Review

A Review on Data-Driven Process Monitoring Methods: Characterization and Mining of Industrial Data

Cheng Ji ![ORCID] and Wei Sun *![ORCID]

College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing 100029, China; 2020400007@mail.buct.edu.cn
* Correspondence: sunwei@mail.buct.edu.cn; Tel.: +86-010-6444-5826

Abstract: Safe and stable operation plays an important role in the chemical industry. Fault detection and diagnosis (FDD) make it possible to identify abnormal process deviations early and assist operators in taking proper action against fault propagation. After decades of development, data-driven process monitoring technologies have gradually attracted attention from process industries. Although many promising FDD methods have been proposed from both academia and industry, challenges remain due to the complex characteristics of industrial data. In this work, classical and recent research on data-driven process monitoring methods is reviewed from the perspective of characterizing and mining industrial data. The implementation framework of data-driven process monitoring methods is first introduced. State of art of process monitoring methods corresponding to common industrial data characteristics are then reviewed. Finally, the challenges and possible solutions for actual industrial applications are discussed.

Keywords: chemical industrial process; fault detection and diagnosis; feature extraction; machine learning; multivariate statistical methods; deep learning; hybrid methods; complex nonlinear process; dynamic process; nonstationary process; multimode continuous process; batch process; Tennessee Eastman process; fault propagation analysis

1. Introduction

Chemical process accidents have become a critical threat to sustainable development in our society. With the newly proposed carbon peaking and carbon neutrality goals, governments have strengthened the supervision on production safety, including establishing regulations and carrying out technical training. However, safety accidents still occur every year, resulting in major losses of property and human lives [1,2]. According to numerous statistical analysis studies on chemical accidents, it can be concluded that human error is the main cause of chemical accidents [3,4]. Although distributed control systems (DCS) have been widely applied in the chemical industry, it is still difficult for operators to detect abnormal process deviations and make proper decisions to eliminate them at an early stage due to the increasing scale of chemical production and equipment complexity. The operators can only focus on a few key variables out of a large number of process variables in DCS, and unnecessary alarms can be quite overwhelming without an effective fault detection and diagnosis system. Therefore, the emergence of process monitoring technology is crucial and necessary in the chemical industry.

Process monitoring technology has been developed as a useful tool to assist operators to ensure product quality and production safety. Process monitoring can be implemented in two steps, fault detection and diagnosis (FDD). Fault detection aims at determining whether the process is operating under normal conditions, and fault diagnosis works after a fault is detected to determine the root cause of the fault. Taking advantage of the wide application of DCS, massive historical data, which contain internal process operation mechanism information, could be obtained, making data-driven process monitoring methods a popular

research focus, especially for industrial processes whose mechanistic models are hard to build [5]. In the last 20 years, process monitoring technology has been developed to be an important and indispensable branch of process system engineering (PSE) [6,7]. Comprehensive reviews of process data analytics and its application on process monitoring have been provided and discussed. In Chiang's work in 2001, traditional multivariate statistical methods were introduced and compared based on their performance on process monitoring results of the Tennessee Eastman process (TEP) [8]. In 2003, process history-based methods were reviewed and classified as an important type of process fault detection and diagnosis in a series of reviews by Venkatasubramanian [9–11]. Qin made a deep survey on the advanced development of principal component analysis (PCA) and partial least squares (PLS)-based process monitoring methods [12]. Reis and Gins discussed the industrial process monitoring issues in the era of big data from the perspectives of detection, diagnosis and prognosis [13]. Severson et al. provided perspectives on progress in process monitoring systems by summarizing methods for each step of the process monitoring loop over the last twenty years [14]. Ge introduced the framework of industrial process modeling and monitoring and reviewed data-driven methods according to various aspects of plant-wide industrial processes [15]. Qin and Chiang presented the current development of machine learning and artificial intelligence (AI)-based data analytics and discussed opportunities in AI-based process monitoring [16].

In recent years, with the popularity of the notion of industry 4.0 and the digital factory, data-driven process monitoring technologies have gradually attracted attention from industry. Unlike simulated data, industrial process data show more complex characteristics due to various practical operating conditions, which provides a huge challenge to industrial process monitoring. Figure 1 provides a statistic regarding the publication of proposed process monitoring methods in the past five years. This statistic was generated from the Web of Science by searching the keywords, "chemical process monitoring" and "data-driven".

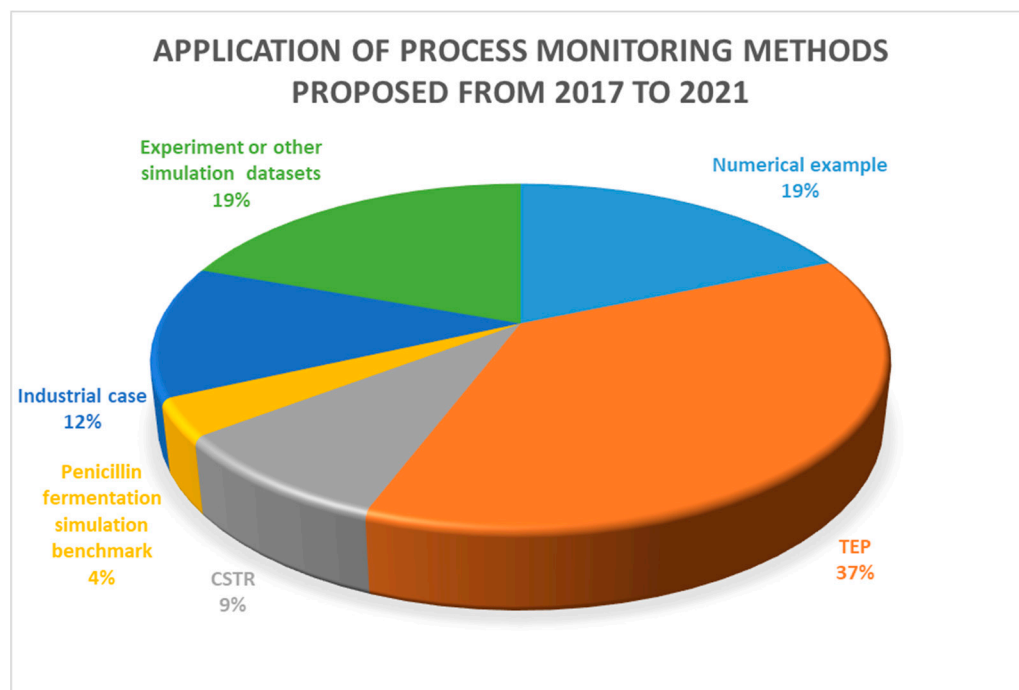**Figure 1.** A statistic on the application of process monitoring methods.

It can be seen that most proposed data-driven process monitoring methods were applied to simulated cases, such as numerical examples, TEP, continuous stirred tank reactor (CSTR), and penicillin fermentation simulation benchmarks, and only 12 percent of the studies were applied to practical industrial processes, indicating the huge difficulty in

industrial process monitoring. Process monitoring methods that achieve good performance in simulation processes cannot be directly applied to industrial processes, because there are significant differences between practical industrial data and simulation data. Simulation process data are generally simulated under a single ideal operating condition, while industrial process data show complex characteristics due to various factors in practical production, as summarized in Figure 2. Industrial processes are not limited to a single operating condition, and characteristics of normal operating conditions also vary with different processes, which brings great challenges to the monitoring of industrial processes. On the other hand, it also indicates that industrial process monitoring can be well implemented if the normal operating conditions can be correctly defined and the characteristics of the normal operating conditions data can be effectively and completely extracted. The focus of this work is to provide an overview of the current status and recent progress in data-driven process monitoring from the perspective of the characterization and mining of industrial data. This paper is organized as follows: a general data-driven process monitoring framework is presented in Section 2. In Section 3, data-driven process monitoring methods are reviewed according to the classification of corresponding industrial process data characteristics. In Section 4, challenges and prospects regarding the application of industrial process monitoring are provided and discussed. Conclusions are presented in Section 5.
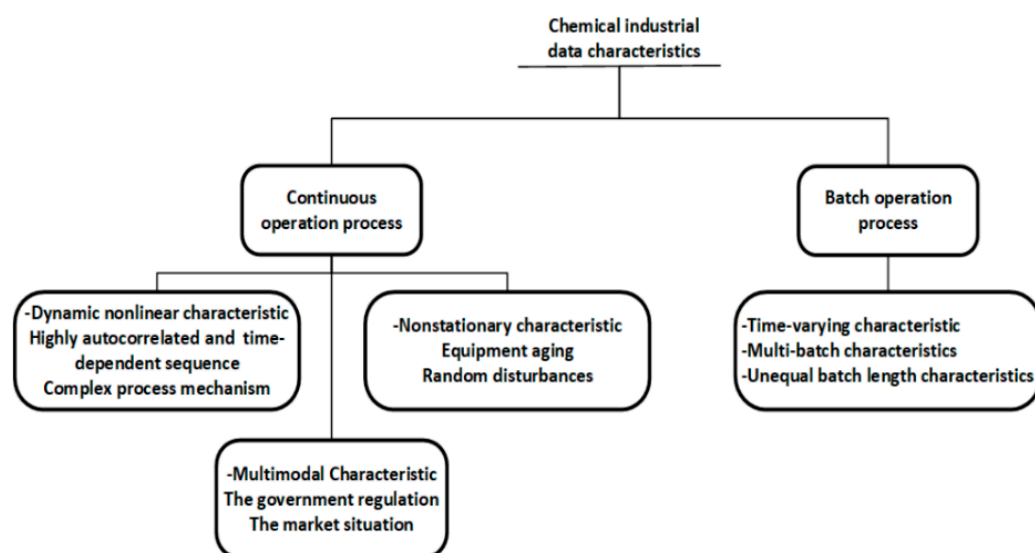
Figure 2. The characteristics of data from the chemical process industry.

2. Data-Driven Process Monitoring

In this section, a data-driven process monitoring framework is presented from the perspective of defining normal operating conditions, data preprocessing, feature extraction, monitoring statistics, and fault diagnosis. The flow diagram of the data-driven industrial process monitoring framework is shown in Figure 3. Common methods applied to these key procedures are presented in each corresponding subsection.

Figure 3. Data-driven industrial process monitoring framework.

2.1. Definition of Normal Operating Conditions

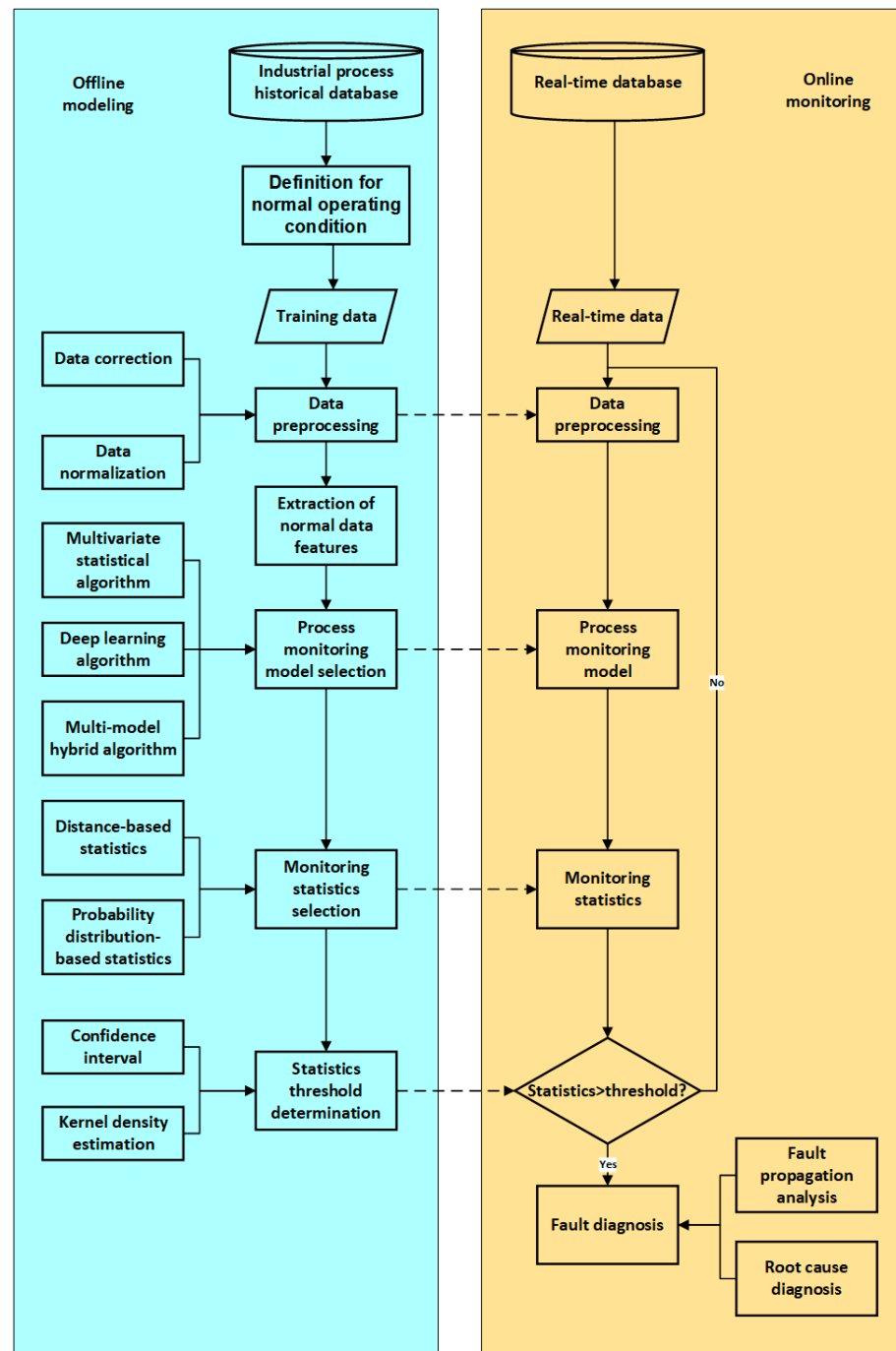Different from commonly used simulation processes, such as CSTR and TEP, there are usually multiple complex operating conditions existing in industrial processes. In the academic field of process monitoring, the performance of the newly proposed method can be easily tested by comparing the process monitoring results using fault datasets, when a novel feature extraction method is proposed to extract the common feature of normal datasets and calculate the monitoring statistic. However, in industrial applications, the normal operating conditions and fault conditions are usually not so clearly distinguished as those in the simulation processes. The definition of normal operating conditions and the labelling of data are usually the most important and time-consuming tasks in building

an industrial process monitoring model. With the increasing scale of modern industrial processes, huge historical datasets with high dimensionality are available, while the information for upcoming process situations is poor. Various types of random variations and unexpected disturbances are mixed up in the large amount of historical operating data. If these normal random variations are not fully captured when training a process monitoring, a large number of false alarms will be caused during online monitoring. On the other hand, if the unexpected disturbances in historical data are included in offline modeling, these kinds of abnormal conditions will be regarded as normal, resulting in alarm missing for real faults. Therefore, it is important to effectively separate abnormal disturbances from normal random variations. A certain process internal mechanism for information needs to be incorporated to help characterize data features under normal operating conditions. More importantly, it is necessary to consider the complex data characteristics in industrial processes. Among them, the multimode characteristics and nonstationary characteristics should be considered in the definition of normal operating conditions, these are analyzed in the following sections.

2.1.1. Multimode Characteristics

In industrial processes, production load is frequently adjusted due to fluctuations in the market price of the product and the upgradation of government regulations, especially in the context of carbon emission restrictions. Therefore, data in actual production often display multimode characteristics, which could be defined as at least one variable that does not follow a single steady operating condition due to various changes in production loads, feed flow, and set points [17,18]. However, traditional statistic process monitoring models are established under the assumption that the process is operated at a single stable working point. When the operating condition is switched, the mean and variance of the data change significantly, and massive false alarms will be triggered, which can be categorized to multimode continuous process monitoring. In industrial scenarios, future modes are hard to estimate and are usually not available in historical datasets.

