IET Smart Grid

Research Article

Data-driven online distributed disturbance location for large-scale power grids

Zekun Yang1, Yu Chen1, Ning Zhou1, Aleksey Polunchenko2, Yilu Liu3

1Department of Electrical and Computer Engineering, Binghamton University, Binghamton, NY 13902, USA
2Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA
3Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA

E-mail: ychen@binghamton.edu

Abstract: Timely detecting disturbances and locating their sources are critical to the reliable operation of power grids. This capability enables operators to effectively diagnose disturbances over wide areas and earns time for remedial reactions. In this study, a travelling-wave based scheme, namely data-driven online distributed disturbance location (DODDL), is proposed to quickly detect disturbances and determine their geographic location in large-scale power grids when the grids’ topology is not available. The proposed DODDL scheme consists of two function blocks: (i) a singular spectrum analysis-based change-point detection method, which can quickly detect disturbances and determine their arrival time at distributed sensors, and (ii) a novel temporal scanning algorithm, which can accurately determine the geographic location of the disturbance source point. Utilising field measurement data sets recorded by the frequency disturbance recorders from the frequency monitoring network, it is shown that the DODDL scheme is not only quicker and more robust to grid non-homogeneity than existing approaches, but also can capture and locate more subtle and concealed disturbances.

1 Introduction

Timely and efficiently detecting and locating disturbances are critical to the reliability and security of power grid operations. Once a fault occurs in power systems, operators need to detect the disturbance and locate the hypocenter before launching appropriate remedial countermeasures. Although many fine-grained methods have been developed [1, 2] to detect and locate disturbances, they may not be reliable all the way. For example, in the Northeast blackout of 2003 [3], a software ‘bug’ in a computer at First Energy prevented its energy management system from detecting and reporting important transmission line tripping events to the operators. The operators were unaware of the malfunction, the transmission line tripping events, and the needs of re-dispatching power generation. Consequently, a transmission line tripping event led to widespread power outage. If a backup disturbance detection and location system had worked in parallel with the current detection system and increased the situational awareness (SAW) over a wide geographic area, the risk of cascading failures like the Northeast blackout of 2003 could have been significantly reduced.

Several considerations to building the backup disturbance detection and location system are listed as follows:

(i) Wide coverage. To obtain SAW in wide areas, the method should be able to monitor disturbance over the entire grid, and give a warning if the local fault detection fails.
(ii) Lightweight. The method should not rely on a high volume of sensors over the whole system.
(iii) Independence to system models. Due to the administrative constraints, it is very difficult to build a detailed operational model in real time for a large interconnected power grid. Therefore, the method should function properly even without a detailed model.

Frequency monitoring network (FNET) is a platform for building a backup disturbance detection and location system. FNET is an Internet-based real-time, wide-area frequency monitoring network proposed in 2000 [4], which measures and records frequencies of typical office outlet at 120 V voltage level using GPS-synchronised clocks. The objective is to create a low cost and rapidly deployable wide-area frequency measurement system with high accuracy and minimal installation cost. The frequency disturbance recorders (FDRs), as a major part of FNET, calculate frequency using algorithms of phasor analysis and signal sampling techniques. The wide area measurement data obtained from different FDRs has revealed valuable information about power system disturbances. Researchers have proposed to leverage the FDR data to locate disturbance sources based on the principle that disturbances travel as an electromechanical wave in power grid network [4]. When a disturbance happens, the contingency frequency perturbation will propagate in the entire power grid through transmission lines. The sampled frequency measurement data at sensors carries information about disturbance characteristics, such as disturbance categories and intensities. In general, a disturbance can lead to frequency changes with the same properties at different terminals, such as frequency decrease following a generator tripping or frequency increases following a load shedding. However, the wide-area frequency measurements have a time difference of arrival (TDOA) of frequency perturbation by the distance between the measurement locations and disturbance source [5]. Consequently, it is a critical step to accurately detect the arrival time of the disturbance signals at sensors that are deployed at different geographic locations [6–8] and then the disturbance source can be located. Recently, researchers have leveraged the novel edge/fog computing paradigm to tackle the early anomaly detection problem in wide-area phasor measurement unit (PMU) network [9], and the results are pretty encouraging because the PMU fog can effectively reduce the data flow end-to-end delay without sacrificing data completeness.

In this paper, a data-driven online distributed disturbance location (DODDL) scheme is proposed to detect disturbances and locate their sources. The major contribution lies in two folds: (i) a singular spectrum analysis (SSA) [10] based change-point detection method is proposed to quickly detect disturbances and determine their arrival time; and (ii) a novel temporal scanning (TeSc) algorithm is proposed to identify the geographic location of the disturbance source. Utilising the FNET data, we have validated the DODDL scheme with satisfying results. Not only had it instantly detected the change in frequency with high accuracy, but it also located the source precisely.
The rest of the paper is organised as follows. A brief review of related work is presented in Section 2. Section 3 describes the proposed DODDL scheme. Section 4 discusses the rationale and design of the SSA based change-point detection method to quickly detect a disturbance and determine its arrival time at distributed sensors, along with conceptually-proof simulation experimental results. Section 5 introduces the TeSc algorithm, which can determine the geographic location of disturbance sources. Section 6 discusses the computational complexity issue. Finally, Section 7 concludes the paper with some discussions and our on-going efforts.

