\langle \mathbf{h}_\cdot \rangle$, $\left\langle \hat{\mathbf{x}}_{j|\cdot} \right\rangle = \mathbf{x}_{j|\cdot} - \langle \Delta\mathbf{G} \rangle \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle$, $\left\langle \bar{\mathbf{x}}^*_{j|\cdot} \right\rangle = \left\langle \mathbf{x}^*_{j|\cdot} \right\rangle - \boldsymbol{\mu}_\cdot$, and

$$\begin{aligned}
\left\langle \Delta\mathbf{G}^T \Delta\mathbf{G} \right\rangle &= \sum_{k=1}^{m} \left\langle \Delta\mathbf{g}_k^T \right\rangle \left\langle \Delta\mathbf{g}_k \right\rangle + \boldsymbol{\Phi}_k, \\
\mathbf{P}_\cdot &= \left(\mathbf{I} + \left\langle \gamma^{-1} \right\rangle \left\langle \Delta\mathbf{G}^T \Delta\mathbf{G} \right\rangle + \left\langle \nu^{-1} \right\rangle \mathbf{G}_\cdot^T \mathbf{G}_\cdot\right)^{-1}.
\end{aligned}$$

Finally, the posterior expectations of scaled factors and other parameters are obtained as follows.

$$\langle z_k \rangle = \sqrt{\langle a_k \rangle \big/ \mathrm{tr}\left(\left\langle \Delta\mathbf{g}_k^T \right\rangle \left\langle \Delta\mathbf{g}_k \right\rangle + \boldsymbol{\Phi}_k\right)} \tag{44}$$

$$\langle a_k \rangle = \left(\lambda + (D_G + 1)/2\right) \big/ \left(\eta + \langle z_k \rangle /2\right) \tag{45}$$

$$\left\langle \gamma^{-1} \right\rangle = \frac{ml/2 + \alpha}{\beta + \left\langle \left\|\tilde{\mathbf{x}}_{t|\cdot} - \tilde{\mathbf{x}}^*_{t|\cdot} - \Delta\mathbf{G}\tilde{\mathbf{w}}^*_{t|\cdot}\right\|_F^2 \right\rangle} \tag{46}$$

$$\left\langle \nu^{-1} \right\rangle = \frac{ml/2 + \rho}{\kappa + \left\langle \left\|\tilde{\mathbf{x}}^*_{t|\cdot} - \tilde{\boldsymbol{\mu}}_\cdot - \tilde{\mathbf{b}}_{t|\cdot}\right\|_F^2 \right\rangle} \tag{47}$$

where $\tilde{\mathbf{b}}_{t|\cdot} = \left[\mathbf{G}_\cdot\mathbf{w}^*_{t-l+1|\cdot}, \cdots, \mathbf{G}_\cdot\mathbf{w}^*_{t|\cdot}\right]$, $\tilde{\boldsymbol{\mu}}_\cdot = [\boldsymbol{\mu}_\cdot, \cdots, \boldsymbol{\mu}_\cdot]$, and

$$\begin{aligned}
&\left\langle \left\|\tilde{\mathbf{x}}_{t|\cdot} - \tilde{\mathbf{x}}^*_{t|\cdot} - \Delta\mathbf{G}\tilde{\mathbf{w}}^*_{t|\cdot}\right\|_F^2 \right\rangle = \\
&\left\|\tilde{\mathbf{x}}_{t|\cdot}\right\|_F^2 + \sum_{j=t-l+1}^{t} \mathrm{tr}\left[\left\langle \mathbf{x}^*_{j|\cdot} \mathbf{x}^{*T}_{j|\cdot} \right\rangle\right] + \mathrm{tr}\left[\left\langle \Delta\mathbf{G}^T \Delta\mathbf{G} \right\rangle \left\langle \mathbf{w}^*_{j|\cdot} \mathbf{w}^{*T}_{j|\cdot} \right\rangle\right] \\
&-2\left[\mathbf{x}^T_{j|\cdot}\left(\left\langle \mathbf{x}^*_{j|\cdot} \right\rangle + \langle \Delta\mathbf{G} \rangle \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle\right) - \left\langle \mathbf{x}^{*T}_{j|\cdot} \right\rangle \langle \Delta\mathbf{G} \rangle \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle\right], \\
&\left\langle \left\|\tilde{\mathbf{x}}^*_{t|\cdot} - \tilde{\boldsymbol{\mu}}_\cdot - \tilde{\mathbf{b}}_{t|\cdot}\right\|_F^2 \right\rangle = \\
&\left\|\tilde{\boldsymbol{\mu}}_\cdot\right\|_F^2 + \sum_{j=t-l+1}^{t} \mathrm{tr}\left[\left\langle \mathbf{x}^*_{j|\cdot} \mathbf{x}^{*T}_{j|\cdot} \right\rangle + \mathbf{G}_\cdot \left\langle \mathbf{w}^*_{j|\cdot} \mathbf{w}^{*T}_{j|\cdot} \right\rangle \mathbf{G}_\cdot^T\right] \\
&-2\left[\left\langle \mathbf{x}^{*T}_{j|\cdot} \right\rangle \left(\boldsymbol{\mu}_\cdot + \mathbf{G}_\cdot \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle\right) - \boldsymbol{\mu}_\cdot^T \mathbf{G}_\cdot \left\langle \mathbf{w}^*_{j|\cdot} \right\rangle\right].
\end{aligned}$$

