A Novel Spatiotemporal Process Feature Learning Method Based On the Pseudo-Siamese Network for Complex Chemical Process Concurrent Condition Monitoring

Yuemei Xu, Mingxing Jia,* and Zhizhong Mao

ABSTRACT: The deep learning-based process monitoring method has attracted great attention due to its ability to deal with nonlinear correlation. However, the further modeling of learned deep features from process data to better depict typical process features to obtain more precise monitoring results remains a challenge. In this paper, a novel nonlinear spatiotemporal process feature learning method is proposed to extract high-value slow-varying spatiotemporal process features, with an explicit temporal relationship model for the concurrent monitoring of the static deviation and the dynamic anomaly of complex chemical processes. Different from directly mixed spatiotemporal information methods, the pseudo-Siamese autoencoder network is designed with two different subencoders to separately describe the nonlinear spatial and temporal relationships of the nonlinear dynamic input data. Correspondingly, a cost function including three losses and one orthogonal constraint is proposed to make sure that the extracted spatiotemporal process features change as slowly as possible and contain the most nonlinear dynamic information on the input data. With the explicit spatial and temporal relationship submodel, predictions are utilized to shrink the variability of the nonlinear temporal correlated data and focus on the unpredictable variabilities to improve process monitoring performance. Meanwhile, the linear dynamic information is further extracted in the reconstructed residual space by the general slow feature analysis (SFA) method to provide a more detailed analysis of the process characteristics and improve the monitoring results. The case study monitoring results demonstrate the effectiveness and superiority of the proposed method over other compared methods for concurrent process monitoring.

1. INTRODUCTION

Process monitoring, which aims to detect abnormal operating conditions of a process, plays an essential role in complex chemical processes by improving product quality and ensuring production safety.1 With the rapid progress in measuring and sensing techniques, modern process condition monitoring has cuddled the dawn of a data-based method era due to the difficulty of deriving a precise physical model compared to the ease of process data collection for complicated processes.2 In the last few decades, many data-driven process condition monitoring techniques have been developed to facilitate safe operation and efficient production in industrial processes.3,4

Data-based process monitoring methods first use a model to transfer the original process condition data to a new feature space to extract the high-value features, then a statistical model is further built with the corresponding control limit to monitor whether the online new condition is consistent with the reference process condition.5,6 Most of the traditional widely

Received: August 6, 2022
Accepted: September 28, 2022
Published: October 7, 2022
used process condition monitoring methods are based on multivariable statistical analysis methods, such as principal component analysis (PCA),\textsuperscript{7} independent component analysis (ICA),\textsuperscript{8} canonical correlation analysis (CCA),\textsuperscript{9} and slow feature analysis (SFA).\textsuperscript{10} SFA is a concurrent process monitoring method. Compared to the other traditional multivariable statistical analysis methods, SFA can concurrently monitor the steady-state deviation and dynamic anomalies, providing a more detailed interpretation of the detected fault.\textsuperscript{11,12} Although those monitoring methods have been widely researched and successfully applied in many actual cases, they are essentially linear methods and are based on the tacit assumption that the applied actual processes are linear. These methods may achieve poor monitoring performances when dealing with nonlinear industrial processes. In order to obtain a superior monitoring performance in a nonlinear process, there are many improvements on the original linear multivariable statistical analysis methods. The kernel trick, which is the most widely used trick to eliminate the nonlinear correlation between variables in the original dimensional space compared to the linear one, can be introduced in combination with a common multivariable analysis method, for example KPCA,\textsuperscript{13} KICA,\textsuperscript{14} KCCA,\textsuperscript{15} or KSFA.\textsuperscript{16} The specific implementation process is to implicitly map the original nonlinear correlation data to the high-dimensional space through appropriate kernel functions, such as the Gaussian kernel function, and reduce the dimension of this space through a basic linear multivariable statistical analysis method. However, in the kernel-based method it is difficult to determine the parameters in the kernel function, as they are not optimal for a nonlinear process. Moreover, the kernel function needs to be predefined, which could not be adapted to the measured data in the industrial process. Later, a manifold learning-based nonlinear process monitoring method was developed to self-teach the kernel matrix by setting up some constraints such as maximum variance unfolding and local preservation projection. The method no longer requires an explicit kernel function but is still limited by the sample size.\textsuperscript{17} Besides, finding the mapping relationship between the latent feature space and the original data space also limits the development of manifold learning-based nonlinear process monitoring methods.

Recently, deep learning methods have attracted considerable attention due to their advantages in extracting hierarchical and deep representations.\textsuperscript{18} Furthermore, deep neural network-based modeling methods describe the nonlinearity using an activation function such as hyperbolic tangent function and learn the nonlinear relationships among data by adjusting relevant weights and bias parameters to obtain an optimal nonlinear model for the nonlinear process.\textsuperscript{19,20} Considering that the data produced by an industrial process are usually unlabeled, process condition monitoring, which is the detection of process abnormalities without the further identification of the fault source, usually uses the normal operation condition data to build the monitoring model, which makes supervised deep learning methods inappropriate for process monitoring.\textsuperscript{21} Autoencoders (AEs), a typical unsupervised deep neural network that uses the two processes of encoding and decoding to cleverly realize the feature learning of unlabeled data, has been applied for nonlinear process condition monitoring. For example, Yan, Guo, et al.\textsuperscript{22} combined denoising autoencoders (DAEs) and contractive autoencoders for the robust monitoring of nonlinear processes; Zhang, Jiang, et al.\textsuperscript{23} used a stacked denoising autoencoder and the k-nearest-neighbor rule to learn automated features for nonlinear process monitoring; Yu, Zhao, et al.\textsuperscript{24} proposed a novel robust process monitoring and fault isolation method for a nonlinear industrial process that used a denoising autoencoder and elastic net; and Yu, Zhang, et al.\textsuperscript{25} introduced sparsity and manifold regularity into a stacked autoencoder to detect faults in complex industrial processes. Although AE-based methods, which have shown promising capabilities for process condition monitoring, fully consider the nonlinear correlation among variables, they ignore the dynamically interdependence among process samples at different times. In order to provide more sensitive monitoring of complex nonlinear dynamic industrial processes, recurrent neural network (RNN)\textsuperscript{26} was developed to expand the traditional AE methods by taking into consideration both spatial relationships that are variable correlations and temporal relationships that are dynamic correlations in variables. Using convolutional LSTM neural network autoencoders, Essien and Giannetti\textsuperscript{27} built a deep learning model for smart manufacturing to capture the temporal and spatial distributions of input time-series signals. Cheng and He et al.\textsuperscript{28} proposed a new process monitoring method based on a variational recurrent autoencoder (VRAE) that took the temporal dependency between variables into account. In the RAE-based process monitoring framework, both spatial relationships and temporal relationships are considered to be nonlinear to achieve more representative process characteristics; however, the extracted latent variables pay no attention to the temporal correlation only focused on the minimization of the reconstruction error. The implicit temporal correlation information in the extracted inner latent variable will result in a high false alarm rate during process monitoring, since the statistic of the extracted latent variable does not satisfy the assumption of time-independence. Since the existing neural network-based method cannot obtain the explicit temporal correlation model of the latent variable, it is hard to utilize the prediction in temporal relationships to further shrink the data variability and focus on the unpredictable variability to improve the capability for abnormality detection. Besides, the above methods mainly extract nonlinear dynamic process features. In practice, there are not only linear or nonlinear dynamic features but also mixed dynamic features in industrial processes. Although the consideration of only nonlinear dynamic features has led to the achievement of significant monitoring performance for complex process condition monitoring, this method is still insufficient for process characteristics analysis and has the space to improve the detection rate. Furthermore, it is also impossible to further interpret whether the fault detection result belongs to a steady-state deviation or a dynamic abnormality.

In this paper, we propose a novel spatiotemporal process feature learning method to extract high-value slow features with the explicit presentation of temporal relationships based on the proposed pseudo-Siamese network architecture. Under the proposed pseudo-Siamese network architecture, nonlinear spatial information that reflects the correlation among process variables and nonlinear temporal information that reveals the dynamic character in the temporal correlation of the inner process variables are considered simultaneously to provide separate spatial and temporal models. Inspired by the SFA\textsuperscript{10} method, we introduced the minimization of various speed and orthogonal constraints of the latent features to form a
multiojective optimization to adaptively aggregate spatial and
temporal models and minimize the reconstruction error in the
original AE unsupervised learning. The extracted spatiotem-
poral process features are deemed to be nonlinear and
temporal-correlated with slow variations, and not only their
spatial models but also their temporal models are built
explicitly. Meanwhile, the linear dynamic information is further
extracted in the reconstructed residual space using the general
SFA method to provide more detailed process characteristic
analysis. For process condition monitoring, compared to the
existing monitoring methods for complex processes with
nonlinear and dynamic characteristics, the proposed process
monitoring method has several advantages. First, with the
explicit spatial and temporal relationship model, predictions
are utilized to shrink the nonlinear temporal-correlated data
variability and focus on the unpredictable variables for process
monitoring. Second, by considering all the characteristics of
the process comprehensively, including nonlinear dynamic
characteristics and linear dynamic characteristics, the estab-
lished model can more accurately represent the complex
process. Third, the proposed process monitoring method could
provide separate descriptions of the distribution of the
extracted latent variables and the distribution of the variation
speeds of extracted latent variables to further interpret whether
the fault detection result belongs to a steady-state deviation or
dynamic abnormality.

