

Toward statistical real-time power fault detection

Mantautas Rimkus

Piotr Kokoszka

Colorado State University

Colorado State University

Kumaraguru Prabakar

Haonan Wang

National Renewable Energy Laboratory

Colorado State University

March 15, 2023

Abstract

We propose statistical fault detection methodology based on high-frequency data streams that are becoming available in modern power grids. Our approach can be treated as an online (sequential) change point monitoring methodology. However, due to the mostly unexplored and very nonstandard structure of high-frequency power grid streaming data, substantial new statistical development is required to make this methodology practically applicable. The paper includes development of scalar detectors based on multichannel data streams, determination of data-driven alarm thresholds and investigation of the performance and robustness of the new tools. Due to a reasonably large database of faults, we can calculate frequencies of false and correct fault signals, and recommend implementations that optimize these empirical success rates.

Keywords: high-frequency streaming data, power grid faults, sequential change point detection.

1 Introduction

Faults in power systems cause excessive currents and pose safety threats to people and property, and even cause major fires with substantial economic and social impacts. Fast and reliable detection of faults is therefore of paramount importance. Traditional methods use high fault current magnitude and power flow direction to differentiate between normal operating conditions and faulted conditions. Integration of inverter based (solar, wind) distributed energy resources (DERs) in the distribution system has created bi-directional current flows under normal operation and reduced the effectiveness of the traditional methods based on current direction. This challenge has created great interest in new methods of detecting faults in power grids. Our objective is to contribute to the recent intensive research in this area by proposing a relatively simple, but effective, way to detect power grid faults. Our approach is different from existing engineering approaches and points toward possibilities of using high-frequency measurements generated by modern measurement units. It is anchored in statistical sequential change point monitoring methodology that goes back to the work of Page (1954) and has been developed in many directions, see Lorden (1971), Lai (1998) and the book of Tartakovsky *et al.* (2015), among hundreds of significant contributions. However, due to the mostly unexplored and very nonstandard structure of high-frequency power grid data, a great deal of exploration and many adjustments need to be made to make the approach practically applicable. The objective of this paper is to propose a general statistical paradigm for power grid fault detection using the type of data that are becoming available in real time. In addition to the new statistical methodology, an advance over very extensive engineering literature is the computation of success rates, similar to type I and II errors, and incorporation of incomplete data (some measurements may be unavailable).

Existing fault detection methods in power systems fall into the following three categories: qualitative model-based, quantitative model-based, and data-driven approaches, see Jiang *et al.* (2011). Data-driven approaches have recently been receiving considerable attention for the

following reasons: (1) Various intelligent electronic devices (IEDs) have been widely adopted and installed in the power grid (Moghaddass and Wang, 2017; Ghosal and Conti, 2019; Chen *et al.*, 2020). These devices collect large amounts of different data at many nodes across the grid. (2) Compared to the traditional model-based approaches, the data-driven approaches are more resilient against system model misspecification and adapt better to the variations in system components and/or topology (Chen *et al.*, 2018; Chen *et al.*, 2019; Tripathi and De, 2018; Yin *et al.*, 2014). To illustrate the scope of research on power grid fault detection, without aiming at a complete list of references, we also cite (Zhou *et al.*, 2019; Chow *et al.*, 2007; Xie *et al.*, 2014; Kaci *et al.*, 2014; Kim *et al.*, 2017; Cui *et al.*, 2019; Nguyen *et al.*, 2015; Rafferty *et al.*, 2016).

Recent new approaches generally test a proposed method to a single fault. For example, using the magnitude of the voltage, Gholami *et al.* (2019) develop advanced fault detection methods based on multiple detectors. The approach of Hannon *et al.* (2021), based on machine learning, is very promising, but the application uses only one bus. The focus of Ardakanian *et al.* (2016) is the estimation of the admittance matrix. A change in this matrix and its localization provide information on the timing and the localization of a fault. The algorithm is evaluated on a 13 bus grid, similar to the one we use, but only on a single fault. Recent methods for fault localization, e.g. Li *et al.* (2019) and Khushwant *et al.* (2021), require the knowledge of the time the fault occurs, the issue we address. The importance of incorporating PMU measurement before, during and after a fault was emphasized by Chen *et al.* (2016) who proposed a method to detect and identify faults in near real-time by exploiting the statistical properties of voltage phase-angle measurements. We take a similar view. Our work has a statistical dimension because we apply many versions of our method to a relatively large database of 55 faults, and report success and failure rates, similar to size and power of a significance test, to identify the best methods. As we explain in Section 2, the 55 faults we consider are basically all possible faults in a 13 bus system we use as a test bed.

To summarize, we develop a basically complete statistical methodology to detect a fault in a small grid based on high-frequency measurements at selected nodes of the grid. Since the temporal resolution of data streams is very high, the detection is almost immediate. Such an immediate detection can be expected of many other methods, if a fault in fact did occur. We therefore focus on the correct decision. It is important to avoid false alarms because cutting off power has serious consequences. On the other hand, failing to detect a fault may have even more serious consequences. We therefore focus on frequencies of correct and incorrect decisions, similarly as in the Neyman–Pearson testing paradigm. For this, reasonably many different faults are needed. We address the following items: (1) preprocessing of irregularly recorded data streams, (2) development of scalar detectors based on multichannel data streams, (3) determination of dynamic alarm thresholds, (4) investigation of the performance and robustness of our methodology. In the end, we are able to recommend specific implementations that work very well in our test-bed grid. However, our process of statistical methodology development can be even more valuable, as it can be viewed as a recipe that could be applied in similar, but different, settings.

The remainder of the paper is organized as follows. Section 2 is dedicated to the description of the data we study, Section 3 to the development of the methodology and Section 4 to the examination of its performance. We summarize and provide an illustrative algorithm in Section 5.

2 IEEE 13 bus feeder: structure and data generation

Real power grid fault data are generally not publicly available; power companies do not share them, chiefly for legal reasons. However, the mechanisms and types of faults that occur in real power grids are well understood. National Renewable Energy Laboratory (NREL) has hardware and software to generate faults of any type at specified times and locations within a grid that

can be designed to correspond to a problem to be studied. In this paper, we use a fairly standard grid design, the IEEE 13 bus feeder, to generate streaming data and faults. The small grid we consider is the usual test bed in power grid research that can be identified with a small subgrid within which a fault can occur. We now proceed to describe it and the data that have been generated at NREL to develop a statistical fault detections technique.

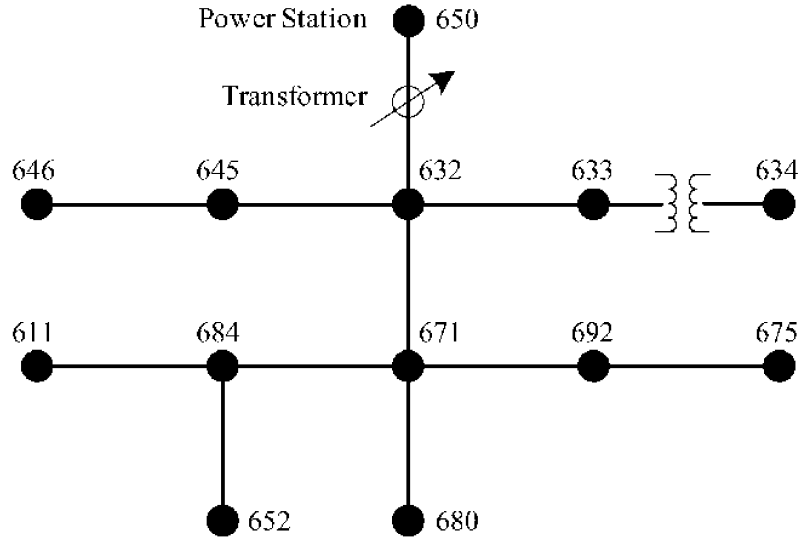


Figure 1: IEEE 13 bus feeder representation.