In summary, the alternating updates of all listed posterior expectations constitute the VB algorithm. At each update step, the posterior distribution of each variable is inferred conditioned on the fixed distributions for the other variables. As for the estimation of $\mathbf{R}$, a similar VB inference algorithm can be developed.

3) Fault Scoring: To facilitate the characterization of different fault levels, two fault scores for each variable are defined as

$$\delta_k = \left\|\Delta\langle \mathbf{g}_k \rangle\right\|_1 \tag{48}$$

$$\tilde{\delta}_k = \left\|\langle \tilde{\mathbf{r}}_k \rangle\right\|_1 \tag{49}$$

where $\tilde{\mathbf{r}}_k$ is the k-th row of $\mathbf{R}$, $\|\cdot\|_1$ denotes the l_1-norm of a vector, δ_k corresponds to fault score obtained from faulty samples violating the T^2 statistic, and $\tilde{\delta}_k$ corresponds to those causing violations in the SPE. For faulty samples violating the T^2 statistics, a fault score vector $\delta = [\delta_1, \cdots, \delta_m]^T$ is obtained via Eq.(48), which can be normalized by its maximum value. If faulty samples violate the SPE statistics, another fault score vector can be obtained.

V. Numerical Simulation Example

This section illustrates the proposed process monitoring strategy based on I-PLDA. A process with six variables driven by two hidden variables is generated as follows.

$$\begin{aligned} \mathbf{x} &= \boldsymbol{\Omega}\boldsymbol{\upsilon} + \mathbf{e} \\ \boldsymbol{\Omega} &= [\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2] \\ \boldsymbol{\kappa}_1 &= [0, 0, 0.9835, 0.8979, 0, 0.7482]^T \\ \boldsymbol{\kappa}_2 &= [0.8921, 0.5856, 0, 0, 0.9154, 0.0581]^T \end{aligned} \tag{50}$$

Here $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6]^T$ are process variables to be monitored, 2 hidden variables are denoted as $\boldsymbol{\upsilon} = [\upsilon_1, \upsilon_2]^T$. The observation noise $\mathbf{e} = [e_1, e_2, e_3, e_4, e_5, e_6]^T$ follows a Gaussian distribution of $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(0.01 \times \mathbf{I}))$. To simulate a process with multiple operation modes, three different Gaussian distributions are imposed on the hidden variables as

Mode 1: $\upsilon_1 \sim \mathcal{N}(10, 0.8^2)$, $\upsilon_2 \sim \mathcal{N}(12, 1.3^2)$;
Mode 2: $\upsilon_1 \sim \mathcal{N}(5, 1.4^2)$, $\upsilon_2 \sim \mathcal{N}(20, 1.5^2)$;
Mode 3: $\upsilon_1 \sim \mathcal{N}(16, 2.0^2)$, $\upsilon_2 \sim \mathcal{N}(30, 2.5^2)$.

