IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 June 2022, accepted 18 July 2022, date of publication 21 July 2022, date of current version 5 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3192972

RESEARCH ARTICLE

Improving Fault Detection in Industrial Processes by Event-Driven Data Acquisition

DANIEL GUTIERREZ-ROJAS[1], (Student Member, IEEE), PEKKA RÄSÄNEN[1], ANA BRANDÃO BELISÁRIO[2], MERIM DZAFERAGIC[3], GUSTAVO MATHEUS DE ALMEIDA[2], AND PEDRO H. J. NARDELLI[1], (Senior Member, IEEE)

[1]Department of Electrical Engineering, School of Energy Systems, LUT University, 53850 Lappeenranta, Finland
[2]Department of Chemical Engineering, School of Engineering, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil
[3]CONNECT, Trinity College Dublin, Dublin 2, D02 PN40 Ireland

Corresponding author: Daniel Gutierrez-Rojas (daniel.gutierrez.rojas@lut.fi)

This work was supported in part by the Academy of Finland through the EnergyNet Research Fellowship under Grant 321265 and Grant 328869, in part by the Academy of Finland through the FIREMAN Consortium under Grant 326270, and in part by the CHIST-ERA under Grant CHIST-ERA-17-BDSI-003.

ABSTRACT Data acquisition in process industries usually takes place at each sampling. The disadvantage is that a considerable amount of data without new information about the state of the process is continuously transmitted and processed. This negatively affects the communication system and computational power, which is more critical nowadays given the number of variables measured, even in seconds. One solution concerns the event-driven paradigm, in which only relevant data according to a pre-defined criterion is forwarded for further processing. This work investigated the event-based threshold and delta methods in the context of fault detection. The data transmission rate was also analyzed. The well-know Tennessee Eastman problem (TEP) was used as a case study. The fault detection system was based on PCA (principal component analysis), which is widely used for this purpose in this benchmark. The results were compared with the commonly used time-based approach, for a fixed false alarm rate. The threshold rule provided similar results, but with much less data. For the delta rule, significant MDR (missed detection rate) gains of up to 74% were obtained for five of the six hard-to-detect faults, and of up to 69%, for two of the three very hard-to-detect faults. MDR values very close to zero were reached for two of the three intermediate detection faults and two of the hard-to-detect faults. The detection time was also evaluated. In this regard, considerably lower values were obtained for all intermediate detection faults, three of the hard-to-detect faults and all very hard-to-detect faults. In short, the delta method was able to improve fault detection performance, especially for hard-to-detect faults, with a considerably lower data transmission rate, around 20% on average. Event-driven data acquisition can be very attractive for process industries.

INDEX TERMS Data acquisition, event-driven, signal reconstruction, fault detection, PCA, Tennessee benchmark.

I. INTRODUCTION

Data acquisition based on a fixed time interval is commonly used in continuous process industries worldwide [1]. The disadvantage of this procedure is that a fair amount of data without new information about the state of the process is continuously transmitted and processed. This data can overwhelm system communication and limit computing power, resulting in information loss. Currently, high frequency data sampling nowadays through a multitude of industrial sensors, even within seconds, exacerbates this problem. In this context, the challenge is to select only data that carry new information for transmission and further processing.

Event-based data acquisition is an alternative to periodic procedure, since events are asynchronous in nature [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai.

An event refers to a significant change in a signal according to a predefined threshold [3]. In process industries, a considerable change in a measured variable is often associated with new information about the state of the process, which would actually require processing. Variables measurements can be used directly, or their variations (derived). This strategy favors communication and data processing, which in turn can result in energy savings and running more tasks in parallel [4]. The event-based paradigm emerged in the context of system control [5]–[10]. More information can be found in [3], [4].

Several potential advantages using event-driven data acquisition strategies have been reported in the literature. They pertain to data storage, channel bandwidth, signal reconstruction, communication system, and sensor power consumption, to name a few. Below are examples of works from different areas of knowledge. [11] used an event-driven mechanism to update information about unreliable links in network switched systems. The authors reported a gain in network security and bandwidth. [12] described an event-based analysis of a solar distribution feeder integrated into distribution substations. The authors highlighted an improvement in event detection, analysis of the impact of solar production, and understanding of the dynamics of the solar feeder control system. [13] employed smart meters through event-driven sampling for data acquisition and feature extraction to disregard redundant information. The authors reported that this data filtering facilitated the identification of energy consumption patterns of devices by vector support machines. [14] designed an event-based model predictive control strategy to more efficiently reconfigure production logic in real-time. An industrial sewing machine production plant was used as a case study, focusing on smart factories. [15] investigated event-based sampling techniques for more efficient node communication and power consumption in wireless sensor networks (WSN). The application involved air quality monitoring. According to the authors, smarter data acquisition would contribute to increasing the life cycle of sensor nodes, since data transmission is currently the main source of energy consumption. [16] compared time- and event-based sampling strategies in electricity grids. The authors reported that the latter was able to perform the signal reconstruction satisfactorily. [17] used Monte Carlo simulation to perform component reliability analysis in complex and dynamic systems using dynamic fault tree. As this is time consuming, an event-driven simulation approach was applied disregarding the gate simulations without significant changes in the output. The authors reported that this data filtering increased computational efficiency. [18] presented an event-driven architecture in the context of Industry 4.0 for manufacturing systems. The focus was on more efficient integration of data from devices and services at all levels for more flexible and timely decisions. [19] developed an information system to detect significant events in collaborative processes in modern business environments. The authors aimed to provide more agile responses to improve interaction between people, devices and organizations. [20] applied an event-driven strategy, in combination with MPCA (multiway principal component analysis), to monitor a SBR (sequencing batch reactor) wastewater process, which outperformed the time-based approach. [21] investigated techniques for developing event-oriented control systems for batch processes. The authors reported gains in engineering and economics aspects. [22] proposed a reduction in the number of events and decision variables using a context-based definition of what an event would be. This more rational concept was used in a MILP (mixed integer linear programming) problem to model batch scheduling with respect to multi-product/stage machine processes. [23] considered a series of sensors to monitor the thermal control system of a space station. To reduce the amount of data to be processed, the authors adopted an event-driven simulation, computing only the events that could actually contribute to reaching an alarm level. In summary, all works reported that the event-based paradigm can be an efficient strategy to reduce the amount of data before further processing. This is also the case for fault detection ([15], [17], [20]), which is the focus of this work.

Process monitoring, and more specifically fault detection, is an essential activity in process industries. Due to the inherent complexity of the processes, being multivariable, nonlinear and only partially known, this task is still challenging from a practical point of view [24]. Many process variables are currently measured continuously, even within seconds, which is very beneficial on the one hand [25]. However, problems with the resulting massive amount of data also arise from the other side. Packet collision and high latency are examples involving data transmission in communication systems [26]. As far as information is concerned, more data does not necessarily mean more quality. Process noisy, missing values, measurement errors and imbalanced classes (operation modes), to name a few, occur frequently [27]. In this context, process industries can benefit greatly from the event-driven paradigm, reducing the amount of data to be transmitted and processed. Unlike the fixed-time data acquisition procedure, the input data for the fault detection system would only be updated in the case of new process information, according to a predefined criterion.

This work investigates the event-driven data acquisition strategy for fault detection in continuous industrial systems. The threshold [28] and delta [29] event-based methods were evaluated under different data transmission rates. The well-known Tennessee Eastman Process (TEP) benchmark served as the case study [30]. Several techniques and strategies have been used for fault detection purposes in this benchmark. To name a few, DPCA (dynamic principal component analysis) with decorrelated residuals [31], a combination of PCA (principal component analysis) and k-NN (k-nearest neighbors) [32], a combination of PCA with fuzzy logic [33], a sparse auto-encoder [34], RNN (recurrent neural network) and CNN (convolutional neural network) deep learning models [35], and image processing using MLP (multilayer perceptron) and RBF (radial basis function) neural networks

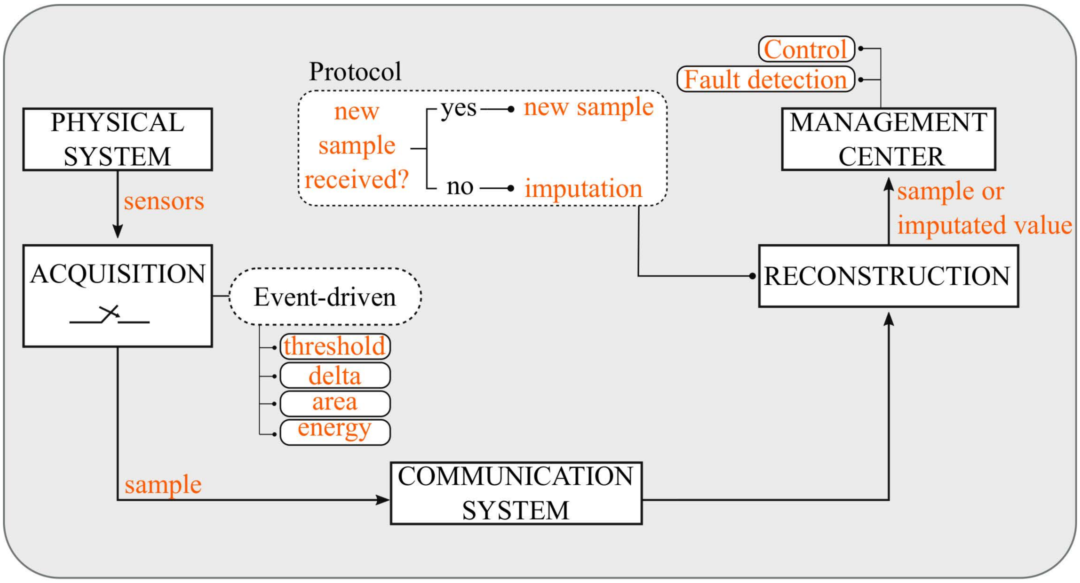


FIGURE 1. Proposed approach for event-driven data acquisition.

based classifiers [36]. A common point among all these works is the use of the fixed-time approach. The following works are reviews on fault detection and diagnosis using data-driven modeling in various fields of engineering: photovoltaic system [37], power transformer [38], HVAC (heating, ventilation and air conditioning) system [39], building energy system [40], marine current turbine [41], gear system [42], thermal system [43], industrial system [44] and high-speed trains in intelligent transportation [45]. The present work made use of PCA, as it is the most applied technique for fault detection in the TEP benchmark. [31], [32], [46]–[48] are examples of recent works, which also employed periodic data acquisition. In this sense, the results obtained with the event-based methods were compared with this commonly used fixed-time approach. The classic work in [49] was used for this purpose. To have the same basis of comparison, the percentage of variance explained by the PCA model was approximately 55% and the false alarm rate was set at 1%. In addition, the missed detection rate (MDR) and detection time (delay) were used as performance metrics. This work shows, as its main contribution, the possibility of obtaining a better performance in fault detection with considerably low data transmission rates. Event-driven applications in process industries are often related to process control and instrumentation. To the best of the authors' knowledge, there is no systematic study of its use for fault detection in continuous industrial systems.

