Condition-Driven Data Analytics and Monitoring for Wide-Range Nonstationary and Transient Continuous Processes

Chunhui Zhao, *Senior Member, IEEE*, Junhao Chen, and Hua Jing

Abstract— **Frequent and wide changes in operation conditions are quite common in real process industry, resulting in typical wide-range nonstationary and transient characteristics along time direction. The considerable challenge is, thus, how to solve the conflict between the learning model accuracy and change complexity for analysis and monitoring of nonstationary and transient continuous processes. In this work, a novel condition-driven data analytics method is developed to handle this problem. A condition-driven data reorganization strategy is designed which can neatly restore the time-wise nonstationary and transient process into different condition slices, revealing similar process characteristics within the same condition slice. Process analytics can then be conducted for the new analysis unit. On the one hand, coarse-grained automatic condition-mode division is implemented with slow feature analysis to track the changing operation characteristics along condition dimension. On the other hand, fine-grained distribution evaluation is performed for each condition mode with Gaussian mixture model. Bayesian inference-based distance (BID) monitoring indices are defined which can clearly indicate the fault effects and distinguish different operation scenarios with meaningful physical interpretation. A case study on a real industrial process shows the feasibility of the proposed method which, thus, can be generalized to other continuous processes with typical wide-range nonstationary and transient characteristics along time direction.**

Note to Practitioners—**Industrial processes in general have nonstationary characteristics which are ubiquitous in real world data, often reflected by a time-variant mean, a time-variant autocovariance, or both resulting from various factors. The focus of this study is to develop a universal analytics and monitoring method for wide-range nonstationary and transient continuous processes. Condition-driven concept takes the place of time-driven thought. For the first time, it is recognized that there are similar process characteristics within the same condition slice and changes in the process correlations may relate to its condition modes. Besides, the proposed method can provide enhanced physical interpretation for the monitoring results with concurrent analysis of the static and dynamic information which carry different information, analogous to the concepts of "position" and "velocity" in physics, respectively. The static information can tell the current operation condition, while the dynamic information can clarify whether the process status is switching between different steady states. It is noted that the condition-driven concept is universal and can be extended to other applications for industrial manufacturing applications.**

Index Terms— **Bayesian inference-based distance (BID), condition slices, Gaussian mixture model (GMM), slow feature analysis (SFA), temporal analytics, transient process, wide-range nonstationarity.**

Manuscript received January 2, 2020; revised April 27, 2020; accepted July 15, 2020. Date of publication August 4, 2020; date of current version October 6, 2021. This article was recommended for publication by Associate Editor P. Burget and Editor F.-T. Cheng upon evaluation of the reviewers' comments. This work was supported in part by the Zhejiang Key Research and Development Project under Grant 2019C01048 and in part by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant U1709211. *(Corresponding author: Chunhui Zhao.)*

The authors are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: chhzhao@zju.edu.cn).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2020.3010536

I. Introduction

IN THE last few decades, machine learning techniques [1]–[6], such as multivariate statistical analysis methods [1], [2], support vector machine [3], neural network [4], and Bayesian inference methods [5], [6], have been widely used for industrial process analytics and monitoring as the amount of measurement data increases. These techniques design feature engineering in different ways to reveal the underlying characteristics from measurement data. The normal variation is, thus, defined within a certain region under some confidence level. For online application, when the process moves outside the desired operating region, it is concluded that an "unusual and faulty" change in the process behavior has occurred. Therefore, the key is how close the model and confidence region describe the normal process. The underlying assumption is that the distribution of the observed data is stationary which relies on the fact that we can generalize from a sample to the population. However, industrial processes in general have nonstationary characteristics which are ubiquitous in real world data, often reflected by a time-variant mean, a time-variant autocovariance, or both [7], [8] resulting from various factors, such as operation condition changes, frequent product changing, facility aging, and unmeasured disturbances. Besides, for industrial processes with time-variant conditions, the process status is multimodal with not only multifarious steady states but also frequent switchings between different steady states. It is termed wide-range nonstationary here since nonstationary is significant which is different from the conventional nonstationary process. As the properties of nonstationary variables significantly vary along time, the conventional machine learning-based process monitoring methods experience a model mismatch problem where the process behaviors to be evaluated vary along time and show difference from the reference data. Besides, the transient changes that in fact embody rich dynamic information are not well addressed and, thus, may lose high resolution into the frequent process fluctuations. The nonstationary and transient changes may cause frequent false alarms even though the process is under nominal operation status. Therefore, nonstationary and transient continuous process monitoring is a difficult task, which, however, has only been sporadically studied.

Some existing machine learning methods may be used to address this problem. One is the adaptive methods [9]–[11]. They, however, try to capture the frequent changes along time directions and, thus, may require frequent model updating, which, thus, may damage the model performance since fault data may be falsely included. Just in time learning (JITL) [12] is deemed to be one special case of adaptive methods. Without prior models, it employs database technology and nearest neighbors to dynamically build local models upon query in which the most relevant samples are searched from the historical database around a query sample. However, the accuracy of data searching can influence the resulting local model which, thus, may not be sensitive to the abnormality. Besides, it simply regards the out-of-control samples to be abnormality and cannot distinguish different types of normal changes from real faults. A multimode process-monitoring strategy [13]–[15] separates the wide-scale nonstationary processes into different steady modes and developed multiple models for different operation statuses. Each model is deemed to represent a specific operation mode and explains the local process characteristics with a high resolution, which can effectively improve the monitoring reliability. The key point is how to divide the whole process into different modes and online judge the mode affiliation of the new sample so that the proper mode model can be correctly adopted for monitoring. Clustering is a typical procedure that is used to identify the observations that belong to each operating mode. Some popular clustering algorithms [16]–[19] can be used for mode division, including K-means [16], K-nearest neighbor (KNN) clustering [17], [18], and fuzzy c-means [19]. However, they in general consider the distance between separate samples, which cannot comprehensively describe their different characteristics. The quality of mode division may directly influence the model performance and, thus, influence the monitoring performance. For online application, it is difficult to determine which mode the current sample belongs to and which model should be used for calculation of monitoring statistics. In general, different models are tried in order to check which can best accommodate the current sample. If a wrong model is adopted, it may result in inaccurate monitoring results, including both missing alarms and false alarms. Besides, the existing multimode modeling strategies only consider the steady-state distribution and detect steady-state deviations from each operation mode, which do not explicitly represent the temporal information of process data.

In recent years, cointegration analysis (CA) [8] which is an effective method to investigate the long-term cointegration relationship between nonstationary variables has received increasing attentions. Despite the time-varying changes in process trajectories, the underlying long-term equilibrium relation keeps invariable and, thus, only one CA model is needed to describe the robust characteristics. Several applications in industrial field have been reported [20]–[23], including process monitoring and fault diagnosis. Zhao and Sun [20] and Zhao and Huang [21] considered both the steady-state distribution and temporal process behaviors to detect deviations from the long-term equilibrium relation which can distinguish changes in the normal operating condition and real faults for nonstationary processes. Despite some successful reports of CA algorithm to industrial application, it assumes that the nonstationary variables are integrated of the same order, which, however, may not be well satisfied in practice. In that case, CA may fail to handle nonstationary process variables.

For real industrial processes, the operation status is commonly transient which means they stay at some steady state for only a period of time and may have transition patterns between different states, revealing typical dynamic characteristics. Different from the steady-state operation condition that could be abstracted as the static distribution, process dynamics can be perceived as the temporal distribution [24]. They carry different information, analogous to the concepts of "position" and "velocity" in physics [25]. Therefore, the static and dynamic variations ought to be explained differently and separately monitored since they are completely different from each other. Zhang and Zhao [26] proposed a concurrent batch process monitoring strategy to distinguish operation condition deviations and process dynamics anomalies with slow feature analysis (SFA) algorithms [27]–[29]. Besides, Yu and Zhao [9] further developed the recursive version of SFA algorithm which can implement updating with newly available samples to adapt for the new normal slow changes. Although SFA can probe into the dynamic information for process analysis and monitoring, it inherently assumes that the operation condition is unique, which cannot hold for nonstationary processes. For transient process analysis, both the static information and its dynamic counterpart are important to reveal its real case. Either a single or a multiple steady model cannot well describe the dynamic information hidden in a state of flux. It, thus, spurs further fine-grained research for transient process modeling and monitoring. A study on advanced machine learning is of necessity which requires deep analytics of physical process nature for designing the variant of machine learning models to capture the specific changes in industrial process mechanism. It is different from the conventional machine learning methods that are directly applied to industrial processes and, thus, is termed to be advanced machine learning here.