At first, data from multiple operating conditions cannot be simply integrated into one single training dataset, because the normal working point determined in this way is just an average of different operating conditions. At the same time, data associated with abnormal transitions between normal operating conditions may also belong to the assumed normal data range, which will affect the early identification of abnormal process deviations. Another challenge is that the operating conditions contained in training data will not cover all possible situations in actual production. Even if a corresponding model is established for each normal operating condition, when a brand-new operating condition appears, it will still be mistaken as a fault by such methods. In addition, the consideration of the transition states is inevitable since the switching of operating conditions cannot be completed instantaneously. The data characteristics in transition states are significantly different from steady states, as the mean of certain variables keeps changing until a new operating condition is reached. It is difficult to extract the common features of transition states, and the monitoring of transition states is the most critical challenge in multimode process monitoring. Detailed reviews of different methods for multimode process monitoring are illustrated in Section 3.2.

2.1.2. Dynamic and Nonstationary Characteristics

In addition to multimode characteristics, different types of dynamic and nonstationary characteristics caused by various factors need to be considered in the definition of normal operating conditions. The concept of "dynamic" comes from process control and monitoring in PSE, and the concept of "nonstationary" is mostly defined in the field of time series analysis. Both "dynamic" and "nonstationary" can be considered as the time-variate nature of the variables in the process. In the category of process monitoring and fault diagnosis, the dynamic characteristics are obtained because sequences of certain variables are highly autocorrelated due to internal mechanisms and the response of control systems.

Moreover, nonstationary characteristics are also reflected in practical production due to equipment aging, and random disturbances in process or environment. Resulting from the existence of these complex data characteristics, the means and variances of the variables are time-varying even in a single normal operating condition, particularly in a batch process. The time-varying characteristics violate the assumption of traditional multivariate statistic process monitoring that the process is time-independent, and therefore, limit the application of industrial process monitoring. In industrial practice, these time-varying characteristics caused by process dynamics and non-stationarity shall be defined as normal conditions, otherwise they will be regarded as process faults in real time monitoring, resulting in massive false alarms. In addition, minor abnormal changes that happen at the early stage of certain faults can be buried by these time-varying characteristics, which also should be considered when defining the range of normal operating conditions. The development status of process monitoring methods aiming at solving process dynamics, non-stationarity and batch processes are reviewed in Sections 3.1.1, 3.1.3 and 3.3, respectively.

2.2. Data Preprocessing

Data preprocessing is a relatively straightforward but indispensable step in process modeling. There are two main purposes of data preprocessing, data reconciliation and data normalization.

During the process of data acquisition and data transmission, there will inevitably be missing values and outliers due to mechanical problems with data acquisition equipment and sensors. Data reconciliation technology is designed to deal with these problems through data removal or data supplement. When most variables at a measurement point are missing, these samples can be deleted, and if a small number of variables are missing, various data processing methods can be used to supplement the data. The simplest way is to supplement the current value with the value collected at the last moment or the average value of the previous time period. If there is a certain conservation relationship between the missing variables and other variables, the missing value can be supplemented through mass balance, energy balance, etc. To further improve the accuracy of data supplement, soft sensing technology can be applied to build regression models for missing values using historical data. The missing values can be replaced by predicted values through the regression relationship between other measurements and the missing measurements.

After the missing values have been effectively supplemented and outliers have been properly dealt, data normalization should be applied to balance the contribution scale of each variable before modeling. The most commonly used data normalization method, z-score normalization, aims to scale data samples to zero means and unit variance. Z-score normalization is suitable for processing normally distributed data and has become the default standardized method in many data analysis tools, such as PCA, and statistical product and service software (SPSS). Min-max normalization, logarithmic function conversion and logistic/SoftMax transformation are also commonly used data normalization methods. Different data normalization methods have various effects on the results of the model, but there is no general rule to follow in the selection of data normalization methods. In industrial process monitoring, data normalization methods can be selected according to the characteristics of process data and monitoring algorithms. In most cases, industrial data conform to a normal distribution and Z-score normalization is the most reliable method. When the process is highly nonlinear, logarithmic function conversion and kernel functions transformation should be considered. Furthermore, data expansion and batch normalization need to be implemented to unfold three-dimensional data into two dimensions and normalize data in different batches. The criterion to pick a proper data preprocessing method is decided by the requirement of feature extraction. Usually, data preprocessing is considered together with feature extraction. A summary of different data preprocessing methods is shown in Table 1.

Table 1. A summary of data preprocessing methods.

Data Preprocessing Methods	
Data Reconciliation	**Data Normalization**
Data removal; Data reconciliation with mass balance, energy balance; Data reconciliation with soft sensor technology	Z-score normalization; Min-max normalization; Logarithmic function conversion; Logistic/SoftMax transformation; Kernel functions transformation; Data expansion

2.3. Selection of Feature Extraction Method

Feature extraction is the problem of obtaining a new feature space through data conversion or data mapping. When sufficient training data from normal operating conditions have been selected and normalized, a proper feature extraction method should be applied to project or map the data into a low-dimensional feature space where most of the information from the original data can be retained. For example, the traditional statistic method, PCA, is applied to transform a set of originally correlated variables into a set of linearly uncorrelated variables, which are called principal components, through orthogonal transformation.

Based on the characteristics of data under normal operating conditions defined and analyzed in previous steps, corresponding feature extraction methods should be selected for modeling. For a simple linear process in which data are normally distributed, traditional PCA or PLS can be applied for feature extraction. If the process data do not conform to Gaussian distribution, methods that do not require assumptions about the distribution of the training data, such as independent component analysis (ICA) and support vector data description (SVDD), can be selected for feature extraction. When the process variables are highly nonlinear, kernel-based methods or neural network-based methods can be applied to extract process nonlinearity by nonlinear mapping. To extract process dynamic characteristics, time series analysis and autoregressive tools can be included in feature extraction methods. When dealing with more complex data characteristics in nonstationary processes, multimode processes and batch processes, more complicated feature extraction procedures need to be proposed by extracting the common normal trends in complex data features, and sometimes even a combination of multiple methods is required. A simple summary of different feature extraction methods for process monitoring is shown in Table 2. A detailed review of feature extraction methods for various industrial process data characteristics is provided in Section 3. Once the feature extraction method has been determined, the model needs to be trained and a corresponding statistic and its confidence interval should be constructed for fault detection. An introduction to monitoring statistics will be presented in the next section.

Table 2. A summary of commonly used feature extraction methods.

Methods	Usage
PCA/PLS	Linear Gaussian process monitoring
ICA/SVDD	Non- Gaussian process monitoring
Kernel-based methods/neural network	Nonlinear process monitoring
Dynamic PCA/PLS	Dynamic process monitoring
Common trends analysis/cointegration	Nonstationary process monitoring

2.4. Monitoring Statistics

Feature extraction methods can only be applied to obtain a mapping or projection direction to transform the original data into a feature space, but it cannot be directly used to monitor the change in operating conditions [19]. The construction of monitoring statistics is required for fault detection. Monitoring statistic is a measure of the distance between the measurement point and the original point in the feature space. Distance indicators have been widely used in constructing monitoring statistics, such as Euclidean distance

and Mahalanobis distance. Typically, T^2 statistic is calculated to measure the variations in principal component subspace using Mahalanobis distance. Considering data under normal operating conditions following a multivariate normal distribution, the T^2 statistic can be regarded as an *F*-distribution, and the control limits can be determined at different confidence levels [20]. When the data do not conform to normal distribution, the control limits of T^2 statistic will not be reliable [12]. In this case, the squared prediction error (SPE) statistic calculated based on Euclidean distance generally performs better to measure the projection on the residue subspace. The control limits can be obtained by approximating SPE statistics to a normal distribution with zero mean and unit variance [21]. With the development of feature extraction technology, many novel monitoring statistics have been proposed accordingly. In moving PCA, a novel monitoring statistic is proposed to measure the difference in the angle of principle components between the current operating condition and the reference normal operating conditions [19,22,23]. The difference in the probability distribution of data can also be applied as a monitoring statistic. Kano et al. proposed a process monitoring method by monitoring the distribution of process data, and therefore a dissimilarity index is introduced to quantify the difference in distribution between two sequences [24]. As a common measure of the similarity between two probability distribution functions, Kullback–Leibler divergence (KLD) was introduced by Harrou et al. as a fault decision statistic to measure the variations of residues from PLS [25]. Zeng et al. developed statistics based on KLD to measure variations in the probability distribution functions of data using a moving window [26]. In the work of Cheng et al., negative variational score is derived from KLD as a novel statistic to monitor the distribution probability of data in probability space [27].

To simplify the fault detection, multiple monitoring statistics can be combined and integrated into a single statistic. Decision fusion strategies are usually applied to fuse results from various methods to combine their strengths [28]. Raich and Cinar proposed a combined statistic by introducing a weighting factor to the T^2 statistic and the SPE statistic [29]. Yue and Qin also proposed a new statistic combing the T^2 statistic and the SPE statistic, but the control limits are constructed using the distribution of quadratic forms rather than using their respective thresholds [30]. Different from a simple combination of the T^2 statistic and the SPE statistic, in the work of Choi and Lee, the T^2 statistic and the SPE statistic are combined into a unified index from the estimated likelihood of sample distribution [31]. Another widely applied fault decision fusion method is Bayesian inference, by which an ensemble probabilistic index can be generated from multiple monitoring statistics [32]. Moreover, SVDD has also been reported in multiblock methods to integrating monitoring statistics of multiple sub-blocks for plant-wide process monitoring [33]. Based on the fusion of multiple monitoring statistics, variations in each feature space or sub-block could be comprehensively measured with a single statistic, which significantly simplifies the monitoring of process operating conditions and the shortcomings of each statistic could be overcome to a certain extent and an objective process monitoring result could be obtained. A summary of different types of statistics for process monitoring is shown in Table 3.

Table 3. A summary of different statistics for process monitoring.

| | Process Monitoring Statistics | |
Distance-Based Statistics	Probability Distribution-Based Statistics	Other Statistics
Mahalanobis distance: T^2 statistic; Euclidean distance: SPE statistic	Kullback–Leibler divergence; Negative variational score	Distribution of process data; Angle of principle components; Fault decision fusion strategy

2.5. Fault Diagnosis

After the monitoring statistics calculated from real time data exceed their thresholds determined from offline modeling, it is critical to isolate the root cause of the fault through fault diagnosis. Fault diagnosis includes two consecutive steps: identification of fault

variables and analysis of fault propagation path. The identification of fault variables aims to find out which variables are beyond their normal ranges and how much they contribute to the fault. The challenge is that the fault can be rapidly propagated from root variable to other variables with mass transfer and heat transfer, etc. A large number of variables will be identified and the variable with the largest contribution to the fault may not be the root cause, but just a result caused by the fault. Therefore, the fault propagation path among these identified fault variables needs to be further analyzed. If causal reasoning among these variables could be obtained, the fault propagation could be displayed in a network diagram and then the real root cause of the fault could be located. The acquisition of causal reasoning models among process variables has always been a challenge. Especially as industrial processes become more and more complex, variables are highly correlated with each other, and the causal logic among process variables may change due to the response of the control systems, which makes it difficult to capture the real time causal correlations among process variables. The solutions to these challenges and the recent development of data-driven fault diagnosis methods are presented in Section 3.4.

Generally, fault diagnosis has an inseparable connection with fault detection. The fault variable's isolation is usually realized on the basis of the fault detection model. For example, the fault diagnosis is implemented by finding the value of which variable exceeds its set point and which is determined offline for fault detection. In addition, fault variables identified in contribution plots are obtained by calculating their degree of deviation in corresponding feature spaces, which are defined by fault detection methods.

3. Review of Data-Driven Process Monitoring Methods in the Process Industry

In this section, data-driven process monitoring methods corresponding to industrial data characteristics are reviewed, including traditional multivariate statistical methods and advanced machine learning methods. The development of continuous process monitoring methods for a single steady state are first investigated. Advances in monitoring strategies for industrial nonlinear processes, dynamic processes, nonstationary processes, multimode processes, and batch processes are then reviewed separately.

3.1. Continuous Process Monitoring

In the 1980s, DCS began to be applied, which significantly facilitated the development of data-driven process monitoring methods. Based on historical data stored in the process operation, a proper data analysis algorithm is adopted in data-driven FDD methods to extract the steady state characteristics under normal operating conditions. In 1993, TEP was simulated by Eastman chemical company, which contained five chemical units, 52 process variables and 20 different types of faults [34]. It provided a benchmark for researchers in the field of process monitoring to compare the performance of their newly proposed methods. Data-driven process monitoring methods can be divided into two categories, multivariate statistics methods and novel machine learning or artificial intelligence methods [12,27,35].

Multivariate statistics methods are usually adopted by projecting high-dimensional data into a low-dimensional space, where most of the original data information can be retained. The range of normal operating conditions can be determined in this low-dimensional feature space for fault detection. In 1991, PCA and PLS were successfully applied to the process monitoring of two chemical units by Kresta et al. [36]. In PCA, original data are transformed into a low-dimensional linearly independent space and a residual space by linear projection. The T^2 statistic, the SPE statistic and their control limits are established in the corresponding space for fault detection. PCA has been widely developed in process monitoring as an effective feature extraction and dimensionality reduction tool. Multiscale PCA was proposed by Bakshi to monitor the wavelet coefficients at each scale by PCA [37]. Sun et al. proposed a fault detection method by capturing the interdependence of wavelet coefficients using hidden Markov models [38]. Due to the nature of the hidden Markov model, autocorrelation is also considered by hidden state transition. Kano et al. proposed moving window PCA and applied it to online process

monitoring by monitoring changes in the correlation structure of process variables [22]. Considering that significant serial correlation contained in industrial process data cannot be extracted by PCA and PLS, canonical variate analysis (CVA) was employed to generate accurate state-space models from serially correlated data [39–41]. In addition, industrial data usually do not conform to multivariate normal distribution, which will influence the determination of the control limits of the monitoring statistics in PCA and PLS. Different from PCA, ICA was applied to process monitoring by projecting the correlated variables into independent space with no orthogonality constraint, making it more applicable to non-Gaussian processes [42,43]. Another method to deal with non-Gaussian processes is the Gaussian mixture model (GMM), by which multiple Gaussian models are established to approximate non-Gaussian data. It has been widely applied for process monitoring in arbitrary datasets that do not conform to normal distribution [44–46]. For processes with non-Gaussianity, a fault detection method based on statistics pattern analysis (SPA) was further developed by He and Wang [47]. Unlike traditional multivariate statistical method such as PCA, SPA monitors the different statistics of process variables rather than variables themselves, then high-order statistics are applied to capture the process characteristics. The advantages of SPA have been demonstrated by a comprehensive evaluation of process monitoring performance using TEP [48]. For large datasets in industrial processes, it is difficult to interpret the feature corresponding to each principal component because the principal components are linear combinations of the original variables [49,50]. Sparse principal component analysis (SPCA) was proposed to solve this problem by restricting some of the loadings on principal components to zero, and therefore main parts of principal components can be displayed [49–51]. Various sparse-based methods have also been applied to dimension reduction and process monitoring, such as sparse PLS and sparse CVA [52,53]. Further improvements of traditional multivariate statistics methods, which have been proposed to deal with industrial process data, will be introduced later according to various industrial data characteristics.

