IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 31, 2017, accepted September 22, 2017, date of publication September 26, 2017, date of current version October 25, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2756872

Data Mining and Analytics in the Process Industry: The Role of Machine Learning

ZHIQIANG GE[1], (Senior Member, IEEE), ZHIHUAN SONG[1], STEVEN X. DING[2], AND BIAO HUANG[3], (Senior Member, IEEE)

[1]State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
[2]Institute for Automatic Control and Complex Systems, University of Duisburg-Essen, 47057 Duisburg, Germany
[3]Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G2G6, Canada

Corresponding author: Zhiqiang Ge (gezhiqiang@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61722310 and Grant 61673337 and in part by the Alexander von Humboldt Foundation.

ABSTRACT Data mining and analytics have played an important role in knowledge discovery and decision making/supports in the process industry over the past several decades. As a computational engine to data mining and analytics, machine learning serves as basic tools for information extraction, data pattern recognition and predictions. From the perspective of machine learning, this paper provides a review on existing data mining and analytics applications in the process industry over the past several decades. The state-of-the-art of data mining and analytics are reviewed through eight unsupervised learning and ten supervised learning algorithms, as well as the application status of semi-supervised learning algorithms. Several perspectives are highlighted and discussed for future researches on data mining and analytics in the process industry.

INDEX TERMS Data mining, data analytics, machine learning, process industry.

I. INTRODUCTION

In recent years, many unions or countries have announced a new round of development plans in manufacturing. For example, the European Union proposed 20-20-20 goals to achieve a sustainable future, which means 20% increase in energy efficiency, 20% reduction of CO2 emissions, and 20% renewables by 2020. Those goals can only be realized by incorporating more intelligence into industrial manufacturing process. The US government has proposed a new industrial internet framework for developing the next generation of manufacturing. Japan announced a revitalization plan for manufacturing. Germany proposed the concept of industry 4.0, the main aim of which is to develop smart factories for producing smart products. Similarly, China has announced a new manufacturing plan more recently, which is known as China Manufacturing 2025, the aim of which is also to make the manufacturing process more intelligent. As an important part of manufacturing, the process industry needs to be re-formulized in the era of new information technologies. How to improve process understanding and accumulate effective knowledge plays an important role in all aspects of the process industry, such as sustainability design, system integration, advanced process/quality control, decision supports, etc. From the viewpoint of automation, data mining and analytics may serve as a basic tool to promote the process industry from machine automation to information automation and then to knowledge automation.

In the past several decades, a large amount of data has been achieved in the process industry, due to the wide use of distributed control systems. While it becomes more and more difficult to build first-principle models in those increasingly complex processes, data-driven process modeling, monitoring, prediction and control have received much attention in recent years. At the very beginning, those large amounts of data have rarely been used for detailed analyses, which are instead only used for routinely technical checks and process log fulfillments. Later, awareness of the importance in extracting information from data has taken a leading role in the process industry. At the same time, the technical development of database and data analyses has been in a fast pace during those years. Therefore, researches on data mining and analytics in the process industry have become quite popular since then [1]–[8]. By analyzing the patterns of process data and relationships among variables, useful information can be extracted, based on which statistical models can then be developed for various applications, such as process monitoring, fault diagnosis, mode clustering, soft sensing of key variables/quality variables, etc. In a word, the main aim of data mining and data analyses is to extract useful information from process data, and transfer it to effective knowledge for improvement of understanding and decision supports of the process.

To effectively carry out data mining and analytics in the process industry, machine learning algorithms have always played an important role. There are four different types of machine learning algorithms, termed as unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning [9]. While the reinforcement learning is used in robotics, gaming and navigation areas, other three types of machine learning have all been widely used for data mining and analytics in the process industry. According to the introduction provided by Wikipedia, machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. However, the theory of probability and statistics also plays a very important role in modern machine learning. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed" [10], and later Tom M. Mitchell provided a widely quoted, more formal definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [11]. Compared to other definitions, this definition is more notable. Since its definition, machine learning has attracted lots of attention in various areas, such as pattern recognition, industrial applications, bioinformatics, image analyses, etc. Although the research of machine learning sunk for several times, the ship has been re-sailed and will become increasingly more powerful since we entered the 21 century.

During past several decades, machine learning has been playing an important role in constructing experience based models from process data, based on which useful information can be extracted, new patterns of data can be identified, predictions can be made more easily for new data samples, decisions can be made more quickly and effectively, etc. All these applications of machine learning have fundamentally changed the manufacturing style of the process industry. For data mining and analytics in the process industry, lots of works have been reported in the past years, among then there are several reviews and books which provided more detailed state-of-the-art of the research on this topic [1]–[8], [12]–[16]. However, there exists no such a report that details data mining and analytics from the viewpoint of machine learning, although the terminology has been mentioned many times in various papers and books.

The first aim of this paper is to provide a systematic review on data mining and analytics in process industry up to date, from the viewpoint of machine learning. This will include reviews on the topics of unsupervised learning, supervised learning and semi-supervised learning methods for applications in the process industry. Under each specific review topic, various applications of machine learning algorithms will be introduced, with detailed discussions among different methods. The second aim of this paper is to provide new perspectives for future applications of machine learning algorithms in the process industry. With the rapid development of the modern industry, new phenomena of process data have emerged, e.g. big data problem, which need to be carefully addressed when applying machine learning methods. On the other hand, the technology of machine learning itself continues to evolve all the time. For example, neural networks have been used for data mining for many years. With the increased computing power, we can create neural networks with many layers, which is known as deep learning neural networks. Both of the change in industry environments and the increased power of modern technologies have made the application of machine learning more challenging, more exciting, and thus need more future investigations and more ambitious researchers to explore in this area.

The rest of this paper is organized as follows. In section II, the main methodologies of data mining and analytics in the process industry are provided. Detailed machine learning reviews for data mining and analytics are carried out in section III, in which discussions and some remarks are also provided. In section IV, some new perspectives on the topic of machine learning applications in the process industry are presented. Finally, conclusions are made.

II. METHODOLOGY OF DATA MINING AND ANALYTICS

In this section, the methodology of data mining and analytics is illustrated, with detailed descriptions and discussions in each of the main procedures, including data preparation, data pre-processing, model selection, training, performance evaluation, and data mining and analytics based on the trained model. It should be noted that there are too many application aspects of data mining and analytics in the process industry, only some commonly used ones are discussed here. An overview of the data mining and analytics methodology is presented in Fig.1, in which the areas related to machine learning are highlighted in yellow.

A. DATA PREPARATION

Data preparation is an initial step for machine learning model development, and the aim of this step is to get an overview of the process data and then select the most appropriate data samples for modeling. The main tasks of this step are to extract the dataset from the historical database, examine the structure of the dataset, and make data selections through sample and variable directions, etc. In order to extract an effective dataset from the historical database, the operating regions of the process need to be analyzed, and any changes of operating condition also need to be identified. To ensure the efficiency for the information extraction step, the natures or characteristics of the process data should be analyzed, such as

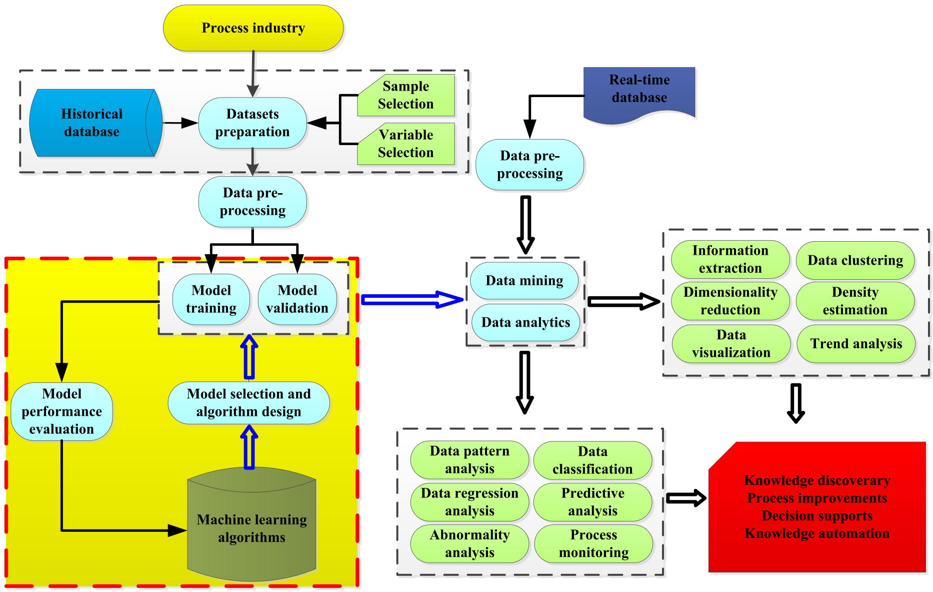


FIGURE 1. Overview of the data mining and analytics methodology.

non-Gaussianity, linear/nonlinear relationships among different variables, time-series correlations, etc. Another important issue of this step is to carry out sample and variable selections. Actually, this is highly related to the step of machine learning model development [9]. It depends on what kind of model we want to develop, and what is the main task of the model. For example, if the machine learning model is developed for monitoring a specific part of the process, e.g. PCA model, then a dataset from a stable operating condition of the process needs to be selected, and the corresponding variables in this specific part should be selected for modeling. If the aim of the model is for quality prediction, then those variables which are highly related to the predicted quality variables/indices needs to be picked out for development of the machine learning model. In summary, data preparation is a fundamental step for machine learning of the process data.

B. DATA PRE-PROCESSING

After the dataset has been prepared in the previous step, data pre-processing needs to be carried out in order to improve the quality of the data, and some appropriate data transformations may be needed to make the data modeling more efficient [17]. First, the time axis of the process data should be carefully checked, any inconsistency needs to be eliminated. To do this, some methods for time warping/dynamic time warping may be useful [18]. Second, outliers and gross errors should be removed from the modeling dataset, which will otherwise greatly deteriorate the performance of the machine learning model [19]. Third, there may be some missing data in the prepared dataset, those missing values need to be addressed, e.g. deletion of the sample, missing value estimation, Bayesian inference, etc [20], [21]. Fourth, the scale difference among process variables needs to be considered. For example, process variables are always normalized to zero means and unit variance in the PCA modeling method, in which the scale difference among different variables has been removed. As a result, the data model will not be inclined to any process variable, which means all variables are equally considered no matter how big or how small the absolute value is. However, this data-preprocessing step needs to be addressed carefully, since some models or learning tasks may have requirements on different scales for different variables. In this case, there are also other data-scaling and data transformation methods which may enhance the performance of the data model.

C. MODEL SELECTION, TRAINING, AND PERFORMANCE EVALUATION

When the training dataset is ready to use, we are in the position to select an appropriate machine learning algorithm for data model construction. Based on the detailed analyses of data characteristics, the complexity of the data model can be evaluated [9]; for example, what kind of model we should employ for machine learning of the training dataset? How much complexity the model structure should provide to describe the data information? Is a single model structure enough? Or is a multiple model structure needed? And so on. Take the regression based modeling for an example, if the relationships among the two types of variables are linear, then a multivariate linear regression learning algorithm can be selected; if there are strong nonlinear relationships, a nonlinear learning algorithm such as artificial neural networks or support vector regression model is required; if multiple operating conditions have been identified in the training dataset, a multiple regression model structure needs to be selected; or if the time-series dynamic feature is obvious in some process variables, a dynamical learning method should be employed to include the dynamic information of the process data [22], [23]. Although there is no unified model selection method, the machine learning community provides several useful rules or criterions which could be considered as promising tools for model selection, such as Akaike Information Criterion, Bayesian Information Criterion, Hannan-Quinn Criterion, etc [9], [11].

Once the structure of the model has been selected, the model parameters can be determined by implementing the machine learning algorithm on the training dataset. In this step, different data models depend on their own learning algorithms, most of which can actually be formulated as optimization problems [11], [12]. Before putting the model for online utilization, its performance needs to be evaluated. There are many well defined methods for model validation and performance evaluation, such as cross-validation, model stability analysis, model robust analysis, parameter sensitivity analysis, etc [9]. To do model validation and performance evaluation, another separate testing or validation dataset is always required. However, in some particular cases, we may not be able to obtain sufficient data to carry out separate model validation and performance evaluation. In this case, some re-sampling methods such as bagging and boosting, and the Bayesian learning algorithms are particularly useful. With the change of the process operating condition, model selection, training and performance evaluations should be carried out periodically, in order to provide a satisfactory model for data mining and analytics.

D. DATA MINING AND ANALYTICS

Now, suppose the machine learning procedure is completed, and the data model has been trained and validated. The model can now be used for both offline and online data mining and analytics, such as data clustering, dimensionality reduction, data visualization, trend analysis, process monitoring and fault diagnosis, fault classification, online soft sensing and quality prediction, and so on. Several examples are provided as follows.

For process monitoring and fault diagnosis, appropriate statistics can be constructed for online monitoring the operating condition of the process [24]. For example, two monitoring statistics T^2 and SPE are usually constructed in multivariate statistical monitoring methods. By defining control limits for those two monitoring statistics, the good and bad operating condition of the process can be well differentiated. If the values of the monitoring statistics exceed their corresponding control limits, a fault alarm should be triggered, which denotes the abnormal condition of the process. In this case, further analyses need to be carried out, in order to find the root cause of the abnormal event. This is the main task of fault diagnosis, which has been researched in the process industry for many years [25]. Fault diagnosis intends to provide detailed reasons for the announced fault in the process. Depending on different fault diagnosis methods, the root causes of the fault may be located in a small part of the process or precisely to some specific sensors/actuators [26]. As soon as the fault diagnosis results have been obtained, an evaluation report about the operating performance of the process can be generated, based on which appropriate process improvements or optimizations can be made afterward [27].

There may be various abnormal events or faults that can occur in the process. Therefore, when a fault alarm has been trigged in the process, we may want to know which type of faults it belongs to. This is corresponding to the data/fault classification problem [28]. For data classification modeling, machine learning has provide many useful algorithms, detailed reviews of which will be provided in the section 3. In order to develop an effective classification model, acquirements of data information from different fault types are important. If the detected data pattern/fault can be successfully identified, it may provide a great convenience for engineers to understand the condition of the process, and thus appropriate maintenance strategy can be quickly formulated.

In order to examine the data pattern of the process, e.g. how many clusters are among the whole dataset, clustering analysis can be carried out [29]. This can be done in both offline and online stages. In the offline stage, different clusters can be identified, which may correspond to various operating modes of the process. This clustering result can help one to understand the operating pattern/mode of the process, depending on which specific data mining and analytics can take place for different clusters. As a result, different process improvements or decision supports can be made for different operating regions of the process. In the online stage, the pattern of the new data sample can be identified and located in a specific operating region of the process. Therefore, the corresponding data model or process knowledge can be employed for information mining and analytics of the new data. This will greatly improve the efficiency for online data mining and analytics.

Usually, the dimensionality of the process data is very high, due to the wide use of various instrumentation tools in modern industry processes. Carrying out data mining and analytics from the original high-dimensional process variables is difficult, since those variables are highly correlated with each other, and data information is strongly overlapped, making data visualization difficult. Therefore, dimensionality reduction is necessary for mining effective data information from the process [30]. With the help of dimensionality reduction, the key data information can be extracted while simultaneously the computational complexity of the following data analytics procedures could be significantly reduced. Furthermore, with the compression of data information, data visualization will become much easier which is of great importance for information expression and results release.

Another interesting application of data mining and analytics is for soft sensing or prediction of key performance indices in the process. While those ordinary variables such as temperature and flow can be routinely recorded from the process, some key performance indices are very difficult to measure online [2]. For example, measuring the melt index in the polypropylene production process is very difficult, the value can only be obtained through laboratory analysis. If those key performance indices cannot be measured on time, the control system will be introduced by a significant delay. Fortunately, based on soft sensing or prediction modeling techniques, those key performance indices may be estimated online. Through mining the regression relationships between the ordinary process variables and those key performance indices, predictive data models can be derived and applied for online prediction of new data samples. As a result, real-time predictions can be obtained through regression analytics based on the predictive model. Actually, this predictive data mining and analytics can be applied throughout the whole process, including quality estimation, key performance index prediction, energy consumption estimation, environment performance evaluation, economic performance calculation, etc. All of those results from predictive data mining and analytics can help us to improve the efficiency and understanding of the process system.

III. MACHINE LEARNING METHODS APPLIED IN THE PROCESS INDUSTRY

This section carries out a systematic review on machine learning methods that have been applied in the process industry in the past years. As have been mentioned, machine learning methods can be partitioned into four different categories: unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning. Two of the most widely adopted machine learning methods are supervised learning and unsupervised learning, which may account 80-90 percent of all industry applications. While semi-supervised learning methods have recently been used for data classification and regression in the process industry, the reinforcement learning has rarely been used in this area.

Cases in which the data consists of samples of the inputs along with their corresponding targets are termed as supervised learning applications. For example, for process fault classification, we try to classify the faults into different categories. If the desired output consists of one or more continuous variables, then the task of supervised learning is called regression. A typical example of the data regression problem is the prediction of the key performance in the process, in which the inputs contain routinely recorded process variables such as temperature and pressure. Those applications in which the data only consists of a set of inputs without any corresponding targets are known as unsupervised learning. The goal of such unsupervised learning problems may be to discover groups of similar examples within the data which is called clustering, or to determine the distribution of data within the input space, known as density estimation, or to project the data from a high-dimensional space down to low-dimensional space for the purpose of dimensionality reduction and data visualization. Semi-supervised learning has recently attracted much attention in the process industry, which is applied for similar purposes as supervised learning. In semi-supervised learning, a small amount of labeled data and a large amount of unlabeled data are typical assumed for modeling. Semi-supervised learning is particularly useful when the cost associated with labeling is too high to allow for a fully labeled training process. Instead, unlabeled data is much cheaper and takes less effort to acquire from the process. This type of machine learning methods can also be used for data classification, regression and prediction of key performance indices of the process. With appropriate information integrations, model structure modifications and training improvements, both unsupervised learning and supervised learning methods can be made semi-supervised. Therefore, semi-supervised learning can be considered as a bridge connecting unsupervised learning and supervised learning.

Machine learning methods that have been widely used in the process industry are illustrated in Fig.2. Detailed literature reviews on specific topics are provided in the following subsections.

A. UNSUPERVISED LEARNING METHODS

As have been introduced, the unsupervised learning methods are used against data which have no labels. The main goal of the unsupervised learning method is to explore the data and find some hidden structure among them. For industry applications, this type of machine learning methods are mainly used for dimensionality reduction, information extraction, data visualization, density estimation, outlier detection, process monitoring, etc. Conventionally applied unsupervised learning methods in the process industry include principal component analysis, independent component analysis, k-means clustering, kernel density estimation, self-organizing map, Gaussian mixture models, manifold learning, support vector data description and so on. Detailed reviews of those representative unsupervised learning methods are provided as follows.

1) PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of correlated variables into linearly uncorrelated variables (principal components) [31]. Usually, the number of principal components is less than the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. Fig.3 provides an intuitive view of a three-dimensional data transformation example based on PCA. Compared to the original variables, principal components will not influence to each other due to the uncorrelated nature, which may greatly improve the efficiency for data analytics. Along the past 15 years, lots of applications of PCA have been made in the process industry, such as process monitoring, dimensionality reduction, data visualization, novelty detection, abnormality detection, outlier detection, etc.

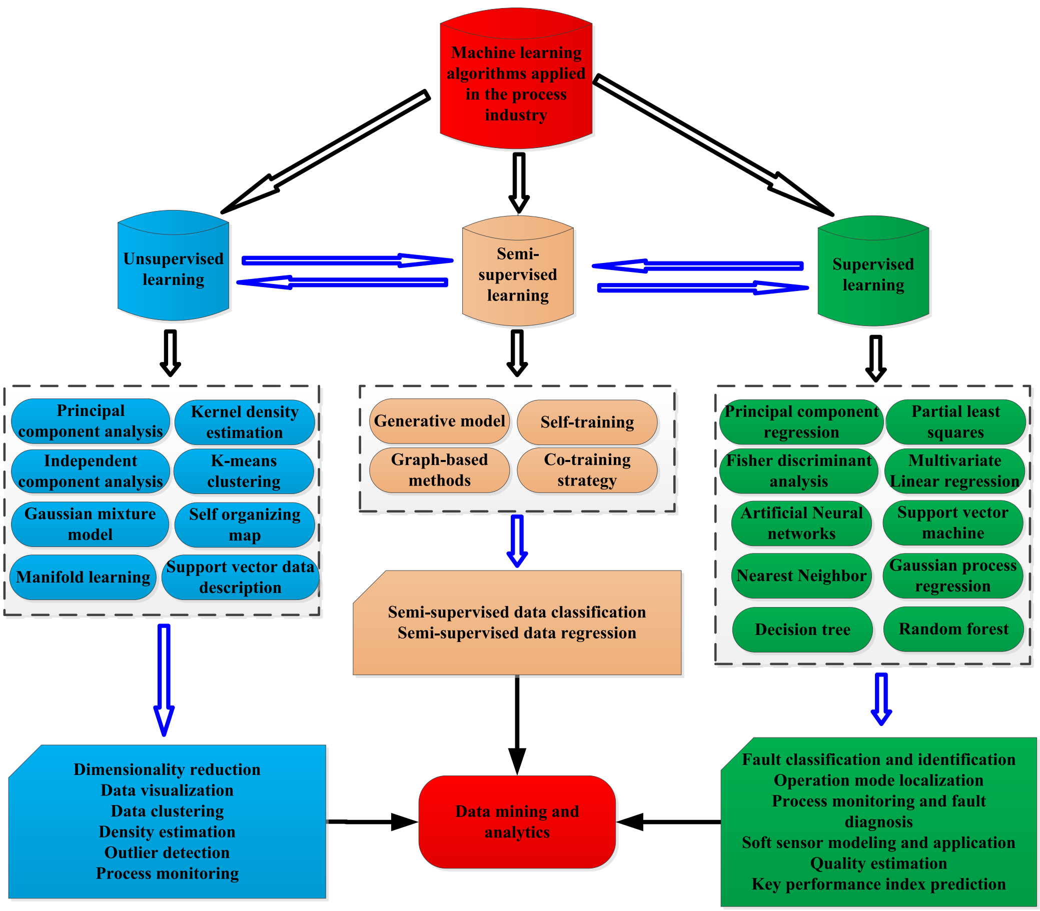


FIGURE 2. Machine learning methods applied in the process industry.

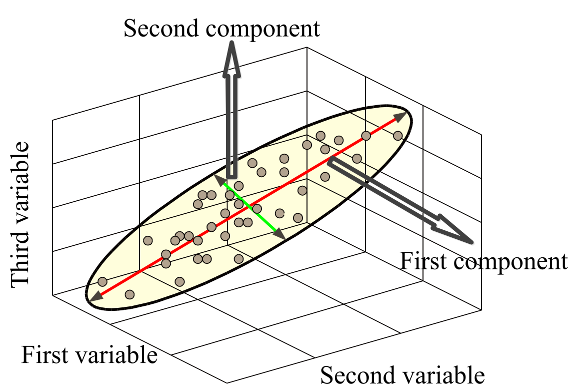


FIGURE 3. Illustration of the PCA method.

The most popular application of PCA has probably been made for the process monitoring purpose. This is because process monitoring has become a hot research spot since 2000, and PCA serves as a basic modeling tool. In order to meet the requirements for monitoring processes under different conditions, various modifications and improvements have been made for the PCA method, such as nonlinear PCA, multi-model/mixture PCA, adaptive/recursive PCA, dynamic PCA, multiway PCA, and so on [32]–[41]. Due to the limited space of the paper, only a small part of the related and relatively new literatures are provided in the reference list.

Dimensionality reduction is another interesting application topic of the PCA method, which is quite useful in the process industry. For high-dimensional correlated process variables, data information can be greatly compressed through dimensionality reduction by the PCA method, while the main data information can be well reserved. After the dimensionality of the process variables has been reduced, further data analytics could become much easier. For example, data visualization will be physically possible if the dimensionality of the data is reduced to 2 or 3 dimensions. In the past years, there are lots of applications of PCA that have been used for dimensionality reduction and data visualization [42]–[45].

Besides, the PCA method can be used for novelty detection, abnormality detection, and outlier detection, which are also common applications in the process industry. For example, a recent review of novelty detection was proved by Pimentel *et al.* [46]. A recursive dynamic PCA model has been developed for online novelty detection under drift conditions of gas sensor arrays [47]. An improved PCA method has been proposed for anomaly detection and applied to an emergency department [48]. An Improved Methodology has been formulated for Outlier Detection in dynamic industrial process datasets [49]. Besides, several robust PCA methods were developed and applied for outlier detection in the process industry [50]–[52].

2) INDEPENDENT COMPONENT ANALYSIS

Independent component analysis is an emerging technique for finding independent latent components from the measured variables, which is firstly proposed to handle the blind source separation problem [53]. The aim of ICA is to recover the real components from observed variables through optimization algorithms. Different from the PCA method, ICA tries to find those latent components that are both statistically independent and non-Gaussian. Therefore, ICA may reveal more meaningful information in the non-Gaussian data than PCA. Similar to the PCA method, ICA can also be used for various applications in the process industry, such as dimensionality reduction, non-Gaussian information extraction, process monitoring, data visualization, and so on.

ICA was first introduced for process monitoring and dimensionality reduction by Li and Wang [54] and Kano *et al.* [55]. Later, Lee *et al.* [56] developed a process monitoring method based on the ICA method, in which three statistics have been used for process monitoring. More recently, the ICA-based process monitoring method has been modified and improved, which can be adopted in more complex processes, such as time-varying processes, nonlinear processes, dynamic processes and batch process monitoring [57]–[63]. Other applications of the ICA method in the process industry include Albazzaz and Wang [64], Cai and Tian [65], Xia *et al.* [66], Peng *et al.* [67], Xu *et al.* [68], etc.

3) K-MEANS CLUSTERING

K-means clustering is a method of vector quantization, which was originally developed for signal processing [69]. To date, k-means clustering serves as a popular unsupervised machine learning method for cluster analysis in datasets. The aim of k-means clustering is to partition the data samples into k clusters in which each data sample belongs to the cluster with the nearest mean. The main applications of this method in the process industry are for dividing the process data into various operating modes, different fault types, or different grades of products. For example, k-means clustering was employed to derive the segmentation rule of variable subspace for batch process monitoring [70]; the K-means model was applied to build the hillslope prediction model in a wireless sensor network application [71]; a kernel k-means clustering based local support vector domain description method has been developed for fault detection of multimodal processes [72]; Zhao *et al.* [73] used the k-mean clustering algorithm for monitoring the offshore pipeline; Zhou *et al.* [74] combined the k-means clustering method and the PCA method for fault detection and identification in multiples processes; Tong *et al.* [75] developed an adaptive multimode process monitoring strategy based on the k-means clustering algorithm.

4) KERNEL DENSITY ESTIMATION

Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. In cases where the distributions of the data are not known which are usually non-Gaussian, the kernel density estimation is used as a fundamental data smoothing method which can provide an inference about the population, given a finite number of training data samples [76]. In the process industry, the kernel density estimation method has been used for estimating the distributions of process variables, monitoring statistics, or other related quantities that are used for describing the nature of the process. Here are some representative applications of the kernel density estimation method in the process industry. Chen *et al.* [77] proposed a regularized kernel density estimation method for clustered process data; Jiang and Yan [78] used the kernel density estimation approach in the weighted kernel PCA model for monitoring nonlinear chemical processes; He *et al.* [79] applied the KDE method for estimating the distribution of the froth color texture, in order to monitor the sulphur flotation process; a control-loop diagnosis approach has been developed by continuous evidence through kernel density estimation [80]; KDE was combined with the nearest neighbor method for process fault detection [81]; in the PCA-based control chart application, kernel density estimation has been used for multivariate non-normal distribution estimation [82]; Gonzalez *et al.* [83] combined kernel density estimation with Bayesian networking for process monitoring.

5) SELF-ORGANIZING MAP

Self-organizing map (SOM) is a type of artificial neural network (ANN) that is trained by using unsupervised learning approach, in order to produce a low-dimensional, discretized representation of the input space of the training samples [84]. A self-organizing map consists of components called nodes or neurons. Each node is associated with a weight vector of the same dimension as the input data vectors, and a position in the map space. Usually, the self-organizing map describes a mapping from a high-dimensional data space to a low-dimensional data space. The nature of the self-organizing map makes it possible to be utilized for various industrial applications, such as dimensionality reduction, data visualization, process monitoring, etc.

In the past years, lots of applications of the SOM method have been made to the process industry. Here are some examples. Corona *et al.* [85] used the SOM method for topological modeling and analysis of industrial process data. Merdun [86] made a SOM application in multidimensional soil data analysis. Dominguez *et al.* [87] carried out industrial process monitoring with SOM-based dissimilarity maps. Voyslavov *et al.* [88] applied the SOM method together with the Hasse diagram technique for surface water quality assessment. Tikkala and Jamsa-Jounela [89] used the SOM method for monitoring of caliper sensor fouling in a board machine. An improved SOM method was proposed for fault diagnosis for chemical industrial processes [90]. An advanced monitoring platform has been developed for industrial wastewater treatment, which is based on the SOM driven multivariate approach [91]. Yu *et al.* [92] developed a SOM-based fault diagnosis technique for non-Gaussian processes. The SOM method has been used for exploring the information in the soil databases [93]. A SOM-based topological preservation technique has been proposed for nonlinear process monitoring [94].

6) GAUSSIAN MIXTURE MODEL

Gaussian mixture model is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs [9]. For those processes which have multiple operating modes or aim to produce multiple grades of products, the Gaussian mixture model is particularly useful to characterize those multi-modal data natures. Besides, the Gaussian mixture model would also be helpful in those processes which have highly nonlinear relationships among process variables, or the distributions of some process variables are non-Gaussian. In those cases, the Gaussian mixture model serves as local linearization or local Gaussianity tools for data descriptions, based on which basic linear or Gaussian sub-models can then be developed for further data mining and analytics.

Similarly, Gaussian mixture model can also be used for general process applications, such as data clustering analysis, process monitoring, dimensionality reduction, data visualization, etc. Some recent examples of Gaussian mixture model applications in the process industry are given as follows. Choi *et al.* [95] used a Gaussian mixture model via principal component analysis and discriminant analysis for process monitoring; A Gaussian mixture model has been developed for multivariate statistical process control [96]; Yu and Qin [97] proposed a multimode process monitoring scheme based on Bayesian inference with finite Gaussian mixture models; Chen and Zhang [98] used the Gaussian mixture model for online multivariate statistical monitoring of batch processes; Yu [99] proposed a Gaussian mixture model and Bayesian method to incorporate both local and nonlocal information for monitoring the semiconductor manufacturing process; Zhu *et al.* [100] developed a robust mixture model for process monitoring; Wen *et al.* [101] combined the Gaussian mixture model with the canonical variate analysis method for monitoring dynamic multimode processes; Yan *et al.* [102] developed a semi-supervised mixture model for discriminant monitoring of batch processes; Yu [103] developed a particle filter driven dynamic Gaussian mixture model for monitoring and fault diagnosis of complex processes; Liu and Chen [104] employed the Gaussian mixture model to extract a series of operating modes from the historical process data, and then formulated a nonstationary fault detection and diagnosis method for multimode processes; Feital *et al.* [105] discussed the application of Gaussian mixture model for Modeling and Performance Monitoring of Multivariate Multimodal Processes; Yu *et al.* [106] combined the Gaussian mixture model with multiway independent component analysis for fault detection and diagnosis of multiphase batch processes.

7) MANIFOLD LEARNING METHODS

For dimensionality reduction of high-dimensional data for the process industry, the manifold learning method has recently been introduced. High-dimensional data requires more than two or three dimensions to represent, which is practically difficult to interpret. One approach to simplification is to assume that the data of interest lie on an embedded nonlinear manifold within the higher-dimensional space [107]. If the manifold is of low enough dimension, the data can be visualized in the low-dimensional space. Conventionally used manifold learning methods include principal curves, generative topographic mapping, Gaussian process latent variable model, maximum variance unfolding, isomap, locally linear embedding, laplacian eigenmaps, diffusion maps, neighborhood preserving embedding, locality preserving projections, etc. Based on the nonlinear dimensionality reduction procedure provided by the manifold learning method, further data analytics can be carried out, such as process monitoring, fault detection and fault diagnosis, soft sensor modeling and applications, and so on.

Shao and Rong [108] proposed a nonlinear process monitoring method based on the maximum variance unfolding projection method. Ge and Song [109] developed a nonlinear probabilistic process monitoring scheme based on generative topographic mapping, and also formulated a nonlinear probabilistic fault detection method by introducing the Gaussian process latent variable model [110]. Zhang *et al.* [111] developed a global-local structure analysis model and made an application for fault detection and identification in the process industry. Miao *et al.* [112] proposed a new fault detection method based on an improved neighborhood preserving embedding method. Tong and Yan [113] developed a statistical process monitoring scheme based on a multi-manifold projection algorithm. Liu *et al.* [114] extended the maximum variance unfolding method and used it for process monitoring and fault isolation. Miao *et al.* [115] proposed a nonlocal structure constrained neighborhood preserving embedding algorithm and used for fault detection in the process industry. Ma *et al.* [116] carried out fault detection via local and nonlocal embedding. Miao *et al.* [117] developed a data regression model based on the neighborhood preserving embedding method, and used it for soft sensor modeling and applications.

8) SUPPORT VECTOR DATA DESCRIPTION

The method of support vector data description is similar to one-class support vector machine, based on which the boundary of a dataset can be used to detect novel data or outliers [118]. SVDD obtains a spherically shaped boundary around a dataset, which defines a region for description of normal data samples. By introducing the kernel trick, the SVDD method can be made flexible to use various kernel functions and thus is able to describe highly nonlinear data. The main features of the SVDD method are demonstrated in Fig.4. According to the feature of the SVDD method, it can be used for novel detection, outlier detection, monitoring abnormality of data, etc. Along the past years, the SVDD method has been introduced into the process industry, and lots of applications of this method have been reported.

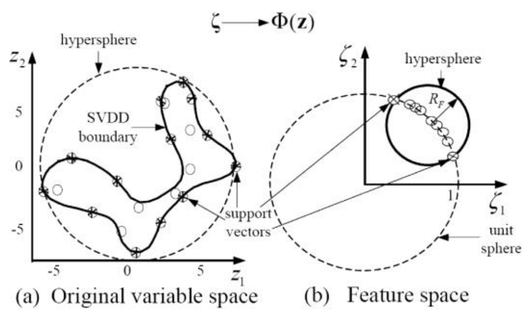


FIGURE 4. **Illustration of SVDD method [119].**

Liu *et al.* [119] used the SVDD method to characterize the non-Gaussian data information extracted by the ICA model, and defined the control limit of the developed statistic for monitoring non-Gaussian systems. A similar SVDD based non-Gaussian statistical modeling method has been proposed for process monitoring by Ge *et al.* [120]. Later, a new non-Gaussian fault reconstruction method has been formulated upon the SVDD model, which can be used for sensor fault identification and isolation in the process industry [121]. Ge *et al.* [122] extended the application of SVDD for batch process monitoring, and later boosted its performance by introducing the bagging strategy [123]. Liu *et al.* [124] improved the nonlinear PCA method with introduction of the SVDD model, and applied it for process monitoring. Zhu *et al.* [125] developed a multiclass SVDD model and combined it with a dynamic ensemble clustering method for transition process modeling and monitoring. Jiang and Yan [126] proposed a probabilistic weighted NPE-SVDD method for process monitoring. Yao *et al.* [127] developed a batch process monitoring based on functional data analysis and support vector data description. Du *et al.* [128] developed a monitoring scheme based on lazy learning, SVDD, and modified receptor density algorithm, and used it in nonlinear multiple modes processes.

9) SUMMARY

In summary, unsupervised learning methods have been widely used in the process industry in the past two decades. The most applications of the unsupervised method were made through the PCA model. Fig.5 provides a general application status about all of the 8 main unsupervised learning methods between 2000 and 2015, which is based on the key database in the web of science. It can be seen that the applications based on the PCA model accounts over half of the all applications. In recent years, with more and more manifold learning algorithms introducing into the process industry, the applications of these methods grow rapidly, which accounts for 12% of all industrial applications. Particularly, detailed applications of all unsupervised learning methods in terms of dimensionality reduction, outlier detection, process monitoring, and data visualization are shown in Fig.6. It can be seen that PCA plays as a dominant role in each application aspect. For dimensionality reduction, the manifold learning method has also served a lot of applications, due to its great ability in nonlinear dimensionality reduction. The self-organizing map method has provided a comparative application number in terms of data visualization, compared to PCA. This is mainly because self-organizing map describes a mapping from a high-dimensional data space to a low-dimensional data space, which makes data visualization more convenient for the process.

B. SUPERVISED LEARNING METHODS

Different from the unsupervised learning method, supervised learning deals with labeled data samples, discrete or continuous. When the label of the data sample is a discrete value, the supervised learning can be used for classification of process data, e.g. fault classification, or operating mode classification. Otherwise, when the label of the data sample is a continuous value, regression models can be constructed for prediction and estimation purposes. Main applications of the supervised machine learning method include process monitoring, fault classification and identification, online operating mode localization, soft sensor modeling and online applications, quality prediction and online estimation, key performance index prediction and diagnosis, etc. Some well used supervised learning methods in the process industry include principal component regression, partial least squares, fisher discriminant analysis, multivariate linear regression, neural networks, support vector machine, nearest neighbor, Gaussian process regression, decision tree, random forest, and so on. Detailed reviews of those representative supervised learning methods are provided as follows.

1) PRINCIPAL COMPONENT REGRESSION

Principal component regression (PCR) is a regression analysis technique that is based on principal component analysis method. In this method, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables which are extracted by the PCA method are used as regressors [129]. Therefore, the major use of the PCR method is similar to PCA, which lies in overcoming the multicollinearity problem which arises when two or more of the explanatory variables are close to being collinear. This is the common situation that happens in the process industry. As a result, both of the PCA and PCR methods have been widely used in the past years.