The general structure of the IEEE 13 bus feeder, PES (2020), is shown in Figure 1; it is a connected graph with 13 nodes. In power grid research a node is called a bus. For example, we refer to node 650 as Bus 650. Measurement devices are placed at buses. They cannot be placed at transmission lines that connect the buses. We use only data from 6 buses, similarly to Onaolapo *et al.* (2019) and Hu *et al.* (2013). We assume that each of these 6 buses represents a different part of the system:

1. Bus 650 (the substation)
2. Bus 632 (includes buses 645, 646, 633)
3. Bus 634

4. Bus 671 (includes buses 684, 611, 652, 692)

5. Bus 675

6. Bus 680

The above partition of the grid is dictated by the way the buses are connected and power transmitted. For example, there is only a circuit breaker between Bus 671 and Bus 692, so they are identical buses when the circuit breaker is closed. The data at Bus 611 and Bus 652 can be assumed to be identical to the data at Bus 671 because the lines connecting Bus 671 to Bus 611 and Bus 652 are short. Similar arguments hold for Bus 632 with the surrounding buses; voltage drops are insignificant from the perspective of the target application.

To explain what data are available at each of the six buses, we must briefly describe how power is generated and transmitted. We provide only the most essential facts that are well-known to power grid engineers and researchers, but may be less known to statisticians. Detailed background is available in dozens of textbooks, Glover *et al.* (2022) is a recent edition of a popular textbook. Alternating current (AC) is generated by a generator that has a large electromagnet spinning inside stationary coils of wire (windings). Industrial generators have three separate windings, each producing its own current. These separate currents are referred to as A, B and C phases. Measurement devices at buses can measure each phase separately and all three phases combined. If there are no faults, the voltage and current of the phases are shifted exactly by $\pi/3$ radians. This corresponds to the equi-spaced locations of the three windings inside which the electromagnet rotates at a constant speed, generating currents of constant frequency (nominally 60 Hz in the United States). *At each bus, we thus have measurements of 12 variables: voltage (in kilovolts), current (in kiloamperes) and frequency (in hertz) for phases A, B, C, and three-phase.* From these measurements, voltage and current Root Mean Square (RMS) can be computed, which is the peak voltage or current divided by $\sqrt{2}$. This convention is routinely used in sinusoidal AC power transmission systems because then the transmitted power can be calculated as the product of the RMS current and RMS voltage. *If there is a*

fault, the measured current, voltage and frequency will be impacted, but they also exhibit small fluctuations in the absence of any faults.

This IEEE 13 bus systems is simulated in a digital real time simulator (DRTS) in electromagnetic transients (EMT) domain. The RMS of voltage and current are calculated in real-time. The frequency measurements are calculated using phase locked loop (PLL) approach. The measurements are stored as csv files through UDP (User Datagram Protocol) streaming from DRTS.

For this study, 55 different simulations of faults were conducted. These include all possible bus faults in this test grid. We used a fault model from the RSCAD library which uses an impedance in the path to the ground. We used the value of 0.1 Ohms for the impedance value. In each simulation, different types of faults were applied to buses 632, 634, 646, 650, 671, 675, 680, 684. Additionally, faults were applied at the middle of the line between buses 632 and 671. In each simulation, the fault was applied at the beginning of second 10 and the simulation was running for another 5 seconds. The total length of each simulation is thus 15 seconds. It is assumed that the earliest time when a fault can be detected is the start of second 10. If a fault is detected earlier, it is treated as a false detection. Depending on the type of the faulted bus (two-phase or three-phase), up to seven different types of faults are possible. These types are (1) Phase A fault (2) Phase B fault (3) Phase C fault (4) Phases A and B fault (5) Phases A and C fault (6) Phases B and C fault (7) Three-phase fault. Also, additionally, one 35 second long simulation without a fault was conducted. It is used for training certain aspects of our algorithm, as elaborated on in Section 4.1.

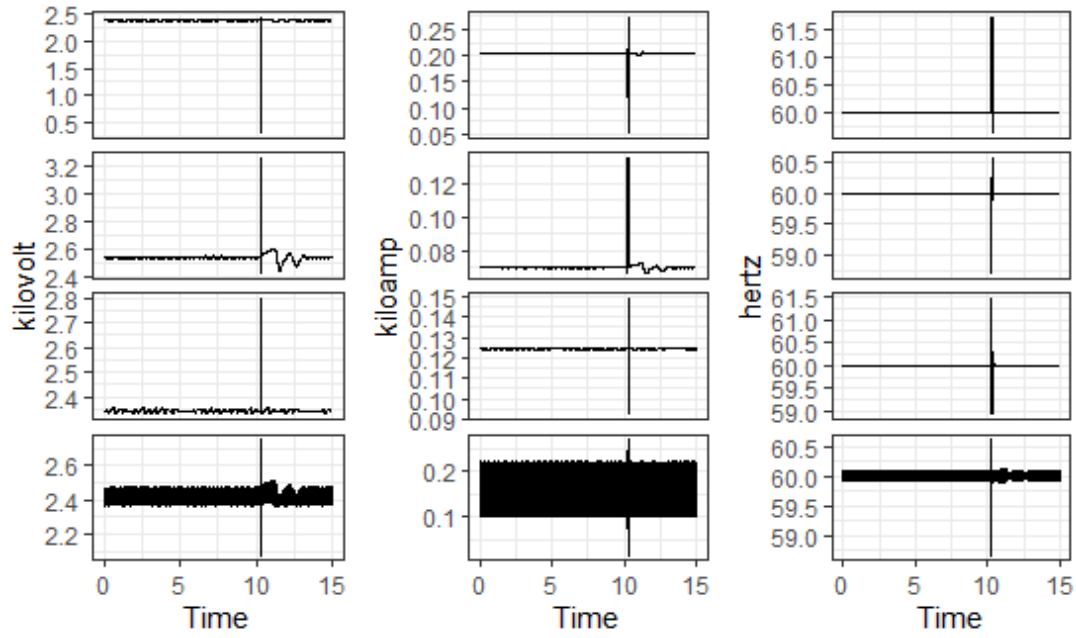


Figure 2: Measurements at Bus 675 with phase A fault applied at Bus 632. The fault is applied at the beginning of second 10. From Top to Bottom - phases A, B, C, and Three-phase measurements. From Right to Left - voltage, current, and frequency measurements.

Figure 2 presents a part of the data for Phase A fault at Bus 632. The presented data contains the measurements at Bus 675. Each time series goes through noticeable perturbations after the fault. Figure 3 presents the same data, but focuses on the first 0.3 seconds after the start of second 10, i.e. after the fault. Notice that the first visible reaction can be seen only around 0.25 seconds later.

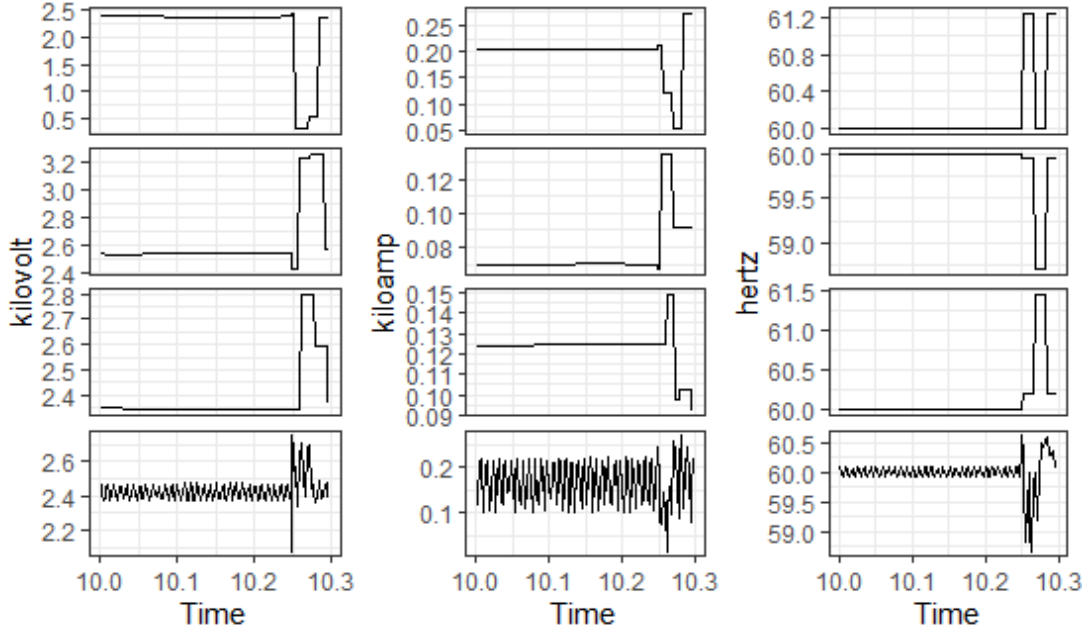


Figure 3: Close-ups of the data shown in Figure 2, the first 0.3 seconds after the fault.