This paper is organized as follows. Section II shows the proposed event-driven framework for data acquisition. The methodology is depicted in section III. The event-based methods used in this work are described in section IV. Section V refers to the TEP (Tennessee Eastman Process) benchmark used as a case study. Section VI presents the results and discussion of the fault detection performance, including a comparison between the event- and the time-based approaches. An analysis of the data transmission rate is given in section VII. Final considerations are given in section VIII.

II. EVENT-DRIVEN DATA ACQUISITION FRAMEWORK

To properly design an event-driven data acquisition procedure, it is necessary to characterize the measured signals. This can be done explicitly by knowing the math function or directly from the data. The latter is generally used due to the complexity of industrial processes. The main question in this case concerns how to "filter" the signals to consider only the sampling points associated with new information about the state of the process. Predefined thresholds based on individual values, or signal variations or accumulations over time, are used to identify these relevant events [3], [4]. As part of the data acquisition phase, event-driven strategies have the potential to significantly reduce the amount of data to be transmitted and processed. In this sense, applications of Cyber-Physical Systems (CPS) involving large amounts of data can be greatly benefited [50]. Figure 1 depicts the proposed event-based data acquisition system with a focus on continuous industrial processes. Each element is described below.

A. PHYSICAL SYSTEM

This initial element refers to the real-world system of interest. It is very important to have information about the measured variables, the location of the sensors and the dynamics of the process.

B. DATA ACQUISITION

The input to this step is the sampled values collected continuously by a series of sensors in the physical system using a fixed time interval. The selection of samples to be transmitted containing new information about the state of the process is based on a predefined event-based strategy. This data filtering can be applied locally in the case of smart sensors or at a remote center. Examples of event-based methods are given by the threshold, delta, area, and energy approaches (described in the IV section). The first two were used in this work with different data compression rates, which determines the amount of data to be transmitted for further processing.

C. COMMUNICATION SYSTEM

This system is responsible for transmitting the previously selected values. Wired or wireless systems can be used. This data filtering contributes to more efficient management of memory, latency and packet collision, to name a few, which is most critical in the case of networks with limited bandwidth. The communication system infrastructure and its challenges are outside the scope of this work.

D. SIGNAL RECONSTRUCTION

The signal at this point may contain missing values as the values of some variables may not have been transmitted. This occurs when they do not exceed the respective event-based method limits. For signal reconstruction, a data imputation protocol should then be used [51]. The present work adopted the last value sent by the communication system. This procedure is also beneficial in case of signal loss.

E. MANAGEMENT CENTER

This unit receives and processes the previously reconstructed signal, according to the purpose of the application. This can refer to process modeling, simulation, control, optimization or monitoring. Given the objective of this work, a fault detection system is part of this unit in the present work. Central units like this are even more important nowadays, given the concept of a cyber-physical system in the context of Industry 4.0 [50]. This system integration can be useful for offline and real-time applications. Event-based strategies also favor the computational power needed in this case, given the smaller amount of data to be processed.

III. METHODOLOGY

Figure 2 depicts the methodology adopted in this work. There are three main steps: model identification, control limits definition, and fault detection itself. Each step is described below.

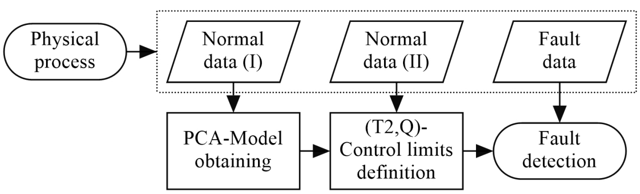


FIGURE 2. Methodology steps.

A. MODEL IDENTIFICATION

The fault detection system was based on PCA (Principal Component Analysis). This technique have been the most used for fault detection in industrial applications [52]–[55]. The case study in this work used the Tennessee benchmark problem [30], described in section V. In this case, a first data set (normal data I), characteristic of the normal operating condition, was used to obtain the PCA model. The approximate number of twelve principal components, which explain about 55% of the total variance in the original data, was adopted for comparison purposes [49].

B. CONTROL LIMITS DEFINITION

The multivariate control charts for the T^2 and Q statistics were used for fault detection [44], [56], [57]. Their respective control limits were calculated from a second data set (normal data II), also characteristic of the normal operating condition. For this, the false alarm rate was set at 1% [49], being determined by the 99th percentile. Both previous steps considered fixed-time sampling. *Fault detection:* The event-based data acquisition framework (Figure 1) was used in this step. At each sampling, the measured value of the variable is transmitted only if it is considered an event. Otherwise, the last value sent to the variable is used during signal reconstruction. This procedure is repeated for each variable. The reconstructed signal is then fed into the previous PCA model, and the corresponding values of T^2 and Q are calculated and plotted on the respective control charts. All twenty-one faults available in the benchmark were investigated (using the respective fault data set). The results were compared with the classical time-based data acquisition approach [49]. The missed detection rate and detection time performance metrics were used for this purpose. The event-based strategies used in this work are described in section IV.

In short, PCA is a dimensionality reduction technique belonging to multivariate statistics [58]. Its principle resides in an orthogonal rotation of the coordinate system given by the original variables. The rotated axes, called principal components, are defined according to a criterion of maximum variance. The first component explains as much of the total variance of the original data as possible, the second, which is orthogonal to the first, as much of the remaining unexplained variance, and so on. Given a required amount of explained variance, the first k components are used as the final PCA model. With regard to fault detection, this model is obtained with normal data, being, therefore, characteristic of the normal operating condition. When fed with fault data, it is expected to recognize such a condition, which is normally accomplished through control charts for the T^2 and Q statistics.

IV. EVENT-DRIVEN METHODS

Event-based methods are generally provided by the following strategies: send-on-delta (SoD) [29], send-on-prediction (SoP) [59], [60], send-on-area (SoA) [61], send-on-energy (SoE) [62], and a simple threshold definition. In SoD, the current value is transmitted when a minimum difference from the last value sent is reached. SoP is an extension of the previous strategy, in which a predicted value from the last update

TABLE 1. Process disturbances [30], [49].

Fault	Process variable	Fault type
1	A/C feed ratio, B composition constant (stream 4)	Step
2	B composition, A/C ratio constant (stream 4)	Step
3	D feed temperature (stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss (stream 1)	Step
7	C header pressure loss - Reduced availability (stream 4)	Step
8	A, B, C feed composition (stream 4)	Random variation
9	D feed temperature (stream 2)	Random variation
10	C feed temperature (stream 4)	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	The valve for Stream 4 was fixed at the steady state position	Constant position

is used. SoA and SoE are two other common extensions. The triggering criterion in the first is given by the integral of the absolute difference and, in the second, by the energy of the difference. Another method is provided by the use of a predefined threshold [28]. The event trigger in this case occurs when the current value crosses a cut-off point. This work investigated the threshold and delta strategies. The use of the others is straightforward.

A. THRESHOLD-BASED METHOD

The threshold method employs a cut-off point as a decision rule for data acquisition. The measured value is only transmitted if it exceeds this reference, whose definition uses the mean (S_{avg}), minimum (S_{min}) and maximum (S_{max}) statistics of the signal (S) under the normal operating condition. The resulting lower (T_l) and upper (T_u) thresholds are shown in Equation 1. To investigate the effect of the magnitude of the differences ($S_{max} - S_{avg}$; $S_{avg} - S_{min}$) on data transmission rate and fault detection performance, a parameter p ($0 < p < 1$) was varied as follows: $[5, 10 : 10 : 90, 95]\%$. The higher, the lower the number of values transmitted. This procedure is repeated for each variable separately.

$$T_u = S_{avg} + (S_{max} - S_{avg}) \times p \quad (1a)$$
$$T_l = S_{avg} - (S_{avg} - S_{min}) \times p \quad (1b)$$

The reconstruction of the signal (S'_t) is performed according to Equation 2. The value of the variable (S_t) is transmitted if it exceeds the lower or upper thresholds; otherwise, the last value sent (S'_{t-1}) is used. The threshold definition and signal reconstruction steps follow the event-based data acquisition framework shown in Figure 1.

$$S'_t = \begin{cases} S_t, & \text{if } (S_t < T_l) \text{ or } (S_t > T_u) \\ S'_{t-1}, & \text{if } T_l \leq S_t \leq T_u \end{cases} \quad (2)$$

The threshold method (with $p = 50\%$) is illustrated for feed A (stream 1), which is one of the variables of the TEP benchmark. Figures 3a and 3b show the original and filtered signals, respectively, given fault 5 (Table 1).