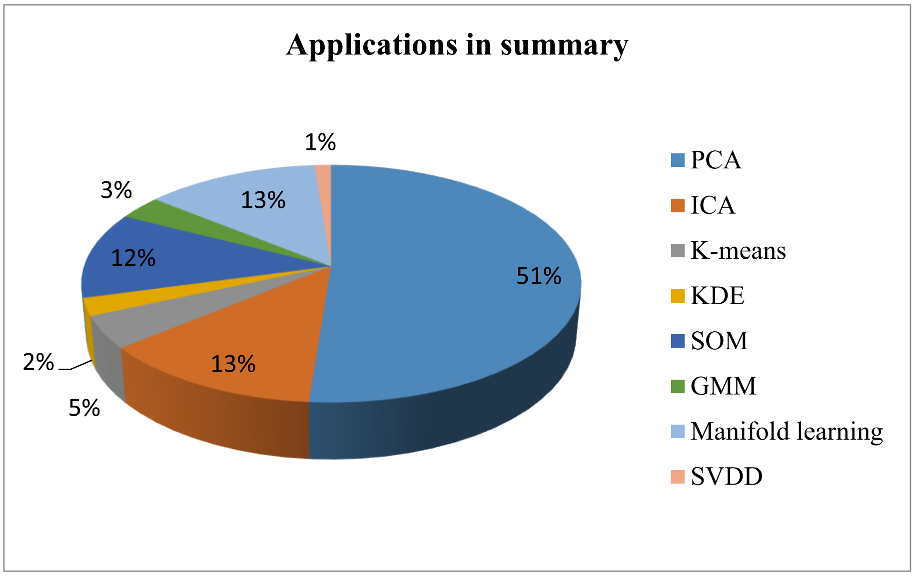


FIGURE 5. Application status of unsupervised learning methods.

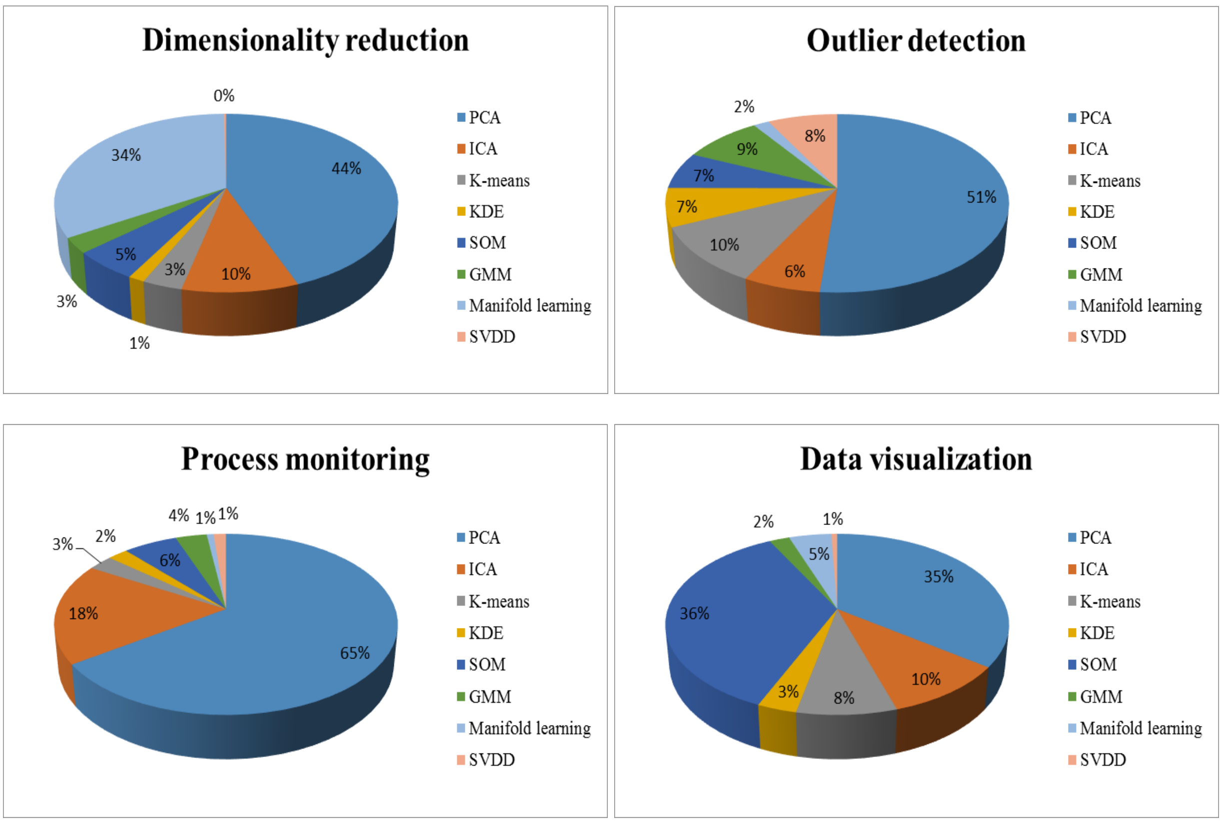


FIGURE 6. Application status of unsupervised learning methods in different aspects.

Compared to the PCA method, further information has been incorporated into the PCR method. Based on the regression modeling between the extracted principal components from PCA and the dependent variable, various applications have been made in the process industry, such as quality-related process monitoring, soft sensor modeling, online product quality prediction, etc.

Here are some typical and recent application examples in the process industry. Xie and Kalivas [130] developed local prediction models based on the PCR method.

Hartnett *et al.* [131] made online dynamic inferential estimation for the process by using the PCR model. Yang and Gao [132] developed a nonlinear form of the PCR model, and used it for online prediction and control for the product weigh in the injection molding process. PCR was combined with the residual analysis method for determination of multivariate concentration by Keithley *et al.* [133]. A mixture probabilistic form of the PCR model has been proposed for online soft sensor development and was applied to multimode processes [134]. Li *et al.* [135] used the PCR method for near-infrared spectroscopic calibration modeling and quality analysis. Yuan *et al.* [136] developed a locally weighted form the PCR model for online soft sensing of nonlinear time-varying processes. Lee *et al.* [137] applied the PCR method for identification of key factors which influenced primary productivity in two river-type reservoirs. Other related applications of the PCR methods include Khorasani *et al.* [138], Kolluri *et al.* [139], Popli *et al.* [140], Ghosh *et al.* [141], etc.

2) PARTIAL LEAST SQUARES

Partial least squares regression (PLS regression) is a statistical modeling method, which is quite similar to the PCR method. However, instead of finding hyperplanes of minimum variance between the response and independent variables, PLS finds a linear regression model by simultaneously projecting the predicted variables and the observable variables to the latent variable space [142]. Therefore, the predicted variables and the observable variables are connected through the latent variables. Fig.7 shows the main idea of the PLS method, where X and Y correspond to observable and predicted variables, T and U are components in the latent variable space, and P and Q are loading matrices of X and Y, respectively.

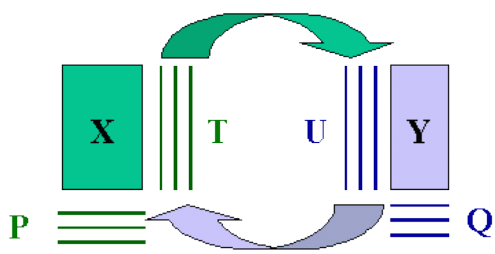


FIGURE 7. Illustration of the PLS method [143].

To date, PLS has been widely used in chemometrics and related areas, although the original applications of this method were in the social sciences. In the process industry, PLS has been used for quality-related process monitoring, fault classification, soft sensor development and online applications, product quality predictions, and so on. Here are some examples of the PLS method applied in the process industry. Kresta *et al.* [144] applied the PLS method for development of inferential process models. Kruger *et al.* [145] proposed an extended PLS approach for enhanced condition monitoring of industrial processes. Zhang and Zhang [146] proposed a modified PLS method and combined with the independent component regression model for monitoring complex processes. Yu [147] developed an adaptive kernel PLS regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes. Zhang *et al.* [148] proposed a kernel partial least squares model for nonlinear multivariate quality estimation and prediction. Zhang *et al.* [149] carried out quality prediction in complex processes by using a modified kernel PLS method. Ni *et al.* [150] proposed a localized and adaptive recursive PLS regression method for dynamic system modeling. Liu [151] developed a soft sensor based on sparse PLS with variable selection. Shao *et al.* [152] developed an online soft sensor design method by using local PLS models with adaptive process state partition. A MATLAB toolbox was recently been developed for class modeling based on one-class PLS classifiers [153]. Vanlaer *et al.* [154] used the PLS method for quality assessment of a variance estimator in prediction of batch-end quality. A multiway interval form of the PLS model was proposed for batch process performance monitoring by Stubbs *et al.* [155]. Godoy *et al.* [156] made several new contributions to nonlinear process monitoring through the use of the kernel PLS model. A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring has been carried out between the PLS method and other related approaches [157]. Zhou *et al.* [158] proposed a total projection form of the PLS model for the purpose of process monitoring. Qin and Zheng [159] formulated a concurrent form of the PLS model, and used it for quality-relevant and process-relevant fault monitoring. Peng *et al.* [160] developed a quality-related prediction and monitoring for multi-mode processes based on multiple PLS method, and applied it to an industrial hot strip mill.

3) FISHER DISCRIMINANT ANALYSIS

Fisher Discriminant Analysis (FDA) is a supervised linear dimensionality reduction technique designed for data classification. The basic idea of FDA is to seek a transformation matrix which maximizes the between-class scatter and minimizes the within-class scatter simultaneously [161]. Fig.8 provides an illustration of the FDA method, where Sw means within-class covariance and Sb is the between-class covariance. Due to the ability of the FDA method in dimensionality reduction and data classification, it has been widely used in the process industry. The method can be used directly for classification of different operating modes in the process, or to differentiate various faults that happen in the process. On the other hand, FDA can be applied as a dimensionality reduction step, which can efficiently extract data information that makes the subsequent data classification step more convenient.

In the past years, FDA has mainly been used for process monitoring, fault classification and fault diagnosis purposes in the process industry. For example, Chiang *et al.* [162]

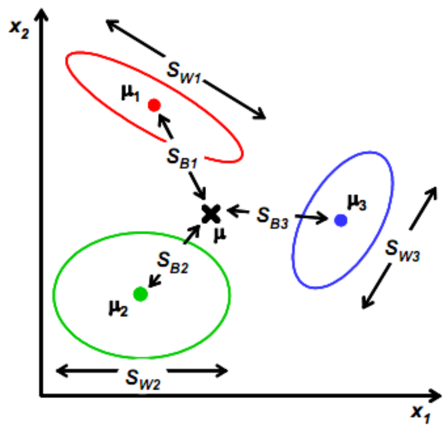


FIGURE 8. Illustration of the FDA method.

proposed a combined method with fisher discriminant analysis and support vector machines for fault diagnosis in industrial processes; Yu [163] applied the localized FDA method for monitoring complex chemical processes; Zhu and Song [164] developed a novel fault diagnosis system by using pattern classification on kernel FDA subspace; He *et al.* [165] proposed a new fault diagnosis method using fault directions in FDA; Zhang *et al.* [166] developed a nonlinear real-time process monitoring and fault diagnosis scheme based on PCA and kernel FDA; Yu [167] extended the localized FDA model to the multiway kernel form, and used it for nonlinear bioprocess monitoring; Huang *et al.* [168] made a mixture discriminant monitoring application based on the FDA method; Jiang *et al.* [169] combined the FDA method with canonical variate analysis, and developed a fault diagnosis approach; Zhu and Song [170] modified the basic kernel FDA model for imbalance datasets and used the developed method for fault diagnosis; Sumana *et al.* [171] proposed an improved fault diagnosis method by using dynamic kernel scatter-difference-based discriminant analysis; Zhao and Gao [172] developed a nested-loop FDA algorithm for industrial applications; Rong *et al.* [173] extended the FDA method for locality preserving discriminant analysis, and also developed its kernel form for fault diagnosis.

4) MULTIVARIATE LINEAR REGRESSION

Multivariate linear regression is a generalization of linear regression by considering more than one dependent variable [174]. In the process industry, there are lots of multivariate linear regression problems, such as soft sensor modeling between key indices/variables and ordinary process variables, quality prediction in the final product in batch processes, etc. In the past years, main applications of the multivariate linear regression methods are focused on soft sensor developments and prediction of product quality. For example, Park and Han [175] developed a nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns; Liu *et al.* [176] compared linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis-NIR diffuse reflectance spectroscopy; The performance of multivariate linear regression in PM10 forecasting has been compared with artificial neural networks by Caselli *et al.* [177]; Multivariate linear regression has been used for high-throughput quantitative biochemical characterization of algal biomass by NIR spectroscopy [178]; An evaluation has been made between multivariate linear regression and artificial neural networks in prediction of water quality parameters [179]; Aleixandre-Tudo *et al.* [180] applied the multivariate linear regression method to predict sensory quality in red wines; Different multivariate linear regression methods were compared in micro-Raman spectrometric quantitative characterization [181]. Besides, there are also some applications for process monitoring based on the multivariate linear regression methods, such as Amiri *et al.* [182], Noorossana *et al.* [183], Eyvazian *et al.* [184], etc.

5) ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANNs) are a family of models inspired by biological neural networks, which are presented as systems of interconnected "neurons" that exchange messages between each other [185]. The connections have numeric weights that can be tuned based on experience, making them adaptive to inputs and capable of learning. Like other machine learning methods, neural networks have been used to solve a wide variety of tasks in different application area. Theoretically, neural networks are able to approximate any linear/nonlinear function by learning from observed data. Along the past several decades, various types of artificial neural networks have been developed. Most of them were designed for supervised learning purposes, such as data classification and regression modeling. Fig.9 gives a description for a typical three layers neural network.

Due to the modeling ability of the artificial neural network, it has served as a basic tool for various applications in the process industry, such as process monitoring, fault classification, soft sensor modeling and applications, etc. Here, only a very small part of application examples are demonstrated. Lee *et al.* [186] applied a moving-window based adaptive neural network for modeling of a full-scale anaerobic filter process. Gonzaga *et al.* [187] developed a ANN-based soft sensor for real-time process monitoring and control of an industrial polymerization process. Radhakrishnan and Mohamed [188] used neural networks for the identification and control of blast furnace hot metal quality. Shi *et al.* [189] developed a ICA and multi-scale analysis based neural network modeling method for melt index prediction. Jamil *et al.* [190] proposed a generalized neural network and wavelet transform based approach for fault location estimation of a transmission line. Lliyas *et al.* [191] developed a RBF neural network inferential sensor for process emission monitoring. Nagpal and Brar [192] made a performance comparison among various artificial neural networks for fault classification. A probabilistic form of the neural network has been proposed for fault diagnosis of gas turbine [193].

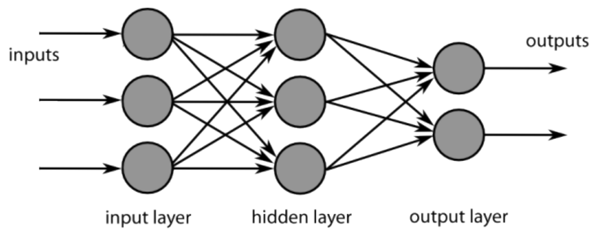


FIGURE 9. Structure of a three-layer neural network.

A feed-forward neural network with rolling mechanism and gray model has been developed for prediction of particular matter concentrations [194]. Sun *et al.* [195] developed a variable selection method for soft sensor by using neural network and nonnegative garrote. Xu and Liu [196] proposed a dynamic fuzzy neural network based method for melt index prediction purpose. A hybrid soft sensor for measuring hot-rolled strip temperature in the laminar cooling process has been developed by Pian and Zhu [197]. The resilient back propagation neural network has been combined with least square support vector regression for online monitoring and control of particle size in the grinding process [198]. A tool condition monitoring method has been proposed based on the neural networks by Liu and Jolley [199]. Gholami and Shahbazian [200] designed as soft sensor based on fuzzy C-Means and RFN_SVR and applied it in a stripper column.

6) SUPPORT VECTOR MACHINE

Support vector machine is a supervised machine learning method, which was originally developed by Vapnik in 1995 [201], [202], with associated learning algorithms that analyze data and recognize patterns, conventionally used for classification and regression analysis. Given a set of training examples, an SVM training algorithm builds a model that assigns new samples into one category or the other. The main idea of SVM is to construct a hyperplane or set of hyperplanes in a high or infinite-dimensional space, based on which different types of data samples can be well separated. When the target data is continuous, support vector regression model can be developed for regression analysis. In addition to performing linear classification/regression, SVMs can efficiently perform a non-linear classification/regression using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. More detailed descriptions of SVM classification and regression algorithms can be found in Vapnik [202], Schölkopf *et al.* [203], and Schölkopf and Smola [204].

Since SVM was proposed by Vapnik, it has become more and more popular in various application areas, due to its great modeling abilities in data classification and regression. Compared to other widely used supervised learning methods such as artificial neural networks, SVM may have better generalizations under many cases. In the process industry, particularly, SVM has also gained lots of successful applications, in terms of process monitoring, fault classification, soft sensor developments, product quality regression modeling and predictions, etc.

Due to the data classification ability of SVM, it has been used for process monitoring, fault classification and diagnosis in the past years. Due to the limited space, only some recent application examples are presented here. Zhang [205] proposed a fault detection and diagnosis method for nonlinear processes based on kernel ICA and SVM. Du *et al.* [206] used wavelet transform decomposition and multiclass support vector machines for recognition of concurrent control chart patterns. Tian *et al.* [207] applied SVM for fault diagnosis in steel plants. You *et al.* [208] used SVM for Laser Welding Process Monitoring and Defects Diagnosis. Xiao *et al.* [209] designed two methods for selecting Gaussian kernel parameters in one-class SVM and made an application to fault detection. SVM was used for monitoring continuous decision function in incipient fault diagnosis by Namdari and Jazayeri-Rad [210]. Multi class SVM has been used for simultaneous fault diagnosis in a Dew Point process [211]. A weighted form of the SVM method was developed for control chart pattern recognition [212]. Saidi *et al.* [213] applied higher order spectral features in SVM modeling and used it for bearing faults classification. Jing and Hou [214] developed PCA and SVM based fault classification approaches for complicated industrial processes. Wu *et al.* [215] carried out fault detection and diagnosis in process data by using SVM. Zhang *et al.* [216] used SVM for online monitoring of precision optics grinding.

