Dynamic probabilistic predictable feature analysis for high dimensional temporal monitoring

A Preprint

Wei Fan
Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education
Southeast University
Nanjing 210096, China
wfan@seu.edu.cn

Qinqin Zhu*
Department of Chemical Engineering
University of Waterloo
ON N2L 3G1, Canada
qinqin.zhu@uwaterloo.ca

Shaojun Ren*
Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education
Southeast University
Nanjing 210096, China
rsj@seu.edu.cn

Liang Zhang
School of Instrument Science and Engineering
Key Laboratory of Micro-Inertial Instrument and Advanced Navigation Technology
of Ministry of Education
Southeast University
Nanjing 210096, China
zhangliang418@seu.edu.cn

Fengqi Si
Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education
Southeast University
Nanjing 210096, China
fqsi@seu.edu.cn

October 1, 2021

ABSTRACT

Dynamic statistical process monitoring methods have been widely studied and applied in modern industrial processes. These methods aim to extract the most predictable temporal information and develop the corresponding dynamic monitoring schemes. However, measurement noise is widespread in real-world industrial processes, and ignoring its effect will lead to sub-optimal modeling and monitoring performance. In this article, a probabilistic predictable feature analysis (PPFA) is proposed for high dimensional time series modeling, and a multi-step dynamic predictive monitoring scheme is developed. The model parameters are estimated with an efficient expectation-maximum algorithm, where the genetic algorithm and Kalman filter are designed and incorporated. Further, a novel dynamic statistical monitoring index, Dynamic Index, is proposed as an important supplement of $T^2$ and_SPE to detect dynamic anomalies. The effectiveness of the proposed algorithm is demonstrated via its application on the three-phase flow facility and a medium speed coal mill.

Keywords Dynamic process monitoring, probabilistic predictable feature analysis, EM algorithm, genetic algorithm, Kalman filter.

*co-corresponding authors, equal contribution
1 Introduction

In the current era of big data, industrial processes are equipped with a large number of sensors to measure different process variables. At the same time, massive amounts of historical data are collected and stored. On this basis, data-driven process monitoring has become a popular research topic due to its reliable performance and easy-to-implement characteristics. Multivariate statistical process monitoring method is a representative kind, including principal component analysis (PCA) and canonical component analysis. As a dimensionality reduction algorithm, PCA decomposes the measurement space into principal component subspace and residual subspace, and realizes industrial process monitoring by designing relevant statistical indices. However, traditional PCA does not take temporal information into account, and thus tends to obtain sub-optimal performance.

To tackle the inevitable dynamics in the data samples, several extensions have been proposed. Ku et al. proposed a dynamic PCA (DPCA), which employs augmented measurements with time lags and tries to explore the serial correlations between current and previous observations. With the derived relations, statistical indices such as $T^2$ and SPE are employed to monitor abnormal condition of industrial processes. Rato and Reis designed a dynamic PCA (DiPCA) method to capture the most dynamic variations from time series data. Similarly, Richthofer and Wiskott proposed the predictable feature analysis (PFA) to extract dynamic latent variables that are as predictable as possible. Both DiPCA and PFA are efficient auto-regressive models, and their difference lies in the design of their objective functions. DiPCA builds a model by maximizing the covariance between the actual value and estimated value of the latent variable, while PFA aims to minimize the auto-regressive prediction error of the latent variable.

In practical engineering applications, the variables are inevitable to be polluted by random noise, which, however, are not considered in the aforementioned methods. To provide the complete distribution of the data, the dynamic characteristics of process variables should be extracted through statistical patterns rather than deterministic manner. Based on the above analysis, this article is dedicated to exploring the integration of probabilistic models and traditional VAR-based methods, and applying them for dynamic predictive process monitoring. Inspired by the state space expressions designed in probabilistic PCA (PPCA) and probabilistic slow feature analysis (PSFA), we extend DiPCA and PFA to a probabilistic structure, referred to as probabilistic predictable feature analysis (PPFA). PPFA takes a high-order linear Markov Gaussian state-space form, and multi-step time lags of latent variables are involved in PPFA. It is worthwhile to point out that PPFA has intrinsic advantages over DiPCA and PFA. First, deterministic methods fail to grasp the distribution of measurement noise, while PPFA overcomes this problem with a fully probabilistic framework. On the other hand, based on the established PPFA model, a novel monitoring statistic, Dynamic Index (DI), can be derived to demonstrate the dynamic changes of studied systems, thereby providing a reliable guidance for improving control performance. The Expectation-Maximization (EM) algorithm and Kalman filter are adopted to estimate the model parameters of PPFA. The main contributions of this work are

- A novel probabilistic extension of DiPCA and PFA, termed as PPFA, is designed, which provides a full interpretation of the dynamic characteristics of both measurements and latent variables.
- Multiple time lags are designed in PPFA to capture the actual dynamics involved in the data.
- In M-step of the EM algorithm, the genetic algorithm (GA) is employed to optimize the weight coefficients of latent variables, which are difficult to be solved analytically.
- During the procedure of E-step, an expansion method of latent structure is proposed to estimate relevant expectations of the high-order dynamic system with Kalman filter.
- A novel monitoring statistic, Dynamic Index (DI), is defined as an important supplement of $T^2$ and SPE for timely anomaly detection in dynamic systems.

The rest of this paper is organized as follows. Section 2 presents a brief introduction of DiPCA and PFA. In Section 3, the detailed derivation of PPFA as well as the novel PPFA-based dynamic process monitoring framework are
where \( \langle \cdot \rangle \)

where \( t \)

\[ t_k = \sum_{j=1}^{s} \beta_j t_{k-j} + e_k \]  

where \( t_k = x_k^T w \) is the latent score for the observed sample \( x_k \) at time \( k \), \( w \) is the weight vector, \( e_k \) is the modeling error, \( s \) is the time lag, and \( \beta_j \) is the auto-regressive coefficient. It is assumed that \( e_k \) is white noise if \( s \) is long enough, such that the estimated prediction of latent variable is expressed as

\[ \hat{t}_k = x_{k-1}^T w \beta_1 + \cdots + x_{k-s}^T w \beta_s \]

\[ = [x_{k-1} \cdots x_{k-s}] (\beta \otimes w) \]  

where \( \beta = [\beta_1, \beta_2, \ldots, \beta_s]^T \), and \( \beta \otimes w \) is the Kronecker product.