2 Related work

2.1 Disturbance arrival time detection

Quick and accurate detection of disturbances is the prerequisite of successful disturbance location. Thanks to the development of synchronised system measurement and communication technologies, many data analysis and machine learning algorithms have been introduced to detect disturbance. These methods can be roughly divided into model-based methods and non-model based methods. The non-model based methods directly catch the instantaneous changes from the observed waveform, induced by the abrupt transition when the disturbance occurs. For example, principal component analysis has been used on PMU data to extract features and then obtain the disturbance information like type and intensity [11]. In [12], support vector machine and wavelet analysis are adopted to identify the faulty-section and faulty-half. In another work [13], the FDRs data was described as a hidden Markov model, and then disturbances were identified using matching pursuit decomposition. All of these machine learning and data analysis methods show significant potential in disturbance detection.

Under the same framework of TDOA using FDR data, current disturbance arrival time methods are simple and straightforward. A method has been proposed that employed a morphological filter to remove the noise before extracting wave-front arrival time [7]. The arrival time was defined as the first sample that presents several consecutive and large-enough differences. Another disturbance detection method is the event start time (EST) method [6], which also pre-processed the noises and spikes with an adaptive median filter. After de-noising, the EST method detects the EST as the last point with the most ‘normal’ value in a section of frequency data series. These methods suffer from not only single detection capability but also non-trivial detection delay. Therefore, they are good candidates for postmortem analysis, but not suitable for online detection.

In this paper, to achieve the goal of online disturbance detection, we considered the SSA, a non-model time-series analysis technique. Because the measurement data is synchronised, it is appropriate to consider it as a time series. Intuitively, arrival time detection can be considered as a change detection problem, which identifies the change point of the time series. And similar time-series analysis techniques have been tentatively used for other applications in the smart grid. Although the time-series analysis exists in the voltage quality monitoring [14, 15], anomaly detection [16], false data injection attack detection [17] and so on, to date there is not reported research that has tried to leverage the SSA method in power system data analysis.

2.2 Geolocation methods

In this subsection, we focus the discussion on the travelling-wave based disturbance location methods. The major challenge of this kind of methods is the high heterogeneity in the network. For instance, the propagation velocities are not the same in the entire power grid. Most of the reported approaches tried to reduce the location error resulted from the heterogeneity through the additional information of the network topology, or just using single-end data. For example, the approach proposed in [13] detected the impact of the fault on each recorded bus and combined with topology information to give a fault contour map. In [12], the authors firstly identified the transmission configuration (either cable or line) and then used the corresponding wave travelling velocity to estimate the distance from the single end. These methods either require the topology information or only cover limited local grid network.

To locate disturbances in wide areas without using power grid topology, we used geographic distance instead of actual transmission line distance. The reported method using the TDOA to identify the geographic location of a disturbance source is defined by the equation below [8, 18, 19],

\[(x_i - x_h)^2 + (y_i - y_h)^2 = V^2(t_i - t_h)^2\] (1)

where \(V\) is the mean velocity of electromagnetic waves propagating in the power grid, \(t_i\) is the disturbance arrival time at a frequency measurement unit \(i\), and \(t_h\) is the actual disturbance start time. \((x_i, y_i)\) is the Cartesian coordinates of the \(i\)th frequency measurement unit and \((x_h, y_h)\) represents the disturbance hypocentral location.

The equation set for estimating the disturbance source location can be solved using Newton’s method or the least squares (LS) method [8, 18]. The LS method linearises the model by making a difference between neighbour equations. One concern is that the root close to the real situation is often eliminated because of the order reduction. Newton’s method is a gradient approaching method. However, because of the limited, unknown number, only a subset of the data items is needed. The method itself cannot take advantage of the abundant collected measurement information. The location estimation errors under these methods were at the level of 100 km. Besides the TDOA based methods, there are other approaches using triangulation location. Following the principle of local minimum, researchers have tried to triangulate disturbance location with only three measurement units, and they successfully detected in which nation a disturbance happened [16]. Researchers have estimated the location of disturbance sources by studying the frequency oscillations [19]; the smallest estimation error is still >160 km.

Because the major estimation errors result from the unequal velocities and simplification of electrical distance to geographic distance, we propose a location estimation method to mitigate these two issues by leveraging redundant measurements.

3 DODDL: system level overview

The DODDL scheme is a travelling-wave based method that aims at detecting disturbances and determining their geographic locations in real time using measurement data. A frequency perturbation will be observed at monitors distributed at different geographic locations in the identical morphology. Fig. 1 shows the frequency responses recorded by six different FDRs after a generator tripping event. The figure illustrates that FDRs at different locations observed different disturbance arrival time. The proposed disturbance location scheme is based on the geometrical triangulation algorithm which leverages the measurement of TDOA. According to the event triangulation principle, the DODDL scheme consists of two function blocks as shown in Fig. 2. Table 1 summarises the symbols to be used in the paper.

3.1 Arrival time estimation

An SSA-based change-point detection method is implemented to quickly detect the arrival of a disturbance at distributed sensing units. This detection method calculates the disturbance arrival time without any additional pre-processing steps for de-noising the data sequence. Consequently, it achieves much higher efficiency and is applicable to detect different categories of disturbances.