In this work, a novel viewpoint of process analytics is presented to handle the monitoring problem of wide-range nonstationary and transient continuous processes. First, a new data analysis unit is reorganized by splitting the samples into different condition slices following the changes in condition indicator. In each condition slice, the process characteristics stay similar since the samples are around the same condition, which can be readily revealed by a two-way SFA model. Second, different condition modes are identified automatically by following the changes in the SFA projection matrices, reflecting the rule of underlying characteristics changing along condition direction. That is, the condition-slice matrices will have similar SFA models within each condition mode; different condition modes result in different SFA models, reflecting that process correlation changes over different condition modes. Third, SFA-based condition-mode model is built to extract both the static and dynamic information of the samples. The underlying distribution within each condition mode is further and comprehensively explored by Gaussian mixture model (GMM), including both the static and temporal information. Four new Bayesian inference-based distance (BID) statistics are designed for monitoring different types of process changes with meaningful physical interpretation. Case study on a real nonstationary continuous industrial process shows the feasibility of the proposed method which can be generalized to other nonstationary continuous processes.

The main contributions of this article can be summarized as follows.

1) A novel data reorganization strategy is designed to transform the time-driven wide-range nonstationary process into different condition slices. It is the first time that the condition-driven idea is proposed based on such a recognition that process characteristics may stay similar within the same condition slice.
2) The changing rule of process characteristics along condition direction is revealed by developing an automatic sequential condition-mode division (SCMD) algorithm. It reveals that despite changes in operation conditions over time, the underlying variable correlation will be largely similar within the same condition mode and significantly different for different condition modes.
3) BID monitoring indices are defined with joint analysis of the static and dynamic information which can clearly indicate the fault effects and distinguish different operation scenarios.

The remainder of this article is presented as follows. First, problem statement and the motivation for this work are discussed. Then the proposed condition-driven data analytics methodology is described with reference to, condition-slice data array reorganization, automatic condition-mode division, fine-grained model development, and online application. Third, we present the application results of the proposed modeling and monitoring method to a real wide-range nonstationary and transient industrial process. Comparison is made with the conventional methods. The conclusion is drawn in Section IV.

II. Methodology

In this section, the necessity of this study is analyzed first. Then, an advanced machine learning-based process analytics and monitoring strategy is proposed, including both coarse-grained and fine-grained levels. The major factor that drives the changes in operation conditions can point to the changes in process characteristics which can, thus, be used as an indicator to reorganize the data array. It tends to be the input to the concerned system. For example, for coal mill facility, the coal feed rate is the key input and chosen as the condition indicator since the process variables, such as outlet pressure, change with time under the influences of coal feed rate. An automatic condition-mode division method is then developed to preliminarily distinguish the operation statuses. It is termed coarse-grained modeling and analysis. This allows

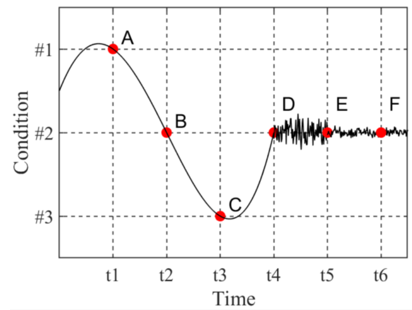


Fig. 1. Illustration of the wide-range nonstationary and transient process.

two-way SFA to be "directly" applied to a nonstationary process after a proper dimension reorganization and automatic condition-mode division. Further, a fine-grained analysis is conducted to further explore the underlying distribution for each condition mode and its corresponding online monitoring strategy.

A. Problem Statement and Motivation

In practice, the industrial process is commonly regulated to different operating levels following nonstationary operating trajectories, often reflected by a time-variant mean, a time-variant autocovariance, or both. With multifarious operation conditions that change widely and frequently each dominated by different physical and chemical phenomena, industrial processes present typical wide-range nonstationary characteristics. Besides, the process may not persistently stay at some steady point and can be typically transient with frequent dynamic switches. Even for the operation points which lie at the same level, the process status may not be steady since the process may be going across the current operation point from the other operation points.

Fig. 1 illustrates different operation points of a wide-range nonstationary and transient process. For different time, on the one hand, they may lie at different operation conditions but reveal the same process dynamics; on the other hand, the process may show different process dynamics along time direction although they may lie at the same operation condition. For example, A and C stay at different operation conditions, i.e., Condition #1 and Condition #3, but they reveal the same process dynamics. B, D, E, and F stay at the same operation condition, i.e., Condition II, which, however, has quite different temporal characteristics. First, B and D transfer from different operation conditions. Point B is changing from Condition #1 down to Condition #2, while Point D is changing from Condition #3 up to Condition #2. Second, they may have different changing speeds with different magnitudes of changes within the same time span. For Points E and F, although they both operate at Condition #2, they present different temporal variations as evaluated by the changing speed.

From Fig. 1, it is clear that a single static information may not comprehensively describe the process characteristics. In particular, the time-varying process trajectories reveal rich temporal distribution information. Both the static and dynamic information are important to distinguish different process statuses and changes. How to comprehensively describe and model the transient characteristics in presence of wide-range nonstationary process variations is imperative, which has been seldom discussed before from the data-driven perspective.

For wide-range nonstationary and transient processes subject to time-variant operation conditions, the following considerations are first recognized: 1) although the process characteristics change with time, the process may follow certain relations within the same condition; 2) the dynamic information is important to distinguish different process changes considering the transient nature; and 3) despite changes in operation conditions over time, the underlying variable correlation will be largely similar within the same condition mode and a process may be divided into several condition modes as indicated by changes in its inherent process correlations over conditions.

Based on the above recognition, a new data analysis object is needed to prepare the modeling data. The changes in operation conditions can point to the changes in process characteristics which can, thus, be used as an indicator to reorganize the data array. The data array reorganization strategy is presented as below.

B. Condition-Driven Data Array Reorganization

For nonstationary and transient continuous process, a two-way data matrix $\mathbf{X}(K \times J)$ is available assuming that J process variables are measured online at $k = 1, 2, \ldots, K$ time instances. Resulting from frequent changes in operation conditions along time as illustrated in Fig. 1, the process shows the wide-range nonstationary and transient characteristics, in which the nonstationary variables are not integrated of the same order and, thus, cannot be modeled by CA. Since the conventional methods cannot be directly used to well enclose the process variations and derive the underlying characteristics from the data, one question naturally arises. That is, whether we can reorganize the measurement data to reduce the influences of nonstationarity to a certain extent. Clearly, the conventional time dimension cannot be directly used and new analytical unit should be designed. Here, the analysis dimensionality is reorganized which can neatly translate the time-wise nonstationary and transient process into condition-wise well-regulated data array. Besides, to reveal the important temporal distribution information, the changing speed is considered as the counterpart of the static data reorganization. The data array reorganization is performed as below.

First, each measurement vector is extended at each time t to have the time-series difference information. Considering the data $\mathbf{x}_t \in \mathrm{R}^J$, the difference vector $\dot{\mathbf{x}}_t$ is defined as $\dot{\mathbf{x}}_t^{\mathrm{T}}(1 \times J) = \mathbf{x}_t^{\mathrm{T}}(1 \times J) - \mathbf{x}_{t-1}^{\mathrm{T}}(1 \times J)$. In this way, the separate measurement vector $\mathbf{x}_t$ at each time t is extended to have its temporal counterpart, i.e., $\{\mathbf{x}_t, \dot{\mathbf{x}}_t\}$. It is noted that the first difference vector $\dot{\mathbf{x}}_t$ is calculated by using the measurement information at $t = 0$.

Second, the operation condition, instead of time, is used to reorganize the measurement data $\mathbf{x}_t$. Therefore, it is termed

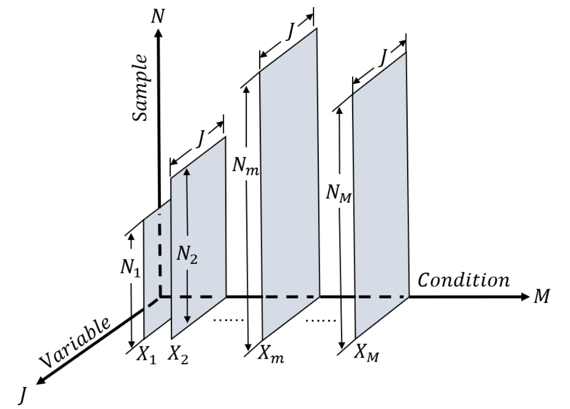


Fig. 2. Reorganization of a three-way data array (each gray slice represents the data samples within the same condition interval, termed condition slice).

condition indicator here. The values of condition are reordered to increase monotonously and the condition interval β is defined to separate all the condition values into M conditions. The measurement data within the same operation condition interval are collected together for both the static and temporal parts. The new analysis unit is termed condition slice here. It is noted that for an uncommon condition, the samples may not be enough for modeling. For that situation, the size of the condition interval β can be adjusted to cover more samples considering that the process characteristics cannot significantly change with similar conditions.