On the other hand, SVM has also been used for regression modeling in the process industry, which is known as support vector regression (SVR). Some application examples for soft sensor developments and quality prediction purposes are given as follows. Jain *et al.* [217] used SVR to develop a soft sensor for a batch distillation column. Liu *et al.* [218] developed a soft chemical analyzer by using adaptive least-squares support vector regression with selective pruning and variable moving window size. Yu [219] proposed a Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. Ge and Song [220] compared the performance of SVR with other regression methods under the framework of just-in-time-learning. Liu *et al.* [221] proposed a just-in-time kernel learning method with adaptive parameter selection for soft sensor modeling of batch processes. Kaneko and Funatsu [222] made an application of online SVR model for soft sensor development. Liu and Chen [223] proposed an integrated soft sensor by using just-in-time support vector regression method and carried out probabilistic analysis for quality prediction of multi-grade processes. Kaneko and Funatsu [224] developed an adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants. Zhang *et al.* [225] made an online quality prediction for cobalt oxalate synthesis process using least squares support vector regression approach with dual updating. Wang and Liu [226] applied an adaptive mutation fruit fly optimization algorithm into the SVM model, and used it for melt index prediction. Jin *et al.* [227] developed a multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes. Shamshirband *et al.* [228] carried out an comparative study on sensor data fusion by SVR. Lv *et al.* [229] developed an adaptive least squares support vector machine model with a novel update for NOx emission prediction.

7) NEAREST NEIGHBORS

Nearest neighbors algorithm also known as k-nearest neighbors is a non-parametric method in machine learning, which can be used for both classification and regression [230]. In both cases, the input consists of the k closest training samples in the feature space. The output depends on whether the method is used for classification or regression. For data classification, the output is a class membership. A data sample is classified by a majority vote of its neighbors. If k is selected as 1, the output of the data sample will simply be assigned to the class of its nearest neighbor. For data regression, the output is a continuous value for the data sample. Normally, the output value for the predicted data sample is determined through averaging the values of its k nearest neighbors. In this method, the model structure is totally open to any type. For both classification and regression, NN or kNN has been considered as one of the simplest method of machine learning algorithms. Fig.10 gives examples of the NN method for data classification and regression.

In the past years, like other supervised machine learning algorithms, NN or kNN has also been used for various applications in the process industry, e.g. process monitoring, fault classification and soft sensor development. For example, Lee and Scholz [232] made a comparative study on Prediction of constructed treatment wetland performance with K-nearest neighbors and neural networks; Facco *et al.* [233] used the nearest neighbor method for the automatic maintenance of multivariate statistical soft sensors in batch processing; He and Wang [234] developed a statistical pattern analysis and process monitoring method based on the nearest neighbor method, and applied it in Semiconductor Batch Processes; Penedo *et al.* [235] carried out hybrid incremental modeling based on least squares and fuzzy K-NN for monitoring tool wear in turning processes; Kaneko *et al.* [236] developed a novel soft sensor for detecting completion of transition in industrial polymer processes; K-nearest neighbor was used for accelerating wrapper-based feature selection by Wang *et al.* [237]; Fuchs *et al.* [238] built nearest neighbor ensembles for interpretable feature selection of functional data; Su *et al.* [239] proposed a nonlinear fault separation method for redundancy process variables based on FNN in multi-kernel FDA Subspace; Li and Zhang [240] developed a diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection;

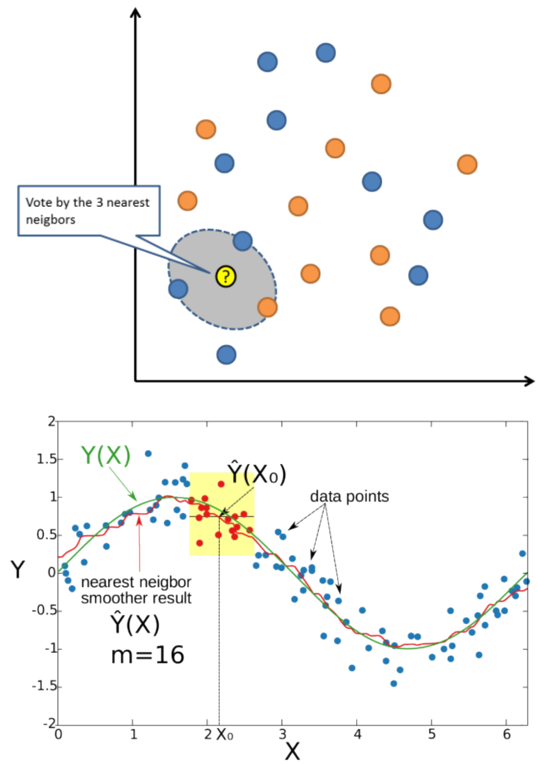


FIGURE 10. **Illustrations of NN for classification and regression [231].**

The k-nearest-neighbor method was used to improve the performance of fault diagnosis in batch processes [241].

8) GAUSSIAN PROCESS REGRESSION

The concept of Gaussian processes is named after Carl Friedrich Gauss, since it is based on the notion of the Gaussian distribution. Statistically, it can be seen as an infinite-dimensional generalization of multivariate normal distributions. In a Gaussian process, every point in the input space is associated with a normally distributed random variable. The distribution of a Gaussian process is the joint distribution of all those random variables, and thus it is actually a distribution over functions [242]. It has been demonstrated that a large class of methods will finally converged to an approximate Gaussian process. Inference of continuous values with a Gaussian process prior is known as Gaussian process regression. In this case, the Gaussian process can be used as a prior probability distribution over functions in Bayesian inference. Thus, given some available data samples, the distribution over functions can be inferred through Bayesian method, providing the posterior for the distribution. An illustration for the Gaussian process regression method is shown in Fig.11, where the left figure denotes samples from the prior distribution, and the right figure presents samples from posterior distribution through Bayesian method.

To date, the Gaussian process regression method has mainly been used for soft sensor modeling, quality-related monitoring and prediction purposes in the process industry.

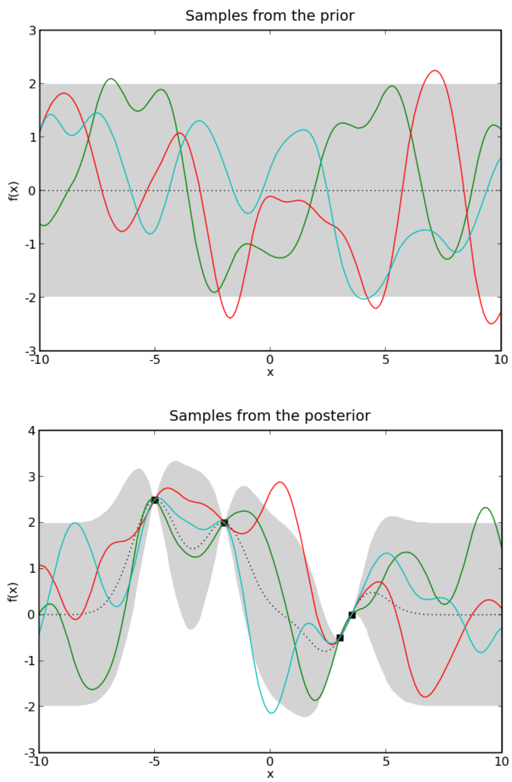


FIGURE 11. Illustration of Gaussian process regression method for both prior and posterior distributions [9].

Here are some application examples based on the Gaussian process regression method. A combined local Gaussian process regression model has been proposed for quality prediction of polypropylene production process [243]. Yu [244] developed an online quality prediction method for nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach. Song *et al.* [245] proposed an independent component regression-Gaussian process algorithm for real-time mooney-viscosity prediction of mixed rubber. Chen and Ren [246] introduced the bagging approach into Gaussian process regression model, in order to improve its performance. Grbic *et al.* [247] developed an adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. Chen *et al.* [248] proposed a multivariate video analysis and Gaussian process regression model based soft sensor for online estimation and prediction of nickel pellet size distributions. Zhou *et al.* [249] developed a recursive Gaussian process regression model for adaptive quality monitoring in batch processes. Auto-switch Gaussian process regression-based probabilistic soft sensors have been proposed for industrial multi-grade processes with transitions by Liu *et al.* [250]. Jin *et al.* [251] developed an adaptive soft sensor for nonlinear time-varying batch processes, which is based on an online ensemble Gaussian process regression model.

9) DECISION TREE

As its name indicates, a decision tree is a decision support tool that uses a tree-like graph or model to describe relationships among different variables and makes decisions. Decision trees are commonly used in operations research, particularly in decision analysis, in order to help identify a strategy that most likely to reach the aim [252]–[254]. Recently, decision tree has also been introduced into the process industry. The most common applications have been made in process monitoring, fault diagnosis, and quality prediction. For example, Kuo and Lin [255] used neural network and decision tree for prediction of machine reliability; Yeh *et al.* [256] used the decision algorithm to extract classification knowledge in mold tooling test; A fuzzy decision tree method was proposed by Zio *et al.* [257] for fault classification in the steam generator of a pressurized water reactor; Ma and Wang [258] carried out inductive data mining based on genetic programming for automatic generation of decision trees from data, and used it for process historical data analysis; He *et al.* [259] applied the decision tree learning technique for online monitoring and fault identification of mean shifts in bivariate processes; Demetgul [260] proposed a fault diagnosis method for production system with support vector machine and decision trees algorithms; An approach for automated fault diagnosis based on a fuzzy decision tree and boundary analysis of a reconstructed phase space has been proposed by Aydin *et al.* [261]; An improved decision tree construction method was developed and used for fault diagnosis in rotating machines, which is based on attribute selection and data sampling [262].

10) RANDOM FOREST

The general method of random forest was first proposed by Ho in 1995 [263]. It is a notion of the general technique of random decision forests, which can be used for classification, regression and other tasks. Usually, the method of random forest is carried out by constructing a multitude of decision trees based on the training dataset [264]. When used for new data samples, the method outputs the mode class (when it is used for classification) or the mean prediction value (when it is used for regression) of the individual trees. Similar to the decision tree method, random forest has also been introduced into the process industry. Here are some application examples. Pardo and Sberveglieri [265] used random forests and nearest shrunken centroids for the classification of sensor array data. Yang *et al.* [266] developed a random forest classifier for machine fault diagnosis. Auret and Aldrich [267] applied the random forest method for change point detection in multivariate time-series data. A random forest based unsupervised fault detection scheme has been developed by Auret and Aldrich [268]. Ahmad *et al.* [269] used the random forest method in gray-box modeling for prediction and control of molten steel temperature in tundish.

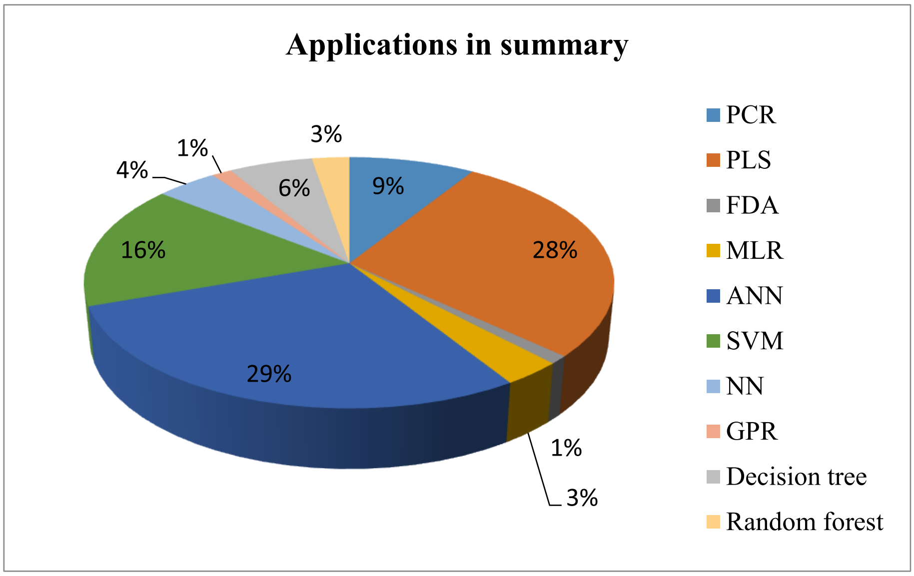


FIGURE 12. Application status of supervised learning methods.

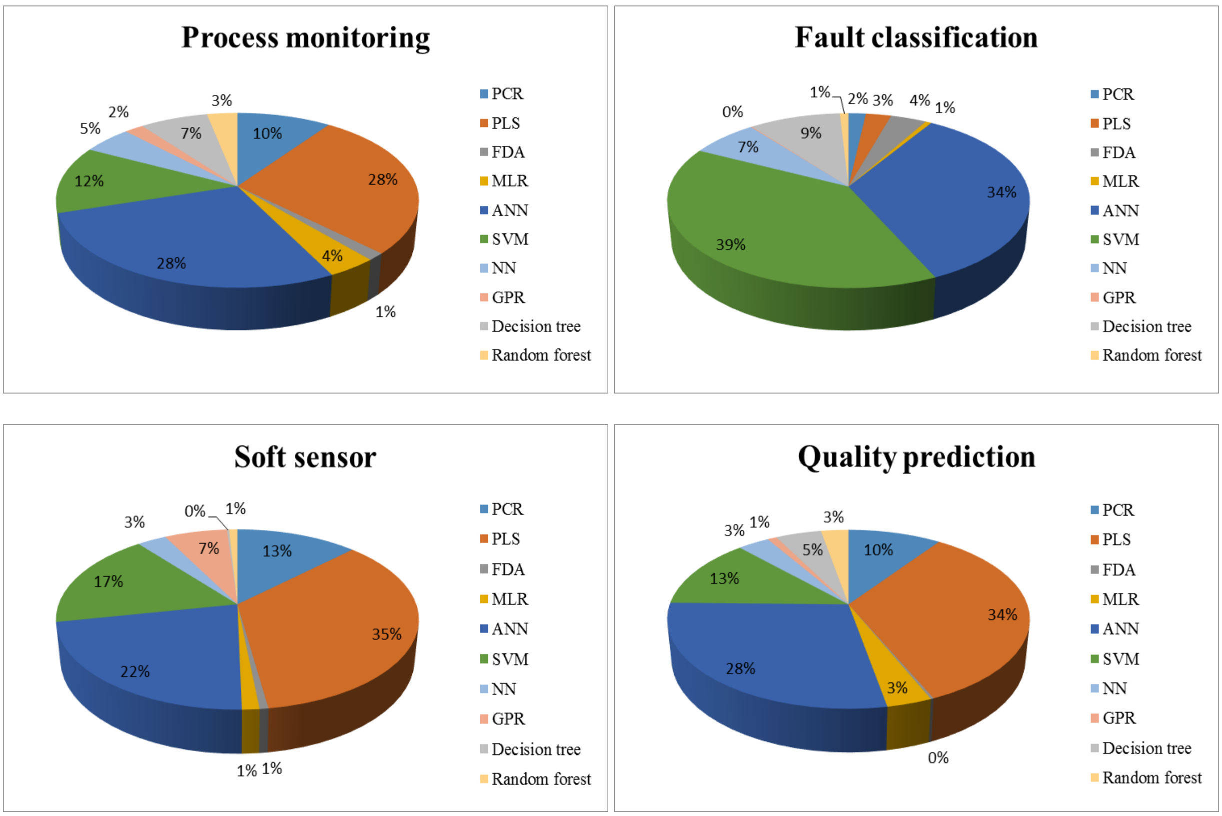


FIGURE 13. Application statuses of supervised learning methods in different aspects.

11) SUMMARY

To summarize the supervised machine learning methods used in the process industry since 2000, Fig.12 provides an overview of application status for 10 different supervised learning methods, the results of which are from the key database of Web of Science. As can be seen, artificial neural networks and partial least squares are two most widely used methods in the past 15 years, which account for over half applications in the process industry. More specifically, application statuses of all those supervised learning methods in terms of process monitoring, fault classification, soft sensor and quality prediction are presented in Fig.13. For the purpose of process monitoring, PLS and ANN have played dominant roles. Due to the great ability of SVM in data classification, it has gained the most applications in fault classification. For soft sensor developments and quality predictions, PLS and ANN are also the two most popular methods which occupy about 60% of all applications.

C. SEMI-SUPERVISED LEARNING METHODS

For applications in the process industry, the semi-supervised learning method is particularly useful when the cost of labeling data samples is expensive or time consuming. For example, assigning the type of detected faulty data samples is a difficult task, which may require process knowledge and experiences of process engineers, thus will be costly and may need a lot of time. As a result, there are only a small number of faulty samples that are assigned to their corresponding types. Although most faulty samples are unlabeled, they still contain important information of the process. If those unlabeled samples can be used in a good way, the efficiency of the fault classification system could be greatly improved. In this case, semi-supervised learning with both labeled samples and unlabeled samples will provide an improved fault classification model, compared to the model that only depends on a small part of labeled data samples. Another example is for quality prediction, in which a regression model needs to be developed between the ordinary process variables and quality variables. Compared to the ordinary process variable, the quality variable is much more difficult to obtain. Here, the quality variable serves as the label for the ordinary process variable. While the samples for ordinary process variables are quite easy to obtain from the process, acquirements of quality related variables are much more difficult, which are usually resorted to lab analysis, expensive analytical tools, additional human efforts, etc. As a result, the input and output samples of the regression model are imbalanced, which means there are a lot of missing output values in the historical database. Similar to the case in fault classification, those samples which only have input values also contain important information of the process. From a probabilistic modeling perspective, the distribution of the input variables will become more accurate if more sampled data are used for modeling, which may then provide a positive support to the predictive modeling performance for the quality variable. Semi-supervised learning in this case is to construct the regression model based on a small number of quality data samples and a large number of data samples from ordinary process variables. Compared to the supervised learning method, those additional unlabeled data samples may greatly improve the regression performance of the prediction model.

Compared to unsupervised and supervised learning, the application of semi-supervised learning has not received too much attention in the process industry until in the recent several years. Represented semi-supervised learning schemes include generative model based method, self-training, co-training, graph-based method, etc [270], [271]. There are already some semi-supervised application examples in both data classification and regression. Ge and Song [272] proposed a semi-supervised Bayesian method for soft sensor development, which can successfully incorporate the information from unlabeled data. Shao *et al.* [273] proposed a Bayesian method for multirate data synthesis and model calibration. Ji *et al.* [274] developed a recursive weighted kernel regression for semi-supervised soft sensing modeling of fed-batch processes. Ge *et al.* [275] introduced the self-training strategy for statistical quality prediction of batch processes with limited quality data. Deng and Huang [276] developed an identification method for nonlinear parameter varying systems with missing output data, which is based on a particle filter under the framework of the expectation-maximizaiton algorithm. Zhong *et al.* [277] proposed a semi-supervised fisher discriminant analysis model for fault classification in industrial processes. Yan *et al.* [102] proposed a semi-supervised mixture discriminant monitoring method for batch processes. Jin *et al.* [278] developed a multiple model based LPV soft sensor under the condition of irregular and missing process output measurements. Ge *et al.* [279] proposed a mixture semi-supervised principal component regression model for soft sensor application in the process industry, and later extended it to the nonlinear form [280]. Raju and Cooney [281] developed an active learning method for process data modeling and applications. Ge [282] also developed an active learning strategy for smart soft sensor development under a small number of labeled data samples. Zhou *et al.* [283] proposed a semi-supervised PLVR models for process monitoring with unequal sample sizes of process variables and quality variables. Zhu *et al.* [284] proposed a robust form of mixture semi-supervised principal component regression model for soft sensor applications. Bao *et al.* [285] combined the co-training strategy with PLS for development of a semi-supervised soft sensor.

