Annual Reviews in Control 42 (2016) 190–200

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Review

Perspectives on process monitoring of industrial systems☆,☆☆

Kristen Severson, Paphonwit Chaiwatanodom, Richard D. Braatz*

Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 1 September 2016
Accepted 1 September 2016
Available online 13 September 2016

Keywords:
Fault detection
Fault identification
Fault tolerant control
Process monitoring

ABSTRACT

Process monitoring systems are necessary for ensuring the long-term reliability of the operation of industrial systems. This article provides some perspectives on progress in the design of process monitoring systems over the last twenty years. Methods for each step of the process monitoring loop are summarized. The challenges in the field and opportunities for future research are discussed. When looking into the future, it is argued that advances are likely to come from combining different methods to exploit the strengths of various techniques while minimizing their weaknesses.

1. Introduction

Process monitoring is an important component in the long-term reliable operation of any automated controlled system. To distinguish between different types of disruptions on operations, this article adopts the definitions of Isermann and Ballé (1997). A **disturbance** is an unknown and uncontrolled input acting on a system. A **fault** is an unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable/usual/standard operating conditions. A **failure** is a permanent interruption of a system's ability to perform a required function under specified operating conditions. Traditional control systems are designed to return the system to normal operations in the presence of disturbances but not in the presence of faults or failures. *Fault-tolerant control* (FTC) systems refer to control systems that have been designed to explicitly account for some class of specified faults in the closed-loop system. FTC systems must act in the time between a fault and a system failure.

In chemical systems, a fault is an extreme event such as catalyst deactivation, valve blockage or compressor failure. Due to the increasing complexity of facilities, faults are inevitable and occur more often. Monitoring is complicated by recycle streams that cause bidirectional interactions as well as by control systems which can mask the effect of faults. Additionally faults will commonly occur together, known as multiple faults (see Fig. 1). However, even a relatively simple modern facility, in terms of its operations, will have a large sensor network which can be used for process monitoring (see Fig. 2). The key of fault detection and diagnosis (FDD) is how to use these sensors effectively to minimize the impact of faults.

Many process monitoring systems are implemented in the form of a loop that consists of fault detection, fault isolation, fault identification, and process recovery (see Fig. 3). Sometimes the combined steps of fault isolation and identification are referred to as fault diagnosis. The steps are to progressively determine: (1) whether a fault occurred, (2) the location and time of the fault, (3) the magnitude the fault, and (4) how to reverse the effects of the fault (Gertler, 1998).

Process monitoring has been a growing field for nearly a half century. Relevant works on process monitoring in the 1970s include the application by Mehra and Peschon (1971) of systems and statistical decision theory to dynamic systems, the review paper by Willsky (1976) on publications up to the mid 1970s, and the textbook by Himmelblau (1978). Over the years, much of the literature has been focused on particular applications including to aerospace, chemical, nuclear, and automotive systems (Hwang, Kim, Kim, & Seah, 2010). The growing complexity and degree of integration in these systems has increased the possibility that faults occurring locally somewhere in a system can have their effects propagate to other parts of the system, and has made the consequences of designing a poor process monitoring system greater, therefore making the design of process monitoring systems more challenging. As such, many reviews have been published over the last twenty years, e.g. (Alcala & Qin, 2011; Frank & Ding, 1997; Hwang et al., 2010; Isermann, 2005; Isermann & Ballé, 1997; Qin, 2003; Russell, Chiang, & Braatz, 2000a; Venkatasubramanian, Rengaswamy, & Kavuri, 2003a; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003b; Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003c; Yin, Ding, Haghani, Hao, & Zhang, 2012).

☆ BP is acknowledged for funding.

☆☆ A preliminary version of this manuscript was presented at the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (Severson, Chaiwatanodom, and Braatz, 2015a).

* Corresponding author.

E-mail addresses: kseverso@mit.edu (K. Severson), pchaiwat@mit.edu (P. Chaiwatanodom), braatz@mit.edu (R.D. Braatz).

http://dx.doi.org/10.1016/j.arcontrol.2016.09.001

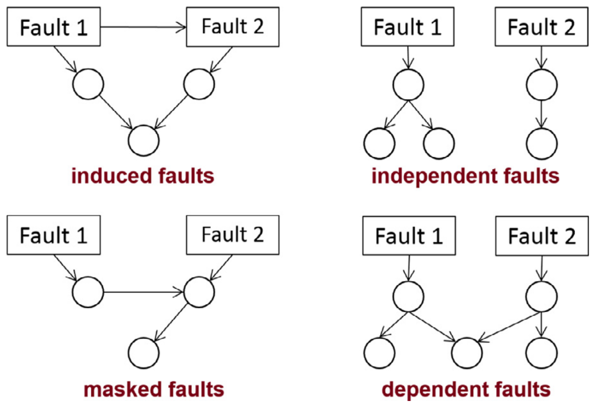

Fig. 1. The four classes of multiple faults (Chiang et al., 2015).

This article does not review the entire process monitoring field which, according to the Web of Science in March 2015, has had over 34,000 publications since the 1970s. This article provides some perspectives on the current state of process monitoring systems as well as current challenges and promising future directions for the field.

2. Process monitoring – background

Modern process monitoring systems are designed based on a model of some form that is developed using process data. The model allows process operators to make informed decisions about whether or not there is a fault. Different fault detection methods provide information of different quality and quantity to the fault diagnosis steps. In this section, each step in the process monitoring loop is presented.

2.1. Fault detection

The design of a fault detection system generally begins with the development of a model that characterizes the normal operating signature of a process. Faults are then typically defined as a deviation from this normal operation above a threshold. As such, the design of a fault detection system can be described as consisting of two steps: building a process model and choosing metrics to test for faults. Active fault detection and identification is an exception to this pattern and is discussed later in the section on process monitoring.

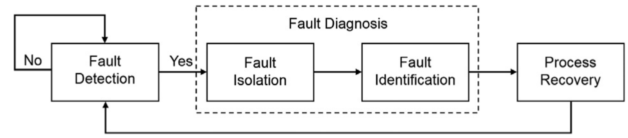

Fig. 3. Process monitoring loop (Isermann & Ballé, 1997; Russell et al., 2000a).

Many types of process models have been employed in fault detection. Principal component analysis (PCA) is one of the most commonly applied fault detection methods for industrial systems. PCA is a linear dimensionality reduction technique that produces lower dimensional representations of the original data that maximize the retained variance (Hotelling, 1933; Jolliffe, 2002). In the absence of noise and disturbances, data from normal operating conditions operate in a much lower dimensional manifold due to physical, chemical, and biological constraints such as Euler's laws of motion, stoichiometry in chemical and/or metabolic reaction networks, and mass, energy, molar species, and fluid momentum balances. In the presence of noise and disturbances, the data from normal operating conditions will approximately lie within a lower dimensional manifold, and data-based dimensionality reduction techniques such as PCA attempt to construct the manifold purely from data.

Variance is a useful metric for fault detection, since it is often reasonable to assume that an outlier as compared to historical operation would indicate a fault. PCA calculates a set of orthogonal vectors, called *loading vectors*, ordered by the amount of variance explained in each loading vector direction using a singular value decomposition. This set of vectors is then truncated, retaining the columns corresponding to the largest singular values. New observations can then be projected into lower dimensional space using the reduced set of loading vectors. The aim of this dimensionality decrease is to keep systematic variations while removing random variations (Wise, Ricker, Veltkamp, & Kowalski, 1990). The technique can be extended to nonlinear systems by us-