By the above data reconstruction method, the new analysis subject is prepared. As shown in Fig. 2, a three-way irregular data array is rearranged along the condition dimension for both the static and dynamic information, including three dimensionalities, i.e., condition, sample, and variables. Here, for simplicity, only the static data array is presented. For dynamic data, it is arranged in the same way. The collected measurement and difference matrices are defined for each operation condition interval, $\mathbf{X}_m(N_m \times J)$ and $\dot{\mathbf{X}}_m(N_m \times J)$, which are termed condition slices here. They are the smallest analysis unit. N_m denotes the number of samples for each condition which can be different, and $K = \sum_{m=1}^{M} N_m$. The subscript m denotes conditions and $m = 1, 2, \ldots, M$. $\mathbf{X}_m$ is termed the static condition slice and $\dot{\mathbf{X}}_m$ is termed the dynamic condition slice. They denote the static process status at different operation conditions and the affiliated temporal information. With the proposed data reorganization strategy, the conventional measurement data are transformed into the 3-D data structure which is often used to describe batch processes [30]. In this way, the monitoring methods are smartly connected between batch process and continuous process. The difference is that for batch processes, it is time driven instead of condition driven, revealing changes in process characteristics along time.

C. Automatic SCMD

Here, condition slices with similar underlying process characteristics are collected to define a condition mode. In SFA analysis, the projection model represents the information of the process correlations. The developed modeling strategy begins with analyzing the projection matrix at each condition slice. The condition slices will have similar SFA models within each condition mode; different condition modes result in different SFA models, reflecting that process correlation changes over different condition modes. Likewise, changes in the SFA models, reflecting changes in the underlying process behavior, can be used to determine the condition mode. A new condition-mode division method is proposed based on the reorganized data analysis subject. Both the similar characteristics within a local condition region and the condition sequence are considered for automatic condition-mode division. The basic analysis and modeling unit is condition slice. The modeling procedure is described as below.

Step 1 (Data Preparation): Prepare the static condition slice $\mathbf{X}_m(N_m \times J)$ and the corresponding temporal condition slice $\dot{\mathbf{X}}_m(N_m \times J)$ as mentioned before. For each condition-slice data matrix $\mathbf{X}_m$, the variables are preprocessed to have zero mean. Denote the preprocessed condition slice $\bar{\mathbf{X}}_m$.

Step 2 (Condition-Slice-Based Modeling): Perform the two-way SFA algorithm on the static and dynamic condition slices and get the initial condition-slice models. The number of retained slow features (SFs) is determined [25] to separate slow and fast process information. Then find the number of SFs that occurs most throughout all conditions and set it as the unified dimension of condition-slice SFA models, R. Calculate the monitoring statistic values of both the dominant and residual SFs to form an orthogonal decomposition and complete monitoring of the data space at each condition and determine the condition-slice confidence limits, $\mathrm{Ctr}_{\mathrm{s},m}$ and $\mathrm{Ctr}_{\mathrm{f},m}$, by the kernel density estimation (KDE) method [31]. They represent the explanatory ability of condition-slice SFA models.

Step 3 (Condition-Segment-Based SFA Modeling): From the beginning of the reorganized data array, iteratively add next condition slice one by one to the existing ones and variable-unfold them within the current condition region up to the kth condition as shown in Fig. 3 to form the condition-segment data matrix, $\mathbf{X}_{\upsilon,k}(\sum_{i=1}^{k} N_i \times J)$. The corresponding dynamic matrix is also obtained, $\dot{\mathbf{X}}_{\upsilon,k}(\sum_{i=1}^{k} N_i \times J)$. Perform SFA on the rearranged data matrix and get the condition-segment SFA models for both slow and fast parts $\mathbf{W}_{\upsilon,s,k}(J \times R)$ and $\mathbf{W}_{\upsilon,f,k}(J \times (J-R))$. Calculate the slow and fast monitoring values for each condition-slice data matrix by the explanation of the current time-segment SFA models and determine the confidence limits, $\mathrm{Ctr}_{\upsilon,\mathrm{s},k}$ and $\mathrm{Ctr}_{\upsilon,\mathrm{f},k}$, by KDE [31]. It represents the reconstruction power of the time-segment SFA models to each condition slice within the concerned condition region.

Step 4 (Compare Model Accuracy): Compare $\mathrm{Ctr}_{\upsilon,\mathrm{s},k}$ with $\mathrm{Ctr}_{\mathrm{s},k}$ and $\mathrm{Ctr}_{\upsilon,\mathrm{f},k}$ with $\mathrm{Ctr}_{\mathrm{f},k}$ for each condition slice within the concerned condition region. Find the condition k^* from which three consecutive samples show $\mathrm{Ctr}_{\upsilon,\mathrm{s},k} > \alpha * \mathrm{Ctr}_{\mathrm{s},k}$ or $\mathrm{Ctr}_{\upsilon,\mathrm{f},k} > \alpha * \mathrm{Ctr}_{\mathrm{f},k}$, where α is a constant attached to the original condition-slice control limits, termed relaxing factor here. It means that the addition of the current condition slice has imposed great influences on the condition-segment SFA monitoring models and the resulting monitoring performance. The accuracy of condition-segment models is, thus, significantly worse than that of condition-slice models.

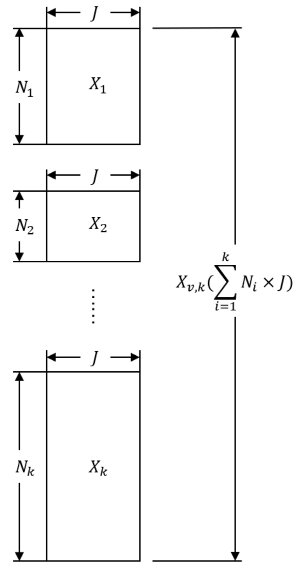


Fig. 3. Variable-unfolding data arrangement up to the kth time.

So α determines how much the condition-segment SFA model is allowed to be less representative than condition-slice SFA models, i.e., insufficient explanatory power in comparison with the same-dimensional condition-slice SFA models. The condition slices before k^* are denoted as one condition mode.

Step 5 (Data Updating and Recursive Implementation): Remove the first condition mode and the left process data are now employed as the new input data in step 3. Recursively repeat steps 3–4 from the updated beginning of the process to find the following condition modes.

The output is a partition of samples along condition direction and the alternation of different condition modes is consecutive along condition direction. The basic idea is to find those condition slices with similar variable correlations so that they can be modeled by the same model while those different ones will be characterized by different models. It is noted that each condition slice does not necessarily follow Gaussian distribution. Assume C condition modes are available here. In each condition mode, the condition-mode data, $\mathbf{X}_c^{\upsilon}(N_c^{\upsilon} \times J)$, is organized in the same way as shown in Fig. 3 by adding all condition slices $\bar{\mathbf{X}}_m$ within the same condition mode, where $N_c^{\upsilon} = \sum_{m \in C} N_m$. Correspondingly, the temporal counterpart, $\dot{\mathbf{X}}_c^{\upsilon}(N_c^{\upsilon} \times J)$, is also organized based on $\dot{\mathbf{X}}_m$ in the same way.

For different operation conditions, the proposed automatic SCMD algorithm is to separate the whole process with time-varying conditions into different condition modes which are deemed to have similar static characteristics within the same condition mode and present different variable correlations for different condition modes. It is termed coarse-grained mode division here since the distribution within each condition slice is assumed to be Gaussian and the temporal information is not used for evaluation. By dividing the complex nonstationary process into different condition modes, the nonlinearity can be reduced to a certain extent. Therefore, linear method is used in this work. However, if nonlinearity cannot be omitted, nonlinear methods can be considered for modeling within each condition mode.

D. Fine-Grained Distribution Evaluation Algorithm

The above condition-mode division is performed only using the static information. The same operation condition does not mean that they have exactly the same process characteristics or Gaussian distribution. It is clear that they may show different changes at the specific operation condition as shown in Fig. 1 where the process is going through the same condition in a different way. Therefore, a fine-grained process analytics will be conducted to further and comprehensively explore the underlying process characteristics, including both the static and dynamic distribution information.