3.2 Disturbance source location

Leveraging the disturbance arrival time reported by each sensor, a novel TeSc algorithm is introduced to accurately identify the geographic location of the disturbance source point. There are two main challenges to accurately locate disturbances. Firstly, the propagation speed of the electromechanical waves may vary from
100 to 1000 km/s among different paths in the power grid due to variant transmission technologies. Secondly, it is the absolute geological distance that is used to simplify the calculation. However, in the real world, the electromechanical waves propagate along power grid transmission lines. To mitigate the negative impacts of the module simplification, the proposed TeSc algorithm takes advantage of the abundant FDRs data.

### Table 1 Symbol table

| Symbol | SSA based change-point detection method symbol | SSA based change-point detection method symbol | SSA based change-point detection method symbol |
|--------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| $f$    | frequency measurement time series            | $y$                                          | $y$-axis of FDR i’s location                   |
| $f_{\text{test}}$ | target time series of frequency measurement     | $t_i$                                | disturbance arrival time at FDR i |
| $F_{\text{target}}$ | target matrix                                   | $V$                                      | propagation velocity                        |
| $N$    | length of target matrix                       | $x_i$                                    | x-axis of disturbance location               |
| $K$    | column of target matrix                        | $y_i$                                    | y-axis of disturbance location               |
| $\lambda_i$ | eigenvalue i of target matrix                  | $t_i$                                    | disturbance start time                      |
| $U_i$  | right eigenvector i of target matrix           | $\text{Lat}_i$                      | latitude of FDR i’s location                |
| $\mathcal{D}_i$ | l-dimensional subspace                      | $\text{Lon}_i$                      | longitude of FDR i’s location               |
| $l$    | selected eigenvector number                    | $\Delta t$                                 | scanning step size                         |
| $F_{\text{target},l}$ | target matrix in l-dimensional subspace              | $d_{\text{min}}$                 | largest distance in grid                    |
| $F_{\text{test},l}$ | test time series of frequency measurement                  | $t_{\text{min}}$                       | lower bound of disturbance start time       |
| $s$    | interval of target matrix and test matrix      | $t_{\text{max}}$                       | upper bound of disturbance start time       |
| $\rho$ | column of test matrix                           | $\lambda_i$                                  | evaluation of disturbance start time        |
| $\delta$ | sum of squared Euclidean distance             | $\lambda_i$                                  | evaluation of disturbance y-axis            |
4 Estimation of disturbance arrival time

One fundamental step for disturbance location is that each sensor should accurately and quickly estimate the time point at which a disturbance starts to impact. With synchronisation, it is feasible to compare the disturbance record data marked with time stamps. Therefore, a time series is naturally suitable for corresponding data analysis technique. In the normal state, the collected frequency measurement time series will operate stably. When a disturbance happens and affects the local condition, the frequency measurement time series will present different patterns. Consequently, the detection of disturbance arrival problem is formulated as a change point detection problem, and the SSA algorithm is adopted to solve it.

4.1 Change-point detection method based on the SSA

Change-point detection [20–22] is concerned with the design and analysis of procedures for on-the-go detection of possible changes in the characteristics of a running (random) process. The time instance at which the state of the process changes is referred to as the change-point, which is not known in advance. The SSA theory, proposed in 1986 by Broomhead and King [10], has shown its significant ability in a wide field of practical applications, such as finding data structure, extracting periodic patterns and complex trends, smoothing and change point detection [17, 23]. Actually, the SSA method has shown good potential in the detection of the sustained oscillation in the power grid [24]. Applied by a branch of applications, the SSA approach features lower false alarm rate and higher detection accuracy. Practically, matrix operation-based SSA conversion is more regular and easier for implementation than hash-function-based cardinality estimator. Results of SSA procedure have already outlined the potential point of change, which simplifies the following operation, to benefit the overall process [21, 22, 25].

The main idea of SSA-based change-point detection method is as follows. Assume \( f_1, f_2, \ldots, f_t, \ldots \) is an ongoing frequency measurement time series. At the time point of \( t \), a section of previous data \( f_1, f_2, \ldots, f_t \) will provide a template of normal data operation, while a section of current data \( f_{t+1}, f_{t+2}, \ldots, f_{t+s} \) is supposed to describe the situation at the current time point. These two sections will be transformed into two trajectory matrices to keep the completed information. Meanwhile, the matrix of previous data, i.e. the target matrix, will be projected to an \( l \)-dimensional subspace \( \mathcal{B}_l \) by the singular value decomposition (SVD). After that, the distance between the mentioned subspace and the matrix of the current situation, i.e. test matrix, will be calculated. A change point will be declared if the distance between the previous situation and the current situation is larger than a pre-determined threshold.

This SVD-based change-point detection method is resistant to noise. When projected by the SVD, the target trajectory matrix can be divided into different components. Each component can be represented by one or a set of eigenvalues along with their corresponding right eigenvectors and left eigenvectors. The signal containing disturbance information usually is either slowly-change trend or simple oscillation. So the signal can be described by several low-rank components with larger eigenvalues than others, while the noise is complex and its power is distributed to several components. Therefore, even with low signal-to-noise ratio, extracting the primary components (components with large eigenvalues) can effectively retain the disturbance information and eliminate noise.