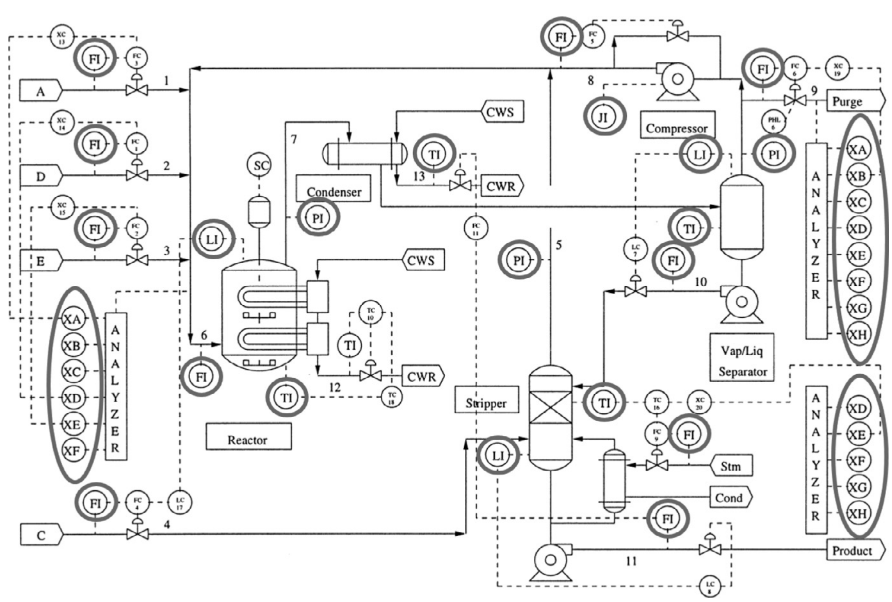

Fig. 2. The process diagram for the Tennessee Eastman (TE) benchmark problem (Downs and Vogel, 1993). The process is a reactor/separator/recycle with two simultaneous gas-liquid exothermic reactions. The process has 12 valves for manipulation and 41 measurements for monitoring and control. The sensors are circled in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

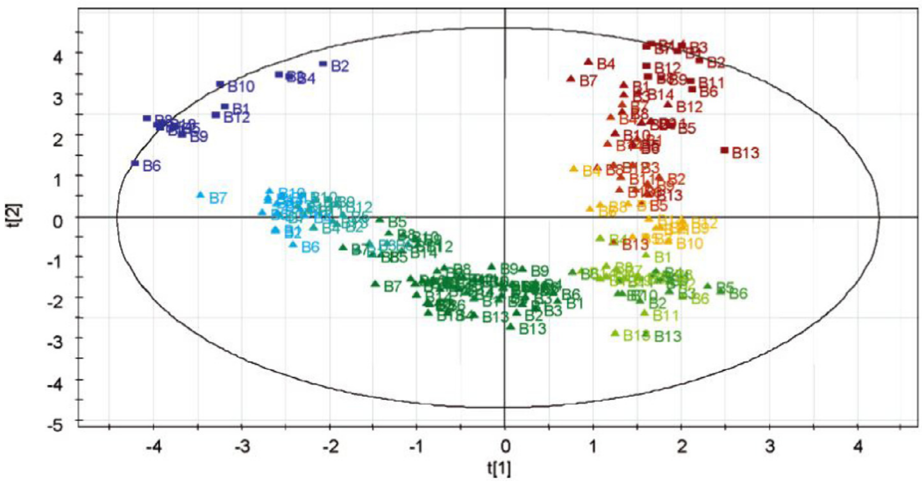

Fig. 4. A scatter plot of experimental data projected onto the leading two loading vectors for a cell culture, with an ellipse designed to contain at least 95% of normal operating data. Reprinted with permission from Kirdar et al. (2007).

ing kernel functions within the PCA formulation (Choi, Lee, Lee, Park, & Lee, 2005). PCA has been applied in a variety of fields including (bio)pharmaceutical manufacturing (Gunther, Conner, & Seborg, 2007; Kirdar, Conner, Baclaski, & Rathore, 2007; Kourti, 2006; Undey, Ertunç, & Çinar, 2003), the chemicals industry (Nomikos & MacGregor, 2014; Zhu & Braatz, 2014), and semiconductor manufacturing (Cherry & Qin, 2006). Fig. 4 shows an example of data plotted in terms of its principal components for a cell culture process.

Partial least squares (PLS, aka projection to latent structures) is another linear dimensionality reduction technique (Wold, Ruhe, Wold, & Dunn III, 1984) widely applied for fault detection in industrial systems. PLS maximizes the covariance between the input and output data in the reduced space (Geladi & Kowalski, 1986). Unlike PCA, PLS does not have a closed-form solution but instead uses an iterative algorithm such as NIPALS (Wold, 1975). PLS is widely applied in the chemicals, petrochemicals, and refining industries (Russell et al., 2000a; Zhou, Li, & Qin, 2010) and in pharmaceutical and biologic drug manufacturing (Kirdar et al., 2007; Severson et al., 2015b). The low cost of entry of chemometrics (PCA, PLS) methods and the lack of dynamic models for most plant operations are the main reasons for their dominance in these industries. Both their current heavy usage and the ever-increasing quantity of real-time data (Reis, Braatz, & Chiang, 2016) suggests that chemometrics methods will continue to dominate those industries for the foreseeable future.

An alternative to fault detection methods that rely on dimensionality reduction are methods based on state-space models. The most commonly used model is the discrete-time linear stochastic state-space model

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + G\mathbf{u}_k + \mathbf{w}_k \tag{1}$$

$$\mathbf{y}_k = H\mathbf{x}_k + A\mathbf{u}_k + B\mathbf{w}_k + \mathbf{e}_k \tag{2}$$

where k is the sampling index; $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{y}$ are the system states, inputs, and outputs, respectively; and $\mathbf{w}$ and $\mathbf{e}$ denote the sensor and process noise of the system (Larimore, 1983; Simoglou, Martin, & Morris, 2002). Such models are typically constructed from subspace identification techniques, such as canonical variate analysis (CVA) (Larimore, 1990), multivariable output-error state-space (MOESP) (Verhaegen, 1993a; 1993b; 1994; Verhaegen & Dewilde, 1992a; 1992b), and numerical algorithm subspace-based state-space system identification (N4SID) (Van Overshcee & De Moor, 1994). The subspace identification techniques most applied to industrial systems is CVA, which was pioneered by Akaike (1976) and promoted and further developed by Larimore (1990). The objective of CVA is to identify a linear combination of past inputs and outputs that are most predictive of future outputs. CVA relies on minimizing the prediction error using a singular value decomposition of the covariance matrix for past inputs and outputs. CVA has been reported to produce near maximum-likelihood solutions (Juricek, Seborg, & Larimore, 2001). Another type of identification technique uses fuzzy rule-based models. In this approach, fuzzy clustering techniques are used to partition the data into linear subsets (Tomohiro and Sugeno, 1985). This approach was originally proposed for modeling and control and then extended to fault diagnosis (Simani, 2013).

Another class of fault detection models relies on graphical models, which are typically directed and often lumped into the broader class of knowledge-based methods. These methods employ some form of expert knowledge in their construction. A decision tree is a type of graphical model developed via inductive learning that aims to map measured data to classes of operating conditions. These models are able to describe normal and abnormal operations during complicated startup, shutdown, and changeover procedures (such systems are often called *mixed continuous-discrete systems* or *hybrid systems*). Feature selection and extraction are important considerations for the success of decision trees and are facilitated by process understanding. A benefit of this approach is that a well-developed graphical model has an easily interpreted physical meaning (e.g., see Fig. 5), and that the same model can be used in fault identification and isolation (Bakshi & Stephanopoulos, 1994).

Several other types of graphical models have also been applied in the field, with representative examples being causal maps, Petri nets, bond graphs, and neural networks. A *causal map* is a directed graph where the nodes represent process variables and the directed edges represent cause-and-effect relationships (Chiang, Russell, & Braatz, 2001). A model of this type has a clear physical interpretation, and can be constructed from a piping and instrumentation diagram or process flow diagram embedded in the distributed control system. A *Petri net* is a graphical model that is suitable for modeling transitions/events that may occur in the operation of the system and is most well-suited to graphs that have parallel or concurrent events (Murata, 1989). The graph consists of transitions, places, and arcs in which nodes can be transitions or places (marked with different symbols, typically bars and circles) and arcs connect nodes of different type only. Petri

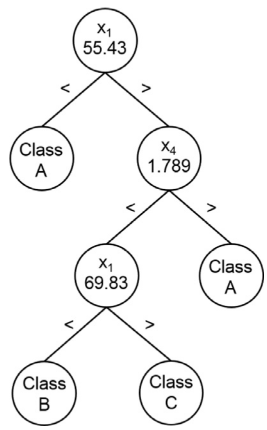

Fig. 5. An example of a decision tree as applied to a process for maintaining octane number of a gasoline product adapted from Bakshi and Stephanopoulos (1994).

nets were first introduced by Petri (1962) and conferences and several tutorials helped popularize the technique (Murata, 1989). Viswanadham (1988) was one of the first researchers to apply Petri nets to fault detection applications, which was followed by a large number of studies (e.g., see Boubour, Jard, Aghasaryan, Fabre, and Benveniste, 1997; Cabasino, Giua, and Seatzu, 2010; Srivinvasan and Jafari, 1993 and citations therein). Rather than focus on transitions, *a bond graph* is a graphical representation of a physical dynamical system that represents its energy flows Borutzky (2009). Bond graphs were first introduced by Paynter (1961), and examples of the application of bond graphs to fault detection include Feebstra, Mosterman, Biswas, and Breedveld (2001); Low, Wang, Arogeti, and Luo (2010). A *neural network* is a graphical model that is characterized by input, output and hidden nodes. When applied to fault diagnosis problems, often the input nodes represent the measurement space, the hidden nodes represent the feature space, and the output nodes represent the decision space (Venkatasubramanian et al., 2003a). Examples of the application of neural networks can be found in Korbicz, Kośielny, Kowalczuk, and Cholewa (2004); Mrugalski (2013, 2014).