For the purpose of model training, 400 samples are generated from each mode and a total of 1200 training samples have been collected and stored in $\mathbf{X}_0 \in \mathscr{R}^{1200\times 6}$. The first 400 samples are generated from Mode 1, followed by the 800 samples generated from Mode 2 and Mode 3. In addition, a test dataset $\mathbf{X}_{f1} \in \mathscr{R}^{1200\times 6}$ is generated from the 3 modes in a similar way and a sensor fault is introduced in the hidden variable for samples from $101 \sim 400$ as

$$\text{Samples } 101 \sim 400: \ \mathbf{X}_{f1} = \boldsymbol{\Omega}\left(\boldsymbol{\upsilon} + [2, 0]^T\right) + \mathbf{e};$$

Based on the training data, Algorithm 1 is used to estimate the model parameters of the I-PLDA model. For parameters setting, the dimensions of loading matrices are set as $D_F = 2$ and $D_G = 6$. For other parameters, according to Yang *et al.* [32], large α and ρ encourage small noise variances and the hyperparameters setting includes $\alpha = \rho = 0.5$ and $\beta = \kappa = 10^{-6}$. Also, Yang *et al.* [32] demonstrates that a larger λ results in a more sparsity-encouraging prior and the similar settings for the other hyperparameters are $\lambda = 0.5$ and $\eta = 10^{-6}$.

The process monitoring strategy proposed in Section IV is now tested on the faulty samples. For fault detection, the monitoring statistics T^2 and SPE of both I-PLDA/PLDA and PPCA are provided. For fault isolation, based on the analysis by Liu *et al.* [31], the interval of $5 \le l \le 20$ for window length is suggested. Here it is set as $l = 10$ through experiments. For comparison, the standard PLDA, reconstruction based contribution (RBC) [33], and PPCA are considered.

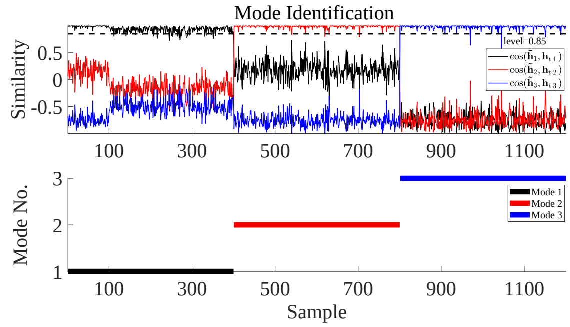

Fig. 1. Mode identification results based on I-PLDA for $\mathbf{X}_{f1}$

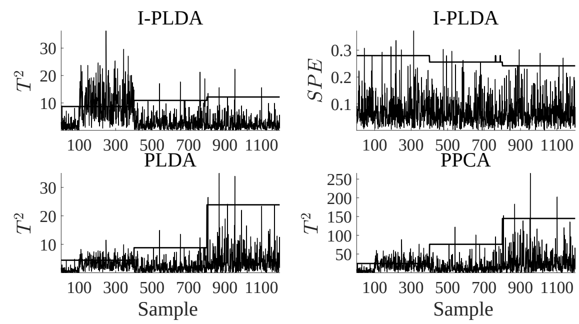

Fig. 2. Monitoring results using I-PLDA, PLDA, and PPCA for $\mathbf{X}_{f1}$

For each of the test samples, the mode identification method proposed in Section IV-A is applied and the results are shown in Figure 1. The upper plot of Figure 1 shows the cosine similarities, with the black line, red line, and blue line corresponding to those of Mode 1, Mode 2, and Mode 3, respectively. The lower plot shows the assigned mode for each sample. It can be seen from the upper plot of Figure 1 that the maximal cosine similarity always corresponds to the correct process mode. A clearer inspection shows that the cosine similarity $\cos(\tilde{\mathbf{h}}_1, \mathbf{h}_{t|1})$ fluctuated more significantly from the 101-st to the 400-th samples, this is due to the introduction of the sensor bias in the hidden variable. Despite the fluctuations, faulty samples can still be correctly assigned into Mode 1, which shows the good performance of the proposed method in mode assignment.