Denote the data matrix as \( X = [x_1^T, x_2^T, \cdots, x_n^T]^T \in \mathbb{R}^{(n+s) \times m} \), and define a new data matrix containing temporal information \( Z_s = [X_1, X_2, \cdots, X_s] \), where \( X_j = [x_j^T, x_{j+1}^T, \cdots, x_{n+j-1}^T]^T, j = 1, 2, \cdots, s \). Then the objective of DiPCA in Eq. (1) can be represented as the following matrix form.

\[ \max_{w, \beta} w^T X_{s+1}^T Z_s (\beta \otimes w) \]

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

More details of extracting dynamic components for DiPCA can be found in ref. [Dong and Qin, 2018].

2.2 Predictable Feature Analysis

Similar to DiPCA, PFA is also an auto-regressive model, and it focuses on measuring the predictability of the extracted latent variables [Richthofer and Wiskott, 2015]. PFA defines that a good predictability is achieved when the latent score can be well predicted by a linear combination of \( s \) past values. Given an \( m \)-dimensional temporal measurement \( x_s \), PFA aims to find an orthogonal transformation \( W \in \mathbb{R}^{m \times r} \), such that the projection \( t_k = x_k^T W \) obtains the highest predictability. Mathematically, given the coefficient matrices \( B_i \in \mathbb{R}^{r \times r}, 1 \leq i \leq s \), the prediction of \( t_k \) is

\[ \hat{t}_k = \sum_{i=j}^{s} B_j t_{k-j} = \sum_{j=1}^{s} B_j W^T x_{k-j} \]  

where the modeling error is omitted. It is observed that both Eq. (1) and Eq. (4) share similar structure, and both of them aim to extract the most predictable latent variables. Their main difference is that the goal of DiPCA is to maximize the covariance between \( t_k \) and \( \hat{t}_k \), while PFA solves the auto-regressive issue by minimizing the prediction error, which is expressed as

\[ \min_{W} \langle \| W^T x_k - \sum_{j=1}^{s} B_j W^T x_{k-j} \|^2 \rangle \]  

where \( \langle \cdot \rangle \) means the average of signals over time. Readers can refer to ref. [Richthofer and Wiskott, 2015] for the detailed PFA algorithm.

3 Methodology

3.1 Probabilistic Predictable Feature Analysis

In real-world industrial processes, the variables are inevitable to be polluted by random noise, which are not considered in both DiPCA and PFA. To improve the prediction performance, the dynamic characteristics of process variables demonstrated. Then, experiments on the three-phase flow facility and a medium speed coal mill are presented to testify the effectiveness of the proposed PPFA-based modeling and monitoring method. Finally, conclusions are drawn in Section 5.
should be extracted through statistical patterns rather than deterministic manner [Zheng et al., 2016]. Therefore, this work is to extend DiPCA and PFA to a probabilistic structure with a state-space form, referred to as probabilistic predictable feature analysis (PPFA), which describes the process dynamics in a more compact and clearer manner.

Given the collected samples \( X = [x_1^T, x_2^T, \ldots, x_{n+s}^T] \), PPFA takes the following generative state-space form

\[
\begin{align*}
\{ & t_k = \sum_{j=1}^{s} B_j t_{k-j} + e_k \\
& x_k = H t_k + \varepsilon_k \\
\text{s.t.} & \ E [t_k] = 0, \ E [t_k t_k^T] = \mathbb{I}_r
\end{align*}
\] (6)

where \( H \in \mathbb{R}^{m \times r} \) represents the emission matrix, \( B_j = \text{diag} \{ \beta_j^1, \ldots, \beta_j^r \} \), \((1 \leq j \leq s)\) denotes the transition matrix, \( e_k \sim \mathcal{N}(0, \Gamma) \) with \( \Gamma = \text{diag} \{ \tau^2_1, \ldots, \tau^2_r \} \) is the Gaussian distributed noise, \( \varepsilon_k \sim \mathcal{N}(0, \Sigma) \) with \( \Sigma = \text{diag} \{ \sigma^2_1, \ldots, \sigma^2_r \} \) is the measurement noise, and \( \mathbb{I}_r \) is an \( r \times r \) identity matrix.

**Lemma 1** Given the constraints in Eq. (6) that \( E (t_k) = 0 \) and \( E (t_k t_k^T) = \mathbb{I}_r \), the relation between \( B_j \) and \( \Gamma \) is

\[
\tau^2_i = 1 - \sum_{j=1}^{s} \beta_j^i \gamma_j^i \geq 0, \quad 1 \leq i \leq r
\] (7)

where \( \gamma_j^i = \text{cov} (t_k^i, t_{k-j}^i) = \text{cov} (t_{k+j}^i, t_k^i) \) is the autocovariance, and \( t_k^i \) with \( 1 \leq i \leq r \) represents the \( i^{th} \) latent variable at time \( k \).

The proof of Lemma 1 is given in APPENDIX A. For PPFA, the parameters needed to be estimated are summarized as \( \Theta = \{ B_j \ (1 \leq j \leq s), \ H, \ \Gamma, \ \Sigma \} \).

The maximum-likelihood method is widely used for parameter optimization of probabilistic models. Given the temporal series \( X \), the complete data log-likelihood of the dynamic system in Eq. (6) can be given by

\[
L (\Theta) = \log p (X, T | \Theta)
\]

\[
= \log \prod_{k=s+1}^{n+s} \left[ p (t_k | \{ t_{k-j} \}_{j=1}^{s}) p (x_k | t_k) \right]
\times \prod_{k=1}^{s} [p (t_k) p (x_k | t_k)]
\]

\[
= \sum_{k=1}^{s} \log p (t_k) + \sum_{k=1}^{n+s} \log p (x_k | t_k)
+ \sum_{k=s+1}^{n+s} p (t_k | \{ t_{k-j} \}_{j=1}^{s})
\] (8)

where \( T = [t_1^T, t_2^T, \ldots, t_{n+s}^T] \in \mathbb{R}^{(n+s) \times r} \) denotes the latent variables matrix.

For each sample \( x_k \), based on the property of conditional independence, the corresponding high-order linear Markov Gaussian dynamic system \( p \left( t_k | \{ t_{k-j} \}_{j=1}^{s} \right) \) in Eq. (8) is expressed as

\[
p \left( t_k | \{ t_{k-j} \}_{j=1}^{s} \right) = p \left( t_k | \{ t_{k-j} \}_{j=1}^{s}, \{ B_j \}_{j=1}^{s} \Gamma \right)
= \prod_{i=1}^{r} p \left( t_k^i | \{ t_{k-j}^i \}_{j=1}^{s}, \{ \beta_j^i \}_{j=1}^{s} \tau^2_i \right)
\] (9)

where

\[
p \left( t_k^i | \{ t_{k-j}^i \}_{j=1}^{s}, \{ \beta_j^i \}_{j=1}^{s} \tau^2_i \right) = \mathcal{N} \left( \sum_{j=1}^{s} \beta_j^i t_{k-j}^i, \tau^2_i \right)
\]

\[
p (t_1) = p (t_2) = \cdots = p (t_s) = \mathcal{N}(0, \mathbb{I}_r)
\] (11)
The probabilistic distribution of the mapping from \( t_k \) to \( x_k \) in Eq. (8) is

\[
p(x_k|t_k) \equiv p(x_k|t_k, H, \Sigma) = \mathcal{N}(Ht_k, \Sigma)
\]

The complete specification of PPFA is defined in Eqs. (8) to (12), and the parameter set \( \Theta \) can be obtained via maximizing \( L(\Theta) \) in Eq. (8).