The remainder of the paper is organized as follows. Section
2 gives a brief introduction to some preliminaries. Then, the
proposed spatiotemporal process feature learning method
based on a pseudo-Siamese network for complex nonlinear
dynamic process concurrent condition monitoring is described
in detail in section 3. The effectiveness of the proposed
method is demonstrated on a simulation data set and a
conclusions are summarized in section 5. Finally, the

2. PRELIMINARIES

2.1. Autoencoder. An AE is an unsupervised neural
network structure for dimensionality reduction and feature
extraction. A conventional autoencoder consists of an encoder
and a decoder, as shown in Figure 1(a). Consider the input
data \( X \in \mathbb{R}^{J \times n} \) with the \( J \)-dimensional variable and a number
of samples \( n \). The encoding process transforms the input data
into a group of latent features as follows:

\[
H = \delta(W^T X + b)
\]

where \( H \) is the latent feature, the \( W \) is the encoding weight
matrix, \( b \) is an offset vector, and \( \delta(\cdot) \) is an element
activation function, such as a sigmoid function or a hyperbolic tangent
function. The decoding process is reconstructed the input data
using the latent feature as follows:

\[
\hat{X} = \beta(W^T H + \bar{b}) = \beta(W^T (\delta(W^T X + b)) + \bar{b})
\]

where \( \hat{X} \) is the reconstructed value of the input data, \( W \) and \( \bar{b} \)
are the decoding weight matrix and the offset vector different
from the encoder part, respectively, and \( \beta(\cdot) \) is also an
element activation function. The weight matrix sets \( W, b, \hat{W} \)
and \( \bar{b} \) of the encoder and decoder are learned simultaneously
during the task of reconstructing the original input data. The
aim of an AE is to minimize the reconstruction error, which is
defined as follows:

\[
\{W, b, \hat{W}, \bar{b}\} = \arg\min_{W,b,\hat{W},\bar{b}} \lVert X - \hat{X} \rVert_2^2
\]

Similar to the convolutional neural network (CNN), an AE can
also expand its receptive field and extract deeper information
by adding more hidden layers, as shown in Figure 1(b). The
encoding process can be described as follows:

\[
\begin{align*}
H_i &= \delta(W_i^T X + b_i) \quad i = 1 \\
H_i &= \delta(W_i^T H_{i-1} + b_i) \quad i > 1
\end{align*}
\]

where \( \{W_i, b_i\} \) is the parameter of the encoder of the \( i \)-th layer.
Similarly, the decoding process could be described as follows:

\[
\begin{align*}
\hat{X}_i &= \beta(W_i^T \hat{H}_i + \bar{b}_i) \quad i = 1 \\
\hat{H}_{i-1} &= \beta(\hat{W}_i^T \hat{H}_i + \bar{b}_i) \quad i > 1
\end{align*}
\]

where \( \hat{H}_m = \hat{H}_m \) is the top-layer representation of the deep AE,
which is the output of the encoder. The top layer
representations of the deep AE are highly abstract and contain
the most important information for the reconstruction of the
original data. Deep AE efficiently extracts representations from
a set of data. Finally, the average reconstruction error is
minimized using a backpropagation algorithm to fine-tune
the whole network. In the training process, all parameters in
the deep AE model, which are the key factors in the formation
of the nonlinear function, are learned from the original data.
This property ensures that the deep AE model is able to
automatically learn the nonlinear relationships among process
variables to get the crucial features, which avoids the problem of manually selecting inappropriate nonlinear functions.

2.2. Long Short-Term Memory Network. A RNN is an extension of the conventional feedforward neural network for modeling sequence data that preserves the temporal dimension of sequential data by connecting neurons to and from a network. The long short-term memory (LSTM) network is a variant of the traditional recurrent neural network that could eliminate the vanishing gradient problem in general recurrent neural networks. Figure 2 shows the basic structure of an LSTM memory cell. As shown in Figure 2, the basic LSTM unit contains input, output, and forget control gates to perform write, read, and reset functions in each cell.

\[
\begin{align*}
    \hat{f}_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\
    \hat{i}_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\
    \hat{c}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) \\
    c_t &= f_t \odot c_{t-1} + i_t \odot \hat{c}_t \\
    \hat{o}_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\
    h_t &= o_t \odot \tanh(c_t)
\end{align*}
\]

Where \( h_t \) is the extracted hidden state at time \( t \), \( x_t \) is the input vector at time \( t \), \( \hat{f}_t \) is the output of the forget gate, \( \hat{i}_t \) is the input, \( \hat{c}_t \) is the state of the input gate, \( c_t \) is the long-term state, and \( \hat{o}_t \) is the output of the output gate. \( W_f, W_i, W_c, \) and \( W_o \) and \( b_f, b_i, b_c, \) and \( b_o \) are the weight and biases coefficient matrices. \( \sigma \) is the activation function, and \( \tanh \) is the tanh activation function.

2.3. Siamese Neural Network Architecture. A Siamese neural network is a coupling architecture based on two neural networks that is used to discriminate the similarity of two data sets. The basic idea of a Siamese neural network is to use the same neural network to extract features of the two data and then a selected distance metric to determine if the two data sets are similar. The specific architecture is shown in Figure 3. A Siamese neural network contains two identical subnetworks. Both the structure and parameters of the subnetworks are identical to each other. Thus, each of the subnetworks provides an identical mapping of inputs to the

\[
\|d\| = \|H^1 - H^2\|
\]
latent features. $X^1$ and $X^2$ are the inputs that will be fed into the same neural network for feature extraction. The extracted feature vectors of the two input data are denoted as $H^1$ and $H^2$. Then, the inputs of each network are compared by computing the “semantic” distance of the highest-level feature representations. The output of the network is a difference metric, $d(H^1 - H^2)$, between the two latent features. The object of the Siamese neural network is as follows:

$$j(W) = \begin{cases} d(H^1 - H^2) & \text{if } X^1 \text{ and } X^2 \text{ are similar} \\ \max(0, \sigma - d(H^1 - H^2)) & \text{otherwise} \end{cases}$$

(7)

In the training process, the above loss is minimized by constructing a Siamese neural network with shared weights whose inputs are pairs of samples. The loss is minimized with respect to the trainable parameters using stochastic gradient descent via a common back-propagation strategy.

3. PROCESS CONDITION MONITORING BASED ON A NOVEL PSEUDO-SIAMESE UNSUPERVISED SLOW FEATURE EXTRACTION NETWORK

3.1. Problem Statement and Model Representation.

In a modern industrial process, closed-loop control systems have been widely and massively fitted out to compensate disturbances as a central rule, ensuring process safety and pursuing maximal profits.32 Process concurrent condition monitoring is crucially important for industrial processes with feedback control systems to distinguish static deviation and dynamic abnormalities and detect real process faults.11,12 However, considering nonlinear characteristics in both spatial and temporal relationships is still a challenging task for process concurrent condition monitoring. Nonlinear characteristics, including the nonlinear correlation between variables and the nonlinear unvariable autocorrelation, are ubiquitous in industrial processes. The original AE-based unsupervised method only focuses on representing the nonlinear correlation between variables (spatial relationships) to reconstruct the input data and extract the high-value latent features. Although AE-based methods fully consider the nonlinear spatial relationship, they ignore the dynamically interdependent relationships among process samples at different times. The original AE is almost unable to directly represent temporal relationships, while the RAE is devoted to representing them. RAE-based unsupervised process modeling could extract the representative latent feature to consider both spatial and temporal relationships. The latent features can be described as follows:

$$h = f_1(X^1, f_2(X^{l-1}), f_3(X^{l-2}, \ldots, f_L(X^{l-1})))) \in R^{1 \times A}$$

(8)

where $h$ represents the number of $A$ latent features, $f_i$ ($i = 1, 2, \ldots, L$) represents the nonlinear function, and $X^{l-1}$ in $R^{1 \times X}$ ($i = 1, 2, \ldots, L$) is the input data with $L$-order time dependence. It is clear that both correlations and autocorrelations are considered to be nonlinear in the RAE-based method, as shown in eq 8. However, the models cannot separate the two kinds of correlations explicitly. The temporal information has two notable effects on process monitoring. One effect is that the temporal characteristic represents the process dynamics; since the process dynamics are reflected in data mainly in temporal correlations and variation speed, separately monitoring the temporal behavior makes the process dynamic abnormal monitoring be real. This is important for monitoring industrial processes with a closed-loop control systems, which are widely used to compensate for disturbances and pursue maximal profits. Another effect is that the temporal-correlated data could further build the autoregressive prediction model. For process monitoring, autoregressive prediction is utilized to shrink the data variability and focus on the unpredictable variability, which improves the ability to detect faults. In order to take the above two effects into consideration, it is necessary to extract the latent feature with an explicit temporal correlation. Considering the presence of both nonlinear correlation and nonlinear autocorrelation in a complex industrial process, our goal is to extract the high-value latent features with the explicit nonlinear temporal relationship model. The latent features model is described as follows:

$$\begin{cases} X^1 = f^{-1}(S^1) \in R^{1 \times X} \\ S^1 = g_1(S^{l-1}, g_2(S^{l-2}, \ldots, g_q(S^{l-q}))) + e_1 \in R^{1 \times A} \end{cases}$$

(9)

where $S^{l-i} \in R^{1 \times A}$ ($i = 1, 2, \ldots, q$) is the value of the extracted latent feature at time $t - i$, $X^l \in R^{1 \times X}$ is the input measurement vector at time $t$, $f(\bullet)$ represents the spatial nonlinear mapping function, $g_q(\bullet)$ ($i = 1, 2, \ldots, q$) is the nonlinear function of the nonlinear temporal mapping function with $q$-order time dependence, and $e_1$ is the prediction error, which represents the stationary stochastic process that drives the process dynamics. In our proposed latent feature model, the first part $X^1 = f^{-1}(S^1)$ is the spatial model, which is a nonlinear mapping function used to extract the latent features from the original observed data; the extracted latent features are temporally correlated. The second part $S^1 = g_1(S^{l-1}, g_2(S^{l-2}, \ldots, g_q(S^{l-q}))) + e_1$ is a temporal model to give an explicit nonlinear autocorrelation representation of the extracted slow features, which represent the dynamics of the process. As temporal model building is dependent on the extracted latent feature, it is difficult to obtain the spatial and temporal model simultaneously. To bring the latent feature spatial extraction model into the temporal prediction model, we reorganized the latent feature model as follows:

$$\begin{cases} \hat{S}' = f(X') \\ \hat{S}' = g_1(f(X'^{-1}), g_2(f(X'^{-2}), \ldots, g_q(f(X'^{-q})))) \end{cases}$$