After the mode identification step, the monitoring results based on I-PLDA, PLDA, and PPCA are presented in Figure 2. The fault only affects the T^2 statistics and not the SPE statistics of all methods. Hence only the SPE statistics of I-PLDA are shown, and those of PLDA and PPCA are not shown as they do not produce significant alarms. As can be seen from Figure 2, significant number of violations are observed in the T^2 statistics of all three models after the 101-st sample, indicating the occurrence of a process fault in Mode 1. A clearer inspection of Figure 2, however, yields that higher sensitivity can be observed for I-PLDA. This is verified by the fault detection rates (FDR) of I-PLDA being 53.3%, comparing to those of 29.3% for PLDA and 43.3% for PPCA. This is expected, as the flexibility is enhanced by introducing a specific within-class loading matrix for each mode.

After the fault is successfully detected, it is essential to

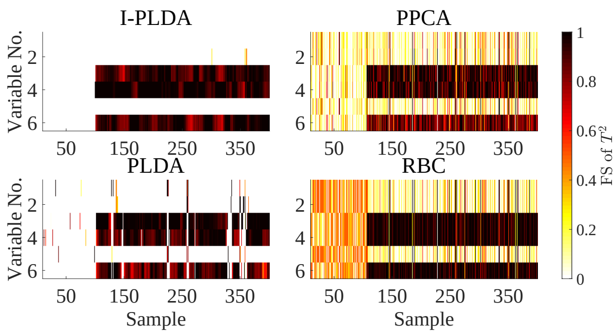

Fig. 3. Isolation results using I-PLDA, PLDA, PPCA, and RBC for $\mathbf{X}_{f1}$

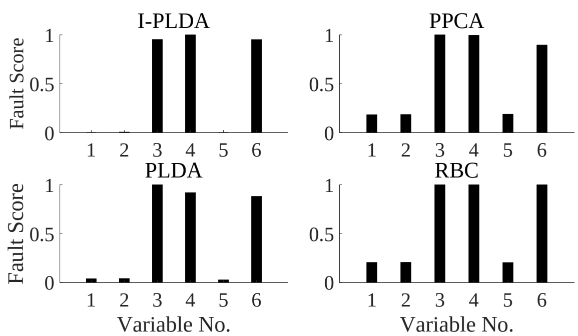

Fig. 4. Aggregated fault scores using I-PLDA, PLDA, PPCA, and RBC for $\mathbf{X}_{f1}$

localize the faulty variables. The moving window based fault isolation strategy proposed in Section IV-C is now applied on the 300 faulty samples and the results are shown in Figure 3, where deeper color indicates greater fault score. This time, comparison is made between I-PLDA, PLDA, PPCA, and RBC. In Figure 3, the sample-by-sample fault scores obtained from I-PLDA are recorded in the left-upper plot and those for PLDA, PPCA, and RBC are shown in the left-lower, right-upper, and right-lower plots, respectively. Figure 4 presents the aggregated fault scores for the 101 $\sim$ 400-th samples. Comparing the plots in Figure 3, one can observe that I-PLDA/PLDA shows better isolation capability. This can be confirmed by the plots of Figure 4 as the aggregated fault scores for I-PLDA/PLDA clearly indicate that x_3, x_4, and x_6 are faulty variables. In contrast, the aggregated contribution of PPCA and RBC also identify x_3, x_4, and x_6 as faulty, however, the contributions of x_1, x_2, and x_5 cannot be neglected.

VI. Application to Industrial Processes

This section demonstrates the performance of the proposed method using its applications to the Tennessee (TE) process and a blast furnace ironmaking process, in comparison with PLDA/PPCA. Comparisons with other methods showed similar results and hence are not included.

A. TE process application

The TE process involves five units including an exothermic reactor, a stripper, a flash separator, a recycle compressor, and a condenser. In the TE process, 41 measured variables and 12 manipulated variables are collected. In this paper, a total

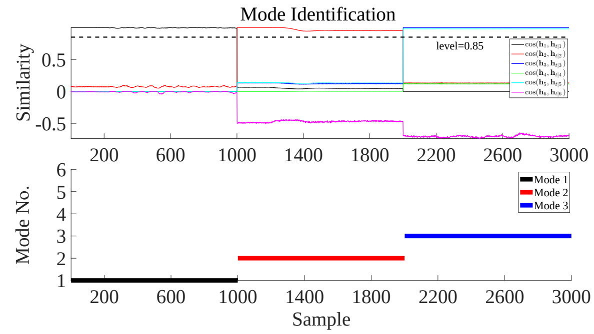

Fig. 5. Mode identification results based on I-PLDA for $\mathbf{X}_f$

of 16 continuous variables are selected and listed in Table I. According to [6], there are six normal operation conditions,