Once the model has been determined, a metric for detecting faults is required. In PCA, PLS, and related models, faults are usually detected using the T^2 statistic, which is the Euclidean norm of the deviation of an observation vector from its mean in the reduced space, scaled by its variance. A fault is detected when the T^2 value exceeds a specified threshold. Alternatively, the Q statistic (also known as standard prediction error or SPE), which measures the total sum of variations in the residual space, can also be used to identify faults. In extensive simulations, the Q statistic has been observed to be usually more effective at detecting faults than the T^2 statistic (Chiang et al., 2001). The explanation for this observation is that most faults push the process operations outside of the normal linear relationships between variables rather than magnify the extent of operation within the normal linear relationships between variables. Some researchers have used a weighted combination of Q and T^2 statistic (Yue & Qin, 2001).

For knowledge-based models, faults are detected if the measured variables result in a prediction of a fault, based on the model.

If a state-space model is used, fault detection typically occurs via a similar residual generation, which compares model predictions and measurements (often referred to as *output estimation* approaches). Alternatively, the difference between nominal and estimated parameters has been used to detect faults (often referred to as *parameter estimation* approaches). In the particular case of CVA, a series of different statistics have been proposed for fault detection Chiang and Braatz (2003); Juricek, Seborg, and Larimore (2004). In state-space models, fault detection and diagnosis are closely coupled, as discussed below.

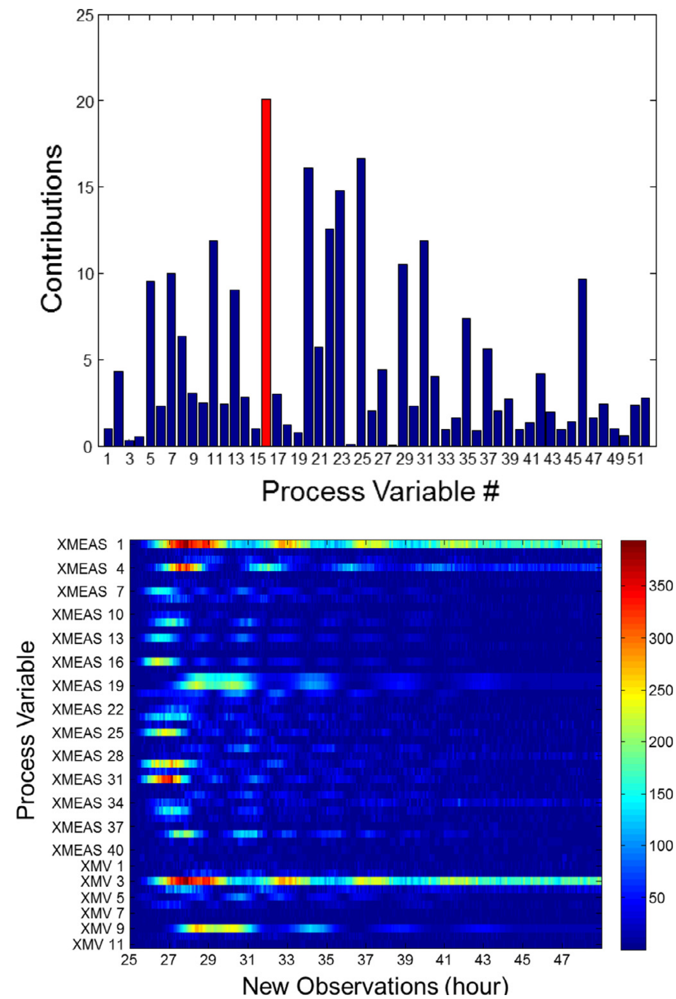

Fig. 6. Classic contribution chart (top) and 2D contribution map (bottom). The contributions are for a fault in a simulated chemical manufacturing facility (Zhu & Braatz, 2014).

2.2. Fault isolation

Once the fault has been detected, the next step is to determine the location of the fault. One fault isolation method widely used in industrial systems is the contribution chart. Contribution charts are typically used in concert with dimensionality reduction techniques, such as PCA and PLS. The contribution chart projects the data back into the higher dimensional observation space, which can be used by an operator to identify which process variables are deviating from their historical values. This approach exploits correlations between variables to reduce the effects of process and sensor noise on identifying which observation variables are most likely associated with the fault.

As an example, a contribution chart of the TE process is shown in Fig. 6, both in the form of a classic contribution chart at one time instance and in the form of a 2D contribution map with the contributions in each column as a function of time in the form of a color map. The 2D contribution map, introduced by Zhu and Braatz (2014), allows the operator to visualize the dynamic propagation of the effects of a fault on the observation variables through the facility. The 2D plot shows which deviations are suppressed by the various control systems and which deviations are persistent. The variables which have deviations can be compared to the process flow diagram or piping and instrumentation diagram to better track down the location of the root cause of the fault. The ability to visualize the data so that the fault can be located is a crucial element for the success of a method. Methods that allow for easy interpretation of the data are much more valuable in industrial application.

An alternative to the classic contributions chart, referred to as *reconstruction-based contribution chart*, has been proposed (Alcala and Qin, 2008). This method finds the contribution of each monitored variable to the fault detection metric, for example T^2. An example is given where the reconstruction-based contribution chart provides an accurate fault isolation while the classic contribution chart cannot.

2.3. Fault identification

Fault identification can be very challenging if fault detection and isolation have been carried out using PCA or PLS, as the quality of information that can be extracted from models constructed from normal operating data is limited. If the training data has been characterized from past experience into normal operating conditions and specific faulty conditions, then Fisher discriminant analysis (FDA) is a dimensionality reduction method that can be used for fault identification.

FDA maximizes the separation (aka scatter) among different classes while minimizing the scatter within each class (Duda & Hart, 1973). The formulation of the problem is

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\neq 0} \frac{\mathbf{v}^\top S_b \mathbf{v}}{\mathbf{v}^\top S_w \mathbf{v}} \tag{3}$$

where

$$S_b = \sum_{j=1}^{p} n_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^\top, \tag{4}$$

$$S_w = \sum_{j=1}^{p} \sum_{\mathbf{x}_i \in \mathcal{X}_j} (\mathbf{x}_i - \bar{\mathbf{x}}_j)(\mathbf{x}_i - \bar{\mathbf{x}}_j)^\top, \tag{5}$$

$\mathbf{x} \in \mathcal{R}^m$, $\bar{\mathbf{x}}$ is the total mean vector, $\bar{\mathbf{x}}_j$ is the mean vector for class j, n_j is the number of observations in class j, and p is the number of classes, and sufficient data have been collected that the matrix S_w is nonsingular. For FDA to be well-defined, at least two sets of characterized data are required (e.g., normal operating conditions and data collected during one fault).

FDA models can be more specific about *which* fault is occurring, if they have been trained using multiple fault classes and that set is comprehensive. Often the amount of faulty data is limited in practice and each fault will require its own investigation once the fault isolation step using the contribution chart is complete.

Another data-based fault diagnosis technique that attempts to reduce the dimensionality of the problem is support vector machines (SVMs). SVM methods find a separating hyperplane which is specified by a number of support vectors (samples). These support vectors typically represent a small subset of the complete dataset used for analysis. The separating hyperplane is oriented in such a way as to maximize the distance, called the *margin*, between the plane and the nearest point of each class (Bishop, 2007). SVMs can be formulated using a kernel, which is amenable to feature selection. This technique has gained increased interest in the past 15 years due to efficient optimization formulations (Platt, 1998). Since then, SVMs have been tested in mechanical engineering applications (Baccarini, Rocha E Silva, De Menezes, & Caminhas, 2011; Widodo & Yang, 2007) and semiconductor manufacturing (Mahadevan & Shah, 2009). Like FDA, SVMs are typically trained with labeled target data and therefore require data that have been collected during past faults and, for best results, specific faults must be associated with each data set. Both FDA and SVMs are ineffective for fault diagnosis if data have not been collected during past fault states.