If we take all simulations and all buses into account, we would end up with $6 \times 55 = 330$ Figures analogous to Figure 3. Visual examination of some of them reveals that while delays are similar, the shapes of responses differ. A relationship between the location of a fault and a location of the bus at which it is measured, as well as the responses of the 12 measured variables and the type of the fault, are not easy to establish. In the next section, we describe a general approach that uses all available information and work out the details of specific, most effective approaches. The effectiveness is measured by a low rate of false alarms and a high rate of correct fault detections.

3 Methodology

The goal of statistical change-point detection is to determine if a change in the structure of data has occurred and to estimate the time of the change. The context of the streaming power grid data we consider falls into the field of on-line, real-time or sequential change-point detection, or

change point monitoring. All these terms are almost synonymous. There are now many monographs on change point detection, e.g. Brodsky and Darkhovsky (1993), Gustafsson (2000), Chen and Gupta (2011), Basseville *et al.* (2012) and Brodsky (2017). There is a continuous stream of publications within this field that focus on specific aspects, models and applications. With the awareness of listing a subjectively selected small sample, we cite Aminikhanghahi *et al.* (2019), Horváth *et al.* (2021), Zhang *et al.* (2021), Estrada Gómez *et al.* (2022), who also provide many references specific to the directions they advance.

Our purpose is to develop an effective method of detecting a fault in power grid streaming data. We will explain the relevant statistical concepts as we progress with the methodology development. In our approach, we use a moving window change-point detection technique.

This section is organized as follows. First, in Section 3.1, we focus on the irregular structure of the data and present our regularization approach. In Section 3.2, we describe the derivation of the methodology leading to change-point detection. We explore a few different approaches. The proposed methodology requires a suitable normalization of the data streams, which is described in Section 3.3. In Section 3.4, we describe the derivation of the threshold for decision-making, the approach for determining the final form of our fault detection methodology, and criteria for parameter tuning.

3.1 Regularization

The data capturing mechanism used by the NREL hardware and software provides unevenly spaced time series, so a regularization approach is needed. The 55 simulations share an almost identical data capture mechanism. Within a simulation, all 12 variables, like phase A current or three-phase voltage, are recorded at the same time. Thus, at any moment, if the value of one variable is recorded, the values of for the remaining 11 variables are recorded as well. The recording times are however not identical for the 55 simulations. Moreover, if different devices are used in a monitoring of a power grid, the assumption of the identical recording times for all

variables might not hold. The regularization approach proposed below would work even in this more general setting.

To illustrate the irregularity of the measurement time points, we use a simulation with A phase fault at Bus 632 and analyze differences between subsequent time points. The histogram is presented in Figure 4. The minimum of differences is 0.00096s, the maximum is 0.00798s (more than a 700% difference). Most of the values group around 0.001, 0.002 and 0.003 seconds. For other simulations, the broad picture is similar, but the distributions of the time points are slightly different.

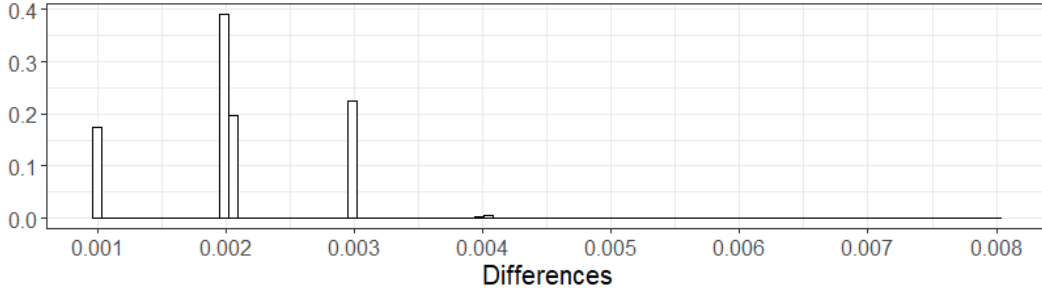


Figure 4: The unit area histogram of differences between subsequent time points of A phase fault at Bus 632 simulation.

The proposed regularization procedure consists of dividing the total time of a simulation into a set of fixed length subintervals. Each second is divided into D equal intervals of length $d = 1/D$ seconds. If D is too large, the intervals will be too small, some would contain no data points, eventually leading to additional noise in the data. If D is too small, intervals would contain many data points, leading to a loss of valuable information. In our analysis, we used $D = 500$, to reflect the most typical time points separations described above.

Denote by $t_1, t_2, \dots$ the end points of the intervals, where $t_{i+1} - t_i = d$. Denote by X_t a variable, e.g phase A voltage, observed at time t . The value X_{t_j} is calculated as the average of the X_t with $t_j - d < t \leq t_j$. If such X_t do not exist (there are no observations in the interval $(t_j - d; t_j]$), we assign $X_{t_j} := X_{t_j-d}$. Recall that each simulation with a fault is 15 seconds long

and starts at second 0, thus this procedure leads to a maximum of 7,500 (with $D = 500$) different equally spaced time points, and our time series are defined at times $\{0.002, 0.0004, \dots\}$ seconds. If a fault is detected, we no longer need to continue to calculate the regularized X_{t_j} .

3.2 Moving window detection algorithm

We begin by introducing a convenient notation. Fix a simulation and consider all-time series $X_{(b',k',f')}$. We use the primers to distinguish the indexes identifying the available data streams from parameters of a statistical fault detection procedure. The index b' refers to a bus number, so $b' \in \{650, 632, 634, 671, 675, 680\}$. The index k' refers to a type of variable and k' can be either Voltage, Current, or Frequency. The index f' refers to a type of phase and can be either A, B, C, or three-phase. For example, $X_{(650, \text{Voltage}, \text{A})}$ refers to A-phase voltage measurements at Bus 650. and $X_{(650, \text{Voltage}, \text{A})}(t)$ is to A-phase voltage measurement at Bus 650 at time t . As described before, we have six buses, three types of variables and four types of phases. Our purpose is to determine the time of a fault using a suitable change-point detection technique based on a moving window. A moving window ensures that a detection statistic can be calculated in real-time, as it is based on only a limited number of observations, and that it adjusts to the most recent state of the system. The slow evolution of system readings does not indicate a fault of the type we are aiming to detect.