TABLE I
PROCESS VARIABLES AND DESCRIPTIONS

Variable	Description	Variable	Description
x_1	A Feed, *kscmh*	x_9	Separator temperature, °C
x_2	D Feed, kg/h	x_{10}	Separator pressure, *kPa*
x_3	E Feed, kg/h	x_{11}	Separator underflow, m^3/h
x_4	A+C Feed, *kscmh*	x_{12}	Stripper pressure, *kPa*
x_5	Recycle flow, *kscmh*	x_{13}	Stripper underflow, m^3/h
x_6	Reactor feed, *kscmh*	x_{14}	Stripper temperature, °C
x_7	Reactor temperature, °C	x_{15}	React. cool temperature, °C
x_8	Purge rate, *kscmh*	x_{16}	Cond. cool temperature, °C

each corresponding to a different process mode. To obtain the training dataset, a normal dataset $\mathbf{X}_0$ consisting of 3000 data points are sampled every 3 minutes under the 6 modes, so that each mode involves 500 samples. In addition, a faulty dataset $\mathbf{X}_f$ undergoing two random variations is introduced. For the sake of simplicity, the first three modes are included in $\mathbf{X}_f$, with each mode containing 1000 samples. For Mode 1, a Gaussian variation with zero mean and variance of 10 is added to the reactor temperature from 201-st time instance. For Mode 2, another Gaussian variation with zero mean and variances of 5 are added to both the purge rate and stripper pressure at the 1161-st sample. The parameters of I-PLDA are set as $D_F = 4$ and $D_G = 20$ via trial and error. The other parameters remain the same as before. For PPCA, the number of retained PCs is set as 4.

The I-PLDA parameters are then determined using Algorithm 1 based on $\mathbf{X}_0$ and the results of mode identification for $\mathbf{X}_f$ are shown in Figure 5. It can be seen from Figure 5 that the samples in $\mathbf{X}_f$ are successfully identified into the correct modes using the cosine similarity. As the mode assignments in the lower plot of Figure 5 are exactly the same as the initial mode settings.

Next, the monitoring procedures proposed in Section IV-B are used to detect the faults and the results are shown in Figure 6. It can be seen from Figure 6 that the fault of random variation in Mode 1 is successfully detected by the *SPE* statistics of both I-PLDA and PPCA. The random variation in Mode 2, on the other hand, are also successfully detected by both the T^2 and *SPE* statistics. For fault isolation, the procedures proposed in Section IV-C are used. This time, for simplicity, only the aggregated fault scores are presented in

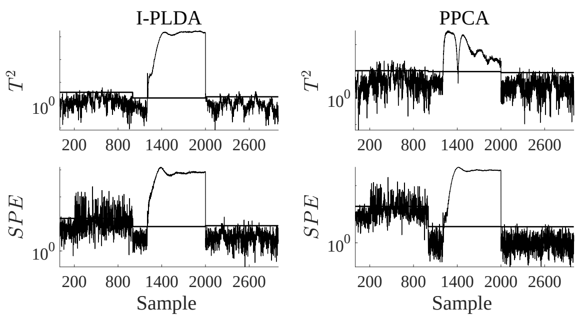

Fig. 6. Monitoring results using I-PLDA and PPCA for $\mathbf{X}_\mathrm{f}$

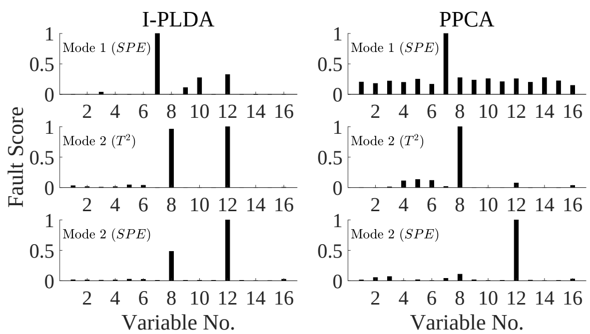

Fig. 7. Aggregated fault scores using I-PLDA and PPCA for $\mathbf{X}_\mathrm{f}$

