A Gaussian Feature Analytics-Based DISSIM Method for Fine-Grained Non-Gaussian Process Monitoring

Jie Wang and Chunhui Zhao

Abstract— **Dissimilarity analysis (DISSIM) has been widely used to monitor the Gaussian processes. However, its further application is hindered due to its unavailability to non-Gaussian processes whose data do not satisfy the hypothesis of the Gaussian distributions. To sensitively detect faults and enhance understanding of the non-Gaussian processes, a Gaussian feature analytics-based DISSIM (GDISSIM) method is proposed to monitor both Gaussian information and non-Gaussian information concurrently. The key lies in the separation of information with different statistical properties mixed in the process data. Hence, Gaussian-feature-based analytics is first proposed to devise the extraction, representation, and analysis of the Gaussian information. Besides, multiple Gaussian clusters are estimated for the remained non-Gaussian information integrated with posterior probabilities, enabling both Gaussian information and non-Gaussian information to be readily monitored. Different from the methods based on specific assumptions or approximations, the proposed GDISSIM scheme investigates both non-Gaussian information and Gaussian information and is, therefore, defined as a fine-grained monitoring method. The practical utility and feasibility of the proposed method are verified by a numerical case and a real thermal power plant process.**

Note to Practitioners—**The focus of this article is to develop a GDISSIM monitoring strategy to detect faults for non-Gaussian industrial processes. The GDISSIM strategy separates Gaussian information and non-Gaussian information. The non-Gaussian distribution is further evaluated by multiple Gaussian clusters with their posterior probabilities. In this way, the monitoring of twofold information can be effectively integrated to provide more fine-grained and practical monitoring performance.**

Index Terms— **Dissimilarity analysis (DISSIM), Gaussian latent feature (GLF), non-Gaussian process, process monitoring.**

I. Introduction

The tendency of modern industrial processes moves toward large-scale and complicated, resulting in the frequent occurrence of abnormalities. Process monitoring, which is responsible for detecting abnormalities, therefore, is gaining importance in industries. However, hundreds of variables need to be monitored in a single plant. A large amount of data makes it difficult for human-based monitoring, but it creates some opportunities for machine learning (ML) techniques, which can extract pivotal information from the ocean of data.

Monitoring methods can be primarily classified into two categories: distance- and distribution-based methods [1]. As a distance-based method, principal component analysis (PCA) [2]–[4] constructs a subspace to mirror the underlying variations. Moreover, control limits are built with the Gaussian assumption [5]. Independent component analysis (ICA) [5]–[7] and some extensions, such as kernel ICA [5] and hybrid ICA (HICA) [7], are presented to extract non-Gaussian components. Slow feature analysis (SFA) [8], [9] is widely used to monitor the process. Besides, many deep learning techniques [10]–[14], for example, the deep discriminative representation method [10], a 2-D deep correlated representation learning (2D-DCRL) method [11], and denoising autoencoder and elastic net (DAE-EN) [12], are good for handling nonlinear process data. As distance-based methods mainly focus on the mean and variance of the data, they may miss changes of correlations among process variables. Recently, improvements have been presented to enhance monitoring performance. Thereinto, a distribution-based method, known as dissimilarity analysis (DISSIM) [15], [16], explores the data distribution and is sensitive to the changes of correlations among variables. Many extensions [1], [17]–[19] have been reported, such as the extended DISSIM (EDISSIM) [17] specifically designed for batch processes, the sparse DISSIM (SDISSIM) [1] for the fault isolation, and the canonical variate DISSIM (CVDA) [18] for the incipient fault detection. Nevertheless, the further application of DISSIM is hindered due to its unavailability to the non-Gaussian process.

The non-Gaussian property is one of the typical characteristics of most real industrial processes. Improper indices or thresholds may give misleading results using the Gaussian-hypothesis-based methods. Besides, ignoring the presence of non-Gaussian information may lose an opportunity to provide better insights into the process. However, most existing methods only aim at the coverage of a single aspect. It is difficult for them to analyze a practical process from Gaussian and non-Gaussian aspects simultaneously. Therefore, handling the non-Gaussian property for real industrial processes becomes a challenge for the applications of many monitoring methods.

The purpose of this article is to develop a Gaussian feature analytics-based DISSIM (GDISSIM) method for non-Gaussian process monitoring, which can separate the Gaussian information and non-Gaussian information and monitor these two parts concurrently. Two critical problems will be addressed here. The first one is how to achieve the separation of Gaussian information and non-Gaussian information mixed in the process data; the other one is how to design monitoring statistics based on two partitioned parts to enhance the comprehension of processes. First, the Gaussian latent features (GLFs) can be extracted from the data by solving the constructed optimization function, leaving the residual subspace as non-Gaussian information. Since the extracted GLFs follow the Gaussian distribution, the DISSIM index can be readily calculated. For the non-Gaussian residuals, multiple Gaussian clusters are estimated and integrated with posterior probabilities, such that the monitoring for the Gaussian and non-Gaussian distributions could be readily synthesized. The main contributions of this work are summarized as follows.

1) A Gaussian-latent-features' extraction method is proposed to separate the Gaussian information and non-Gaussian information coexisting in the process data, overcoming the Gaussian-distributed requirement.

Manuscript received January 10, 2020; revised April 16, 2020 and June 16, 2020; accepted July 28, 2020. Date of publication August 18, 2020; date of current version October 6, 2020. This article was recommended for publication by Associate Editor C.-Y. Lee and Editor X. Xie upon evaluation of the reviewers' comments. This work was supported in part by the NSFC–Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant U1709211 and in part by the Zhejiang Key Research and Development Project under Grant 2019C01048. *(Corresponding author: Chunhui Zhao.)*

The authors are with the College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: 11832018@zju.edu.cn; chhzhao@zju.edu.cn).