With the rapid development of computer technology, various novel machine learning algorithms have been proposed and applied to process monitoring. Aldrich and Auret provided a comprehensive review on unsupervised machine learning-based process monitoring methods [54]. Venkatasubramanian reviewed the application of AI in chemical engineering and classified it into different phases [55]. It can be concluded that AI-based data analysis methods play an increasingly important role in chemical processes as well as process monitoring. Early application of AI-based process monitoring methods was mostly based on process knowledge, such as the expert systems [55], which is an intelligent computer program system that contains a large amount of expert experience and simulates the decision-making process of human experts for reasoning and judgment. With the fast development of computer technology and machine learning, data-driven approaches have become the basic assumption of modern AI. The success of the neural network method has brought the development of artificial intelligence into a new era. Convolutional neural networks have been widely applied for fault classification and diagnosis [56,57]. These classification models require a large number of labeled historical fault data, while fault data are usually unavailable in industrial processes and the fault type that occurs in real time production may not be contained in a historical database. In order to perform fault detection, an autoencoder (AE) was applied by constructing the structure of the neural network to make the output dimension equal to the input [58,59]. Another type of AI-based process monitoring methods are derived from the well-known support vector machine (SVM), which is a powerful binary classification model which divides different types of samples through a hyperplane [60]. For fault detection, one class SVM model can be built to obtain a hypersphere boundary around training data, and the optimization goal is to minimize the volume of the hypersphere while including the most normal training data [61,62]. Different from neural network- and kernel-based methods, tree-based methods are easy to implement and explain. A decision tree can be used to solve regression and classification issues by recursively selecting the optimal feature, and labelling the training data according

to the feature [63,64]. To improve the generalization ability of the decision tree, the random forest was proposed to integrate multiple decision trees into one model and use majority voting to make classification decisions [65]. The random forest has been applied for fault classification in TEP and shows a better classification performance than SVM and neural network-based methods [66]. The random forest can also be applied as an unsupervised anomaly detection method, a single class can be constructed from normal training data, and the distance from a previously defined normal class can be used as a similarity measurement [67]. However, the random forest has not been widely used for a long time because it can be considered as a black box model, the internal operation of the model is difficult to control by trying different parameters and random seeds. Until the proposal of the isolation forest, which is an unsupervised outlier detection algorithm, the tree-based method began to attract more attention in the field of fault detection [68,69]. The isolation forest is composed by isolation trees, and the trees are constructed by binary division of the samples according to the sample value until the data are no longer divisible or the binary tree reaches a limited maximum depth. The isolation forest has a good performance in outlier detection, but it is not suitable for particularly high-dimensional data, because each binary division of the samples is implemented in one random dimension. Considering the importance of local data features represented by data neighborhood information, manifold learning is proposed to provide a novel perspective to preserve the local features of data [70]. The hypothesis of manifold learning is that some observed high-dimensional data are actually a low-dimensional manifold structure embedded in a high-dimensional Euclidean space, and the manifold learning aims at mapping it back to the low-dimensional space and revealing its essential characteristics. Recently, manifold learning-based methods have been applied to process monitoring in industrial processes [71]. More details of the application of AI-based process monitoring methods and their achievements and limitations will be discussed later according to industrial data characteristics.

3.1.1. Dynamic Feature Extraction for Process Monitoring

In industry processes, sequences of process variables generally reflect dynamic characteristics due to internal mechanisms, control response, etc. Aiming to address the time dependency of process data, dynamic PCA (DPCA) is proposed by Ku et al. to extract the autocorrelation of variables [72]. DPCA applies an augmented matrix to describe time-varying dynamic characteristics with the same PCA decomposition. The application of an augmented matrix can be found in many other methods to improve dynamic process monitoring. Li et al. proposed a partial DPCA method to enhance the isolation ability of dynamic process monitoring by structured residuals [73]. Russell et al. combined a residual-based CVA statistic and DPCA to achieve a better monitoring performance on TEP [74]. Li et al. proposed a DPCA-dynamic ICA-based process monitoring method to consider both non-Gaussian and dynamic characteristics [75]. However, there are limitations in these dynamic process monitoring methods using time lagged variables. The dimensions of the data matrix increase significantly as the lags increase, which causes huge computational loads and great difficulties in dimension reduction [76]. Another problem is that all process variables have to be expended to the same level, but the process dynamic has different effects on each variable [76,77]. For this purpose, dynamic latent variable (DLV) models were proposed for process monitoring [78]. DLV-based methods aim to separate the dynamic part and the static part of process variables. For the static part, static PCA can be directly applied for feature extraction, and for the dynamic part, autoregressive models can be applied to capture the relationships of dynamic latent variables. Li et al. proposed a DLV-based process monitoring method, by which auto-regressive PCA is first applied to extract dynamic principal factors, and then a vector autoregressive model is adopted for the latent variables to obtain residues for monitoring [79]. A dynamic-inner PCA algorithm was proposed by Dong and Qin by maximizing the covariance of dynamic latent variables, making them easier predict based on their past values [80]. More DLV-based methods and their application to dynamic process monitoring can be found in the review paper

by Zheng et al. [81]. Another type of method applied in dynamic process monitoring is slow feature analysis (SFA), which is a dimension reduction algorithm which extracts the slowly varying features of process data [82]. Shang et al. applied SFA to monitor CSTR and TEP, better results were obtained by comparing DPCA and DICA [83]. Huang et al. proposed an online feature which reorders and selects strategies to choose slow features, through which fault information is considered in feature selection to optimize process monitoring results [84]. Zheng and Zhao proposed a process monitoring strategy based on enhanced canonical variate analysis with slow feature, by which dynamic relationships can be fully extracted and separated from static information by extracting slow canonical variates [85]. To improve the performance of original SFA in extracting long-term time dependency, long-term dependency slow feature analysis was proposed by extending first-order autoregressive processes to higher-order autoregressive processes [86]. The comparison of the abovementioned dynamic process monitoring methods is shown in Table 4.

Table 4. Comparative analysis of dynamic process monitoring methods.

Method Types	Achievements	Limitations
Dynamic multivariate statistics methods	Easy to implement	Lag order is limited due to high computational loads; Process variables have to be expended to the same level; Poor interpretability
DLV-based methods	The inner dynamic relationships can be captured; Good interpretability	Unpredictable dynamic fluctuations cannot be effectively monitored; Some static information is omitted
SFA-based methods	Slow varying features can be captured; Process disturbance and faults can be simultaneously monitored [87]	Most long-term time dependency is not properly considered

3.1.2. Nonlinear Feature Extraction for Process Monitoring

For most of the multivariate statistical methods discussed above, feature extraction is implemented by analyzing linear correlations among process variables. However, the process nonlinearity contained in modern complex industrial processes cannot be ignored. The kernel function method has been well developed to deal with nonlinear problems by mapping the original data into a higher dimensional space [88]. With the well accepted assumption that the linear inseparable problem can be solved by mapping it to a high-dimensional linearly separable space, the kernel function method is combined with various traditional process monitoring methods to realize nonlinear process monitoring. The idea of this kind of method is to increase and then reduce the dimension of a process for nonlinear feature extraction. Schölkopf et al. proposed a kernel principal component analysis (KPCA) to perform a nonlinear form of PCA with kernel function [89]. KPCA has been widely applied for nonlinear process monitoring, in which the original inputs were first mapped into a higher dimensional space with the kernel functions, and then traditional PCA procedures were applied for process monitoring [90–92]. The same form of data processing has been applied to many other algorithms. Kernel partial least squares were proposed to solve nonlinear prediction and process monitoring problems [93,94]. Kernel ICA and kernel CVA have also been successfully applied to monitor nonlinear processes by combining kernel functions with ICA and CVA [95–98]. Moreover, the kernel function is also applied for process monitoring in another machine learning method named SVDD, which is a one-class classification method derived from SVM [99]. After mapping the original data into a high dimensional space, SVDD aims to find a minimum hypersphere to capture all the normal training data [61]. The distance between samples and the center of the hypersphere is applied as a monitoring statistic to compare with the radius of the hypersphere. The SVDD can also be applied for the fusion of multiple statistics to give

final process monitoring decisions [76,77]. Although a certain degree of nonlinearity can be extracted by kernel function-based methods, there still remains a few difficulties in its applications on industrial process monitoring. As the dimension of original data is already high in complex industrial processes, the mapping of the kernel function will cause a huge computational load, making it inapplicable to large-scale industrial processes. On the other hand, the high dimension of data after mapping is not interpretable, and the results are mostly dependent on the selection of kernel function and parameters.

Recently, neural network-based methods have been developed to be a kind of unsupervised approach for nonlinear process monitoring, which is called autoencoder (AE) [59]. The AE is an artificial neural network that encodes original data into a low dimensional hidden layer and then decodes the hidden layer into the original dimension [100]. As an unsupervised feature extraction method, AE has received broader application prospects in process monitoring than other neural network-based classification algorithms. On the basis of AE, the stacked autoencoder (SAE) was proposed to increase generalization ability by using multilayer neural networks [101]. To further consider the removal of noise in the process, the denoising autoencoder (DAE) was proposed by Vincent et al. to extract the robust features of input data [102]. DAE has been applied to process monitoring as an improvement of basic AE for various processes [103–105]. More recently, deep learning technology has made great progress due to the development of computer science. A deep neural network has also been combined with AE and applied for anomaly detection [106], such as variational autoencoders (VAE), which is a powerful generative model proposed by Kingma and Welling [107]. The purpose of VAE is to build a model that can generate target X from latent variable distribution Z, which is assumed to follow the normal distribution. Therefore, regardless of the distribution of the original data, the features extracted by VAE follow the normal distribution, making it suitable for monitoring both nonlinear and non-Gaussian processes [108,109]. On the basis of VAE, Cheng et al. constructed a variational recurrent autoencoder (VRAE) to consider the process dynamic through the utilization of a recurrent neural network [27]. To improve the interpretability of deep learning-based process monitoring, Bi and Zhao proposed a novel orthogonal self-attentive variational autoencoder (OSAVA) model that can provide interpretable identification of abnormal variables while implementing process monitoring [110]. The AE-based process monitoring methods demonstrate good performance when extracting almost any kind of nonlinear features in the process. However, the features extracted by AE are difficult to interpret because it is a neural network, which is a black box model with weak generalizability. The performance is mostly dependent on the network structure and the number of hidden layer units, and the model needs to be retrained when the operating condition is adjusted.

Another important research area of nonlinear dimensionality reduction is manifold learning. Unlike traditional statistic process monitoring methods that only focus on the global features of the data, manifold learning provides a novel perspective to preserve the local features of data. Assuming that data are evenly sampled in a low-dimensional manifold structure that is embedded in a high-dimensional Euclidean space, the purpose of manifold learning is to map it back to a low-dimensional space and calculate the corresponding embedding mapping [70]. Based on manifold learning, several algorithms have been proposed for dimensionality reduction. Roweis and Saul proposed a locally linear embedding (LLE) algorithm, which assumes that the manifold is locally linear, and therefore each point can be represented by the optimal linear combination of its nearest neighbors by least squares. The coefficients of linear fitting are regarded as the description of the local geometric properties of the manifold in LLE [70]. The method of LLE can be regarded as the beginning of local feature analysis and related theoretical research. Belkin and Niyogi proposed a Laplacian eigenmaps (LE) algorithm, which uses a undirected weighted graph to describe a manifold [111]. The graph formed by the manifold data is mapped from the high-dimensional space to the low-dimensional space under the premise of maintaining the local adjacency of the graph. The above methods are limited to defining the embedding mapping of the training dataset, in order to find the corresponding position

of the new test data point in the subspace after dimensionality reduction, the locality preserving projection (LPP) is proposed by finding the optimal linear approximations to preserve the neighborhood structure of the dataset on the manifold [112]. The LPP has been introduced to the field of process monitoring. Shao et al. proposed a generalized orthogonal locality preserving projection for nonlinear fault detection by imposing orthogonality constraints on LPP, which can preserve the global shape of distribution and is desirable for fault detection [113]. He and Xu combined SPA and LPP to propose a statistics locality preserving projection (SLPP) to extract the local and non-Gaussian features simultaneously for fault detection [114]. Although local features represented by neighborhood information can be extracted, the global features expressed by variance information is ignored in the above methods. By considering the extraction of both global and local features, Zhang et al. proposed a global-local structure analysis model (GLSA) and applied it to process monitoring [115]. In GLSA, the objective functions of PCA and LPP are integrated, and a tuning parameter is introduced to balance the corresponding structure preserving [115]. Yu proposed a local and global PCA (LGPCA) by constructing an objective function based on the ratio of LPP to PCA [116]. To reveal the intrinsic relationships with PCA and LPP, Luo proposed a process monitoring method named global−local preserving projections (GLPP) to extract global and local features based on the distance relationship between neighbors and non-neighbors [71]. GLPP is a brand-new data projection algorithm, and under certain conditions, GLPP is consistent with the normal PCA and LPP. The method has promising application prospects in industry because there is no assumption of data distribution. Therefore, GLPP has been extended to solve the problems of nonlinearity, sparsity, and ensemble learning [117–119]. More recently, Li et al. proposed a global-local marginal discriminant preserving projection method, by which GLPP was integrated with multiple marginal fisher analysis to extract the discriminative feature of historical fault data. This method can better separate normal data from fault data under the circumstance that the fault data is available [120]. The manifold learning-based methods have an ability to describe nonlinear features through linear approximation projection. Combined with the traditional global structure preservation projection method, both the global and local data features can be extracted for process monitoring. The analysis and interpretation of the extracted features are easier than other nonlinear process monitoring methods. However, the computational complexity of manifold learning is high, which hinders its practical application. In addition, little research has been reported to systematically study the determination of the size of the projected dimension and the neighborhood in the field of process monitoring. The achievements and limitations of the abovementioned nonlinear feature extraction methods are shown in Table 5.

Table 5. Comparative analysis of nonlinear process monitoring methods.

Method Types	Achievements	Limitations
Kernel function-based methods	The extraction of nonlinear features is easy to implement.	Poor interpretability, the results are mostly dependent on the selection of kernel function and parameters; The mapping of kernel function results in a high data dimension in industrial processes.
Neural network-based methods	Almost any kind of nonlinear features can be effectively extracted.	The features extracted are hard to interpret; The selection of the network structure and the number of hidden layer units is difficult; The generalizability of the model is limited
Manifold learning-based methods	Nonlinear features can be extracted by linear approximation projection; Local feature represented by neighborhood information is considered; Good interpretability	The computational complexity is high; The size of the projected dimension and the neighborhood are difficult to determine.

3.1.3. Nonstationary Process Monitoring

In addition to dynamic characteristics, certain process variables display nonstationary trends due to various other factors, such as equipment aging, catalyst consumption, and different operating modes [121]. The means and variances of process measurements are time-varying, which challenges the assumption that the measured processes of traditional multivariate statistical methods are stationary [122]. Therefore, massive false alarms can be caused by nonstationary trends, and on the other hand, a faulty signal could be buried in the nonstationary trends of process variables at its early stage, leading to a long fault detection time [123]. The simplest way to deal with nonstationary series is differential processing. Most nonstationary series will become stationary after one or two differences. Based on the differential processing, the autoregressive integrated moving average model (ARIMA) has been proposed and widely used in time series analysis [124]. The idea of ARIMA is to process the original nonstationary time series into a stationary series through differential processing with an appropriate order, and then a well-developed autoregressive model can be used to fit the process. The ARIMA has a good performance on time series prediction, Zhou and Huang applied ARIMA to predict the remaining life of a lithium-ion battery [125]. As the coronavirus disease, 2019 (COVID-19) has received widespread attention from scientists around the world, Duan and Zhang modeled and predicted the spread of COVID-19 based on epidemic data from Japan and South Korea [126]. The ARIMA has also been applied to the field of process monitoring by Mu et al., a quality fault diagnosis model of TEP was established by predicting the quality variables. The model could be used to describe the change of quality to produce quick fault detections and warnings [127]. However, useful information from the data could also be removed by the differential preprocessing operation, although the nonstationary trends in the original series can be eliminated through differential processing with a sufficient order in any nonstationary series. In particular, small changes at the early stage of faults cannot be revealed, resulting in a longer fault detection time.

Another approach which addresses the difficulties of nonstationary process monitoring are methods based on recursive adaptation. Slow time-varying statistical characteristics, such as means and variances, can be captured by incorporating the concept of recursive adaptation within a moving window. Qin et al. proposed two recursive PCA algorithms to adapt normal process changes by recursively calculating the number of principal components and control limits for process monitoring [128]. Wang et al. proposed a fast-moving window PCA to adapt the nonstationary trends by adjusting the PCA model. A high computation speed and the ability to detect slow changing faults can be realized by applying the PCA model, which is N-steps earlier identified, to analyze the current observation [129]. Wang and Jiang proposed an online model updating criterion on the basis of canonical correlation analysis for adaptive monitoring of slowly varying process [130]. Several other recursion-based methods have also been proposed to combine various traditional process monitoring methods, such as PLS and SFA [131–133]. However, the application of moving windows in most recursion-based methods will reduce the sensitivity of early fault identification, and the selection of the length of the moving window has a great effect on the process monitoring results. If the size of window is too small, the abnormal behavior may be adapted by the model, and if the size of window is too large, the fault cannot be identified at an early stage.