Most fault diagnosis methods based on state-space models assume that the fault is either *additive* or *multiplicative* (Chen & Patton, 1999; Kościelny & Łabęda-Grudziak, 2013). An additive fault is assumed to be well represented in terms of a vector added to the fault-free state-space equations, whereas a multiplicative fault is assumed to be well represented by a deviation in a parameter in the state-space matrices; for this reason, multiplicative faults are also commonly referred to as *parametric faults*.

Multiplicative faults can be diagnosed by determining which online parameter estimates have the largest deviations from nominal values. This method is sufficiently general to be applicable to nonlinear dynamical systems. A weakness of this method is that it requires that the data are sufficiently rich in information to be able to accurately estimate parameters online. Fault diagnosis occurs via the link between the parameters and the physical system. If the model parameters are not tied to physical parameters, diagnosis abilities are limited. In particular, deviations in the elements of state-space models constructed from subspace identification methods are not tied to any physical parameters, so this approach provides little value for such models.

Observer-based methods are most commonly used for diagnosing additive faults. In observer-based methods, the residuals between estimated and measured outputs are used for detection and diagnosis. As an example, consider the full-order state estimator

$$\hat{\mathbf{x}}_{k+1} = A\hat{\mathbf{x}}_k + B\mathbf{u}_k + H(\mathbf{y}_k - \hat{\mathbf{y}}_k) \tag{6}$$

$$\hat{\mathbf{y}}_k = C\hat{\mathbf{x}}_k \tag{7}$$

where $\hat{\mathbf{x}}$ is the predicted state, $\hat{\mathbf{y}}$ is the predicted output, $\mathbf{y}$ is the measured output, and the observer gain H is chosen to satisfy design criteria such as stability, fault sensitivity, and robustness. For a linear process with additive faults, the residuals are

$$\begin{aligned}\Delta\mathbf{x}_{k+1} = &(A - HC)\Delta\mathbf{x}_k + (B_f - HD_f)\mathbf{f}_k \\ &+ (B_d - HD_d)\mathbf{d}_k\end{aligned} \tag{8}$$

$$\mathbf{r}_k = \Delta\mathbf{y}_k = C\Delta\mathbf{x}_k + D_f\mathbf{f}_k + D_d\mathbf{d}_k \tag{9}$$

where $\Delta\mathbf{x}_k$ is the state estimation error. The residuals are a function of both the faults and disturbances. In large-scale systems, disturbances can be significant, which motivates the use of transformed output errors as the residual,

$$\mathbf{r}_k = W\Delta\mathbf{y}_k. \tag{10}$$

where the matrix W is designed such that the residuals are insensitive to disturbances but sensitive to faults. One common method for designing both matrices H and W is the unknown input observer (UIO) method. This method attempts to design the observers such that the effects of disturbances approach zero asymptotically (Simani, Fantuzzi, & Patton, 2003). The isolation and identification steps then occur via a structured residual set, where *structured* implies that each residual is designed to be sensitive to only one particular fault (Chen & Patton, 1999).

The above approach generalizes directly to nonlinear dynamical systems and to models with explicit uncertainty descriptions—the latter known as *robust observer-based fault diagnosis* methods (Gertler, 1998). A challenge in applying the latter methods to industrial systems is that the requirement of having accurate models of the nominal system, the faults, the disturbances, process noise, and the structure of the model uncertainties.

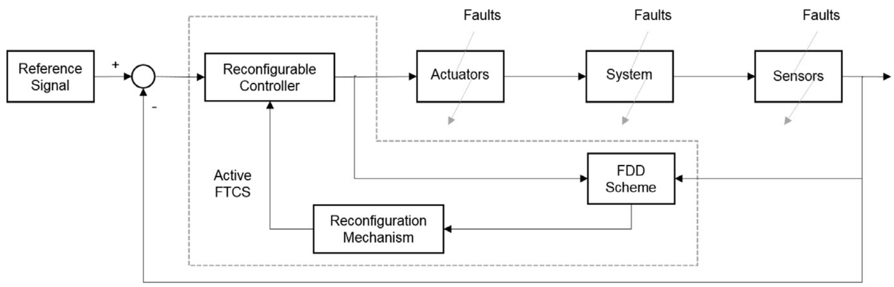

Fig. 7. Architecture of an active FTC, adapted from Jiang and Yu (2012).

2.4. Process recovery

The end goal of all process monitoring is process recovery, where the process is returned to its normal operation. Most FDD methods will require manual intervention once a fault has been diagnosed. Fault-tolerant control (FTC) refers to a control system that automatically performs process recovery, that is, without real-time human intervention (Patton, 1997; Witczak, 2014; Zhang and Jiang, 2008).

The objective of FTC can be interpreted as treating faults as if they are disturbances, to return the system to acceptable operation either via retuning or restructuring the control system (MacGregor & Cinar, 2012). FTC can generally be divided into two methodologies: passive and active. In passive FTC, the process monitoring system observes the process data and decides if a fault has occurred, using methods as described in Section 2, with the fault classes known *a priori*. The control system is designed with redundancies so that it is not necessary to reparameterize or restructure the controller during faulty operation. If there is more than one system fault possible, this approach often leads to a conservative controller design with slow closed-loop performance (Jiang & Yu, 2012).

In active FTC, depending on what conditions are detected, the controller is reconfigured for that scenario (see Fig. 7). A major challenge of active FTC is the coordination of the process monitoring and control systems (Raimondo, Marseglia, Braatz, & Scott, 2013b). Furthermore, most FTC design methods assume that faults are detected and isolated correctly and instantaneously, to allow for computational tractability (Raimondo et al., 2013b).

Some recent algorithms combine active FDD with active FTC. Active FDD uses a test signal, called an *auxiliary input*, to generate data that enables more effective determination of whether a fault has occurred or, if a fault has been detected, which fault has occurred. This approach addresses one of the major issues of passive FDD, which can have difficulties identifying faulty conditions because the process can mask faults, particularly if the process is under control (Nikoukhah, 1998). One method of active FDD is set based, which aims to find the separating inputs which guarantee fault diagnosis (Nikoukhah, 1998; Raimondo, Braatz, & Scott, 2013a; Scott, Findeisen, Braatz, & Raimondo, 2013a; 2014). This technique has been combined with model predictive control to guarantee diagnosability given input and state constraints for linear systems (Raimondo et al., 2013b). These methods are formulated for discrete-time models. Unlike the generalization of many results from discrete-time models to continuous-time models, the generalization of these results to continuous-time models would be challenging.

2.5. Comparisons of classical methods

Each process monitoring method has advantages and disadvantages. The data-based dimensionality reduction techniques of PCA and PLS are easy to implement for fault detection and isolation but of limited value for fault identification. Graphical models have the ability to incorporate expert knowledge, which is a positive if such information is available, but also require expert knowledge in their construction, which is a negative if such information is not available. State-space models require a lot of investment to develop and maintain for an industrial system, but have the potential for including very precise information on faults and disturbances in fault diagnosis procedures. The research area of process monitoring is still very active as researchers aim to tackle some of the drawbacks of various methods.

3. Challenges and opportunities

In the past twenty years, the quantity of data that can be collected and processed for industrial processes has greatly increased. The development of new tools such as smart and wireless sensors, the Internet of Things, smart devices, and smart manufacturing has allowed the amount of available data to grow exponentially (Qin, 2014). Although FDD methods are often categorized as model-, data-, or knowledge-based, all FDD models require process data for validation and successfully utilizing this data is a key challenge and opportunity for the continued improvement of process monitoring. This section presents challenges in the field that could be addressed using this new data and methods tailored to such data.

Increasingly, these new datasets are referred to as *Big Data*. Big Data is characterized by four characteristics referred to as the 4 V's: velocity, volume, variety, and veracity (IBM, 2016). So that our discussion of challenges and opportunities fit into this framework, we will refer back to these characteristics throughout the section.