### 3.2 Parameter Estimation Scheme

The EM algorithm is a powerful method to solve the parameter estimation problem of probabilistic generative models through two steps of continuous iteration, namely E-step and M-step. In this section, we give detailed derivation process of EM algorithm applied in PPFA. The main body of parameter estimation process is the EM algorithm; however, it is difficult to obtain explicit analytical solutions for parameters \( \beta_j^i \). Thus, GA algorithm is embedded into the EM process to construct the complete parameter optimization strategy.

#### 3.2.1 EM Algorithm

Substituting the probability density function in Eqs. (9)-(12) into Eq. (8), we have

\[
L(\Theta) = -\frac{(n+s)(r+m)}{2} \log(2\pi) - \frac{1}{2} \sum_{k=1}^{s} (t_k^\top t_k) \\
- n \sum_{i=1}^{r} \log(\tau_i) - \frac{1}{2} \sum_{k=1}^{s} \left[ (x_k - Ht_k)^\top \Sigma^{-1} (x_k - Ht_k) \right] \\
- \frac{(n+s)}{2} \log|\Sigma| - \frac{1}{2} \sum_{k=s+1}^{s+r} \sum_{i=1}^{s} \frac{1}{\tau_{k-i}} \left[ \tau_k^i - \sum_{j=1}^{s} \beta_j^i \right]
\]

The \( Q \)-function is defined conditioned on the old parameter set \( \Theta^{old} \)

\[
Q(\Theta, \Theta^{old}) = \mathbb{E}_{X, \Theta^{old}} \{ L(\Theta) \}
\]

Then, in the M-step, new parameters are estimated through

\[
\Theta^{new} = \arg \max_{\Theta} Q(\Theta, \Theta^{old})
\]

Since \( B_j \) (\( 1 \leq j \leq s \)) and \( \Gamma \) are related to each other as shown in Lemma [7], the parameter set to be estimated can be simplified to \( \Theta = \{B_j, 1 \leq j \leq s, H, \Sigma\} \). Then take derivatives of \( Q \)-function with respect to different parameters in \( \Theta \). First of all, for \( \beta_j^i \) in \( B_j \), taking the derivation leads to

\[
\frac{\partial Q(\Theta, \Theta^{old})}{\partial \beta_j^i} = \frac{n \gamma_j^i}{2 \left( 1 - \sum_{l=1}^{s} \beta_l^i \gamma_l^i \right)} - \frac{1}{2} \sum_{k=s+1}^{s+r} \sum_{l=1}^{s} \beta_l^i t_{k-l}^i - \frac{n+s}{2} \sum_{k=s+1}^{s+r} \beta_l^i t_{k-l}^i = 0
\]

Then setting the above equation to zero, we have

\[
\left[ n \gamma_j^i - 2 \sum_{l=1}^{s} \beta_l^i \sum_{k=s+1}^{s+r} t_{k-l}^i t_{k-l}^i + 2 \sum_{k=s+1}^{s+r} t_{k-l}^i t_{k-l}^i \right] \\
\times \left( 1 - \sum_{l=1}^{s} \beta_l^i \gamma_l^i \right) = 2 \sum_{l=1}^{s} \sum_{g=l+1}^{s} \beta_l^i \beta_g^j \times \sum_{k=s+1}^{s+r} (t_{k-l}^i t_{k-g}^i) \\
+ \gamma_j^i \sum_{k=s+1}^{s+r} (t_{k-l}^i)^2 - 2 \sum_{l=1}^{s} \beta_l^i \sum_{k=s+1}^{s+r} t_{k-l}^i t_{k-l}^i + \sum_{l=1}^{s} (\beta_l^i)^2 \sum_{k=s+1}^{s+r} (t_{k-l}^i)^2
\]

It is difficult to derive the analytical solution for Eq. (15). Alternatively, GA algorithm is employed to simplify the derivation, which is demonstrated in the next subsection.
For parameter matrix $H$, differentiating $Q(\Theta, \Theta_{\text{old}})$ with respect to it and setting it to zero result in

$$H_{\text{new}} = \left( \sum_{k=1}^{n+s} x_k^t t_k^\top \right) \left( \sum_{k=1}^{n+s} t_k t_k^\top \right)^{-1}$$

(16)

Similarly, the covariance parameter $\Sigma$ is updated with

$$\Sigma_{\text{new}} = \frac{1}{n + s} \sum_{k=1}^{n+s} (x_k - H_{\text{new}} t_k) (x_k - H_{\text{new}} t_k)^\top$$

(17)

It is noted that $\Sigma = \{\sigma_1^2, \cdots, \sigma_m^2\}$, and each $\sigma_i^2$ in $\Sigma$ can be calculated by

$$\sigma_i^2 = \frac{1}{n + s} \sum_{k=1}^{n+s} \left\{ (x_k^i)^2 - 2 t_k^\top x_k^i + t_k^\top t_k h_i \right\}$$

(18)

where $h_i = [h_{i1}, \cdots, h_{it}]^\top$ is the $i^{th}$ row of $H_{\text{new}}$.

During the process of M-step, the following expectations with respect to the latent variables $t_k$ should be evaluated, which will be further utilized in the E-step given the old parameter set $\Theta_{\text{old}}$.

$$\mathbb{E}_{t_k|\Theta, \Theta_{\text{old}}} [t_k]
\mathbb{E}_{t_k|\Theta, \Theta_{\text{old}}} [t_k t_k^\top]
\mathbb{E}_{t_k|\Theta, \Theta_{\text{old}}} [t_k t_{k-l}^\top], 1 \leq l \leq s
\mathbb{E}_{t_k|\Theta, \Theta_{\text{old}}} [t_{k-g} t_{k-g}^\top], 1 \leq l, g \leq s$$

(19)

Eq. (19) will be further investigated in Subsection "Expectation Estimation Strategy".