Recently, cointegration theory has been developed as a more promising method to deal with nonstationary processes without the requirements of differential preprocessing or model updates. Cointegration is a concept which describes the relationship among process variables [134]. For a nonstationary process, certain variables are time-varying, but there may be a long-term stable time-invariant relationship among these nonstationary variables, which is called a cointegration relationship [134]. Considering that these time-invariant features can be extracted as a stationary monitoring index, the cointegration theory has been introduced to industrial process monitoring [135]. Li et al. applied a non-stationarity test to distinguish nonstationary series from stationary series in TEP and establish the

cointegration relationship for nonstationary series. The captured common trends and equilibrium error were used to construct monitoring indices. The results showed a better fault detection performance than the method based on differential processing [122]. Yu et al. proposed a monitoring strategy based on recursive cointegration analysis, which can effectively distinguish between normal changes in cointegration relationships and real abnormal deviations [136]. The method combined cointegration theory and the idea of recursion; the model could be adaptively updated when the cointegration relationships among variables change under normal operating conditions. In order to take both stationary variables and nonstationary variables into consideration, multiple order moments obtained from cointegration analysis have been proposed as a new monitoring index by Wen et al. to capture the statistical features of stationary datasets [137]. Zhao and Huang combined the advantages of cointegration theory and SFA, and proposed a full-condition monitoring method for dynamic nonstationary processes [138]. The nonstationary variables were first separated from stationary variables, and the cointegration theory was applied to extract the long-term equilibrium relations, while SFA was applied to extract the static and dynamic characteristics. The ability to extract and distinguish different latent features make it promising for monitoring complex chemical processes. Considering that cointegration theory mainly focuses on changes occurring in relationships among nonstationary variables, Lin et al. proposed a nonstationary process monitoring method based on cointegration and common trends analysis that could extract both stationary and nonstationary trends, and was sensitive to any fault condition [139,140]. However, cointegration theory can be considered a new method of industrial process monitoring, and its application is still relatively limited. The case studies reported in literature are mostly implemented on numerical simulation cases or on one specific industrial operation unit, few cointegration theory-based methods have even been successfully applied to monitor all the faults in TEP.

It can be concluded that the nonstationary characteristics existing in industrial process data have not yet received enough attention. According to the index result in the database, the methods proposed to solve nonstationary problems can only be sporadically found. So far, there are few methods which can effectively solve the problem of nonstationary process monitoring. A summary of the existing nonstationary process monitoring methods is provided in Table 6.

Table 6. Comparative analysis of nonstationary process monitoring methods.

Method Types	Achievements	Limitations
Differential processing-based methods	Easy to implement; Any series can be processed to stationary	Useful information from data could be removed together
Recursive adaptation methods	The model can be updated to adapt the slow time-varying statistical characteristics	The frequent updates from the model cause huge computational loads; Not sensitive to small changes at the early stage of faults
Cointegration theory-based methods	Long-term equilibrium relations among nonstationary variables are captured and applied as monitoring indices; No differential processing is required and the model does not need to be updated frequently	Only changes in the relationships among nonstationary variables can be monitored by cointegration-based methods; Only the linear relationship among variables is concerned in cointegration analysis

3.2. Multimode Process Monitoring

Multimode characteristics in industrial production have always been a challenge for process monitoring. Although advanced models have been proposed and have achieved satisfactory monitoring results during single-mode processes, a large number of false alarms will be generated by models based on a single mode in industrial application with multiple operating modes. As more and more statistical methods and machine learning methods are successfully applied to process monitoring, research to overcome the multimode issue

has drawn increasing attention in recent years. Generally, data-driven monitoring methods for multimode processes can be categorized into single model methods and multi-model methods through which a monitoring method is built for each mode [17,18,141]. Current research progress and a summary of the achievements and limitations of these multimode process monitoring methods is provided in the following section.

3.2.1. Multi-Model Methods

The main idea of multi-model methods in process monitoring is to establish multiple local monitoring models for each operating mode. Zhao et al. proposed a multiple PCA method by establishing multiple PCA models for each mode. For online monitoring, the real time data are brought into each model and the minimum value of SPE is selected for comparison with its control limit for fault detection [142]. A similar idea has also been combined with PLS to propose the multiple PLS method [143]. However, minimum SPE may mislead the state to a mode it does not belong to. To deal with this issue, various effective clustering methods were introduced to label historical data into different modes and applied for mode identification to determine which mode the new observation belongs to. Once a new observation was grouped to a mode, the corresponding model established under this mode was applied for monitoring. In the work of Yoo et al., the cluster method using a posterior probability was applied for mode identification [144]. Tan et al. applied the similarity of data characteristics in a moving window to realize mode identification for multimode monitoring [145]. Ha et al. combined a k-nearest neighbor (KNN) algorithm with multiple-model methods, in which the new observations are first clustered into a mode by KNN, and then the corresponding PCA model is applied for monitoring [146]. Tong et al. proposed an adaptive strategy to integrate mode clustering and mode unfolding for multimode process monitoring [147]. Mixture modeling techniques can also be applied in multi-modal methods when the mode of new observations cannot be identified [148]. Yu and Qin proposed a finite GMM and Bayesian inference-based process monitoring approach to integrate each operating condition with probability [45]. Zhang et al. applied a process monitoring method which combined multi-model methods and GMM, in which different models built for each mode were probabilistically united by GMM [148]. Xie and Shi proposed a moving window GMM to consider both the multimode characteristics and dynamic characteristics [149]. Cao et al. established a variational Bayesian PCA for each GMM to model multimodality, and KLD between a mixture of Gaussians were used as statistics, it was found that the sensitivity to small faults increased [150]. Multimode methods can provide a more effective feature extraction in each local mode. However, all modes have to be included in the training data, otherwise the model will not function properly if the operating condition shifts to a new mode. Even if a new mode has been identified, it still takes time and enough data to build a model for the new mode, which is not realistic in real time monitoring. Moreover, false alarms will be generated in transition states because data characteristics between two modes are usually not considered in multimode methods.

3.2.2. Single Model Methods

Single model methods are established based on the similar characteristics of all modes, and therefore applicable for each one. The super PCA proposed by Hwang and Han can be considered as the first single model method for multimode process monitoring, by which hierarchical clustering is applied to identify the common characteristics from the hierarchical relations between clusters and a more generalized PCA model is constructed than multiple local PCA models [17]. However, the method is not applicable to nonlinear processes [17]. Han et al. found that the loading matrix of the PCA models built in each mode shows a high similarity, and therefore a single PCA model was established for all modes and a condition recognizer was applied using a moving window to adjust the normalization center of the PCA model [151]. However, faults that occur between two stable modes cannot be identified, and the normalization center of the PCA will be adjusted

incorrectly, resulting in a model failure. Another type of single model method is to scale multimode data to a single distribution by data preprocessing. Ma et al. proposed a local neighborhood standardization (LNS) strategy for multimode process monitoring [152]. Different from the traditional z-score method, the samples are standardized by LNS with mean and standard deviation of their local neighborhoods. Therefore, data from each mode can be standardized to a similar level and the traditional PCA can be directly applied for monitoring. Based on the idea of local standardization, Wang et al. proposed an Euclidean distance-based weighted k neighborhood standardization and applied it for multimode monitoring with PCA [153]. Considering the process nonlinearity, Wu and Zhao proposed a multimode monitoring method combining local adaptive standardization (LAS) with variational auto-encoder [141]. The process monitoring methods based on data standardization with a local neighborhood can be generalized to monitor new operating conditions, which are not included in training data, but the shortcoming is that the methods are not applicable to transitional modes. The achievements and limitations of the multi-modal process monitoring methods are summarized in Table 7.

Table 7. Comparative analysis of multimode process monitoring methods.

Method Types	Achievements	Limitations
Multi-model methods with clustering	The operating mode can be identified; An independent model can be established for each mode; The transitional modes could be considered	The real time operating conditions have to be included in training data; The results rely on the performance of mode identification
Multi-model methods with hybrid modeling	The operating mode does not need to be identified; An independent model can be established for each mode; The transitional modes could be considered	The real time operating conditions have to be included in training data; The operating mode cannot be identified
Single model methods based on data standardization with local neighborhood	Low model complexity and computational loads; Applicable to new operating conditions, which are not included in training data	The window size applied for data standardization is difficult to determine; The transitional modes are not considered

The monitoring of multiple steady modes has been well developed in multimode process monitoring, but as the most challenging part for their obvious time-varying changes, transitional modes have only been studied sporadically [154–158]. Existing approaches towards transitional mode monitoring are mostly dependent on a large number of if-then strategies. Considering that the fault occurring in transition modes will cause more serious consequences, it is of great significance to develop useful methods for the simultaneous monitoring of transition modes and multiple steady modes in future work.

3.3. Batch Process Monitoring

As an important production mode in the chemical industry, batch process methods has been widely used in many industrial productions with a flexible capacity and a high value-added product, including biochemical and pharmaceutical industries. Compared with continuous production, batch processes usually have more complicated characteristics, which provide great challenges to batch process monitoring. On the one hand, data collected from batch process methods is three-dimensional, which has an extra dimension of batch. Traditional continuous process monitoring methods cannot be directly applied because they only fit for a two-dimensional data structure. Moreover, the characteristics and the length of the data are different due to differences in initial conditions and ambient conditions between batches. On the other hand, multiple stages exist within a single batch,

and set points of variables are constantly changing at each stage, making normal operating conditions in batch process a time-varying trajectory rather than a stable operating point. Research on monitoring issues for batch processes began to attract attention in the field of process monitoring since the multiway PCA method was proposed by Nomikos and MacGregor in 1990s [159]. The development of batch process monitoring methods was highly promoted in 2000s by the proposal of the simulation platform for the penicillin fermentation process named Pensim [160], which has developed into a batch process benchmark to evaluate the performance of newly proposed process monitoring methods. Based on this benchmark process, many other methods have been proposed to deal with batch process monitoring. A review of batch process monitoring methods is provided in the following section.

3.3.1. Multiway Methods

The multiway method is a classic way to deal with batch process data by expanding three-dimensional data into two-dimensional data, and therefore traditional multivariate statistics methods can be applied for process monitoring. Nomikos and MacGregor performed batch-wise unfolding on three-dimensional batch data and then applied PCA to normalized data for online monitoring [159]. A multiway PLS method was also proposed by introducing the similar batch processing operation to PLS [161]. The nonlinear and dynamic characteristics in the operation trajectory of each variable within each batch will be eliminated through data normalization, and the variations between batches will be emphasized. Word et al. proposed another variable-wise unfolding method, by which the variations of each variable in batch and time intervals are emphasized [162]. Based on data unfolding, different multiway batch process monitoring methods considering nonlinear and dynamic characteristics are introduced from continuous process monitoring. Chen and Liu introduced the idea of dynamic PCA and PLS in continuous processes to batch processes and proposed dynamic multiway PCA and PLS [163]. Choi et al. proposed a dynamic multiway method based on an autoregressive model and multiway PCA, by which PCA was applied to monitor the residues of expended data calculated from an autoregressive model [164]. Lee et al. proposed a multiway KPCA method to capture nonlinear characteristics by combining data unfolding with KPCA [165]. Zhang and Qin proposed a multiway kernel ICA method to extract dominant independent components in batch processes [166]. More methods introduced previously in continuous process monitoring could be found to be combined with multiway methods, such as AE [167], kernel SFA [168], and manifold learning [168,169].

Multiway methods are easy to implement and can be combined with other methods to deal with complex data characteristics in batch processes by unfolding data into two dimensions. However, there are certain problems with the two data expansion methods according to van Sprang et al.'s critical evaluation [170]. The limitation of batch-wise unfolding is that all batches have to be processed to the same length, and when implementing online monitoring, the future behavior of the new batch has to be filled to the same length [159,170]. A certain error will be introduced no matter which data filling-in procedure is employed. Moreover, the internal process dynamic, variable correlation and time-varying operational trajectory are mostly ignored by batch-wise unfolding since data in a whole batch are expended to a single sequence, which affects the extraction of essential data characteristics to a certain extent. Different from batch-wise unfolding, there is no requirement for the length of each batch in variable-wise unfolding and more samples for modeling can be provided by the variable-wise unfolding, making it a more widely used data expansion method for batch process monitoring. However, the operation trajectory information within each batch still cannot be effectively extracted by just expanding each variable of different batches into a time series in chronological order. On the other hand, operational data show a multiple phase characteristic in most batch processes, and data collected from each phase may have different variable correlations and operation trajec-

tories [171], which cannot be fully captured by a single monitoring method. Therefore, multiphase modeling methods have been developed to solve this problem.

3.3.2. Multiphase Modeling

A large number of multiphase modeling methods have been proposed by establishing different models for each phase. Generally, a proper phase division is an important prerequisite for determining the effectiveness of multiphase modeling [172]. Several phased identification methods have been proposed for phase division. Nomikos and MacGregor suggested that an indicator variable can be applied to indicate the progress of batch process instead of timeline [173]. Undey and Cinar clarified the requirements of indicator variables, and combined a pharmaceutical monitoring case to illustrate its advantages in stage division [171]. Zhao et al. applied an indicator variable to divide batch data into multiple phases, and process variations within each phase which are monitored using sliding windows [174]. To automatically model multiphase batch process, Lu et al. proposed a sub-PCA monitoring strategy by building sub-PCA models for each divided phase based on process correlation structure changes [175]. Considering non-Gaussian and nonlinear data behaviors, Ge et al. proposed a sub-SVDD method, and obtained a better performance than sub-PCA [176]. Camacho and Picó proposed an automatic phase identification algorithm to model multiphase batch processes [177,178]. Based on a top-down recursive strategy and a bottom-up procedure, the phase can be reasonably divided corresponding to the best PCA model predictions. After phase division, multiphase PCA (MPPCA) models can be built for each phase. To reveal the changes in the nonlinear behaviors of different phases, Zhao et al. proposed a phase-based kernel ICA-PCA method to map unfolded data into a high dimensional space before modeling [179]. Assuming that the unfolded process data from each phase can be approximated as a multivariate Gaussian distribution, GMM was applied for phase identification in multiphase batch monitoring [46,180,181]. However, the transition issues between different phases were not considered in these methods. While the switching between different phases is a gradual process, this transitional state also needs to be modeled. Yao and Gao proposed a phase identification and transition identification method by measuring the angles between different PCA score spaces, allowing different models to be established for steady phases and transition phases separately [182]. Sun et al. proposed a multiphase MPCA (MPMPCA) method, by which a more objective phase identification was implemented based on the change of cumulative percent variance for the first principal component between different PCAs, allowing the phase transition to be automatically identified simultaneously [183]. Liu et al. proposed a sequential local-based GMM for multiphase batch process monitoring, in which the transition issues were considered by an adaptive iteration strategy [184].

The current research on batch process monitoring has achieved a certain level of application progress, but most of the efforts of existing solutions are still mostly focused on data preprocessing, such as multiway data unfolding and multiphase identification. Within these data preprocessing procedures, batch-wise data expansion, variable-wise data expansion and data filling are inevitably required, which will affect the extraction of the essential characteristic information from the batch process data under normal operating conditions, and further affect the performance of online monitoring for the batch process. An evaluation of batch processing methods previously mentioned is listed in Table 8.

Table 8. Comparative analysis of batch processing methods.