(10)

where $\hat{S}'$ is the prediction value of $S'$. Since both $g_q(\bullet)$ and $f(\bullet)$ are nonlinear functions, eq 10 can be further described as follows:

$$\begin{cases} \hat{S}' = f(X') \\ \hat{S}' = \bar{g}_1(X'^{-1}, \bar{g}_2(X'^{-2}, \ldots, \bar{g}_q(X'^{-q}))) \end{cases}$$

(11)

where $\bar{g}_q(\bullet)$ ($i = 1, 2, \ldots, q$) is also a nonlinear mapping function that combines the original $g_q(\bullet)$ and $f(\bullet)$. Based on the model in eq 11, both temporal and spatial submodels built are dependent on the input data. It is possible to extract the high-value latent features with the explicit nonlinear temporal relationship model to obtain the process feature and its prediction. The process feature could be directly mapped form the current state by the spatial submodel, which also could be indirectly predicted from the previous observation state by the...
temporal submodel. In order to further use the prediction of the extracted latent features to shrink the variability, focus on the unvariability, and improve the monitoring performance for nonlinear dynamic process data, the autocorrelation relationship of extracted latent variable should be modeled explicitly. That is, the spatial submodel and the temporal submodel should be satisfied simultaneously to obtain the spatiotemporal process feature and its prediction. Both submodels should be accurate to extract the spatiotemporal process features, which include most process information and have explicit nonlinear autocorrelation. Traditional neural network architecture could not obtain the separate nonlinear spatial and temporal models simultaneously. We need to design a new network architecture to accomplish extraction of the high-value latent features with the explicit nonlinear temporal relationship model.

For complex process condition monitoring, both the nonlinear dynamics and the linear dynamics should be taken into consideration. The above model, as shown in eq 11, is used to extract the nonlinear dynamics relationships of the whole process; however, it is focus on the nonlinear dynamics characteristics that inevitably leave some linear dynamics information in the residual information space. Therefore, we would further analyze the linear dynamic characteristics in the reconstruction residual space. The residual is obtained as follows:

$$E = X - f^{-1}(S') \in \mathbb{R}^{n \times f}$$

(12)

The general SFA method could be used to further extract the linear dynamics information in the reconstruction residual space and divide it into the main linear dynamics space and residual space as follows:

$$S_{ed} = W_d^T \in \mathbb{R}^{1 \times p}$$

$$S_{ec} = W_e^T \in \mathbb{R}^{1 \times (f - p)}$$

(13)

where $W_d \in \mathbb{R}^{1 \times p}$ and $W_e \in \mathbb{R}^{1 \times (f - p)}$ are the mapping matrices of the main linear dynamics space and the residual space, respectively.

3.2. A Novel Pseudo-Siamese Unsupervised Slow Feature Extraction Network. In order to extract the spatiotemporal process features as represented in section 3.1 and provide a separate spatial and temporal model for process monitoring, we designed a new network architecture named the pseudo-Siamese unsupervised slow feature extraction network, the structure of which is shown in Figure 4. Autoencoders mainly operate in a vector space, thereby increasing the difficulty of learning the high-dimensional features of an input time series. However, LSTM networks are master sequential learning by passing signal information across time steps. Based on such considerations, we used two different networks to organize a new pseudo-Siamese slow feature extraction network to extract the slow-variation spatiotemporal process features: an autoencoder, which was used to learn the vector space to represent the spatial relationships, and a LSTM network, which was used to learn the temporal relationships. The new definition of the deviation of the extracted slow features with respect to time proposed in our previously research was used to judge the slow degree of the highest-level feature representations to consider the dynamic information, which includes both the variation speed and the temporal correlation. The new derivative of extracted slow features is as follows:

$$S = (S' - S'^{-1}) + (S'^{-1} - \hat{S}'^{-1}) = S' - S'^{-1} + S'^{-1} - \hat{S}'^{-1} = S' - \hat{S}'$$

(14)

where $S'$ represents the value of the extracted slow feature at time $t$, $S'^{-1}$ represents the prediction value of $S'^{-1}$ based on the temporal autoregressive model, $S' - \hat{S}'^{-1}$ is the variation of the extracted slow features, and $S'^{-1} - \hat{S}'^{-1}$ represents the variation of the autoregressive prediction error. In order to extract the high-value slow features that carry the dynamics information, including both the variation speed and the temporal correlation as defined in eq 14, the slow feature $S'$ and its prediction value $\hat{S}'$ need to be obtained simultaneously. Inspired by the pseudo-Siamese neural network, we
proposed a new pseudo-Siamese unsupervised slow feature extraction network.

As shown in Figure 4, the proposed pseudo-Siamese unsupervised slow feature extraction network was organized using two inputs, two different “autoencoder” networks, and three outputs. For the original collected data \( X \in \mathbb{R}^{n \times m} \), the matrix described in eq 15 was organized from original input matrix \( X \in \mathbb{R}^{n \times m} \):

\[
X^k = [X_k, X_{k+1}, X_{k+2}, \ldots, X_{k+q}] \in \mathbb{R}^{(n-q) \times m}
\]

\( k = 1, 2, \ldots, q \)  

Here \( q \) is the time lag and \( X_k \in \mathbb{R}^m \) indicates the input vector corresponding to the \( k \)th sample in original collection data matrix \( X \). The first input of the proposed pseudo-Siamese unsupervised slow feature extraction network is \( X_{m1} = X^1 \in \mathbb{R}^{(n-q) \times m} \). The second input of the proposed pseudo-Siamese unsupervised slow feature extraction network is as follows:

\[
X_{m2} = [X^{q+1}, X^{q+2}, \ldots, X^1] \in \mathbb{R}^{(n-q) \times m(q-2)}
\]

As shown in Figure 3, these two “autoencoder” networks are organized by an encoder and a decoder in each network. The encoders have different structures and parameters, while the decoders are identical to each other in terms of both the structure and the parameters. The first network is the spatial process feature extraction network, which is organized by an encoder and a decoder. The spatial encoding process transforms the input data to a group of high-value latent feature as follows:

\[
S' = f_{le}(X_{m1})
\]

\[
\delta(W_1^T \delta(W_2^T \ldots \delta(W_h^{T}X^q + b_h) + b_2) + b_1)
\]

where \( S' \) indicates the extracted latent feature; \( X_{m1} = X^1 \) is the input data matrix; \( h_1 \) is the number of the hidden layers in the spatial encoder; \( f_{le}(\{\}) \) represents the nonlinear spatial encoding process-mapping function, which could be further obtained by a full-connection neural network with multiple hidden layers, as shown in eq 17; \( \delta(\{\}) \) is an element activation function; and \( \{W_i, b_i\} (i = 1, 2, \ldots, h_1) \) is the parameter for the spatial encoder of the 1th hidden layer. The decoding process is used to reconstruct input data \( X^1 \) from the latent slow feature. The output 1 of the decoding process is described as follows:

\[
\hat{X} = f_{de}(S')
\]

\[
\beta(W_{h_{1}}^{T} \beta(W_{h_{2}}^{T} \ldots \beta(W_{h_{d}}^{T}S' + b_{h_{d-1}}) + b_{h_{d-1}}) + b_{h_{d}})
\]

where \( \hat{X} \) represents the reconstruction data of the input 1 data matrix \( X^1 \), \( f_{de}(\{\}) \) represents the nonlinear decoding process-mapping function, \( h_d \) is the number of hidden layers in the decoder, \( \beta(\{\}) \) is also an element activation function, and \( \{W_i, b_i\} (i = 1, 2, \ldots, h_d) \) is the parameter of the decoder of the 1th hidden layer.

The second network is the temporal process feature extraction network, which is different from the first network. The LSTM neural network is applied in the encoding process instead a traditional neural network because the LSTM neural network can handle the temporal correlation. The temporal encoding process transforms the input data to a group of high-value latent feature as follows:

\[
\hat{S}^{-1} = f_{le}(X_{m2})
\]

\[
= f_{LSTM}(X^{q+1}, f_{LSTM_2}(X^{q+1}), \ldots, f_{LSTM_1}(X^1))
\]

where \( \hat{S}^{-1} \) is the prediction value of the extracted latent slow feature at time \( t - 1 \); \( f_{le}(\{\}) \) represents the nonlinear temporal encoding process-mapping function, which could be obtained by connecting a series LSTM neural units; \( l = q - 2 \) indicates the number of LSTM units; and \( f_{LSTM}(\{\}) \) is the function of the LSTM unit, whose specific expanded formation could be obtained in eq 6. The decoding process does not reconstruct input data; instead, it reconstructs the data \( X^{q+1} \), since the output of the temporal encoding process is the prediction for the value of the slow features extracted by the spatial encoder at the time \( t - 1 \). To make sure that \( \hat{S}^{-1} \) estimates \( S^{-1} \) as accurately as possible, the decoder of the second autoencoder has a structure and parameters identical to the decoder of the first autoencoder. The output 2 of the decoder of the second autoencoder is as follows:

\[
\hat{X}^{q+1} = f_{de}(\hat{S}^{-1})
\]

\[
\beta(W_{h_1}^{T} \beta(W_{h_2}^{T} \ldots \beta(W_{h_d}^{T}\hat{S}^{-1} + b_{h_d}) + b_{h_{d-1}}) + b_{h_{d-1}})
\]

where \( \hat{X}^{q+1} \) represents the reconstruction data of the data matrix \( X^{q+1} \), Output 3 is the key to constraining the extracted latent features, which are slow-varying and have dynamic inner nonlinear correlation, and connecting the above two autoencoders to accomplish spatiotemporal process feature extraction with separate spatial and temporal models. Output 3 is the norm of the derivative defined in eq 14.

\[
d = \|S' - \hat{S}^{-1}\|
\]

