Data Analysis for Real Time Identification of Grid Disruptions

Authors: Varun Chandola, Olufemi Omitaoumu, and Steven J. Fernandez
chandolav@ornl.gov, omitaoumuoa@ornl.gov, fernandezsj@ornl.gov

Abstract: The U.S. electric power system comprises multiple distinct interconnections of generators, high voltage transmission systems, and local distribution systems that maintain a continuous balance between generation and load with impressive levels of efficiency and reliability. This critical infrastructure has served the nation remarkably well, but is likely to see more changes over the next decade than it has seen over the past century. In particular, the widespread deployment of renewable generation, smart-grid controls, energy storage, and new conducting materials will require fundamental changes in grid planning and the way we run the power grid. Two challenges in the realization of the smart grid technology are the ability to visualize the deluge of expected data streams for global situational awareness; as well as the ability to detect disruptive and classify such events from spatially-distributed high-speed power system frequency measurements.

One element of smart grid technology is the installation of a wide-area frequency measurement system on the electric poles in the streets for conditions monitoring of the distribution lines. This would provide frequency measurements about the status of the electric grid and possible information about impending problems before they start compounding and cascading. The ability to monitor the distribution lines is just one facet of proposed smart grid technology. Other elements include the installation of advanced devices such as smart meters, the automation of transmission lines, the integration of renewable energy technologies such as solar and wind, and the advancement of plug-in hybrid electric vehicle technology.

This chapter describes recent advancements in the area of intelligent data analysis for real time detection of disruptive events from power system frequency data collected using an existing internet-based frequency monitoring network (FNET), which is a precursor for future smart-grid systems.

1. Introduction

The U.S. power grid system comprises multiple distinct interconnections of generations, high voltage transmission systems, and local distribution systems that maintain a continuous balance between generation and load with impressive levels of efficiency and reliability. This critical infrastructure is likely to see more changes over the next decade as a result of the proposed modernization of the grid system than it has seen over the past century. In particular, the widespread deployment of renewable generation (centralized and distributed), smart meters, synchrophasors, smart-grid controls, energy storage, and new conducting materials will require fundamental changes in grid planning and the way we run the power grid.

This modernization will lead to an explosive growth in the amount data collected at each level of the power grid system – generation, transmission, distribution, and consumption. The data collected at these levels can be summarized into three classes – field data (collected by the
various devices distributed throughout the system, such as digital recorders); centralized data archives (from monitoring, control, and operation systems, such as supervisory control and data acquisition (SCADA) systems); and data from simulation (carried out in planning or operation environments). At the center of this data collection activity is the ability to process the information from the massive stock of data to support future decision making. It should be noted that the data cannot be left in the hands of field experts alone because of the high latency (for example, as high as 60 data points per second in case of phasor measurement units (PMU)) and high dimensionality of the data sets. Hence, there is a need to develop algorithms capable of synthesizing structures from data. Developing such algorithms, implementing them, and applying them to real problems are the purposes of the so called data mining field.

Data mining is the term for a general approach that is supported to varying degrees by a set of technologies: statistics, visualization, machine learning, and neural networks. Classification, regression, clustering, summarization, dependency modeling, deviation detection, and temporal problems are expected to be solved by data mining tools. Data mining tools will be useful in power grid system because of the characteristics of the collected data: large scale character of power grid (thousands of state variables), temporal (from milliseconds, seconds, minutes, hours, weeks, year) and statistical nature of the data, existence of a discrete (e.g., events such as generator trip or phase changes) and continuous (analog state variables) information mixture, necessity of communication with experts through means of visualization, on-line operation time restriction for fast decision making, and existence of uncertainty (noise, outliers, missing information).

This chapter presents applications of data mining tools to some problems in power grid systems using some of the data described earlier in this section. Section 2 discusses some of the identified research problems in the power grid system. Section 3 describes a methodology for classifying and visualizing frequency data collected using synchrophasors at the distribution level. Section 4 discusses a methodology for detecting and visualizing inter-area oscillatory modes from frequency data. Section 5 describes a spatiotemporal anomaly detection method for fast detection of grid events using a statistically principled change detection technique.

2. Some Research Problems in the Power Grid System

As mentioned earlier, the modernization in the data collection capabilities within the power grid is resulting in an explosive growth in the amount of data collected at each level of the grid system – generation, transmission, distribution, and consumption. As noted earlier, we require new algorithmic approaches as well as parallel formulations to address this data explosion. One of the critical components is the prediction of changes and detection of anomalies. The state-of-the-art algorithms are not suited to handle the demands of streaming data analysis. A recent Department of Energy sponsored workshop on “Mathematics for Analysis of Petascale Data” have identified a few challenges that are important for this proposal: (i) need for events detection algorithms that can scale with the size of data, (ii) need for algorithms that can not only handle multi-dimensional nature of the data, but also model both spatial and temporal dependencies in the data, which, for the most part, are highly nonlinear, (iii) need for algorithms that can operate in an online fashion with streaming data.
As stated above, one element of the modernized power grid system is the installation of a wide-area frequency measurement system on the electric poles in the streets for conditions monitoring of the distribution lines. This would provide frequency measurements that reflect the status of the electric grid and possible information about impending problems before they occur. The timely processing of these frequency data could eliminate impending failures and their subsequent cascading into the entire system. The ability to monitor the distribution lines is just one facet of proposed smart grid technology. Other elements include the installation of advanced devices such as smart meters, the automation of transmission lines, the integration of renewable energy technologies such as solar and wind, and the advancement of plug-in hybrid electric vehicle technology. The overall objective then is to make the electric grid system more robust in view of impending national and global operational challenges.