Method Types	Achievements	Limitations
Multiway methods with batch-wise unfolding	Easy to calculate; The average operation trajectory can be extracted; Variations between different batches can be emphasized	Internal correlation among variables is eliminated; The length of data in different batches and different phases of a batch have to be processed to the same level; Data must be filled to an entire batch in online monitoring
Multiway methods with variable-wise unfolding	Easy to calculate; No requirement for the length of data; More samples can be retained for modeling; Internal correlation among process variables can be retained	Correlations and differences between different batches cannot be reflected; Nonlinear and time-varying characteristics in multiphase batch processes cannot be removed, resulting in poor monitoring performance
Phase division by indicator variable	Low computational complexity; Each phase can be effectively identified when certain process features are available	Prior knowledge is required; Changes in variable correlation are not considered for phase division; The phase division may be unreliable for large-scale industrial processes
Phase division by data analysis	No prior knowledge is required; Phase division is more accurate by considering process correlation in industrial processes	Computational complexity can be overwhelming; Relatively less consideration of transition issues

3.4. Root Cause Diagnosis

As an important part of process monitoring technology, fault diagnosis aims to identify the root cause of the abnormal deviations after a fault is detected. Generally, the fault is usually caused by the abnormal deviation of a single variable, while variables interact with each other under the influence of the equipment, stream loop, and control response, the fault will quickly propagate to other variables, which brings a huge confusion to the root cause diagnosis. Sighed directed graphs (SDG) are commonly used for fault propagation analysis by modelling the cause and effect behavior of a process [185]. The cause and effect information among process variables is represented in the form of signed digraphs to trace back to the root cause. Vedam and Venkatasubramanian combined PCA with SDG to implement FDD on an industrial fluidized catalytic cracking unit [186]. Wang et al. applied SDG as an auxiliary tool for hazard and operability analysis to identify the cause and effect of process deviations [187]. A challenge for traditional SDG is that the modelling is based on expert system and process knowledge, but variable correlation would not remain identical, and could be very different under certain industrial operation conditions. Therefore, data-driven fault diagnosis methods are developed based on historical and real time data.

One type of data-driven fault diagnosis methods is to classify the real time fault into one of the fault types defined in historical data by various classification algorithms. However, such methods are limited to academia because historical fault data are difficult to obtain in industrial processes and the unexpected abnormal disturbances which appear in real-time operations may not be included in historical data, resulting in unreliable fault diagnosis results. How to establish a real time fault diagnosis model to provide correct fault diagnosis results for unknown faults with limited labeled data is an important research focus. The simplest way is to employ a single variable fault diagnosis method, for example, each measurement variable will be preset to an operating range in DCS, when a fault occurs, the variables that exceed the defined normal range are considered as the root cause. Gao et al. applied Euclidean distance between the normal state and real time state to determine

whether the variable is normal or not [57]. Ji et al. proposed a fault diagnosis method, by which the variations in information entropy of each variable sequence was applied to identify variables with abnormal deviations [188]. Yan and Yao introduced the concept of the least absolute shrinkage and selection operator (LASSO) for the identification of faulty variables [189]. However, there are a large number of measurements in industrial processes, and process variables are highly correlated with each other, when a fault is detected, usually multiple variables have deviated from their normal range, and the root cause cannot be further diagnosed. Therefore, the superiority of the multivariable fault diagnosis method is reflected, various data-driven fault diagnosis methods have been proposed based on multivariate statistical analysis, such as contribution plots-based methods, and causal analysis, such as probability-based methods and causal reasoning-based methods. A review of these fault diagnosis methods is presented below.

3.4.1. Contribution Plots-Based Methods

Contribution plots are commonly used fault diagnosis methods based on the contribution of each variable to the occurring fault [190], which are usually combined with multivariate statistical methods, such as PCA [191] and CVA [192]. Contribution plots on the T^2 statistic and the SPE statistic are the most common methods to measure the contribution of each variable to the deviation in the principal component space and the residual space, respectively [12]. The contribution of each variable can be plotted in a histogram for visualization, and the variable with the highest contribution is considered as the root cause of the fault. Considering the challenge in contribution plots, it is difficult to determine whether a variable is related to the fault because there are no control limits in contribution plots. Conlin and Martin proposed confidence limits for contribution plots by estimating the standard deviation of the contributions to the scores to solve this problem [193]. Westerhuis et al. introduced contribution plots to batch processes, and calculated the control limits in the same way as the SPE statistic confidence limit [194]. However, the fault is quickly propagated to other variables due to the high correlation among process variables; the variable with the highest contribution may not be the root cause, which is defined as the smearing effect by Westerhuis et al. [194]. Alcala and Qin proposed reconstruction-based contribution plots to deal with this problem, by which the amount of reconstruction of a monitoring statistic along a variable direction was regarded as the contribution of the variable to the reconstructed statistic [195]. In this way, variables with significant contributions can be guaranteed to be faulty variables rather than noisy effects. To extend contribution plots to nonlinear cases, Alcala and Qin further constructed reconstruction-based contribution plots for KPCA with a weighted combination of the T^2 and SPE statistic [92]. Peng et al. constructed contribution plots for a total kernel PLS model to isolate faulty variables in a nonlinear industrial process [196]. More methods based on reconstruction-based contribution plots have been reported to solve various industrial fault diagnosis issues, such as sparse contribution plots [197], chi-square contribution [198], weighted reconstruction-based contribution plots [199] and a reconstruction-based fault diagnosis strategy for multiphase batch processes with limited batches [200].

The contribution plots are easy to calculate, and the faulty variables can be effectively identified by reconstruction-based contribution plots. However, the smearing effect still exists even if the fault sample is only in a single sensor direction because the fault propagation among faulty variables is not diagnosed [195]. To solve this problem, several methods haven been proposed to capture the causal logic among the process variables, by which the fault propagation among the faulty variables can be further arranged in a network diagram, in order to diagnose the root cause of the fault.

3.4.2. Probability Reasoning-Based Methods

Probability reasoning is a common type of method which represents the direction of information transfer among process variables using conditional probability. The propagation path among faulty variables can be identified with the probability reasoning model estab-

lished from historical data. The Bayesian network is the most commonly used probability reasoning model for fault diagnosis, within which the variable correlation is preliminarily established by prior knowledge and process flow diagrams, and then conditional probability information between node states is determined from historical data including fault data to represent the transfer of faults between variable nodes. The Bayesian network has been widely applied for industrial fault diagnosis [201]. Amin et al. proposed a hybrid process monitoring and fault diagnosis method, by which the Bayesian network was applied to diagnose the fault propagation among faulty variables identified by contribution plots [202]. Gharahbagheri et al. adopted KPCA and the Bayesian network for process monitoring and fault diagnosis in TEP and a catalytic cracking unit [203]. Amin et al. proposed a dynamic Bayesian network model that could be updated with monitored process data, through which the fault could be detected and diagnosed precisely [204]. More Bayesian network-based fault diagnosis methods are covered in the review by Cai et al. [201].

Although the Bayesian network has been widely applied for process monitoring and root cause diagnosis combined with various fault detection methods, it has a significant limitation in that prior process knowledge and a large amount of historical fault data are usually not available to establish an accurate causal network in industrial processes.

3.4.3. Causal Reasoning-Based Methods

Considering that causal logic among process variables can also be described as time delayed correlation between each pair of variables, several correlation analysis methods have been introduced to identify the fault propagation. Bauer and Thornhill applied cross-correlation function to calculate the time delayed relationship between variables for fault propagation analysis on plant-wide industrial processes [205]. Yuan and Qin proposed a fault diagnosis method to reveal the cause and effect relationship between faulty variables using Granger causality, which is a data-driven causal reasoning method based on linear regression [206]. These correlation analysis methods are limited to linear correlation, while the correlation among process variables is highly nonlinear for most systems in industrial processes. It has been proved by Li et al. that the Granger causality is inadequate for the extraction of nonlinear relationships [207]. To deal with nonlinear issues, Chen et al. extended Granger causality to nonlinearity based on Gaussian process regression and proposed a fault diagnosis framework based on LASSO and nonlinear Granger causality [208]. Considering the superiority of information theory methods in extracting nonlinear correlations, Bauer et al. constructed the causal map of process variables by transfer entropy, which is adopted as a measurement of the direction of information transfer between variables. The causal map obtained from historical data was further applied for online analysis of fault propagation [209]. Hajihosseini et al. further applied transfer entropy to the fault isolation of different types of faults in TEP [210]. According to the characteristics of different types of faults, Li et al. proposed a novel fault diagnosis framework by dividing the faults into stationary faults and nonstationary faults, in which Granger causality analysis and transfer entropy were applied to isolate stationary faults and a dynamic time warping-based method was adopted for nonstationary faults [211]. Transfer entropy demonstrates outstanding performance in identifying the relationship between variables based on mutual information, but a significant shortcoming of transfer entropy is its high computational cost. To solve this problem, Lee et al. proposed a fault diagnosis method based on transfer entropy and graphical LASSO [212]. Redundant variables that are not related to the fault are eliminated by the graphical LASSO, thereby reducing the computational cost of transfer entropy. Considering the fault information in real time data, Ji et al. proposed a real time fault diagnosis method based on time delayed mutual information. In their work, both historical data under normal operating conditions and real time data were employed for fault propagation analysis, therefore, the changes in variable correlation when a fault occured could be captured to obtain a more objective root cause diagnosis [213]. More research on fault propagation analysis using several other methods can be found, such as k nearest neighbor for a nonlinear multi-input, single-output

process [214], and convergent cross-mapping for capturing nonlinear relationships with a low computational cost [215].

Causal logic among process variables can be effectively identified by data-driven causal reasoning methods with no prior knowledge or historical fault data. To date, most causal reasoning methods have been established to extract variable correlation data under normal operating conditions, but with the increasingly complex process topology, process variables are highly correlated and their correlation could change with the response of a large number of control loops, resulting in unreliable fault diagnosis results. It is expected to properly employ both historical data and real time data to analyze the variable correlation under multiple operating conditions and complex control loops to obtain more objective fault diagnosis results. If correct fault propagation analysis could be implemented by effective causal reasoning methods when a fault occurs, operators could easily identify the type and root cause of the fault and react quickly to the fault to avoid greater losses. A comparative analysis of the different types of fault diagnosis methods is presented in Table 9.

Table 9. Comparative analysis of fault diagnosis methods.

Method Types	Achievements	Limitations
Contribution plots-based methods	Faulty variables can be effectively identified; Easy to calculate; Easy to be combined with multivariate statistical methods	The fault propagation cannot be identified to obtain the root cause
Probability reasoning-based methods	The propagation path of faulty variables can be identified by establishing variable correlation with conditional probability; Can be easily combined with contribution plots from multivariate statistical method	Historical fault data are required, but usually not available; Fault diagnosis result of unknown faults may be unreliable; Prior knowledge is required
Causal reasoning-based methods	Causal logic can be effectively obtained by analyzing time delayed correlation between variables; No prior knowledge or historical fault data is required	Changes in variable correlation caused by control loops are difficult to identify

In this section, data-driven fault detection methods using multivariate statistics and machine learning are discussed in terms of nonlinear feature extraction, dynamic feature extraction, nonstationary feature extraction, multimode feature extraction, and batch feature extraction. It can be found that the definition of these characteristics and reasons for their formation are partially overlapped, for example, nonlinear characteristics exist in almost any chemical industrial processes, and sometimes dynamic characteristics can be regarded as a type of non-stationarity. Therefore, when dealing with industrial process monitoring, multiple data characteristics need to be simultaneously considered according to the definition of normal operating conditions from historical data.

4. Challenges and Prospects in the Application of Industrial Process Monitoring

As process monitoring technology developed for industrial application has attracted much more attention in recent years, great research progress has been made in industrial process monitoring targeting various complex characteristics of industrial process data. However, there are still several problems that have not been effectively solved yet. In this section, these challenges in industrial process monitoring are first presented. Possible solutions are then proposed accordingly, followed by the potential areas for future research.

4.1. Dynamic Process Monitoring with a Deep Consideration of Variable Autocorrelation

Process dynamics are a common type of behavior in industrial processes, certain variables show high autocorrelation due to internal mechanisms and the response of control systems. Most existing process monitoring methods focus on the extraction of variable correlation, but little research has been carried out which considers autocorrelation. Several dynamic process monitoring methods have been proposed by applying time-lagged process variables or establishing autoregressive models, but the challenge in dynamic process monitoring is that different variables reflect different degrees of autocorrelation, and variable autocorrelation and cross-correlation should be effectively integrated in a feature extraction method for nonlinear and non-Gaussian dynamic processes.

How to establish a process monitoring model that can comprehensively consider the different degrees of variable autocorrelation and complex cross-correlation among variables is still a problem to be solved. For the first challenge, it is promising to apply time-lagged variables with different lag orders for each variable since data of multiple sampling frequencies can be flexibly acquired from DCS. For the other challenge, the process variables could be divided into a dynamic part and a static part, and corresponding monitoring strategies could be carried out respectively. A traditional statistic process monitoring method can be applied to extract variable cross-correlation for the static part, and models on autoregression can be adopted to extract variable autocorrelation in the dynamic part. However, the complex dynamic characteristics reflected in dynamic variables cannot be effectively extracted solely by analyzing variable autocorrelation, as the dynamic characteristics can be regarded as a result of the interaction of both the static and the dynamic components. How to flexibly integrate proper static feature extraction methods and dynamic feature extraction methods to effectively decompose the dynamic and static components of dynamic variables should be further considered in future research.

4.2. Nonstationary Process Monitoring with Cointegration Theory

Compared to dynamic feature extraction, much less research has been carried out on nonstationary process monitoring because the data characteristics of nonstationary process variables are far more complicated. The means of nonstationary variables keep drifting with various internal factors, such as equipment aging and catalyst consumption, and external factors, such as random disturbances of environmental conditions. Once these variables drift out of the range during modeling, a large number of false alarms will be triggered. Therefore, the nonstationary trends should be carefully considered in process monitoring, but how to effectively extract nonstationary characteristics to remove these drifting trends is still a serious challenge in current research.

The simplest way, difference processing, can be applied to effectively process nonstationary series into stationary series. However, the small changes at the early stage of faults, which are buried by nonstationary trends, will be removed simultaneously, meaning that the fault cannot be identified early. Among the limited nonstationary process monitoring research, the cointegration theory can be regarded as the most promising method of processing nonstationary variables into stationary residues. However, its applications are only reported sporadically because there are still many challenges to generalize it to industrial applications. The current research on cointegration theory focuses on capturing the cointegration relationship among nonstationary variables, while variable relationships and data characteristics in stationary variables are mostly not considered. In this way, only the faults which occur in the nonstationary part of the process can be identified, but a large number of faults may occur in stationary variables during industrial production which cannot be detected, thereby, compromising the performance of the model. Moreover, cointegration relations among nonstationary variables are restricted to linear relations in current research, which are insufficient at characterizing the complex nonlinear and nonstationary relations in industrial processes. These challenges greatly limit the application of cointegration theory for nonstationary process monitoring. Optimistically, these challenges also provide a wide range of opportunities for future research. On one hand, multiple options for nonlin-

ear feature extraction algorithms can be integrated into cointegration theory to generalize cointegration relationships among nonstationary variables to nonlinearity. One the other hand, it is feasible to model stationary and nonstationary variables separately. Nonlinear and dynamic feature extraction methods can be established for stationary variables, and cointegration theory can be applied for nonstationary variables. Process monitoring can be implemented through respective monitoring statistics. Furthermore, the stationary residues obtained from cointegration theory can be integrated with original stationary variables, by which the process of non-stationarity is absolutely removed, and therefore corresponding stationary process monitoring methods can be directly adopted.

4.3. Transition Process Monitoring with the Idea of Residues Generation

The multimode characteristic caused by multiple operation modes is a typical problem in industrial process monitoring. Although the multiple steady states could be well monitored by many multi-model methods under the assumption that all possible modes involved in the process are known, new modes still occur inevitably in real time monitoring and incorrect monitoring results will be caused. Several methods applied to data standardization with local neighborhoods solve this problem, but the transition states were not considered. Since the production loads could be frequently adjusted in industrial production, the frequent occurrence of transition states has become the most severe challenge for multimode process monitoring. Different from steady states, the transition state is a constantly time-varying process, the common data features in transition states are difficult to capture. Moreover, the detection of faults that occur in transition states is even harder. Only a few existing multimode process monitoring methods take transition states into consideration based on a large number of if-then strategies. These monitoring strategies have relatively poor generalization, and most of them need to be developed case by case.