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

Digital Object Identifier 10.1109/TASE.2020.3013654

2) A fine-grained monitoring scheme is proposed for non-Gaussian processes by integrating multiple Gaussian clusters with posterior probabilities and evaluating their distributions in the frame of DISSIM, enhancing the monitoring performance of practical processes.

The rest of this article is organized as follows. In Section II, the DISSIM is revisited. The proposed GDISSIM scheme is presented in Section III. The feasibility of the GDISSIM method is demonstrated in numerical and real industrial examples in Section IV. Finally, Section V provides concluding remarks.

II. Dissimilarity Index

DISSIM [15], [16] is a classic method for quantifying the dissimilarity between two data distribution. Here, a brief revisit of DISSIM is presented.

Given two data sets, $\mathbf{X}_1$ and $\mathbf{X}_2$, considered to be a reference and test data, respectively, the reference set is normalized to have zero-mean and unit-variance first. Then, the normalization information is adopted to handle the test one. The processing outlines are summarized as follows.

Step 1: Calculate the covariance matrices of $\mathbf{X}_1$ and $\mathbf{X}_2$ by (1), where $\mathbf{X}_i$ consists of N_i samples of J variables

$$\mathbf{R}_i = \frac{1}{N_i}\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i. \tag{1}$$

Step 2: Calculate the covariance matrix of the mixture of two data sets by the following equation:

$$\mathbf{R} = \frac{N_1}{N_1+N_2}\mathbf{R}_1 + \frac{N_2}{N_1+N_2}\mathbf{R}_2. \tag{2}$$

Step 3: Conduct the eigenvalue decomposition on $\mathbf{R}$, and obtain the transform matrix $\mathbf{P}$

$$\mathbf{R}\mathbf{P}_0 = \mathbf{P}_0\mathbf{\Lambda} \tag{3}$$

$$\mathbf{P} = \mathbf{P}_0\mathbf{\Lambda}^{-1/2}. \tag{4}$$

Step 4: Transform $\mathbf{X}_i$ into $\mathbf{Y}_i$ by (5), and calculate the covariance of $\mathbf{Y}_i$ according to (6). The obtained covariance matrices satisfy the equations of $\mathbf{S}_1 + \mathbf{S}_2 = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix

$$\mathbf{Y}_i = \sqrt{\frac{N_i}{N_1+N_2}}\mathbf{X}_i\mathbf{P} \tag{5}$$

$$S_i = \frac{1}{N_i}\mathbf{Y}_i^{\mathrm{T}}\mathbf{Y}_i. \tag{6}$$

Step 5: Apply the eigenvalue decomposition to $\mathbf{S}_i$, and obtain j pairs of eigenvalues λ_i^j and eigenvectors $\boldsymbol{\omega}_i^j$ by the following equation:

$$\mathbf{S}_i\boldsymbol{\omega}_i^j = \lambda_i^j\boldsymbol{\omega}_i^j. \tag{7}$$

Step 6: Calculate the dissimilarity index by the following equation:

$$D = \mathrm{diss}\left(\mathbf{X}_1, \mathbf{X}_2\right) = \frac{4}{J}\sum_{j=1}^{J}\left(\lambda_i^j - 0.5\right)^2 \tag{8}$$

where λ_i^j denotes eigenvalues of $\mathbf{S}_i$ obtained in Step 5, and the superscript j represents the number of eigenvalues, whereas the subscript i can be set to be 1 or 2 and is consistent with the subscript in $\mathbf{S}_1$ and $\mathbf{S}_2$.

When $\mathbf{X}_2$ is similar to $\mathbf{X}_1$, eigenvalues of $\mathbf{S}_1$ and $\mathbf{S}_2$ are close to 0.5; hence, the DISSIM index tends to be zero. However, when $\mathbf{X}_2$ is quite different from $\mathbf{X}_1$, the index tends to be 1.

III. Methodology

A. Gaussian Latent Feature Extraction

For a multivariate process, the observed process data can be regarded as a linear superposition of many Gaussian and non-Gaussian signals. The proposed extraction method aims to separate these two parts of information by finding a demixing projection, along which the distribution of the projected feature follows the Gaussian distribution. This is motivated by the stationary subspace analysis (SSA) algorithm [20], designed for nonstationary processes [21].

For a process with J variables, N collected samples forms the matrix $\mathbf{X}\,(N \times J)$. The mean and variance of $\mathbf{X}$ is denoted as $\bar{\boldsymbol{\mu}}\,(1 \times J)$ and $\bar{\mathbf{\Sigma}}\,(J \times J)$. The goal is to determine a projection matrix $\mathbf{B}\,(J \times j)$, such that the projected features $\mathbf{S}\,(N \times j)$ follow the Gaussian distribution, where $j < J$