B. DELTA-BASED METHOD

The send-on-delta (SoD) method is based on the difference between the current value and the one previously sent by the data acquisition system. Equation 3 shows the definition of the reference value in this case, which is given by the maximum absolute difference (Δ max) between consecutive values (S_t, S_{t-1}) considering the signal under the normal operating condition. The parameter p was used as before. This procedure is applied to each variable separately.

$$\Delta \max = \max(| S_t - S_{t-1} |) \times p \quad (3)$$

The next step concerns the reconstruction of the signal (S'_t), according to Equation 4. The value of the variable (S_t) is transmitted only if the corresponding difference is greater than Δ max; otherwise, the last value sent (S'_{t-1}) is used. The delta rule is detailed in Figure 4. This procedure is applied to each variable separately. It also follows the event-based data acquisition framework shown in Figure 1.

$$S'_t = \begin{cases} S_t, & \text{if } | S_t - S_{t-1} | > (\Delta \max) \\ S_{t-1}, & \text{if } | S_t - S_{t-1} | \leq (\Delta \max) \end{cases} \quad (4)$$

Figure 3c illustrates the use of the delta rule for A feed (stream 1), with $p = 50\%$, given fault 5 (Table 1). A low resolution signal can be verified in comparison to the threshold rule (Figure 3b).

V. CASE STUDY: TENNESSEE EASTMAN PROCESS (TEP) BENCHMARK

The Tennessee benchmark problem [30] has often been used for the development of fault detection systems [35], [54], [55], [63], [64]. Based on a real industrial process, it involves

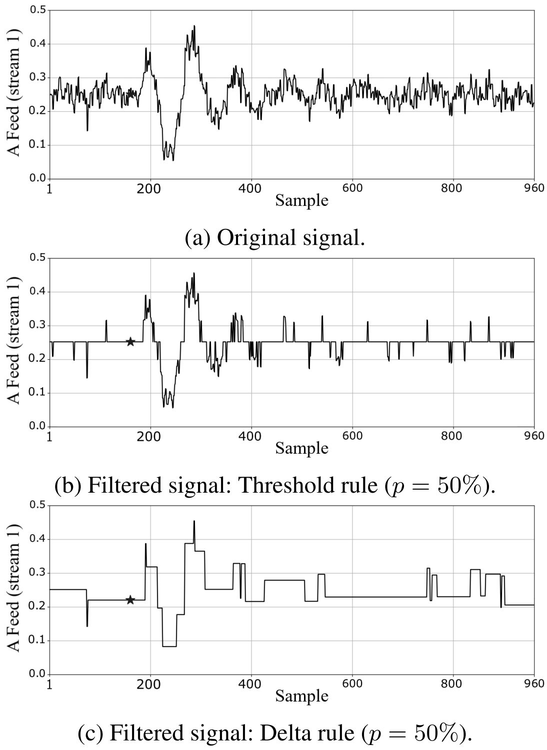


FIGURE 3. **Event-based data acquisition for A feed (stream 1), which is one of the variables of the case study (section V), given fault 5 (Table 1) (⋆: Fault start up at $t = 160$).**

a reactor, a condenser, a vapor-liquid separator, a recycle compressor and a stripper column (Figure 5). The objective is to obtain the liquid (l) products, G and H, from the gaseous (g) reactants, A, C, D and E. This is achieved by a set of four irreversible and exothermic chemical reactions, as shown below. Components B (not shown) and F are an inert and a by-product, respectively.

$$
\begin{aligned}
A_{(g)} + C_{(g)} + D_{(g)} &\rightarrow G_{(l)}(product1)\\
A_{(g)} + C_{(g)} + E_{(g)} &\rightarrow H_{(l)}(product2)\\
A_{(g)} + E_{(g)} &\rightarrow F_{(l)}(byproduct)\\
3\ D_{(g)} &\rightarrow 2\ F_{(l)}(byproduct)
\end{aligned}
$$

There are fifty-two measurements, where twenty-two are process variables, eleven are manipulated variables, and nineteen are laboratory parameters. The variables and parameters are collected every three and six minutes, respectively. Twenty-one faults are available (Table 1) [1], [49], occurring one at a time. The normal and fault data sets are composed of 500 and 960 observations, which correspond to twenty-five and forty-eight hours of simulation, respectively. Each fault occurs at $t = 160$, that is, after eight hours under normal operation. Gaussian noise is introduced into all measurements.

VI. RESULTS AND DISCUSSION

The results obtained with the event-driven methods, namely, threshold and delta rules, were compared with the time-based approach. The PCA model (step 1 in Figure 2), characteristic of the normal operating condition, as well as the respective

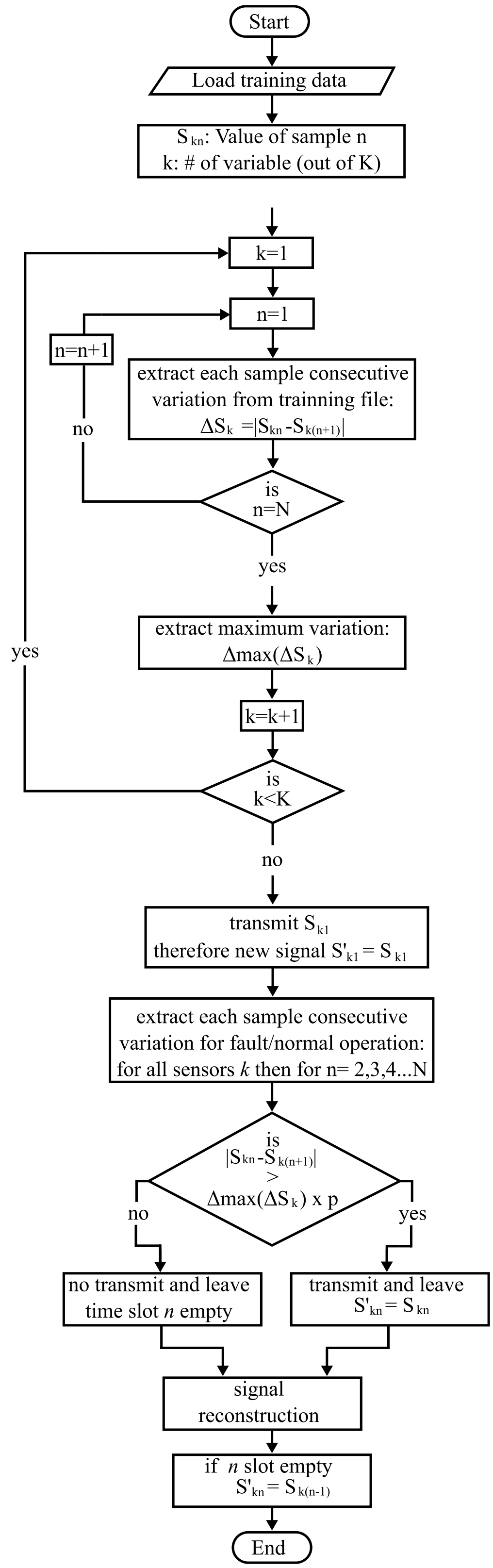


FIGURE 4. **Flowchart of the delta event-driven method for data acquisition.**

upper control limits for the statistics T^2 and Q (step 2), were common to both cases. This model is given by twelve

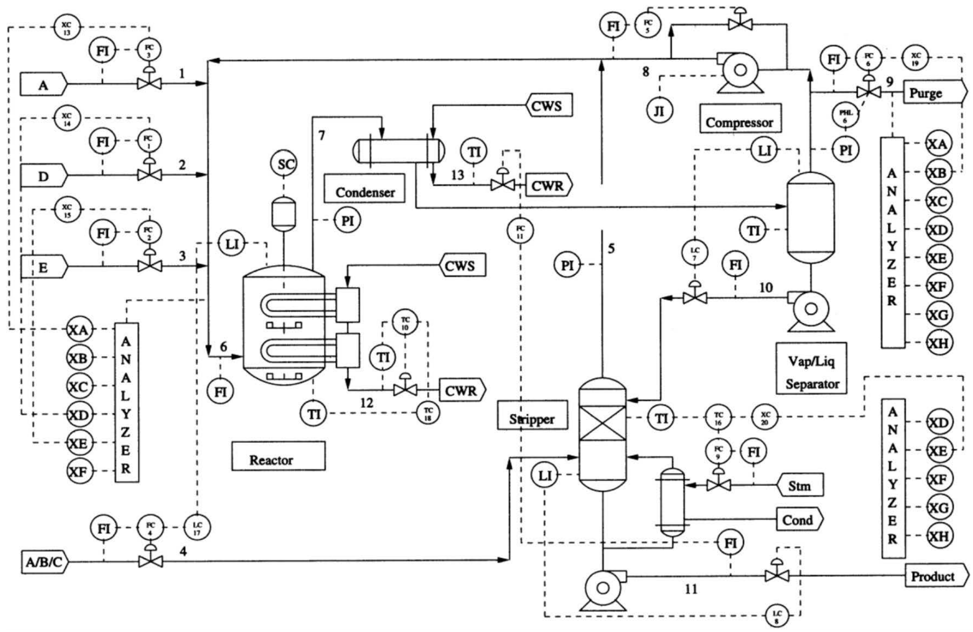


FIGURE 5. Process flowchart of the Tennessee benchmark problem [1].

TABLE 2. MDR (missed detection rate) for the T^2 statistic.