How to remove the time-varying characteristics of transition states plays an important role in multimode process monitoring. Considering the superiority of residual generation-based statistic process monitoring methods, which have the capability of transferring non-Gaussian distributed data into Gaussian distributed residuals [216], can be regarded as a promising solution for multimode process monitoring, including transition states. If the idea of residual generation can be transferred to multimode and transition process monitoring, the time-varying characteristics of data in transition states could be effectively predicted, and the obtained residues will conform to a steady Gaussian distribution, by which the problem is solved by transforming transition states into steady states. However, different from steady states, the transition states show an obvious time-varying dynamic and nonlinear characteristics, traditional soft sensor methods cannot be applied to predict transition states accurately. Regression of time-varying features requires strong fitting ability. It is promising to find an effective dynamic regression model for transition process monitoring in future research. Although modern machine learning methods, such as a few deep neural networks, have a powerful fitting ability, the interpretability of the model is another issue to be considered. The explanation of the machine learning models is also an important problem and may become a focus of future research. As models become more complex, their interpretability becomes more difficult. Although the fitting ability of the model is enhanced, the generalizability is still not sufficient, thus, when a new situation emerges the model may not function as expected. Moreover, the incomprehensibility of black-box models brings great challenges to human-computer interaction, making models unreliable in industrial applications. To deal with this problem, grey-box modeling is a promising research topic that has already attracted significant attention in related fields. For complex chemical processes, process knowledge from operators contains significant useful information on the process, even when there is a mismatch with the actual situation. Therefore, human knowledge of process mechanisms and in-depth mining of operating data may be integrated into machine learning models to continuously design and obtain more interpretable models for industrial applications.

4.4. Batch Process Monitoring with Operation Trajectory Prediction

Batch operation often corresponds to a high requirement for product quality, it is necessary to conduct real time monitoring to detect abnormal deviations from the predetermined production trajectory in time. On one hand, proper measures can be taken to eliminate the fault, on the other hand, if the product quality has been affected, the reaction can be terminated in time to save production time and utility. Compared with continuous process monitoring, the challenges of batch process monitoring are mainly reflected in two aspects, batch operation and multiphase operation. As mentioned in Section 3.4, many methods have been proposed for batch process monitoring based on data preprocessing, including batch expansion and phase division. However, the extraction of batch data characteristics will be affected no matter what kind of data expansion operation is selected. Another problem is that the normal operation within each batch is a time-varying trajectory, data within each phase still show complex dynamic and nonlinear characteristics even though the phases are effectively divided, especially the transition between different phases. The multiphase modeling for each phase still cannot be applied to effectively extract the complex time-varying characteristics in batch process.

A promising way to simultaneously solve problems of batch production and multiphase operation is to predict the operation trajectory of each batch. Although the production time may vary from batch to batch due to different initial conditions and ambient factors, and the data of each phase within a batch will show different characteristics, the variable autocorrelation and cross-correlation among process variables in each batch and each phase of a certain batch should always conform to their internal mechanism relationship, by which the operation trajectory is determined. If the internal mechanism information of the monitored process could be captured from data of multiple historical batch data, real time production trajectories of each batch under different initial conditions can be correctly predicted, and therefore the difficulties in batch process monitoring could be transformed into a continuous process monitoring issue, which avoids the data preprocessing operations of data expansion and phase division. To develop a dynamic and nonlinear method that is suitable for the prediction of operation trajectory in batch processes is a focus of future research.

4.5. Real Time Fault Diagnosis Based on Online Analysis of Fault Propagation

Fault diagnosis is an important and indispensable part of industrial process monitoring to isolate the root cause when a fault is detected. Due to the strong correlation among industrial process variables, the fault will be propagated from the root variable to many other variables when the fault is detected, making the real root cause of the fault difficult to diagnose.

It is of great significance to find the path of fault propagation by establishing the causal network among process variables. With the increasing scale of industrial processes, it is insufficient to apply process knowledge and expert experience to build an effective causal model that could be updated with an ever-changing operating condition, and the switch of control strategies. Although several data-driven methods have been adopted to capture the causal logic among process variables using historical data under normal operating conditions, the variable correlation could be significantly different under different control strategies [217]. Moreover, the response of control systems when a fault occurs will result in the changes in variable correlation, making the diagnosis results inconsistent with practical situations. The causal network established based on historical data under normal operating conditions is still insufficient for online diagnosis of a fault propagation path. However, few studies on the identification of fault propagation paths have been reported to take real time data into consideration when capturing the causal logic among process variables. Since the root cause of the fault must be the first variable, usually an inlet or manipulate variable that deviates from its normal operating conditions, the causal network obtained by identifying chronological order between variable deviations using effective AI methods can interpret the propagation path of the fault, and it can be reflected to the operators

to assist them in quickly identifying the type and the root cause of the fault, in order to quickly take measures to eliminate the fault and restore the system to stable operation. Therefore, how to effectively employ real time data to capture the causal logic among process variables or adaptively update the causal network established using historical data from normal operating conditions can be further investigated for the identification of the fault propagation path.

5. Conclusions

Data-driven process monitoring technology has been well developed in academia over last twenty years. With the recent popularity of the notion of industry 4.0 and the digital factory, multivariate process monitoring has begun to attract interest from industrial fields. The development of multivariate process monitoring technology aims to extract the objective correlation among available process measurements to support operators in their everyday operations. The development of multivariate process monitoring technology, especially AI-based methods, can provide fault information which is more consistent with real-world situations by extracting the internal characteristics of industrial data. Based on the more objective fault information from the machine, the operators could make early decisions to react to the fault with less errors. For the purpose of providing guidance on the implementation of industrial process monitoring, a comprehensive review of data-driven process monitoring was conducted from the perspective of extracting the various complex characteristics of industrial processes data. The procedures of data-driven process monitoring were first summarized to propose a general process monitoring framework for industrial application from the definition of normal operating conditions, data preprocessing, feature extraction, monitoring statistics, and fault diagnosis. Corresponding to the different types of normal operating conditions defined in this framework, current research progress relating to data-driven process monitoring methods for continuous stationary processes, nonstationary processes, multimode processes, batch processes, and fault diagnosis was detailed. Existing challenges in dealing with these problems were analyzed and promising methods for future research were proposed. Since useful methods have been proposed to solve the problems in monitoring the single mode process through the simulation process, future research will focus on the flexible application and improvement of these methods to overcome the challenges in plant-wide industrial production process caused by multiple operating conditions, as well as the incorporation of process knowledge into data-driven methods to improve the performance and interpretability of AI-based models. Finally, process monitoring technology is expected to be integrated into a mature fault detection, management and diagnosis system cooperating with DCS in general process industry, which has already become and will continue to be an important research area in PSE.

Due to the limitation of our literature scope, this review is only based on our understanding of this topic, and may not cover all works in this area, and we acknowledge that we could potentially interpret certain findings incorrectly. However, hopefully we have provided a new angle to discuss this topic and a reference for our peer researchers.

References

1. Jung, S.; Woo, J.; Kang, C. Analysis of severe industrial accidents caused by hazardous chemicals in South Korea from January 2008 to June 2018. *Saf. Sci.* **2020**, *124*, 104580. [CrossRef]
2. Zhao, L.; Qian, Y.; Hu, Q.-M.; Jiang, R.; Li, M.; Wang, X. An Analysis of Hazardous Chemical Accidents in China between 2006 and 2017. *Sustainability* **2018**, *10*, 2935. [CrossRef]
3. Chen, C.; Reniers, G. Chemical industry in China: The current status, safety problems, and pathways for future sustainable development. *Saf. Sci.* **2020**, *128*, 104741. [CrossRef]
4. Zhang, H.-D.; Zheng, X.-P. Characteristics of hazardous chemical accidents in China: A statistical investigation. *J. Loss Prev. Process Ind.* **2012**, *25*, 686–693. [CrossRef]
5. Ge, Z.; Song, Z.; Gao, F. Review of Recent Research on Data-Based Process Monitoring. *Ind. Eng. Chem. Res.* **2013**, *52*, 3543–3562. [CrossRef]
6. Modak, N.M.; Lobos, V.; Merigó, J.M.; Gabrys, B.; Lee, J.H. Forty years of computers & chemical engineering: A bibliometric analysis. *Comput. Chem. Eng.* **2020**, *141*, 106978. [CrossRef]
7. Pistikopoulos, E.N.; Barbosa-Povoa, A.; Lee, J.H.; Misener, R.; Mitsos, A.; Reklaitis, G.V.; Venkatasubramanian, V.; You, F.; Gani, R. Process systems engineering—The generation next? *Comput. Chem. Eng.* **2021**, *147*, 107252. [CrossRef]
8. Chiang, L.H.; Russell, E.L.; Braatz, R.D. *Fault Detection and Diagnosis in Industrial Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000.
9. Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S.N. A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311. [CrossRef]
10. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N. A review of process fault detection and diagnosis Part II: Qualitative models and search strategies. *Comput. Chem. Eng.* **2003**, *27*, 313–326. [CrossRef]
11. 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*, 327–346. [CrossRef]
12. Qin, S.J. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* **2012**, *36*, 220–234. [CrossRef]
13. Reis, M.; Gins, G. Industrial Process Monitoring in the Big Data/Industry 4.0 Era: From Detection, to Diagnosis, to Prognosis. *Processes* **2017**, *5*, 35. [CrossRef]
14. Severson, K.; Chaiwatanodom, P.; Braatz, R.D. Perspectives on process monitoring of industrial systems. *Annu. Rev. Control* **2016**, *42*, 190–200. [CrossRef]
15. Ge, Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 16–25. [CrossRef]
16. Qin, S.J.; Chiang, L.H. Advances and opportunities in machine learning for process data analytics. *Comput. Chem. Eng.* **2019**, *126*, 465–473. [CrossRef]
17. Hwang, D.-H.; Han, C. Real-time monitoring for a process with multiple operating modes. *Control Eng. Pract.* **1999**, *7*, 891–902. [CrossRef]
18. Quiñones-Grueiro, M.; Prieto-Moreno, A.; Verde, C.; Llanes-Santiago, O. Data-driven monitoring of multimode continuous processes: A review. *Chemom. Intell. Lab. Syst.* **2019**, *189*, 56–71. [CrossRef]
19. Kano, M.; Nagao, K.; Hasebe, S.; Hashimoto, I.; Ohno, H.; Strauss, R.; Bakshi, B.R. Comparison of statistical process monitoring methods: Application to the Eastman challenge problem. *Comput. Chem. Eng.* **2000**, *24*, 175–181. [CrossRef]
20. Tracy, N.D.; Young, J.C.; Mason, R.L. Multivariate Control Charts for Individual Observations. *J. Qual. Technol.* **2018**, *24*, 88–95. [CrossRef]
21. Jackson, J.E.; Mudholkar, G.S. Control Procedures for Residuals Associated With Principal Component Analysis. *Technometrics* **1979**, *21*, 341–349. [CrossRef]
22. Kano, M.; Hasebe, S.; Hashimoto, I.; Ohno, H. A new multivariate statistical process monitoring method using principal component analysis. *Comput. Chem. Eng.* **2001**, *25*, 1103–1113. [CrossRef]
23. Kano, M.; Nagao, K.; Hasebe, S.; Hashimoto, I.; Ohno, H.; Strauss, R.; Bakshi, B.R. Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. *Comput. Chem. Eng.* **2002**, *26*, 161–174. [CrossRef]
24. Kano, M.; Hasebe, S.; Hashimoto, I.; Ohno, H. Statistical process monitoring based on dissimilarity of process data. *AlChE J.* **2002**, *48*, 1231–1240. [CrossRef]
25. Harrou, F.; Sun, Y.; Madakyaru, M. Kullback-Leibler distance-based enhanced detection of incipient anomalies. *J. Loss Prev. Process Ind.* **2016**, *44*, 73–87. [CrossRef]
26. Zeng, J.; Kruger, U.; Geluk, J.; Wang, X.; Xie, L. Detecting abnormal situations using the Kullback–Leibler divergence. *Automatica* **2014**, *50*, 2777–2786. [CrossRef]
27. Cheng, F.; He, Q.P.; Zhao, J. A novel process monitoring approach based on variational recurrent autoencoder. *Comput. Chem. Eng.* **2019**, *129*, 106515. [CrossRef]
28. Ghosh, K.; Ng, Y.S.; Srinivasan, R. Evaluation of decision fusion strategies for effective collaboration among heterogeneous fault diagnostic methods. *Comput. Chem. Eng.* **2011**, *35*, 342–355. [CrossRef]
29. Raich, A.; Cinar, A. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AlChE J.* **1996**, *42*, 995–1009. [CrossRef]

30.	Yue, H.H.; Qin, S.J. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res.* **2001**, *40*, 4403–4414. [CrossRef]

31.	Choi, S.W.; Lee, I.-B. Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chem. Eng. Sci.* **2004**, *59*, 5897–5908. [CrossRef]

32.	Tong, C.; Lan, T.; Shi, X. Ensemble modified independent component analysis for enhanced non-Gaussian process monitoring. *Control Eng. Pract.* **2017**, *58*, 34–41. [CrossRef]

33.	Zeng, J.; Huang, W.; Wang, Z.; Liang, J. Mutual information-based sparse multiblock dissimilarity method for incipient fault detection and diagnosis in plant-wide process. *J. Process Control* **2019**, *83*, 63–76. [CrossRef]

34.	Downs, J.J.; Vogel, E.F. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, *17*, 245–255. [CrossRef]

35.	Qin, S.J. Data-driven Fault Detection and Diagnosis for Complex Industrial Processes. *IFAC Proc. Vol.* **2009**, *42*, 1115–1125. [CrossRef]

36.	Kresta, J.V.; MacGregor, J.F.; Marlin, T.E. Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Eng.* **1991**, *69*, 35–47. [CrossRef]

37.	Bakshi, B.R. Multiscale PCA with application to multivariate statistical process monitoring. *AlChE J.* **1998**, *44*, 1596–1610. [CrossRef]

38.	Sun, W.; Palazoğlu, A.; Romagnoli, J.A. Detecting abnormal process trends by wavelet-domain hidden Markov models. *AlChE J.* **2003**, *49*, 140–150. [CrossRef]

39.	Negiz, A.; Çinar, A. Statistical monitoring of multivariable dynamic processes with state-space models. *AlChE J.* **1997**, *43*, 2002–2020. [CrossRef]

40.	Jiang, B.; Zhu, X.; Huang, D.; Braatz, R.D. Canonical variate analysis-based monitoring of process correlation structure using causal feature representation. *J. Process Control* **2015**, *32*, 109–116. [CrossRef]

41.	Jiang, B.; Braatz, R.D. Fault detection of process correlation structure using canonical variate analysis-based correlation features. *J. Process Control* **2017**, *58*, 131–138. [CrossRef]

42.	Kano, M.; Tanaka, S.; Hasebe, S.; Hashimoto, I.; Ohno, H. Monitoring independent components for fault detection. *AlChE J.* **2003**, *49*, 969–976. [CrossRef]

43.	Lee, J.-M.; Yoo, C.; Lee, I.-B. Statistical process monitoring with independent component analysis. *J. Process Control* **2004**, *14*, 467–485. [CrossRef]

44.	Choi, S.W.; Park, J.H.; Lee, I.-B. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Comput. Chem. Eng.* **2004**, *28*, 1377–1387. [CrossRef]

45.	Yu, J.; Qin, S.J. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AlChE J.* **2008**, *54*, 1811–1829. [CrossRef]

46.	Chen, T.; Zhang, J. On-line multivariate statistical monitoring of batch processes using Gaussian mixture model. *Comput. Chem. Eng.* **2010**, *34*, 500–507. [CrossRef]

47.	Wang, J.; He, Q.P. Multivariate statistical process monitoring based on statistics pattern analysis. *Ind. Eng. Chem. Res.* **2010**, *49*, 7858–7869. [CrossRef]

48.	Galicia, H.J.; He, Q.P.; Wang, J. A comprehensive evaluation of Statistics Pattern Analysis based process monitoring. *IFAC Proc. Vol.* **2012**, *45*, 39–44. [CrossRef]

49.	Gajjar, S.; Kulahci, M.; Palazoglu, A. Real-time fault detection and diagnosis using sparse principal component analysis. *J. Process Control* **2018**, *67*, 112–128. [CrossRef]

50.	Gajjar, S.; Kulahci, M.; Palazoglu, A. Use of sparse principal component analysis (SPCA) for fault detection. *IFAC-PapersOnLine* **2016**, *49*, 693–698. [CrossRef]

51.	Zou, H.; Hastie, T.; Tibshirani, R. Sparse Principal Component Analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [CrossRef]

52.	Chun, H.; Keleş, S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Ser. B* **2010**, *72*, 3–25. [CrossRef] [PubMed]

53.	Lu, Q.; Jiang, B.; Gopaluni, R.B.; Loewen, P.D.; Braatz, R.D. Sparse canonical variate analysis approach for process monitoring. *J. Process Control* **2018**, *71*, 90–102. [CrossRef]

54.	Aldrich, C.; Auret, L. *Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 16.

55.	Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally. *AlChE J.* **2019**, *65*, 466–478. [CrossRef]

56.	Wu, H.; Zhao, J. Deep convolutional neural network model based chemical process fault diagnosis. *Comput. Chem. Eng.* **2018**, *115*, 185–197. [CrossRef]

57.	Gao, X.; Yang, F.; Feng, E. A process fault diagnosis method using multi-time scale dynamic feature extraction based on convolutional neural network. *Can. J. Chem. Eng.* **2020**, *98*, 1280–1292. [CrossRef]

58.	Fan, J.; Wang, W.; Zhang, H. AutoEncoder based high-dimensional data fault detection system. In Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (Indin), Emden, Germany, 24–26 July 2017; pp. 1001–1006.