Although this section focuses on methods, it is useful to first comment about data infrastructure. Because the very large size of the data (volume), and the quick rate at which data are collected (velocity), new data systems are required. Data-centric architectures and distributed storage and processors need to be used for the value of Big Data to be realized (Qin, 2014). In other words, the data are useless if the data cannot be accessed and processed reliably with reasonable computational cost. Waiting a longer time to access the data and compute a useful result from the data is not always an option, as the time available for making decisions based on the data is constrained by the time in which such decisions would be useful. This consideration is especially important in process monitoring, as faults need to be detected and diagnosed quickly enough that damage to the system is limited. A technology for improving access to Big Data is Hadoop (O'Malley et al., 2016), which is a distributed file system and distributed computing framework specifically designed to handle Big Data. All modules in Haddop are designed to automatically handle any computer hardware failures, such as crashes of processors within computer clusters, with minimal disruption on the calculations applied to the data. More recently, Spark, an open-source processing engine developed at UC Berkeley, has been gaining popularity as an additional tool for Big Data analytics (databricks, 2016).

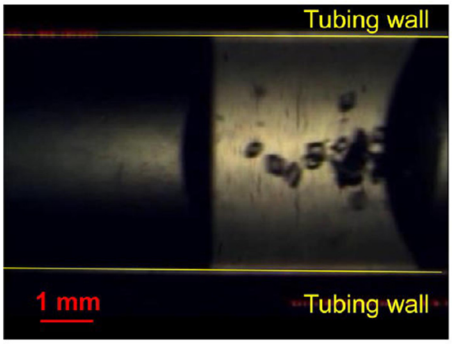

Fig. 8. An in-line stereomicroscope image for the monitoring system of a crystallization process in which particles are in liquid slugs that flow down a tube (Jiang et al., 2014). Many such images are collected each second in real-time video. This type of data highlights the high-order structures occurring in modern datasets.

3.1. Utilizing new data sources

Beyond needing to handle a "black-box" of data as described above, new methods are required to handle new features (variety) of Big Data datasets. One of these features is high-dimensional data. In high-dimensional data, it is often the case that there are many more measurements per sample than samples, which can lead to ill-conditioning. Methods such as PCA address ill-conditioning by projecting the data into a lower dimensional space. However, with the increase in the new of measurements, there may be motivation to select a subset and not a subspace. A subset may allow for a decrease in the number of sensors which can be desirable to decrease maintenance and data storage costs. To find subsets, several avenues exist such as subset selection via optimization, penalty methods, and greedy methods. One approach is to use mutual information as the selection criteria for a greedy approach (Verron, Tiplica, & Kobi, 2008). A drawback of the greedy approach is the lack of optimality guarantees. Mutual information is also not necessarily the best metric. Research is needed in this area to better understand tradeoffs between the number of sensors and the accuracy of the model. This issue is inherently intertwined with design of experiments for new process development. Experiments should be planned with process monitoring in mind such that the most valuable data can be extracted for the lowest cost while still considering standard operations. The issue of the connection of data-based monitoring and process design has not yet been solved.

Another feature of Big Data is the presence of higher-order tensors associated with new types of measurements such as real-time spectroscopic imaging or video. Instead of vectors or matrices, a single "measurement" can consist of third-, fourth-, or higher-order tensors. An example would be an inline imaging system used to characterize the shape properties of crystals in fluid flow (see Fig. 8), in which a single measurement at a time instance is a second-order tensor (aka matrix), with the two dimensions being space along horizontal and vertical axes, with each pixel being a grey-scale value between 0 and 255. Typically such data are collected at many frames per second at time scales much faster than the process time scales, with few particles per image. To obtain statistically reliable measurement, each measurement is treated as a video collected from seconds to minutes, which consists of many individual images (aka frames). This measurement constitutes a third-order tensor with the third dimension being the time axis over a short period of time. For color imaging systems, the order of the tensor increases by one, with the additional dimension being the color axis for red, green, and blue. The data are stored as a number between 0 and 255 for red, green, and blue at each pixel, for a two-dimensional array of pixels that make up an image. When the measurement is video over a short time period, a single measurement is a fourth-order tensor (that is, two physical dimensions, color, and time). Stacking the data into vectors and then applying PCA and PLS methods is suboptimal in practice, and such methods ignore the inherent correlations and internal structure that such datasets possess, such as that neighboring images in a video have dominant signals being shifted slightly in space as particles move. The quality of model predictions based on such data would be improved if higher order correlations and internal structure were explicitly exploited by the methods.

A related feature of Big Data is heterogeneity. New data sources are increasingly heterogeneous in terms of types and time scale. For instance, some data in the bioprocess industry are collected online, such as dissolved oxygen in a bioreactor as a function of time, while other data are collected offline, such as cell density (Charaniya, Hu, & Karypis, 2008). Both sets of data provide valuable information about the status of the bioreactor, and new methods are needed for efficient integration. Some level of integration can be obtained via similarity scores and kernel transformations (Charaniya et al., 2008), but a lot of research is needed to generate optimal methods. Methods developed to apply to Big Data need to be able to handle rare-event data well. In fault detection, because the goal is often to find an anomaly, careful attention must also be given to data cleaning. Data cleaning is a process of removing faulty data while still retaining unexpected values. If an analysis does not take care in handling data cleaning, the behavior of interest can be overlooked.

3.2. Semi-supervised and online learning

Another challenge deals with using all available data. Here, specifically, the interest lies in using unlabeled data that is readily available from operations. Particularly in industrial applications, it is not reasonable, for safety or financial concerns, to purposely generate faulty data for training process monitoring algorithms. Therefore, datasets to be used for process monitoring are inherently unbalanced and methods attempt to characterize nominal operations without access to faults. In a best-case scenario, a small subset of the data is labeled as associated with some fault, but most data are not. In this setting, a state-space model using either parameter or prediction residuals may be successful, but such models are expensive to develop and maintain for complex industrial systems. Data-based methods such as PCA may be successful, but have limited capability for fault identification. Therefore semi-supervised and online learning methods should be a focus of future research.

Unsupervised learning refers to model building without knowledge of the true value of the output. Clustering and density estimation are common examples of unsupervised learning (Bishop, 2007). The opposite approach is supervised learning, where the targets are known. Supervising learning is ideal, but typically unreasonable in fault detection applications for the aforementioned reasons. Semi-supervised learning is in-between, where some but not all targets are known. In online learning, sometimes also referred to as *sequential learning*, the model is continually updated as additional data become available (Bishop, 2007; Murphy, 2012). These methods are more suited to the constraints of the fault detection problem. Some work in these areas is already being done. In Jin and Shi (2001), the set of features that characterize faults are calculated online as new data are streaming. The approach requires limited to no prior fault information. Jin and Shi (2001) apply the approach to the monitoring of stamping tonnage signal analysis and are able to detect faults related to shut height, which is a common process variable in these operations. Another example is Zhao, Ball, Mosesian, de Palma, and Lehman (2015), who also develop a technique that adapts over

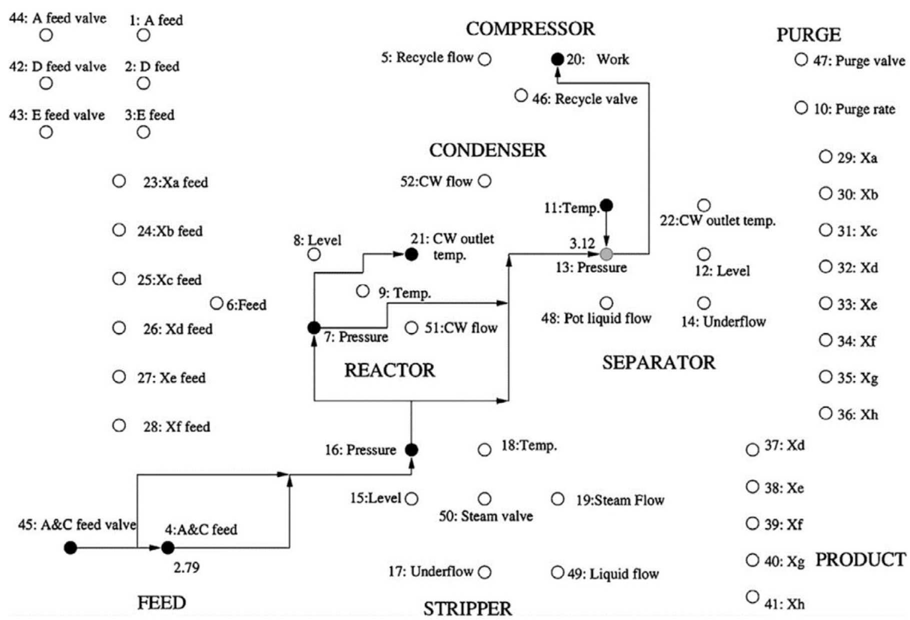

Fig. 9. Visualization of a fault propagation using the Tennessee Eastman benchmark problem (Chiang and Braatz, 2003).