Denote by l the length of the moving window in seconds. We split the moving window into two intervals of lengths pl and $(1 - p)l$ seconds. The first interval is used to evaluate the prior state. The second interval is used to evaluate the current state. It is natural to choose $p > 0.5$, as it provides a more stable evaluation of the prior state. This leads to $(1 - p) < 0.5$. The shorter interval helps better capture the current state and respond to a fault quickly. However, the interval that is too short might put too much emphasis on noisy observations, thus one needs to find optimal values. We explain in the following how this is done. Assume that t_i is the current time point. Notice that if $t_i < l$, the moving window intervals cannot be built. Without

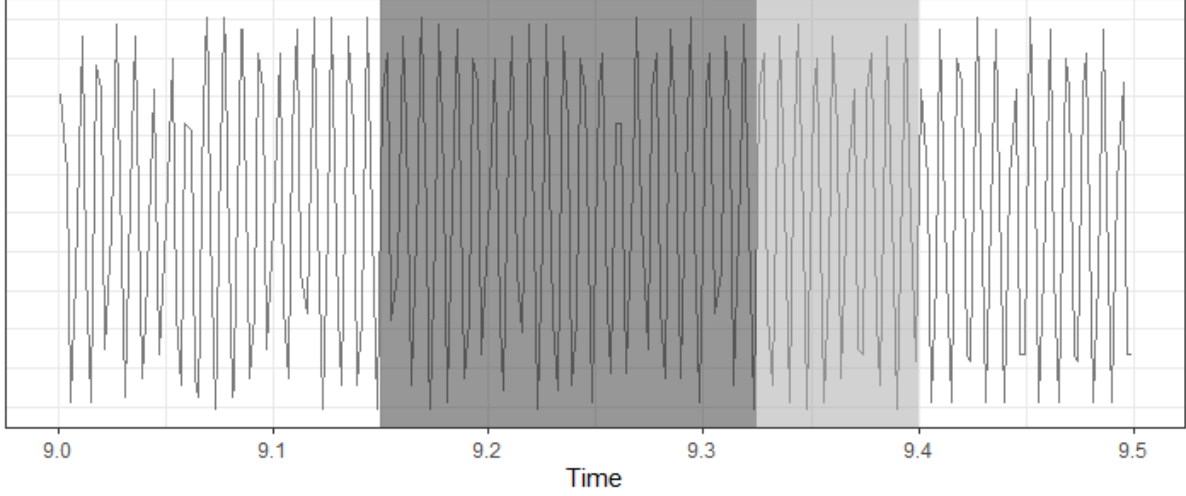


Figure 5: An example of the moving window using $l = 0.25$ seconds and $p = 0.7$. Here, the current time is $t_i = 9.4$ seconds. Dark grey area represents the interval of length pl and light grey - the interval of length $(1 - p)l$.

loss of generality, we assume that $t_i \geq l$, because we do not expect a fault within a fraction of a second after the monitoring has begun. The moving window with the two subintervals is shown in Figure 5.

We now turn to the description of a general class of tests statistics, often referred to as detectors, that are based on the two subintervals of the moving window. Set

$$(3.1) \quad V_{(b',k',f')}(t_i) = [f(X_{(b',k',f')})(t_i) - \tilde{f}(X_{(b',k',f')})(t_i)] ,$$

where $f(X_{(b',k',f')})(t_i)$ is some function f computed from the trajectory $X_{(b',k',f')}$ using the points t_j , such that $t_i - l \leq t_j < t_i - (1 - p)l$ (the first interval) and $\tilde{f}(X_{(b',k',f')})(t_i)$ is the same function on $t_i - (1 - p)l \leq t_j \leq t_i$ (the second interval). Thus, for each combination of indexes, (b',k',f') , we acquire a statistics $V_{(b',k',f')}(t_i)$ at the current time point t_i .

The next step is to specify the functions f that can capture differences between the two intervals well and propose a method to combine the $V_{(b',k',f')}$ over all indexes. One should notice that the pool of choices is wide, and considering all of them is not practical. We considered several

options for f and several ways to combine the $V_{(b',k',f')}(t_i)$ across the indexes. The performance of these different approaches is evaluated in Section 4. The choices of f we considered are listed in Table 1.

Table 1: The functions f considered in this study. The time t_i is the end point of the moving window. The specified functions are computed over the two intervals shown in Figure 5.

Name	$f(X_{(b',k',f')})(t_i)$	Definition
Mean	$\bar{X}_{(b',k',f')}$	The average of observations
SQMean	$\bar{X}^2_{(b',k',f')}$	The average of the squared observations
Range	$\max(X_{(b',k',f')}) - \min(X_{(b',k',f')})$	The range of observations
Median	$\text{median}(X_{(b',k',f')})$	The median of observations

The potential functions f in (3.2) can be viewed as features computed from the data on each of the two subintervals of the moving window. One can clearly come up with many functions, which could also be combined in various ways, e.g. by addition or multiplication. The functions listed in Table 1 are just the most commonly used statistics and, as we shall see, already lead to reliable detection. As illustrated in Figure 2, a fault is “seen” by different measurements at different buses differently, and it is not clear what differences between the two intervals one should look for. Generally, it is some sort of level shift, but the direction of the shift is unclear, and it may consist of several “sub-shifts”. However, if a moving window starts to cover the fault, the differences between the two intervals should become visible for suitable functions f that must be determined empirically. For the streaming data, we think of the window as a filter through which that data stream passes. This filter must be tuned to detect a fault most effectively.

To combine the differences $V_{(b',k',f')}(t_i)$ into suitable real-valued statistics, we considered three methods, which result in statistics we denote as V_{mean} , V_{median} , and V_{trunc} . They are defined as follows. The detector V_{mean} is the average of the squared differences:

$$(3.2) \quad V_{\text{mean}}(t_i) = \frac{\sum_{(b',k',f')} \{V_{(b',k',f')}(t_i)\}^2}{\#(b',k',f')}.$$

Formula (3.2) shows that if one or more of the differences $V_{(b',k',f')}(t_i)$ between the two subintervals of the moving window is large, then $V_{\text{mean}}(t_i)$ will be large. On the other hand, one unusually large value of $V_{(b',k',f')}(t_i)$, that can be due, for example, to the instrument failure, can lead to a false detection. It can be argued that in smaller systems like the IEEE 13, if a fault happens, at least half of the measurements (collected variables) should react to it. Thus, alternatively, we considered the median:

$$(3.3) \quad V_{\text{median}}(t_i) = \text{median}_{(b',k',f')} \{V_{(b',k',f')}(t_i)\}^2.$$

A statistics that combines the characteristics of the mean and median is the truncated mean:

$$(3.4) \quad V_{\text{trunc}}(t_i) = \frac{\sum_{(b',k',f') \in Q} \{V_{(b',k',f')}(t_i)\}^2}{\#(b',k',f') \in Q}.$$

The set Q in (3.4) consists of the indexes (b',k',f') such that the values $(V_{(b',k',f')}(t_i))^2$ are between the $q/2$ th and the $(1 - q/2)$ th quantiles. Here, q can be considered as a tuning parameter. Figure 6 illustrates the concept of truncated mean. In the following, all three methods $V_{\text{trunc}}, V_{\text{mean}}$ and V_{median} are generically denoted as V_{stats} , if the same considerations apply to all of them.

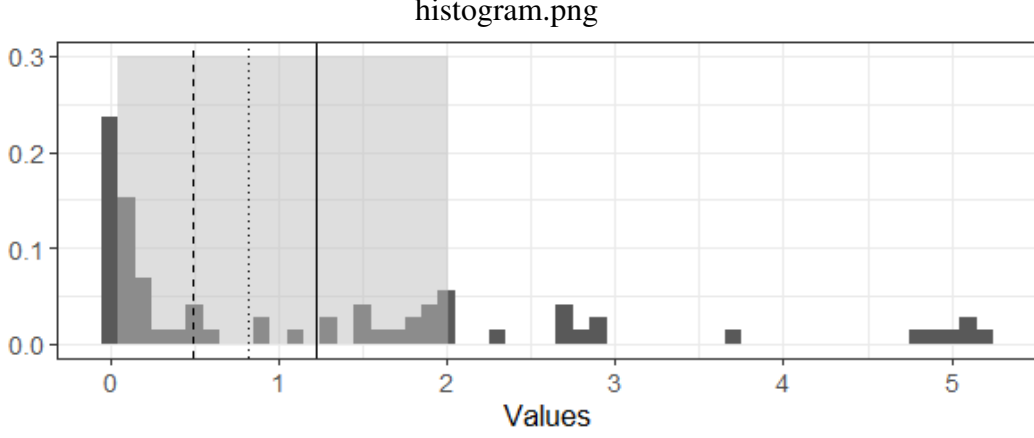


Figure 6: Histogram of $(V_{(b',k',f')}(t_i))^2$ at $t_i = 0.5$ seconds using Mean for f with $l = 0.25$ and $p = 0.95$. The data is from the simulation without a fault and were normalized as described Section 3.3 before computing the $V_{(b',k',f')}(t_i)$ values. The shaded area shows the central 60% of the $(V_{(b',k',f')}(t_i))^2$ values, corresponding to $q = 0.4$. The truncated mean is the average of observations in the shaded area. The solid line represents $V_{\text{mean}}(t_i)$, the dashed line $V_{\text{median}}(t_i)$, and the dotted line $V_{\text{trunc}}(t_i)$.