59.	Sakurada, M.; Yairi, T. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis—MLSDA'14, Gold Coast, Australia, 2 December 2014; pp. 4–11.

60. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
61. Tax, D.M.; Duin, R.P. Support vector data description. *Mach. Learn.* **2004**, *54*, 45–66. [CrossRef]
62. Sukchotrat, T.; Kim, S.B.; Tsung, F. One-class classification-based control charts for multivariate process monitoring. *IIE Trans.* **2009**, *42*, 107–120. [CrossRef]
63. Song, Y.Y.; Lu, Y. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatry* **2015**, *27*, 130–135. [CrossRef]
64. Dorgo, G.; Palazoglu, A.; Abonyi, J. Decision trees for informative process alarm definition and alarm-based fault classification. *Process Saf. Environ. Prot.* **2021**, *149*, 312–324. [CrossRef]
65. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
66. Liu, Y.; Ge, Z. Weighted random forests for fault classification in industrial processes with hierarchical clustering model selection. *J. Process Control* **2018**, *64*, 62–70. [CrossRef]
67. Puggini, L.; Doyle, J.; McLoone, S. Fault detection using random forest similarity distance. *IFAC-PapersOnLine* **2015**, *48*, 583–588. [CrossRef]
68. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Washington, DC, USA, 15–19 December 2008; pp. 413–422.
69. Wang, H.; Jiang, W.; Deng, X.; Geng, J. A new method for fault detection of aero-engine based on isolation forest. *Measurement* **2021**, *185*, 110064. [CrossRef]
70. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef] [PubMed]
71. Luo, L. Process Monitoring with Global–Local Preserving Projections. *Ind. Eng. Chem. Res.* **2014**, *53*, 7696–7705. [CrossRef]
72. Ku, W.; Storer, R.H.; Georgakis, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179–196. [CrossRef]
73. Li, R.; Rong, G. Fault Isolation by Partial Dynamic Principal Component Analysis in Dynamic Process. *Chin. J. Chem. Eng.* **2006**, *14*, 486–493. [CrossRef]
74. Russell, E.L.; Chiang, L.H.; Braatz, R.D. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 81–93. [CrossRef]
75. Li, S.; Zhou, X.; Shi, H.; Pan, F. Dynamic Non-Gaussian hybrid serial modeling for industrial process monitoring. *Chemom. Intell. Lab. Syst.* **2021**, *216*, 1–15. [CrossRef]
76. Li, Z.; Yan, X. Complex dynamic process monitoring method based on slow feature analysis model of multi-subspace partitioning. *ISA Trans.* **2019**, *95*, 68–81. [CrossRef] [PubMed]
77. Huang, J.; Ersoy, O.K.; Yan, X. Fault detection in dynamic plant-wide process by multi-block slow feature analysis and support vector data description. *ISA Trans.* **2019**, *85*, 119–128. [CrossRef] [PubMed]
78. Qin, S.J.; Dong, Y.; Zhu, Q.; Wang, J.; Liu, Q. Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring. *Annu. Rev. Control* **2020**, *50*, 29–48. [CrossRef]
79. Li, G.; Liu, B.; Qin, S.J.; Zhou, D. Dynamic latent variable modeling for statistical process monitoring. *IFAC Proc. Vol.* **2011**, *44*, 12886–12891. [CrossRef]
80. Dong, Y.; Qin, S.J. A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *J. Process Control* **2018**, *67*, 1–11. [CrossRef]
81. Zheng, J.; Zhao, C.; Gao, F. Retrospective comparison of several typical linear dynamic latent variable models for industrial process monitoring. *Comput. Chem. Eng.* **2022**, *157*, 107587. [CrossRef]
82. Wiskott, L.; Sejnowski, T.J. Slow feature analysis: Unsupervised learning of invariances. *Neural Comput.* **2002**, *14*, 715–770. [CrossRef] [PubMed]
83. Shang, C.; Yang, F.; Gao, X.; Huang, X.; Suykens, J.A.K.; Huang, D. Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis. *AIChE J.* **2015**, *61*, 3666–3682. [CrossRef]
84. Huang, J.; Ersoy, O.K.; Yan, X. Slow feature analysis based on online feature reordering and feature selection for dynamic chemical process monitoring. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 1–11. [CrossRef]
85. Zheng, J.; Zhao, C. Enhanced canonical variate analysis with slow feature for dynamic process status analytics. *J. Process Control* **2020**, *95*, 10–31. [CrossRef]
86. Gao, X.; Shardt, Y.A.W. Dynamic system modelling and process monitoring based on long-term dependency slow feature analysis. *J. Process Control* **2021**, *105*, 27–47. [CrossRef]
87. 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. [CrossRef]
88. Apsemidis, A.; Psarakis, S.; Moguerza, J.M. A review of machine learning kernel methods in statistical process monitoring. *Comput. Ind. Eng.* **2020**, *142*, 106376. [CrossRef]
89. Schölkopf, B.; Smola, A.; Müller, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [CrossRef]
90. Lee, J.-M.; Yoo, C.; Choi, S.W.; Vanrolleghem, P.A.; Lee, I.-B. Nonlinear process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **2004**, *59*, 223–234. [CrossRef]

91. Cho, J.-H.; Lee, J.-M.; Wook Choi, S.; Lee, D.; Lee, I.-B. Fault identification for process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **2005**, *60*, 279–288. [CrossRef]
92. Alcala, C.F.; Qin, S.J. Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Ind. Eng. Chem. Res.* **2010**, *49*, 7849–7857. [CrossRef]
93. Zhang, X.; Yan, W.; Shao, H. Nonlinear multivariate quality estimation and prediction based on kernel partial least squares. *Ind. Eng. Chem. Res.* **2008**, *47*, 1120–1131. [CrossRef]
94. Godoy, J.L.; Zumoffen, D.A.; Vega, J.R.; Marchetti, J.L. New contributions to non-linear process monitoring through kernel partial least squares. *Chemom. Intell. Lab. Syst.* **2014**, *135*, 76–89. [CrossRef]
95. Lee, J.M.; Qin, S.J.; Lee, I.B. Fault detection of non-linear processes using kernel independent component analysis. *Can. J. Chem. Eng.* **2007**, *85*, 526–536. [CrossRef]
96. Fan, J.; Wang, Y. Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. *Inf. Sci.* **2014**, *259*, 369–379. [CrossRef]
97. Fan, J.; Qin, S.J.; Wang, Y. Online monitoring of nonlinear multivariate industrial processes using filtering KICA–PCA. *Control Eng. Pract.* **2014**, *22*, 205–216. [CrossRef]
98. Chen, Q.; Wang, Y. Key-performance-indicator-related state monitoring based on kernel canonical correlation analysis. *Control Eng. Pract.* **2021**, *107*, 104692. [CrossRef]
99. Huang, J.; Yan, X. Related and independent variable fault detection based on KPCA and SVDD. *J. Process Control* **2016**, *39*, 88–99. [CrossRef]
100. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]
101. Wan, F.; Guo, G.; Zhang, C.; Guo, Q.; Liu, J. Outlier Detection for Monitoring Data Using Stacked Autoencoder. *IEEE Access* **2019**, *7*, 173827–173837. [CrossRef]
102. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 1096–1103.
103. 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*, 1083–1091. [CrossRef]
104. Zhu, J.; Shi, H.; Song, B.; Tao, Y.; Tan, S.; Zhang, T. Nonlinear process monitoring based on load weighted denoising autoencoder. *Measurement* **2021**, *171*, 108782. [CrossRef]
105. Zhang, Z.; Jiang, T.; Li, S.; Yang, Y. Automated feature learning for nonlinear process monitoring—An approach using stacked denoising autoencoder and k-neare.est neighbor rule. *J. Process Control* **2018**, *64*, 49–61. [CrossRef]
106. An, J.; Cho, S. Variational autoencoder based anomaly detection using reconstruction probability. *Spec. Lect. IE* **2015**, *2*, 1–18.
107. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
108. Zhang, Z.; Jiang, T.; Zhan, C.; Yang, Y. Gaussian feature learning based on variational autoencoder for improving nonlinear process monitoring. *J. Process Control* **2019**, *75*, 136–155. [CrossRef]
109. Lee, S.; Kwak, M.; Tsui, K.-L.; Kim, S.B. Process monitoring using variational autoencoder for high-dimensional nonlinear processes. *Eng. Appl. Artif. Intell.* **2019**, *83*, 13–27. [CrossRef]
110. Bi, X.; Zhao, J. A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification. *Process Saf. Environ. Prot.* **2021**, *156*, 581–597. [CrossRef]
111. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic January, Vancouver, BC, Canada, 3–8 December 2001; pp. 585–591.
112. He, X.; Niyogi, P. Locality preserving projections. *Adv. Neural Inf. Process. Syst.* **2004**, *16*, 153–160.
113. Shao, J.-D.; Rong, G.; Lee, J.M. Generalized orthogonal locality preserving projections for nonlinear fault detection and diagnosis. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 75–83. [CrossRef]
114. He, F.; Xu, J. A novel process monitoring and fault detection approach based on statistics locality preserving projections. *J. Process Control* **2016**, *37*, 46–57. [CrossRef]
115. Zhang, M.; Ge, Z.; Song, Z.; Fu, R. Global–Local Structure Analysis Model and Its Application for Fault Detection and Identification. *Ind. Eng. Chem. Res.* **2011**, *50*, 6837–6848. [CrossRef]
116. Yu, J. Local and global principal component analysis for process monitoring. *J. Process Control* **2012**, *22*, 1358–1373. [CrossRef]
117. Luo, L.; Bao, S.; Mao, J.; Tang, D. Nonlinear Process Monitoring Using Data-Dependent Kernel Global–Local Preserving Projections. *Ind. Eng. Chem. Res.* **2015**, *54*, 11126–11138. [CrossRef]
118. Bao, S.; Luo, L.; Mao, J.; Tang, D. Improved fault detection and diagnosis using sparse global-local preserving projections. *J. Process Control* **2016**, *47*, 121–135. [CrossRef]
119. Zhan, C.; Li, S.; Yang, Y. Enhanced Fault Detection Based on Ensemble Global–Local Preserving Projections with Quantitative Global–Local Structure Analysis. *Ind. Eng. Chem. Res.* **2017**, *56*, 10743–10755. [CrossRef]
120. Li, Y.; Ma, F.; Ji, C.; Wang, J.; Sun, W.F. Fault Detection Method Based on Global-Local Marginal Discriminant Preserving Projection for Chemical Process. *Processes* **2022**, *10*, 122. [CrossRef]
121. Kwak, S.; Ma, Y.; Huang, B. Extracting nonstationary features for process data analytics and application in fouling detection. *Comput. Chem. Eng.* **2020**, *135*, 106762. [CrossRef]

122. Li, G.; Qin, S.J.; Yuan, T. Nonstationarity and cointegration tests for fault detection of dynamic processes. *IFAC Proc. Vol.* **2014**, *47*, 10616–10621. [CrossRef]
123. He, Z.; Zhou, H.; Wang, J.; Chen, Z.; Wang, D.; Xing, Y. An Improved Detection Statistic for Monitoring the Nonstationary and Nonlinear Processes. *Chemom. Intell. Lab. Syst.* **2015**, *145*, 114–124. [CrossRef]
124. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
125. Zhou, Y.; Huang, M. Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. *Microelectron. Reliab.* **2016**, *65*, 265–273. [CrossRef]
126. Duan, X.; Zhang, X. ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. *Data Brief* **2020**, *31*, 105779. [CrossRef]
127. Mu, W.; Zhang, A.; Gao, W.; Huo, X. Application of ARIMA Model in Fault Diagnosis of TEP. In Proceedings of the 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), Liuzhou, China, 20–22 November 2020; pp. 393–398.
128. Qin, S.J.; Li, W.; Henry Yue, H. Recursive PCA for Adaptive Process Monitoring. *IFAC Proc. Vol.* **1999**, *32*, 6686–6691. [CrossRef]
129. Wang, X.; Kruger, U.; Irwin, G.W. Process monitoring approach using fast moving window PCA. *Ind. Eng. Chem. Res.* **2005**, *44*, 5691–5702. [CrossRef]
130. Wang, Y.; Jiang, Q. Recursive correlated representation learning for adaptive monitoring of slowly varying processes. *ISA Trans.* **2020**, *107*, 360–369. [CrossRef] [PubMed]
131. Qin, S.J. Recursive PLS algorithms for adaptive data modeling. *Comput. Chem. Eng.* **1998**, *22*, 503–514. [CrossRef]
132. Wang, X.; Kruger, U.; Lennox, B. Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Eng. Pract.* **2003**, *11*, 613–632. [CrossRef]
133. Shang, C.; Yang, F.; Huang, B.; Huang, D. Recursive Slow Feature Analysis for Adaptive Monitoring of Industrial Processes. *IEEE Trans. Ind. Electron.* **2018**, *65*, 8895–8905. [CrossRef]
134. Engle, R.F.; Granger, C.W.J. Co-integration and error correction: Representation, estimation, and testing. *Econom. J. Econom. Soc.* **1987**, *55*, 251–276. [CrossRef]
135. Chen, Q.; Kruger, U.; Leung, A.Y.T. Cointegration testing method for monitoring nonstationary processes. *Ind. Eng. Chem. Res.* **2009**, *48*, 3533–3543. [CrossRef]
136. Yu, W.; Zhao, C.; Huang, B. Recursive cointegration analytics for adaptive monitoring of nonstationary industrial processes with both static and dynamic variations. *J. Process Control* **2020**, *92*, 319–332. [CrossRef]
137. Wen, J.; Li, Y.; Wang, J.; Sun, W. Nonstationary Process Monitoring Based on Cointegration Theory and Multiple Order Moments. *Processes* **2022**, *10*, 169. [CrossRef]
138. Zhao, C.; Huang, B. A full-condition monitoring method for nonstationary dynamic chemical processes with cointegration and slow feature analysis. *AlChE J.* **2017**, *64*, 1662–1681. [CrossRef]
139. Lin, Y.; Kruger, U.; Chen, Q. Monitoring Nonstationary Dynamic Systems Using Cointegration and Common-Trends Analysis. *Ind. Eng. Chem. Res.* **2017**, *56*, 8895–8905. [CrossRef]
140. Lin, Y.; Kruger, U.; Gu, F.; Ball, A.; Chen, Q. Monitoring nonstationary and dynamic trends for practical process fault diagnosis. *Control Eng. Pract.* **2019**, *84*, 139–158. [CrossRef]
141. Wu, H.; Zhao, J. Self-adaptive deep learning for multimode process monitoring. *Comput. Chem. Eng.* **2020**, *141*, 107024. [CrossRef]
142. Zhao, S.J.; Zhang, J.; Xu, Y.M. Monitoring of processes with multiple operating modes through multiple principle component analysis models. *Ind. Eng. Chem. Res.* **2004**, *43*, 7025–7035. [CrossRef]
143. Zhao, S.J.; Zhang, J.; Xu, Y.M. Performance monitoring of processes with multiple operating modes through multiple PLS models. *J. Process Control* **2006**, *16*, 763–772. [CrossRef]
144. Yoo, C.K.; Villez, K.; Lee, I.B.; Rosen, C.; Vanrolleghem, P.A. Multi-model statistical process monitoring and diagnosis of a sequencing batch reactor. *Biotechnol. Bioeng.* **2007**, *96*, 687–701. [CrossRef] [PubMed]
145. Tan, S.; Wang, F.; Peng, J.; Chang, Y.; Wang, S. Multimode Process Monitoring Based on Mode Identification. *Ind. Eng. Chem. Res.* **2011**, *51*, 374–388. [CrossRef]
146. Ha, D.; Ahmed, U.; Pyun, H.; Lee, C.-J.; Baek, K.H.; Han, C. Multi-mode operation of principal component analysis with k-nearest neighbor algorithm to monitor compressors for liquefied natural gas mixed refrigerant processes. *Comput. Chem. Eng.* **2017**, *106*, 96–105. [CrossRef]
147. Tong, C.; Palazoglu, A.; Yan, X. An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding. *J. Process Control* **2013**, *23*, 1497–1507. [CrossRef]
148. Zhang, S.; Wang, F.; Tan, S.; Wang, S.; Chang, Y. Novel Monitoring Strategy Combining the Advantages of the Multiple Modeling Strategy and Gaussian Mixture Model for Multimode Processes. *Ind. Eng. Chem. Res.* **2015**, *54*, 11866–11880. [CrossRef]
149. Xie, X.; Shi, H. Dynamic Multimode Process Modeling and Monitoring Using Adaptive Gaussian Mixture Models. *Ind. Eng. Chem. Res.* **2012**, *51*, 5497–5505. [CrossRef]
150. Cao, Y.; Jan, N.M.; Huang, B.; Fang, M.; Wang, Y.; Gui, W. Multimodal process monitoring based on variational Bayesian PCA and Kullback-Leibler divergence between mixture models. *Chemom. Intell. Lab. Syst.* **2021**, *210*, 104230. [CrossRef]
151. Han, X.; Tian, S.; Romagnoli, J.A.; Li, H.; Sun, W. PCA-SDG based process monitoring and fault diagnosis: Application to an industrial pyrolysis furnace. *IFAC-PapersOnLine* **2018**, *51*, 482–487. [CrossRef]