A wide-area frequency disturbance recorder (FDR) is already in use at both the transmission and distribution levels of the power grid system [1]. These recorders are used to monitor and record the changes in voltage frequency in real time at various locations. The FDRs perform local GPS synchronized frequency measurements and send data to a central server through the internet, and the information management system handles data collection, storage, communication, database operations, and a web service. There are currently more than 50 FDRs deployed around the United States. Each FDR measures the voltage phasor and frequency at the distribution level using a 110-V outlet and streams 10 data points per second, with future models expected to have higher streaming rates. One immediate challenge with the massive amount of data streams collected from FDRs is how to detect and classify an impending failure of the grid from multiple high-speed data streams in real-time while minimizing false alarms and eliminating missed detection, and how to identify and evaluate the impacts of the detected failures.

In the next three sections we describe three different applications of data mining for addressing two electric grid related problems. The first problem deals with identifying a specific type of pattern in the data (pattern discovery) while the second problem deals with identifying disruptive events from grid data. Section 3 addresses the first problem while Sections 4 and 5 address the second problem.

3. Detection and Visualization of Inter-area Oscillatory Modes

Inter-area mode oscillations in power systems are global small-signal stability problems based on the architecture of the system. Given a specific system topology these oscillatory modes result. Each system will have its own characteristic modes which can be excited under a various number of stress conditions. When excited, these modes create oscillations in power, generating unnecessary power flows which serve to introduce added stress to the system. If under damped or left unchecked these oscillations have the potential to incite cascading blackouts.

Due to these hazards several different control schemes have been implemented specifically targeting modes of known oscillation frequencies. A major pitfall of these methods is that they rely on knowing the specific inter-area modal frequencies in advance to design the control scheme. These mode frequencies can be calculated from the system topology but in many cases the system models maybe inaccurate, incomplete or unavailable. Additionally as the system topology changes over time the mode frequencies will change with it. These factors can contribute to an improperly tuned control system.
With the proliferation of high time resolution power system measurement devices it becomes possible to observe these inter-area modes as they occur. As more devices are installed across the system a more complete picture of the different regions is achieved. The increased number of measurement vectors allows for identification of coherent generator groups, and their associated geographic regions.

The work presented in this sub-section outlines the procedures in developing a solution to extract an inter-area mode and determine its amplitude, phase and damping at each measurement point, and it is based on the paper by [8].

### 3.1 Signal Pre-Processing

In order to provide a better fit of the oscillatory content the measurement data needs to be properly conditioned first. An example of measured frequency data is given in Figure 7. It is desired to extract the properties of this oscillation.

The data for this example is drawn from FDR measurements which capture the system response to a generation trip. The resulting system frequency drop is seen in Figure 7 as a sharp decline from about 60.05 Hz to 59.96 Hz. During this drop period a strong oscillation is also observed with power grid systems in Maine oscillating 180 degrees out of phase with systems in North Dakota. This example data set will be used throughout this section to demonstrate the operation of the modal identification procedure.

Although this oscillation is easily observed by visual inspection it is hard to qualify numerically. The main reason for this is the high DC component which dominates the frequency spectra. The inter-area band of 0.1 Hz to 0.8 Hz is completely overshadowed by the lower frequency components. To resolve this problem, the DC component needs to be mathematically removed from the signal. A simple solution to this problem is to subtract the median from the original signal, doing this detrends the data and centers the distribution of data points around zero. Although doing this removes the DC component there are still low frequency components over shadowing the inter-area band. These remaining unwanted components will be further attenuated
by a non-linear band pass filter in a subsequent stage.

At this point the noise content of the input signals will be considered. Observing Figure 7, it is seen that a noticeable amount of noise is still present in the example data. From previous work [13] the nature of noise in the FNET system has been characterized as being Laplacian in nature. It has also been shown that a moving median filter provides an optimal filter for these noise elements. Since the inter-area oscillation components occupy the 0.1 Hz to 0.8 Hz band the selected low pass filter should pass these frequencies with minimal distortion. Thus the low pass filter used to denoise the FDR data for this application should have a break frequency greater than 0.8 Hz and less than 2 Hz. Based on this a moving median filter with a window size of 5 points was selected to satisfy these requirements.