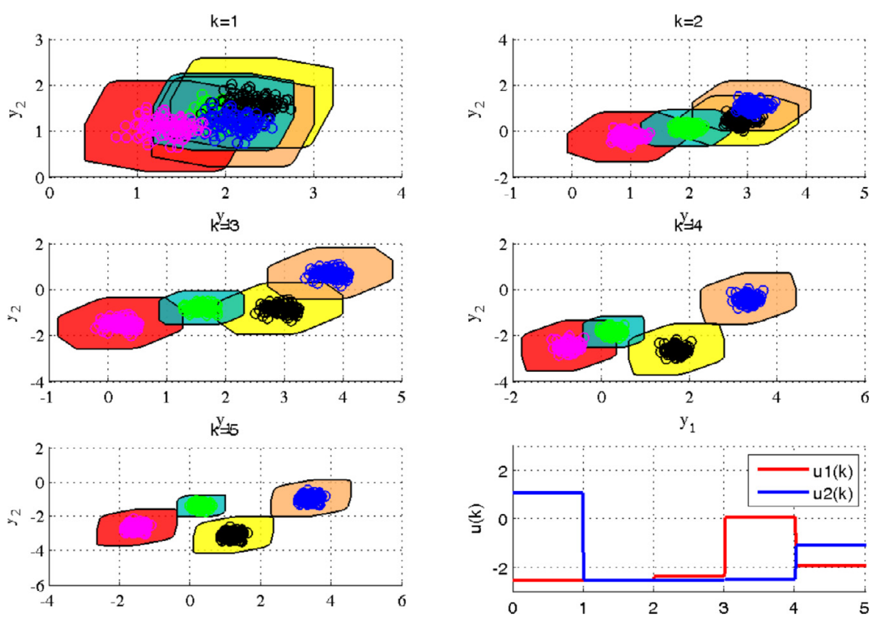

Fig. 10. Reachable output sets of nominal and faulty models using the hybrid stochastic-deterministic approach implemented with the input $\tilde{u}_1$, which guarantees separation in five steps and approximately maximizes the probability of diagnosis in three steps (Marseglia et al., 2014).

time. In their work, a small set of labeled data to train the model, i.e. semi-supervised. Ge and Song (2011) also use a semi-supervised approach, although for the goal of process modeling and not fault detection.

Semi-supervised, unsupervised, and online learning methods are gaining increased focus in the machine learning literature. The fault detection and diagnosis community would benefit from leveraging results from the machine learning community, by tailoring the methods to the specific needs of FDD problems. Some examples of methodologies for utilizing unlabeled data are support vector machines (SVM) (Schölkopf, Platt, Smola, & Williamson, 2001; Xu & Schuurmans, 2005) and Parzen density estimates (Parzen, 1962). Many advances have been made more recently in deep learning (LeCun, Bengio, and Hinton, 2015) and the leveraging of such advances in FDD would be interesting.

3.3. Addressing process uncertainty

Another challenge is most closely related to data veracity. Reliable process monitoring can often be limited due to process uncertainties, which inhibit interpretation of process data (Campbell & Nikoukhah, 2004). Much of the past work has focused on deterministic bounded uncertainties, while some newer work has shifted that focus towards formulations that utilize probability distributions to characterize the uncertainties. For example, Zhong, Ding, Lam, and Wang (2003) consider uncertainty in the inputs and parameters of linear systems and propose reducing the robust fault detection problem to a standard H_∞ model-matching problem. The central concept of the work is to find a robust fault detection filter. As an example of handling probabilistic uncertainties, Mesbah, Streif, Findeisen, and Braatz (2014) proposed one such system that treats probabilistic uncertainties in the parameters and initial conditions of a nonlinear system, and utilizes polynomial chaos theory for uncertainty propagation. The input design is then performed using a constrained nonlinear optimization. Readers interested in robust process monitoring methods are encouraged to read the papers cited in the above publications.

4. Hybrid methods

The next generation of process monitoring systems need to meet a variety of needs including reliability, ability to handle uncertainty, and ability to utilize large quantities of data. An important technique for handling these demands is the use of hybrid methods that capture the strengths of different methods while minimizing their weaknesses. This section highlights some examples of hybrid models.

One example is the approached used by Chiang and Braatz (2003) as applied to the Tennessee Eastman benchmark problem. Their technique aimed to improve upon PCR/PLS which ignores information on process connectivity by instead using a causal map and a modified distance metric. A causal map is easily developed in many chemical applications using existing process flow or piping and instrumentation diagrams. This causal map is a type of graph that can then be combined with information theory and multivariate statistics to measure changes in the distributions of variables and in relationships between distributions of causally related variables. Furthermore, because the directed graph is directly related to the process, fault propagation could be visualized in real time (see Fig. 9 for an example of this visualization).

Another example of a hybrid method is the CVA-FDA method proposed by Jiang, Zhu, Huang, Paulson, and Braatz (2015c). This method was implemented to tackle the challenge of fault identification and diagnosis in the presence of data overlap. This work was also applied to the Tennessee Eastman benchmark. Initially FDA was applied to the problem but it was determined that the data had too much serial correlation for FDA to provide good separation. Therefore, drawing from the state-space literature, the authors first applied CVA then FDA to handle the serial correlations and then perform fault diagnosis and identification. Using this technique decreased the misclassification rate by approximately 40% compared to using FDA alone (Jiang et al., 2015c).

A third example of the power of hybrid methods relates to active FDD. Active FDD methods are largely either stochastic or set-based. Stochastic methods provide convenient descriptions but do not provide guarantees, whereas set-based methods compute hard bounds but are often based on worst-case uncertainty. Hybrid methods were proposed by Scott, Marseglia, Magni, Braatz, and Raimondo (2013b) and Marseglia, Scott, Magni, Braatz, and Raimondo (2014) to compromise between these two methodologies by using model uncertainties described by pdfs of finite support but also guaranteed correct diagnosis at a given time, N, while maximizing the probability of correct diagnosis at some earlier time (see Fig. 10). These approaches provide better flexibility compared to using purely stochastic or purely deterministic approaches.

Many other examples of hybrid approaches are described in the literature, e.g. Chiang and Braatz (2003); Chiang, Jiang, Zhu, Huang, and Braatz (2015); Chiang, Russell, and Braatz (2000); Jiang, Huang, Zhu, Yang, and Braatz (2015a); Jiang, Zhu, Huang, and Braatz (2015b); Maki, Jiang, and Hagino (2001); Russell, Chiang, and Braatz (2000b). The process monitoring field has been increasingly focused on complex and high-value processes over the past 40 years. Hybrid systems show the most promise for being able to handle the fault scenarios that arise in such systems.

5. Conclusions and future directions

This article provides an overview of process monitoring methods and introduces the major challenges facing the next generation of techniques. The article advocates for the use of hybrid methods to address these challenges in modern and complex facilities and provides some examples of how hybrid methods have been successful in past studies. The process monitoring field would benefit from increased sharing of data for the comparative evaluation of process monitoring systems. The machine learning community has benefited greatly from the availability of public data sources, for example, the Wall Street Journal corpus used for speech recognition and natural language processing (Paul & Baker, 1992), the PASCAL challenge for image recognition (Everingham, Van Gool, Williams, & Zisserman, 2005), and the MNIST dataset for digit recognition (LeCun, Cortes, & Burges, 1998). The FDD community is also heavily dependent on data. Robust and implementable models need real process data for training and testing. The Tennessee Eastman chemical manufacturing facility meets this need in many ways (Chiang et al., 2001). However, the community would benefit from additional data, particularly real data or from a different manufacturing setting such as pharmaceutical manufacturing or oil well data. Progress in process monitoring systems would benefit from the availability of public datasets for comparative studies to focus on the most promising directions in algorithm development.

References

Akaike, H. (1976). Canonical correlations analysis of time series and the use of an information criterion. *Mathematics in Science and Engineering, 126*, 27–96.

Alcala, C., & Qin, S. J. (2008). Reconstruction-based contribution for process monitoring. In *Proceedings of the IFAC world congress* (pp. 7889–7894).

Alcala, C. F., & Qin, J. S. (2011). Analysis and generalization of fault diagnosis methods for process monitoring. *Journal of Process Control, 21*(3), 322–330.