Due to the space of this paper, lots of machine learning methods have not been mentioned here, although they are also quite important for data mining and analytics in the process industry. For example, the Markov model is a very important class of machine learning method, which has been used for dynamic modeling of process data. To date, lots of applications has already been made in the process industry. Several regularization techniques such as L2 (Tikhonov) and L1 (lasso) have also been applied for data mining and analytics in the past decades. Like Markov model, linear dynamic systems is also a class of dynamic machine learning methods, which can be used for dynamic process data modeling and analytics. Besides, ensemble learning methods such as bagging and boosting have also been introduced into the process industry in the past years.

IV. PERSPECTIVES FOR FUTURE RESEARCH

Due to the fast development of new computing technologies, machine learning today is quite different from its past. The ever increased power of modern computers has also helped data mining and analytics techniques make better use of machine learning algorithms. For example, neural networks have long been used in data mining and analytics applications. However, the computational burden could be a problem if neural networks are used to handle big computing problems. With more computing power, we can create deep neural networks with many layers, which enable fast processing and automated learning of training dataset. Another example is about the big data problem, which has recently become very popular in various areas. While many machine learning algorithms have been proposed for a long time, the ability to handle big data is a recent development, e. g. self-driving Google car, online recommendation offers like those from Amazon, knowing what customers are saying about you on Twitter, etc.

For the applications of machine learning in the process industry, increasingly more new developed machine learning algorithms have been introduced. However, different from other application areas of machine learning, the process industry has its own nature which should be taken into account for promoting the grade of data mining and analytics. How to efficiently utilize the knowledge discovered from data and transfer it to effective decisions need more re-investigations under emergence of new industrial natures. Also, some fundamental issues related to machine learning based data mining and analytics such as data quality evaluation, data acquirements, model selection should be reconsidered and paid more attentions. The followings are some perspectives that we believe may lead to more future researches on the topic of data mining and analytics in the process industry.

A. MACHINE LEARNING OF BIG DATA IN THE PROCESS INDUSTRY

With the development of modern process industry, the data volume becomes much larger, which poses great challenges to information capture, data management and storage. More importantly, it becomes difficult to efficiently interpret the information hidden within those data. With more measuring devices introduced into the process industry, different types of structured and unstructured data have been collected, such as sensor data, pictures, videos, audios, log files and so on. Qin [286] details the process data analytics in modern industrial processes in the era of big data. Usually, big data refers to the size and variety of data sets that challenge the ability of traditional software tools to capture, store, manage, and analyze. Often four V's are used to characterize the essence of big data [287]–[289], termed as Volume, Variety, Value, and Velocity. According to this definition, the modern process industry is actually driven by big data.

In the era of big data, machine learning has also faced a big challenge that has never met before [290], [291]. To date, lots of strategies have been proposed to address the machine learning problems for big data, among which the distributed learning strategy is a very attractive one [291]–[295]. For the application of machine learning in the process industry, the distributed learning strategy is particularly useful. Distributed natures of operating units, process equipment, and manufacturing plants have provided a basic application environment for the distributed machine learning strategy. Recently, a distributed modeling framework has been proposed for plant-wide process monitoring [296]. In this framework, the whole industrial process is divided into various blocks, in which machine learning algorithms are employed for sub-model developments. Then, the final decision is made through fusing the results obtained from different blocks. Besides, several research challenges and promising issues have also been discussed in this paper. For distributed machine learning in the process industry, the clouding computing technology may provide an efficient tool to facilitate data mining and analytics in large-scale/plant-wide process industries. Cloud computing solutions would not only handle huge amounts of process data but also minimize operational costs.

B. SUSTAINABILITY DRIVEN DATA MINING AND ANALYTICS

While the process industry has gained lots of developments in the past several decades, the sustainability problem has recently received great attentions, not only in energy efficiency research, but also for those issues related to environment pollution and protection, such as gas emission monitoring, sewage treatment and pollution analytics, etc [297], [298]. Both of the energy efficiency and environmental sustainability of the process industry depend on optimal operations, which require advanced process control, monitoring and optimization. To this end, deep mining and analytics of process data may discover more useful knowledge, making the process more understandable and efficient for sustainability analyses. As a result, the process could be able to forecast the abnormal events which may cause sustainability degradation, improve the energy efficiency throughout the whole process, prediction and diagnose of key performance indices related to process safety, environment pollution, energy consumption, etc.

To carry out those advanced data mining and analytics in the process industry, machine learning always plays an important role in model training and prediction. First of all, machine learning can serve as a predictive modeling tool for advanced process and quality control. Those machine learning algorithms which can efficiently describe dynamic relationships among process data are particularly useful to real-time process control and optimization. Second, hierarchical machine learning algorithms can be employed for multi-level modeling of different key performance indices, such as operating performance, energy consumption, gas emission, recycle efficiency, and so on. Based on this developed model in different levels of the process, the sustainability indices can be connected together throughout the whole process. Therefore, the sustainability of the process can be predicted and evaluated through observing some easy-to-measure process variables. Any performance degradation of the process sustainability can then be detected, and the root causes of which could be located for further investigations and improvements. Besides, machine learning algorithms can also be employed for sustainability performance evaluation and prediction along the evolving of the process industry from past, present to future.

C. PROCESS CAUSALITY MODELING AND ANALYTICS

Like other complex systems such as biological systems and social networks, the modern process industry is also featured by connecting different elements in terms of operating units, process equipment, or different sensors/measurement devices. The whole process behavior is determined by the relationships among different process elements as well as the local behavior within each element. For effective data mining and analytics, it is essential to identify connection relationships or process causality among different elements. For fault/abnormality detection and diagnosis, identification of process causality is particularly important, since the abnormality can easily propagate among various process units. Based on the effective identification of process causality, root causes of the abnormality can be located more quickly and efficiently as well as the track to its propagation path among the process.

The process causality can be captured in two ways, namely knowledge-based and data-based methods [299]. However, for large-scale modern industrial processes, identification of process causality could be a very troublesome task, which in most time may be impossible to be captured from process knowledge or experiences from engineers. Instead, through effective mining and analytics from history data, the process causality may also be captured, which is relatively much easier for practical application. To date, several methods have been introduced or proposed for causality identification in the process industry, such as Granger causality, frequency domain methods, cross-correlation analysis, information-theoretic methods, Bayesian networks and so on [300]–[305]. However, there are still some open issues that require more deep investigations. As an extension of Bayesian networks from the view of machine learning, the probabilistic graphic modeling approach may provide a more general way for description of process causality, based on which more advanced data mining and analytics could be expected.

D. DATA CLEANING AND QUALITY EVALUATION

This is a quite fundamental step for data mining and analytics in the process industry, since the quality of data determines the effectiveness of the following machine learning model development step. If the quality of the process data cannot be well guaranteed, even a best machine learning algorithm could generate a misleading model. Therefore, before applying the machine learning algorithms for model development, process data should be cleaned, and its quality needs to be evaluated. More specifically, the missing data should be addressed appropriately; the outliers need to be detected; the problem of different sampling rates among process variables needs to be taken into account. Besides, noises from various measurement devices and process itself should also be well considered during this step, e. g. removal, reduced, well modeled, etc. In summary, how to improve the quality of data is a critical step for machine learning, data mining and analytics for knowledge discovery.

To date, researches have been carried out on specific issues for improvements of data quality, e.g. missing data, outlier, multi-rate sampling, etc [100]–[306]. Recently, Xu *et al.* [17] carried out an review on data cleaning in the process industry, in which different data cleaning methods have been introduced, including missing data imputation, outlier detection, noise removal and time alignment. Several future research directions on data cleaning have also been highlighted. Here, an open question is how to evaluate the quality of the process data? This is actually related to the aim of machine learning and data mining. For example, if we intend to implement a process monitoring method, outlier detection is particularly important since it will greatly deteriorate the performance of the monitoring algorithm; if we want to develop a soft sensor for predictions of key process variables, the problems of missing values and multiple sampling rates need to be taken good care of, since they play critical roles in the development of soft sensor model. To our best knowledge, those methods which can simultaneously handle several data cleaning issues have rarely been reported. Due to the advantages in dealing with irregular data, the robust machine learning algorithms may be considered as a promising way for data cleaning and quality improvement, thus should be more introduced for data mining and analytics in the process industry. Also, how to combine the data cleaning method with online machine learning algorithms is also an interesting issue for future research, as indicated in Xu *et al.* [17].

E. SENSOR NETWORK DESIGN FOR THE PROCESS INDUSTRY

For data acquirement from the process, the design of sensor network shall be considered as an important task. This is because the sensors serve as eyes of the data system to observe the whole process. Therefore, the quality and completeness of the process data both depends on the design of sensor network. With the proposal of industry 4.0 in the past several years, the concepts of industrial internet, cyber physical systems and internet of things become more and more popular. In the same time, more and more wireless sensing technologies have been introduced into the process industry. As can be expected, most existing process sensors may be replaced by wireless sensors in the near future. For design of sensor network in the process industry, a lot of methods that have been developed in the areas of wireless sensor networks, industrial internet, cyber physical systems and internet of things can be referred and employed for utilization [316]–[318].

Based on a well-designed sensor network, the information of the process can be efficiently captured and transferred to useful knowledge by machine learning, data mining and analytics. Simultaneously, the obtained topology of sensor network can also be used for capturing the process causality, based on which the process behavior can be deeply described and analyzed. In the current context of industrial internet/industrial 4.0, data mining and analytics driven by new emergences of machine learning technologies may be promoted to a new level.

V. CONCLUSIONS AND REMARKS

In this paper, methodologies of data mining and analytics in the process industry are introduced, with a systematical review through the viewpoint of machine learning. It is no doubt that machine learning plays a key role in data mining and analytics in the process industry. While both unsupervised and supervised machine learning methods have already been widely used in the process industry, which approximately accounts for 90%-95% of all applications, the semi-supervised machine learning has been introduced in recent years, thus its application will become more popular in the near future. A total of 8 specific unsupervised learning methods and 10 supervised learning methods have been introduced in detail, with corresponding reviews in data mining and analytics in the process industry. Furthermore, several perspectives have been illustrated and discussed for future research on this topic. It can be expected that data mining and analytics will play increasingly important roles in the process industry, with the development of new machine learning technologies.

However, it should be noted that data mining and analytics is quite a multi-disciplinary research area, which may require expertise from process control, chemometrics, machine learning, pattern recognition, statistics, etc. In order to make efficient data mining and analytics in practice, barriers among various disciplines will have to be eliminated, new educational and training strategies will need to be developed for industry and the level of cooperation between universities and the industrial sectors will need to be increased as well. In this respect, the role of data mining and analytics is to bring researchers and practitioners from different cultures to work together.

REFERENCES

[1] M. Kano and M. Ogawa, "The state of the art in chemical process control in Japan: Good practice and questionnaire survey," *J. Process Control*, vol. 20, no. 9, pp. 969–982, Oct. 2010.

[2] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.

[3] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 327–346, 2003.

[4] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, 2013.

[5] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.

[6] S. Khatibisepehr, B. A. Huang, and S. Khare, "Design of inferential sensors in the process industry: A review of Bayesian methods," *J. Process Control*, vol. 23, no. 10, pp. 1575–1596, Nov. 2013.

[7] J. MacGregor and A. Cinar, "Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods," *Comput. Chem. Eng.*, vol. 47, pp. 111–120, Dec. 2012.

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

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[10] P. Simon, *Too Big to Ignore: The Business Case for Big Data*. Hoboken, NJ, USA: Wiley, 2013.

[11] T. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[12] X. Z. Wang, *Data Mining and Knowledge Discovery for Process Monitoring and Control*. London, U.K.: Springer-Verlag, 2012.

[13] L. H. Chiang, R. D. Braatz, and E. L. Russell, *Fault Detection and Diagnosis in Industrial Systems*. New York, NY, USA: Springer-Verlag, 2001.

[14] U. Kruger and L. Xie, *Statistical Monitoring of Complex Multivariate Processes*. West Sussex, U.K.: Wiley, 2012.

[15] Z. Ge and Z. Song, *Multivariate Statistical Process Control: Process Monitoring Methods and Applications*. London, U.K.: Springer-Verlag, 2013.

[16] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London, U.K.: Springer-Verlag, 2014.

[17] S. Xu *et al.*, "Data cleaning in the process industries," *Rev. Chem. Eng.*, vol. 31, no. 5, pp. 453–490, 2015.

[18] A. Kassidas, J. F. MacGregor, and P. A. Taylor, "Synchronization of batch trajectories using dynamic time warping," *AIChE J.*, vol. 44, no. 4, pp. 864–875, 1998.

[19] H. Liu, S. Shah, and W. Jiang, "On-line outlier detection and data cleaning," *Comput. Chem. Eng.*, vol. 28, no. 9, pp. 1635–1647, 2004.

[20] S. A. Imtiaz and S. L. Shah, "Treatment of missing values in process data analysis," *Can. J. Chem. Eng.*, vol. 86, no. 5, pp. 838–858, 2008.

[21] P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor, "Missing data methods in PCA and PLS: Score calculations with incomplete observations," *Chemometrics Intell. Lab. Syst.*, vol. 35, no. 1, pp. 45–65, 1996.

[22] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 51, no. 1, pp. 81–93, 2000.

[23] S. W. Choi, J. Morris, and I.-B. Lee, "Dynamic model-based batch process monitoring," *Chem. Eng. Sci.*, vol. 63, no. 3, pp. 622–636, 2008.

[24] T. Kourti, J. Lee, and J. F. MacGregor, "Experiences with industrial applications of projection methods for multivariate statistical process control," *Comput. Chem. Eng.*, vol. 20, no. 1, pp. S745–S750, 1996.

[25] N. F. Thornhill, J. W. Cox, and M. A. Paulonis, "Diagnosis of plant-wide oscillation through data-driven analysis and process understanding," *Control Eng. Pract.*, vol. 11, no. 12, pp. 1481–1490, 2003.

[26] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 50, no. 2, pp. 243–252, Mar. 2000.

[27] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, 2003.

[28] A. Kassidas, P. A. Taylor, and J. F. MacGregor, "Off-line diagnosis of deterministic faults in continuous dynamic multivariable processes using speech recognition methods," *J. Process Control*, vol. 8, pp. 381–393, Oct./Dec. 1998.

[29] T. Chen, J. Morris, and E. Martin, "Probability density estimation via an infinite Gaussian mixture model: Application to statistical process monitoring," *J. Roy. Statist. Soc. C-Appl. Statist.*, vol. 55, pp. 699–715, Nov. 2006.

[30] N. F. Thornhill, M. A. A. S. Choudhury, and S. L. Shah, "The impact of compression on data-driven process analyses," *J. Process Control*, vol. 14, pp. 389–398, Jun. 2004.

[31] I. T. Jolliffe, *Principal Component Analysis* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[32] S. J. Qin, "Statistical process monitoring: Basics and beyond," *J. Chemometrics*, vol. 17, pp. 480–502, Aug./Sep. 2003.

[33] S. W. Choi, E. B. Martin, and A. J. Morris, "Fault detection based on a maximum-likelihood principal component analysis (PCA) mixture," *Ind. Eng. Chem. Res.*, vol. 44, no. 7, pp. 2316–2327, 2005.

[34] U. Kruger, D. Antory, J. Hahn, G. W. Irwin, and G. McCullough, "Introduction of a nonlinearity measure for principal component models," *Comput. Chem. Eng.*, vol. 29, pp. 2355–2362, Oct. 2005.