In order to extract the high-value slow feature variables in an unsupervised way in the proposed pseudo-Siamese unsupervised slow feature extraction network, the following three objectives need to be satisfied simultaneously:

\[
J_1(\theta) = \arg\min_\theta \|S' - \hat{S}^{-1}\|_2
\]

\[
J_2(\theta) = \arg\min_\theta \|X^0 - \hat{X}^0\|_2
\]

\[
J_3(\theta) = \arg\min_\theta \|X^{q+1} - \hat{X}^{q+1}\|_2
\]

s. t. \( SS^T = I \)

where the first objective \( J_1 \) minimizes the output 3, which is the variation of the new deviation with respect to time; the second objective \( J_2 \) minimizes the reconstruction error between \( X^0 \) of output 1 and \( X^0 \) of input 1; the third objective \( J_3 \) minimizes the reconstruction error between \( X^{q+1} \) and \( X^0 \) of output 1. It is easy to understand that the first objective function ensures that the extracted potential features retain as much information on the original process as possible and the second objective function is constrains extracted potential features and causes them to change as slowly as possible. Here we want to emphasize the
third objective function. As shown in Figure 4, the temporal encoder outputs an estimate of the value of the latent slow features at the previous time. If this estimate is as accurate as possible, it can ensure that the new derivative we define is meaningful rather than simply a representation of the estimation errors of $S_{t-1}$ and $\hat{S}_{t-1}$. However, it is difficult to directly define the objective to make the $\hat{S}_{t-1}$ as close as possible to $S_{t-1}$, since the previous latent slow spatiotemporal process feature $S_{t-1}$ cannot be obtained in advance. We consider that since the reconstruction error of $S'$ is decoded to $\hat{X}^d$ and the reconstruction error $J_2$ is as small as possible, if the value of $\hat{S}_{t-1}$ is as close as possible to $S_{t-1}$, the reconstruction error $J_3$ should also be as small as possible under the same decoder. Therefore, we designed a decoder with the same structure and the same parameters and introduced the third objective to make sure that the value of $\hat{S}_{t-1}$ was as close as possible to $S_{t-1}$. Further, in order to avoid a trivial solution of the above optimization problem where the obtained latent features would be constant, the regularization layer was embedded behind the encoder to constrain the extracted SFs that were decorrelated and had unit variance. Finally, the global loss function was formed by weighted sum of the subobjective functions $J_1(\theta)$, $J_2(\theta)$, and $J_3(\theta)$ and the constraint. The specific formation is as follows:

![Diagram](https://doi.org/10.1021/acsomega.2c05028)

**Figure 5.** Offline modeling scheme of the proposed concurrent process condition monitoring method.
\[
L = \arg \min_{\theta} [J_1(\theta) + \alpha J_2(\theta) + \alpha J_3(\theta) + \alpha \|SS^T - I\|_F^2]
\]  
(23)

In order to determine parameter \( \theta \) of the proposed pseudo-Siamese slow feature extraction network, the commonly used Adam optimizer method was adopted to minimize eq 23.

### 3.3. Parameter Determination Discussion

Some parameters need to be determined in the proposed pseudo-Siamese unsupervised slow feature extraction network training process, namely, the weight coefficients \( \alpha_1, \alpha_2 \), and \( \alpha_3 \); the time lag \( q_t \), the number of hidden layers and units in each layer, and the learning rate.

In this paper, the objective of the proposed algorithm is to extract the slow features in an unsupervised way. It is important to emphasize that the extracted slow features should be as slow as possible to contain the main information; one should not blindly pursue the slow change speed. Therefore, we should determine the structure parameter of the deep AE to make sure the deep AE can preserve the principal information on the process at first. However, the general question of how to determine the structure of a neural network for a given process data set is still an open problem. In order to make sure that the extracted latent feature preserves the principal information about the input data, we determined the structure via a trail to make the reconstruction precious of the deep AE achieve 90%. We also used the same method to determine the structure of the LSTM network to make the prediction error smaller than 10%. Once the structures of the deep AE and LSTM networks are obtained, we will further adjust the weight coefficients \( \alpha_1, \alpha_2, \text{ and } \alpha_3 \). Since \( \alpha_1 \) and \( \alpha_2 \) both indicate the weight of the reconstruction error, we let \( \alpha_1 = \alpha_2 \) in this paper to simplify the parameter determination process. It is notable that if \( \alpha_1 = \alpha_2 = 0 \), the optimization problem described in eq 18 can not ensure that the extracted slow features preserve the most process information; conversely, if \( \alpha_1 = 0 \), the extracted slow features are not uncorrelated. The orthogonal regularization will influence how much process information is preserved in the extracted latent slow feature; that is, the reconstruction parameter will decrease. In this paper, we determine the weight coefficients \( \alpha_1 \) and \( \alpha_2 \) by trial according to the principle that the reconstruction parameter does not decrease further than 80%. That is adjust the \( \alpha_2 \) to make the reconstruction parameter is higher than 80% at first, then \( \alpha_2 \) should be adjusted such that the extracted slow feature is as uncorrelated as possible but the reconstruction parameter does not decrease below 80%.

### 3.4. Concurrent Process Monitoring Based On The Proposed Method

In this section, we propose a new process concurrent monitoring scheme for monitoring the static deviation and the dynamic abnormal simultaneously based on the proposed pseudo-Siamese unsupervised slow feature extraction network model; the offline modeling process is described in detail in Figure 5. In the proposed pseudo-Siamese unsupervised slow feature extraction network model, the extracted latent slow features are nonlinear and temporal-correlated and have the following explicit spatial and temporal relation model:

\[
\begin{align*}
\mathbf{S'} &= f_{se}(\mathbf{X}) \\
\mathbf{S} &= f_{se}(\mathbf{X}^{q-1}, \mathbf{X}^{q-2}, ..., \mathbf{X}^2)
\end{align*}
\]  
(24)

For complex industrial processes, the extracted slow feature based on the proposed pseudo-Siamese unsupervised slow feature extraction network preserves the most process information rather than all the industrial process information. The remaining process information and noise disturbance are retained in the reconstruction error.

\[
\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X} - f_{se}(f_{se}(\mathbf{X}))
\]  
(25)

As shown in eq 25, the reconstruction error still contains some process information, and the variables also correlated. Therefore, in order to more precisely monitor the complex industrial process, the reconstruction error needs to be further analyzed to divide it into main features and a residual, where main features represent the process information preserved in the reconstruction error and the residual represents the noise. In this paper, we use the SFA method to analyze reconstruction error and decompose the extracted SFs into two groups according to their varying speed as follows:

\[
\begin{align*}
\mathbf{S}_{ed} &= \mathbf{E}\mathbf{W}_d^T \\
\mathbf{S}_{ec} &= \mathbf{E}\mathbf{W}_e^T
\end{align*}
\]  
(26)

where \( \mathbf{S}_{ed} \) represents the slow-varying features that indicate the process information, \( \mathbf{S}_{ec} \) represents the fast-varying features that indicate the noise, and \( \mathbf{W}_d \in \mathbb{R}^{l \times M} \) and \( \mathbf{W}_e \in \mathbb{R}^{(M-\ell) \times M} \) are coefficient matrices. The value of \( l \) can be determined based on the criterion that the \( \mathbf{S}_{ec} \) SFs will be faster than all input variables. More detailed information on the criterion can be found in.\(^{12}\)

In order to monitor the process condition, the following three statistics were designed to measure the static variation of the whole process:

\[
\begin{align*}
\mathbf{T}_d^2 &= \frac{\mathbf{S}_d^T}{\phi_1} + \frac{(\mathbf{S}^T - \mathbf{S}_d^T)(\mathbf{S}^T - \mathbf{S}_d^T)^T}{\phi_2} \\
\mathbf{T}_{ed}^2 &= \mathbf{S}_{ed}\mathbf{S}_{ed}^T \\
\mathbf{T}_{ee}^2 &= \mathbf{S}_{ec}\mathbf{S}_{ec}^T
\end{align*}
\]  
(27)

These three statistics describe the process condition information. If one of the statistics exceeds the limit, the state of the process is inconsistent with the state corresponding to the training data. This state anomaly needs to be further distinguished between a dynamic anomaly and static deviation. Therefore, three statistics are designed to monitor the dynamic information on the process as follows.

\[
\begin{align*}
\mathbf{S}_{ed}^2 &= (\mathbf{S}^T - \mathbf{S}_d^{T-1})(\mathbf{S}^T - \mathbf{S}_d^{T-1})^T \\
\mathbf{S}_{ed}^2 &= \mathbf{S}_{ed}\Omega_d^{-1}\mathbf{S}_{ed}^T \\
\mathbf{S}_{ec}^2 &= \mathbf{S}_{ec}\Omega_e^{-1}\mathbf{S}_{ec}^T
\end{align*}
\]  
(28)

where \( \Omega_d = \text{cov}(\mathbf{S}_{ed}) \in \mathbb{R}^{l \times l} \) and \( \Omega_e = \text{cov}(\mathbf{S}_{ec}) \in \mathbb{R}^{(M-\ell) \times (M-\ell)} \) are the diagonal matrices of the eigenvalues. The correspond-}

For the online application of the proposed monitoring method, the procedure is as follows: Step 1, obtain a new online process data sample \( \mathbf{X}_{new} \) and form the input data pair \( \{\mathbf{X}_{new}, \mathbf{X}_{new}^{q-1}, ..., \mathbf{X}_{new}^{2}\} \) by combining the previous \( q - 1 \) sample. Step 2, extract the spatiotemporal process feature using the
proposed pseudo-Siamese unsupervised slow feature analysis network as follows:

\[
\begin{aligned}
S_{\text{new}}^t &= f_s(X_{\text{new}}^q) \\
\hat{S}_{\text{new}}^t &= f_s(X_{\text{new}}^q, X_{\text{new}}^{q-1}, \ldots, X_{\text{new}}^2)
\end{aligned}
\]  

(29)