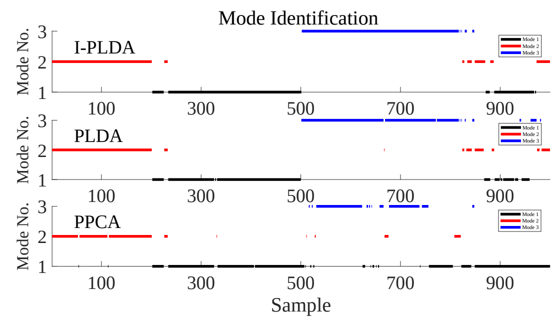

Fig. 8. Mode identification using I-PLDA, PLDA, and PPCA for blast furnace data

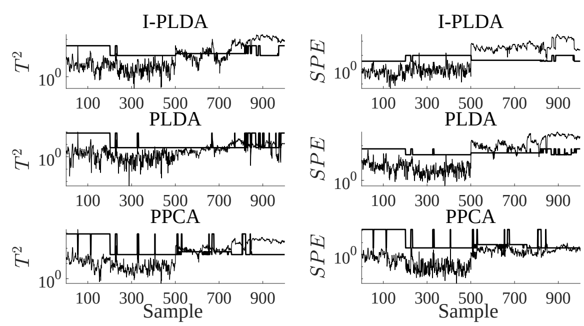

Fig. 9. Monitoring results of blast furnace fault using I-PLDA, PLDA, and PPCA

Figure 7. The fault isolation results in Figure 7 demonstrate that x_7 to be the dominant faulty variables in Mode 1 and x_8, x_{12} for Mode 2. A clearer inspection, however, shows that the contributions of other variables are not negligible for PPCA. This shows the better fault isolation performance of the proposed method.

B. Application to a blast furnace ironmaking process

A blast furnace is used to produce hot metal for steelmaking. During the operation of a blast furnace, iron ore and coke dropping from the top meet hot air and coal powder blowing from the bottom, resulting in a series of chemical reactions and gradually forming the product-liquid hot metal. Meanwhile, slag exits from the bottom and the flue gas escapes from the top. In this example, practical data collected from a blast furnace with 2500 m^3 in China is considered and a total of 10 process variables related to gas flow are considered and listed in Table II. The dataset consists of 1500 samples covering 3 different operation conditions and an abnormal dataset with 1000 samples. This fault involves a fluctuation in the CO concentration in the flue gas after the 500-th sample due to excessive blast of coal powder.

TABLE II
MONITORED BLAST FURNACE PROCESS VARIABLES

Variable	Description	Variable	Description
x_1	Quality of blast	x_6	Permeability index
x_2	Temperature of blast	x_7	Quantity of blasted oxygen
x_3	Pressure of blast	x_8	Coke ratio
x_4	Quantity of coal powder	x_9	CO concentration in the flue gas
x_5	Top gas pressure	x_{10}	CO_2 concentration in the flue gas

The sampling period for these samples is 2 min. Based on the normal dataset, the model parameters of I-PLDA are obtained with $D_F = 3$ and $D_G = 10$ via trial and error. The other parameters are the same as the TE process, and the retained PCs for PPCA is set as 3. After model training, the mode identification for the faulty dataset is carried out. For PPCA, the modes of test samples are identified using the maximum likelihood criterion. The results of mode identification using I-PLDA/PLDA and PPCA are shown in Figure 8, from which it can be seen that the mode identification results of all the three methods are very similar, indicating the identifications are appropriate. A clearer inspection shows that the identified modes of the $500 \sim 1000$-th samples shows higher fluctuation than the first 500 samples. This is expected, as the fault may affect the process characteristics and hence the mode identification. With the mode identified, fault detection are performed and the corresponding results are presented in Figures 9.