$$\mathbf{S} = \mathbf{X}\mathbf{B}. \tag{9}$$

Considering that the demixing matrix $\mathbf{B}$ is not unique, any projections that eliminate the Gaussian contributions in the remained residuals are valid. As for GLFs, their joint distribution should remain unchanged over time so that the obtained GLFs are stationary. To measure the distribution of the input data over a while, the matrix $\mathbf{X}$ is divided into L segments, where L cannot be less than half the number of variables [20]. The distribution of each segment $i\,(1 < i < L)$ can be characterized by its empirical mean and covariance matrix, denoted as $\left(\hat{\boldsymbol{\mu}}_i, \hat{\mathbf{\Sigma}}_i\right)$, while the distribution of the projected data in each segment can be represented as $\aleph\left(\hat{\boldsymbol{\mu}}_{S,i}, \hat{\mathbf{\Sigma}}_{S,i}\right)$, where $\hat{\boldsymbol{\mu}}_{S,i} = \hat{\boldsymbol{\mu}}_i\mathbf{B}$ and $\hat{\mathbf{\Sigma}}_{S,i} = \mathbf{B}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_i\mathbf{B}$. Based on the Kullback–Leibler (KL) divergence [22], which is adapted to quantify the difference between two probability densities, the optimization function is constructed as follows:

$$\mathbf{B} = \arg\min\left\{\sum_{i=1}^{L} D_{KL}\left[\aleph\left(\hat{\boldsymbol{\mu}}_{S,i}, \hat{\mathbf{\Sigma}}_{S,i}\right)\right] || N(\bar{\boldsymbol{\mu}}\boldsymbol{B}, \mathbf{B}^{\mathrm{T}}\bar{\mathbf{\Sigma}}\mathbf{B})\right\}. \tag{10}$$

The target is to minimize the summation of KL divergences between the projected data with a Gaussian distribution over L segments. This allows the distribution of GLFs to be as similar as possible to the given Gaussian distribution. The solution of $\mathbf{B}$ can be obtained by the gradient descent algorithm [23].

As for determining the number of GLFs, set the number j equal to one first. The Kolmogorov–Smirnov (KS) test [24] is employed to examine the properties of GLFs qualitatively. If all the j GLFs are proved to have the Gaussian property, increase the number of j one by one until non-Gaussian signals are detected in the GLFs. It indicates that the Gaussian information has been completely extracted, while the information remained follows the non-Gaussian distribution, termed non-Gaussian residuals.

After that, the above-extracted GLFs can be utilized to reconstruct the original data matrix by the principle of least squares (LS) [25]. The space reconstructed by GLFs is denoted as $\hat{\mathbf{X}}$. The differences between $\hat{\mathbf{X}}$ and $\mathbf{X}$ are defined as non-Gaussian residuals $\mathbf{E}$, which can be calculated by the following equation:

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X} - \mathbf{S}\mathbf{P}^{\mathrm{T}} = \mathbf{X} - \mathbf{S}\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{X}. \tag{11}$$

B. Analysis of Non-Gaussian Information

In this section, the Gaussian mixture model (GMM) [26] is utilized to approximate non-Gaussian densities with a weighted sum of multiple Gaussian probability density functions (pdfs). An important merit of GMM is that an arbitrary probability density can be approximated by a mixture of multiple Gaussian pdfs.

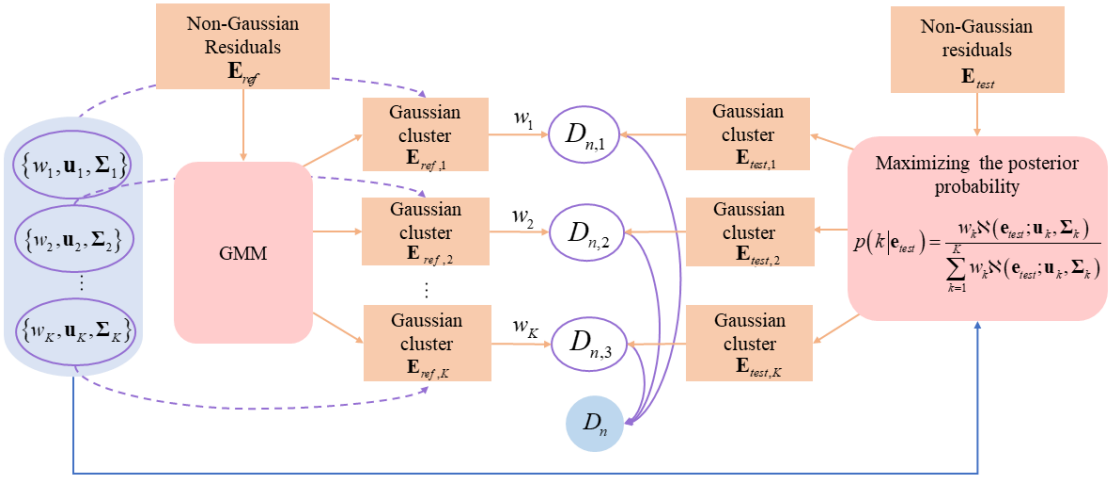

Fig. 1. Schematic of the monitoring index for the non-Gaussian information.

Consider that $\mathbf{e}$ is a sample of non-Gaussian residuals $\mathbf{E}$, and K is the number of Gaussian clusters. The pdf provided by GMM is given as follows:

$$P(\mathbf{e}|\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k p(\mathbf{e}|k) \tag{12}$$

$$p(\mathbf{e}|k) = \frac{\exp\left[-\frac{1}{2}(\mathbf{e}-\mathbf{u}_k)^{\mathrm{T}} \mathbf{\Sigma}_k^{-1}(\mathbf{e}-\mathbf{u}_k)\right]}{(2\pi)^{J/2}|\mathbf{\Sigma}_k|^{1/2}} \tag{13}$$

where w_k is the prior probability of the sample belonging to the kth Gaussian component, and $\mathbf{u}_k$ and $\mathbf{\Sigma}_k$ are the mean vector and covariance matrix of the kth Gaussian component, respectively. The parameter vector $\boldsymbol{\theta}$ is denoted by the following equation:

$$\boldsymbol{\theta} = \{w_k, \mathbf{u}_k, \mathbf{\Sigma}_k; k = 1, 2, \ldots, K\}. \tag{14}$$