152. Ma, H.; Hu, Y.; Shi, H. A novel local neighborhood standardization strategy and its application in fault detection of multimode processes. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 287–300. [CrossRef]

153. Wang, G.; Liu, J.; Zhang, Y.; Li, Y. A novel multi-mode data processing method and its application in industrial process monitoring. *J. Chemom.* **2015**, *29*, 126–138. [CrossRef]

154. Zhao, C.; Yao, Y.; Gao, F.; Wang, F. Statistical analysis and online monitoring for multimode processes with between-mode transitions. *Chem. Eng. Sci.* **2010**, *65*, 5961–5975. [CrossRef]

155. Wang, F.; Tan, S.; Peng, J.; Chang, Y. Process monitoring based on mode identification for multi-mode process with transitions. *Chemom. Intell. Lab. Syst.* **2012**, *110*, 144–155. [CrossRef]

156. Wang, F.; Tan, S.; Yang, Y.; Shi, H. Hidden Markov Model-Based Fault Detection Approach for a Multimode Process. *Ind. Eng. Chem. Res.* **2016**, *55*, 4613–4621. [CrossRef]

157. Wang, F.; Tan, S.; Shi, H. Hidden Markov model-based approach for multimode process monitoring. *Chemom. Intell. Lab. Syst.* **2015**, *148*, 51–59. [CrossRef]

158. Wang, L.; Yang, C.; Sun, Y. Multimode Process Monitoring Approach Based on Moving Window Hidden Markov Model. *Ind. Eng. Chem. Res.* **2017**, *57*, 292–301. [CrossRef]

159. Nomikos, P.; MacGregor, J.F. Monitoring batch processes using multiway principal component analysis. *AlChE J.* **1994**, *40*, 1361–1375. [CrossRef]

160. Birol, G.; Ündey, C.; Cinar, A. A modular simulation package for fed-batch fermentation: Penicillin production. *Comput. Chem. Eng.* **2002**, *26*, 1553–1565. [CrossRef]

161. Nomikos, P.; MacGregor, J.F. Multi-way partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 97–108. [CrossRef]

162. Wold, S.; Kettaneh, N.; Fridén, H.; Holmberg, A. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 331–340. [CrossRef]

163. Chen, J.; Liu, K.-C. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem. Eng. Sci.* **2002**, *57*, 63–75. [CrossRef]

164. Choi, S.W.; Morris, J.; Lee, I.-B. Dynamic model-based batch process monitoring. *Chem. Eng. Sci.* **2008**, *63*, 622–636. [CrossRef]

165. Lee, J.-M.; Yoo, C.; Lee, I.-B. Fault detection of batch processes using multiway kernel principal component analysis. *Comput. Chem. Eng.* **2004**, *28*, 1837–1847. [CrossRef]

166. Zhang, Y.; Qin, S.J. Fault detection of nonlinear processes using multiway kernel independent component analysis. *Ind. Eng. Chem. Res.* **2007**, *46*, 7780–7787. [CrossRef]

167. Gao, X.; Xu, Z.; Li, Z.; Wang, P. Batch process monitoring using multiway Laplacian autoencoders. *Can. J. Chem. Eng.* **2020**, *98*, 1269–1279. [CrossRef]

168. Zhang, H.; Tian, X.; Deng, X. Batch Process Monitoring Based on Multiway Global Preserving Kernel Slow Feature Analysis. *IEEE Access* **2017**, *5*, 2696–2710. [CrossRef]

169. Hui, Y.; Zhao, X. Multiway dynamic nonlinear global neighborhood preserving embedding method for monitoring batch process. *Meas. Control* **2020**, *53*, 994–1006. [CrossRef]

170. Van Sprang, E.N.; Ramaker, H.-J.; Westerhuis, J.A.; Gurden, S.P.; Smilde, A.K. Critical evaluation of approaches for on-line batch process monitoring. *Chem. Eng. Sci.* **2002**, *57*, 3979–3991. [CrossRef]

171. Undey, C.; Cinar, A. Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Syst. Mag.* **2002**, *22*, 40–52. [CrossRef]

172. Yao, Y.; Gao, F. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu. Rev. Control* **2009**, *33*, 172–183. [CrossRef]

173. Nomikos, P.; MacGregor, J.F. Multivariate SPC charts for monitoring batch processes. *Technometrics* **1995**, *37*, 41–59. [CrossRef]

174. Zhao, L.; Zhao, C.; Gao, F. Inter-batch-evolution-traced process monitoring based on inter-batch mode division for multiphase batch processes. *Chemom. Intell. Lab. Syst.* **2014**, *138*, 178–192. [CrossRef]

175. Lu, N.; Gao, F.; Wang, F. Sub-PCA modeling and on-line monitoring strategy for batch processes. *AlChE J.* **2004**, *50*, 255–259. [CrossRef]

176. Ge, Z.; Gao, F.; Song, Z. Batch process monitoring based on support vector data description method. *J. Process Control* **2011**, *21*, 949–959. [CrossRef]

177. Camacho, J.; Picó, J. Multi-phase principal component analysis for batch processes modelling. *Chemom. Intell. Lab. Syst.* **2006**, *81*, 127–136. [CrossRef]

178. Camacho, J.; Picó, J. Online monitoring of batch processes using multi-phase principal component analysis. *J. Process Control* **2006**, *16*, 1021–1035. [CrossRef]

179. Zhao, C.; Gao, F.; Wang, F. Nonlinear Batch Process Monitoring Using Phase-Based Kernel-Independent Component Analysis-Principal Component Analysis (KICAPCA). *Ind. Eng. Chem. Res.* **2009**, *48*, 9163–9174. [CrossRef]

180. Yu, J. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chem. Eng. Sci.* **2012**, *68*, 506–519. [CrossRef]

181. Yu, J.; Qin, S.J. Multiway Gaussian mixture model based multiphase batch process monitoring. *Ind. Eng. Chem. Res.* **2009**, *48*, 8585–8594. [CrossRef]

182. Yao, Y.; Gao, F. Phase and transition based batch process modeling and online monitoring. *J. Process Control* **2009**, *19*, 816–826. [CrossRef]
183. Sun, W.; Meng, Y.; Palazoglu, A.; Zhao, J.; Zhang, H.; Zhang, J. A method for multiphase batch process monitoring based on auto phase identification. *J. Process Control* **2011**, *21*, 627–638. [CrossRef]
184. Liu, J.; Liu, T.; Chen, J. Sequential local-based Gaussian mixture model for monitoring multiphase batch processes. *Chem. Eng. Sci.* **2018**, *181*, 101–113. [CrossRef]
185. Chang, C.C.; Yu, C.C. On-line fault diagnosis using the signed directed graph. *Ind. Eng. Chem. Res.* **1990**, *29*, 1290–1299. [CrossRef]
186. Hiranmayee, V.; Venkatasubramanian, V. PCA-SDG based process monitoring and fault diagnosis. *Control. Eng. Pract.* **1999**, *7*, 903–917. [CrossRef]
187. Wang, H.; Chen, B.; He, X.; Tong, Q.; Zhao, J. SDG-based HAZOP analysis of operating mistakes for PVC process. *Process Saf. Environ. Prot.* **2009**, *87*, 40–46. [CrossRef]
188. Ji, C.; Zhu, X.; Ma, F.; Wang, J.; Sun, W. Fault Diagnosis Algorithm of Chemical Process Based on Information Entropy. *Chem. Eng. Trans.* **2020**, *81*, 541–546. [CrossRef]
189. Yan, Z.; Yao, Y. Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO). *Chemom. Intell. Lab. Syst.* **2015**, *146*, 136–146. [CrossRef]
190. Miller, P.; Swanson, R.E.; Heckler, C.E. Contribution plots: A missing link in multivariate quality control. *Appl. Math. Comput. Sci.* **1998**, *8*, 775–792.
191. Kourti, T.; MacGregor, J.F. Multivariate SPC Methods for Process and Product Monitoring. *J. Qual. Technol.* **2018**, *28*, 409–428. [CrossRef]
192. Jiang, B.; Huang, D.; Zhu, X.; Yang, F.; Braatz, R.D. Canonical variate analysis-based contributions for fault identification. *J. Process Control* **2015**, *26*, 17–25. [CrossRef]
193. Conlin, A.K.; Martin, E.B.; Morris, A.J. Confidence limits for contribution plots. *J. Chemom. A J. Chemom. Soc.* **2000**, *14*, 725–736. [CrossRef]
194. Westerhuis, J.A.; Gurden, S.P.; Smilde, A.K. Generalized contribution plots in multivariate statistical process monitoring. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 95–114. [CrossRef]
195. Alcala, C.F.; Qin, S.J. Reconstruction-based contribution for process monitoring. *Automatica* **2009**, *45*, 1593–1600. [CrossRef]
196. Peng, K.; Zhang, K.; Li, G.; Zhou, D. Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process. *Control Eng. Pract.* **2013**, *21*, 360–369. [CrossRef]
197. Ning, C.; Chen, M.; Zhou, D. Sparse contribution plot for fault diagnosis of multimodal chemical processes. *IFAC-PapersOnLine* **2015**, *48*, 619–626. [CrossRef]
198. Ji, H.; He, X.; Zhou, D. On the use of reconstruction-based contribution for fault diagnosis. *J. Process Control* **2016**, *40*, 24–34. [CrossRef]
199. Xu, H.; Yang, F.; Ye, H.; Li, W.; Xu, P.; Usadi, A.K. Weighted Reconstruction-Based Contribution for Improved Fault Diagnosis. *Ind. Eng. Chem. Res.* **2013**, *52*, 9858–9870. [CrossRef]
200. Zhao, C.; Zhang, W. Reconstruction based fault diagnosis using concurrent phase partition and analysis of relative changes for multiphase batch processes with limited fault batches. *Chemom. Intell. Lab. Syst.* **2014**, *130*, 135–150. [CrossRef]
201. Cai, B.; Huang, L.; Xie, M. Bayesian Networks in Fault Diagnosis. *IEEE Trans. Ind. Inf.* **2017**, *13*, 2227–2240. [CrossRef]
202. Amin, M.T.; Imtiaz, S.; Khan, F. Process system fault detection and diagnosis using a hybrid technique. *Chem. Eng. Sci.* **2018**, *189*, 191–211. [CrossRef]
203. Gharahbagheri, H.; Imtiaz, S.A.; Khan, F. Root Cause Diagnosis of Process Fault Using KPCA and Bayesian Network. *Ind. Eng. Chem. Res.* **2017**, *56*, 2054–2070. [CrossRef]
204. Amin, M.T.; Khan, F.; Imtiaz, S. Fault detection and pathway analysis using a dynamic Bayesian network. *Chem. Eng. Sci.* **2019**, *195*, 777–790. [CrossRef]
205. Bauer, M.; Thornhill, N.F. A practical method for identifying the propagation path of plant-wide disturbances. *J. Process Control* **2008**, *18*, 707–719. [CrossRef]
206. Yuan, T.; Qin, S.J. Root cause diagnosis of plant-wide oscillations using granger causality. *IFAC Proc. Vol.* **2012**, *45*, 160–165. [CrossRef]
207. Li, S.; Xiao, Y.; Zhou, D.; Cai, D. Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information. *Phys. Rev. E* **2018**, *97*, 052216. [CrossRef]
208. Chen, H.-S.; Zhao, C.; Yan, Z.; Yao, Y. Root cause diagnosis of oscillation-type plant faults using nonlinear causality analysis. *IFAC-PapersOnLine* **2017**, *50*, 13898–13903. [CrossRef]
209. Bauer, M.; Cox, J.W.; Caveness, M.H.; Downs, J.J.; Thornhill, N.F. Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy. *IEEE Trans. Control Syst. Technol.* **2007**, *15*, 12–21. [CrossRef]
210. Hajihosseini, P.; Salahshoor, K.; Moshiri, B. Process fault isolation based on transfer entropy algorithm. *ISA Trans.* **2014**, *53*, 230–240. [CrossRef] [PubMed]
211. Li, G.; Qin, S.J.; Yuan, T. Data-driven root cause diagnosis of faults in process industries. *Chemom. Intell. Lab. Syst.* **2016**, *159*, 1–11. [CrossRef]
212. Lee, H.; Kim, C.; Lim, S.; Lee, J.M. Data-driven fault diagnosis for chemical processes using transfer entropy and graphical lasso. *Comput. Chem. Eng.* **2020**, *142*, 107064. [CrossRef]

213. Ji, C.; Ma, F.; Wang, J.; Wang, J.; Sun, W. Real-Time Industrial Process Fault Diagnosis Based on Time Delayed Mutual Information Analysis. *Processes* **2021**, *9*, 1027. [CrossRef]
214. Stockmann, M.; Haber, R. Determination of fault propagation by time delay estimation using k nearest neighbour imputation. In Proceedings of the 2010 Conference on Control and Fault-Tolerant Systems (SysTol), Nice, France, 6–8 October 2010; pp. 185–190.
215. Luo, L.; Cheng, F.; Qiu, T.; Zhao, J. Refined convergent cross-mapping for disturbance propagation analysis of chemical processes. *Comput. Chem. Eng.* **2017**, *106*, 1–16. [CrossRef]
216. Tong, C.; Lan, T.; Yu, H.; Peng, X. Distributed partial least squares based residual generation for statistical process monitoring. *J. Process Control* **2019**, *75*, 77–85. [CrossRef]
217. Ji, C.; Ma, F.; Zhu, X.; Wang, J.; Sun, W. Fault Propagation Path Inference in a Complex Chemical Process Based on Time-delayed Mutual Information Analysis. In *30th European Symposium on Computer Aided Process Engineering*; Computer Aided Chemical Engineering; Elsevier Science: Amsterdam, The Netherlands, 2020; pp. 1165–1170.