The detection method involves the following three steps.

Assuming \( N, M, s \) and \( \alpha \) be constant integers, \( 1 < M \leq N/2, \ 0 < s \leq \alpha, N \leq o \) and \( N = 1, 2, \ldots \).

**Step 1: Construction of the target matrix.** Define \( f_{\text{target}}^{(0)} \) as the target time series starting from time point \( t \) with length \( N_r \):

\[
f_{\text{target}}^{(0)} = (f_{t+1}, f_{t+2}, \ldots, f_{t+N_r})
\]

Map the vector \( f_{\text{target}}^{(0)} \) into an \( M \times K \) matrix \( F_{\text{target}}^{(0)} \) as follows:

\[
F_{\text{target}}^{(0)} = [F_1, F_2, \ldots, F_K]
\]

where \( F_i = (f_{t+1+i}, f_{t+2+i}, \ldots, f_{t+N_i}) \) and \( i = 1, \ldots, K \) and \( K = N - M + 1 \).

Apply the SVD operation on \( F_{\text{target}}^{(0)} \). Assume \( R_{\text{target}} = F_{\text{target}}^{(0)} U_{\text{target}} \), and get \( M \) eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_M \) (in decreasing order) and the corresponding eigenvectors \( U_1, U_2, \ldots, U_M \), which are biorthogonal, i.e. \( (U_i, U_j) = 0 \), when \( i \neq j \). Then, let's choose the first \( l \) eigenvalues, calculate their corresponding rank-one biorthogonal matrices \( F_i \) and get l-dimensional target matrix \( F_{\text{target}}^{(l)} \), which is the sum of \( F_i \).

**Step 2: Construction of the test matrix.** Similar to Step 1, construct the test time series

\[
f_{\text{test}}^{(0)} = (f_{t+1+s}, f_{t+2+s}, \ldots, f_{t+N_r})
\]

as follows

\[
F_{\text{test}}^{(0)} = [F_1, F_2, \ldots, F_P]
\]

where \( F_i = (f_{t+1+i}, \ldots, f_{t+N_i+s}) \) and \( i = 1, \ldots, P \). So the size of the test matrix \( F_{\text{test}}^{(0)} \) is \( M \times P \), and \( o = s + P + M \).

**Step 3: Computing distance.** Compute the difference between the two matrices measured via calculating the sum of squared Euclidean distance

\[
\mathcal{D} = \sum_{i = s, s + 1}^{P} \left( F_{\text{test}}^{(0)} - F_{\text{test}}^{(0)} U U^{T} F_{\text{test}}^{(0)} \right)^T,
\]

where \( U \) is composed by the chosen eigenvectors \( U_1, U_2, \ldots, U_M \) of \( R_{\text{test}} \) from step 1, and \( t \) indicates the \( i \)th column of the matrix. \( \mathcal{D} \) is used to detect the change points through a specific threshold.

4.2 Online estimation of disturbance arrival time

The SSA-based change-point detection method is developed following the above principle. As shown in Algorithm 1 (see Fig. 3), it is straightforward to detect the disturbance arrival time. In a power grid, frequency data measured by each monitoring unit is collected as a one-dimensional time series. When an event occurs, such as a generator tripping or a transmission line tripping, a frequency perturbation is introduced. A generator tripping will result in decreasing of frequency, while a transmission line tripping will create a short-term oscillation of frequency and then the frequency will come back to the original value. Thus the detection of the change-point of the frequency data sequence identifies the arrival time of the disturbance.

In general, the longer the target/test window is, the more details can be extracted from the working pattern of that particular point. However, the tradeoff includes both higher computing complexity and longer processing time. Considering the length of the time series and the sizes of both target matrix and test matrix, a relatively small value was selected to control the computational complexity to a tolerable level with little sacrifice of the accuracy. In this paper, the parameters were experimentally set as: \( N = 36, M = 18, s = 36, o = 66 \) and \( l = 6 \).

4.3 Experimental validation

Because the actual disturbance arrival time is not known in the real-world FDRs data, simulation data were used to validate the proposed method. A generator tripping case and a transmission line tripping case were reported in this subsection. In our experimental
time detected by the two algorithms at the 15 generators (i.e. G02, ..., G16). When the noise level grew, the start time estimated by the SSA was consistently close to the actual start time 10.1 s, with some ignorable increasing. In contrast, the start time estimated by the EST algorithm suffered obvious longer time delays. This degradation was due to two factors: (i) the effect of the disturbance was initially weakened by the median filter; and (ii) in the EST the disturbance start time was defined as the last intersection with the benchmark, while the vibration after the generator tripping made the last intersection away from the yielding point.

Fig. 5a also compares the detection rate of the SSA and EST algorithms. Given the generator tripping start time at 10.1 s. The detection rate $R_d$ in Fig. 5 is defined as

$$R_d = \frac{\text{No. of events detected by the algorithm}}{\text{No. of total events}}$$

Generally, both the EST and SSA algorithms were able to successfully detect the generator tripping. The relatively lower detection rate of the EST may be caused by its median filter and its approximate treatment of the yielding point, which may lead to detection before the disturbance.