The major purpose of GMM is to estimate the model parameter vector. The expectation–maximization (EM) [27] and the Figueiredo–Jain (FJ) algorithm [28] are often used to maximize the log-likelihood function. Here, the FJ algorithm is adopted to automatically choose the number of K. With the normalized confidence score calculated for $\mathbf{e}_n$ by (15), we can achieve the classification results based on the estimated parameters by finding k that maximizes the posterior probability

$$p(k|\mathbf{e}_n) = \frac{w_k \aleph(\mathbf{e}_n; \mathbf{u}_k, \mathbf{\Sigma}_k)}{\sum_{k=1}^{K} w_k \aleph(\mathbf{e}_n; \mathbf{u}_k, \mathbf{\Sigma}_k)}. \tag{15}$$

The non-Gaussian residuals $\mathbf{E}$ can be described as multiple weighted Gaussian distributions, i.e., $\mathbf{E}_k(t), k = 1, 2, \ldots, K$.

C. Monitoring Indices Design Based on GDISSIM

1) Monitoring Index for Gaussian Information: After the separation of GLFs and non-Gaussian residuals for both reference data and test data, the GLFs, on behalf of the Gaussian information, are used for the construction of the monitoring index. To explore data characteristics and structures, we use DISSIM to monitor the distribution of GLFs, such that the proposed monitoring index can not only detect faults that cause the variations in the mean and variance of data but also identify the changes of data structures. Due to the Gaussian statistical properties, the dissimilarity index [16] between two distributions of GLFs can be naturally defined by the following equation:

$$D_g = \mathrm{diss}(\mathbf{S}_{\mathrm{ref}}, \mathbf{S}_{\mathrm{test}}) \tag{16}$$

where the subscript g indicates that the index is for the Gaussian part.

2) Monitoring Index for Non-Gaussian Information: A compact DISSIM index is designed based on the estimated multiple Gaussian clusters for the reference data to endow the development of the non-Gaussian part with the ease of implementation.

As introduced earlier, K Gaussian clusters can be estimated for the reference non-Gaussian residuals $\mathbf{E}_{\mathrm{ref}}$ with parameters denoted as $\boldsymbol{\theta}_{\mathrm{ref}} = \{w_k, \mathbf{u}_k, \mathbf{\Sigma}_k; k = 1, 2, \ldots, K\}$. Especially, the posterior probability that each sample of $\mathbf{E}_{\mathrm{test}}$ belongs to K Gaussian clusters can be calculated according to the pdf based on $\mathbf{u}_k$ and $\mathbf{\Sigma}_k$, such that $\mathbf{E}_{\mathrm{test}}$ can be classified into $\mathbf{E}_{\mathrm{test},k}$ too. The value of k regarding the maximum posterior probability is adopted as the cluster label for each sample.

After that, two classified non-Gaussian residuals, $\mathbf{E}_{\mathrm{ref},k}$ for the reference and $\mathbf{E}_{\mathrm{test},k}$ for the test, can be obtained. To construct the monitoring statistic for non-Gaussian information, it is an intuitive idea that by applying DISSIM to $\{\mathbf{E}_{\mathrm{ref},k}, \mathbf{E}_{\mathrm{test},k}\}$, the index $D_{n,k}$ can be computed and used to measure the similarity between each pair of Gaussian clusters

$$D_{n,k} = \mathrm{diss}(\mathbf{E}_{\mathrm{ref},k}, \mathbf{E}_{\mathrm{test},k}), \quad k = 1, 2, \ldots, K. \tag{17}$$

Moreover, D_n can be designed by integrating a series of $D_{n,k}$ with prior weighting factor w_k

$$D_n = \sum_{k=1}^{K} w_k D_{n,k}. \tag{18}$$

In summary, the diagram of the proposed monitoring index for non-Gaussian information is shown in Fig. 1.

D. GDISSIM-Based Concurrent Monitoring Strategy

1) Off-Line Modeling: For applying the proposed method, off-line models and control limits are built by the following procedures.

Step 1 *(Data construction and Normalization):* For the reference data set with M variables, N archived normal samples are composed of a matrix $\bar{\mathbf{X}}_{\mathrm{ref}}(N \times J)$. Then, scale the raw data to have the zero mean and unit variance to obtain $\mathbf{X}_{\mathrm{ref}}$.

Step 2 *(Separation of GLFs and Non-Gaussian Residuals):* Perform the Gaussian feature extraction on $\mathbf{X}_{\mathrm{ref}}$ to separate GLFs $\mathbf{S}_{\mathrm{ref}}$ and non-Gaussian residuals $\mathbf{E}_{\mathrm{ref}}$ by projection $\mathbf{B}$.