Here, GMM [32]–[34] is constructed to characterize the multiple operating regions in each condition mode. To reduce the modeling complexity, the orthogonal features are used instead of measurement data for GMM. Therefore, SFA is used first to extract the features for each condition mode as follows:

$$\min \ \mathrm{tr}\left[\mathbf{W}_c^{\mathrm{T}}\frac{\dot{\mathbf{X}}_c^{v\mathrm{T}}\dot{\mathbf{X}}_c^{v}}{N_c^v-1}\mathbf{W}_c\right]$$
$$\text{s.t. } \mathbf{W}_c^{\mathrm{T}}\frac{\mathbf{X}_c^{v\mathrm{T}}\mathbf{X}_c^{v}}{N_c^v-1}\mathbf{W}_c=\mathbf{I} \tag{1}$$

where $\mathbf{I}$ denotes the identity matrix. N_c^v is the number of samples that belong to the same condition mode. $\mathbf{W}_c$ denotes the condition-mode SFA model.

It is clear that for SFA, it minimizes the temporal variation of the output SF. In comparison with principal component analysis (PCA) which only describes the static variations by seeking the largest variances of the output features, both the temporal and static variations are extracted by SFA. $\mathbf{W}_c$ will work as the initial condition-mode model to evaluate both the static and dynamic information of the samples within each mode c. SFs are then calculated by projection within each condition mode

$$\mathbf{S}_c^v=\mathbf{X}_c^v\mathbf{W}_c,\ \dot{\mathbf{S}}_c^v=\dot{\mathbf{X}}_c^v\mathbf{W}_c. \tag{2}$$

Here, for simplicity, all SFs are kept except those that have zero variances. J_z denotes the number of features. The static features $\mathbf{S}_c^v$ denote the static process variations within the same condition mode. The dynamic features $\dot{\mathbf{S}}_c^v$ denote the temporal process variations within the cth condition mode. Based on the extracted static features, fast features and SFs, $\mathbf{S}_{f,c}^v$ and $\mathbf{S}_{s,c}^v$, are separated from each other by the rule proposed by Shang *et al.* [25]. Correspondingly, the dynamic features, $\dot{\mathbf{S}}_{f,c}^v$ and $\dot{\mathbf{S}}_{s,c}^v$, are also separated from each other. Therefore, the slow model and fast model are separated, described by $\mathbf{W}_{c,s}$ and $\mathbf{W}_{c,f}$, respectively. It is noted that for the above SCMD algorithm, Gaussian distribution is assumed to evaluate the changes in process characteristics. However, this assumption may not be well satisfied in reality. Besides, considering that the four parts of features have different distribution characteristics, GMM [33] is conducted for each part of SFA features to further perform fine-grained exploration. In this way, the distribution of each condition mode for each part of features can be separated into several submodes, revealing different operation statuses and transient dynamics. The modeling procedure is described as below.

The slow and static (SS) features within the same condition mode c are separated from the static features $\mathbf{S}_c^v$, which is described by a two-way matrix, $\mathbf{S}_{s,c}^v(N_c^v\times R_{s,c}^v)$. The dynamic counterpart, i.e., slow and dynamic (SD) features, $\dot{\mathbf{S}}_{s,c}^v(N_c^v\times R_{s,c}^v)$, is obtained correspondingly. The left features are fast and static (FS), described by $\mathbf{S}_{f,c}^v(N_c^v\times(J_z-R_{s,c}^v))$, which are organized in the same way, as well as the fast and dynamic (FD) features, $\dot{\mathbf{S}}_{f,c}^v(N_c^v\times R_{f,c}^v)$, are obtained correspondingly. In the following, GMM-based monitoring strategy is designed for the static features for example.

Conduct GMM for the SS features and FS features. Assume that M_{ss}^c and M_{fs}^c Gaussian components are available for the slow and fast parts in each static condition mode c. For an arbitrary feature $\mathbf{s}_{s,t}\in\mathbf{S}_{s,c}^v$ and feature $\mathbf{s}_{f,t}\in\mathbf{S}_{f,t}^v$, the probability density function can be expressed as follows:

$$g\left(\mathbf{s}_{s,t}|\theta_{\mathrm{ss}}^c\right)=\sum_{\ell=1}^{M_{\mathrm{ss}}^c}\omega_{\mathrm{ss},\ell}^c\,g\left(\mathbf{s}_{s,t}|\theta_{\mathrm{ss},\ell}^c\right) \tag{3}$$
$$g\left(\mathbf{s}_{f,t}|\theta_{\mathrm{fs}}^c\right)=\sum_{\ell=1}^{M_{\mathrm{fs}}^c}\omega_{\mathrm{fs},\ell}^c\,g\left(\mathbf{s}_{f,t}|\theta_{\mathrm{fs},\ell}^c\right) \tag{4}$$

where $\omega_{\mathrm{ss},\ell}^c$ denotes the prior probability of the ℓth Gaussian component for the SS part in the cth condition mode, satisfying $0\le\omega_{\mathrm{ss},\ell}^c\le 1$ and $\sum_{\ell=1}^{M_{\mathrm{ss}}^c}\omega_{\mathrm{ss},\ell}^c=1$. $\omega_{\mathrm{fs},\ell}^c$ denotes the prior probability of the ℓth Gaussian component for the FS part in the cth condition mode, satisfying $0\le\omega_{\mathrm{fs},\ell}^c\le 1$ and $\sum_{\ell=1}^{M_{\mathrm{fs}}^c}\omega_{\mathrm{fs},\ell}^c=1$. $\theta_{\mathrm{ss},\ell}^c=\{\mu_{\mathrm{ss},\ell}^c,\Sigma_{\mathrm{ss},\ell}^c\}$ which consists of the mean $\mu_{\mathrm{ss},\ell}^c$ and the covariance matrix $\Sigma_{\mathrm{ss},\ell}^c$. Similarly, $\theta_{\mathrm{fs},\ell}^c=\{\mu_{\mathrm{fs},\ell}^c,\Sigma_{\mathrm{fs},\ell}^c\}$ which consists of the mean $\mu_{\mathrm{fs},\ell}^c$ and the covariance matrix $\Sigma_{\mathrm{fs},\ell}^c$.

$g(\mathbf{s}_{s,t}|\theta_{\mathrm{ss},\ell}^c)$ and $g(\mathbf{s}_{f,t}|\theta_{\mathrm{fs},\ell}^c)$ denote the multivariate Gaussian probability density function for the ℓth SS Gaussian component and FS Gaussian component shown as

$$g\left(\mathbf{s}_{s,t}|\theta_{\mathrm{ss},\ell}^c\right)=\frac{\exp\left[-\frac{1}{2}\left(\mathbf{x}-\mu_{\mathrm{ss},\ell}^c\right)^T\Sigma_{\mathrm{ss},\ell}^{c-1}\left(\mathbf{x}-\mu_{\mathrm{ss},\ell}^c\right)\right]}{\sqrt{(2\pi)^{R_{s,c}^v}|\Sigma_{\mathrm{ss},\ell}^c|}} \tag{5}$$
$$g\left(\mathbf{s}_{f,t}|\theta_{\mathrm{fs},\ell}^c\right)=\frac{\exp\left[-\frac{1}{2}\left(\mathbf{x}-\mu_{\mathrm{fs},\ell}^c\right)^T\Sigma_{\mathrm{fs},\ell}^{c-1}\left(\mathbf{x}-\mu_{\mathrm{fs},\ell}^c\right)\right]}{\sqrt{(2\pi)^{J_z-R_{s,c}^v}|\Sigma_{\mathrm{fs},\ell}^c|}}. \tag{6}$$

The parameters of GMM can be estimated using the electromagnetic (EM) algorithm and the number of Gaussian components can be adjusted by the F–J algorithm [32]. In order to monitor the industrial processes, two BID monitoring indices are calculated for the SS and FS features of the cth condition mode, which is denoted as $\mathrm{BID}_{\mathrm{ss},t}$ and $\mathrm{BID}_{\mathrm{fs},t}$, respectively,

$$\mathrm{BID}_{\mathrm{ss},t}=\sum_{\ell=1}^{M_{\mathrm{ss}}^c}p\left(\Theta_{\mathrm{ss},\ell}^c|\mathbf{s}_{s,t}\right)D_L^\ell\left(\mathbf{s}_{s,t},\Theta_{\mathrm{ss},\ell}^c\right) \tag{7}$$
$$\mathrm{BID}_{\mathrm{fs},t}=\sum_{\ell=1}^{M_{\mathrm{fs}}^c}p\left(\Theta_{\mathrm{fs},\ell}^c|\mathbf{s}_{f,t}\right)D_L^\ell\left(\mathbf{s}_{f,t},\Theta_{\mathrm{fs},\ell}^c\right) \tag{8}$$

where $p(\Theta_{\text{ss},\ell}^{c}|\mathbf{s}_{s,t})$ denotes the posterior probability of the feature $\mathbf{s}_{s,t}$ to the ℓth SS Gaussian component $(\Theta_{\text{ss},\ell}^{c})$. $p(\Theta_{\text{fs},\ell}^{c}|\mathbf{s}_{f,t})$ denotes the posterior probability of the feature $\mathbf{s}_{f,t}$ to the ℓth FS Gaussian component. $D_L^{\ell}(\mathbf{s}_{s,t}, \Theta_{\text{ss},\ell}^{c})$ represents the local Mahalanobis distance of the sample $\mathbf{s}_{s,t}$ to the ℓth SS Gaussian component within the cth condition mode. $D_L^{\ell}(\mathbf{s}_{f,t}, \Theta_{\text{fs},\ell}^{c})$ represents the local Mahalanobis distance of the training sample $\mathbf{s}_{f,t}$ to the ℓth FS Gaussian component within the cth condition mode.