It is natural to set the value t_i at which $V(t_i)$ exceeds a suitable threshold τ as the estimated time of the fault. In principle, a slightly earlier time could be used, e.g. $(1 - p)l$ seconds earlier, but this is a minor point and a delay of a small fraction of a second may be inconsequential. One could shift the detection time even more because there is a delay of about 0.25 s in the reaction of the test grid to a fault. Such a delay may however be different in a different grid and would generally be unknown. Adjustments of this type are easy to add to our procedure and do not affect the correct and false detection rates. Thus, we define the time of the fault as

$$(3.5) \quad t_f = \min \{t_j : V_{\text{stats}}(t_j) > \tau(t_j)\},$$

where $\tau(t_j)$ is a threshold that is adjusted in real time. The methodology for the derivation of such a threshold is described in Section 3.4. First, we explain the normalization of the data

streams in Section 3.3.

The performance of detection procedures based on different choices of $V_{\text{stats}}(t_i)$ as well as different options for $f(X)_{(b',k',f')}(t_i)$ is discussed in Section 4.

3.3 Normalization

The method described in Subsection 3.2 compares the two parts of the moving window (Figure 5) by evaluating V_{stats} . Thus for each variable included in the data, e.g. phase A voltage RMS, three-phase current RMS, etc., we compute separate $V_{(b',k',f')}(t_i)$ at time point t_i . Notice that each $V_{(b',k',f')}(t_i)$ has an equal weight in the formula for computing $V_{\text{stats}}(t_i)$. Figure 2 shows that different variables have different ranges, averages and variances. Thus, without further adjustments, our proposed methodology would favor some variables merely based on units of measurement and other irrelevant scale factors. Additionally, if units of measurements or other aspects of data collection are changed, this can lead to completely different different performance of our method. For example, instead of measuring a current using the RMS, one could decide using the peak to peak range, but leave voltage and frequency measures unadjusted, possibly getting very different results in V_{stats} . Using amperes instead of kiloamperes would basically eliminate all variables except the current. At this point, we do not know which variables are important for fault detection, and we see that they all react to a fault, so it is judicious to give them equal weight in some suitable sense.

To ensure that each bus, type of variable (voltage, current, and frequency), and phase has an equal weight in the equations of Section 3.2, we use normalization. Traditionally, normalization uses an estimate of standard deviation and of the mean to calculate a Z score. In the case of streaming data, one cannot take the mean and standard deviation estimates of the whole data. Moreover, the potential for a fault implies that measurements over a period considered fault-free should be used. There is no perfect and widely accepted solution. After initial exploration, we propose using the following moving window technique.

We have observations $X_{(b',k',f')}(t_i)$ at time point t_i . We compute the sample mean and the sample standard deviation over the time period $t_i - l \leq t_j < t_i - (1 - p)l$ and denote them as $m(X_{(b',k',f')}(t_i))$ and $SD(X_{(b',k',f')}(t_i))$. The interval over which these statistics are computed is the same as the first interval in (3.2). The normalized value $Z_{(b',k',f')}(t_i)$ at the point t_i is calculated as:

$$(3.6) \quad Z_{(b',k',f')}(t_i) = \frac{X_{(b',k',f')}(t_i) - m(X_{(b',k',f')}(t_i))}{SD(X_{(b',k',f')}(t_i))}.$$

The normalized values $Z_{(b',k',f')}(t_i)$ depend on the choice of p and l . Figures 7 contain examples of normalization applied to Bus 632 voltage trajectories with $l = 0.25$ s and $p = 0.9$. *In the methodology of fault detection, we use the $Z_{b',k',f'}$ as input (instead of the $X_{b',k',f'}$).*

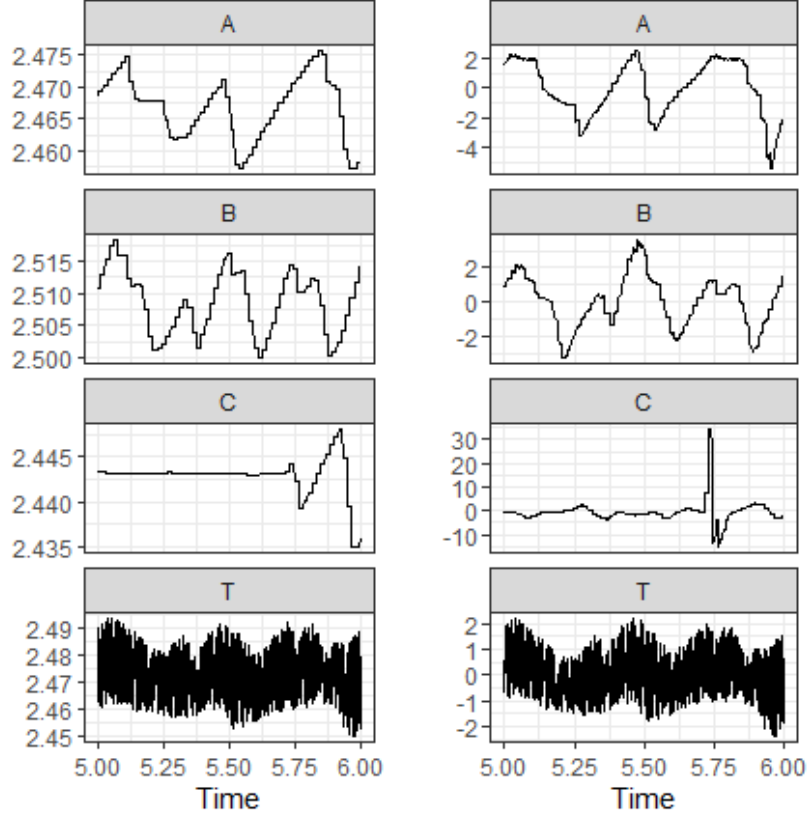


Figure 7: Illustration of the normalization procedure applied to Bus 632 voltage data (no fault). The right four panels contain the regularized data (original data with irregularity removed) and the left four panels show the corresponding normalized values using $l = 0.25$ second and $p = 0.9$. Note the different vertical scales between the left and right columns.

3.4 Threshold determination methodology

The performance of the method described in Section 3.2 depends on how well we can acquire the dynamic threshold τ . We want to determine the τ that balances two criteria: 1) if there is no fault, false alarms should be rare, 2) if there is a fault, it should be detected with a large probability and the delay of detection (time of detection - time of fault) should be small. A larger τ ensures that criterion 1) is met, and a smaller τ ensures that criterion 2) is met. There is no guidance on how to balance these two criteria for power grid streaming data. In this

section, we explain our approach to the determination of τ . Our method relies on several tuning parameters, and it is a priori not clear which of them will work best for our data. The winners are determined after the application of our approach in Section 4.1.

First, we explored how different tuning parameters affect the $V_{\text{stats}}(t_i)$ values. The properties we are looking for are small variance and moderate percentage change between $V_{\text{stats}}(t_i - 1)$ and $V_{\text{stats}}(t_i)$ if there is no fault. Such properties would lead to a robust threshold calculation based on a moving window. The tuning parameters are:

- the length of the moving window l ,
- the proportion parameter p ,
- in case of V_{trunc} - cut off parameter q .