TABLE I
PARAMETERS DEFINED IN THE GDISSIM METHOD

Parameter	Meaning	Parameter	Meaning
j	The number of GLFs	$\mathbf{S}$	The matrix of GLFs
L	The number of segments	$\mathbf{E}$	Non-Gaussian residuals
K	The number of Gaussian clusters	D_g	The Gaussian monitoring index
J	The number of variables	D_n	The non-Gaussian monitoring index
N	The number of samples	w_k	The prior probability of GMM
$\mathbf{X}$	The process data	$\mathbf{u}_k$	The mean vector in GMM
$\mathbf{B}$	The projection matrix	$\mathbf{\Sigma}_k$	The covariance matrix in GMM

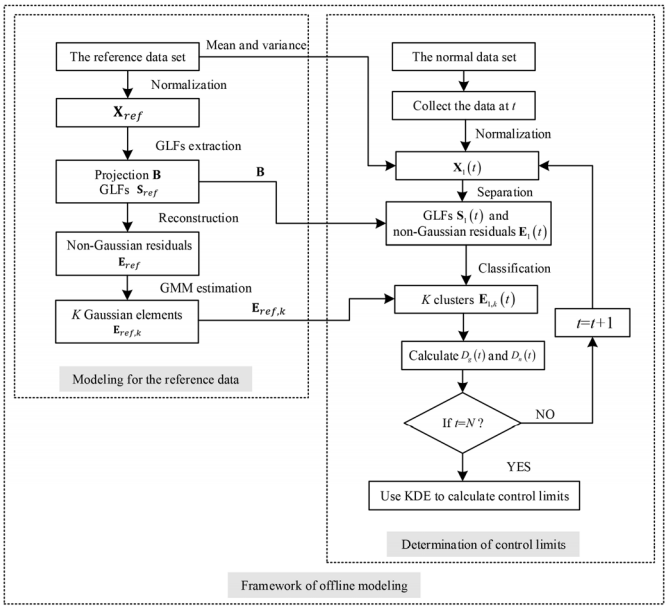

Fig. 2. Flow diagram of GDISSIM-based off-line modeling.

Step 3 *(Analysis of Non-Gaussian Information):* The non-Gaussian residuals are analyzed by GMM to obtain parameters $\boldsymbol{\theta}$ and K Gaussian clusters $\mathbf{E}_{\text{ref},k}$.

Step 4 *(Determination of Control Limits):* Another normal data set is formed and normalized to be $\mathbf{X}_1$ using the information of reference data. A fixed-length moving window is used to obtain $\mathbf{X}_1(t)$ at instance t. Use matrix $\mathbf{B}$ to separate $\mathbf{S}_1(t)$ and $\mathbf{E}_1(t)$. Obtain K clusters of $\mathbf{E}_{1,k}(t)$ based on parameters $\boldsymbol{\theta}$, and calculate $D_g(t)$ and $D_n(t)$. When the sliding window reaches the end of $\mathbf{X}_1$, a series of $D_g(t)$ and $D_n(t)$ is obtained. To improve the monitoring accuracy, the kernel density estimation (KDE) approach [29] is used to extract an empirical pdf and determine the control limits with a specific significance level.

The flow diagram of GDISSIM-based off-line modeling is shown in Fig. 2. The parameters defined in GDISSIM are summarized in Table I.

2) Online Modeling:

Step 1 *(Data Construction and Normalization):* For a test data set $\bar{\mathbf{X}}_2(t)$, scale it using the normalization information of the reference data to obtain $\mathbf{X}_2(t)$.

Step 2 *(Separation of GLFs and Non-Gaussian Residuals for Test Data):* The predesigned projection matrix $\mathbf{B}$ is used to separate $\mathbf{S}_2(t)$ and $\mathbf{E}_2(t)$, respectively.

TABLE II
PROCESS STATICS INDICATED BY MONITORING STATISTICS OF GDISSIM ("UNDER" AND "BEYOND" REFER TO THE SITUATIONS THAT THE STATISTIC IS UNDER AND BEYOND THE CONTROL LIMIT, RESPECTIVELY)

Case	D_g	D_n	Process status
1	Under	Under	Process is operating in a normal condition.
2	Under	Beyond	An anomaly in non-Gaussian information is detected.
3	Beyond	Under	An anomaly in Gaussian information is detected.
4	Beyond	Beyond	An anomaly with both influence in Gaussian and non-Gaussian information is detected.

Step 3 *(Monitoring for the Gaussian Information):* $D_g(t)$ is calculated by applying DISSIM to $\mathbf{S}_{\text{ref}}$ and $\mathbf{S}_2(t)$.

Step 4 *(Monitoring for the Non-Gaussian Information):* The index $D_n(t)$ for the non-Gaussian information is calculated based on $\mathbf{E}_{\text{ref},k}$ and $\mathbf{E}_{2,k}(t)$ for $k = 1, 2, \ldots, K$.

Step 5 *(Data Update):* If one of two statistics exceeds the corresponding limit, faults have been detected; otherwise, let $t = t+1$ to update data, and begin a new cycle from step 1.

Two indices are integrated to monitor the process. In-control statistics reveal that the process is normal. Otherwise, it indicates that the process gets disrupted. Corresponding to alarms issued by different monitoring indices, a different piece of information is provided and summarized in Table II.

IV. CASE STUDY

A. Numerical Example

To verify the effectiveness of the proposed method, a numerical example is designed here. The basic idea is that the Gaussian information is the same in the reference and the test data, while the non-Gaussian information is different. Three univariate Gaussian signals each containing 1000 samples are generated as the Gaussian sources using $\aleph(6, 2)$, $\aleph(0, 2)$, and $\aleph(-1, 1)$. Two non-Gaussian sources for the reference and the test data are generated by the different pdfs of 2-D GMM with random mean vectors and covariance matrices. To make it more practical, a mixing matrix $\mathbf{W}$ is generated in random to mix the sources to form the observed data $\mathbf{X}$ and $\mathbf{Y}$ for the reference and the test.