Fault	Time-based approach	Event-based delta approach (p in %)											Absolute gain
		5	10	20	30	40	50	60	70	80	90	95	
1	0.008	**0.008**	**0.008**	**0.008**	**0.008**	**0.008**	**0.008**	0.009	**0.008**	**0.008**	0.010	0.010	0%
2	0.020	**0.020**	**0.020**	**0.020**	**0.020**	0.023	0.023	**0.020**	0.023	0.023	0.023	0.023	0%
3	0.998	0.999	0.996	0.996	1.000	**0.993**	1.000	1.000	0.996	1.000	1.000	1.000	1%
4	0.956	0.961	0.959	0.960	0.973	0.969	0.939	0.934	**0.905**	0.999	0.999	0.999	**5%**
5	0.775	0.774	0.771	0.770	0.766	0.764	0.766	0.773	0.729	0.683	**0.604**	0.701	**17%**
6	0.011	0.011	0.011	0.011	0.011	0.011	0.011	**0.008**	0.009	0.009	0.009	0.009	0%
7	0.085	0.079	0.076	0.079	0.065	0.045	0.051	0.161	**0.000**	**0.000**	**0.000**	**0.000**	**9%**
8	0.034	0.033	**0.031**	0.034	**0.031**	0.033	**0.031**	**0.031**	**0.031**	0.043	0.054	0.045	0%
9	0.994	0.998	0.996	0.991	0.993	0.998	0.984	0.984	**0.979**	1.000	1.000	1.000	2%
10	0.666	0.653	0.653	**0.646**	0.655	0.671	0.678	0.679	0.694	0.744	0.806	0.826	2%
11	0.794	0.789	0.786	0.796	0.793	0.794	0.761	**0.688**	0.876	0.886	0.899	0.926	**11%**
12	0.029	0.021	0.021	0.023	**0.019**	0.024	0.026	0.021	0.035	0.034	0.059	0.056	1%
13	0.060	**0.060**	**0.060**	0.061	0.061	0.061	0.061	0.063	0.061	0.063	0.076	0.078	0%
14	0.158	0.120	0.125	0.111	0.114	0.096	**0.094**	0.106	0.095	0.154	0.149	0.154	**6%**
15	0.988	0.979	0.978	0.984	0.983	0.993	**0.940**	0.974	0.984	0.996	1.000	1.000	**5%**
16	0.834	0.833	0.833	0.851	0.844	0.834	0.860	0.809	**0.733**	0.920	0.934	0.946	**10%**
17	0.259	0.254	0.251	0.258	0.260	**0.249**	0.261	0.255	0.273	0.338	0.261	0.324	1%
18	0.113	0.113	0.113	0.111	**0.110**	0.113	0.116	0.118	0.113	0.113	0.121	0.121	0%
19	0.996	0.999	0.999	1.000	0.999	0.996	0.993	0.998	**0.885**	0.988	0.996	1.000	**11%**
20	0.701	0.704	0.710	0.709	0.718	0.725	0.684	0.731	**0.650**	0.766	0.754	0.805	**5%**
21	0.736	0.694	0.693	0.691	0.691	0.696	0.690	**0.651**	0.716	0.786	0.865	0.876	**8%**

Gray cells: False alarm rate > 5% (Table 4).

principal components, which explains about 56% of the total variance of the original data. This percentage was adopted for comparison purposes [49]. The difference between them concerns the input information for the fault detection system (step 3). All values sampled periodically are used in the time-based procedure, while they are filtered in the event-based approach (Figure 1).

The threshold method performed slightly better compared to the fixed-time approach over the entire range of the p parameter (results not shown due to lack of space). The benefit in this case concerns the use of a much smaller amount of data, which favors online applications, especially nowadays, given the era of big data.

Tables 2 and 3 show the missed detection rate (MDR) obtained with the delta rule for the T^2 and Q statistics, respectively. The best result for each fault is in bold. For example, the best MDR values for fault 5 were equal to 0.604 (for T^2 and $p = 0.90$) and 0.001 (for Q and $p = 0.80$). Absolute differences from the respective MDR values obtained with the periodic approach are also presented. The greater this difference, the better the result obtained by the delta rule. For example, the gains for fault 5 were equal to about 17% (=0.775 – 0.604) and 74% (=0.746 – 0.001) for the statistics T^2 and Q, respectively. For the Q statistic, it can be seen that the MDR value obtained with the event-driven delta rule is very close to zero (0.001), as desired, while the

TABLE 3. MDR (missed detection rate) for the Q statistic.

Fault	Time-based approach	Event-based delta approach (p in %)										Absolute gain
		5	10	20	30	40	50	60	70	80	90	95	
1	0.003	0.003	0.003	0.003	0.003	**0.001**	0.003	0.003	0.003	0.006	0.004	0.004	0%
2	0.014	0.014	0.014	0.015	0.011	0.011	**0.010**	0.018	0.000	0.023	0.023	0.023	0%
3	0.991	0.990	0.990	0.985	0.975	0.830	0.949	0.160	0.461	**0.876**	0.926	0.995	**11%**
4	0.038	0.035	**0.031**	0.040	0.038	0.031	0.033	0.046	0.006	0.063	0.000	0.470	1%
5	0.746	0.745	0.743	0.743	0.720	0.644	0.303	0.164	0.000	**0.001**	0.000	0.008	**74%**
6	0.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0%
7	0.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0%
8	0.024	0.025	0.025	0.021	0.023	**0.019**	0.025	0.000	0.025	0.025	0.026	0.026	1%
9	0.981	0.979	0.983	0.979	0.966	0.915	0.675	0.711	0.484	**0.501**	1.000	1.000	**48%**
10	0.659	0.650	0.653	0.648	0.578	0.465	0.255	**0.061**	0.065	0.226	0.454	0.464	**60%**
11	0.356	0.340	0.335	0.335	0.329	0.290	0.198	**0.099**	0.320	0.313	0.519	0.635	**26%**
12	0.025	0.025	0.025	0.029	0.031	0.018	**0.009**	0.005	0.014	0.016	0.016	0.016	2%
13	0.045	0.045	0.045	0.045	0.048	**0.041**	0.044	0.054	0.053	0.055	0.055	0.060	0%
14	0.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.001	0.001	0.001	0.001	0.001	0%
15	0.973	0.970	0.971	0.968	0.948	0.855	0.756	0.251	**0.279**	0.954	0.956	0.884	**69%**
16	0.755	0.744	0.740	0.728	0.661	0.574	**0.290**	0.210	0.174	0.536	0.633	0.546	**47%**
17	0.108	0.098	0.103	0.100	0.100	0.064	**0.034**	0.054	0.043	0.083	0.086	0.094	**7%**
18	0.101	0.103	0.100	0.101	0.096	0.076	0.031	0.048	**0.034**	0.063	0.114	0.114	**7%**
19	0.873	0.863	0.856	0.869	0.821	0.666	0.539	**0.526**	0.628	0.845	0.711	0.740	**35%**
20	0.550	0.546	0.538	0.529	0.509	0.448	0.286	0.169	**0.135**	0.138	0.376	0.354	**42%**
21	0.570	0.578	0.571	0.570	0.535	**0.456**	0.309	0.505	0.555	0.610	0.614	0.704	**11%**

Gray cells: False alarm rate > 5% (Table 5).

TABLE 4. FAR (false alarm rate) for the T^2 statistic.

Fault	Time-based approach	Event-based delta approach (p in %)									
		5	10	20	30	40	50	60	70	80	90	95
1	0	**0**	**0**	**0**	**0**	**0**	**0**	0	**0**	**0**	0	0
2	0	**0**	**0**	**0**	**0**	0.013	0	**0**	0	0	0	0
3	0.006	0.006	0.006	0	0	**0**	0	0	0	0	0	0
4	0.006	0.006	0.006	0.006	0.006	0.006	0	0	**0**	0	0	0
5	0.006	0.006	0.006	0.006	0.006	0.006	0	0	0	0	**0**	0
6	0	0	0	0	0	0	0	**0**	0	0	0	0
7	0	0	0	0	0	0	0	0	**0**	**0**	**0**	**0**
8	0	0	**0**	0	**0**	0	**0**	**0**	**0**	0	0	0
9	0.019	0.019	0.019	0.025	0.013	0	0	0	**0**	0	0	0
10	0	0	0	**0**	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	**0**	0	0	0	0
12	0	0	0	0	**0**	0	0.013	0	0	0	0	0
13	0	**0**	**0**	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	**0**	0	0	0	0	0
15	0	0	0	0	0	0	**0**	0	0	0	0	0
16	0.044	0.050	0.056	0.056	0.044	0.038	0	0	**0**	0	0	0
17	0	0	0	0	0	**0**	0	0	0	0	0	0
18	0	0	0	0	**0**	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	**0**	0	0	0
20	0	0	0	0	0	0	0	0	**0**	0	0	0
21	0	0	0	0	0	0	0	**0**	0	0	0	0

Gray cells: False alarm rate > 5%.

corresponding value for the time-fixed procedure is considerably high (0.746). A gray cell means the result was not considered valid due to a false alarm rate (FAR) above 5%. Tables 4 and 5 show the FAR values for the T^2 and Q statistics, respectively. For example, for fault 5 and statistic T^2, FAR is equal to 0.006 and zero ($p = 0.90$) for the temporal and delta approaches, respectively. The corresponding values for the Q statistic are equal to 0.006 and 0.013 ($p = 0.80$), respectively. The vast majority of FAR values were comparable to those obtained with the time-based approach. This point is critical when comparing fault detection strategies. Tables 6 and 7 show the detection time (or detection delay) for the statistics T^2 and Q, respectively. A fault was recognized after the occurrence of six consecutive points beyond the control limit, the detection time being calculated from the first [49]. For fault 5 and statistic T^2, it was equal to 16 and 15 ($p = 0.90$) sampling units given the periodic and delta procedures, respectively. That is, there was a gain of one sampling unit in this case. A detailed analysis of the event-based delta strategy is presented below, along with a comparison with the time-fixed approach.

From the MDR results obtained from the statistics T^2 and Q with PCA for the fixed-time approach (2nd column of Tables 2 and 3), the set of twenty-one faults of the Tennessee benchmark problem can be grouped into four subsets according to the level of detection difficulty. Group 1 refers to easy faults (1, 2, 4, 6, 7, 8, 12, 13, 14); group 2, to intermediate faults (11, 17, 18); group 3, to hard faults (5, 10, 16, 19, 20, 21); and group 4, to very hard faults (3, 9, 15).

First, significant MDR gains were verified for eight faults, all related to the Q statistic. Namely, faults 5 (74%; $p = 0.80$), 9 (48%; $p = 0.80$), 10 (60%; $p = 0.60$), 11 (26%; $p = 0.60$),

TABLE 5. FAR (false alarm rate) for the Q statistic.