The two posterior probability indices in (7) and (8) are calculated by the following two equations:

$$p\left(\Theta_{\text{ss},\ell}^{c}|\mathbf{s}_{s,t}\right) = \frac{p\left(\Theta_{\text{ss},\ell}^{c}\right)p\left(\mathbf{s}_{s,t}|\Theta_{\text{ss},\ell}^{c}\right)}{\sum_{i=1}^{M_{\text{ss}}^{c}} p\left(\mathbf{s}_{s,t}|\Theta_{\text{ss},i}^{c}\right)p\left(\Theta_{\text{ss},i}^{c}\right)} = \frac{\omega_{\text{ss},\ell}^{c} g\left(\mathbf{s}_{s,t}|\theta_{\text{ss},\ell}^{c}\right)}{g\left(\mathbf{s}_{s,t}|\theta_{\text{ss}}^{c}\right)} \tag{9}$$

$$p\left(\Theta_{\text{fs},\ell}^{c}|\mathbf{s}_{f,t}\right) = \frac{p\left(\Theta_{\text{fs},\ell}^{c}\right)p\left(\mathbf{s}_{f,t}|\Theta_{\text{fs},\ell}^{c}\right)}{\sum_{i=1}^{M_{\text{ss}}^{c}} p\left(\mathbf{s}_{f,t}|\Theta_{\text{fs},i}^{c}\right)p\left(\Theta_{\text{fs},i}^{c}\right)} = \frac{\omega_{\text{fs},\ell}^{c} g\left(\mathbf{s}_{s,t}|\theta_{\text{fs},\ell}^{c}\right)}{g\left(\mathbf{s}_{s,t}|\theta_{\text{fs}}^{c}\right)} \quad ss. \tag{10}$$

The value of BID index can quantify the changes in fault effects which is different from the conventional Bayesian inference probability (BIP) index [33] in which one is the largest monitoring value for all abnormalities and, thus, cannot present the different fault effects. For a prespecified confidence level $(1-\alpha)$, the control limits for $\text{BID}_{\text{ss},t}$ and $\text{BID}_{\text{fs},t}$ can be calculated using KDE [31]. The process can be considered to be different from the reference status when either of BID indices is larger than the predefined confidence limit with the specific confidence level.

The above modeling procedure is also conducted for the dynamic features to get $\text{BID}_{\text{sd},t}$ and $\text{BID}_{\text{fd},t}$ indices. Therefore, for the proposed method, four monitoring statistics are constructed for both the static and dynamic levels, which are summarily described in Table I. The different indices of the proposed monitoring method are used with different physical meanings.

E. Online Application

During online application, for any new sample vector $\mathbf{x}_{new}$, its condition is used to indicate which condition mode the current sample lies in. Then the corresponding mode-specific SFA model is adopted to extract the static and dynamic features for both the slow and fast parts

$$\begin{aligned} \mathbf{s}_{s,\text{new}} &= \mathbf{x}_{\text{new}}^{\text{T}}\mathbf{W}_{c,s}, \mathbf{s}_{f,\text{new}} = \mathbf{x}_{\text{new}}^{\text{T}}\mathbf{W}_{c,f} \\ \dot{\mathbf{s}}_{s,\text{new}} &= \dot{\mathbf{x}}_{\text{new}}^{\text{T}}\mathbf{W}_{c,s}, \dot{\mathbf{s}}_{f,\text{new}} = \dot{\mathbf{x}}_{\text{new}}^{\text{T}}\mathbf{W}_{c,f}. \end{aligned} \tag{11}$$

Then for both static and dynamic features, the corresponding condition-mode GMM models are adopted to calculate the BID indices

$$\text{BID}_{\text{ss,new}} = \sum_{\ell=1}^{M_{\text{ss}}^{c}} p\left(\Theta_{\text{ss},\ell}^{c}|\mathbf{s}_{s,\text{new}}\right) D_L^{\ell}\left(\mathbf{s}_{s,\text{new}}, \Theta_{\text{ss},\ell}^{c}\right) \tag{12}$$

$$\text{BID}_{\text{fs,new}} = \sum_{\ell=1}^{M_{\text{fs}}^{c}} p\left(\Theta_{\text{fs},\ell}^{c}|\mathbf{s}_{f,\text{new}}\right) D_L^{\ell}\left(\mathbf{s}_{f,\text{new}}, \Theta_{\text{fs},\ell}^{c}\right) \tag{13}$$

TABLE I

FOUR CASES INDICATED BY MONITORING STATISTICS ("√" REFERS TO THE SITUATION THAT THE CONCERNED STATISTICS ARE UNDER CONTROL LIMITS AND "✗" REFERS TO THE SITUATION THAT AT LEAST ONE STATISTIC IS BEYOND THE CONTROL LIMITS)

Case #	Two Static monitoring statistics $BID_{ss,t}$ $BID_{fs,t}$	Two dynamic monitoring statistics $BID_{sd,t}$ $BID_{fd,t}$	Description
1	√	√	The process is operating with no unexpected condition change and control action.
2	√	✗	The controller is regulating the process with the switch of operation conditions and the regulation results in unregular dynamic behaviors different from those shown in reference data. Besides, the deviation from the current condition is not significant and has not been detected by the static monitoring statistics.
3	✗	√	A deviation from the current condition is noticed but the dynamic varying speed remains similar with that in reference data which means the controller is regulating the process the same as that in reference data.
4	✗	✗	The controller is regulating the process in a quite different way from that in reference data, resulting in unregular dynamic behaviors. Besides, the deviation from the current condition is significant and has been detected by static monitoring statistics.

$$\text{BID}_{\text{sd,new}} = \sum_{\ell=1}^{M_{\text{fs}}^{c}} p\left(\Theta_{\text{sd},\ell}^{c}|\dot{\mathbf{s}}_{s,\text{new}}\right) D_L^{\ell}\left(\dot{\mathbf{s}}_{s,\text{new}}, \Theta_{\text{fs},\ell}^{c}\right) \tag{14}$$

$$\text{BID}_{\text{fd,new}} = \sum_{\ell=1}^{M_{\text{fs}}^{c}} p\left(\Theta_{\text{fd},\ell}^{c}|\dot{\mathbf{s}}_{f,\text{new}}\right) D_L^{\ell}\left(\dot{\mathbf{s}}_{f,\text{new}}, \Theta_{\text{fd},\ell}^{c}\right). \tag{15}$$

For online application, the four monitoring statistics are combined to reveal the current process changes in comparison with their respective control limits for each new sample. If no alarming signals can be found for all monitoring statistics, the process is considered operating under the normal condition. Otherwise, if consistent alarming signals are issued for any statistics, the current process is revealed to have some different behaviors as described in Table I. For the changes in operation conditions, the controller will continuously regulate the process, resulting in time series correlations (i.e., typical dynamics). Therefore, for nonstationary process monitoring, analysis of process dynamics is necessary to distinguish different process changes. Unlike the steady operation condition that could be abstracted as the static distribution, process dynamics can be perceived as the temporal distribution. They carry different information, analogous to the concepts of "position" and "velocity" in physics [21]. The static information can tell the current operation condition which the process lies in, while the dynamic information can clarify whether the process status is changing under the regulation of operation conditions. Therefore, the steady and dynamic variations ought to be