Also, we separately discuss the differences between the $V_{\text{trunc}}, V_{\text{mean}}, V_{\text{median}}$ values. (The median appears in two different contexts: in Equation (3.3) and in the definition of the statistic $f(X)_{(b',k',f')}(t_i) = \text{median}(X_{(b',k',f')}(t_i))$.) For each combination of tuning parameters, we explore for different statistics - Mean, SQMean, Range, Median, cf. Table 1. The properties we are looking for are small variance and moderate percentage change difference between $V_{\text{stats}}(t_{i-1})$ and $V_{\text{stats}}(t_i)$ if there is no fault at time point t_i . On the other hand, some variability in V_{stats} should be expected, as otherwise, the detector would not respond to changes and the detection of a fault would be practically impossible.

While many threshold formulas are reasonable, to narrow down our choices, we considered 3 different ways to calculate τ . To display the formulas, we first introduce the interval

$$I_1(t_i) = \{t_j : t_i - l \leq t_j \leq t_{i-1}\}.$$

The above interval resembles the interval considered in the $V_{\text{stats}}(t_i)$ calculations, but it does not contain the final time point t_i ; if a fault happens, its effects should be noticed comparing current $V_{\text{stats}}(t_i)$ value to a threshold based on previous values. We considered:

1. $\tau_1(t_i) = 3\max_{t_j \in I_1(t_i)} V_{\text{stats}}(t_j)$;
2. $\tau_2(t_i) = \max_{t_j \in I_1(t_i)} V_{\text{stats}}(t_j) + 3\text{range}_{t_j \in I_1(t_i)} V_{\text{stats}}(t_j)$;
3. $\tau_3(t_i) = \max_{t_j \in I_1(t_i)} V_{\text{stats}}(t_j) + 3\text{SD}_{t_j \in I_1(t_i)} V_{\text{stats}}(t_j)$.

Using the coefficient 3 is motivated heuristically by the 99.7% quantile of the normal distribution. In the following, all three thresholds τ_1, τ_2, τ_3 are referred to as τ . After choosing tuning parameters, V_{stats} , and a procedure for computing τ , the fault is signaled at the first time point t_j such that $V_{\text{stats}}(t_j) > \tau(t_j)$, as expressed in equation (3.5).

4 Results

In Section 4.1, we explore the performance of the methodology derived in Section 3 assuming all data streams are available. Based on the observed performance, we select the final form of the method. In Section 4.2, we investigate to what extent our method is robust to the elimination of certain data streams or buses at which they are measured. The first consideration may be viewed as a further tuning of our method because there may be sufficient information only in one of two data stream, and it may not be necessary to use all streams. This is however not clear a priori. The second consideration is very practical because if all buses react to a fault, it may be enough to place measurement devices only at some of them. Again, the answer is not clear a priori and the question must be investigated.

We use the following quantities for the assessment:

F_1 - Fraction of simulations in which a fault is detected over the first 10 seconds. This quantity is similar to the size of a statistical significance test or type I error.

F_2 - Fraction of simulations in which a fault is detected over the whole 15 seconds. This quantity is similar to power under an alternative or 1 minus type II error.

In statistical significance tests, F_1 equal to 5 or 1 percent would generally be acceptable. Since in power grids a false alarm may be expensive to investigate, we may target a different

value of F_1 . Even $F_1 = 0$ might be reasonable. We proceed analogously as in the traditional Neyman–Pearson paradigm. We first determine method parameters that give satisfactory F_1 . From those, we select those that give the best F_2 .

4.1 Detection based on complete data

In accordance with the plan outlined above, we first applied our method to a simulation that does not contain a fault. We tested different combinations of tuning parameters and different choices of V_{stats} and f . We noticed that different values l do not have a large impact on F_1 . One should not choose the value that is too small, as it increases variance in V_{stats} and leads to false detections. The value we propose to use is $l = 0.25$ s, but one can choose relatively close values and get basically identical results. Regarding the selection of p , our statistical experiments showed that p should be close to 1. We have observed better properties of V_{stats} when the first interval of the moving window is longer, and the second interval is relatively short. The value we propose to use is $p = 0.95$, but one can choose relatively close values, like 0.9, and similar results. Notice that with $l = 0.25$ and $p = 0.95$, the effective interval used to detect the faults has the length of $(1 - p)l = 0.0125$ s. The explorations regarding V_{stats} revealed that using V_{mean} leads to many false-detection cases. Random variability in the data combined with the sensitivity of the standard error to large and small values leads to high volatility in V_{mean} values, thus resulting high percentage change from $V_{\text{stats}}(t_{i-1})$ to $V_{\text{stats}}(t_i)$. On the other hand, V_{median} solves the aforementioned problems, but introduces some other issues; it is too conservative, as it is calculated using the middle of input values. It is basically not sensitive to a change in 50% of the largest input values. This method had little variation and values proved to be insensitive even if there are some changes in the data. This leads to undetected faults. The best results were obtained using V_{trunc} . This statistic proved to be less volatile than V_{mean} , but more flexible than V_{median} . The value we propose to use is $q = 0.1$ (90% of the "middle" data). To summarize, we

recommend

$$(4.1) \quad l = 0.25 \text{ s}, p = 0.95, \text{ and } V_{\text{stats}} = V_{\text{trunc}} \text{ with } q = 0.1.$$

All tables that follow are produced using these settings.

The above settings produced the best overall results for the faults we considered. With additional knowledge of the type of fault, it is possible that slightly different settings might be optimal.

Next, for each proposed threshold τ , and each statistic f in the Table 1, we evaluated how many times the detector was triggered over the course of the simulation without a fault (Table 2). The best results (0 false detections) were achieved by using Mean and Range for $f(X_{(b',k',f')})$ and either τ_1 or τ_2 . We observed no false detections in either of these four cases. However, even in the worst cases we observed only 21 false alarms out of potential 17,500 time points where they could be signaled. As note above, in power grids false alarms are expensive, and we have found methods that basically do not produce them.

Table 2: Counts of false detections over the simulation with no fault using settings (4.1). The last column shows the count of false detections over the whole length of the interval (35s) which contains $35 \cdot 500 = 17,500$ regularized time points.

f	τ	Count	f	τ	Count
Mean	τ_1	0	Range	τ_1	0
Mean	τ_2	0	Range	τ_2	0
Mean	τ_3	4	Range	τ_3	5
Median	τ_1	3	SQMean	τ_1	3
Median	τ_2	3	SQMean	τ_2	2
Median	τ_3	21	SQMean	τ_3	21

We now show how the methods with the best parameters perform over the simulations with faults (55 simulations). The results are presented in Table 3. Notice that three of four combinations showed perfect results in terms of fault detection ($F_1 = 0$ and $F_2 = 1$). The combination of τ_2 and Mean gave results of $F_2 = 0.98$ (in one of the simulations, the fault were missed). The fractions F_1 of false detections were computed using the first 10 seconds because we know that there were no faults in the initial 10 s.

Table 3: Fault detection evaluation over 55 simulations with a fault using setting (4.1). Results are presented for the combinations of f and τ that showed the best results in Table 2.

f	τ	F_1	F_2
Mean	τ_1	0	1.00
Mean	τ_2	0	0.98
Range	τ_1	0	1.00
Range	τ_2	0	1.00

Allowing for uncertainty associated with using a database of 55 faults, our results show that any of the four combinations in Table 3 can be used. For additional certainty, One can choose to use these four combinations simultaneously. Additionally, we tested the timing of the fault detection to see if there exist any differences between these four combinations. The results show that τ_1 and τ_2 have no effect on the detection timing and are identical for both f choices. **With additional information about the type of fault, e.g. its impedance or path, it might be possible to recommend an optimal τ most suitable for the specific faults.** In 29% of the simulations, Range led to a faster fault detection than Mean, but only by 0.002s. In the other 71% simulations, both statistics gave identical time of the fault.