Fault	Time-based approach	Event-based delta approach (p in %)										
		5	10	20	30	40	50	60	70	80	90	95
1	0	0	0.006	0.013	0.031	**0.044**	0.006	0	0	0	0	0
2	0.006	0.006	0.006	0.013	0.025	0.069	**0.006**	0.006	0.375	0	0	0
3	0.013	0.013	0.006	0.019	0.050	0.100	0	0.475	0.106	**0**	0	0
4	0.006	0.006	**0.006**	0.006	0.013	0.094	0.150	0.044	0.200	0.013	0.294	0
5	0.006	0.006	0.006	0.006	0.013	0.094	0.150	0.044	0.200	**0.013**	0.294	0
6	0	**0**	**0**	**0**	**0**	**0**	0.056	**0**	**0**	**0**	**0**	**0**
7	0	**0**	**0**	**0**	**0**	**0.031**	**0.031**	**0.031**	**0**	**0**	**0**	**0**
8	0.006	0	0.006	0.013	0.025	**0.019**	0	0.069	0	0	0	0
9	0.006	0.006	0.006	0.006	0.013	0.106	0.400	0.044	0.169	**0**	0	0
10	0	0	0	0	0.019	0.038	0.006	**0.006**	0.038	0	0	0
11	0.006	0.006	0.006	0.006	0.019	0.056	0.131	**0**	0	0	0	0
12	0	0	0	0.006	0.006	0.006	**0.031**	0.188	0	0	0	0
13	0	0	0	0	0	**0**	0	0	0	0	0	0
14	0	**0**	**0**	**0.006**	**0**	**0.013**	**0**	0	0	0	0	0
15	0.006	0.006	0	0	0	0.038	0.013	0.063	**0.038**	0	0	0
16	0.006	0.006	0.006	0.006	0.019	0.100	**0**	0.188	0.106	0	0	0
17	0.013	0.013	0.013	0.006	0.019	0.044	**0.044**	0.038	0	0	0	0
18	0.006	0.006	0.006	0.006	0.006	0.019	0.063	0	**0**	0	0	0
19	0	0	0	0.006	0.006	0.019	0	**0.019**	0	0	0	0
20	0	0	0	0	0.006	0.044	0	0.006	**0**	0	0	0
21	0.019	0.019	0.019	0.019	0.038	**0.038**	0.369	0	0	0	0	0

Gray cells: False alarm rate > 5%.

15 (69%; $p = 0.70$), 16 (47%; $p = 0.50$), 19 (35%; $p = 0.60$) and 20 (42%; $p = 0.70$) (Table 3). For example, for fault 10, the MDR values were equal to 0.659 and 0.061 ($p = 0.60$) for the periodic and delta strategies, respectively. The difference between them corresponds to the gain of 60%. The respective FAR values were considerably low, equal to 0.013, zero, 0.006, zero, 0.038, zero, 0.019 and zero, respectively (Table 5). Faults 5, 10, 16, 19 and 20 are hard-to-detect (group 3), and faults 9 and 15 are very hard-to-detect (group 4). Figure 6 shows the Q control charts obtained with the event-based delta rule for faults 5, 10, 16 and 20 (on the left). Corresponding charts for the usual fixed-time approach are also presented for comparison purposes (on the right).

Furthermore, it can be noted that the best MDR values are a function of the parameter p. However, they are concentrated at $0.50 < p < 0.80$. This result suggests an optimal range for p, which is positive in the sense of defining a single value. For example, the MDR for fault 5, initially equal to 0.001 ($p = 0.80$), is equal to 0.164 for the most usual value of p, equal to 0.60 (Table 3). The gain in the latter would still be high, around 58% (=0.746 − 0.164), with a FAR value still considerably low, equal to 0.044 (Table 5). An ensemble approach using an interval for p can also be considered.

There are cases where the gain is not relatively large; however, MDR tends to zero as desired. This was verified for fault 7, with an MDR of zero for $p = 0.70$, 0.80, 0.90 and 0.95, given the T^2 statistic (Table 2). The value corresponding to the periodic approach is relatively higher, equal to 0.085. For the Q statistic, it occurred for faults 17 and 18, both with an MDR of 0.034 for $p = 0.50$ and 0.70, respectively (Table 3). The corresponding values for the cyclic procedure are relatively larger, equal to 0.108 and 0.101, respectively. The MDR values for faults 5, 10 and 11, which showed significant gains (previous analysis), were also relatively low. Namely, 0.001 ($p = 0.80$), 0.061 ($p = 0.60$) and 0.099 ($p = 0.60$), respectively (Table 2). On the other hand, the values corresponding to the time-based approach are considerably high, equal to 0.746, 0.659 and 0.356, respectively. The FAR values in all cases were close to zero (Tables 4 and 5). Faults 11, 17 and 18 belong to group 2 (intermediate detection), and faults 5 and 10 to group 3 (hard detection).

Another performance metric concerns detection time. Tables 6 and 7 summarize the results for the statistics T^2 and Q, respectively. As before, a gray cell means it was not considered valid due to a false alarm rate above 5%. Shorter detection times were generally associated with lower MDR values (Tables 2 and 3), as expected. Relatively high MDRs and therefore low gains were obtained for faults 11, 16 and 21, given the T^2 statistic. The respective MDR gains were equal to 11% (=0.794 − 0.688; $p = 0.60$), 10% (=0.834 − 0.733; $p = 0.70$) and 8% (=0.736 − 0.651; $p = 0.60$) (Table 2), with FAR values equal to zero for all (Table 4). However, the main result in this case concerns the detection time, with gains of 208 (=304 − 96), 26 (=312 − 286) and 49 (=563 − 514) sampling units, respectively, in relation to the fixed-time approach (Table 6). That is, these faults were detected in advance, through the delta rule. Fault 11 belongs to group 2 (intermediate detection) and faults 16 and 21 to group 3 (hard detection). More, the detection times for faults 4, 9, 15 and 19 were equal to 447 ($p = 0.70$), 734 ($p = 0.70$), 646 ($p = 0.50$) and 444 ($p = 0.70$) sampling units, respectively, given the T^2 statistic (Table 6), with FAR values equal to zero for all (Table 4). These values are still relatively high, but these faults were not detected (ND) by the time-fixed approach (Table 6). For the Q statistic, this occurred for faults 3, 9 and 19, with detection times of 307 ($p = 0.80$), 233 ($p = 0.80$) and 217 ($p = 0.60$) sampling units, respectively (Table 7), with all FAR values close to zero (Table 5). These faults were also not detected (ND) by the time-based approach (Table 7). Fault 19 is hard-to-detect (group 3) and faults 3, 9 and 15 are very hard-to-detect (group 4). Furthermore, as described earlier, faults 10,

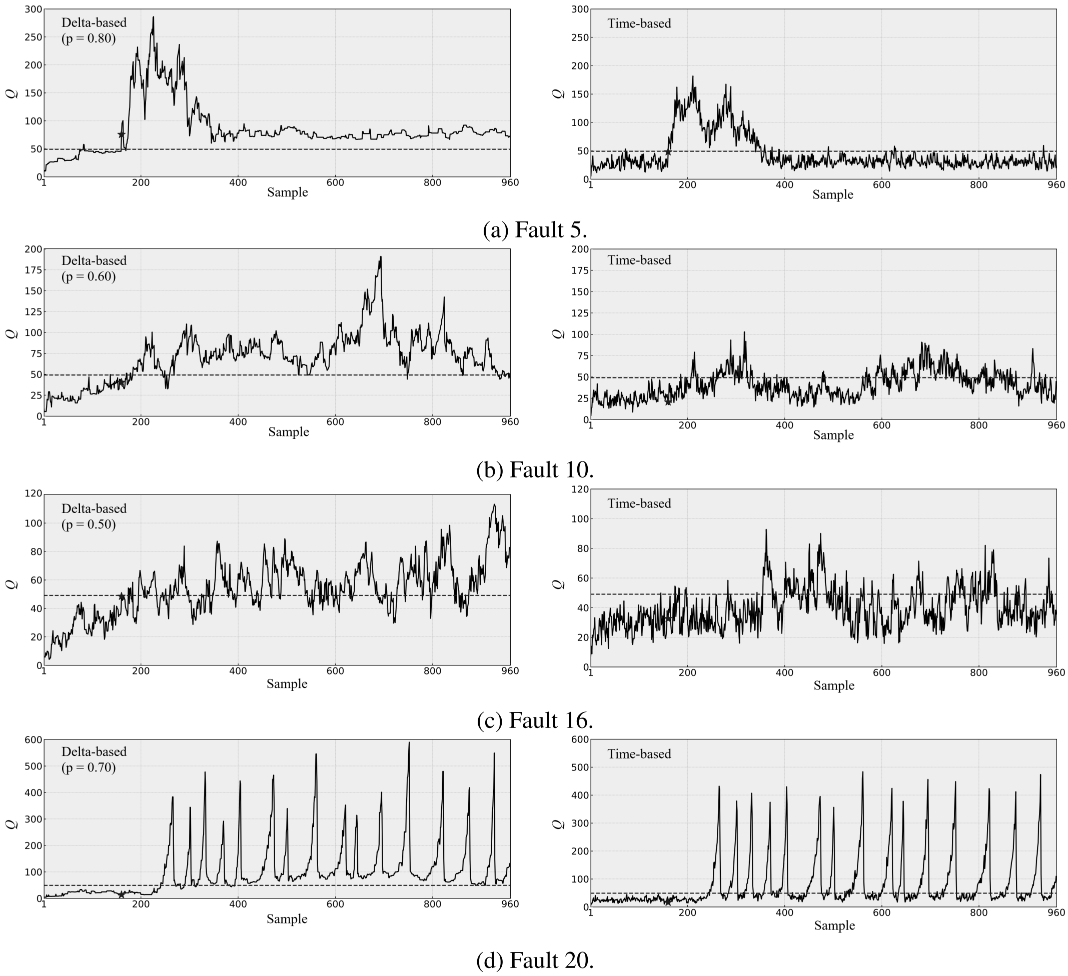


FIGURE 6. ***Q* control charts for the event-based delta method (on the left) and time-based approach (on the right) for hard-to-detect faults ($\star$: Fault start up at $t = 160$).**

15 and 16 showed significant MDR gains, given the Q statistic (Table 3). Thus, they also showed considerable gains in terms of detection time. Namely, 25 (=49 − 24; $p = 0.60$), 662 (=740 − 78; $p = 0.70$) and 161 (=197 − 36; $p = 0.50$) sampling units, respectively (Table 7), with FAR values close to zero for all (Table 5). Faults 10 and 16 are hard-to-detect (group 3), and fault 15 are very hard-to-detect (group 4). Intermediate faults 17 and 18 (group 2), with MDR gains of 7% (Table 3), also presented reasonable gains for detection time, given the Q statistic. Namely, 16 (=25 − 9; $p = 0.50$) and 72 (=84 − 12; $p = 0.70$) sampling units, respectively (Table 7), with FAR values close to zero (Table 5). Slightly longer detection times were observed for some faults in relation to the periodic procedure. Namely, for fault 14, equal to −1 sampling unit (=4 − 5; $p = 0.50$), and for fault 20, equal to −3 sampling units (=87 − 90; $p = 0.70$), given the T^2 statistic (Table 6). However, the respective MDR gains were equal to 6% (=0.158 − 0.094) and 5% (=0.701 − 0.650) (Table 2), with FAR values equal to zero (Table 4).