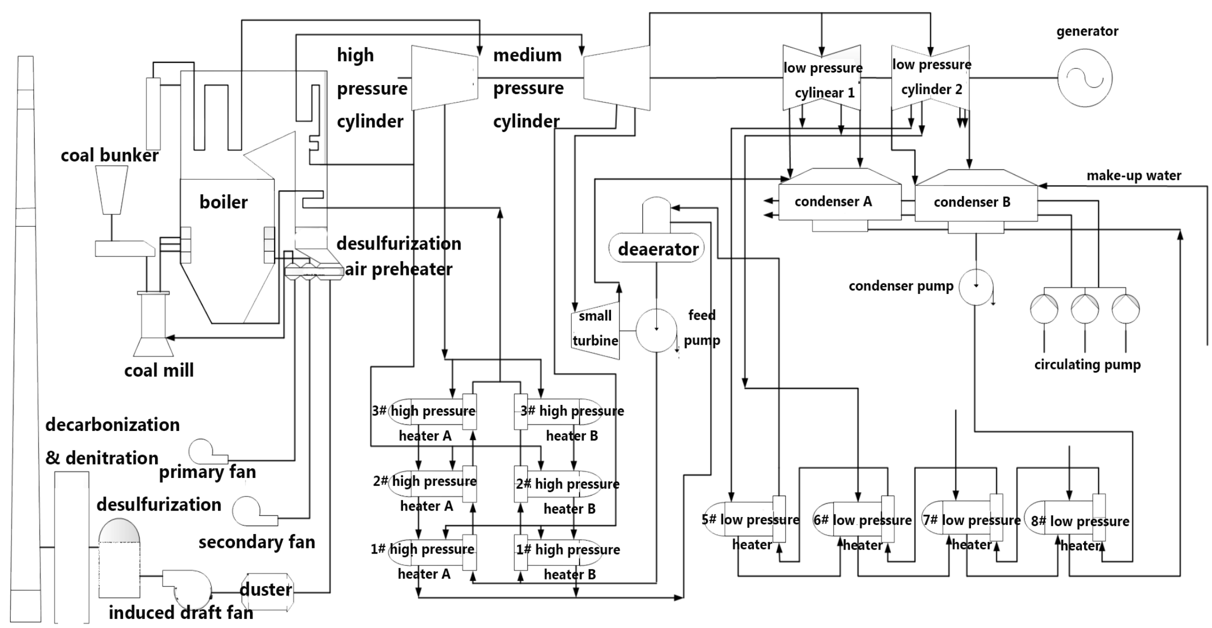


Fig. 4. Schematic of the 1000-MW USC thermal power unit.

explained differently and concurrently monitored, which can infer the physical interpretation of the process changes. Without considering the process dynamics, the traditional multivariate statistical analysis methods may not effectively distinguish between normal status changes and abnormal disturbances.

III. Illustration and Discussion

A. Thermal Power Plant Description

In this section, the proposed method will be illustrated for its ability of dealing with wide-range nonstationary and transient processes. The 1000-MW ultrasupercritical (USC) unit, as shown in Fig. 4, is a highly complex industrial process, exhibiting time-varying, dynamic characteristics and nonstationary nature. For the specific description of process mechanism, readers can be referred to previous work [20], [35] for more details. In this work, coal mill, one of the important machines for USC unit of the thermal power plant, is used as the test bed for the evaluation of the proposed method. This machine works under varying operation conditions, where the coal feed rate as the key input to the coal mill frequently changes with time, leading to typical nonstationary and transient characteristics. This provides a good platform to verify the proposed method regarding data reorganization, condition-mode division and modeling, as well as online monitoring. In this work, 35 measured variables are used for analysis and monitoring purpose. Here, only the first 15 variables are listed in Table II for simplicity. The sampling interval is 1 min. Three testing data sets are collected, where one contains normal and frequent working point switches and the other two cases are real faults that have occurred during the nonstationary operation.

B. Modeling and Analysis

There are more than 15 000 samples for modeling which can cover sufficient changes in operation conditions.

TABLE II

Description of Some Process Variables for Coal Mill

No.	Description	Units
1	Motor coil temperature	ºC
2	Motor bearing temperature	ºC
3	Motor current	A
4	Planetary gear bearing temperature	ºC
5	Rotary separator motor current	A
6	Rotary separator bearing temperature	ºC
7	Rotary separator speed	r/min
8	Tank temperature	ºC
9	Output temperature	ºC
10	Outlet pressure	kPa
11	Import primary air volume	t/h
12	Import primary air temperature	ºC
13	Sealed air pressure	kPa
14	Import primary air pressure	kPa
15	Coal feed rate	t/h

From Fig. 5, it is clear that the coal feed rate changes frequently within a wide range along time direction and correspondingly, the process variable, such as outlet pressure, also changes with time under the influences of coal feed rate, revealing typical nonstationary feature. Considering that the process changes with the input, coal feed rate, it is chosen as the condition indicator. The condition interval β is set to be 0.48 ton/h for the coal feed rate which is about three times of the standard deviation of this variable under the same condition based on the 3α rule [36]. The condition slice is then obtained by collecting the samples within the same condition interval. Within each condition slice, both static samples and the dynamic counterpart are collected. It is deemed that process characteristics stay similar within the same condition slice. Then the proposed partition method is used to evaluate the changes in process characteristics. For different slices, 25 SFs are separated from nine fast features. The constant α is set to be 1.5 here by cross-validation which

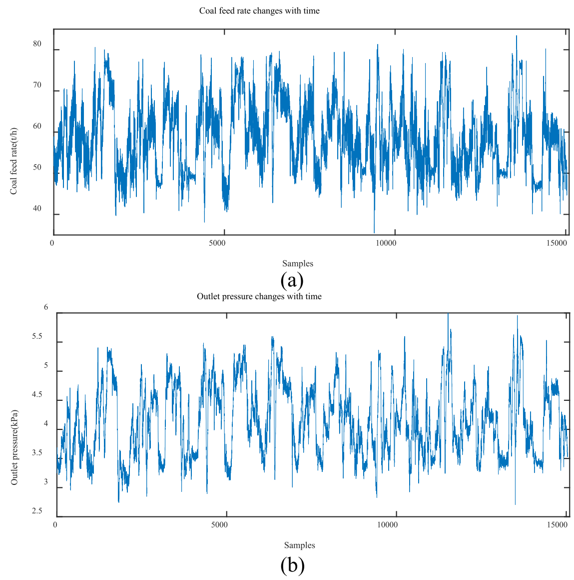


Fig. 5. Trajectories of nonstationary measurement variables for (a) coal feed rate and (b) outlet pressure.

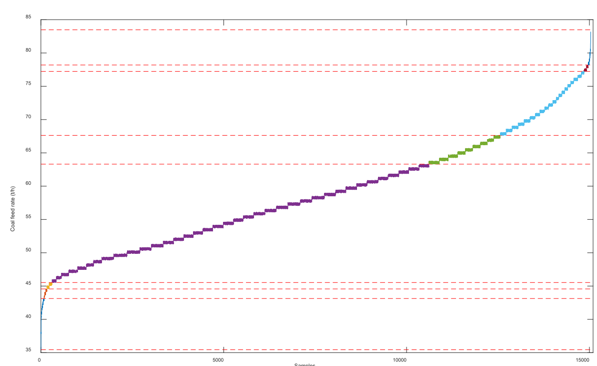


Fig. 6. Condition-mode division result based on the coal feed rate (different condition modes are separated by the red dashed line).

can achieve the best monitoring performance for validation data. Using the proposed algorithm, the whole nonstationary process is automatically partitioned into eight condition modes with the increase in the coal feed rate in which the number of samples is 86, 80, 138, 10 308, 1950, 2290, 113, and 70 for each condition mode. The partition results are shown in Fig. 6. It is clear that the condition slices with similar values of coal feed rates are grouped into the same condition mode. Besides, a large condition mode which has 10 308 samples is observed with the coal feed rate ranging from 45 to 65 ton/h. That is, the process stays within this condition mode most of the time.

As mentioned before, for each condition mode separated by SCMD, it does not mean that the data have the Gaussian distribution. For each condition mode, the sub-SFA model is developed by variable-unfolding slices with the same mode affiliation in which four parts of orthogonal features are separated from each other. In each condition mode, Gaussian distribution is evaluated based on the samples for four types of features as shown in Fig. 7 in which deviations are observed to different extents. It is noted that for the static features,

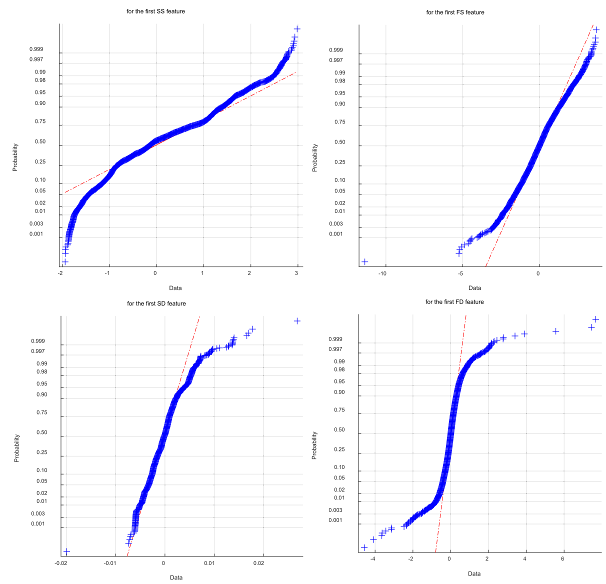


Fig. 7. Normal probability plot of the data for different features in the fourth condition mode.