4.2 Detection with partial data streams

In this Section, we explore how robust our methodology is to the reduction of the available data streams. We consider only the best parameters settings, those defined by (4.1) and Table 3. Recall that the data streams are indexed by the triples (b', k', f') , where b' is the bus number, k' is the variable (voltage, current, frequency) and f' is the phase (A, B, C, 3 phase). There are a large number of ways in which data streams could be restricted. To provide information in some systematic way, we restrict each of the three coordinates separately and investigate what the impact of such a data reduction is. In Section 4.2.1, we consider the practically most important setting of measurement devices placed only at some buses. In Sections 4.2.2 and 4.2.3 we investigate, respectively, what happens if only some variables or some phases are used. Information of this type is also relevant because it might, for example, be the case that it is enough to use only voltage. Since the data streams we use have not been explored from such angles, the answers are not a priori clear. As a byproduct of the investigations in this section, we may be able to give preference to some of the methods in Table 3.

4.2.1 Limited buses

In Table 3 data from all buses were used. In this section, we investigate what happens if data from subsets of cardinality K of the set $\{650, 632, 634, 671, 675, 680\}$ are used. If $K = 6$, there is only one subset, if $K = 5$, there are 6 subsets, if $K = 4$, there are 15 subsets, etc. (number of combinations). For example if $K = 2$, a possible combination is $\{650, 632\}$, and we have 15 combinations in total for $K = 2$. Thus for $K = 2$, we can compute F_1 and F_2 using $15 \cdot 55 = 825$ cases. The results are displayed in Table 4. Notice that for smaller K , the F_1 values are higher and F_2 values are smaller. With $K \geq 3$, the F_1 stays at 0, while F_2 varies from 0.97 to 0.99. For low K values Mean gives better results in terms of F_1 than Range. For $K \geq 3$, Range gives better results than Mean with $F_1 = 0$ and larger values of F_2 . *The main conclusion is that to ensure no false alarms, it is enough to use any three of the six buses, but to ensure that each fault is*

detected all six buses must be used.

Table 4: Faults detection evaluation using complete data streams from K buses.

f	τ	K	F_1	F_2	K	F_1	F_2
Mean	τ_1	1	0.07	0.93	4	0.00	0.97
Mean	τ_2	1	0.00	0.88	4	0.00	0.96
Range	τ_1	1	0.08	0.95	4	0.00	0.98
Range	τ_2	1	0.12	0.95	4	0.00	0.98
Mean	τ_1	2	0.00	0.94	5	0.00	0.98
Mean	τ_2	2	0.00	0.92	5	0.00	0.98
Range	τ_1	2	0.12	0.96	5	0.00	0.99
Range	τ_2	2	0.22	0.96	5	0.00	0.99
Mean	τ_1	3	0.00	0.95	6	0.00	1.00
Mean	τ_2	3	0.00	0.94	6	0.00	0.98
Range	τ_1	3	0.00	0.97	6	0.00	1.00
Range	τ_2	3	0.00	0.97	6	0.00	1.00

4.2.2 Limited variables

We now explore if it is necessary to consider all three variables. Table 5 shows the results of our procedure based on only one of the three available variables. We see that *using only current gives perfect results*, even slightly better than using all three variables, cf. Table 3.

Table 5: Performance of procedures based on single variables

f	τ	Variable	F_1	F_2
Mean	τ_1	Frequency	0.00	0.96
Mean	τ_2	Frequency	0.00	0.93
Range	τ_1	Frequency	0.00	0.96
Range	τ_2	Frequency	0.00	0.96
Mean	τ_1	Current	0.00	1.00
Mean	τ_2	Current	0.00	1.00
Range	τ_1	Current	0.00	1.00
Range	τ_2	Current	0.00	1.00
Mean	τ_1	Voltage	0.00	0.91
Mean	τ_2	Voltage	0.00	0.87
Range	τ_1	Voltage	0.02	1.00
Range	τ_2	Voltage	0.04	1.00

4.2.3 Limited phases

We finally explore how our procedure performs if only one phase measurement is used. The phases are A, B, C and 3 (three-phase measurement). The last case revers to a single value derived from all three phases, for example the sum of currents flowing in phases A, B, C at a given time. The results displayed in Table 6 are not clear-cut. The performance is perfect for $f = \text{Mean}$ and $\text{phase} = 3$, but replacing $f = \text{Mean}$ with $f = \text{Range}$ gives a lot of false alarms. Using a truly single phase, A, B or C gives practically no false alarms but can miss up to 10% of faults.

Table 6: Performance of procedures based on single phase measurements. Phase 3 refers to the simultaneous measurement of all three phases reported as a single number.

f	τ	Phase	F_1	F_2	Phase	F_1	F_2
Mean	τ_1	A	0.00	0.91	C	0.00	0.91
Mean	τ_2	A	0.00	0.91	C	0.00	0.91
Range	τ_1	A	0.00	0.93	C	0.04	0.95
Range	τ_2	A	0.00	0.93	C	0.04	0.95
Mean	τ_1	B	0.00	0.93	3	0.00	1.00
Mean	τ_2	B	0.00	0.91	3	0.00	1.00
Range	τ_1	B	0.00	0.93	3	0.38	1.00
Range	τ_2	B	0.00	0.93	3	0.24	1.00

5 Summary and an algorithm

Algorithm 1 Detection procedure using $f = \text{Mean}$, threshold τ_1 and 3 phase current

Input: Trajectories $X_{(b',k',f')}$ with $k' = \text{current}$ and $f' = 3$ phase

Output: Fault = 1 or 0 (1 if fault detected); t_f (time of fault)

- 1: **Initialization**
 - 2: Set $l = 0.25$ s, $p = 0.95$, and $V_{\text{stats}} = V_{\text{trunc}}$ (3.4) with $q = 0.1, d = 1/500$.
 - 3: **while** Fault= 0 **do**
 - 4: Set $t_i := t_i + d$
 - 5: Apply regularization to each $X_{(b',k',f')}(t)$, $t \in [t - 2l, t_i]$ (Section 3.1)
 - 6: Find $Z_{(b',k',f')}(t_j)$ (3.6) for $t_j \in [t_i - l, t_i]$ using regularized $X_{(b',k',f')}(t_j)$
 - 7: Find $V_{\text{trunc}}(t_j)$ (3.4) for $t_j \in [t_i - l, t_i]$ using $Z_{(b',k',f')}(t_j)$ and $f = \text{Mean}$
 - 8: Calculate $\tau_1(t_i) = 3\max_{t_j \in I_1(t_i)} V_{\text{trunc}}(t_j)$ (Section 3.4)
 - 9: **if** $V_{\text{trunc}}(t_i) > \tau_1(t_i)$ **then**
 - 10: Set Fault= 1 and $t_f := t_i$
 - 11: **end if**
 - 12: **end while**
-

We have proposed a methodology for fault detection in a small power distribution system based on a suitably developed moving window change-point analysis technique. We investigated the properties of our technique and determined optimal parameter and other settings, including data stream regularization and normalization as well as dynamic threshold selection. Investigations reported in Section 4, indicate that three phase current is the most important variable. In fact, a simple version of our procedure that uses only three phase current at all six buses produces perfect results; there are no false alarms and all faults are detected. This is true for both $f = \text{Mean}$ and $f = \text{Range}$ and for thresholds τ_1 and τ_2 . Other options also produce perfect or nearly perfect results for our data set of 55 faults, showing that the general approach is sound. Algorithm 1 summarizes the whole procedure with $f = \text{Mean}$, threshold τ_1 and three phase current. It is possible that some false alarms and missed faults might occur if a grid with a different topology is used, but our methodology shows how to find a nearly perfect algorithm for any grid.

Our research proposes a general, data-driven, statistical approach to instantaneous and correct detection of faults in a subgrid of a distribution system that is based on high-frequency measurements, like those generated by PMUs. This approach looks ahead to the increasing penetration of renewable energy sources that generate more random variability and bidirectional power flows. Our methodology is obviously not a definite engineering solution, but it shows a potential of data-driven, statistical approaches to fault detection and develops a scalable paradigm.