[35] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, ''Nonlinear process monitoring using kernel principal component analysis,'' *Chem. Eng. Sci.*, vol. 59, no. 1, pp. 223–234, 2004.
[36] N. F. Thornhill, S. L. Shah, B. Huang, and A. Vishnubhotla, ''Spectral principal component analysis of dynamic process data,'' *Control Eng. Pract.*, vol. 10, no. 8, pp. 833–846, 2002.
[37] C.-Y. Cheng, C.-C. Hsu, and M.-C. Chen, ''Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes,'' *Ind. Eng. Chem. Res.*, vol. 49, no. 5, pp. 2254–2262, 2010.
[38] K. De Ketelaere, M. Hubert, and E. Schmitt, ''Overview of PCA-based statistical process monitoring methods for time-dependent, high-dimensional data,'' *J. Quality Technol.*, vol. 47, pp. 318–335, Jan. 2015.
[39] Y. Guo, K. Li, Z. Yang, J. Deng, and D. M. Laverty, ''A novel radial basis function neural network principal component analysis scheme for PMU-based wide-area power system monitoring,'' *Electr. Power Syst. Res.*, vol. 127, pp. 197–205, Oct. 2015.
[40] Y. Zhang, S. Li, and Y. Teng, ''Dynamic processes monitoring using recursive kernel principal component analysis,'' *Chem. Eng. Sci.*, vol. 72, pp. 78–86, Apr. 2012.
[41] M. Yao and H. Wang, ''On-line monitoring of batch processes using generalized additive kernel principal component analysis,'' *J. Process Control*, vol. 103, pp. 338–351, Apr. 2015.
[42] G. Ivosev, L. Burton, and R. Bonner, ''Dimensionality reduction and visualization in principal component analysis,'' *Anal. Chem.*, vol. 80, no. 13, pp. 4933–4944, 2008.
[43] R. C. Wang, T. F. Edgar, M. Baldea, M. Nixon, W. Wojsznis, and R. Dunia, ''Process fault detection using time-explicit Kiviat diagrams,'' *AIChE J.*, vol. 61, pp. 4277–4293, Dec. 2015.
[44] R. Dunia, T. F. Edgar, and M. Nixon, ''Process monitoring using principal components in parallel coordinates,'' *AIChE J.*, vol. 59, no. 2, pp. 445–456, 2013.
[45] A. Prieto-Moreno, O. Llanes-Santiago, and E. García-Moreno, ''Principal components selection for dimensionality reduction using discriminant information applied to fault diagnosis,'' *J. Process Control*, vol. 33, pp. 14–24, Sep. 2015.
[46] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, ''A review of novelty detection,'' *Signal Process.*, vol. 99, pp. 215–249, Jun. 2014.
[47] A. Perera, N. Papamichail, N. Barsan, U. Weimar, and S. Marco, ''On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions,'' *IEEE Sensors J.*, vol. 6, no. 3, pp. 770–783, Jun. 2006.
[48] F. Harrou, F. Kadri, S. Chaabane, C. Tahon, and Y. Sun, ''Improved principal component analysis for anomaly detection: Application to an emergency department,'' *Comput. Ind. Eng.*, vol. 88, pp. 63–77, Oct. 2015.
[49] S. Xu, M. Baldea, T. F. Edgar, W. Wojsznis, T. Blevins, and M. Nixon, ''An improved methodology for outlier detection in dynamic datasets,'' *AIChE J.*, vol. 61, no. 2, pp. 419–433, 2015.
[50] M. Daszykowski, K. Kaczmarek, Y. Heyden, and B. Walczak, ''Robust statistics in data analysis—A review basic concepts,'' *Chemometrics Intell. Lab. Syst.*, vol. 85, no. 2, pp. 203–219, 2007.
[51] L. H. Chiang, R. J. Pell, and M. B. Seasholtz, ''Exploring process data with the use of robust outlier detection algorithms,'' *J. Process Control*, vol. 13, no. 5, pp. 437–449, 2003.
[52] T. Chen, E. Martin, and G. Montague, ''Robust probabilistic PCA with missing data and contribution analysis for outlier detection,'' *Comput. Statist. Data Anal.*, vol. 53, no. 10, pp. 3706–3716, 2009.
[53] A. Hyvärinen and E. Oja, ''Independent component analysis: Algorithms and applications,'' *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, 2000.
[54] R. F. Li and X. Z. Wang, ''Dimension reduction of process dynamic trends using independent component analysis,'' *Comput. Chem. Eng.*, vol. 26, no. 3, pp. 467–473, 2002.
[55] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, ''Monitoring independent components for fault detection,'' *AIChE J.*, vol. 49, no. 4, pp. 969–976, Apr. 2003.
[56] J.-M. Lee, C. Yoo, and I.-B. Lee, ''Statistical process monitoring with independent component analysis,'' *J. Process Control*, vol. 14, no. 5, pp. 467–485, 2004.
[57] Z. Ge and Z. Song, ''Process monitoring based on independent component analysis-principal component analysis (ICA–PCA) and similarity factors,'' *Ind. Eng. Chem. Res.*, vol. 46, no. 7, pp. 2054–2063, 2007.
[58] Y. Zhang and S. Qin, ''Fault detection of nonlinear processes using multiway kernel independent component analysis,'' *Ind. Eng. Chem. Res.*, vol. 46, no. 23, pp. 7780–7787, 2007.
[59] C.-C. Hsu, M.-C. Chen, and L.-S. Chen, ''A novel process monitoring approach with dynamic independent component analysis,'' *Control Eng. Pract.*, vol. 18, no. 3, pp. 242–253, 2010.
[60] H. Yu, F. Khan, and V. Garaniya, ''Nonlinear Gaussian belief network based fault diagnosis for industrial processes,'' *J. Process Control*, vol. 35, pp. 178–200, Nov. 2015.
[61] Y. Liu and G. Zhang, ''Scale-sifting multiscale nonlinear process quality monitoring and fault detection,'' *Can. J. Chem. Eng.*, vol. 93, no. 8, pp. 1416–1425, 2015.
[62] H. Yu, F. Khan, and V. Garaniya, ''Modified independent component analysis and Bayesian network-based two-stage fault diagnosis of process operations,'' *Ind. Eng. Chem. Res.*, vol. 54, no. 10, pp. 2724–2742, 2015.
[63] Q. Jiang, B. Wang, and X. Yan, ''Multiblock independent component analysis integrated with Hellinger distance and Bayesian inference for non-Gaussian plant-wide process monitoring,'' *Ind. Eng. Chem. Res.*, vol. 54, no. 9, pp. 2497–2508, 2015.
[64] H. Albazzaz and X. Wang, ''Introduction of dynamics to an approach for batch process monitoring using independent component analysis,'' *Chem. Eng. Commun.*, vol. 194, no. 2, pp. 218–233, 2007.
[65] L. Cai and X. Tian, ''A new process monitoring method based on noisy time structure independent component analysis,'' *Chin. J. Chem. Eng.*, vol. 23, no. 1, pp. 162–172, 2015.
[66] L. Xia, H. Fang, and X. Liu, ''Hidden Markov model parameters estimation with independent multiple observations and inequality constraints,'' *IET Signal Process.*, vol. 8, no. 9, pp. 938–949, 2014.
[67] K. Peng, K. Zhang, X. He, G. Li, and X. Yang, ''New kernel independent and principal components analysis-based process monitoring approach with application to hot strip mill process,'' *IET Control Theory Appl.*, vol. 8, no. 16, pp. 1723–1731, 2014.
[68] Y. Xu, H. Wang, Z. Cui, F. Dong, and Y. Yan, ''Separation of gas–liquid two-phase flow through independent component analysis,'' *IEEE Trans. Instrum. Meas.*, vol. 59, no. 5, pp. 1294–1302, May 2010.
[69] S. Lloyd, ''Least squares quantization in PCM,'' *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[70] Z. Lv, X. Yan, and Q. Jiang, ''Batch process monitoring based on just-in-time learning and multiple-subspace principal component analysis,'' *Chemometrics Intell. Lab. Syst.*, vol. 137, pp. 128–139, Oct. 2014.
[71] C.-I. Wu, H.-Y. Kung, C.-H. Chen, and L.-C. Kuo, ''An intelligent slope disaster prediction and monitoring system based on WSN and ANP,'' *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4554–4562, 2014.
[72] I. B. Khediri, C. Weihs, and M. Limam, ''Kernel k-means clustering based local support vector domain description fault detection of multimodal processes,'' *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2166–2171, 2012.
[73] X. Zhao, W. Li, L. Zhou, G.-B. Song, Q. Ba, and J. Ou, ''Active thermometry based DS18B20 temperature sensor network for offshore pipeline scour monitoring using K-means clustering algorithm,'' *Int. J. Distrib. Sensor Netw.*, vol. 2013, Jun. 2013, Art. no. 852090, doi: 10.1155/2013/852090.
[74] J. Zhou, A. Guo, B. Celler, and S. Su, ''Fault detection and identification spanning multiple processes by integrating PCA with neural network,'' *Appl. Soft Comput.*, vol. 14, pp. 4–11, Jan. 2014.
[75] C. Tong, A. Palazoglu, and X. Yan, ''An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding,'' *J. Process Control*, vol. 23, no. 10, pp. 1497–1507, 2013.
[76] M. Wand and M. Jones, *Kernel Smoothing*. London, U.K.: Chapman & Hall, 1995.
[77] Q. Chen, U. Kruger, and A. T. Y. Leung, ''Regularised Kernel density estimation for clustered process data,'' *Control Eng. Pract.*, vol. 12, no. 3, pp. 267–274, 2004.
[78] Q. Jiang and X. Yan, ''Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical process monitoring,'' *Chemometrics Intell. Lab. Syst.*, vol. 127, pp. 121–131, Aug. 2013.
[79] M. He, C. Yang, X. Wang, W. Gui, and L. Wei, ''Nonparametric density estimation of froth colour texture distribution for monitoring sulphur flotation process,'' *Minerals Eng.*, vol. 53, pp. 203–212, Nov. 2013.
[80] R. Gonzalez and B. Huang, ''Control-loop diagnosis using continuous evidence through kernel density estimation,'' *J. Process Control*, vol. 24, no. 5, pp. 640–651, May 2014.
[81] G. Wang, J. Liu, Y. Li, and L. Shang, ''Fault detection based on diffusion maps and k nearest neighbor diffusion distance of feature space,'' *J. Chem. Eng. Jpn.*, vol. 48, no. 9, pp. 756–765, 2015.

[82] P. Phaladiganon, S. B. Kim, V. C. P. Chen, and W. Jiang, "Principal component analysis-based control charts for multivariate nonnormal distributions," *Expert Syst. Appl.*, vol. 40, no. 8, pp. 3044–3054, 2013.
[83] R. Gonzalez, B. Huang, and E. Lau, "Process monitoring using kernel density estimation and Bayesian networking with an industrial case study," *ISA Trans.*, vol. 58, pp. 330–347, Sep. 2015.
[84] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.
[85] F. Corona, M. Mulas, R. Baratti, and J. Romagnoli, "On the topological modeling and analysis of industrial process data using the SOM," *Comput. Chem. Eng.*, vol. 34, no. 12, pp. 2022–2032, 2010.
[86] H. Merdun, "Self-organizing map artificial neural network application in multidimensional soil data analysis," *Neural Comput. Appl.*, vol. 20, no. 8, pp. 1295–1303, 2011.
[87] M. Domínguez, J. J. Fuertes, I. Díaz, M. A. Prada, S. Alonso, and A. Morán, "Monitoring industrial processes with SOM-based dissimilarity maps," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7110–7120, 2012.
[88] T. Voyslavov, S. Tsakovski, and V. Simeonov, "Surface water quality assessment using self-organizing maps and Hasse diagram technique," *Chemometrics Intell. Lab. Syst.*, vol. 118, pp. 280–286, Aug. 2012.
[89] V.-M. Tikkala and S.-L. Jämsä-Jounela, "Monitoring of caliper sensor fouling in a board machine using self-organising maps," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11228–11233, 2012.
[90] X. Chen and X. Yan, "Using improved self-organizing map for fault diagnosis in chemical industry process," *Chem. Eng. Res. Des.*, vol. 90, no. 12, pp. 2262–2277, 2012.
[91] M. Liukkonen, I. Laakso, and Y. Hiltunen, "Advanced monitoring platform for industrial wastewater treatment: Multivariable approach using the self-organizing map," *Environ. Model. Softw.*, vol. 48, pp. 193–201, Oct. 2013.
[92] H. Yu, F. Khan, V. Garaniya, and A. Ahmad, "Sediagnosis technique for non-Gaussian processes," *Ind. Eng. Chem. Res.*, vol. 53, no. 21, pp. 8831–8843, 2014.
[93] D. Rivera, M. Sandoval, and A. Godoy, "Exploring soil databases: A self-organizing map approach," *Soil Use Manage.*, vol. 31, no. 1, pp. 121–131, 2015.
[94] G. Robertson, M. C. Thomas, and J. A. Romagnoli, "Topological preservation techniques for nonlinear process monitoring," *Comput. Chem. Eng.*, vol. 76, pp. 1–16, May 2015.
[95] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1377–1387, Jul. 2004.
[96] U. Thissen, H. Swierenga, A. P. de Weijer, R. Wehrens, W. J. Melssen, and L. M. C. Buydens, "Multivariate statistical process control using mixture modelling," *J. Chem.*, vol. 19, no. 1, pp. 23–31, 2005.
[97] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models," *AIChE J.*, vol. 54, no. 7, pp. 1811–1829, 2008.
[98] T. Chen and J. Zhang, "On-line multivariate statistical monitoring of batch processes using Gaussian mixture model," *Comput. Chem. Eng.*, vol. 34, no. 4, pp. 500–507, Apr. 2010.
[99] J. Yu, "Semiconductor manufacturing process monitoring using Gaussian mixture model and Bayesian method with local and nonlocal information," *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 3, pp. 480–493, Aug. 2012.
[100] J. Zhu, Z. Ge, and Z. Song, "Robust modeling of mixture probabilistic principal component analysis and process monitoring application," *AIChE J.*, vol. 60, no. 6, pp. 2143–2157, 2014.
[101] Q. Wen, Z. Ge, and Z. Song, "Multimode dynamic process monitoring based on mixture canonical variate analysis model," *Ind. Eng. Chem. Res.*, vol. 54, no. 5, pp. 1605–1614, 2015.
[102] Z. Yan, C. Huang, and Y. Yao, "Semi-supervised mixture discriminant monitoring for chemical batch processes," *Chem. Intell. Lab. Syst.*, vol. 134, pp. 10–22, May 2014.
[103] J. Yu, "A particle filter driven dynamic Gaussian mixture model approach for complex process monitoring and fault diagnosis," *J. Process Control*, vol. 22, no. 4, pp. 778–788, 2012.
[104] J. Liu and D.-S. Chen, "Nonstationary fault detection and diagnosis for multimode processes," *AIChE J.*, vol. 56, no. 1, pp. 207–219, 2010.
[105] T. Feital, U. Kruger, J. Dutra, J. Pinto, and E. Lima, "Modeling and performance monitoring of multivariate multimodal processes," *AIChE J.*, vol. 59, no. 5, pp. 1557–1569, 2013.
[106] J. Yu, J. Chen, and M. M. Rashid, "Multiway independent component analysis mixture model and mutual information based fault detection and diagnosis approach of multiphase batch processes," *AIChE J.*, vol. 59, no. 8, pp. 2761–2779, 2013.
[107] J. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. New York, NY, USA: Springer, 2007.
[108] J. Shao and G. Rong, "Nonlinear process monitoring based on maximum variance unfolding projections," *Expert Syst. Appl.*, vol. 36, no. 8, pp. 11332–11340, 2009.
[109] Z. Ge and Z. Song, "A nonlinear probabilistic method for process monitoring," *Ind. Eng. Chem. Res.*, vol. 49, no. 4, pp. 1770–1778, 2010.
[110] Z. Ge and Z. Song, "Nonlinear probabilistic monitoring based on the Gaussian process latent variable model," *Ind. Eng. Chem. Res.*, vol. 49, no. 10, pp. 4792–4799, 2010.
[111] M. Zhang, Z. Ge, Z. Song, and R. Fu, "Global–local structure analysis model and its application for fault detection and identification," *Ind. Eng. Chem. Res.*, vol. 50, no. 11, pp. 6837–6848, 2011.
[112] A. Miao, Z. Ge, Z. Song, and L. Zhou, "Time neighborhood preserving embedding model and its application for fault detection," *Ind. Eng. Chem. Res.*, vol. 52, no. 38, pp. 13717–13729, 2013.
[113] C. Tong and X. Yan, "Statistical process monitoring based on a multi-manifold projection algorithm," *Chem. Intell. Lab. Syst.*, vol. 130, pp. 20–28, Jan. 2014.
[114] Y. Liu, T. Chen, and Y. Yao, "Nonlinear process monitoring and fault isolation using extended maximum variance unfolding," *J. Process Control*, vol. 24, no. 6, pp. 880–891, 2014.
[115] A. Miao, Z. Ge, Z. Song, and F. Shen, "Nonlocal structure constrained neighborhood preserving embedding model and its application for fault detection," *Chem. Intell. Lab. Syst.*, vol. 142, pp. 184–196, Mar. 2015.
[116] Y. Ma, B. Song, H. Shi, and Y. Yang, "Fault detection via local and nonlocal embedding," *Chem. Eng. Res. Des.*, vol. 94, pp. 538–548, Feb. 2015.
[117] M. Aimin, L. Peng, and Y. Lingjian, "Neighborhood preserving regression embedding based data regression and its applications on soft sensor modeling," *Chem. Intell. Lab. Syst.*, vol. 147, pp. 86–94, Oct. 2015.
[118] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.
[119] X. Liu, L. Xie, U. Kruger, T. Littler, and S. Wang, "Statistical-based monitoring of multivariate non-Gaussian systems," *AIChE J.*, vol. 54, no. 9, pp. 2379–2391, 2008.
[120] Z. Ge, L. Xie, and Z. Song, "A novel statistical-based monitoring approach for complex multivariate processes," *Ind. Eng. Chem. Res.*, vol. 48, no. 10, pp. 4892–4898, 2009.
[121] Z. Ge, L. Xie, U. Kruger, L. Lamont, Z. Song, and S. Wang, "Sensor fault identification and isolation for multivariate non-Gaussian processes," *J. Process Control*, vol. 19, no. 10, pp. 1707–1715, 2009.
[122] Z. Ge, F. Gao, and Z. Song, "Batch process monitoring based on support vector data description method," *J. Process Control*, vol. 21, no. 6, pp. 949–959, Jul. 2011.
[123] Z. Ge and Z. Song, "Bagging support vector data description model for batch process monitoring," *J. Process Control*, vol. 23, no. 8, pp. 1090–1096, 2013.
[124] X. Liu, K. Li, M. McAfee, and G. W. Irwin, "Improved nonlinear PCA for process monitoring using support vector data description," *J. Process Control*, vol. 21, no. 9, pp. 1306–1317, 2011.
[125] Z. Zhu, Z. Song, and A. Palazoglu, "Transition process modeling and monitoring based on dynamic ensemble clustering and multiclass support vector data description," *Ind. Eng. Chem. Res.*, vol. 50, no. 24, pp. 13969–13983, 2011.
[126] Q. Jiang and X. Yan, "Probabilistic weighted NPE-SVDD for chemical process monitoring," *Control Eng. Pract.*, vol. 28, pp. 74–89, Jul. 2014.
[127] M. Yao, H. Wang, and W. Xu, "Batch process monitoring based on functional data analysis and support vector data description," *J. Process Control*, vol. 24, pp. 1085–1097, Jul. 2014.
[128] W. Du, Y. Tian, and F. Qian, "Monitoring for nonlinear multiple modes process based on LL-SVDD-MRDA," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 4, pp. 1133–1148, Oct. 2014.
[129] I. T. Jolliffe, "A note on the use of principal components in regression," *J. Roy. Statist. Soc., Ser. C*, vol. 31, no. 3, pp. 300–303, 1982.
[130] Y.-L. Xie and J. H. Kalivas, "Local prediction models by principal component regression," *Anal. Chim. Acta*, vol. 348, pp. 29–38, Aug. 1997.
[131] M. K. Hartnett, G. Lightbody, and G. W. Irwin, "Dynamic inferential estimation using principal components regression (PCR)," *Chem. Intell. Lab. Syst.*, vol. 40, no. 2, pp. 215–224, 1998.