From the monitoring results in Figure 9, it can be seen that the fault is successfully detected by all three methods. The fault detection rates of the T^2 and SPE statistics for I-PLDA are 69.2% and 100%. In contrast, the FDRs of the T^2 and SPE statistics for PLDA are 35.2% and 96.2% and those for PPCA are 84% and 11.8% respectively. This can be explained, as the introduction of separate within-class loading matrices enhances the fault detection capability. After the fault is detected, fault isolation is performed. The sample-by-sample fault scores are shown in Figures 10 and the aggregated fault scores are shown in Figure 11. It is clearly shown in Figure 10 that I-PLDA identifies x_4 and x_9 as the dominant faulty variables in both T^2 and SPE statistics. This is in

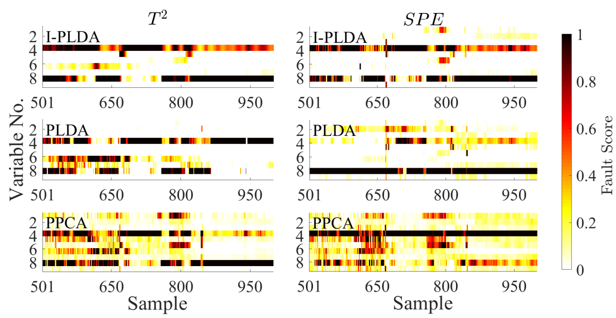

Fig. 10. Isolation results using I-PLDA, PLDA, and PPCA for blast furnace fault

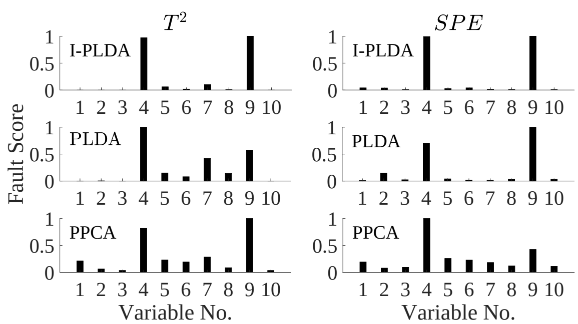

Fig. 11. Aggregated fault scores using I-PLDA, PLDA, and PPCA for blast furnace fault

accordance with the later finding that excessive coal injection caused increased CO concentration in the flue gas after the 500-th sample. In contrast, PLDA and PPCA also identify x_4 and x_9 as the dominant faulty variables. However, the contribution of other variables cannot be neglected. This is further confirmed by the aggregated fault scores shown in Figure 11.The application to blast furnace data demonstrates the advantages of the proposed method in monitoring of multi-mode process over PLDA and PPCA.

VII. Conclusions

This paper proposed a unified probabilistic monitoring framework for multi-mode processes based on probabilistic linear discriminant analysis. To better handle large within-class variance, an improved PLDA is developed by introducing a separate within-class loading matrix for each mode. For mode identification, the cosine similarity is applied and a fault detection and isolation strategy is proposed. The fault isolation procedures involve solution of two sparse probabilistic generative models. Application studies to simulation examples and industrial processes show the proposed method has better fault detection and isolation performance than competitive methods. Future work can be focused on extending the developed framework to monitoring of nonlinear processes using kernel trick.

References

[1] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.

[2] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics and Intelligent Laboratory Systems*, 2017.

[3] M. Quinones-Grueiro, A. Prieto-Moreno, C. Verde, and O. Llanes-Santiago, "Data-driven monitoring of multimode continuous processes: a review," *Chemometrics and Intelligent Laboratory Systems*, vol. 189, pp. 56–71, 2019.

[4] C. Tong, A. Palazoglu, and X. Yan, "An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding," *Journal of Process Control*, vol. 23, pp. 1497–1507, 2013.

[5] R. Grbic, D. Sliskovic, and P. Kadlec, "Adaptive soft sensor for online prediction and process monitoring based on a mixture of gaussian process models," *Computers and Chemical Engineering*, vol. 58, pp. 84–97, 2013.

[6] S. Tan, F. Wang, J. Peng, Y. Chang, and S. Wang, "Multimode process monitoring based on mode identification," *Industrial & Engineering Chemistry Research*, vol. 51, no. 1, pp. 374–388, 2011.

[7] Y. Zhang, Y. Fan, and N. Yang, "Fault diagnosis of multimode processes based on similarities," *IEEE Transactions on Industrial Electronics*, vol. 63, pp. 2606–2614, 2016.

[8] Y. Zhang and S. Li, "Modeling and monitoring between-mode transition of multimodes processes," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2248–2255, 2012.