Step 3, divide the reconstruction error into the main slow features and the residual features using the built mapping coefficient matrix \( W_d \in \mathbb{R}^{nd \times nd} \) and \( W_e \in \mathbb{R}^{(M-2) \times M} \) as follows:

\[
\begin{aligned}
S_{\text{new,ed}} &= (X_{\text{new}}^q - f_d(f_s(X_{\text{new}}^q)))W_d^T \\
S_{\text{new,ee}} &= (X_{\text{new}}^q - f_d(f_s(X_{\text{new}}^q)))W_e^T
\end{aligned}
\]  

(30)

Step 4, calculate the static and dynamic statistics using eqs 27 and 28 as follows:

\[
\begin{aligned}
T_{d_{\text{new}}} &= \frac{S_{\text{new}}^t(S_{\text{new}}^t)^T}{\phi_i} + \frac{(S_{\text{new}}^t - \hat{S}_{\text{new}}^t)(S_{\text{new}}^t - \hat{S}_{\text{new}}^t)^T}{\phi_2} \\
T_{ed_{\text{new}}} &= S_{\text{new,ed}}S_{\text{new,ed}}^T \\
T_{ee_{\text{new}}} &= S_{\text{new,ee}}S_{\text{new,ee}}^T
\end{aligned}
\]  

(31)

\[
\begin{aligned}
S_{d_{\text{new}}} &= (S_{\text{new}}^t - S_{\text{new}}^{t-1})(S_{\text{new}}^t - S_{\text{new}}^{t-1})^T \\
S_{ed_{\text{new}}} &= S_{\text{new,ed}}\Omega_{\text{ed}}^{-1}s_{\text{new,ed}}^T \\
S_{ee_{\text{new}}} &= \hat{S}_{\text{new,ee}}\Omega_{e}^{-1}\hat{s}_{\text{new,ee}}^T
\end{aligned}
\]  

(32)

Step 5, compare the above six statistics to their control limits \( L_{Td}, L_{Ted}, L_{Tee}, L_{Sd}, L_{Sed}, \) and \( L_{Se} \) and judge whether there is a condition deviation; if yes, further judge if there is a dynamic anomaly. If one of six statistics exceeds its control limit, the process condition is different from the training data and a process condition abnormality occurs. If only the \( Td^2 \), \( Ted^2 \), and \( Tee^2 \) statistics are outside their control limits and the \( Sd^2 \), \( Sed^2 \), and \( See^2 \) statistics under their control limits, just the static deviation occurs. If one of the \( Sd^2 \), \( Sed^2 \), \( See^2 \) statistics goes beyond its control limit, the process dynamic abnormality occurs, which should attract attention and a quick response.

### Table 1. Description of Process variables for the WWT Process

| symbol | description | unit |
|--------|-------------|------|
| \( S_1 \) | soluble nondegradable organic matter | g·COD/m³ |
| \( S_2 \) | soluble rapidly biodegradable organic matter | g·COD/m³ |
| \( X_1 \) | insoluble particulate non biodegradable organic matter | g·COD/m³ |
| \( X_2 \) | insoluble slowly biodegradable organic matter | g·COD/m³ |
| \( X_{\text{BA}} \) | biomass of active heterotrophic bacteria | g·COD/m³ |
| \( X_{\text{RA}} \) | active autotrophic biomass | g·COD/m³ |
| \( X_P \) | inert substances produced by the decay of biological solids | g·COD/m³ |
| \( S_2 \) | dissolved oxygen | g·COD/m³ |
| \( S_{NO} \) | nitrate nitrogen | g·N/m³ |
| \( S_{NH} \) | ammonia nitrogen | g·N/m³ |
| \( S_{ND} \) | soluble biodegradable organic nitrogen | g·N/m³ |
| \( X_{ND} \) | granular biodegradable organic nitrogen | g·N/m³ |
| \( S_{AL} \) | alkalinity | mol/L |
| \( Q_s \) | input circulate | m³/d |
| \( Q_e \) | inter circulate | m³/d |

### Table 2. Fault Description for the WWT Process

| no. | description | type | dynamic property |
|-----|-------------|------|------------------|
| 1   | increased flow caused by rain | pulse | short-lived dynamics anomalies |
| 2   | change in the maximum specific growth rate of autotrophic bacteria | step | without dynamics anomalies |
| 3   | blocked internal circulation pipeline | slope | dynamics anomalies when serious blocked |

### 4. CASE STUDY

In this section, we introduce the wastewater treatment process and the widely applied TE process to demonstrate the effectiveness of the proposed pseudo-Siamese unsupervised slow feature extraction method. State-of-art methods such as kernel dynamic principal component analysis (KDPCA), kernel dynamic slow feature analyses (KDSFA), an original autoencoder with single hidden layer (AE), and a recurrent autoencoder (RAE) with a LSTM unit are simulated and compared with the proposed pseudo-Siamese unsupervised slow feature extraction network-based process monitoring method.

#### 4.1. Wastewater Treatment Process

The wastewater treatment process has made outstanding contributions to strengthening wastewater treatment, promoting the use of renewable water, and saving and protecting water resources.
The wastewater treatment process is a complex industrial process that includes typical nonlinear and temporal characteristics. As shown in Figure 6, the wastewater treatment process is mainly made up of a biochemical reaction tank, a secondary sedimentation tank, and the circulatory system. The biochemical reaction tank includes five tanks, the first two of which are anoxic tanks and the last three of which are aerated tanks. The denitrification process in the wastewater treatment process is carried out in the anoxic tank. By adjusting the internal return flow $Q_a$, the nitrate nitrogen concentration in...
the second unit is kept at a constant value of 1 mg/L. The nitrification process is carried out in the aerobic tank, and the dissolved oxygen concentration is kept at a constant value of 2 mg/L by controlling the oxygen transfer coefficient $K_l$ of the fifth tank. The secondary settler tank has ten layers, and each layer is 0.4 m thick. The circulation system includes internal circulation, external circulation, and sludge discharge. Internal circulation refers to the return of the mixed liquid in the aerobic tank to the anoxic tank, and external circulation refers to the return of the sludge in the secondary sedimentation tank to the anoxic tank. A more detail description can be obtained in refs 35 and 36.

**Figure 8.** Monitoring results of the wastewater treatment process, case 2: (a) PCA, (b) CCA, (c) SFA, (d) KDPCA, (e) KDSFA, (f) AE, (g) RAE, and (h) our proposed method.
In this section, we choose 15 process variables to monitor the condition of the WWTP. The process variables are described in Table 1. The 1344 sample input data points taken with a sampling interval of 15 min on a sunny day were used as the training data to model the process monitoring model. In order to demonstrate the effectiveness of the proposed method, three abnormal cases as described in Table 2 were used for online monitoring.

In the offline modeling stage, the dynamic order was selected as 3 and the significant level for control limits was set as 0.01 uniformly. The Gaussian kernel function was selected as the kernel function with the kernel parameter $c = 96000$, and the number of principal components was determined according to the 95% cumulative contribution rate in the KDPCA and KDSFA methods. The structure of the AE was set to $15-40-30-20-30-40-15$. The structure of the RAE was

Figure 9. Monitoring results of the wastewater treatment process, case 3: (a) PCA, (b) CCA, (c) SFA, (d) KDPCA, (e) KDSFA, (f) AE, (g) RAE, and (h) our proposed method.
steady state, and gradually decrease to the normal state after the rain stops. The monitoring results of each method are gradually increased when the rain starts, reach the new −9.

In the online monitoring stage, 744 samples were collected. The encoder structure in the AE. In the training process of the decoder structure is set to 30

| no. | description type | dynamic property |
|-----|-----------------|------------------|
| 1   | A/C feed ratio, B composition constant (stream 4) | step | short-lived dynamics anomaly |
| 2   | B composition, A/C feed ratio (stream 4) | step | short-lived dynamics anomaly |
| 3   | D feed temperature (stream 2) | step | without dynamics anomaly |
| 4   | reactor cooling water inlet temperature | step | short-lived dynamics anomaly |
| 5   | condenser cooling water inlet temperature | step | short-lived dynamics anomaly |
| 6   | A feed loss (stream 1) | step | dynamics anomaly |
| 7   | C header pressure loss-reduced availability (stream 4) | step | short-lived dynamics anomaly |
| 8   | A, B, and C feed composition (stream 4) | random | dynamics anomaly |
| 9   | D feed temperature (stream 2) | random | without dynamics anomaly |
| 10  | C feed temperature (stream 4) | random | dynamics anomaly |
| 11  | reactor cooling water inlet temperature | random | dynamics anomaly |
| 12  | condenser cooling water inlet temperature | random | dynamics anomaly |
| 13  | reaction kinetics | slow drift | dynamics anomaly |
| 14  | reactor cooling water valve | sticking | without dynamics anomaly |
| 15  | condenser cooling water valve | random | dynamics anomaly |
| 16−20 | unknown | unknown | dynamics anomaly |
| 21  | unknown | unknown | without dynamics anomaly |

Table 3. Fault Description for the TE Process

to (3 × 15)−(3 × 20)−(3 × 20)−20−(3 × 15). In the proposed method, the encoders in the AE and the RAE are used in the spatial and temporal encoders, respectively. The decoder structure is set to 30−40−15, the same as the structure of decoder in the AE. In the training process of the AE, the RAE, and the proposed method, the tanh function was chosen as the activation function, the learning rate was set at 0.001, the epoch was set at 300, and the batch size was set as 30. In the online monitoring stage, 744 samples were collected for each case. The monitoring results are shown in Figures 7−9.