### 3.2.2 Genetic Algorithm

The genetic algorithm is adopted to get the values of $\beta_j^i$ in Eq. (15). GA is an adaptive heuristic optimization algorithm proposed on the basis of natural selection in the theory of evolution [Goldberg and Holland, 1988]. The main idea of GA is to model the system that satisfies the natural evolution conditions, and solve the optimization problem by randomly searching the pre-constructed search space. During the iterative process, the population is mutated and reorganized, and the fitness function is used to evaluate the reliability of individuals. In the end, individuals who adapt to the system environment are evolved, and the best one is selected as the parameter candidate [Garg, 2016].

To estimate $\beta_j^i$, the optimization model is constructed as follows. Based on Eq. (15), two reference indices $A$ and $B$ are defined as follows

$$A = \left[ \frac{n \gamma_j^i - 2 \sum_{l=1}^{s} \beta_l^i \sum_{k=s+1}^{n+s} t_{k-j}^i t_{k-1-j}^i + 2 \sum_{k=s+1}^{n+s} t_k^i t_{k-j}^i}{1 - \sum_{l=1}^{s} \beta_l^i \gamma_l^i} \right]$$

(20)

$$B = 2 \sum_{l=1}^{s} \sum_{g=l+1}^{s} \beta_l^i \beta_g^i \sum_{k=s+1}^{n+s} \left( t_{k-j}^i t_{k-g}^i \right)$$

$$+ \gamma_j^i \left[ \sum_{k=s+1}^{n+s} \left( t_k^i \right)^2 - 2 \sum_{l=1}^{s} \beta_l^i \sum_{k=s+1}^{n+s} t_k^i t_{k-j}^i \right.$$  

$$+ \sum_{l=1}^{s} \left( \beta_j^i \right)^2 \sum_{k=s+1}^{n+s} \left( t_{k-j}^i \right)^2 \right]$$

(21)

Combined with Lemma 2, the optimization problem for $\beta_j^i$ is formulated as

$$\min_{\beta_j^i} f(\beta_j^i) = (A - B)^2$$

s.t. $1 - \sum_{j=1}^{s} \beta_j^i \gamma_j^i \geq 0$

(22)
To handle the inequality constraints, we further combine the Lagrange method to adjust the objective function as follows.

$$\min_{\beta_j^i} g(\beta_j^i) = \begin{cases} f(\beta_j^i), & 1 - \sum_{j=1}^{s} \beta_j^i \gamma_j^i \geq 0 \\ f(\beta_j^i) - \lambda \left(1 - \sum_{j=1}^{s} \beta_j^i \gamma_j^i\right), & \text{otherwise} \end{cases} \quad (23)$$

where $\lambda$ denotes a regularization coefficient which is greater than zero. The detailed modeling procedure of GA is provided in ref. [Goldberg and Holland, 1988].

### 3.2.3 Expectation Estimation Strategy

For a general first-order linear dynamic system, as proposed by Shang et al. [Shang et al., 2015], the expectations in Eq. (6) can be easily estimated with Kalman smoothing [Bishop, 2006]. Typically, the forward recursion and the backward recursion are employed to solve the problem, which, however, cannot be directly used to estimate the desired expectations for the high-order system proposed in this work.

To address this issue, we expand the dimension of the dynamic system designed in Eq. (6) as follows, and adjust it to the general form of a first-order linear dynamic system. Therefore, the general Kalman smoothing method can be easily adapted to perform the E-step.

$$t_k = [t_k^T, t_{k-1}^T, \ldots, t_{k-s+1}^T]^T \in \mathbb{R}^{rs},$$

$$e_k = [e_k^T, 0, \ldots, 0]^T \in \mathbb{R}^{rs},$$

$$\Phi_k = \begin{bmatrix} B_1, & B_2, & \ldots, & B_{s-1}, & B_s \\ I, & 0, & \ldots, & 0, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots, & \vdots \\ 0, & 0, & \ldots, & I, & 0 \end{bmatrix} \in \mathbb{R}^{rs \times rs},$$

$$H_k = [H, 0, \ldots, 0] \in \mathbb{R}^{m \times rs},$$

$$\Gamma_k = \begin{bmatrix} T, & 0, & \ldots, & 0, & 0 \\ 0, & 0, & \ldots, & 0, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots, & \vdots \\ 0, & 0, & \ldots, & 0, & 0 \end{bmatrix} \in \mathbb{R}^{rs \times rs} \quad (25)$$

Thereafter, the dynamic system proposed in Eq. (6) is adapted into

$$\begin{cases} t_k^s = \Phi_k t_{k-1}^s + e_k^s \\ x_k = H_k t_k^s + e_k \end{cases} \quad (26)$$

where the Gaussian noise $e_k^s \sim \mathcal{N}(0, \Gamma_k)$.

Correspondingly, the initial conditions given by Eq. (11) become

$$t_0^s = [t_0^T, t_1^T, \ldots, t_1^T] \sim \mathcal{N}(0, I_{rs}) \quad (27)$$

where $I_{rs} \in \mathbb{R}^{rs \times rs}$ is an identity matrix.

After the modification, the forward recursions of Kalman smoothing [Bishop, 2006] are first applied to estimate the posterior distribution of $t_k^s$ given the total past-time $x_k$ and $\Theta^{old}$, termed as $p(t_k^s|x_0, \ldots, x_k, \Theta^{old}) \sim \mathcal{N}(\mu_k, V_k)$, sequentially.

$$\mu_k = \Phi_k \mu_{k-1} + K_k (x_k - H_k \Phi_k \mu_{k-1}),$$

$$P_{k-1} = \Phi_k V_{k-1} \Phi_k^T + \Gamma_k,$$

$$V_k = (I - K_k H_k) P_{k-1},$$

$$K_k = P_{k-1} H_k^T (H_k P_{k-1} H_k^T + \Sigma)^{-1} \quad (28)$$

with the initialization

$$\mu_s = K_1 x_s,$$

$$V_s = I - K_1 H_s,$$

$$K_s = H_s^T (H_s H_s^T + \Sigma)^{-1} \quad (29)$$
where $\mu_k$ and $V_k$ represent the mean vector and covariance matrix of the posterior distribution, respectively. $P_k$ is the error covariance matrix of state estimate $t_{s_k}^1$, and $K_k$ denotes the Kalman gain matrix.