Two event-based systems were analyzed in this work, namely, the threshold and delta rules. The threshold method provided at least the same fault detection performance compared to the time-based approach, but with much less data. On the other hand, the delta rule yielded significantly superior performance, including hard-to-detect faults. This analysis also showed that the delta rule led to significantly superior performance, including hard-to-detect faults. This performance can be explained as follows. For the usual case of fixed-time data acquisition procedure, at each sampling interval, all measured samples of all monitored variables compose the input vector fed to the PCA-based fault detection system, which is characteristic of normal operating condition. With respect to the event-based data acquisition approach, only values beyond their respective limits are used in the input vector. The values not transmitted are replaced by the last ones sent for signal reconstruction. This second input vector generally presents more discrepant deviations from the normal condition compared to the first one provided by the fixed-time approach. This fact contributes to improve the fault detection performance. In other words, the most relevant changes in fault signals come earlier, favoring fault detection. This is in agreement with the best results generally obtained

TABLE 6. Time to detection (in sampling units) for the T^2 statistic.

Fault	Time-based approach	Event-based delta approach (p in %)											Absolute gain
		5	10	20	30	40	50	60	70	80	90	95	
1	7	**7**	**7**	**7**	**7**	**7**	**7**	7	**7**	**7**	9	9	0
2	17	**17**	**17**	**17**	**17**	19	19	**17**	19	19	19	19	0
3	–	–	–	–	–	–	–	–	–	–	–	–	**ND**
4	–	–	–	–	–	–	65	470	**447**	–	–	–	**ND**
5	16	13	13	13	16	15	14	1	13	1	**15**	19	1
6	10	10	10	10	10	10	10	**7**	8	8	8	8	3
7	1	1	1	1	1	1	1	1	**1**	**1**	**1**	**1**	0
8	23	23	**23**	23	**23**	25	**23**	**25**	**25**	25	27	27	0
9	–	–	–	–	–	–	781	782	**734**	–	–	–	**ND**
10	96	96	96	**96**	73	74	96	95	96	99	101	102	0
11	304	304	304	304	197	197	197	**96**	325	146	146	146	**208**
12	22	22	22	22	**22**	22	22	22	22	22	30	22	0
13	49	**49**	**49**	49	49	49	49	51	50	51	58	59	0
14	4	4	4	4	4	5	**5**	6	6	6	6	6	-1
15	–	–	–	–	–	–	**646**	781	–	–	–	–	**ND**
16	312	312	312	313	312	310	308	305	**286**	310	310	311	**26**
17	29	29	29	30	29	**29**	29	31	29	31	31	31	0
18	93	93	93	93	**89**	93	94	99	99	94	99	99	4
19	–	–	–	–	–	–	–	–	**444**	–	–	–	**ND**
20	87	87	87	87	87	87	90	87	**90**	90	91	91	-3
21	563	563	563	558	560	550	546	**514**	564	566	696	702	**49**

ND = Not detected by the time-based approach.
Gray cells: False alarm rate > 5%.

TABLE 7. Time to detection (in sampling units) for the Q statistic.

Fault	Time-based approach	Event-based delta approach (p in %)											Absolute gain
		5	10	20	30	40	50	60	70	80	90	95	
1	3	3	3	3	3	**2**	3	3	3	6	6	4	1
2	12	12	12	13	11	11	**12**	15	1	19	19	19	0
3	–	–	–	–	–	86	-	43	88	**307**	320	–	**ND**
4	3	5	**3**	3	3	3	1	15	1	1	1	30	0
5	1	2	2	2	1	1	1	1	1	**1**	1	11	0
6	1	**1**	**1**	**1**	**1**	**1**	1	**1**	**1**	**1**	**1**	**1**	0
7	1	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0
8	20	22	22	19	20	**16**	21	1	21	21	22	22	4
9	–	–	–	–	–	685	1	386	199	**233**	–	–	**ND**
10	49	49	49	49	48	45	47	**24**	27	43	75	75	**25**
11	11	11	11	10	10	6	6	**7**	48	25	25	12	4
12	8	8	8	8	7	7	**7**	8	8	22	22	22	1
13	37	37	37	37	41	**35**	38	44	43	45	45	49	2
14	1	**1**	**1**	**1**	**1**	**1**	**1**	2	2	2	2	2	0
15	740	–	–	–	–	573	99	1	**78**	-	746	586	**662**
16	197	196	196	196	196	13	**36**	36	13	195	258	247	**161**
17	25	25	25	25	24	24	**9**	24	24	24	24	24	**16**
18	84	85	84	85	85	15	1	8	**12**	17	92	92	**72**
19	–	–	–	–	–	15	18	**217**	110	550	478	557	**ND**
20	87	87	87	87	85	82	75	15	**82**	87	87	87	5
21	285	285	285	285	285	**283**	1	398	245	489	492	564	2

ND = Not detected by the time-based approach.
Gray cells: False alarm rate > 5%.

for intermediate values for the parameter p (Equation 4). While values of p close to zero produce similar results to the fixed-time approach, values close to one constitute a strong constraint for the initial detection of the faults. That is, the data transmission rate plays a considerable role in fault detection, and p can be seen as a sensitivity parameter. The highest performance was mostly for hard-to-detect faults.

VII. DATA TRANSMISSION RATE ANALYSIS

Data transmission in industrial systems has become increasingly critical due to the huge amount of continuously measured variables nowadays [65]. Greater efficiency is therefore crucial for better management of networking and computing aspects, among others. For example, with respect to latency and bandwidth, as well as memory allocation and processing power. These issues are also essential for advancing cyber-physical systems (CPS) applications. One way to reduce the amount of data to be transmitted and processed is through the event-based paradigm [4]. The motivation is that significant process changes are usually not periodic. This approach can also improve fault detection performance, as shown in this work mainly for hard-to-detect faults.

Figure 7a shows the average number of values transmitted in each sampling interval as a function of the parameter p (Equation 3). All fifty-two variables and twenty-one faults of the TEP benchmark were considered. The time-based approach ($p = 0$), in which all sampled values are passed

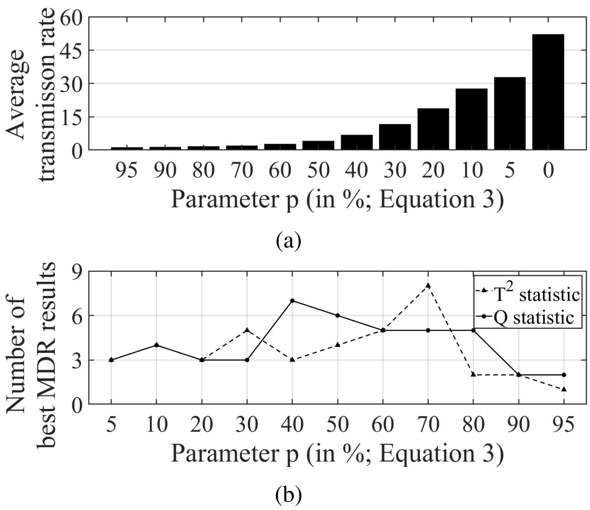


FIGURE 7. Effect of the parameter p (Equation 3) (a) on the average number of sampled values transmitted per sampling unit and (b) on the number of best MDR results, given the event-based delta rule (Equation 4).

continuously, serves as a reference. For the more restrictive case ($p = 0.95$) of the event-based delta method (Equation 4), less than five measured values are forwarded on average. The higher the value of this parameter, the lower the data transmission rate.

The best MDR results (section VI) were obtained for a range of values for p considering all twenty-one faults. Its definition must be a compromise between data transmission rate and fault detection performance. Figure 7b shows the number of best MDR values as a function of the p parameter. For example, there were three best MDR values for $p = 0.05$ (or 5%), given the T^2 statistic. They refer to faults 1, 2 and 13 (Table 2). This was also the case for the Q statistic, with faults 6, 7 and 14 (Table 3). A best result can be counted more than once, as it can occur for an interval of p, given a fault. For example, for fault 13, this occurs for $p = 0.05$ and 0.10 (MDR = 0.06) (Table 2). For T^2, the value of p with the highest number of best MDR values was equal to 0.70 (with eight faults), followed by 0.30 and 0.60 (with five). Given Q, the best score occurred for $p = 0.40$ (with seven faults), followed by $p = 0.50$ (with six) and $p = 0.60$, 0.70 and 0.80 (with five). In general, the best results were found for intermediate values of p, for both statistics. On the one hand, this means that it is not necessary to forward all sampled values ($p = 0$), and on the other hand, values of p close to one degrade the fault detection performance as it is very restrictive. The data transmission rate for intermediate values of p ($0.30 < p < 0.80$) varied between less than five and about ten observations on average, that is, around a maximum of 20% (≈10/52; Figure 7a). This number is considerably lower in relation to the full set of fifty-two values sampled. In addition to the benefit in terms of data transmission rate, an intermediate value of p also improved the fault detection performance compared to the usual fixed-time approach (section VI).