Baccarini, L. M. R., Rocha E Silva, V. V., De Menezes, B. R., & Caminhas, W. M. (2011). SVM practical industrial application for mechanical faults diagnostic. *Expert Systems with Applications, 38*(6), 6980–6984.

Bakshi, B., & Stephanopoulos, G. (1994). Representation of process trends – IV. Induction of real-time patterns from operating data for diagnosis and supervisory control. *Computers & Chemical Engineering, 18*(4), 303–332.

Bishop, C. (2007). *Pattern recognition and machine learning*. New York: Springer.

Borutzky, M. (2009). Bond graph modelling and simulation of multidisciplinary systems – An introduction. *Simulation Modelling Practice and Theory, 17*, 3–21.

Boubour, R., Jard, C., Aghasaryan, A., Fabre, E., & Benveniste, A. (1997). A Petric net approach to fault detection and diagnosis in distributed systems. Part I: Application to telecommunication networks, motivations and modelling. In *Proceedings of the IEEE conference on decision and control* (pp. 720–725).

Cabasino, M. P., Giua, A., & Seatzu, C. (2010). Fault detection for discrete event systems using Petri nets with unobservable transitions. *Automatica, 46*, 1531–1539.

Campbell, S. L., & Nikoukhah, R. (2004). *Auxiliary signal design for failure detection*. New Jersey: Princeton University Press.

Charaniya, S., Hu, W. S., & Karypis, G. (2008). Mining bioprocess data: Opportunities and challenges. *Trends in Biotechnology, 26*(12), 690–699.

Chen, J., & Patton, R. J. (1999). *Robust model-based fault diagnosis for dynamic systems*. Boston: Springer.

Cherry, G. A., & Qin, S. (2006). Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. *IEEE Transactions on Semiconductor Manufacturing, 19*(2), 159–172.

Chiang, L. H., & Braatz, R. D. (2003). Process monitoring using causal map and multivariate statistics: Fault detection and identification. *Chemometrics and Intelligent Laboratory Systems, 65*, 159–178.

Chiang, L. H., Jiang, B., Zhu, X., Huang, D., & Braatz, R. D. (2015). Diagnosis of multiple and unknown faults using the causal map and multivariate statistics. *Journal of Process Control, 28*, 27–39.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2000). Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares and principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 50*, 240–252.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems*. London: Springer.

Choi, S., Lee, C., Lee, J., Park, J., & Lee, I. (2005). Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems, 75*, 55–67.

databricks (2016). Apache Spark. https://databricks.com/spark/.

Downs, J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering, 17*, 245–255.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley and Sons.

Everingham, M., Van Gool, L., Williams, C., & Zisserman, A. (2005). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision, 88*(2), 303–338.
Feebstra, P. J., Mosterman, P. J., Biswas, G., & Breedveld, P. C. (2001). Bond graph modeling procedures for fault detection and isolation of complex flow processes. *Simulation Series, 33*, 77–84.
Frank, P. M., & Ding, X. (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control, 7*(6), 403–424.
Ge, Z., & Song, Z. (2011). Semisupervised Bayesian methods for soft sensor modeling with unlabeled data samples. *AIChE Journal, 57*, 2109–2119.
Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta, 185*, 1–17.
Gertler, J. (1998). *Fault detection and diagnosis in engineering systems*. New York: Mercel Dekker.
Gunther, J. C., Conner, J. S., & Seborg, D. E. (2007). Fault detection and diagnosis in an industrial fed-batch cell culture process. *Biotechnology Progress, 23*(4), 851–857.
Himmelblau, D. M. (1978). *Fault detection and diagnosis in chemical and petrochemical processes*. New York: Elsevier.
Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417–441.
Hwang, I., Kim, S., Kim, Y., & Seah, C. E. (2010). A survey of fault detection, isolation, and reconfiguration methods. *IEEE Transactions on Control Systems Technology, 18*(3), 636–653.
IBM (2016). The four v's of big data. http://www.ibmbigdatahub.com/infographic/four-vs-big-data.
Isermann, R. (2005). Model-based fault-detection and diagnosis – Status and applications. *Annual Reviews in Control, 29*, 71–85.
Isermann, R., & Ballé, P. (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice, 5*, 709–719.
Jiang, B., Huang, D., Zhu, X., Yang, F., & Braatz, R. D. (2015a). Canonical variate analysis-based contributions for fault identification. *Journal of Process Control, 26*, 17–25.
Jiang, B., Zhu, X., Huang, D., & Braatz, R. D. (2015b). Canonical variate analysis-based monitoring of process correlation structure using casual feature representation. *Journal of Process Control, 52*, 109–116.
Jiang, B., Zhu, X., Huang, D., Paulson, J. A., & Braatz, R. D. (2015c). A combined canonical variate analysis and fisher discriminant analysis (CVA–FDA) approach for fault diagnosis. *Computers & Chemical Engineering, 77*, 1–9.
Jiang, J., & Yu, X. (2012). Fault-tolerant control systems: A comparative study between active and passive approaches. *Annual Reviews in Control, 36*(1), 60–72.
Jiang, M., Zhu, Z., Jimenez, E., Papageorgiou, C. D., Waetzig, J., Hardy, A., Langston, M., & Braatz, R. D. (2014). Continuous-flow tubular crystallization in slugs spontaneously induced by hydrodynamics. *Crystal Growth & Design, 14*, 851–860.
Jin, J., & Shi, J. (2001). Automatic feature extraction of waveform signals for in-process diagnositic performance improvement. *Journal of Intelligent Manufacturing, 12*, 257–268.
Jolliffe, I. T. (2002). *Principal component analysis*. New York: John Wiley and Sons, Ltd.
Juricek, B. C., Seborg, D. E., & Larimore, W. E. (2001). Identification of the Tennessee Eastman challenge process with subspace methods. *Control Engineering Practice, 9*(12), 1337–1351.
Juricek, B. C., Seborg, D. E., & Larimore, W. E. (2004). Fault detection using canonical variate analysis. *Industrial & Engineering Chemistry Research, 43*, 458–474.
Kirdar, A. O., Conner, J. S., Baclaski, J., & Rathore, A. S. (2007). Application of multivariate analysis toward biotech processes: Case study of a cell-culture unit operation. *Biotechnology Progress, 23*(1), 61–67.
Korbicz, J., Kośielny, J. M., Kowalczuk, Z., & Cholewa, W. (Eds.). (2004). *Fault diagnosis: Models, artificial intelligence, applications*. Berlin: Springer-Verlag.
Kościelny, J. M., & Łabęda-Grudziak, Z. M. (2013). Double fault distinguishability in linear systems. *International Journal of Applied Mathematics and Computer Science, 23*, 395–406.
Kourti, T. (2006). The process analytical technology initiative and multivariate process analysis, monitoring and control. *Analytical and Bioanalytical Chemistry, 384*, 1043–1048.
Larimore, W. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In *Proceedings of the ieee conference on decision and control* (pp. 596–604).
Larimore, W. E. (1983). System identification, reduced-order filtering and modeling via canonical variate analysis. In *Proceedings of the american control conference* (pp. 445–451).
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
LeCun, Y., Cortes, C., & Burges, C.J.C. (1998). The MNIST database of handwritten digits. Available for download at http://yann.lecun.com/exdb/mnist/, retrieved on May 22, 2015.
Low, C. B., Wang, D., Arogeti, S., & Luo, M. (2010). Quantitative hybrid bond graph-based fault detection and isolation. *IEEE Transactions on Automation Science and Engineering, 7*, 558–569.
MacGregor, J., & Cinar, A. (2012). Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods. *Computers & Chemical Engineering, 47*, 111–120.
Mahadevan, S., & Shah, S. L. (2009). Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control, 19*(10), 1627–1639.
Maki, M., Jiang, J., & Hagino, K. (2001). A stability guaranteed active fault-tolerant control system against actuator failures. In *Proceedings of the ieee conference on decision and control* (pp. 1893–1898).
Marseglia, G. R., Scott, J. K., Magni, L., Braatz, R. D., & Raimondo, D. M. (2014). A hybrid stochastic-deterministic approach for active fault diagnosis using scenario optimization. In *Proceedings of the IFAC world congress* (pp. 1102–1107).
Mehra, R., & Peschon, J. (1971). An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica, 7*(5), 637–640.
Mesbah, A., Streif, S., Findeisen, R., & Braatz, R. D. (2014). Active fault diagnosis for nonlinear systems with probabilistic uncertainties. In *Proceedings of the IFAC world congress* (pp. 7079–7084).
Mrugalski, M. (2013). An unscented kalman filter in designing dynamic GMDH neural networks for robust fault detection. *International Journal of Applied Mathematics and Computer Science, 23*, 157–169.
Mrugalski, M. (2014). *Advanced neural network-based computational schemes for robust fault diagnosis*. Cham: Springer.
Murata, T. (1989). Petri nets: Properties, analysis and applications. In *Proceedings of the IEEE: 77* (pp. 541–580).
Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge: MIT Press.
Nikoukhah, R. (1998). Guaranteed active failure detection and isolation for linear dynamical systems. *Automatica, 34*(11), 1345–1358.
Nomikos, P., & MacGregor, J. F. (2014). Multivariate SPC charts for batch monitoring processes. *Technometrics, 37*(1), 41–59.
O'Malley, O. et al. (2016). Apache Hadoop. https://hadoop.apache.org/.
Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*, 1065–1076.
Patton, R. J. (1997). Fault-tolerant control systems: The 1997 situation. *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, 3*, 1033–1054.
Paul, D. B., & Baker, J. M. (1992). The design for the Wall Street Journal-based CSR corpus. In *Proceedings of the workshop on speech and natural language* (pp. 357–362).
Paynter, H. (1961). *Analysis and design of engineering systems*. Cambridge, MA: MIT Press.
Petri, C. A. (1962). *Kommunikation mit automaten*. Bonn: Institut für Instrumentelle Mathematik Ph.D. thesis..
Platt, J. C. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Technical Report, MSR-TR-98-14*. Cambridge, MA: Microsoft Research.
Qin, S. J. (2003). Statistical process monitoring: Basics and beyond. *Journal of Chemometrics, 17*, 480–502.
Qin, S. J. (2014). Process data analytics in the era of Big Data. *AIChE Journal, 60*, 3092–3100.
Raimondo, D. M., Braatz, R. D., & Scott, J. K. (2013a). Active fault diagnosis using moving horizon input. In *Proceedings of the european control conference* (pp. 3131–3136).
Raimondo, D. M., Marseglia, G. R., Braatz, R. D., & Scott, J. K. (2013b). Fault-tolerant model predictive control with active fault isolation. In *Proceedings of conference on control and fault-tolerant systems* (pp. 444–449).
Reis, M. S., Braatz, R. D., & Chiang, L. H. (2016). Big data challenges and future research directions. *Chemical Engineering Progress, 112*(3), 46–50.
Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000a). *Data-driven techniques for fault detection and diagnosis in chemical processes*. New York: Springer.
Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000b). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 51*, 81–93.
Schölkopf, B., Platt, J. C., S.-T. J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support a high-dimensional distribution. *Neural Computation, 13*, 1443–1471.
Scott, J. K., Findeisen, R., Braatz, R. D., & Raimondo, D. M. (2013a). Design of active inputs for set-based fault diagnosis. In *Proceedings of american control conference* (pp. 3561–3566).
Scott, J. K., Findeisen, R., Braatz, R. D., & Raimondo, D. M. (2014). Input design for guaranteed fault diagnosis using zonotopes. *Automatica, 50*, 1580–1589.
Scott, J. K., Marseglia, G. R., Magni, L., Braatz, R. D., & Raimondo, D. M. (2013b). A hybrid stochastic-deterministic input design method for active fault diagnosis. In *Proceedings of the ieee conference on decision and control* (pp. 5656–5661).
Severson, K., Chaiwatanodom, P., & Braatz, R. D. (2015a). Perspectives on process monitoring of industrial systems. In *Ifac symposium on fault detection, supervision and safety for technical processes* (pp. 931–939).
Severson, K., VanAntwerp, J. G., Natarajan, V., Antoniou, C., Thommes, J., & Braatz, R. D. (2015b). Elastic net with Monte Carlo sampling for data-based modeling in biopharmaceutical manufacturing facilities. *Computers and Chemical Engineering, 80*, 30–36.
Simani, S. (2013). Residual generator fuzzy identification for automotive diesel engine fault diagnosis. *International Journal of Applied Mathematics and Computer Science, 23*, 419–438.
Simani, S., Fantuzzi, C. F., & Patton, R. J. (2003). *Model-based fault diagnosis in dynamic systems using identification techniques*. New York: Springer.
Simoglou, A., Martin, E. B., & Morris, A. J. (2002). Statistical performance monitoring of dynamic multivariate processes using state space modelling. *Computers & Chemical Engineering, 26*(6), 909–920.