Afterward, utilizing the backward recursions [Bishop 2006] to further estimate the parameters of posterior distribution $p(T|X, \Theta^{old})$.

\[ \hat{\mu}_k = \mu_k + J_k (\hat{\mu}_{k+1} - \Phi_k \mu_k), \]
\[ \hat{V}_k = V_k + J_k \left( \hat{V}_{k+1} - P_k \right) J_k^T, \]

where $J_k = V_k \Phi_k P_k^{-1}$, and the initialization is expressed as

\[ \hat{\mu}_{n+s} = \mu_{n+s}, \quad \hat{V}_{n+s} = V_{n+s} \]

The derivation results between Eqs. (24) and (31), especially $\mu_k$ and $V_k$, are important referent variables to evaluate the expectation terms in Eq. (19).

The schematic diagram of the proposed process monitoring method is shown below:

![Figure 1: The schematic diagram of the proposed process monitoring method.](image_url)

3.3 Monitoring Statistics Design

Once the PPFA model is established, monitoring statistical indices are of great significance when proceeding to the online monitoring part. For different indices, $T^2$ and SPE are the most popular ones, which assume that the data follows a Gaussian distribution [Joe Qin 2003]. With the parameters set $\Theta$ obtained with EM and genetic algorithm, the augmented latent variable $t_{s_k}^k = \Phi_k t_{s_k-1}^k + K_k (x_k - H_k \Phi_k t_{s_k-1}^k)$ can be applied to define statistical indices.

\[ T^2 = t_{s_k}^k t_{s_k}^{kT} \]
\[ SPE = ||x_k - H_k \Phi_k t_{s_k-1}^k||^2 \]

where the term $x_k - H_k \Phi_k t_{s_k-1}^k$ represents the estimated error given the previous one-step latent variable $t_{s_k-1}^k$. 

The sample is regarded normal if all indices stay inside the normal range.
Based on the established PPFA model, the derivative of latent variables which demonstrates the system dynamic changes can be easily obtained.

\[ \delta(t^s_k) = t^s_k - t^a_{k-1} \]  

(34)

Besides, during the execution of E-step, the term \( D = \mathbb{E} \left[ \delta(t^s_k) \delta(t^s_k)^\top \right] \) can be obtained as follows.

\[ D = \mathbb{E} \left[ t^s_k t^s_k^\top \right] - 2\mathbb{E} \left[ t^s_k t^a_{k-1}^\top \right] + \mathbb{E} \left[ t^a_{k-1} t^s_k^\top \right] \]  

(35)

A novel monitoring index, named as Dynamic Index (DI), is defined as a supplement of \( T^2 \) and SPE to reflect the dynamics of the process.

\[ \text{DI} = \delta(t^s_k)^\top D^{-1} \delta(t^s_k) \]  

(36)

The control limits of \( T^2 \), SPE and DI, denoted as \( \psi_{T^2} \), \( \psi_{SPE} \) and \( \psi_{DI} \) respectively, are determined with the kernel density estimation (KDE), which is an effective non-parametric tool [Botev et al., 2010, Martin and Morris, 1996]. Using KDE, the probability density function (pdf) of monitoring statistics is defined as

\[ f(\nu) = \frac{1}{nh} \sum_{i=1}^{n} K\left( \frac{\nu - \nu_i}{h} \right) \]  

(37)

where \( h \) is the bandwidth parameter, \( \nu \) represents the monitoring statistics, and \( K(\cdot) \) is a kernel function selected as Gaussian in this paper.

With the estimated pdf, the cumulative density function is defined as

\[ P(\nu < \psi) = \int_{-\infty}^{\psi} f(\nu) d\nu \]  

(38)

where the control limit \( \psi \) is obtained by setting a confidence level \( \alpha \), for example, 95%.

### 3.4 Proposed Monitoring Framework

The proposed PPFA based process monitoring method consists of two interrelated parts, namely offline modeling and online monitoring. Figure [Ruiz-Cárceles et al., 2015] illustrates the schematic diagram of the whole monitoring framework. For the offline modeling procedure, the first step is to normalize the normal training data. In this work, the whitening procedure is applied

\[ x_{\text{norm}} = \Lambda^{-1/2} U^\top x \]  

(39)

where \( \Lambda \) represents the singular values of the covariance matrix \( \Omega = \langle x_i x_k^\top \rangle \), and \( U \) is an orthogonal matrix composed of eigenvectors of \( \Omega \). It is noted that the matrices \( \Lambda \) and \( U \) are calculated by applying SVD to \( \Omega \), where \( \Omega = U \Lambda U^\top \).

After the normalization step, EM algorithm is applied to estimate the parameter set \( \Theta \). It is noted that in the E-step, as shown in Eq. (24), original data as well as parameters are augmented to simplify the Kalman smoothing process. Moreover, when executing the M-step, GA algorithm is applied to provide support for estimating \( \Theta \). When the E-and-M-step loop iterates to the maximum number of iterations or convergence is acquired, the final parameter set \( \Theta \) is obtained. Thereafter, according to Eqs. (32) to (36), the control limits \( \psi_{T^2} \), \( \psi_{SPE} \) and \( \psi_{DI} \), along with the monitoring statistics \( T^2 \), SPE and DI can be calculated.

When it comes to the online monitoring process, the newly observed sample is first scaled with Eq. (39). Then applying the established PPFA model to calculate the augmented latent variable \( t^s_k \) conditioned on the newly measurement. Further, compute the monitoring indices \( T^2 \), SPE and DI according to Eqs. (32) to (36). Finally, monitor if \( T^2_{\text{new}} \), \( \text{SPE}_{\text{new}} \) or \( \text{DI}_{\text{new}} \) exceeds its corresponding control limit.

- A dynamic process-relevant fault or a operating condition shift is detected with \( (1 - \alpha) \times 100\% \) confidence level if \( T^2_{\text{new}} > \psi_{T^2} \) or \( \text{DI}_{\text{new}} > \psi_{DI} \).
- If \( \text{SPE}_{\text{new}} > \psi_{SPE} \), a fault that breaks the correlation of the established model is declared.
- The newly observed data is regarded as normal if \( T^2_{\text{new}} < \psi_{T^2} \), \( \text{SPE}_{\text{new}} < \psi_{SPE} \) and \( \text{DI}_{\text{new}} < \psi_{DI} \).

### 4 Experiments and Discussion

Two industrial processes, a three-phase flow facility and a medium speed coal mill, are employed in this section to illustrate the performance of the proposed monitoring scheme.
Dynamic probabilistic predictable feature analysis for high dimensional temporal monitoring

4.1 The Three-Phase Flow Facility

4.1.1 Process Description

The three-phase flow facility is a pressured system suggested by Granfield University, and it is designed for feeding a controlled and measured flow rate of water, oil and air to a pressurized system [Ruiz-Cárcelet al., 2015]. Recently, this complex system has received considerable attention and been successfully applied as a useful industrial process to evaluate the effectiveness of process monitoring methods [Zhang and Zhao, 2017, Yang et al., 2018]. As illustrated in the process diagram in Figure 2, the system is mainly composed of a two-phase separator and a three-phase separator which are connected with pipelines. It can be used to provide different products, among them are single-phase water, air and oil, or their mixtures.