Extensions of this work might provide further useful insights. More complex grid topologies could be explored. Fault behavior can vary based on the fault path impedance and the source impedance. For this work, we did not vary the source impedance and we did not vary the fault impedance either. Events such as a direct on line (DOL) motor start can cause under voltage for more than a second. Learning on a more extensive data base of faults and various transient but normal events can refine our approach.

Funding Details

This research was partially supported by the following grants: NSF DMS– 2123761, NSF DMS–1914882, NSF DMS–1923142. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding is provided by the U. S. DoE Office of Energy Efficiency and Renewable Energy via the Energy Systems Integration Facility. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Disclosure statement

The authors report there are no competing interests to declare.

Data availability statement

Data used in this research are available at

<https://github.com/MantautasRimkus/TOWARDREALTIMEFAULTDETECTION>.

References

- Aminikhanghahi, S., Wang, T. and Cook, D. (2019). Real-time change point detection with application to smart home time series data. *IEEE Transactions on Knowledge and Data Engineering*, **31**, 1010–1023.
- Ardakanian, O., Yuan, Y., Dobbe, R., von Meier, A., Low, S. H. and Tomlin, C. J. (2016). Event detection and localization in distribution grids with phasor measurement units. *CoRR*.
- Basseville, M., Nikifirov, I. V. and Tartakovsky, A. (2012). *Sequential Analysis: Hypothesis Testing and Change–Point detection*. Chapman & Hall/CRC.
- Brodsky, B. E. (2017). *Change-Point Analysis in Nonstationary Stochastic Models*. CRC Press.
- Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change Point Problems*. Kluwer.

- Chen, H., Jiang, B., Chen, W. and Yi, H. (2018). Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains. *IEEE Trans. Ind. Electron.*, **66**, 4716–4725.
- Chen, H., Yi, H., Jiang, B., Zhang, K. and Chen, Z. (2019). Data-driven detection of hot spots in photovoltaic energy systems. *IEEE Trans. Syst. Man Cybern. Syst.*, **49**, 1731–1738.
- Chen, J. and Gupta, A. K. (2011). *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Birkhäuser.
- Chen, T., Hill, D. J. and Wang, C. (2020). Distributed fast fault diagnosis for multimachine power systems via deterministic learning. *IEEE Trans. Ind. Electron.*, **67**, 4152–4162.
- Chen, Y., Banerjee, T., Domnguez-Garca, A. and Veeravalli, V. (2016). Quickest line outage detection and identification. *IEEE Transactions on Power Systems*, **31**, 749–758.
- Chow, J. H., Chakraborty, A., Arcak, M., Bhargava, B. and Salazar, A. (2007). Synchronized phasor data based energy function analysis of dominant power transfer paths in large power systems. *IEEE Transactions on Power Systems*, **22**, 727–734.
- Cui, M., Wang, J., Tan, J., Florita, A. R. and Zhang, Y. (2019). A novel event detection method using pmu data with high precision. *IEEE Transactions on Power Systems*, **34**, 454–466.
- Estrada Gómez, A. M., Li, D. and Paynabar, K. (2022). An adaptive sampling strategy for online monitoring and diagnosis of high-dimensional streaming data. *Technometrics*, **64**, 253–269.
- Gholami, A., Srivastava, A. and Panday, S. (2019). Data-driven failure diagnosis in transmission protection system with multiple events and data anomalies. *Journal of Modern Power Systems and Clean Energy*, **7**, 767–778.
- Ghosal, A. and Conti, M. (2019). Key management systems for smart grid advanced metering infrastructure: A survey. *IEEE Commun. Surv. Tutor.*, **21**, 2831–2848.
- Glover, J. D., Sarma, M. S., Overbye, T. and Birchfield, A. (2022). *Power System Analysis and Design*, 7th edn. Cengage.
- Gustafsson, F. (2000). *Adaptive filtering and change detection*. Wiley.
- Hannon, C., Deka, D., Jin, D., Vuffray, M. and Lokhov, A. Y. (2021). Real-time anomaly detection and classification in streaming PMU data. In *2021 IEEE Madrid PowerTech*.
- Horváth, L., Kokoszka, P. and Wang, S. (2021). Monitoring for a change point in a sequence of distributions. *Annals of Statistics*, **49**, 2271–2291.
- Hu, B., She, J. and Yokoyama, R. (2013). Hierarchical fault diagnosis for power systems based on equivalent-input-disturbance approach. *IEEE Transactions on Industrial Electronics*, **60**, 3529–3538.
- Jiang, J., Chuang, C., Wang, Y., Hung, C., Wang, J., Lee, C. and Hsiao, Y. (2011). A hybrid framework for fault detection, classification, and location, Part I: Concept, structure, and methodology. *IEEE Trans. Power Delivery*, **26**, 1988–1998.
- Kaci, A., Kamwa, I., Dessaint, L. and Guillon, S. (2014). Synchrophasor data baselining and mining for online monitoring of dynamic security limits. *IEEE Transactions on Power Systems*, **29**, 2681–2695.
- Khushwant, R., Hojatpanahand, F., Ajaei, F. B. and Grolinger, K. (2021). Deep learning for high-impedance fault detection: Convolutional autoencoders. *Energies*, **14**, 3623.

- Kim, D., Chun, T. Y., Yoon, S., Lee, G. and Shin, Y. (2017). Wavelet-based event detection method using pmu data. *IEEE Transactions on Smart Grid*, **8**, 1154–1162.
- Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information theory*, **44**, 2917–2929.
- Li, W., Deka, D., Chertkov, M. and Wang, M. (2019). Real-time faulted line localization and PMU placement in power systems through convolutional neural networks. *IEEE Transactions on Power Systems*, **34**, 4640–4651.
- Lorden, G. (1971). Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, **42**, 1897–1908.
- Moghaddass, R. and Wang, J. (2017). A hierarchical framework for Smart Grid anomaly detection using large-scale smart meter data. *IEEE Trans. Smart Grid*, **9**, 5820–5830.
- Nguyen, D., Barella, R., Wallace, S. A., Zhao, X. and Liang, X. (2015). Smart grid line event classification using supervised learning over pmu data streams. In *2015 Sixth International Green and Sustainable Computing Conference (IGSC)*, pp. 1–8.
- Onaolapo, A. K., Akindeji, K. and Adetiba, E. (2019). Simulation experiments for faults location in smart distribution networks using ieee 13 node test feeder and artificial neural network. In *Journal of Physics Conference Series 1378:032021*.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–105.
- PES, IEEE (2020). IEEE 13 node test feeder document. <http://sites.ieee.org/pestestfeeders/resources/>. Accessed March 30, 2021.
- Rafferty, M., Liu, X., Lavery, D. M. and McLoone, S. (2016). Real-time multiple event detection and classification using moving window pca. *IEEE Transactions on Smart Grid*, **7**, 2537–2548.
- Tartakovsky, A., Nikifirov, I. V. and Basseville, M. (2015). *Sequential Analysis: Hypothesis Testing and Change-Point detection*. CRC Press.
- Tripathi, S. and De, S. (2018). Dynamic prediction of powerline frequency for wide area monitoring and control. *IEEE Trans. Ind. Inf.*, **14**, 2837–2846.
- Xie, L., Chen, Y. and Kumar, P.R. (2014). Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis. *IEEE Transactions on Power Systems*, **29**, 2784–2794.
- Yin, S., Li, X., Gao, H. and Kaynak, O. (2014). Data-based techniques focused on modern industry: An overview. *IEEE Trans. Ind. Electron.*, **62**, 657–667.
- Zhang, Y., Wang, R. and Shao, X. (2021). Adaptive inference for change points in high-dimensional data. *Journal of the American Statistical Association*, **000**,; Published online: 27 Apr 2021.
- Zhou, M., Wang, Y., Srivastava, A. K., Wu, Y. and Banerjee, P. (2019). Ensemble-based algorithm for synchrophasor data anomaly detection. *IEEE Transactions on Smart Grid*, **10**, 2979–2988.