VIII. CONCLUSION

Data transmission in process industries usually occurs at every sampling interval. The problem is that a considerable amount of data without new process information is continually forwarded and processed. This adversely affects the communication system and computing power. For example, regarding network latency and bandwidth, and memory allocation and processing capacity. This is more critical nowadays, given the era of big data and cyber-physical systems (CPS) in the context of Industry 4.0. An efficient way to reduce the amount of data to be transmitted can be provided by the event-based paradigm. In this strategy, only sampled data associated with significant changes (events) in the process state are forwarded for further processing.

This work proposed an event-driven data acquisition framework for continuous industrial systems. Two methods were considered, namely, threshold and delta rules.Furthermore, the level of data filtering was varied in both. They were applied in the context of fault detection, using the Tennessee benchmark problem as a case study. The fault detection system was based on PCA (principal component analysis), which has been widely used for process monitoring. The threshold method provided similar results to the classical data acquisition time-based approach, with the advantage of using much less data. On the other hand, the delta method was generally better. Significant results, including mostly hard-to-detect faults, were achieved for considerably low data transmission rates, around 20% on average. In short, the event-based delta rule was able to reduce the amount of data to be transmitted and, at the same time, improve the fault detection performance compared to the fixed-time procedure.

Future work may be related to the search for an optimal value for the parameter p used for data filtering, sensor technology and communication systems, and signal reconstruction. This work adopted the PCA technique, which is commonly used for fault detection in the TEP benchmark problem. Artificial intelligence techniques can also be evaluated. Furthermore, while fault detection is still a major practical challenge, how the event-driven paradigm affects fault diagnosis is another point of investigation.

Concluding, event-driven data acquisition can be very attractive for process industries, given the large amount of variables measured continuously, even in seconds. The proposed event-based data acquisition framework can be applied directly to similar systems.

REFERENCES

[1] L. H. Chiang, R. D. Braatz, and E. L. Russel, *Fault Detection and Diagnosis in Industrial Systems*. 2001.

[2] F. Marvasti, *Nonuniform Sampling: Theory and Practice*. Springer, 2001.

[3] Y. Tsividis, "Event-driven data acquisition and digital signal processing—A tutorial," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 8, pp. 577–581, Aug. 2010.

[4] M. Miskowicz, *Event-Based Control and Signal Processing*. Boca Raton, FL, USA: CRC Press, 2015.

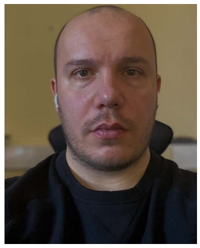
[5] X.-M. Zhang, Q.-L. Han, and B.-L. Zhang, ''Adaptive sampling based on amplitude sensitivity,'' *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 4–16, 2017.
[6] Z. Heng, P. Chen, Z. Jin, and Z. Chu, ''Event-triggered control in networked control systems: A survey,'' in *Proc. 27th Chin. Control Decis. Conf. (CCDC)*, May 2015, Art. no. 15340402.
[7] K.-E. Årzén, ''A simple event-based PID controller,'' in *Proc. 14th IFAC World Congr.*, vol. 18, 1999, pp. 423–428.
[8] R. Tomovic and G. Bekey, ''Adaptive sampling based on amplitude sensitivity,'' *IEEE Trans. Autom. Control*, vol. 11, no. 2, pp. 282–284, Apr. 1966.
[9] R. Dorf, M. Farren, and C. Phillips, ''Adaptive sampling frequency for sampled-data control systems,'' *IRE Trans. Autom. Control*, vol. 7, no. 1, pp. 38–47, Jan. 1962.
[10] P. Ellis, ''Extension of phase plane analysis to quantized systems,'' *IRE Trans. Autom. Control*, vol. AC-4, no. 2, pp. 43–54, Nov. 1959.
[11] W. Xie, X. Zhang, and K. Shi, ''Resilient security filtering for networked switched systems with event-driven measurements,'' *IEEE Access*, vol. 9, pp. 43434–43443, 2021.
[12] P. Khaledian, A. Aligholian, and H. Mohsenian-Rad, ''Event-based analysis of solar power distribution feeder using micro-PMU measurements,'' in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2021, pp. 1–5.
[13] S. M. Qaisar, F. Alsharif, A. Subasi, and A. Bensenouci, ''Appliance identification based on smart meter data and event-driven processing in the 5G framework,'' *Proc. Comput. Sci.*, vol. 182, pp. 103–108, Jan. 2021.
[14] W. Chen, H. Liu, and E. Qi, ''Discrete event-driven model predictive control for real-time work-in-process optimization in serial production systems,'' *J. Manuf. Syst.*, vol. 55, pp. 132–142, Apr. 2020.
[15] C. Santos, J. Jiménez, and F. Espinosa, ''Effect of event-based sensing on IoT node power efficiency. Case study: Air quality monitoring in smart cities,'' *IEEE Access*, vol. 7, pp. 132577–132586, 2019.
[16] M. de Castro Tomé, P. H. J. Nardelli, and H. Alves, ''Long-range low-power wireless networks and sampling strategies in electricity metering,'' *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1629–1637, Feb. 2019.
[17] E. Gascard and Z. Simeu-Abazi, ''Quantitative analysis of dynamic fault trees by means of Monte Carlo simulations: Event-driven simulation approach,'' *Rel. Eng. Syst. Saf.*, vol. 180, pp. 487–504, Dec. 2018.
[18] A. Theorin, K. Bengtsson, J. Provost, M. Lieder, C. Johnsson, T. Lundholm, and B. Lennartson, ''An event-driven manufacturing information system architecture for industry 4.0,'' *Int. J. Prod. Res.*, vol. 55, no. 5, pp. 1297–1311, Mar. 2017.
[19] A.-M. Barthe-Delanoë, S. Truptil, F. Bénaben, and H. Pingau, ''Event-driven agility of interoperability during the run-time of collaborative processes,'' *Decis. Support Syst.*, vol. 59, pp. 171–179, Mar. 2014.
[20] L. Zhao, T. Chai, and Q. Cong, ''Event-driven operation process monitoring of SBR wastewater process,'' in *Proc. 16th Triennial World Congr.* Amsterdam, The Netherlands: Elsevier, 2005, pp. 149–154.
[21] A. Sanchez, G. Rotstein, N. Alsop, J. P. Bromberg, C. Gollain, S. Sorensen, S. Macchietto, and C. Jakeman, ''Improving the development of event-driven control systems in the batch processing industry. A case study,'' *ISA Trans.*, vol. 41, no. 3, pp. 343–363, Jul. 2002.
[22] S. Wand and M. Guignard, ''Redefining event variables for efficient modeling of continuous-time batch processing,'' *Ann. Oper. Res.*, vol. 116, pp. 113–126, Oct. 2002.
[23] R. Doyle, L. Charest, N. Rouquette, J. Wyatt, and C. Robertson, ''Causal modeling and event-driven simulation for monitoring of continuous systems,'' in *Proc. 9th Comput. Aerosp. Conf.* Reston, VA, USA: American Institute of Aeronautics and Astronautics, 1993, p. 4525.
[24] S. Yin and O. Kaynak, ''Big data for modern industry: Challenges and trends [point of view],'' *Proc. IEEE*, vol. 103, no. 2, pp. 143–146, Feb. 2015.
[25] V. Venkatasubramanian, ''Drowning in data: Informatics and modeling challenges in a data-rich networked world,'' *AIChE J.*, vol. 55, no. 1, pp. 2–8, 2009.
[26] F. Zezulka, P. Marcon, Z. Bradac, J. Arm, T. Benesl, and I. Vesely, ''Communication systems for industry 4.0 and the IIoT,'' *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 150–155, Jan. 2018.
[27] J. Liu, J. Li, W. Li, and J. Wu, ''Rethinking big data: A review on the data quality and usage issues,'' *ISPRS J. Photogramm. Remote Sens.*, vol. 115, pp. 134–142, May 2016.
[28] K. J. Astrom and B. M. Bernhardsson, ''Comparison of Riemann and Lebesgue sampling for first order stochastic systems,'' in *Proc. 41st IEEE Conf. Decis. Control*, vol. 2, Dec. 2002, pp. 2011–2016.
[29] M. Miskowicz, ''Send-on-delta concept: An event-based data reporting strategy,'' *Sensors*, vol. 6, no. 1, pp. 49–63, 2006.
[30] J. J. Downs and E. F. Vogel, ''A plant-wide industrial process control problem,'' *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
[31] T. J. Rato and M. S. Reis, ''Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR),'' *Chemometrics Intell. Lab. Syst.*, vol. 125, pp. 101–108, Jun. 2013.
[32] C. Zhang, Q. Guo, and Y. Li, ''Fault detection in the Tennessee Eastman benchmark process using principal component difference based on k-nearest neighbors,'' *IEEE Access*, vol. 8, pp. 49999–50009, 2020.
[33] M. Ammiche, A. Kouadri, and A. Bakdi, ''A combined monitoring scheme with fuzzy logic filter for plant-wide Tennessee Eastman process fault detection,'' *Chem. Eng. Sci.*, vol. 187, no. 12, pp. 269–279, 2018.
[34] H. Ren, Y. Chai, J. Qu, K. Zhang, and Q. Tang, ''An intelligent fault detection method based on sparse auto-encoder for industrial process systems: A case study on Tennessee Eastman process chemical system,'' in *Proc. 10th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, Aug. 2018, pp. 2037–2042.
[35] I. Lomov, M. Lyubimov, I. Makarov, and L. E. Zhukov, ''Fault detection in Tennessee Eastman process with temporal deep learning models,'' *J. Ind. Inf. Integr.*, vol. 23, Sep. 2021, Art. no. 100216.
[36] P. Hajihosseini, M. M. Anzehaee, and B. Behnam, ''Fault detection and isolation in the challenging Tennessee Eastman process by using image processing techniques,'' *ISA Trans.*, vol. 79, pp. 137–146, Aug. 2018.
[37] Y.-Y. Hong and R. A. Pula, ''Methods of photovoltaic fault detection and classification: A review,'' *Energy Rep.*, vol. 8, pp. 5898–5929, Nov. 2022.
[38] A. R. Abbasi, ''Fault detection and diagnosis in power transformers: A comprehensive review and classification of publications and methods,'' *Electr. Power Syst. Res.*, vol. 209, Aug. 2022, Art. no. 107990.
[39] J. Chen, L. Zhang, Y. Li, Y. Shi, X. Gao, and Y. Hu, ''A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems,'' *Renew. Sustain. Energy Rev.*, vol. 161, Jun. 2022, Art. no. 112395.
[40] L. Zhang, M. Leach, Y. Bae, B. Cui, S. Bhattacharya, S. Lee, P. Im, V. Adetola, D. Vrabie, and T. Kuruganti, ''Sensor impact evaluation and verification for fault detection and diagnostics in building energy systems: A review,'' *Adv. Appl. Energy*, vol. 3, no. 25, 2021, Art. no. 100055.
[41] T. Xie, T. Wang, Q. He, D. Diallo, and C. Claramunt, ''A review of current issues of marine current turbine blade fault detection,'' *Ocean Eng.*, vol. 218, no. 15, 2020, Art. no. 108194.
[42] O. D. Mohammed and M. Rantatalo, ''Gear fault models and dynamics-based modelling for gear fault detection—A review,'' *Eng. Failure Anal.*, vol. 117, Nov. 2020, Art. no. 104798.
[43] G. Faure, M. Vallée, C. Paulus, and T. Q. Tran, ''Fault detection and diagnosis for large solar thermal systems: A review of fault types and applicable methods,'' *Sol. Energy*, vol. 197, pp. 472–484, Feb. 2020.
[44] Y.-J. Park, S.-K. S. Fan, and C.-Y. Hsu, ''A review on fault detection and process diagnostics in industrial processes,'' *Ind. Eng. Chem. Res.*, vol. 8, no. 9, pp. 1–26, 2020.
[45] H. Chen and B. Jiang, ''A review of fault detection and diagnosis for the traction system in high-speed trains,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 450–465, Feb. 2020.
[46] Y. Du and D. Du, ''Fault detection and diagnosis using empirical mode decomposition based principal component analysis,'' *Comput. Chem. Eng.*, vol. 115, pp. 1–21, Jul. 2018.
[47] S. Gajjar, M. Kulahci, and A. Palazoglu, ''Real-time fault detection and diagnosis using sparse principal component analysis,'' *J. Process Control*, vol. 67, pp. 112–128, Jul. 2018.
[48] Y. Liu, G. Zhang, and B. Xu, ''Compressive sparse principal component analysis for process supervisory monitoring and fault detection,'' *J. Process Control*, vol. 50, pp. 1–10, Feb. 2017.
[49] 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.
[50] D. Serpanos, ''The cyber-physical systems revolution,'' *Computer*, vol. 51, no. 3, pp. 70–73, Mar. 2018.
[51] C. Kalalas and J. Alonso-Zarate, ''Sensor data reconstruction in industrial environments with cellular connectivity,'' in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Aug. 2020, pp. 1–6.