A total of 24 process variables are included in the three-phase flow facility, and at a sample rate of one second, 16 sets of data are collected, the first three of which represent the normal operation status, and the remaining sets denote faulty cases of different operating conditions. For the 24 variables, the first 23 are included in all data sets, while the last is only used in fault case 6. To generate representative normal data sets, four different set points of air flow rates and five different water flow rate set points are adopted and combined into 20 different operating conditions. Set points of air flowrate and water flowrate are listed in TABLE 1. More details of this process are provided in Ref. [Ruiz-Cárcelet al., 2015].

4.1.2 Monitoring Results and Discussion

In this case study, 3200 normal samples under the combination of 0.047 m³/s (air flow rates) and 2 kg/s (water flow rates) are selected as the training set. In addition, two typical fault cases, as listed in TABLE 2, are selected to verify the performance of the proposed method. To further illustrate the superiority of PPFA model, DiPCA [Dong and Qin, 2018] and PFA [Richthofer and Wiskott, 2015] are utilized to make comparisons.

Through cross validation, parameters of different models are selected: for DiPCA, the dynamic order $s$ is set to be 3, and the number of dynamic latent variables $r = 10$. For PFA, the dimension of latent variables is 10, and the time lags is chosen as 5. In the proposed PPFA, the dynamic order $s$ is set to be 2, and the dimension of $t_k$ is 10. Besides, the confidence level of all methods is chosen as 99%.

### Table 1: Set working conditions of air flow rates (m³/s) and water flow rates (kg/s)

| Air flowrate | 0.0208 | 0.0278 | 0.0347 | 0.0417 |
|--------------|--------|--------|--------|--------|
| Water flowrate | 0.5 | 1 | 2 | 3.5 | 6 |

Figure 2: Sketch of three-phase flow facility.
To further test the effectiveness of PPFA, two different practical fault cases happened in a ZGM-113N medium speed coal mill are employed in our work. Figure 5 depicts the schematic diagram and actual pictures of the studied coal mill. For a general medium speed coal mill, the raw materials are transported by the coal feeder into the internal space through the inlet pipe line. The raw coal falls on a grinding table rotating at a constant speed, which is further ground to coal fines. The powder that meets a certain fineness requirement is blown into the furnace by the hot primary air for combustion. The unqualified particles fall back to the coal mill under the action of gravity and inertia, and continue to be ground. As the essential auxiliary equipment, the operating status of the coal mill has an important impact on the safety and economy of the coal-fired power plant [Fan et al., 2021]. There are mainly four typical failures during the operation process, namely intrusion of foreign materials, choking, shortage of coal and fire or explosion in the mill [Agrawal et al., 2016]. Once a fault occurs, it is crucial to detect it as soon as possible, otherwise it may deteriorate to an irreversible damage. To sum up, discovering and maintaining different faults in time does help to ensure the safe operation of the unit and avoid economic losses.

### 4.2 The Medium Speed Coal Mill

#### 4.2.1 Brief Description

To further test the effectiveness of PPFA, two different practical fault cases happened in a ZGM-113N medium speed coal mill are employed in our work. Figure 5 depicts the schematic diagram and actual pictures of the studied coal mill. For a general medium speed coal mill, the raw materials are transported by the coal feeder into the internal space through the inlet pipe line. The raw coal falls on a grinding table rotating at a constant speed, which is further ground to coal fines. The powder that meets a certain fineness requirement is blown into the furnace by the hot primary air for combustion. The unqualified particles fall back to the coal mill under the action of gravity and inertia, and continue to be ground. As the essential auxiliary equipment, the operating status of the coal mill has an important impact on the safety and economy of the coal-fired power plant [Fan et al., 2021]. There are mainly four typical failures during the operation process, namely intrusion of foreign materials, choking, shortage of coal and fire or explosion in the mill [Agrawal et al., 2016]. Once a fault occurs, it is crucial to detect it as soon as possible, otherwise it may deteriorate to an irreversible damage. To sum up, discovering and maintaining different faults in time does help to ensure the safe operation of the unit and avoid economic losses.

| Index             | Water flowrate | Air flowrate | No. samples | Description   |
|-------------------|----------------|--------------|-------------|---------------|
| Training data     |                |              | 3200        | Normal        |
| Test case 1       | 2              | 0.0417       | 4467        | Airline blockage |
| Test case 2       | 2              | 0.0417       | 3851        | Open direct bypass |

Table 2: Description of data samples for modeling and monitoring
4.2.2 Monitoring Results and Discussion

In this process, two practical fault cases, mechanical failure and coal blockage, are studied. As listed in Table 3, 15 different variables are selected to build models according to the prior knowledge of coal mills [Fan et al., 2021, Cortinovis et al., 2013]. A total of normal 2385 samples are selected as the training set, and the sampling interval of the data samples is 20s. The parameters of different models are selected via cross validation: for DiPCA, \( r = 10, s = 4 \); for PFA, \( r = 8, s = 5 \); and for PPFA, \( r = 8, s = 3 \). For the fault of mechanical failure, it lasts for two hours from the 94th sample to the 452th sample. Figure 6 shows the historical trends of the coal mill current and the coal flow rate. It is observed that the current fluctuates in a larger range during the fault process compared with the normal state. After maintaining the fault, some foreign matters with considerable size were found inside the mill.

The monitoring results of DiPCA, PFA and PPFA are presented in Figure 7, in which the red dot line denotes the control limit with a 99% confidence level. For DiPCA, as reflected in Figure 7(a), both \( T^2 \) and SPE begin exceeding the control limits from around the 110th sample, and return to normal region at the 452th sample. During the process when the fault occurs, a considerable portion of the two indices stay below the control limit, leading to a relatively
low fault detection rate. As observed in Figure 7(b), for PFA, the numerical value of $T^2$ is in an opposite trend to the actual fault, and it is lower than the red line during the occurrence of the fault, but it exceeds the threshold after the fault is removed. Though SPE follows the fault process in a good manner, the overall monitoring performance is unreliable. From the plots of PPFA's results illustrated in Figure 7(c), it can be seen that within the two hours of failure, $T^2$ and SPE detect the fault timely and ensure a high fault detection rate compared with DiPCA and PFA. In addition, when the coal mill current fluctuates greatly, the newly proposed index DI responds timely, and hence the dynamic characteristics of the system during the fault period are well reflected. For the second fault case, as shown in Figure 8, the coal blockage starts from the 195th sample and ends at the 435th sample. When the fault exists, the mill’s current and the difference between the inlet pressure and outlet pressure have an upward trend until it is discovered and handled at about the 379th sample. During this period, the primary air flow rate continues to decrease, and starts increasing when the fault begins recovering. The coal-air mixture temperature at the export of the mill drops significantly at the beginning of the fault, and then remains at a low value until the failure ends. During the actual operation of the unit, it takes more than one hour for the fault to be discovered, which significantly delays the maintenance of the mill and causes unnecessary economic losses.
Dynamic probabilistic predictable feature analysis for high dimensional temporal monitoring

Figure 5: Schematic diagram of the medium coal mill (b) sketch, (a)(c) actual picture.