GLFs of two data sets, i.e., $\mathbf{S}_x$ and $\mathbf{S}_y$, can be gained by the proposed method. The quantile–quantile (QQ) plots of $\mathbf{S}_x$ and $\mathbf{S}_y$ are shown in Fig. 3, verifying that the extracted GLFs obey the Gaussian distribution. With the aid of GMM, $\mathbf{E}_x$ is divided into $\mathbf{E}_{x,1} \in \mathbb{R}^{905\times5}$ and $\mathbf{E}_{x,2} \in \mathbb{R}^{95\times5}$, while $\mathbf{E}_y$ is composed of $\mathbf{E}_{y,1} \in \mathbb{R}^{832\times5}$ and $\mathbf{E}_{y,2} \in \mathbb{R}^{168\times5}$, as shown in Fig. 4. Two indices D_g and D_n are compared with the conventional DISSIM index D in Table III, from which we can see that the small value of D_g implies the large similarity between two data sets in terms of Gaussian information. The large value of D_n shows that the dominant difference occurs in the non-Gaussian information. However, the DISSIM method can only detect the enhancement in D without other detailed information.

B. Thermal Power Plant

As a complex industrial process, the thermal power plant [30] is used to verify the performance of the proposed method here. The thermal system of a power plant consists of the steam turbine system and boiler system. Fault-free and faulty data sets were sampled from the Zhejiang Energy Group; 33 variables are included in the data set,

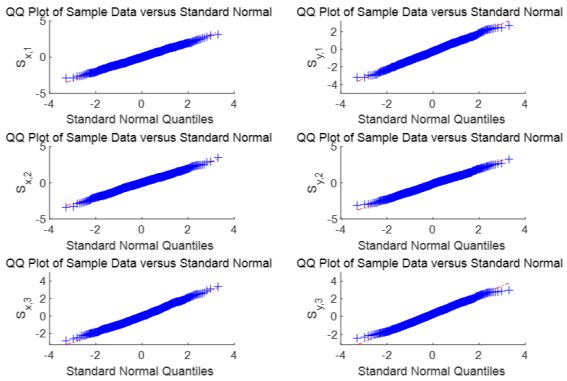

Fig. 3. GLFs and QQ plots of (a) reference data and (b) test data in Case #1.

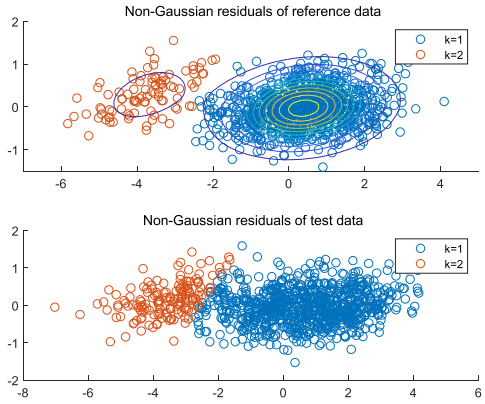

Fig. 4. Estimated Gaussian clusters of the non-Gaussian residuals in Case #1.

including pressure, current, and so on. The sampling interval for all variables is 1 min.

Two normal data sets containing 800 samples are picked as the reference and off-modeling data to calculate control limits. For the test set, 360 samples are included, in which the tearing temperature of the blower increases from the 297th sample, indicating the occurrence of an abnormality. During the off-line modeling, the length of the moving window is set to be 100. The control limits are built with 99% as the significance level. As introduced earlier, the number of GLFs is set to be one first and is increased one by one until non-Gaussian signals are detected in the GLFs. In this case, the number of GLFs is 20.

This section compares the proposed GDISSIM method with DISSIM and two related methods from recent works: HICA and CVDA. Thresholds in HICA are chosen by trial and error. For CVDA, lags included in the Hankel matrices are both set to 15, and the number of dominant features is selected as 25 according to the Akaike criterion [31]. During the online monitoring, two statistics of GDISSIM are shown in Fig. 5. The monitoring charts of DISSIM, HICA, and CVDA are shown in Figs. 6–8. In terms of the detection time, GDISSIM is seen to be the most sensitive as it detects the fault at the 307th sample, illustrating that the Gaussian information has been affected by the fault. This index merits the free of the interference of non-Gaussian information. The CVDA comes the second with detection at the 317th sample, while DISSIM and HICA both suffer from a large delay (at the 338th and 322nd samples). Due to the Gaussian assumption, which can be satisfied in reality, the DISSIM loses its sensitivity by approximating the non-Gaussian

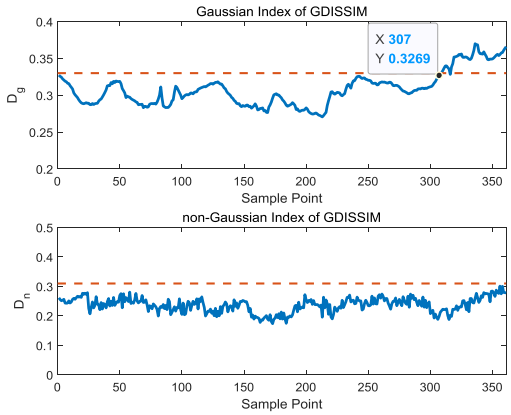

Fig. 5. Monitoring results of the GDISSIM method in the thermal power plant.

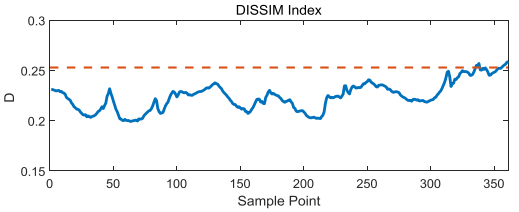

Fig. 6. Monitoring results of the DISSIM method in the thermal power plant.