### 4.3.2 Scenario 2: transmission line tripping

Another tested power system event is a transmission line tripping, which generally speaking does not result in frequency changes as significant as the generator tripping does. Instead, the frequency profile will experience a transitory oscillation with smaller magnitudes than those in generator tripping. Then, the frequency will return to its original value as it was before the disturbance. In the simulation, we tripped the transmission line between bus 47 and bus 48. The fault was cleared at the near end at 10.15 s and at the remote end at 10.20 s. And the total simulation window length was 20 s.

The general performance evaluations actually concur with the intuition. Fig. 5b shows the average estimated start time of the disturbance and the detection rate for each algorithm. The EST algorithm suffered from a huge delay of detection and a much lower detection rate with higher noise levels. One reason lies in the denoising scheme of the EST, which mistakenly removed the head of the disturbance. Also, in the transmission line tripping, the frequency returned to the original value after the disturbance, which invalidates the assumption used by the EST detection.

The SSA outperformed the EST in both the generator tripping and line tripping scenarios. In the generator tripping simulation, although both algorithms were able to detect the disturbance with a high detection rate, the SSA was outstanding with shorter delays. In the transmission line tripping case, the SSA maintained a capability of quick and accurate detection. Meanwhile the EST gradually lost its efficacy of detection as the noise levels got higher.

For the generator tripping event shown in Fig. 1, the disturbance arrival time at FDRs was estimated by the SSA. The results shown in Table 2 match with the real data.

### 5 Location of disturbance sources

Once the arrival time of the disturbance has been accurately determined at distributed FDRs, the second phase is to find out the actual beginning time $t_b$ and the geographic location $(x_0, y_0)$ of the disturbance source based on the location of FDRs.

#### 5.1 Propagation model analysis

Before building the disturbance propagation model, two assumptions are made as follows:

- The frequency perturbation propagates to every measurement unit (i.e. FDR) at different locations as a continuum. It spreads over the entire electric grid at a uniform speed.
- Although the perturbation propagates along electric transmission lines, geological distance is used instead of electric distance.

Thus the disturbance propagation model is written as...
where $V$ is the mean velocity of electromechanical waves, which is treated as a constant according to the first assumption. $t_i$ is the disturbance arrival time at the frequency measurement unit $i$, while $t_h$ represents the actual disturbance or event start time. $(x_i, y_i)$ is the Cartesian coordinates of the $i$th frequency measurement unit, and $(x_h, y_h)$ represents the Cartesian coordinates of the disturbance hypocentral location.

As the location information of the frequency measurement unit was provided in the form of latitude with longitude, a transform is needed before the computation. The Greenwich coordinate of each unit (Lat$_i$, Lon$_i$) can be converted into Cartesian coordinate $(x_i, y_i)$ using the following equation:

$$
egin{align*}
    x_i &= \cos(Lat_i) \times \text{Lon}_i \times 111,300 \\
    y_i &= \text{Lat}_i \times 111,300
\end{align*}
$$

where 111,300 m is the approximate distance between two locations with the same longitudes and 1° difference in the latitudes.

The assumptions have simplified the propagation model but also introduced some systemic errors. One issue is that the actual propagation velocities are measured in the range from 100 to 1000 km/s along different paths. Meanwhile, earlier studies have shown that for a given path, the propagation velocity depends on time-varying system conditions [27]. Also, the difference between the real electric distance and the geometric distance can be unignorably large. However, addressing these issues requires considering the actual propagation velocity and the real-world electric network configuration and topology, which will lead to very high computational costs. It is worthy of considering how to reduce the negative effect of these noise by leveraging the abundant FDRs data.

Another issue that calls for attention is the overdetermined equation set. Normally the number of the measurement units is larger than the number of the unknowns ($x_h, y_h, t_h$ and $V$ at most), which means the result can be attained using only part of the data set. Meanwhile, it is critical to select data points that are less affected by the underlying noise. If the selected data points were with a high level of noise, it is possible that the measurement units far away from the disturbance source will mistakenly detect the disturbance earlier than the ones that actually are located closer. Then the location result will point to the opposite direction.

Considering the measurement samples were more than the unknown variables, we firstly provided some estimates about the disturbance location and then used all the measurement data to evaluate and determine the estimation of the most possible. Although the individual measurement data had propagation error and modification error, according to the law of large number, with a large enough sample size, the estimation will be sufficiently close to the actual situation. Following this rationale, the proposed TeSc algorithm firstly solved the equation set with partial selected data set and start time assumption and got a series of estimations. Then the estimates were evaluated by all measurement samples using the mean squared error (MSE).

### 5.2 Temporal scanning algorithm

Based on the following observations, the TeSc algorithm is proposed to locate the source of the detected disturbance.

1. Although the accurate start time of a disturbance is unknown, a reasonable range of the actual time point can be estimated.
2. Similarly, the propagation velocity of the electromechanical wave in the power grid can be assumed reasonably well.
3. Intuitively, the noises and interferences cannot make the signals propagate faster but only slower.
4. The actual location $(x_h, y_h)$ and start time $t_h$ should be able to fit in (2) with small error.