[52] M. El Koujok, A. Ragab, H. Ghezzaz, and M. Amazouz, "A multiagent-based methodology for known and novel faults diagnosis in industrial processes," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3358–3366, May 2021.
[53] J. Jiao, W. Zhen, W. Zhu, and G. Wang, "Quality-related root cause diagnosis based on orthogonal kernel principal component regression and transfer entropy," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6347–6356, Sep. 2021.
[54] R. Marino, C. Wisultschew, A. Otero, J. M. Lanza-Gutierrez, J. Portilla, and E. D. L. Torre, "A Machine-Learning-Based distributed system for fault diagnosis with scalable detection quality in industrial IoT," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4339–4352, Mar. 2021.
[55] Z. Chen, Y. Cao, K. Zhang, T. Koenings, T. Peng, C. Yang, W. Gui, and S. X. Ding, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2710–2720, May 2019.
[56] K. Serverson, P. Chaiwatanodom, and R. D. Braatz, "Perspectives on process monitoring of industrial systems," *Annu. Rev. Control*, vol. 42, pp. 190–200, Dec. 2016.
[57] 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.
[58] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2002.
[59] K. Staszek, S. Koryciak, and M. Miskowicz, "Performance of send-on-delta sampling schemes with prediction," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jun. 2011, pp. 2037–2042.
[60] Y. S. Suh, "Send-on-delta sensor data transmission with a linear predictor," *Sensors*, vol. 7, no. 4, pp. 537–547, 2007.
[61] V. H. Nguyen and Y. S. Suh, "Networked estimation with an area-triggered transmission method," *Sensors*, vol. 8, no. 2, pp. 897–909, 2008.
[62] M. Miskowicz, "Efficiency of event-based sampling according to error energy criterion," *Sensors*, vol. 10, no. 3, pp. 2242–2261, Mar. 2010.
[63] J. Zhou and Y. Zhu, "Identification based fault detection: Residual selection and optimal filter," *J. Process Control*, vol. 105, pp. 1–14, Sep. 2021.
[64] A. Hamadouche, "Model-free direct fault detection and classification," *J. Process Control*, vol. 87, pp. 130–137, Mar. 2020.
[65] S. J. Qin, "Process data analytics in the era of big data," *AIChE J.*, vol. 60, no. 9, pp. 3092–3100, 2014.

DANIEL GUTIERREZ-ROJAS (Student Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Antioquia, Colombia, in 2016, and the M.Sc. degree in protection of power systems from the University of São Paulo, Brazil, in 2017. He is currently pursuing the Ph.D. degree with the School of Energy Systems, LUT University, Finland. From 2017 to 2019, he worked as a Security of Operation and Fault Analyst with Colombia's National Electrical Operator. His research interests include predictive maintenance, power systems, micro-grids, mobile communication systems, and electrical protection systems.

PEKKA RÄSÄNEN received the B.Sc. degree in automation from the EVTEK University of Applied Sciences of Vantaa, Finland, in 2008, and the M.Sc. degree in electrical engineering (the industrial IoT) from the LUT University of Lappeenranta, Finland, in 2020. He worked with ABB Drives as an Electrical Designer, from 2010 to 2015; with ABB Marine as an Electrical Designer, from 2015 to 2019; and with the Electrical Steering Unit, Azipod, as a Lead Engineer, from 2019 to 2021. Since 2021, he has been working with ABB Drives as a Global Process Owner of electrical engineering as well as a Global Owner for electrical, mechanical and thermal tools.

ANA BRANDÃO BELISÁRIO received the B.Sc. and M.Sc. degrees in chemical engineering from the Federal University of Minas Gerais, Brazil, in 2017 and 2020, respectively. During the bachelor's degree, she received an Exchange Scholarship to study at Hochschule Niederrhein, Germany, and during the master's degree, a research grant to develop a part of her work at the LUT University of Lappeenranta, Finland. She currently works with data science, machine learning and statistics in chemistry and chemical engineering. Her research interests include mixture-process design of experiments and predictive chemistry.



MERIM DZAFERAGIC received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of Sarajevo, Bosnia and Herzegovina, in 2011 and 2013, respectively, and the Ph.D. degree in telecommunications engineering from Trinity College Dublin, Ireland, in 2020. His work has focused on complex system science and machine learning applied to communication networks. His work has been focused on the impact of topological structures on the overall network performance, as well as the use of different machine learning approaches to support the operation of the industrial IoT networks. As part of his Ph.D., he has been working on the concept of emergence, distributed optimization, and self-organization solutions for communication networks. His most recent work focuses on intelligent control in cellular networks.

GUSTAVO MATHEUS DE ALMEIDA received the B.Sc. and M.Sc. degrees in chemical engineering from the Federal University of Minas Gerais, Brazil, in 2003 and 2006, respectively, and the Ph.D. degree in chemical engineering from the University of São Paulo, Brazil, in 2006. He was a Postdoctoral Fellow at the Federal University of Minas Gerais, from 2006 to 2008; a Professor at the Federal University of São João del-Rei, Brazil, from 2009 to 2014; and a Visiting Professor at the University of Coimbra, Portugal, in 2011. Since 2015, he has been a Professor at the Department of Chemical Engineering, Federal University of Minas Gerais, and since 2017, he has been a Co-ordinator of the Research Group on Data Analysis. His research interests include data-driven modeling, industrial statistics, data visualization, and process monitoring in continuous process industries.

PEDRO H. J. NARDELLI (Senior Member, IEEE) received the B.S. and M.Sc. degrees in electrical engineering from the State University of Campinas, Brazil, in 2006 and 2008, respectively, and the Ph.D. degree from the University of Oulu, Finland, and the State University of Campinas, following a Dual Degree Agreement, in 2013. He is currently an Assistant Professor (Tenure Track) in the IoT in energy systems at LUT University, Finland, and holds a position of the Academy of Finland Research Fellow with a project called Building the Energy Internet as a Large-Scale IoT-Based Cyber-Physical System that manages the energy inventory of distribution grids as discretized packets via machine-type communications (EnergyNet). He leads the Cyber-Physical Systems Group, LUT, and is a Project Co-ordinator of the CHIST-ERA European Consortium Framework for the Identification of Rare Events via Machine Learning and IoT Networks (FIREMAN). He is also an Adjunct Professor at the University of Oulu in the topic of communications strategies and information processing in energy systems. His research interest includes wireless communications particularly applied in industrial automation and energy systems. He received the Best Paper Award of IEEE PES Innovative Smart Grid Technologies Latin America 2019 in the track Big Data and Internet of Things. More information can be found at https://sites.google.com/view/nardelli/.