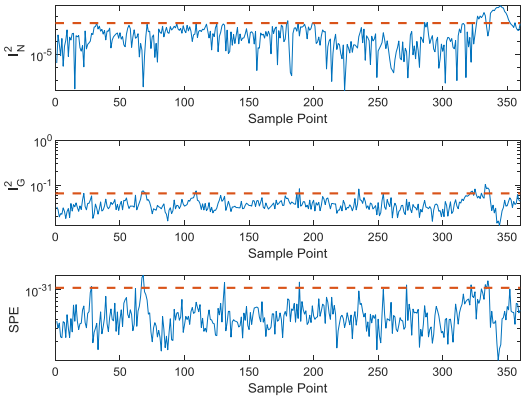

Fig. 7. Monitoring results of the HICA method in the thermal power plant.

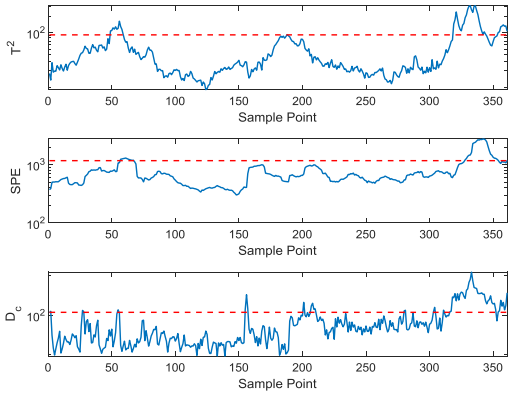

Fig. 8. Monitoring results of the CVDA method in the thermal power plant.

distribution into the Gaussian one. The large delay of the HICA-based index may result from its distance-based statistics. These results validate the importance of the proper separation of the Gaussian

TABLE III
COMPARISON OF INDICES BETWEEN THE TWO DATA SETS

DISSIM	GDISSIM		
	Gaussian	Non-Gaussian	
D=0.09	D_g=0.09	$D_{n,1}=0.15$	$w_1=0.91$
		$D_{n,2}=0.34$	$w_2=0.09$
		$D_n=0.17$	

TABLE IV
PERFORMANCE COMPARISON OF FOUR MONITORING METHODS

Method	FDR	FAR	Offline modeling time (ms)	Online execution time (ms)
DISSIM	14.06%	0.00%	1813.84	0.96
HICA	67.92%	5.74%	303.53	0.07
CVDA	76.56%	11.15%	806.44	0.03
GDISSIM	84.38%	0.3378%	8268.44	1.30

information and non-Gaussian information for fault detection because faults can bring out changes in data structures, such that process data become differently correlated and non-Gaussian distributed.

A good detection index should have a low false alarm rate (FAR) and a high fault detection rate (FDR). The results in Table IV are reported to show that the FDR of GDISSIM is as high as 84.38%, while the other three methods all missed the fault more or less after the detection time with lower FDR. Besides, the comparison of FAR reflects that GDISSIM has fewer false alarms than HICA and CVDA due to the separate and concurrent monitoring of two parts of information with different statistical properties. Besides, the average run length (ARL) [32] is a well-known indicator to measure the quality of the control chart. Here, the in-control ARL is adopted to compare the abovementioned four methods. The ARL for GDISSIM, DISSIM, HICA, and CVDA is 200.39, 305.29, 120.86, and 131.14. Since larger ARL indicates higher monitoring quality, the proposed GDISSIM is verified to be more reliable than HICA and CVDA. Though the DISSIM has the largest ARL and the lowest FAR, the detection sensitivity and efficiency are severely sacrificed as evaluated by FDR.

As for modeling time shown in Table IV, GDISSIM needs the most time for building the off-line model due to the optimization of the projection matrix for extracting GLFs and the estimation of GMM for processing non-Gaussian residuals. Yet, online execution time consumed by GDISSIM is slightly more than that of the other three methods. It is because the GLFs of the test data set are directly projected and the multiple Gaussian clusters are estimated based on the GMM parameters of the reference data rather than exploring from scratch.

Being the most sensitive and reliable one in terms of FAR, FDR, and detection time, the proposed GDISSIM method outperforms the other three methods. With the proper separation of the Gaussian and non-Gaussian distributions and the fine-grained monitoring of two parts of information, monitoring indices can be more sensitive and reliable. However, as a common issue of DISSIM-based methods, there are still time delays of GDISSIM-based indices. The off-line modeling time of GDISSIM is a little bit more than other related methods. Moreover, exploring the contributions of variables to the GLFs and non-Gaussian residuals may provide more useful information for tracing the dominant factors of faults and abnormalities. This is left for future work.

V. CONCLUSION

In this work, a fine-grained monitoring method, termed GDISSIM, has been developed with concurrent analytics of GLFs and non-Gaussian residuals for non-Gaussian processes. The proposed method can not only extract GLFs but also explore the non-Gaussian information. On the basis of GLFs and non-Gaussian residuals, a process monitoring strategy with two kinds of dissimilarity statistics has been developed to detect abnormalities. Meanwhile, the proposed method can provide more reliable and comprehensive information about process statues. In comparison with other methods, the proposed method relaxes the Gaussian assumption of the process data and broadens the application scope. Both applications to numerical simulation and a real industrial process demonstrate the feasibility and effectiveness of the proposed method.