![Fig. 5](http://creativecommons.org/licenses/by/3.0/)  
Detected arrival time and detection rate  
(a) Generator tripping, (b) Transmission line tripping
Accordingly, the TeSc algorithm consists of four steps:

1. Assume a reasonable propagation velocity value $V$ and the range of the disturbance start time $t_0$. 
2. The TeSc scans the entire estimated time range, from the earliest time point to the latest time point with a step size $\Delta t$. In each step, a start time $t_0$ is used to calculate the corresponding $[(x_o, y_o)]_{t_0}$ with the assumed $V$ using the locations of two measurement units and the reported disturbance arrival times.
3. For each calculated $[(x_o, y_o)]_{t_0}$, the MSE is adopted to estimate the error from the measurement units that are not used in step 2.
4. Finally, $[(x_o, y_o)]_{t_0}$ with the minimal MSE is selected as the $(x_o, y_o)$ and $t_0$.
5. Below is a detailed explanation of the operations.

### 5.2.1 Value range of the disturbance start time

As the disturbance event must have started earlier than the time point when the signal is detected, the upper bound of start time was selected as the earliest of the detected arrival times, $t_{min}$. The lower bound $t_{max}$ was defined in the following equation:

$$t_{max} = t_{max} - \frac{d_{max}}{V} \quad (4)$$

where $d_{max}$ is the largest distance between two nodes in the power grid. It is the extreme case when the measurement unit and the disturbance source are at the two furthest ends of the network, such that the unit is the last one to detect the disturbance. Hence, in any situation, the disturbance should happen between $t_{min}$ and $t_{max}$.

### 5.2.2 Dataset selection algorithms

Given a particular value of event start time $t_0$ and a reasonably assumed velocity $V$, there are only two unknowns. As such, two data points are sufficient. More than enough data is collected by measurement units to locate the source of a disturbance, but arbitrarily picking two may end up with a huge error. The fewer errors or noises, the more accurate the spatiotemporal location will be. Once a $[(x_o, y_o)]_{t_0}$ is solved, the rest of the redundant measurement units are used to calculate the MSE of this $[(x_o, y_o)]_{t_0}$ point. It is not practical to expect an MSE of zero, but the minimum MSE does indicate the best estimates. For this purpose, the following two-step data selection method is proposed. On the one hand, the entire map is divided into regions with equal size along the longitude and latitude, and an average disturbance arrival time detected by the measurement units in each region was calculated. Intuitively, an average of noises which lead to inconsistency of the detection sequence concerning the distance between the measurement units and the disturbance source point were also used. By applying the average, the influence of the random noises on the individual data items can be effectively reduced.

On the other hand, to solve (2) to find the geographic location, data points with the minimal disturbance arrival times were selected from two regions in which the average arrival times were minimal. The main source of the error is the assumption that the disturbance signals propagate along a straight line, while they actually follow the grid transmission lines. Hence, it is reasonable to assume that the earlier a sensor detects a disturbance, the less effect from noise or distance error is involved. Unconditionally choosing the single point that is the earliest detection of the disturbance will increase the risk of false detection. Therefore, it is proposed to select the earliest detected points from two regions with minimal average arrival time.

The above-mentioned rationale had been validated by a generator tripping case. In Fig. 6a, the green dots represent the 53 FDRs; and three monitoring units that detected the disturbance are marked as $U_A$, $U_B$, and $U_C$. We divided the regions by each $10^\circ$ in longitude and $5^\circ$ in latitude and calculated the average arrival time, show them in Fig. 6b.

### 5.2.3 MSE calculation

The MSE assesses the quality of an estimator (i.e. a mathematical function mapping a sample of data to a parameter of the population from which the data is sampled) or a predictor (i.e. a function mapping arbitrary inputs to a sample of values of some random variable). The definition of the MSE is

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

where $\hat{Y}$ is a vector for $n$ predictions, and $Y$ is the vector of observed values corresponding to the inputs to the function that generates the predictions.

---

Table 2 Location and arrival time for FDRs in Case 1

| FDRs       | Latitude, °N | Longitude, °W | Arrival time, s |
|------------|--------------|---------------|-----------------|
| UsFlPlantcity | 28.0183      | 82.1130       | 17.1000         |
| UsILChicago  | 41.8781      | 87.6297       | 19.8000         |
| UsMaBoston   | 42.3583      | 71.0602       | 19.6000         |
| UsNyRpi      | 42.7315      | 73.6839       | 19.8000         |
| UsMliHoughton| 47.1218      | 88.5690       | 19.5000         |
| UsVaRichmond | 37.5407      | 77.4360       | 18.9000         |

---

Fig. 6 Disturbance arrival geospatial distribution

(a) FDRs distribution, (b) Regional average arrival time

A reasonable guess is that the source of the disturbance should be located between 70° and 90° in longitude and 35° and 45° in latitude. Although the calculation using $U_A$ claims that the disturbance is located in between 90° and 100° in longitude and 40° and 45° in latitude, considering the regional large average arrival time of 18.22 s, the detection of $U_A$ probably was not accurate. Meanwhile, the second and the third points, $U_B$ and $U_C$, belong to the two regions with minimal average arrival times, which made them more trustworthy. The calculation result also verified that higher accuracy was achieved by using the data at $U_B$ and $U_C$ rather than selecting the first two FDRs at $U_A$ and $U_B$ to solve the equation set.
Let’s define the error function of one set of the estimator $(\tilde{x}_n, \tilde{y}_n, t_n)$, such that

$$f_{x, y, t}(x_n, y_n) = (x_n - \tilde{x}_n)^2 + (y_n - \tilde{y}_n)^2 - V^2(t_n - \tilde{t}_n)^2$$

Ideally, if there is not any noise or error, it is expected that $f_{x, y, t}(x_n, y_n, t_n) = 0$.