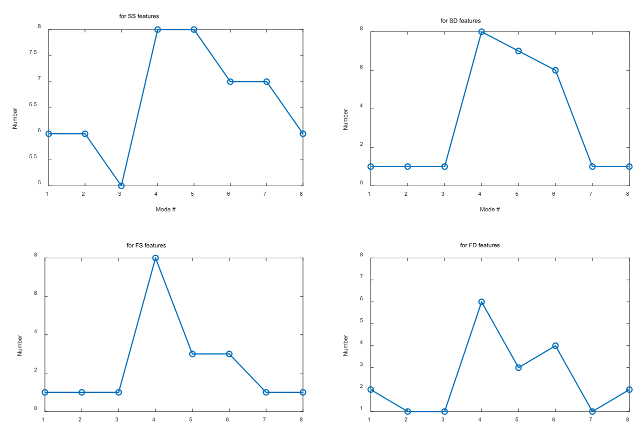


Fig. 8. Number of Gaussian components for different features and different condition modes.

the deviation from Gaussian distribution is larger than that for the dynamic features since they in general capture the changing trajectories. Therefore, it is necessary to conduct a further fine-grained analysis by GMM. For different condition modes, the number of Gaussian components is plotted in Fig. 8 for different features. For the dynamic features, the number of Gaussian components is in general smaller than that of the static features, which agrees well with the real case. That is, the time-varying trend tends to be covered in the static features, which is different from Gaussian distribution more significantly in comparison with the dynamic features.

Based on the condition mode division and SFA-GMM evaluation in each condition mode, monitoring system is developed which is then applied for online application. For the normal case, the monitoring result is presented in Fig. 9 in which almost all samples stay well within the normal region for four monitoring BID indices, revealing a little number of false alarms. It is noted that around the 240th sample, the dynamic

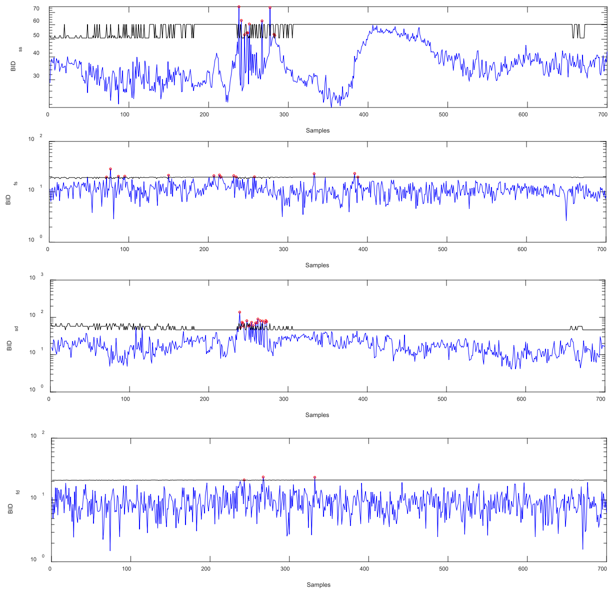


Fig. 9. Monitoring results for the normal case and four types of features (black line denotes the control limit; blue line denotes the BID monitoring statistic; and the red dot denotes the values that are out of control).

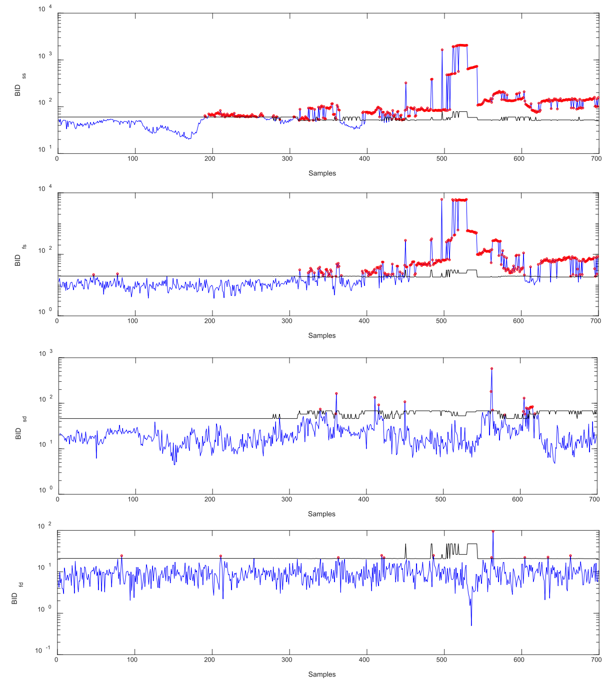


Fig. 10. Monitoring results for Fault #1 and four types of features (black line denotes the control limit; blue line denotes the BID monitoring statistic; and the red dot denotes the values that are out of control).

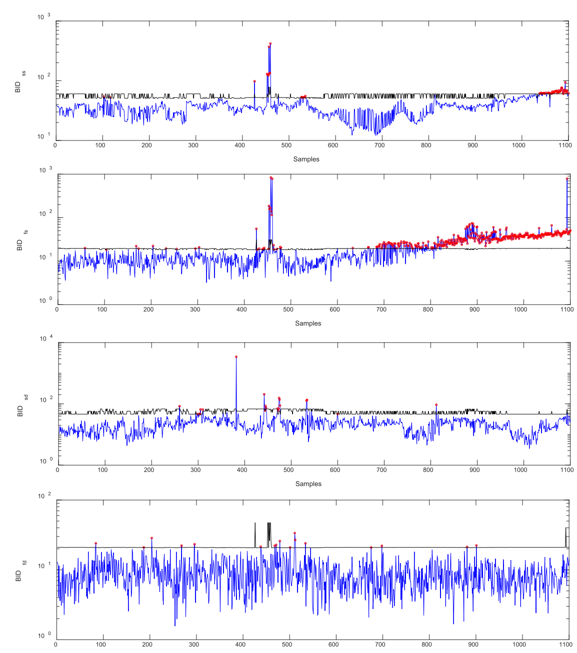


Fig. 11. Monitoring results for Fault #2 and four types of features (black line denotes the control limit; blue line denotes the BID monitoring statistic; and the red dot denotes the values that are out of control).

monitoring statistic BID_{sd} shows disturbances and goes beyond the normal region which agrees well with the real case. At that time, the coal feed rate shows a very fast change from 48 to 70 ton/h, the speed of which is quite different from the case in training data. This change causes the control action which tries to regulate the process to quick follow the switch of operation conditions and the regulation results in irregular dynamic behaviors. In comparison, the static process status still stays well within the normal operation region despite the rapid switch of operation condition as indicated by the static monitoring statistics.

Two fault cases are considered to verify the performance for sensitive fault detection. One fault case is that the outlet temperature shows slow decrease from the 150th sample. Fig. 10 presents the online monitoring performance for Fault Case #1 where from the 195th sample, the disturbance is first detected by the BID index for SS features and then by the BID index for FS features. It reveals that a deviation from the current condition is noticed but the dynamic varying speed remains similar to that in the reference data which means the controller is regulating the process the same as that in the reference data. The second fault case is that the sealed air pressure shows a slow change from the 600th sample. Fig. 11 presents the online monitoring performance for Fault Case #2. From the 689th sample, the deviation is first detected by FS features and then SS features, revealing that the disturbance does not cause different controller regulation. Besides, it is noted that around the 450th sample, the static features are out of control. By checking the coal feed rate, it is found that this is in response to the operation condition that does not exist in the reference data, revealing a deviation from the predefined normal conditions.

For comparison, two methods are used to reveal the necessity of condition-mode division and GMM for fine-scale distribution evaluation. Global SFA method does not conduct mode division and performs SFA for variable unfolding all condition slices. Condition-mode SFA method does not consider GMM for each condition mode. Both of them use four types of statistics including slow and fast, static, and dynamic features for monitoring. For normal cases, the two methods show fewer

TABLE III
COMPARISON OF MONITORING PERFORMANCE EVALUATED BY DETECTION TIME DELAY (SAMPLES) FOR FAULT CASES AND THREE DIFFERENT METHODS

Case #	The proposed method	Global SFA method	Condition-mode SFA method
Fault #1	45	310	112
Fault #2	89	138	119

false alarms because they describe the normal variations with looser confidence limits. However, they cannot indicate the fast changes in the coal feed rate around the 240th sample. Besides, the monitoring performance is compared in Table III for two fault cases as evaluated by the detection time delay (samples). With all condition slices for global SFA modeling, different characteristics are mixed together and different variations are not described closely. Therefore, it is the least sensitive for fault detection. By comparing the global SFA method with the condition-mode SFA method, it reveals the role of mode division. By comparing the proposed method with the condition-mode SFA method, it reveals the role of GMM for distribution evaluation in each condition mode. It is clear that for both fault cases that the proposed method presents more sensitive detection performance than the other methods as evaluated by detection time delay. Besides, the proposed method can provide enhanced physical interpretation for the monitoring results, which can help to clarify which type of process variations is happening.