[132] Y. Yang and F. Gao, "Injection molding product weight: Online prediction and control based on a nonlinear principal component regression model," *Polymer Eng. Sci.*, vol. 46, no. 4, pp. 540–548, 2006.
[133] R. B. Keithley, M. L. Heien, and R. M. Wightman, "Multivariate concentration determination using principal component regression with residual analysis," *TrAC-Trends Anal. Chem.*, vol. 28, no. 9, pp. 1127–1136, 2009.
[134] Z. Ge, F. Gao, and Z. Song, "Mixture probabilistic PCR model for soft sensing of multimode processes," *Chem. Intell. Lab. Syst.*, vol. 105, no. 1, pp. 91–105, Jan. 2011.
[135] W. Li, W. Gao, P. Chen, and B. Sun, "Near-infrared spectroscopy and principal components regression for the quality analysis of glass/epoxy prepreg," *Polymers Polymer Composites*, vol. 19, pp. 15–20, Jan. 2011.
[136] X. Yuan, Z. Ge, and Z. Song, "Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes," *Ind. Eng. Chem. Res.*, vol. 53, no. 35, pp. 13736–13749, 2014.
[137] Y. Lee, S.-Y. Ha, H.-K. Park, M.-S. Han, and K.-H. Shin, "Identification of key factors influencing primary productivity in two river-type reservoirs by using principal component regression analysis," *Environ. Monitor. Assessment*, vol. 187, p. 213, Apr. 2015.
[138] M. Khorasani, J. M. Amigo, C. C. Sun, P. Bertelsen, and J. Rantanen, "Near-infrared chemical imaging (NIR-CI) as a process monitoring solution for a production line of roll compaction and tableting," *Eur. J. Pharmaceutics Biopharmaceutics*, vol. 93, pp. 293–302, Jun. 2015.
[139] S. S. Kolluri, I. J. Esfahani, P. S. N. Garikiparthy, and C. K. Yoo, "Evaluation of multivariate statistical analyses for monitoring and prediction of processes in an seawater reverse osmosis desalination plant," *Korean J. Chem. Eng.*, vol. 32, no. 8, pp. 1486–1497, 2015.
[140] K. Popli, M. Sekhavat, A. Afacan, S. Dubljevic, Q. Liu, and V. Prasad, "Dynamic modeling and real-time monitoring of froth flotation," *Minerals*, vol. 5, no. 3, pp. 570–591, 2015.
[141] A. Ghosh *et al.*, "Monitoring the fermentation process and detection of optimum fermentation time of black tea using an electronic tongue," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6255–6262, Nov. 2015.
[142] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemometrics Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001.
[143] *Partial Least Squares Regression*. [Online]. Available: http://www.stzdatenanalyse.de/stz/plsr.html
[144] J. V. Kresta, T. E. Marlin, and J. F. MacGregor, "Development of inferential process models using PLS," *Comput. Chem. Eng.*, vol. 18, no. 7, pp. 597–611, 1994.
[145] U. Kruger, Q. Chen, D. J. Sandoz, and R. C. McFarlane, "Extended PLS approach for enhanced condition monitoring of industrial processes," *AIChE J.*, vol. 47, no. 9, pp. 2076–2091, 2001.
[146] Y. Zhang and Y. Zhang, "Complex process monitoring using modified partial least squares method of independent component regression," *Chemometrics Intell. Lab. Syst.*, vol. 98, no. 2, pp. 143–148, 2009.
[147] J. Yu, "Multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes," *Ind., Eng. Chem. Res.*, vol. 51, no. 40, pp. 13227–13237, 2012.
[148] X. Zhang, W. Yan, and W. Shao, "Nonlinear multivariate quality estimation and prediction based on kernel partial least squares," *Ind., Eng. Chem. Res.*, vol. 47, pp. 1120–1131, 2008.
[149] Y. Zhang, Y. Teng, and Y. Zhang, "Complex process quality prediction using modified kernel partial least squares," *Chem. Eng. Sci.*, vol. 65, no. 6, pp. 2153–2158, 2010.
[150] W. Ni, S. K. Tan, W. J. Ng, and S. D. Brown, "Localized, adaptive recursive partial least squares regression for dynamic system modeling," *Ind. Eng. Chem. Res.*, vol. 51, no. 23, pp. 8025–8039, 2012.
[151] J. Liu, "Developing a soft sensor based on sparse partial least squares with variable selection," *J. Process Control*, vol. 24, no. 7, pp. 1046–1056, 2014.
[152] W. Shao, X. Tian, P. Wang, X. Deng, and S. Chen, "Online soft sensor design using local partial least squares models with adaptive process state partition," *Chemometrics Intell. Lab. Syst.*, vol. 144, pp. 108–121, May 2015.
[153] L. Xu, M. Goodarzi, W. Shi, C. Cai, and J. Jiang, "A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers," *Chemometrics Intell. Lab. Syst.*, vol. 139, pp. 58–63, Dec. 2014.
[154] J. Vanlaer, G. Gins, and J. F. M. Van Impe, "Quality assessment of a variance estimator for partial least squares prediction of batch-end quality," *Comput. Chem. Eng.*, vol. 52, pp. 230–239, May 2013.
[155] S. Stubbs, J. Zhang, and J. Morris, "Multiway interval partial least squares for batch process performance monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 35, pp. 12399–12407, 2013.
[156] J. L. Godoy, D. A. Zumoffen, J. R. Vega, and J. L. Marchetti, "New contributions to non-linear process monitoring through kernel partial least squares," *Chemometrics Intell. Lab. Syst.*, vol. 135, pp. 76–89, Jul. 2014.
[157] K. Zhang, H. Hao, Z. Chen, S. X. Ding, and K. Peng, "A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches," *J. Process Control*, vol. 33, pp. 112–126, Sep. 2015.
[158] D. Zhou, G. Li, and S. J. Qin, "Total projection to latent structures for process monitoring," *AIChE J.*, vol. 56, no. 1, pp. 168–178, 2010.
[159] S. J. Qin and Y. Y. Zheng, "Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures," *AIChE J.*, vol. 59, no. 2, pp. 496–504, Feb. 2013.
[160] K. Peng, K. Zhang, B. You, and J. Dong, "Quality-related prediction and monitoring of multi-mode processes using multiple PLS with application to an industrial hot strip mill," *Neurocomputing*, vol. 168, pp. 1094–1103, Nov. 2015.
[161] G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. Hoboken, NJ, USA: Wiley, 2004.
[162] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1389–1401, Jul. 2004.
[163] J. Yu, "Localized Fisher discriminant analysis based complex chemical process monitoring," *AIChE J.*, vol. 57, no. 7, pp. 1817–1828, 2011.
[164] Z.-B. Zhu and Z.-H. Song, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6895–6905, 2011.
[165] Q. P. He, S. J. Qin, and J. Wang, "A new fault diagnosis method using fault directions in Fisher discriminant analysis," *AIChE J.*, vol. 51, no. 2, pp. 555–571, 2005.
[166] X. Zhang, W.-W. Yan, X. Zhao, and H.-H. Shao, "Nonlinear real-time process monitoring and fault diagnosis based on principal component analysis and kernel Fisher discriminant analysis," *Chem. Eng. Technol.*, vol. 30, no. 9, pp. 1203–1211, 2007.
[167] J. Yu, "Nonlinear bioprocess monitoring using multiway kernel localized Fisher discriminant analysis," *Ind. Eng. Chem. Res.*, vol. 50, no. 6, pp. 3390–3402, 2011.
[168] C.-C. Huang, T. Chen, and Y. Yao, "Mixture discriminant monitoring: A hybrid method for statistical process monitoring and fault diagnosis/isolation," *Ind. Eng. Chem. Res.*, vol. 52, no. 31, pp. 10720–10731, 2013.
[169] B. Jiang, X. Zhu, D. Huang, J. A. Paulson, and R. D. Braatz, "A combined canonical variate analysis and Fisher discriminant analysis (CVA–FDA) approach for fault diagnosis," *Comput. Chem. Eng.*, vol. 77, pp. 1–9, Jun. 2015.
[170] Z.-B. Zhu and Z.-H. Song, "Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis," *Chem. Eng. Res. Des.*, vol. 88, no. 8, pp. 936–951, 2010.
[171] C. Sumana, B. Mani, C. Venkateswarlu, and R. D. Gudi, "Improved fault diagnosis using dynamic kernel scatter-difference-based discriminant analysis," *Ind. Eng. Chem. Res.*, vol. 49, no. 18, pp. 8575–8586, 2010.
[172] C. Zhao and F. Gao, "A nested-loop Fisher discriminant analysis algorithm," *Chemometrics Intell. Lab. Syst.*, vol. 146, pp. 396–406, Aug. 2015.
[173] G. Rong, S.-Y. Liu, and J.-D. Shao, "Fault diagnosis by locality preserving discriminant analysis and its kernel variation," *Comput. Chem. Eng.*, vol. 49, pp. 105–113, Feb. 2013.
[174] J. O. Rawlings, S. G. Pantula, and D. Dickey, *Applied Regression Analysis* (Springer Texts in Statistics). New York, NY, USA: Springer-Verlag 1998.
[175] S. Park and C. Han, "A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns," *Comput. Chem. Eng.*, vol. 24, nos. 2–7, pp. 871–877, 2000.
[176] Y. Liu, X. Sun, J. Zhou, H. Zhang, and C. Yang, "Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis–NIR diffuse reflectance spectroscopy," *Math. Comput. Model.*, vol. 51, nos. 11–12, pp. 1438–1443, 2010.
[177] M. Caselli, L. Trizio, G. de Gennaro, and P. Ielpo, "A simple feedforward neural network for the PM_{10} forecasting: Comparison with a radial basis function network and a multivariate linear regression model," *Water, Air, Soil Pollution*, vol. 201, nos. 1–4, pp. 365–377, 2009.

[178] L. Laurens and E. Wolfrum, "High-throughput quantitative biochemical characterization of algal biomass by NIR spectroscopy; multiple linear regression and multivariate linear regression analysis," *J. Agricult. Food Chem.*, vol. 61, no. 50, pp. 12307–12314, 2013.
[179] H. Abyaneh, "Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters," *J. Environ. Health Sci. Eng.*, vol. 12, p. 40, Jan. 2014.
[180] J. L. Aleixandre-Tudó, I. Alvarez, M. J. García, V. Lizama, and J. L. Aleixandre, "Application of multivariate regression methods to predict sensory quality of red wines," *Czech J. Food Sci.*, vol. 33, no. 3, pp. 217–227, 2015.
[181] A. Farkas *et al.*, "Comparison of multivariate linear regression methods in micro-Raman spectrometric quantitative characterization," *J. Raman Spectrosc.*, vol. 46, no. 6, pp. 566–576, 2015.
[182] A. Amiri, A. Saghaei, M. Mohseni, and Y. Zerehsaz, "Diagnosis aids in multivariate multiple linear regression profiles monitoring," *Commun. Stat.-Theory Methods*, vol. 43, no. 14, pp. 3057–3079, 2014.
[183] R. Noorossana, M. Eyvazian, A. Amiri, and M. Mahmoud, "Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application," *Quality Rel. Eng. Int.*, vol. 26, no. 3, pp. 291–303, 2010.
[184] M. Eyvazian, R. Noorossana, A. Saghaei, and A. Amiri, "Phase II monitoring of multivariate multiple linear regression profiles," *Quality Rel. Eng. Int.*, vol. 27, no. 3, pp. 281–296, 2011.
[185] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press. 1995.
[186] M. W. Lee, J. Y. Joung, D. S. Lee, J. M. Park, and S. H. Woo, "Application of a moving-window-adaptive neural network to the modeling of a full-scale anaerobic filter process," *Ind. Eng. Chem. Res.*, vol. 44, no. 11, pp. 3973–3982, 2005.
[187] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, and R. M. Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Comput. Chem. Eng.*, vol. 33, no. 1, pp. 43–49, 2009.
[188] V. R. Radhakrishnan and A. R. Mohamed, "Neural networks for the identification and control of blast furnace hot metal quality," *J. Process Control*, vol. 10, no. 6, pp. 509–524, 2000.
[189] J. Shi, X. Liu, and Y. Sun, "Melt index prediction by neural networks based on independent component analysis and multi-scale analysis," *Neurocomputing*, vol. 70, pp. 280–287, Dec. 2006.
[190] M. Jamil, A. Kalam, A. Q. Ansari, and M. Rizwan, "Generalized neural network and wavelet transform based approach for fault location estimation of a transmission line," *Appl. Soft Comput.*, vol. 19, pp. 322–332, Jun. 2014.
[191] S. A. Iliyas, M. Elshafei, M. A. Habib, and A. A. Adeniran, "RBF neural network inferential sensor for process emission monitoring," *Control Eng. Pract.*, vol. 21, pp. 962–970, Jun. 2013.
[192] T. Nagpal and Y. S. Brar, "Artificial neural network approaches for fault classification: Comparison and performance," *Neural Comput. Appl.*, vol. 25, pp. 1863–1870, Dec. 2014.
[193] I. Loboda and M. A. O. Robles, "Gas turbine fault diagnosis using probabilistic neural networks," *Int. J. Turbo Jet-Engines*, vol. 32, no. 2, pp. 175–191, 2015.
[194] M. Fu, W. Wang, Z. Le, and M. S. Khorram, "Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model," *Neural Comput. Appl.*, vol. 26, no. 8, pp. 1789–1797, 2015.
[195] K. Sun, J. Liu, J.-L. Kang, S.-S. Jang, D. S.-H. Wong, and D.-S. Chen, "Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote," *J. Process Control*, vol. 24, no. 7, pp. 1068–1075, 2014.
[196] S. Xu and X. Liu, "Melt index prediction by fuzzy functions with dynamic fuzzy neural networks," *Neurocomputing*, vol. 142, pp. 291–298, Oct. 2014.
[197] J. Pian and Y. Zhu, "A hybrid soft sensor for measuring hot-rolled strip temperature in the laminar cooling process," *Neurocomputing*, vol. 169, pp. 457–465, Dec. 2015.
[198] A. K. Pani and H. K. Mohanta, "Online monitoring and control of particle size in the grinding process using least square support vector regression and resilient back propagation neural network," *ISA Trans.*, vol. 56, pp. 206–221, Mar. 2015.
[199] T.-I. Liu and B. Jolley, "Tool condition monitoring (TCM) using neural networks," *Int. J. Adv. Manuf. Technol.*, vol. 78, pp. 9–12, Jun. 2015.
[200] A. R. Gholami and M. Shahbazian, "Soft sensor design based on fuzzy C-means and RFN_SVR for a stripper column," *J. Natural Gas Sci. Eng.*, vol. 25, pp. 23–29, Jul. 2015.
[201] C. Cortes and V. Vapnik, "Support-vector networks," in *Proc. 25th Mach. Lean.*, 1995, vol. 20. no. 3, pp. 273–297.
[202] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1995.
[203] B. Schölkopf, C. Burges, and A. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1999.
[204] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA, USA: MIT Press, 2002.
[205] Y. Zhang, "Fault detection and diagnosis of nonlinear processes using improved kernel independent component analysis (KICA) and support vector machine (SVM)," *Ind. Eng. Chem. Res.*, vol. 47, no. 18, pp. 6961–6971, 2008.
[206] S. Du, D. Huang, and J. Lv, "Recognition of concurrent control chart patterns using wavelet transform decomposition and multiclass support vector machines," *Comput. Ind. Eng.*, vol. 66, no. 4, pp. 683–695, 2013.
[207] Y. Tian, M. Fu, and F. Wu, "Steel plates fault diagnosis on the basis of support vector machines," *Neurocomputing*, vol. 151, pp. 296–303, Mar. 2015.
[208] D. You, X. Gao, and S. Katayama, "WPD-PCA-based laser welding process monitoring and defects diagnosis by using FNN and SVM," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 628–636, Jan. 2015.
[209] Y. Xiao, H. Wang, and L. Zhang, "Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection," *Knowl.-Based Syst.*, vol. 59, pp. 75–84, Mar. 2014.
[210] M. Namdari and H. Jazayeri-Rad, "Incipient fault diagnosis using support vector machines based on monitoring continuous decision functions," *Eng. Appl. Artif. Intell.*, vol. 28, pp. 22–35, Feb. 2014.
[211] N. Pooyan, M. Shahbazian, K. Salahshoor, and M. Hadian, "Simultaneous fault diagnosis using multi class support vector machine in a Dew Point process," *J. Natural Gas Sci. Eng.*, vol. 23, pp. 373–379, Mar. 2015.
[212] P. Xanthopoulos and T. Razzaghi, "A weighted support vector machine method for control chart pattern recognition," *Comput. Ind. Eng.*, vol. 66, pp. 683–695, Apr. 2013.
[213] L. Saidi, J. B. Ail, and F. Friaiech, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA Trans.*, vol. 54, pp. 193–206, Jan. 2015.
[214] C. Jing and J. Hou, "SVM and PCA based fault classification approaches for complicated industrial process," *Neurocomputing*, vol. 167, pp. 636–642, Nov. 2015.
[215] F. Wu, S. Yin, and H. R. Karimi, "Fault detection and diagnosis in process data using support vector machines," *J. Appl. Math.*, vol. 2014, Mar. 2014, Art. no. 732104.
[216] D. Zhang, G. Bi, Z. Sun, and Y. Guo, "Online monitoring of precision optics grinding using acoustic emission based on support vector machine," *Int. J. Adv. Manuf. Technol.*, vol. 80, pp. 761–774, Sep. 2015.
[217] P. Jain, I. Rahman, and B. D. Kulkarni, "Development of a soft sensor for a batch distillation column using support vector regression techniques," *Chem. Eng. Res. Des.*, vol. 85, no. 2, pp. 283–287, 2007.
[218] Y. Liu, N. Hu, H. Wang, and P. Li, "Soft chemical analyzer development using adaptive least-squares support vector regression with selective pruning and variable moving window size," *Ind. Eng. Chem. Res.*, vol. 48, no. 12, pp. 5731–5741, 2009.
[219] J. Yu, "A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses," *Comput. Chem. Eng.*, vol. 41, pp. 134–144, Jun. 2012.
[220] Z. Ge and Z. Song, "A comparative study of just-in-time-learning based methods for online soft sensor modeling," *Chemometrics Intell. Lab. Syst.*, vol. 104, no. 2, pp. 306–317, 2010.
[221] Y. Liu, Z. Gao, P. Li, and H. Wang, "Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes," *Ind. Eng. Chem. Res.*, vol. 51, no. 11, pp. 4313–4327, 2012.
[222] H. Kaneko and K. Funatsu, "Application of online support vector regression for soft sensors," *AIChE J.*, vol. 60, no. 2, pp. 600–612, 2014.
[223] Y. Liu and J. Chen, "Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes," *J. Process Control*, vol. 23, no. 6, pp. 793–804, 2013.
[224] H. Kaneko and K. Funatsu, "Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants," *Chemometr. Intell. Lab. Syst.*, vol. 137, pp. 57–66, Oct. 2014.

[225] S. Zhang, F. Wang, D. He, and R. Jia, "Online quality prediction for cobalt oxalate synthesis process using least squares support vector regression approach with dual updating," *Control Eng. Pract.*, vol. 21, no. 10, pp. 1267–1276, 2013.

[226] W. Wang and X. Liu, "Melt index prediction by least squares support vector machines with an adaptive mutation fruit fly optimization algorithm," *Chemometrics Intell. Lab. Syst.*, vol. 141, pp. 79–87, Feb. 2015.

[227] H. Jin, X. Chen, J. Yang, H. Zhang, L. Wang, and L. Wu, "Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes," *Chem. Eng. Sci.*, vol. 131, pp. 282–303, Jul. 2015.

[228] S. Shamshirband, D. Petkovic, H. Javidnia, and A. Gani, "Sensor data fusion by support vector regression methodology—A comparative study," *IEEE Sensors J.*, vol. 15, no. 2, pp. 850–854, Feb. 2015.

[229] Y. Lv, T. Yang, and J. Liu, "An adaptive least squares support vector machine model with a novel update for NOx emission prediction," *Chemometrics Intell. Lab. Syst.*, vol. 145, pp. 103–113, Jul. 2015.

[230] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.

[231] *Kernel Smoother*. [Online]. Available: https//en.wikipedia.org/wiki/Kernel_smoother

[232] B. Lee and M. Scholz, "A comparative study: Prediction of constructed treatment wetland performance with K-nearest neighbors and neural networks," *Water Air Soil Pollution*, vol. 174, pp. 279–301, Jul. 2006.

[233] P. Facco, F. Bezzo, and M. Barolo, "Nearest-neighbor method for the automatic maintenance of multivariate statistical soft sensors in batch processing," *Ind. Eng. Chem. Res.*, vol. 49, no. 5, pp. 2336–2347, 2010.

[234] Q. P. He and J. Wang, "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," *AIChE J.*, vol. 57, no. 1, pp. 107–121, 2011.

[235] F. Penedo, R. E. Haber, A. Gajate, and R. M. del Toro, "Hybrid incremental modeling based on least squares and fuzzy K-NN for monitoring tool wear in turning processes," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 811–818, Nov. 2012.