Suppose the number of selected data items is $n$, and the estimated event start time $\hat{t}_n = t_{sa} + m\Delta t$, where

$$m \in (0, 1, \ldots, (t_{sa} - t_{sa})/\Delta t)$$

Therefore, the MSE can be calculated as

$$MSE = \frac{1}{n} \sum_{i=1}^{n} f_{x, y, t}(x_i, y_i)$$

### 5.2.4 TeSc algorithm

The operations of the TeSc algorithm are illustrated by the pseudo-code in the following algorithm description table (see Fig. 7).

### 5.3 Validation

We validated the proposed TeSc algorithm using the generator tripping event which took place on 14 November 2013. As shown in Fig. 1, among all FDR units, significant perturbation had been observed by six sets of FDR frequency signals, which were selected for the validation analysis. The SSA based change-point detection method and TeSc location algorithm were evaluated. Generally speaking, the SSA-based change-point detection method should work all the time for each local measurement to monitor if there is any change, while the TeSc location algorithm only runs when a disturbance is detected.

### 6 Computation Complexity

The computational complexity is an important issue to the algorithms for real-time applications. In this section, the computational complexity of the SSA-based change-point detection method and TeSc location algorithm are evaluated. Generally speaking, the SSA-based change-point detection method should work all the time for each local measurement to monitor if there is any change, while the TeSc location algorithm only runs when a disturbance is detected.

#### 6.1 Computational complexity of the SSA-based change-point detection method

In the calculation of the SSA procedure for a single time point, the matrix algebra computation accounts for a very high share. So the matrix size parameters $N$, $M$, $s$, $o$ and subspace dimension $l$ determine the computational complexity. The complexity of each operation in the SSA procedure is shown in Table 3. In practice, the target matrix usually is a square matrix, i.e. $M = N/2 > l$, and

![Fig. 7: TeSc algorithm](http://creativecommons.org/licenses/by/3.0/)

![Fig. 8: Temporal scanning of solution](http://creativecommons.org/licenses/by/3.0/)

**Require:**

- Measurement units number $n$
- The $i$th measurement units’ location $(x_i, y_i)$, $i \in \{1, 2, \cdots, n\}$, and its event arrival time $t_i$
- The disturbance propagation speed $V$
- Time step $\Delta t$
- Event start time upper bounder $t_H$
- Event start time lower bounder $t_L$

**Ensure:**

- $t_H = t_{sa} + d_{max}$
- $t_L = t_{sa} - d_{max}$
- Select two data points $(x_1, y_1, t_1), (x_2, y_2, t_2)$

1. For each $k \in (0, 1, \cdots, m)$ do
2. $t_h = t_L + \Delta t \times k$;
3. Solve $(x_h, y_h)$:
4. For $j \in (3, 4, \cdots, n)$ do
5. $error = (x_j - x_h)^2 + (y_j - y_h)^2 - V^2(t_j - t_h)^2$;
6. $MSE(k) = MSE(k) + error$;
7. End for
8. End for

Find $\min\{MSE(k)\}$

Find $\min\{MSE(k)\}$ corresponding $(x_h, y_h, t_h) = (x_h, y_h, t_h)$

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/).
measurement units. The SSA-based change-point detection method was proposed to the field of power grid monitoring and evaluated through intensive experimental study. The simulation experiments have shown the robustness of the proposed DODDL method in a noisy environment and its high sensitivity to the changes in the power grid, such as line tripping. Quick and accurate detection of the arrival time of disturbance signal at multiple frequency measurement units inspired the TeSc algorithm, which can estimate the geographic location of the source of a disturbance. Using real-world data, the experimental results verified the effectiveness and correctness of the DODDL scheme. Furthermore, the DODDL scheme achieved high accuracy, and effectively avoided the invalid solution brought by the quadratic equation set, and thus, is robust to the errors of propagation velocity.

The SSA method and TeSc algorithm are non-parameter, model-less algorithms. It is the advantage of the proposal as it is insensitive to the changes in the actual grid properties. However, it is also the main constraint that prevents the algorithms from leveraging the characteristics of the grid to achieve better performance. Accordingly, our on-going work mainly consists of the following components:

(i) The effects of a fine-grained electromechanical wave propagation model and more detailed network topology of the power grid will be considered in the TeSc algorithm for more accurate location.
(ii) New digital signal processing techniques will be explored to denoise the data without losing useful information.
(iii) New signal processing algorithms will be developed to detect, distinguish and locate multiple anomalies and disturbances that happen within short time intervals.