REFERENCES

[1] C. Zhao and F. Gao, "A sparse dissimilarity analysis algorithm for incipient fault isolation with no priori fault information," *Control Eng. Pract.*, vol. 65, pp. 70–82, Aug. 2017, doi: 10.1016/j.conengprac.2017.05.005.

[2] Q. Liu, S. J. Qin, and T. Chai, "Decentralized fault diagnosis of continuous annealing processes based on multilevel PCA," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 687–698, Jul. 2013.

[3] C. Zhao and F. Gao, "Fault-relevant principal component analysis (FPCA) method for multivariate statistical modeling and process monitoring," *Chemometric Intell. Lab. Syst.*, vol. 133, pp. 1–16, Apr. 2014, doi: 10.1016/j.chemolab.2014.01.009.

[4] C. Tong and X. Yan, "A novel decentralized process monitoring scheme using a modified multiblock PCA algorithm," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1129–1138, Apr. 2017.

[5] 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.

[6] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999, doi: 10.1109/72.761722.

[7] S. Zhang and C. Zhao, "Hybrid independent component analysis (H-ICA) with simultaneous analysis of high-order and second-order statistics for industrial process monitoring," *Chemometric Intell. Lab. Syst.*, vol. 185, pp. 47–58, Feb. 2019.

[8] J. Zheng and C. Zhao, "Online monitoring of performance variations and process dynamic anomalies with performance-relevant full decomposition of slow feature analysis," *J. Process Control*, vol. 80, pp. 89–102, Aug. 2019, doi: 10.1016/j.jprocont.2019.05.004.

[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, doi: 10.1109/TII.2018.2878405.

[10] Q. Jiang, X. Yan, and B. Huang, "Deep discriminative representation learning for nonlinear process fault detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1410–1419, Jul. 2020.

[11] Q. Jiang, S. Yan, X. Yan, H. Yi, and F. Gao, "Data-driven two-dimensional deep correlated representation learning for nonlinear batch process monitoring," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2839–2848, Apr. 2020, doi: 10.1109/TII.2019.2952931.

[12] 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, doi: 10.1109/tcst.2019.2897946.

[13] C. Zhao and F. Gao, "Online fault prognosis with relative deviation analysis and vector autoregressive modeling," *Chem. Eng. Sci.*, vol. 138, pp. 531–543, Dec. 2015, doi: 10.1016/j.ces.2015.08.037.

[14] 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.

[15] M. Kano, "Process monitoring based on dissimilarity of time series data," *Kagaku Kogaku Ronbunshu*, vol. 25, no. 6, pp. 1004–1009, 1999.

[16] M. Kano, S. Hasebe, I. Hashimoto, and H. Ohno, "Statistical process monitoring based on dissimilarity of process data," *AIChE J.*, vol. 48, no. 6, pp. 1231–1240, Jun. 2002.

[17] C. Zhao, F. Wang, and M. Jia, "Dissimilarity analysis based batch process monitoring using moving windows," *AIChE J.*, vol. 53, no. 5, pp. 1267–1277, 2007.

[18] K. E. S. Pilario and Y. Cao, "Canonical variate dissimilarity analysis for process incipient fault detection," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5308–5315, Dec. 2018, doi: 10.1109/TII.2018.2810822.

[19] J. Huang and X. Yan, "Angle-based multiblock independent component analysis method with a new block dissimilarity statistic for non-Gaussian process monitoring," *Ind. Eng. Chem. Res.*, vol. 55, no. 17, pp. 4997–5005, May 2016, doi: 10.1021/acs.iecr.6b00093.

[20] S. Hara, Y. Kawahara, T. Washio, P. von Bünau, T. Tokunaga, and K. Yumoto, "Separation of stationary and non-stationary sources with a generalized eigenvalue problem," *Neural Netw.*, vol. 33, pp. 7–20, Sep. 2012.

[21] C. 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. 2020, doi: 10.1109/tcst.2019.2908339.

[22] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loève expansion to feature selection and ordering," *IEEE Trans. Comput.*, vol. C-19, no. 4, pp. 311–318, Apr. 1970, doi: 10.1109/T-C.1970.222918.

[23] P. E. Gill, W. Murray, and M. H. Wright, "Practical optimization," *Math. Gazette*, vol. 104, no. 2, p. 180, 1981.

[24] P. Schmid, "On the kolmogorov and smirnov limit theorems for discontinuous distribution functions," *Ann. Math. Statist.*, vol. 29, no. 4, pp. 1011–1027, Dec. 1958, doi: 10.1214/aoms/1177706438.

[25] J. Aldrich, "Doing least squares: Perspectives from gauss and yule," *Int. Stat. Rev.*, vol. 66, no. 1, pp. 61–81, Apr. 1998.

[26] G. J. McLachlan and D. Peel, "Finite mixture models," *Partha Deb*, vol. 39, no. 4, pp. 521–541, 2000.

[27] J. Yu and S. J. Qin, "Multiway Gaussian mixture model based multiphase batch process monitoring," *Ind. Eng. Chem. Res.*, vol. 48, no. 18, pp. 8585–8594, Sep. 2009, doi: 10.1021/ie900479g.

[28] 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, doi: 10.1109/34.990138.

[29] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Ann. Statist.*, vol. 38, no. 5, pp. 2916–2957, Oct. 2010.

[30] 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.

[31] L. H. C. Al, *Fault Detection and Diagnosis in Industrial Systems*. London, U.K.: Springer-Verlag, 2005.

[32] D. C. Montgomery, *Fault Detection and Diagnosis in Industrial Systems*. New York, NY, USA: Wiley,1997.