Table 3: Selected variables for constructing process monitoring model.

| No. | Variables     | Description                                      | Unit          |
|-----|---------------|--------------------------------------------------|---------------|
| M1  | $N_{\text{unit}}$ | Unit load                                       | MW            |
| M2  | $M_{\text{coal}}$ | Coal flow rate transported by coal feeder       | t/h           |
| M3  | $W_{\text{air}}$ | Flow rate of primary air                         | t/h           |
| M4  | $T_{\text{air}}$ | Temperature of primary air                       | °C            |
| M5  | $P_{\text{air}}$ | Pressure of primary air                          | kPa           |
| M6  | $I_{\text{mill}}$ | Current of the mill’s motor                    | A             |
| M7  | $I_{\text{feed}}$ | Current of coal feeder’s motor                  | A             |
| M8  | $\Delta P_{\text{seal}}$ | Difference between the pressure of seal air and primary air | kPa           |
| M9  | $T_{\text{coal-air}}$ | Outlet temperature of the mixture of coal and air | °C            |
| M10 | $N_{\text{coal-air}}$ | Coal-air mixture outlet pressure              | kPa           |
| M11 | $\Delta P_{\text{in-out}}$ | Difference between the pressure of inlet and outlet of coal equipment | kPa           |
| M12 | $T_{\text{oil}}$ | Coal mill’s lubricating oil temperature        | °C            |
| M13 | $T_{\text{tile}}$ | Coal mill’s thrust tile temperature             | °C            |
| M14 | $T_{\text{bearing}}$ | Bearing temperature of the motor               | °C            |
| M15 | $T_{\text{stator}}$ | Stator winding temperature of the motor         | °C            |

Figure 6: Historical trends of two sensors of fault case 1.
Figure 7: Monitoring results of fault case 1 with (a)DiPCA, (b)PFA, and (c)PPFA.

Figure 8: Historical trends of relevant sensors of fault case 2.
Figure 9 presents the monitoring results for the fault of coal blockage. From Figure 9(a), for the statistics $T^2$ and SPE it can be seen that there is a delay for near 100 samples (0.56 hour) after the disturbance happens. For the results of PFA, depicted in Figure 9(b), it shows similar pattern as it does for fault case 1. That is, though SPE retains a satisfying performance, $T^2$ has a poor performance which raises a large number of false alarms and obtains a low fault detection rate. For the proposed PPFA model, it is observed in Figure 9(c) that $T^2$ and SPE successfully raise the alarm once the fault occurs, along with a great fault detection rate. Moreover, combining with the trend of the variables in Figure 8 it can be seen that DI has obvious over-limit response to rapid fluctuations in different stages. In conclusion, considering the monitoring performance, the proposed PPFA model has great advantages over existing methods such as DiPCA and PFA. Its advantages are mainly manifested in that this method can not only realize fault detection quickly and accurately, but also the newly proposed index can be used as a supplement to reflect the dynamic changes of the system.

5 Conclusion

In this paper, a novel probabilistic predictable feature analysis method is proposed for high dimensional time series monitoring. The proposed method takes measurement noise and full interpretation of dynamic characteristics into consideration. During the procedure of EM iterations, GA and Kalman filter are successfully employed to estimate
parameters. In addition, a new monitoring index DI is defined and applied to monitor the dynamic variations of industrial processes. Through applications on the three-phase flow facility and a medium speed coal mill, the PPFA based process monitoring method has shown its effective performance compared with existing deterministic methods. Based on the monitoring results, the PPFA algorithm is worth for further investigation.

A Proof of Lemma 7

According to Eq. (6), we have

\[ t_k^i = \beta_1^i t_{k-1}^i + \cdots + \beta_s^i t_{k-s}^i + e_k^i \]  

(40)

With the constraint \( \mathbb{E} \left[ t_k \right] = 0 \), the expectation of \( t_k^i \) is

\[ \mathbb{E} \left[ t_k^i \right] = \beta_1^i \mathbb{E} \left[ t_{k-1}^i \right] + \cdots + \beta_s^i \mathbb{E} \left[ t_{k-s}^i \right] + \mathbb{E} \left[ e_k^i \right] = 0 \]  

(41)

The variance of \( t_k^i \) can be obtained with the following formula

\[ \text{var} \left[ t_k^i \right] = \mathbb{E} \left[ (t_k^i)^2 \right] - \mathbb{E} \left[ t_k^i \right]^2 = \mathbb{E} \left[ (t_k^i)^2 \right] \]  

(42)

Since \( (t_k^i)^2 = \beta_1^i t_{k-1}^i t_{k-1}^i + \cdots + \beta_s^i t_{k-s}^i t_{k-s}^i + t_k^i e_k^i \), Eq. (42) is computed by

\[ \mathbb{E} \left[ (t_k^i)^2 \right] = \beta_1^i \mathbb{E} \left[ t_{k-1}^i t_{k-1}^i \right] + \cdots + \beta_s^i \mathbb{E} \left[ t_{k-s}^i t_{k-s}^i \right] + \mathbb{E} \left[ t_k^i e_k^i \right] \]  

(43)

where the last term can be further simplified by

\[ \mathbb{E} \left[ t_k^i e_k^i \right] = \mathbb{E} \left[ (\beta_1^i t_{k-1}^i + \cdots + \beta_s^i t_{k-s}^i + e_k^i) e_k^i \right] \]
\[ = \mathbb{E} \left[ (\beta_1^i t_{k-1}^i + \cdots + \beta_s^i t_{k-s}^i) e_k^i \right] + \mathbb{E} \left[ (e_k^i)^2 \right] \]
\[ = \mathbb{E} \left[ (\beta_1^i t_{k-1}^i + \cdots + \beta_s^i t_{k-s}^i) \mathbb{E} \left[ e_k^i \right] \right] + \mathbb{E} \left[ (e_k^i)^2 \right] \]
\[ = \mathbb{E} \left[ (e_k^i)^2 \right] = \tau_i^2 \]  