8 References

[1] Gao, L., Parhi, K.K.: ‘Models for architectural power and power grid noise analysis on data bus’, J. VLSI Signal Process.Syst. Signal, Image Video Technol., 2006, 44, (1-2), pp. 25–46
[2] Kezunovic, M.: ‘Smart fault location for smart grids’, IEEE Trans. Smart Grid, 2011, 2, (1), pp. 11–22
[3] Group, N.S.: ‘Technical analysis of the August 14, 2003, blackout: what happened, why, and what did we learn’, report to the NERC Board of Trustees, 2004
[4] Zhong, Z., Xu, C., Billian, B.J., et al.: ‘Power system frequency monitoring network (fnet) implementation’, IEEE Trans. Power Syst., 2005, 20, (4), pp. 1914–1921
[5] Zhang, Y., Markham, P., Xia, T., et al.: ‘Wide-area frequency monitoring network (fnet) architecture and applications’, IEEE Trans. Smart Grid, 2010, 1, (2), pp. 159–167
[6] Li, W., Tang, J., Ma, J., et al.: ‘Online detection of start time and location for hypocenter in North America power grid’, IEEE Trans. Smart Grid, 2010, 1, (3), pp. 253–260
[7] Ma, J., Li, J., Wang, Z., et al.: ‘Wave-front arrival time analysis using wide-area frequency measurements’. 2009 IEEE Power & Energy Society General Meeting (PES09), Calgary, Canada, 2009, pp. 1–5
[8] Xia, T., Zhang, H., Gardner, R., et al.: ‘Wide-area frequency based event location estimation’. 2007 IEEE Power Engineering Society General Meeting, Tampa, USA, 2007, pp. 1–7
[9] Yang, Z., Chen, N., Chen, Y., et al.: ‘A novel pmu fog based early anomaly detection for an efficient wide area pmu network’. 2nd IEEE Int. Conf. on Fog and Edge Computing (ICFEC), Washington DC, USA, 2018, pp. 1–10
[10] Broomhead, D.S., King, G.P.: ‘Extracting qualitative dynamics from experimental data’, Phys. D. Nonlinear Phenom., 1986, 20, (2–3), pp. 217–236
[11] Xia, T., Overbye, T.: ‘Real-time event detection and feature extraction using pmu measurement data’. 2015 IEEE Int. Conf. on Smart Grid Communications (SmartGridComm), Miami, USA, 2015, pp. 265–270
[12] Livani, H., Evrenosoglu, C.Y.: ‘A machine learning and wavelet-based fault location method for hybrid transmission lines’, IEEE Trans. Smart Grid, 2014, 5, (1), pp. 51–59
[13] Jiang, H., Zhang, J.J., Gao, W., et al.: ‘Fault detection, identification, and location in smart grid based on data-driven computational methods’, IEEE Trans. Smart Grid, 2014, 5, (6), pp. 2947–2956
[14] He, X., Pun, M.-O., Kuo, C.J.: ‘Change-point detection-based power quality monitoring in smart grids’, APSIPA Trans. Signal Inf. Process., 2015, 4, pp. 1–9
[15] Li, S., Wang, X.: ‘Cooperative change detection for voltage quality monitoring in smart grids’, IEEE Trans. Inf. Forensics Sec., 2016, 11, (1), pp. 86–99
[16] Sterpu, S., Rossignol, S., Prestat, B.: ‘Detection and location of generation trips in large transmission grids’. 2007 IEEE Power Engineering Society General Meeting, Tampa, USA, 2007, pp. 1–6
[17] Moskvina, V., Zhigljavsky, A.: ‘An algorithm based on singular spectrum analysis for change-point detection’, Commun. Stat. -Simul. Comput., 2003, 32, (2), pp. 319–352

[18] Gardner, R.M., Wang, J.K., Liu, Y.: ‘Power system event location analysis using wide-area measurements’. 2006 IEEE Power Engineering Society General Meeting, Atlanta, USA, 2006, pp. 7

[19] Zao, J., Baldwin, M., Zhang, H., et al.: ‘Use of frequency oscillations to improve event location estimation in power systems’. 2007 IEEE Power Engineering Society General Meeting, Tampa, USA, 2007, pp. 1–7

[20] Basseville, M., Nikiforov, I.V.: ‘Detection of abrupt changes: theory and application’, vol. 104 (Prentice Hall, Englewood Cliffs, 1993)

[21] Polunchenko, A.S., Sokolov, G., Du, W.: ‘Quickest change-point detection: A bird’s eye view’, arXiv preprint arXiv:1310.3285, 2013

[22] Poor, H.V., Hadjiliadis, O.: ‘Quickest detection’, vol. 40 (Cambridge University Press, Cambridge, 2009)

[23] Golyandina, N., Nekrutkin, V., Zhigljavsky, A.A.: ‘Analysis of time series structure: SSA and related techniques’ (Chapman and Hall/CRC, New York, 2001)

[24] Yang, Z., Ghorbaniparvar, M., Zhou, N., et al.: ‘Enhancing sustained oscillation detection by data pre-processing using ssa’. 2017 North American Power Symp. (NAPS), Morgantown, USA, 2017, pp. 1–6

[25] Yang, Z., Zhou, N., Polunchenko, A., et al.: ‘Quick online detection of start time of disturbance in power grid’. IEEE GlobeCom, San Diego, USA, 2015, pp. 1–6

[26] Rogers, G.: ‘Power system oscillations’ (Springer Science & Business Media, New York, 2012)

[27] Rudez, U., Mihalic, R.: ‘A method of detecting the time of arrival for an electromechanical wave in large power systems’. 2013 IEEE Grenoble PowerTech (POWERTECH), Grenoble, France, 2013, pp. 1–6