Case 1 simulates the operation condition of the wastewater treatment process on rainy days. On rainy days, the inlet flow will gradually increase when the rain starts, reach the new steady state, and gradually decrease to the normal state after the rain stops. The monitoring results of each method are shown in Figure 7. We use the green dotted line to mark the beginning and end of the change of inlet flow caused by rainy days. The blue dots are used to indicate the right monitoring results, that is, the statistics that are below the control limit in the normal state and the statistics that exceed the control limit in the abnormal state. The red and the pink dots are used to indicate the wrong monitoring results. The red dot indicates the abnormal samples that have not been detected, and the pink dot indicates a sample whose normal state was wrongly identified as abnormal. The SPE statistics of PCA, Te statistic of CCA; Td and Te statistic of SFA; SPE statistics of KDPCA; Te statistic of KDSFA; SPE statistics of AE and RAE; and Td, Ted, and Te statistics of our proposed method could detect this abnormality. The SPE statistics of PCA, the Te statistic of CCA, the Td and Te statistics of SFA, the SPE statistics of KDPCA, the Te statistic of KDSFA, and the SPE statistic of AE and RAE detected this abnormality in the middle stage, which is the most obviously changing stage of the inlet flow; however, these statistics could not detect this abnormality in the beginning and end stages. Although the Td, Ted, and Te statistics of our proposed method also could not detect this abnormality, when this abnormality gradually decreased to the normal state as the rain stoped, the detection rate of our method was significantly higher than the SPE statistics of KDPCA, the Te statistic of KDSFA, and the SPE statistics of AE and RAE. In addition, KDSFA and our proposed method could further distinguish the steady-state deviation and the dynamic abnormality. From the monitoring results shown in Figure 7 (e and h), the Sd and Se statistics are both below their control limits, indicating this deviation does not introduce abnormal dynamics. The Sd, Sed, and See statistics of the proposed method have a short process of exceeding the control limit at the beginning on rainy days, and then quickly returns to below the control limit, which indicates that the process reaches a new steady state under the controller adjustment and also does not introduce abnormal dynamics.

In case 2, the maximum specific growth rate of autotrophic bacteria was changed from the original value of 0.5 to 0.7. The change in the maximum specific growth rate of autotrophic bacteria will influence the concentration of the dissolved oxygen, nitrate nitrogen, and ammonia nitrogen in the tank. However, the first controller will dynamically adjust the internal return flow Q to keep the concentration of nitrate nitrogen at 1 mg/L, and the second controller will dynamically adjust the oxygen transfer coefficient to keep the concentration of dissolved oxygen at 2 mg/L. Although the maximum specific growth rate of autotrophic bacteria changed, the WWTP was still a steady operation under control. The monitoring results of this case are shown in Figure 8. Since the incipient abnormality weakly influences the process, the T2 and SPE statistics of the PCA and KDPCA methods (as shown in Figure 8 (a and d, respectively)) and the AE method (as shown in Figure 8 (f)) are below their control limits when the abnormality is introduced, which shows that the PCA, KDPCA, and AE methods could not detect this abnormal. The Te2 statistics of CCA, SFA, and KDSFA (as shown in Figure 8 (b, c, and e, respectively)) and SPE statistic of RAE (as shown in Figure 8 (g)) partial exceed their control limits, which indicates that they can detect this abnormality but not completely, as there are still some missed detection points.

From the monitoring result of our proposed method (Figure 8 (h)), the Td2 statistic could completely detect this abnormality, the detection rate of this abnormality is significantly higher than those of the other compared methods. On the other hand, both Sd2 and Se2 statistics of KDSFA and Sd2, Sed2, and See2 statistics of the proposed method are below their control limits, indicating that the abnormality in case 2 is the steady deviation.

In case 3, a blocked internal circulation pipeline was simulated by introducing a ramp change of the internal circulation flow. The internal circulation flow as the control variable is used to keep the concentration of nitrate nitrogen at 1 mg/L. When the internal circulation flow is changed due to the pipeline blockage, the controller will adjust the output of the controller to ensure the output of the tank follows the setting points. When the blockage degree is small, even if the internal return flow at the outlet is slightly reduced, the system can still operate normally under the action of the controller. When the blockage is more serious, the internal loop flow is seriously reduced; even if the controller gives a large output, it is still difficult to keep the output near the set value because the
Table 4. Fault Detection Rate and Dynamic Properties for the TE Process

| no. | KDPCA | CCA | SFA |
|-----|-------|-----|-----|
|     | T²    | 100 | 100 | 100 | D₂  |
| 1   | 100   | 99.6| 98.7| 3.5 | 20.8|
| 2   | 96.2  | 98.5| 14.5| 9.7 | D₂  |
| 3   | 0.2   | 2.7 | 1.6 | 0.3 | 1.8 | D₁  |
| 4   | 44.2  | 98.2| 18.3| 3.1 | 2.8 | D₂  |
| 5   | 19.3  | 99.5| 0.4 | 13.8|
| 6   | 93.5  | 95.2| 3.2 | 56.1|
| 7   | 100   | 99.6| 42.8| 16.8|
| 8   | 96.2  | 95.8| 20.8|
| 9   | 0.9   | 1.8 | 1.7 | 3.1 | D₁  |
| 10  | 16.1  | 17.1| 26.5| 10.3|
| 11  | 78.2  | 25.1| 79.2| 34.5|
| 12  | 98.9  | 92.1| 99.2| 46.2|
| 13  | 93.7  | 95.4| 94.6| 25.1|
| 14  | 100   | 81.5| 7.2 | 12.8|
| 15  | 0.7   | 0.9 | 1.2 | 2.1 | D₁  |
| 16  | 1.5   | 17.5| 36.8| 9.6 | 15.2|
| 17  | 91.1  | 96.2| 93.1| 60.8|
| 18  | 86.7  | 85.9| 88.6| 36.8|
| 19  | 5.8   | 18.2| 0   | 10.1|
| 20  | 38.2  | 42.5| 47.6| 28.1|
| 21  | 37.1  | 11.6| 35.8| 0.02|

| no. | KDPCA | KDSFA | AE |
|-----|-------|-------|----|
|     | T²    | 100   | D₂  |
| 1   | 100   | 5.6   | 22.7|
| 2   | 97.2  | 24.5  | 12.7|
| 3   | 0.9   | 1.3   | 2.5 | D₁  |
| 4   | 4.2   | 2.7   | 6.1 | D₂  |
| 5   | 23.3  | 3.4   | 9.6 | D₂  |
| 6   | 98.5  | 3.5   | 63.1|
| 7   | 100   | 21.3  | 36.5|
| 8   | 96.4  | 1.5   | 6.7 | D₁  |
| 9   | 0.03  | 2.9   | 3.1 | D₁  |
| 10  | 15.2  | 26.7  | 45.1|
| 11  | 18.2  | 38.4  | 67.8|
| 12  | 99.3  | 72.1  | 96.4|
| 13  | 91.5  | 90.5  | D₃  |
| 14  | 100   | 100   | 36.3|
| 15  | 0.2   | 2.9   | 3.1 | D₁  |
| 16  | 2.3   | 26.7  | 45.1|
| 17  | 90.1  | 63.1  | 67.8|
| 18  | 89.1  | 72.1  | 96.4|
| 19  | 6.7   | 90.5  | D₃  |
| 20  | 43.2  | 90.5  | D₃  |
| 21  | 41.4  | 90.5  | D₃  |

| no. | RAE proposed method |
|-----|---------------------|
|     | T²    | 100   | D₂  |
| 1   | 91.2  | 99.9  | 22.2|
| 2   | 96.7  | 99.9  | 22.6|
| 3   | 10.8  | 99.9  | 22.6|
| 4   | 37.2  | 99.9  | 22.6|
| 5   | 26.2  | 99.9  | 22.6|
| 6   | 72.1  | 99.9  | 22.6|
| 7   | 97.1  | 99.9  | 22.6|
| 8   | 87.2  | 97.8  | 99.8|
| 9   | 3.2   | 97.8  | 99.8|
| 10  | 3.4   | 97.8  | 99.8|
| 11  | 62.5  | 97.8  | 99.8|
| 12  | 95.6  | 97.8  | 99.8|
adjustable range of the internal loop flow is too small, which leads to the dynamic abnormality of the system. The monitoring results for this abnormality are shown in Figure 9. For this ramp abnormality, the linear PCA, CCA, and SFA methods detected this anomaly after 400th sampling point, with long time delay. The $T^2$ statistic of KDPCA, the $T^2$ statistic of KDSFA, the $T^2$ statistic of AE, the SPE statistics of RAE, and the $Tee^2$ statistic of our proposed method detected this abnormality from the 470th, 378th, 383th, 375th, and 358th sampling points, respectively. It is clearly that our proposed method could detect this abnormality earlier than the above five methods. Besides, $Sd^2$ and $Se^2$ statistics of KDFSA and $Sd^3$ and $Se^3$ statistics of our proposed method are below their control limits when the blocked degree is not serious and exceed their control limits at the seriously blocked stage. This further indicates that the process dynamic is not influenced in the initially blocked stage and is broken in the seriously blocked stage under the close-loop control operation.

4.2. TE Process. The TE process proposed by Downs and Vogel is the most widely used and highly simulated process model of a realistic chemical industrial process from the Eastman Chemical Company, and this model has been widely used for complex process monitoring tests. The TE process is organized by a reactor, a product condenser, a vapor-liquid separator, a recycle compressor, and a product stripper. The control strategy proposed by Lyman and Georgakis was adopted to ensure that production met the set standards. There are 53 variables, including 12 manipulated variables and 41 measured variables, in the TE process, and the 41 measured variables could be further divided in to 22 continuous process measurements and 19 composition measurements. In this paper, the 12 manipulated variables and 22 continuous process measurements were chosen as the monitoring variables. More details could be obtained in refs 38–40.