[236] H. Kaneko, M. Arakawa, and K. Funatsu, "Novel soft sensor method for detecting completion of transition in industrial polymer processes," *Comput. Chem. Eng.*, vol. 35, no. 6, pp. 1135–1142, 2011.

[237] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, "Accelerating wrapper-based feature selection with K-nearest-neighbor," *Knowl.-Bases Syst.*, vol. 83, pp. 81–91, Jul. 2015.

[238] K. Fuchs, J. Gertheiss, and G. Tutz, "Nearest neighbor ensembles for functional data with interpretable feature selection," *Chemometrics Intell. Lab. Syst.*, vol. 146, pp. 186–197, Aug. 2015.

[239] Y.-Y. Su, S. Liang, J.-Z. Li, X.-G. Deng, and T.-F. Li, "Nonlinear fault separation for redundancy process variables based on FNN in MKFDA subspace," *J. Appl. Math.*, vol. 2014, Feb. 2014, Art. no. 729763.

[240] Y. Li and X. Zhang, "Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection," *Chemometrics Intell. Lab. Syst.*, vol. 136, pp. 47–57, Aug. 2014.

[241] G. Gins, P. Van den Kerkhof, J. Vanlaer, and I. F. M. Van Impe, "Improving classification-based diagnosis of batch processes through data selection and appropriate pretreatment," *J. Process Control*, vol. 26, pp. 90–101, Feb. 2015.

[242] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[243] Z. Ge, T. Chen, and Z. Song, "Quality prediction for polypropylene production process based on CLGPR model," *Control Eng. Pract.*, vol. 19, no. 5, pp. 423–432, 2011.

[244] J. Yu, "Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach," *Chem. Eng. Sci.*, vol. 82, pp. 22–30, Sep. 2012.

[245] K. Song, F. Wu, T.-P. Tong, and X.-J. Wang, "A real-time Mooney-viscosity prediction model of the mixed rubber based on the independent component regression-Gaussian process algorithm," *J. Chemometrics*, vol. 26, pp. 557–564, Nov./Dec. 2012.

[246] T. Chen and J. Ren, "Bagging for Gaussian process regression," *Neurocomputing*, vol. 72, no. 7, pp. 1605–1610, 2009.

[247] R. Grbić, D. Slisković, and P. Kadlec, "Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models," *Comput. Chem. Eng.*, vol. 58, pp. 84–97, Nov. 2013.

[248] J. Chen, J. Yu, and Y. Zhang, "Multivariate video analysis and Gaussian process regression model based soft sensor for online estimation and prediction of nickel pellet size distributions," *Comput. Chem. Eng.*, vol. 64, pp. 13–23, May 2014.

[249] L. Zhou, J. Chen, and Z. Song, "Recursive Gaussian process regression model for adaptive quality monitoring in batch processes," *Math. Problems Eng.*, vol. 2015, Dec. 2014, Art. no. 761280, doi: 10.1155/2015/761280.

[250] Y. Liu, T. Chen, and J. Chen, "Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions," *Ind. Eng. Chem. Res.*, vol. 54, no. 18, pp. 5037–5047, 2015.

[251] H. Jin, X. Chen, L. Wang, K. Yang, and L. Wu, "Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes," *Ind. Eng. Chem. Res.*, vol. 54, no. 30, pp. 7320–7345, 2015.

[252] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.

[253] J. R. Quinlan, "Discovering rules by induction from large collections of examples," in *Expert Systems in the Micro Electronic Age*, D. Michie, Ed. Edinburgh, U.K.: Edinburgh Univ. Press, 1979.

[254] J. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[255] Y. Kuo and K.-P. Lin, "Using neural network and decision tree for machine reliability prediction," *Int. J. Adv. Manuf. Technol.*, vol. 50, pp. 1243–1251, Oct. 2010.

[256] D.-Y. Yeh, C.-H. Cheng, and S.-C. Hsiao, "Classification knowledge discovery in mold tooling test using decision tree algorithm," *J. Intell. Manuf.*, vol. 22, no. 4, pp. 585–595, 2011.

[257] E. Zio, P. Baraldi, and I. C. Popescu, "A fuzzy decision tree method for fault classification in the steam generator of a pressurized water reactor," *Ann. Nucl. Energy*, vol. 36, no. 8, pp. 1159–1169, 2009.

[258] C. Y. Ma and X. Z. Wang, "Inductive data mining based on genetic programming: Automatic generation of decision trees from data for process historical data analysis," *Comput. Chem. Eng.*, vol. 33, no. 10, pp. 1602–1616, 2009.

[259] S.-G. He, Z. He, and G. A. Wang, "Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques," *J. Intell. Manuf.*, vol. 24, no. 1, pp. 25–34, 2013.

[260] M. Demetgul, "Fault diagnosis on production systems with support vector machine and decision trees algorithms," *Int. J. Adv. Manuf. Technol.*, vol. 67, pp. 2183–2194, Aug. 2013.

[261] I. Aydin, M. Karakose, and E. Akin, "An approach for automated fault diagnosis based on a fuzzy decision tree and boundary analysis of a reconstructed phase space," *ISA Trans.*, vol. 53, no. 2, pp. 220–229, 2014.

[262] N. E. I. Karabadji, H. Seridi, I. Khelf, N. Azizi, and R. Boulkroune, "Improved decision tree construction based on attribute selection and data sampling for fault diagnosis in rotating machines," *Eng. Appl. Artif. Intell.*, vol. 35, pp. 71–83, Oct. 2014.

[263] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, Montreal, QC, USA, Aug. 1995, pp. 278–282.

[264] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[265] M. Pardo and G. Sberveglieri, "Random forests and nearest shrunken centroids for the classification of sensor array data," *Sens. Actuators B, Chem.*, vol. 131, no. 1, pp. 93–99, Apr. 2008.

[266] B. Yang, X. Di, and T. Han, "Random forests classifier for machine fault diagnosis," *J. Mech. Sci. Technol.*, vol. 22, no. 9, pp. 1716–1725, 2008.

[267] L. Auret and C. Aldrich, "Change point detection in time series data with random forests," *Control Eng. Pract.*, vol. 18, no. 8, pp. 990–1002, 2010.

[268] L. Auret and C. Aldrich, "Unsupervised process fault detection with random forests," *Ind. Eng. Chem. Res.*, vol. 49, no. 19, pp. 9184–9194, 2010.

[269] I. Ahmad, M. Kano, S. Hasebe, H. Kitada, and N. Murata, "Gray-box modeling for prediction and control of molten steel temperature in tundish," *J. Process Control*, vol. 24, no. 4, pp. 375–382, 2014.

[270] O. Chapelle, B. Scholkopf, and A. Zien, Eds., *Semi-Supervised Learning*. Cambridge, MA, USA: MIT Press, 2006.

[271] X. Zhu, "Semi-supervised learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., 2008.

[272] Z. Ge and Z. Song, "Semisupervised Bayesian method for soft sensor modeling with unlabeled data samples," *AIChE J.*, vol. 57, no. 8, pp. 2109–2119, 2011.

[273] X. Shao, B. Huang, J. M. Lee, F. Xu, and A. Espejo, "Bayesian method for multirate data synthesis and model calibration," *AIChE J.*, vol. 57, no. 6, pp. 1514–1525, 2011.

[274] J. Ji, H. Wang, K. Chen, Y. Liu, N. Zhang, and J. Yan, "Recursive weighted kernel regression for semi-supervised soft-sensing modeling of fed-batch processes," *J. Taiwan Inst. Chem. Eng.*, vol. 43, no. 1, pp. 67–76, 2012.

[275] Z. Ge, Z. Song, and F. Gao, "Self-training statistical quality prediction of batch processes with limited quality data," *Ind. Eng. Chem. Res.*, vol. 52, no. 2, pp. 979–984, 2013.

[276] J. Deng and B. Huang, "Identification of nonlinear parameter varying systems with missing output data," *AIChE J.*, vol. 58, no. 11, pp. 3454–3467, 2012.

[277] S. Zhong, Q. Wen, and Z. Ge, "Semi-supervised fisher discriminant analysis model for fault classification in industrial processes," *Chemometrics Intell. Lab. Syst.*, vol. 138, pp. 203–211, Nov. 2014.

[278] X. Jin, S. Wang, B. Huang, and F. Forbes, "Multiple model based LPV soft sensor development with irregular/missing process output measurement," *Control Eng. Pract.*, vol. 20, no. 2, pp. 165–172, 2012.

[279] Z. Ge, B. Huang, and Z. Song, "Mixture semisupervised principal component regression model and soft sensor application," *AIChE J.*, vol. 60, no. 2, pp. 533–545, 2014.

[280] Z. Ge, B. Huang, and Z. Song, "Nonlinear semisupervised principal component regression for soft sensor modeling and its mixture form," *J. Chemometrics*, vol. 28, no. 11, pp. 793–804, 2014.

[281] G. K. Raju and C. L. Cooney, "Active learning from process data," *AIChE J.*, vol. 44, no. 10, pp. 2199–2211, 1998.

[282] Z. Ge, "Active learning strategy for smart soft sensor development under a small number of labeled data samples," *J. Process Control*, vol. 24, pp. 1454–1461, Sep. 2014.

[283] L. Zhou, J. Chen, Z. Song, and Z. Ge, "Semi-supervised PLVR models for process monitoring with unequal sample sizes of process variables and quality variables," *J. Process Control*, vol. 26, pp. 1–16, Feb. 2015.

[284] J. Zhu, Z. Ge, and Z. Song, "Robust semi-supervised mixture probabilistic principal component regression model development and application to soft sensors," *J. Process Control*, vol. 32, pp. 25–37, Aug. 2015.

[285] L. Bao, X. Yuan, and Z. Ge, "Co-training partial least squares model for semi-supervised soft sensor development," *Chemometrics Intell. Lab. Syst.*, vol. 147, pp. 75–85, Oct. 2015.

[286] S. J. Qin, "Process data analytics in the era of big data," *AIChE J.*, vol. 60, no. 9, pp. 3092–3100, 2014.

[287] J. Manyika and M. Chui, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institut Report, 2011. [Online]. Available: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[288] V. Mayer-Schonberge and K. Cukier, *Big Data: A Revolution That will Transform how we Live, Work, and Think*. Boston, MA, USA: Houghton Mifflin, 2013.

[289] N. R. Council, *Frontiers in Massive Data Analysis*. Washington, DC, USA: The National Academies Press, 2013.

[290] F. Y. Wang, "A big-data perspective on AI: Newton, merton, and analytics intelligence," *IEEE Intell. Syst.*, vol. 27, no. 5, pp. 2–4, Sep./Oct. 2012.

[291] B. Ratner, *Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data*. Boca Raton, FL, USA: CRC Press, 2012.

[292] T. Condie, P. Mineiro, N. Polyzotis, and M. Weimer, "Machine learning for big data," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2013, pp. 939–942.

[293] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, Jan. 2013.

[294] C. Ma, H. H. Zhang, and X. Wang, "Machine learning for big data analytics in plants," *Trends Plant Sci.*, vol. 19, no. 12, pp. 798–808, 2014.

[295] Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein, "Distributed GraphLab: A framework for machine learning and data mining in the cloud," in *Proc. VLDB Endowment*, vol. 5. 2012, pp. 716–727.

[296] Z. Ge and J. Chen, "Plant-wide industrial process monitoring: A distributed modeling framework," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 310–321, Feb. 2016.

[297] B. R. Bakshi and J. Fiksel, "The quest for sustainability: Challenges for process systems engineering," *AIChE J.*, vol. 49, no. 6, pp. 1350–1358, 2003.

[298] R. J. Hanes and B. R. Bakshi, "Sustainable process design by the process to planet framework," *AIChE J.*, vol. 61, no. 10, pp. 3320–3331, 2015.

[299] L. H. Chiang and R. D. Braatz, "Process monitoring using causal map and multivariate statistics: Fault detection and identification," *Chemometrics Intell. Lab. Syst.*, vol. 65, no. 2, pp. 159–178, 2003.

[300] F. Yang, P. Duan, S. L. Shah, and T. Chen, *Capturing Connectivity and Causality in Complex Industrial Processes*. New York, NY, USA: Springer-Verlag, 2014.

[301] T. Yuan and S. J. Qin, "Root cause diagnosis of plant-wide oscillations using granger causality," *J. Process Control*, vol. 24, no. 2, pp. 450–459, 2014.

[302] R. Landman, J. Kortela, Q. Sun, and S.-L. Jämsä-Jounela, "Fault propagation analysis of oscillations in control loops using data-driven causality and plant connectivity," *Comput. Chem. Eng.*, vol. 71, pp. 446–456, Dec. 2014.

[303] D. Peng, Z. Geng, and Q. Zhu, "A multilogic probabilistic signed directed graph fault diagnosis approach based on Bayesian inference," *Ind. Eng. Chem. Res.*, vol. 53, no. 23, pp. 9792–9804, 2014.

[304] C. B. Low, D. Wang, S. Arogeti, and J. B. Zhang, "Causality assignment and model approximation for hybrid bond graph: Fault diagnosis perspectives," *IEEE Trans. Autom. Sci. Eng.*, vol. 7, no. 3, pp. 570–580, Jul. 2010.

[305] Q. Zhang and S. Geng, "Dynamic uncertain causality graph applied to dynamic fault diagnoses of large and complex systems," *IEEE Trans. Rel.*, vol. 64, no. 3, pp. 910–927, Sep. 2015.

[306] P. D. Allison, "Handling missing data by maximum likelihood," in *SAS Global Forum Satistics and Data Analysis*. Orlando, FL, USA: SAS Institute, 2012, pp. 1–21.

[307] M. Aminghafari, N. Cheze, and J.-M. Poggi, "Multivariate denoising using wavelets and principal component analysis," *Comput. Stat. Data Anal.*, vol. 50, no. 9, pp. 2381–2398, 2006.

[308] A. N. Baraldi and C. K. Enders, "An introduction to modern missing data analyses," *J. School Psychol.*, vol. 48, pp. 5–37, Feb. 2010.

[309] R. L.-N. de la Fuente, S. García-Muñoz, and L. T. Biegler, "An efficient nonlinear programming strategy for PCA models with incomplete data sets," *J. Chemometrics*, vol. 24, no. 6, pp. 301–311, 2010.

[310] E. Eirola, A. Lendasse, V. Vandewalle, and C. Biernacki, "Mixture of Gaussians for distance estimation with missing data," *Neurocomputing*, vol. 131, pp. 32–42, May 2014.

[311] M. P. Gómez-Carracedo, J. M. Andrade, P. López-Mahía, S. Muniategui, and D. Prada, "A practical comparison of single and multiple imputation methods to handle complex missing data in air quality datasets," *Chemometrics Intell. Lab. Syst.*, vol. 134, pp. 23–33, May 2014.

[312] W. Jiang *et al.*, "Comparisons of five algorithms for chromatogram alignment," *Chromatographia*, vol. 76, pp. 1067–1078, Sep. 2013.

[313] Y. Liu and J. Chen, "Correntropy kernel learning for nonlinear system identification with outliers," *Ind. Eng. Chem. Res.*, vol. 53, no. 13, pp. 5248–5260, 2014.

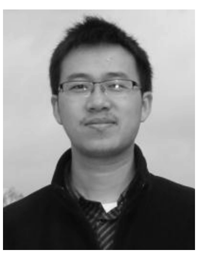
[314] T. Vatanen *et al.*, "Self-organization and missing values in SOM and GTM," *Neurocomputing*, vol. 147, pp. 60–70, Jan. 2015.

[315] Z. Zhao, B. Huang, and F. Liu, "Bayesian method for state estimation of batch process with missing data," *Comput. Chem. Eng.*, vol. 53, pp. 14–24, Jun. 2013.

[316] L. D. Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[317] F. Chen, P. Deng, J. Wan, D. Zhang, V. A. Vasilakos, and X. Rong, "Data mining for the Internet of Things: Literature review and challenges," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 8, pp. 1–14, 2015, doi: 10.1155/2015/431047.

[318] V. J. Hodge, S. O'Keefe, M. Weeks, and A. Moulds, "Wireless sensor networks for condition monitoring in the railway industry: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1088–1106, Jun. 2015.



ZHIQIANG GE received the B.Eng. and Ph.D. degrees from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. He was a Research Associate with the Department of Chemical and Biomolecular Engineering, The Hong Kong University of Science Technology, from 2010 to 2011, and a Visiting Professor with the Department of Chemical and Materials Engineering, University of Alberta, in 2013. He was an Alexander von Humboldt Research Fellow with the University of Duisburg-Essen from 2014 to 2017. He is currently a Full Professor with the College of Control Science and Engineering, Zhejiang University. His research interests include data mining and analytics, process monitoring and fault diagnosis, machine leaning, and industrial big data.

ZHIHUAN SONG received the B.Eng. and M.Eng. degrees in industrial automation from the Hefei University of Technology, Anhui, China, in 1983 and 1986, respectively, and the Ph.D. degree in industrial automation from Zhejiang University, Hangzhou, China, in 1997. Since then, he has been with the Department of Control Science and Engineering, Zhejiang University, where he was at first a Postdoctoral Research Fellow, then an Associate Professor, and currently a Professor. He has authored or co-authored over 200 papers in journals and conferences. His research interests include the modeling and fault diagnosis of industrial processes, embedded control systems, and advanced process control technologies.

STEVEN X. DING received the Ph.D. degree in electrical engineering from the Gerhard-Mercator University of Duisburg, Germany, in 1992. From 1992 to 1994, he was a Research and Development Engineer with Rheinmetall GmbH. From 1995 to 2001, he was a Professor of Control Engineering with the University of Applied Science Lausitz, Senftenberg, Germany, where he served as a Vice President from 1998 to 2000. Since 2001, he has been a Chair Professor of Control Engineering and the Head of the Institute for Automatic Control and Complex Systems with the University of Duisburg-Essen, Germany. His research interests are model-based and data-driven fault diagnosis, control and fault-tolerant systems and their application in industry with a focus on automotive systems, chemical processes and renewable energy systems.

BIAO HUANG received the B.Sc. and M.Sc. degrees in automatic control from Beijing University of Aeronautics and Astronautics in 1983 and 1986, respectively, and the Ph.D. degree in process control from the University of Alberta, Canada, in 1997. He joined the Department of Chemical and Materials Engineering, University of Alberta, in 1997, as an Assistant Professor and is currently a Professor, an NSERC Industrial Research Chair in the control of oil sands processes, and an AITF Industry Chair in process control. His research interests include: process control, system identification, control performance assessment, Bayesian methods, and state estimation. He has applied his expertise extensively in industrial practice particularly in oil sands industry. He is a fellow of the Canadian Academy of Engineering and a fellow of the Chemical Institute of Canada. He was a recipient of the Germany's Alexander von Humboldt Research Fellowship, the Canadian Chemical Engineer Society's Syncrude Canada Innovation and D.G. Fisher awards, the APEGA's Summit Research Excellence Award, the University of Alberta's McCalla and Killam Professorship awards, the Petro-Canada Young Innovator Award, and the Best Paper Award from the *Journal of Process Control*.