IV. CONCLUSION

In this work, a condition-driven data analytics and monitoring method is developed for wide-range nonstationary and transient continuous processes. It is based on the fact that there are similar process characteristics within the same condition slice and changes in the process correlations may relate to its condition modes. For the first time, the condition-driven concept is proposed to handle nonstationary characteristics, which is different from the conventional time-driven thought. Application to a real industrial process demonstrates the feasibility and effectiveness of the proposed method. The proposed method provides a novel analytics viewpoint for continuous process monitoring with typical nonstationary and transient characteristics which reveals promising extension to other applications. The potential future work can be how to conduct condition-driven fault variable isolation and root cause analysis in particular for fault diagnosis with no historical fault samples [37].

REFERENCES

[1] N. Sheng, Q. Liu, S. J. Qin, and T. Chai, "Comprehensive monitoring of nonlinear processes based on concurrent kernel projection to latent structures," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 1129–1137, Apr. 2016.

[2] C. Zhao, "A quality-relevant sequential phase partition approach for regression modeling and quality prediction analysis in manufacturing processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 4, pp. 983–991, Oct. 2014.

[3] S. Cuentas, R. Peñabaena-Niebles, and E. Garcia, "Support vector machine in statistical process monitoring: A methodological and analytical review," *Int. J. Adv. Manuf. Technol.*, vol. 91, nos. 1–4, pp. 485–500, Jul. 2017.

[4] W. Yu and C. Zhao, "Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 3, pp. 1083–1091, May 2020.

[5] W. Yu and C. Zhao, "Online fault diagnosis for industrial processes with Bayesian network-based probabilistic ensemble learning strategy," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1922–1932, Oct. 2019.

[6] R. Raveendran and B. Huang, "Two layered mixture Bayesian probabilistic PCA for dynamic process monitoring," *J. Process Control*, vol. 57, pp. 148–163, Sep. 2017.

[7] P. J. Brockwell, and R. A. Davis, *Time Series: Theory and Methods*. New York, NY, USA: Springer, 2006.

[8] R. F. Engle and C. W. J. Granger, "Cointegration and error-correction: Representation, estimation and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, Aug. 1987.

[9] W. Yu and C. Zhao, "Recursive exponential slow feature analysis for fine-scale adaptive processes monitoring with comprehensive operation status identification," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3311–3323, Jun. 2019.

[10] S. Lee and S. B. Kim, "Time-adaptive support vector data description for nonstationary process monitoring," *Eng. Appl. Artif. Intell.*, vol. 68, pp. 18–31, Feb. 2018.

[11] A. Mitra, K. B. Lee, and S. Chakraborti, "An adaptive exponentially weighted moving average-type control chart to monitor the process mean," *Eur. J. Oper. Res.*, vol. 279, no. 3, pp. 902–911, Dec. 2019.

[12] B. Pan *et al.*, "Just-in-time learning based soft sensor with variable selection and weighting optimized by evolutionary optimization for quality prediction of nonlinear processes," *Chem. Eng. Res. Des.*, vol. 144, pp. 285–299, Apr. 2019.

[13] S. Zhang and C. Zhao, "Stationarity test and Bayesian monitoring strategy for fault detection in nonlinear multimode processes," *Chemometric Intell. Lab. Syst.*, vol. 168, pp. 45–61, Sep. 2017.

[14] C. Zhao, Y. Yao, F. Gao, and F. Wang, "Statistical analysis and online monitoring for multimode processes with between-mode transitions," *Chem. Eng. Sci.*, vol. 65, no. 22, pp. 5961–5975, Nov. 2010.

[15] M. Quiñones-Grueiro, A. Prieto-Moreno, C. Verde, and O. Llanes-Santiago, "Data-driven monitoring of multimode continuous processes: A review," *Chemometric Intell. Lab. Syst.*, vol. 189, pp. 56–71, Jun. 2019.

[16] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[17] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[18] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018.

[19] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, Jan. 1984.

[20] C. Zhao and H. Sun, "Dynamic distributed monitoring strategy for large-scale nonstationary processes subject to frequently varying conditions under closed-loop control," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4749–4758, Jun. 2019.

[21] C. Zhao and B. Huang, "A full-condition monitoring method for nonstationary dynamic chemical processes with cointegration and slow feature analysis," *AIChE J.*, vol. 64, no. 5, pp. 1662–1681, May 2018.

[22] Q. Chen, U. Kruger, and A. Y. T. Leung, "Cointegration testing method for monitoring nonstationary processes," *Ind. Eng. Chem. Res.*, vol. 48, no. 7, pp. 3533–3543, Apr. 2009.

[23] Y. Hu and C. Zhao, "Fault diagnosis with dual cointegration analysis of common and specific nonstationary fault variations," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 237–247, Jan. 2020, doi: 10.1109/TASE.2019.2917580.

[24] Z. Chai and C. Zhao, "A fine-grained adversarial network method for cross-domain industrial fault diagnosis," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1432–1442, Jul. 2020.

[25] C. Shang, F. Yang, X. Gao, X. Huang, J. A. K. Suykens, and D. Huang, "Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis," *AIChE J.*, vol. 61, no. 11, pp. 3666–3682, Nov. 2015.

[26] S. Zhang and C. Zhao, "Slow-feature-analysis-based batch process monitoring with comprehensive interpretation of operation condition deviation and dynamic anomaly," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3773–3783, May 2019.

[27] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comput.*, vol. 14, no. 4, pp. 715–770, Apr. 2002.

[28] T. Blaschke, T. Zito, and L. Wiskott, "Independent slow feature analysis and nonlinear blind source separation," *Neural Comput.*, vol. 19, no. 4, pp. 994–1021, Apr. 2007.

[29] H. Sprekeler, T. Zito, and L. Wiskott, "An extension of slow feature analysis for nonlinear blind source separation," *J. Mach. Learn. Res.*, vol. 15, no. 2, pp. 921–947, 2014.

[30] Y. Yao and F. Gao, "A survey on multistage/multiphase statistical modeling methods for batch processes," *Annu. Rev. Control*, vol. 33, no. 2, pp. 172–183, Dec. 2009.

[31] A. Giantomassi, F. Ferracuti, S. Iarlori, G. Ippoliti, and S. Longhi, "Electric motor fault detection and diagnosis by kernel density estimation and Kullback–Leibler divergence based on stator current measurements," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1770–1780, Mar. 2015.

[32] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

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

[34] L. Cai, X. Tian, and S. Chen, "Monitoring nonlinear and non-Gaussian processes using Gaussian mixture model-based weighted kernel independent component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 122–135, Jan. 2017.

[35] C. H. Zhao, H. Sun, and F. Tian, "Total variable decomposition based on sparse cointegration analysis for distributed monitoring of nonstationary industrial processes," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 4, pp. 1542–1549, Jul. 2019.

[36] R. A. Johnson, and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[37] L. Feng and C. Zhao, "Fault description based attribute transfer for zero-sample industrial fault diagnosis," *IEEE Trans. Ind. Informat.*, early access, Apr. 20, 2020, doi: 10.1109/TII.2020.2988208.

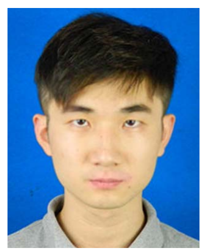
Chunhui Zhao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees from Northeastern University, Shenyang, China, in 2003, 2006, and 2009, respectively.

From January 2009 to January 2012, she was a Post-Doctoral Fellow with the Hong Kong University of Science and Technology, Hong Kong, and the University of California at Santa Barbara, Santa Barbara, CA, USA. Since January 2012, she has been a Professor with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. She has authored or coauthored more than 130 articles in peer-reviewed international journals. Her research interests include statistical machine learning and data mining for industrial application.

Dr. Zhao was a recipient of the National Top 100 Excellent Doctor Thesis Nomination Award, the New Century Excellent Talents in University, China, and the National Science Fund for Excellent Young Scholars, respectively.

Junhao Chen received the B.Eng. degree in mechatronic engineering from the School of Mechanical Engineering, Zhejiang University, Hangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree in control science and engineering with the College of Control Science and Engineering.

His current research interests include process monitoring and pattern recognition.



Hua Jing received the B.Eng. degree in automation from the College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2018. He is currently pursuing the master's degree in control science and engineering with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China.

His current research interests include fault detection and state monitoring.