The data set of the TE process contained 1460 samples, which were drawn from normal conditions. The test data set contained 960 samples, which were used to detect faults. The descriptions and dynamic property taxonomy of this 21 set case are shown in Table 3, where the dynamic property is referred to in ref 11. In this paper, the dynamic property of the TE process is divided into three types, namely, without dynamics anomaly, short-lived dynamics anomaly, and dynamics anomaly according to the monitoring results provided by SFA and the specific control performance analysis. Without dynamic anomaly means that the static deviation occurred but the process remained steady under the closed-loop control. Short-lived dynamics anomaly means that the static deviation occurred and the process was not stabilized rapidly while under the compensation by the controller but was stabilized eventually. Dynamics anomaly means that both the static deviation occurred and the process steady is broken, and the close-loop control compensation also could not bring the process back to steady-state. In these cases, anomalies are introduced in the last 800 samples, and the former 160 samples are still fault-free. In the offline modeling process, the 1460 samples of normal condition data were used to build the monitoring model. The dynamic order was selected as 2 by referring to ref 11. For KDPCA and KDSFA, the Gaussian kernel function was selected as the kernel function with the kernel parameter $c = 75000$, and the number of principal components was determined according to the 95% cumulative contribution rate in the KDPCA and KDSFA methods. The structure of the AE was set to $33 \times 200 \times 50 \times 50 \times 50 \times 100 \times 50 \times 200 \times 33$. The structure of the RA was set to $(2 \times 33)−(2 \times 27)−(2 \times 27)−(2 \times 33)$. In the proposed method, the encoders in the AE and the RA are used as the spatial and temporal encoders, respectively. The decoder structure was set to $50 \times 100 \times 200 \times 33$, the same as the structure of the decoder in the AE. In the training process for the AE, the RA, and the proposed method, tanh function was chosen as the activation function, the learning rate was set at 0.001, the epoch was set at 300, and the batch size was set as 30. In the online monitoring stage, the 21 test data sets were used to test the monitoring performance. The detection rate is shown in Table 4, where $D_1$, $D_2$, and $D_3$ represent without dynamics anomalies, short-lived dynamics anomalies, and dynamics anomalies, respectively.

According to the monitoring results shown in Table 4, we divided the 21 test cases in to three categories: the easily detected abnormalities, the incompletely detected abnormalities, and the difficult to detect abnormalities. For the easily detected abnormalities, including cases 1, 2, 4, 6–8, 12–14, 17, and 18, all compared methods easily detected this abnormality due to the large degree of abnormality, and both KDSFA and our proposed method could clearly judge the dynamic property. For the incompletely detected abnormalities, including cases 4, 5, 10, 11, 16, and 19–21, our proposed method was more sensitive than other compared methods and provided the highest detection rate, since on the one hand our proposed method applied the prediction to shrink the variability in temporal correlations to focus the unvariability and on the other hand our proposed monitoring scheme divide the process monitoring model into a nonlinear spatiotemporal model and a linear model to analysis the process more detail. We take the cases 5 and 19 as examples to illustrate the detailed monitoring results, as shown in Figures 10 and 11.

Table 4. continued

| no. | $T^2$ | SPE | $Td^2$ | $Ted^2$ | $Tee^2$ | $Sd^2$ | $Sd^3$ | $Se^2$ | dynamic property |
|-----|------|-----|--------|---------|---------|-------|-------|-------|-----------------|
| 13  | 72.1 | 94.4| 96.1   | 96.5    | 96.9    | 52.3  | 87.6  | 91.2  | $D_3$           |
| 14  | 89.2 | 100 | 100    | 100     | 100     | 100   | 100   | 100   | $D_3$           |
| 15  | 6.3  | 15.9| 13.6   | 10.5    | 17.8    | 0.00  | 0.01  | 0.00  | $D_1$           |
| 16  | 21.8 | 53.2| 42.1   | 99.1    | 98.7    | 37.1  | 95.6  | 92.3  | $D_3$           |
| 17  | 86.7 | 95.8| 100    | 99.6    | 99.1    | 73.2  | 85.6  | 84.2  | $D_3$           |
| 18  | 75.3 | 89.0| 90.2   | 88.6    | 89.1    | 88.2  | 83.4  | 82.1  | $D_3$           |
| 19  | 6.8  | 38.9| 12.7   | 40.2    | 83.7    | 18.7  | 31.6  | 76.8  | $D_3$           |
| 20  | 63.8 | 76.1| 62.5   | 87.5    | 84.6    | 37.1  | 35.7  | 38.2  | $D_3$           |
| 21  | 17.8 | 52.7| 70.4   | 68.5    | 69.2    | 0.00  | 0.03  | 0.01  | $D_1$           |

Table 3, where the dynamic property is referred to in ref 11. In this paper, the dynamic property of the TE process is divided into three types, namely, without dynamics anomaly, short-lived dynamics anomaly, and dynamics anomaly according to the monitoring results provided by SFA and the specific control performance analysis. Without dynamic anomaly means that the static deviation occurred but the process remained steady under the closed-loop control. Short-lived dynamics anomaly means that the static deviation occurred and the process was not stabilized rapidly while under the compensation by the controller but was stabilized eventually. Dynamics anomaly means that both the static deviation occurred and the process steady is broken, and the close-loop control compensation also could not bring the process back to steady-state. In these cases, anomalies are introduced in the last 800 samples, and the former 160 samples are still fault-free. In the offline modeling process, the 1460 samples of normal condition data were used to build the monitoring model. The dynamic order was selected as 2 by referring to ref 11. For KDPCA and KDSFA, the Gaussian kernel function was selected as the kernel function with the kernel parameter $c = 75000$, and the number of principal components was determined according to the 95% cumulative contribution rate in the KDPCA and KDSFA methods. The structure of the AE was set to $33 \times 200 \times 50 \times 50 \times 50 \times 100 \times 50 \times 200 \times 33$. The structure of the RA was set to $(2 \times 33)−(2 \times 27)−(2 \times 27)−(2 \times 33)$. In the proposed method, the encoders in the AE and the RA are used as the spatial and temporal encoders, respectively. The decoder structure was set to $50 \times 100 \times 200 \times 33$, the same as the structure of the decoder in the AE. In the training process for the AE, the RA, and the proposed method, tanh function was chosen as the activation function, the learning rate was set at 0.001, the epoch was set at 300, and the batch size was set as 30. In the online monitoring stage, the 21 test data sets were used to test the monitoring performance. The detection rate is shown in Table 4, where $D_1$, $D_2$, and $D_3$ represent without dynamics anomalies, short-lived dynamics anomalies, and dynamics anomalies, respectively.

According to the monitoring results shown in Table 4, we divided the 21 test cases in to three categories: the easily detected abnormalities, the incompletely detected abnormalities, and the difficult to detect abnormalities. For the easily detected abnormalities, including cases 1, 2, 4, 6–8, 12–14, 17, and 18, all compared methods easily detected this abnormality due to the large degree of abnormality, and both KDSFA and our proposed method could clearly judge the dynamic property. For the incompletely detected abnormalities, including cases 4, 5, 10, 11, 16, and 19–21, our proposed method was more sensitive than other compared methods and provided the highest detection rate, since on the one hand our proposed method applied the prediction to shrink the variability in temporal correlations to focus the unvariability and on the other hand our proposed monitoring scheme divide the process monitoring model into a nonlinear spatiotemporal model and a linear model to analysis the process more detail. We take the cases 5 and 19 as examples to illustrate the detailed monitoring results, as shown in Figures 10 and 11.
In case 5, the inlet temperature of the condenser cooling water was changed by introducing a step disturbance. The temperature of the cooling water will further influence the final product. However, the flow of the cooling water is manipulated by the PI controller to make sure the products are condensed. Therefore, the change in the inlet temperature of the cooling water will further influence the final product.
condenser cooling water would introduce short-lived dynamics anomalies in the controller adjustment stage, and the system would go back to the without dynamics anomaly state when the temperature of the cooling water reached a new steady state. The monitoring results of this fault are shown in Figure 10. As shown in Figure 10(a), the two statistics of the PCA method detected this anomaly between the 160th and 400th samples and were back to normal after the 400th sample,
indicating that the PCA method could not detect this anomaly completely. Similarly, the Te\(^2\) statistic of CCA, the Td\(^2\) and Te\(^2\) statistics of KDSFA (Figure 10(e)), the T\(^2\) and SPE statistics of AE and RAE (Figure 10(f and g, respectively), and the Td\(^2\) statistic of our proposed method (Figure 10(h)) have the same detection results. None of them could detect this step anomaly. As shown in Figure 10(c), the Td\(^2\) and Te\(^2\) statistics of SFA could detect this abnormal but with some sample delay. It is notable that the Sed\(^2\) and See\(^2\) statistics of our proposed method could completely detect this anomaly from 160th sample to the end as shown in Figure 10(h). At the same time, the Sd\(^2\), Sed\(^2\), and See\(^2\) statistics of the proposed method could further indicate that this anomaly only introduced a short dynamic abnormality and the system gradually returned to normal under the operation of the control.

In case 19, the specific anomaly described is unknown, but this disturbance yields consistent disruption in process dynamics according to the output product inconsistent with the set. The monitoring results of the five methods are shown in Figure 11. As shown in Figure 11, the T\(^2\) and SPE statistics of the PCA method, the Te\(^2\) statistic of the CCA method, and the Td\(^2\) statistic of the SFA method detect this abnormal with many missing detection points. The SPE statistic of the KDPCA method also detected this abnormal with many missing detection points. The Te\(^2\) and Td\(^2\) statistics of the KDSFA method did not exceed their control limits when the abnormality occurred, but the See\(^2\) and Sd\(^2\) statistics detected this dynamics anomaly. However, there were also low detection rates in the Se\(^2\) and Sd\(^2\) statistics. Although the detection rates of the SPE statistic of the AE and RAE methods were higher than those for KDPCA and KDSFA due to the optimal determination parameters, it was still low. As shown in Figure 11(h), the detection rate of the Ttee\(^2\) statistic of our proposed method was significantly higher than those of other methods; at the same time, the See\(^2\) statistic of our proposed method could further indicate the dynamic anomaly with a higher detection rate than the Sd\(^2\) and See\(^2\) statistics of the KDSFA method. In conclusion, the proposed method has the highest fault detection rate compared with other methods for such incompletely detected faults and can accurately determine whether the anomaly belongs to a dynamics anomaly. For the difficult to detect anomalies, including cases 3, 9, 15, it was hard for all compared methods to detect the abnormalities. Although our proposed method has a higher detection rate than other methods, the significantly detection of those anomalies was difficult.