[9] K. Zhang, K. Peng, and J. Dong, "A common and individual feature extraction-based multimode process monitoring method with application to the finishing mill process," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4841–4850, 2018.

[10] X. Peng, Y. Tang, W. Du, and F. Qian, "Multimode process monitoring and fault detection: a sparse modeling and dictionary learning method," *IEEE Transactions on Industrial Electronics*, vol. 64, pp. 4866–4875, 2017.

[11] W. Shao, Z. Ge, Z. Song, and J. Wang, "Semi-supervised robust modeling of multimode industrial processes for quality variable prediction based on student's t mixture model," *IEEE Transactions on Industrial Informatics*, 2019.

[12] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 3, pp. 560–572, 2018.

[13] X. Deng and J. Deng, "Incipient fault detection for chemical processes using two-dimensional weighted slkpca," *Industrial & Engineering Chemistry Research*, vol. 58, no. 6, pp. 2280–2295, 2019.

[14] Z. Ge, "Process data analytics via probabilistic latent variable models: A tutorial review," *Industrial & Engineering Chemistry Research*, vol. 57, no. 3, pp. 12 646–12 661, 2018.

[15] N. Sammaknejad, B. Huang, W. Xiong, A. Fatehi, F. Xu, and A. Espejo, "Operating condition diagnosis based on hmm with adaptive transition probabilities in presence of missing observations," *AIChE journal*, vol. 61, no. 2, pp. 477–493, 2015.

[16] K. H. Raveendran Rahul and H. Biao, "Process monitoring using a generalized probabilistic linear variable model," *Automatica*, vol. 96, pp. 73–83, 2018.

[17] J. Yu and S. J. Qin, "Multimode process monitoring with bayesian inference-based finite gaussian mixture models," *AIChE Journal*, vol. 54, no. 7, pp. 1811–1829, 2008.

[18] Z. Ge and Z. Song, "Mixture bayesian regularization method of ppca for multimode process monitoring," *AIChE journal*, vol. 56, no. 11, pp. 2838–2849, 2010.

[19] M. Fang, H. Kodamana, B. Huang, and N. Sammaknejad, "A novel approach to process operating mode diagnosis using conditional random fields in the presence of missing data," *Computers and Chemical Engineering*, vol. 111, no. 3, pp. 149–163, 2018.

[20] J. Yu, "A new fault diagnosis method of multimode processes using bayesian inference based gaussian mixture contribution decomposition," *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 456–466, 2013.

[21] H. Gharahbagheri, S. A. Imtiaz, and F. Khan, "Root cause diagnosis of process fault using kpca and bayesian network," *Industrial and Engineering Chemistry Research*, vol. 56, pp. 2054–2070, 2017.

[22] B. Rashidi, D. S. Singh, and Q. Zhao, "Data-driven root-cause fault diagnosis for multivariate non-linear processes," *Control Engineering Practice*, vol. 70, pp. 134–147, 2018.

[23] T. Chen and Y. Sun, "Probabilistic contribution analysis for statistical process monitoring: a missing variable approach," *Control Engineering Practice*, vol. 17, pp. 469–477, 2009.

[24] A. Haghani, T. Jeinsch, S. X. Ding, P. Koschorrek, and B. Kolewe, "A probabilistic approach for data-driven fault isolation in multimode

processes," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 8909–8914, 2014.

[25] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[26] K. Grauman and B. Leibe, *Visual Object Recognition*. Morgan & Claypool, 2011.

[27] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, pp. 47–60, 1996.

[28] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[29] S. Prince, P. Li, Y. Fu, U. Mohammed, and J. Elder, "Probabilistic models for inference about identity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 144–157, 2012.

[30] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE transactions on signal processing*, vol. 62, no. 11, pp. 2906–2921, 2014.

[31] Y. Liu, J. Zeng, L. Xie, S. Luo, and H. Su, "Structured joint sparse principal component analysis for fault detection and isolation," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2721–2731, 2018.

[32] L. Yang, J. Fang, H. Cheng, and H. Li, "Sparse bayesian dictionary learning with a gaussian hierarchical model," *Signal Processing*, vol. 130, pp. 93–104, 2017.

[33] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, pp. 1593–1600, 2009.