(44)

For ease of representation, we define

\[ \gamma_j^i = \text{cov} \left( t_k^i, t_{k-j}^i \right) = \text{cov} \left( t_{k-j}^i, t_k^i \right) \]
\[ = \mathbb{E} \left[ t_k^i t_{k-j}^i \right] - \mathbb{E} \left[ t_k^i \right] \mathbb{E} \left[ t_{k-j}^i \right] \]
\[ = \mathbb{E} \left[ t_k^i t_{k-j}^i \right] \]  

(45)

Therefore, the covariance in Eq. (42) simplifies to

\[ \text{var} \left[ t_k^i \right] = \mathbb{E} \left[ (t_k^i)^2 \right] = \beta_1^i \gamma_1^i + \cdots + \beta_s^i \gamma_s^i + \tau_i^2 \]  

(46)

Since the constraint \( \mathbb{E} \left[ t_k^i t_{k-j}^i \right] = I \), has been given in Eq. (6), we have

\[ \beta_1^i \gamma_1^i + \cdots + \beta_s^i \gamma_s^i + \tau_i^2 = 1 \]  

(47)

Therefore, the relation between \( B_j \) and \( \Gamma \) can be expressed by Eq. (7).

Acknowledgment

This work was supported by China Scholarship Council (grant numbers 202006090212), Qinglan Project of Jiangsu Province of China, National Natural Science Foundation of China under Grant 51976031 as well as the University of Waterloo. And this work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.
References

Kristen Severson, Paphonwit Chaiwatanodom, and Richard D Braatz. Perspectives on process monitoring of industrial systems. *Annual Reviews in Control*, 42:190–200, 2016.

Zhiqiang Ge. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 171:16–25, 2017.

Le Zhou, Gang Li, Zhihuan Song, and S Joe Qin. Autoregressive dynamic latent variable models for process monitoring. *IEEE Transactions on Control Systems Technology*, 25(1):366–373, 2016.

Carlos F Alcala and S Joe Qin. Reconstruction-based contribution for process monitoring. *Automatica*, 45(7):1593–1600, 2009.

Gang Li, S Joe Qin, and Donghua Zhou. Geometric properties of partial least squares for process monitoring. *Automatica*, 46(1):204–210, 2010.

Qinqin Zhu, Qiang Liu, and S Joe Qin. Concurrent quality and process monitoring with canonical correlation analysis. *Journal of Process Control*, 60:95–103, 2017.

Wenfu Ku, Robert H Storer, and Christos Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1):179–196, 1995.

Tiago J Rato and Marco S Reis. Advantage of using decorrelated residuals in dynamic principal component analysis for monitoring large-scale systems. *Industrial & Engineering Chemistry Research*, 52(38):13685–13698, 2013.

Erik Vanhatalo, Murat Kulahci, and Bjarne Bergquist. On the structure of dynamic principal component analysis used in statistical process monitoring. *Chemometrics and intelligent laboratory systems*, 167:1–11, 2017.

Gang Li, S Joe Qin, and Donghua Zhou. A new method of dynamic latent-variable modeling for process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11):6438–6445, 2014.

Lingling Guo, Ping Wu, Siwei Lou, Jinfeng Gao, and Yichao Liu. A multi-feature extraction technique based on probabilistic slow feature analysis for nonlinear dynamic process monitoring. *Journal of Process Control*, 85:159–172, 2020.

Yining Dong and S Joe Qin. A novel dynamic pca algorithm for dynamic data modeling and process monitoring. *Journal of Process Control*, 67:1–11, 2018.

Stefan Richthofer and Laurenz Wiskott. Predictable feature analysis. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pages 190–196. IEEE, 2015.

Junhua Zheng, Zhihuan Song, and Zhiqiang Ge. Probabilistic learning of partial least squares regression model: Theory and industrial applications. *Chemometrics and Intelligent Laboratory Systems*, 158:80–90, 2016.

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

Feihong Guo, Chao Shang, Biao Huang, Kangcheng Wang, Fan Yang, and Dexian Huang. Monitoring of operating point and process dynamics via probabilistic slow feature analysis. *Chemometrics and Intelligent Laboratory Systems*, 151:115–125, 2016.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *The Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

Christopher M Bishop. *Pattern recognition and machine learning*. Springer, 2006.

David E Goldberg and John Henry Holland. Genetic algorithms and machine learning. 1988.

Harish Garg. A hybrid PSO-GA algorithm for constrained optimization problems. *Applied Mathematics and Computation*, 274:292–305, 2016.

Chao Shang, Biao Huang, Fan Yang, and Dexian Huang. Probabilistic slow feature analysis-based representation learning from massive process data for soft sensor modeling. *AIChE Journal*, 61(12):4126–4139, 2015.

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

Zdravko I Botev, Joseph F Grotowski, Dirk P Kroese, et al. Kernel density estimation via diffusion. *The annals of Statistics*, 38(5):2916–2957, 2010.

EB Martin and AJ Morris. Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, 6(6):349–358, 1996.
Cristobal Ruiz-Cárce, Yi Cao, D Mba, Liyun Lao, and RT Samuel. Statistical process monitoring of a multiphase flow facility. *Control Engineering Practice*, 42:74–88, 2015.

Shumei Zhang and Chunhui Zhao. Stationarity test and bayesian monitoring strategy for fault detection in nonlinear multimode processes. *Chemometrics and Intelligent Laboratory Systems*, 168:45–61, 2017.

Jian Yang, Zheng Lv, Hongbo Shi, and Shuai Tan. Performance monitoring method based on balanced partial least square and statistics pattern analysis. *ISA transactions*, 81:121–131, 2018.

Wei Fan, Shaojun Ren, Qin Qin Zhu, Zhijun Jia, Delong Bai, and Fengqi Si. A novel multi-mode bayesian method for the process monitoring and fault diagnosis of coal mills. *IEEE Access*, 9:22914–22926, 2021.

Vedika Agrawal, Bijaya Ketan Panigrahi, and PMV Subbarao. Intelligent decision support system for detection and root cause analysis of faults in coal mills. *IEEE Transactions on Fuzzy Systems*, 25(4):934–944, 2016.

Andrea Cortinovis, Mehmet Mercangoez, Tarun Mathur, Jan Poland, and Marcel Blaumann. Nonlinear coal mill modeling and its application to model predictive control. *Control Engineering Practice*, 21(3):308–320, 2013.