5. CONCLUSION

In this paper, we proposed a novel nonlinear spatiotemporal process feature learning-based concurrent process condition monitoring method for complex industrial processes. In this method, a new pseudo-Siamese network was designed to extract the high-value nonlinear dynamic process slow features with separate explicit spatial and temporal models by simultaneously considering the nonlinear spatial information that reflects correlation among process variables and the nonlinear temporal information that reveals the dynamic character in the temporal correlation of process variable inner. The explicit temporal model is further used to shrink the variability of the nonlinear dynamic process feature to focus on the unvariability to improve the motoring efficiency of the nonlinear dynamic process concurrent monitoring method. For real industrial process monitoring, the nonlinear, the linear, and the dynamic relationships often mix. Although our proposed method extracted the high-value nonlinear spatiotemporal process feature to represent the main nonlinear dynamic information over the whole process, it still inevitable left some linear dynamic features in the reconstruction residual space. Therefore, the reconstruction residual space was further analyzed using the SFA method and divided by the linear dynamic slow feature space and the real residual space. On the basis of the above modeling process, the new monitoring scheme is proposed, with six monitoring statistics for the concurrent monitoring of the steady deviation and dynamics anomalies. The experiment results for the WWT process and the TE process demonstrate that the proposed method not only can exhibit the best monitoring performance compared to the KDPCA, KDSFA, AE, and RAE methods but also can further indicate whether the abnormality belongs to the dynamics anomaly. In the future, the proposed method could be improved to monitor more complex process, such as unsteady process or time varying process. Besides, it is meaningful to use the proposed method in combination with the attention mechanism to achieve process fault isolation to explore the key variables that introduce the process anomaly.

## AUTHOR INFORMATION

### Corresponding Author

Mingxing Jia — College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China; Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, Liaoning 110819, China; orcid.org/0000-0003-3305-0357; Email: jiamingxing@ise.neu.edu.cn

### Authors

Yuemei Xu — College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China
Zhizhong Mao — College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China; Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, Liaoning 110819, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c05028

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Key Program of National Natural Science Foundation of China (no. 51634002) and National Natural Science Foundation of China (no. 61773101).

## REFERENCES

(1) Jiang, Q.; Yan, X.; Huang, B. Review and Perspectives of Data-Driven Distributed Monitoring for Industrial Plant-Wide Processes. Ind. Eng. Chem. Res. 2019, 58, 12899–12912.
(2) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N.; Yin, K. A review of process fault detection and diagnosis: Part III: Process history based methods. Comput. Chem. Eng. 2003, 27 (3), 327–346.
(3) Ge, Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemometr. Intell. Lab. Syst. 2017, 171, 16–25.
(4) Yin, S.; Ding, S. X.; Xie, X.; Luo, H.; et al. A review on basic data-driven approaches for industrial process monitoring. IEEE Trans. Ind. Electron. 2014, 61 (11), 6418–6428.

(5) Qin, S. J. Statistical process monitoring: Basics and beyond. J. Chemometrics 2003, 17, 480–502.

(6) Zhao, C.; Gao, F. Fault subspace selection approach combined with analysis of relative changes for reconstruction modeling and multifeature diagnosis. IEEE Trans. Control Syst. Technol. 2016, 24, 928–939.

(7) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. Chemom. Intell. Lab. Syst. 1982, 7, 37–52.

(8) Jiang, Q.; Yan, X. Non-Gaussian chemical process monitoring with adaptively weight independent component analysis and its applications. J. Process Control. 2013, 23 (9), 1320–1331.

(9) Juricek, B. C.; Seborg, D. E.; Larimore, W. E. Fault detection using canonical variate analysis. Ind. Eng. Chem. Res. 2004, 43, 458–474.

(10) Wiskott, L.; Sejnowski, T. J. Slow feature analysis: Unsupervised learning of invariances. Neural computation. 2002, 14, 715–770.

(11) Shang, C.; Yang, F.; Gao, X.; Huang, X.; Suykens, J. A.; Huang, D. Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis. AIChE J. 2015, 61, 3666–3682.

(12) Zhang, S.; Zhao, C.; Huang, B. Simultaneous static and dynamic analysis for fine-scale identification of process operation statuses. IEEE Trans. Industr. Inform. 2019, 15 (8), 5320–5329.

(13) Ge, Z.; Yang, C.; Song, Z. Improved kernel PCA-based monitoring approach for nonlinear processes. Chem. Eng. Sci. 2009, 64 (9), 2245–2255.

(14) Cai, L.; Tian, X.; Chen, S. Monitoring nonlinear and non-Gaussian processes using Gaussian mixture model-based weighted kernel independent component analysis. IEEE Trans. Neural Netw. Learn. Syst. 2017, 28, 122–135.

(15) Pilario, KE S; Cao, Y.; Shahlie, M. Mixed Kernel Canonical Variate Dissimilarity Analysis for Incipient Fault Monitoring in Nonlinear Dynamic Processes. Comput. Chem. Eng. 2019, 123, 143–154.

(16) Zhang, N.; Tian, X.; Cai, L.; Deng, X.; et al. Process fault detection based on dynamic kernel slow feature analysis. Comput. Electr. Eng. 2015, 41, 9–17.

(17) Luo, L.; Bao, S.; Mao, J.; Tang, D. Nonlinear process monitoring based on kernel global–local preserving projections. J. Process Control. 2016, 38, 11–21.

(18) Nasir, V.; Sassani, F. A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges. Int. J. Adv. Manuf. Technol. 2021, 115, 2683–2709.

(19) Kusiak, A. Fundamentals of smart manufacturing: a multithread perspective. Annu. Rev. Control 2019, 47, 214–220.

(20) Jiang, Q.; Ding, S.; Wang, Y.; Yan, X. Data-driven distributed local fault detection for large-scale processes based on GA-regularized canonical correlation analysis. IEEE Trans. Ind. Electron. 2017, 64 (10), 8148–8157.

(21) Onel, M.; Burnak, B.; Pistikopoulos, E. N. Integrated datadriven process monitoring and explicit fault-tolerant multi-parametric control. Ind. Eng. Chem. Res. 2020, 59, 2291–2306.

(22) Yan, W.; Guo, P.; Gong, L.; et al. Nonlinear and robust statistical process monitoring based on variant autoencoders. Chemom. Intell. Lab. Syst. 2016, 158, 31–40.

(23) Zhang, Z.; Jiang, T.; Li, S.; et al. Automated feature learning for nonlinear process monitoring – An approach using stacked denoising autoencoder and k-nearest neighbor rule. J. Process Control 2018, 64, 49–61.

(24) Yu, W.; Zhao, C. Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net. IEEE Trans Control Syst. Technol. 2020, 28, 1083–91.

(25) Zhang, C.; Yu, J.; Ye, L. Sparsity and manifold regularized convolutional auto-encoders-based feature learning for fault detection of multivariate processes. Control Eng. Pract. 2021, 111 (7), 104811.

(26) Bayer, J.; Osendorfer, C. Learning stochastic recurrent networks. arXiv:1411.7610v3 [stat.ML] Statistics.Machine Learning. Mar 5, 2015, 1411.7610. DOI: 10.48550/arXiv.1411.7610

(27) Essien, A.; Giannetti, C. A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. IEEE Trans. Ind. Inf. 2020, 16 (9), 6069–6078.

(28) Cheng, F.; He, Q.; Zhao, J. A novel process monitoring approach based on variational recurrent autoencoder. Comput. Chem. Eng. 2019, 129, 106515.

(29) Hinton, G. E.; Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. Science. 2006, 313, 504–507.

(30) Chipulkara, R.; Huang, B. Siamese Neural Network-Based Supervised Slow Feature Extraction for Soft Sensor Application. IEEE Trans. Indul. Electron. 2021, 68, 8953–8962.

(31) Zhou, X.; Liang, W.; Shimizu, S.; et al. Siamese Neural Network Based Few-Shot Learning for Anomaly Detection in Industrial Cyber-Physical Systems. IEEE Trans. Indul. Inform. 2021, 17, 5790–5798.

(32) Zhao, C.; Sun, H. Dynamic distributed monitoring strategy for large-scale nonstationary processes subject to frequently varying conditions under closed-loop control. IEEE Trans. Indul. Electron. 2019, 66, 4749–4758.

(33) Xu, Y.; Jia, M.; Mao, Z. A novel auto-regressive dynamic slow feature analysis method for dynamic chemical process monitoring. Chem. Eng. Sci. 2022, 248, 117236.

(34) Botev, Z. I.; Grotowski, J. F.; Kroese; et al. Kernel density estimation via diffusion. Ann. Stat. 2010, 38, 2916–2957.

(35) Wang, Y.; Mao, Z.; Jia, M. Feature-Points-Based Multimodel Single Dynamic Kernel Principle Component Analysis (M-SDKPCA) Modeling and Online Monitoring Strategy for Uneven-Length Batch Processes. Ind. Eng. Chem. Res. 2013, 52 (34), 12059–12071.

(36) Cheng, H.; Wu, J.; Huang, D.; et al. Robust adaptive boosted canonical correlation analysis for quality-relevant process monitoring of wastewater treatment. ISA Transactions. 2021, 117, 210–220.

(37) Jeppsson, U.; Rosen, C.; Alex, J.; et al. Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. Water Sci. Technol. 2006, 53, 287–295.

(38) Downs, J. J.; Vogel, E. F. A plant-wide industrial problem process. Comput. Chem. Eng. 1993, 17, 245–55.

(39) Lou, Z.; Wang, Y. New nonlinear approach for process monitoring: neural component analysis. Ind. Eng. Chem. Res. 2021, 60, 387–398.

(40) Li, S.; Luo, J.; Hu, Y. Nonlinear process modeling via unidimensional convolutional neural networks with self-attention on global and local inter-variable structures and its application to process monitoring. ISA Transactions. 2022, 121, 105–118.