Srivinvasan, V. S., & Jafari, M. A. (1993). Fault detection/monitoring using time Petri nets. *IEEE Transactions on Systems, Man, and Cybernetics, 23*, 1155–1162.

Tomohiro, T., & Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics, 15*, 116–132.

Undey, C., Ertunç, S., & Çinar, A. (2003). Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. *Industrial & Engineering Chemical Research, 42*, 4645–4658.

Van Overshcee, P., & De Moor, B. (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica, 30*, 75–93.

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003a). A review of process fault detection and diagnosis. Part II: Qualitative models and search strategies. *Computers & Chemical Engineering, 27*, 313–326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003b). A review of process fault detection and diagnosis. Part III: Process history based methods. *Computers & Chemical Engineering, 27*, 327–346.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. (2003c). A review of process fault detection and diagnosis. Part I: Quantitative model-based methods. *Computers & Chemical Engineering, 27*, 293–311.

Verhaegen, M. (1993a). Identification of the deterministic part of MIMO state space models given in innovation form from input-output data. *Automatica, 30*, 61–74.

Verhaegen, M. (1993b). Subspace model identification. Part III: Analysis of the ordinary output-error state space model identification algorithm. *International Journal of Control, 58*, 555–586.

Verhaegen, M. (1994). Application of a subspace model identification technique to identify LTI systems operating in closed loop. *International Journal of Control, 29*, 1027–1040.

Verhaegen, M., & Dewilde, P. (1992a). Subspace model identification. Part I: The output-error state space model identification class of algorithms. *International Journal of Control, 56*, 1187–1210.

Verhaegen, M., & Dewilde, P. (1992b). Subspace model identification. Part II: Analysis of the elementary output-error state space model identification algorithm. *International Journal of Control, 56*, 1211–1241.

Verron, S., Tiplica, T., & Kobi, A. (2008). Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control, 18*, 479–490.

Viswanadham, N. (1988). Fault detection and diagnosis of automated manufacturing systems. In *Proceedings of the ieee conference on decision and control* (pp. 2301–2306).

Widodo, A., & Yang, B.-S. (2007). Support vector machines in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing, 21*, 2560–2574.

Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica, 12*, 601–611.

Wise, B. M., Ricker, N. L., Veltkamp, D. F., & Kowalski, B. R. (1990). A theoretical basis for the use of principal component models for monitoring multivariate processes. *Process Control and Quality, 1*, 41–51.

Witczak, M. (2014). *Fault diagnosis and fault-Tolerant control strategies for non-linear systems: Analytical and soft computing approaches*. Cham: Springer.

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307–357). New York: Academic Press.

Wold, S., Ruhe, A., Wold, H., & Dunn III, W. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing, 5*(3), 735–743.

Xu, L., & Schuurmans, D. (2005). Unsupervised and semi-supervised multi-class support vector machines. *Association for the Advancement of Artificial Intelligence, 5*, 904–910.

Yin, S., Ding, S. X., Haghani, A., Hao, H., & Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control, 22*, 1567–1581.

Yue, H. H., & Qin, S. J. (2001). Reconstruction-based fault identification using a combined index. *Industrial & Engineering Chemistry Research, 40*, 4403–4414.

Zhang, Y., & Jiang, J. (2008). Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control, 32*, 229–252.

Zhao, Y., Ball, R., Mosesian, J., de Palma, J., & Lehman, B. (2015). Graph-based semi–supervised learning for fault detection and classification in solar photovoltaic arrays. *IEEE Transactions on Power Electronics, 30*, 2848–2858.

Zhong, M., Ding, S. X., Lam, J., & Wang, H. (2003). An LMI approach to design robust fault detection filter for uncertain LTI systems. *Automatica, 39*, 543–550.

Zhou, D., Li, G., & Qin, S. J. (2010). Total projection to latent structures for process monitoring. *AIChE Journal, 56*(1), 168–178.

Zhu, X., & Braatz, R. D. (2014). 2D contribution map for fault detection. *IEEE Control Systems, 33*(4), 72–77.