Application of the described detrending method and moving median filter to the example data set yields the filtered signals of Figure 8. When compared to the original signals of Figure 7 the level to which the noise elements have been attenuated becomes obvious. Figure 8 demonstrates very little high frequency noise content and the oscillatory content is much clearer.

![Figure 8: Example Data after Removal of Trend and Median Filtering](image)

The filtering process up to this point has helped to condition the input data and to isolate the oscillation frequencies of interest. Despite this the data signal still contains a large low frequency component which masks the inter-area band. Our interest at this point is to isolate only those frequencies within the inter-area band so that the dominant oscillation mode can be extracted. This was achieved through the implementation of a non-linear band-pass filter. The filter would be primarily based on an Empirical Mode Decomposition (EMD) of the detrended and denoised input signal.

### 3.1.1 Signal Decomposition

Empirical Mode Decomposition is a data driven method that decomposes a signal in a set of Intrinsic Mode Functions (IMF). Each IMF is an oscillatory signal which consists of a subset of frequency components from the original signal. As opposed to Fourier, Wavelet and similar methods, EMD constructs these component signals directly from the data by identifying local extrema and setting envelopes around the signal in an iterative process. A fit is then performed on the local extrema to create an IMF. After the creation of an IMF it is subtracted from the original
signal and the process repeats to identify the next IMF. This identification and removal process continues until the original signal has been completely described by a set of IMFs. The output of the EMD process is the set of IMFs, generally this set is a small number of signals (usually less than 10 for the data considered here) that when summed together completely match the original signal. The EMD algorithm does not explicitly compute oscillation frequencies, amplitudes or phase angles as with other signal decomposition techniques, instead the IMFs are derived directly from the input signal based on its local extrema. A complete mathematical description of the EMD algorithm is beyond the scope of this chapter but can be found in [9-12]. The EMD and associated Hilbert-Huang Transform have also been recently proposed as methods for isolating inter-area modes in power systems [16-17].

Performing an Empirical Mode Decomposition on the Bangor trace of Figure 8 extracts the seven Intrinsic Mode Functions given in Figure 9. The first and second IMFs extracted in this process are given by the blue and green traces of Figure ; these capture the high frequency components and noise of the input signal. The next three IMFs given by the red, cyan, and violet traces capture the middle frequencies present in the signal. Finally the last two IMFs extracted, those represented by the mustard and black lines, define the low frequency components of the input, representing the event drop itself in this case.

![Image of Intrinsic Mode Functions Computed from Bangor, ME Example Data](image)

**Figure 9: IMFs Isolated From Bangor Example Data**

The Fourier Transform of the IMF signals in Figure 9 is given in Figure 10. The frequency variable is plotted on a log scale to better demonstrate the frequencies in the lower range. The individual FFTs have been scaled by their maximum values so that the low frequency components do not dominate the plot. Inspection of Figure confirms that each IMF is capturing one specific band of the original signal. The first IMF extracted, which is plotted in blue is centered around 3 to 4 Hz and each subsequent one picks up successively lower frequency components.
Since the inter-area band is 0.1 Hz to 0.8 Hz it is wished to preserve only those IMFs which have a significant portion of their power in this band and discard the others. The final implementation of this filter computes the IMFs then performs a Fourier Transform of each one. Using the Fourier Transform results the percentage of power within the inter-area band is computed for each IMF. If this percentage exceeds a given threshold the IMF is retained, otherwise it is discarded. The final filter output is the summation of the retained IMFs. Through testing on several data sets it was determined that cutoff frequencies of 0.1 Hz and 0.8 Hz with a power percentage threshold of 0.75 provided the best response. These settings gave the best preservation of the inter-area band while removing most of the other frequency components.

The total input filtering process for this modal identification application consists of three stages: first, the median detrending stage; followed by a moving median filter; and finally the EMD based filter. This process serves to isolate only the frequency components within the inter-area band so that further processing can extract specified modes within this region. Applying this multistage filtering process to the set of example data results in the signals presented in Figure 11. Comparing this plot with that of Figure 7 it is seen that the low frequency trend is completely removed, leaving a zero centered signal. Additionally the higher frequency components and noise have also been removed from the raw data vectors. The oscillation in the inter-area band observed during the event drop is now the most prominent feature of the data, making it easier to extract from a computational viewpoint.

This filtering procedure was found to function well for several types of data. It is demonstrated here on frequency measurement derived from the FDR system but it performs similarly for angle measurements from FDR and Phasor Measurement Unit (PMU) data sets. Given any of these different types of data the filtering process returns a zero centered signal with the isolated inter-area modes similar to that of Figure 11. The ability of this filtering process to work for various different types of data sets makes the final modal extraction procedure compatible with all these forms of input data. The remainder of the modal extraction procedure is tuned to handle data vectors similar to those of Figure 11, thus any data set that the input filter can reduced to that form can be processed.
3.2 Identification and Extraction of Dominant Oscillatory Mode

One oscillation frequency is selected and the properties of that mode will be extracted over time for each measurement vector. For historic data analysis this extraction will be focused around some particular event containing an oscillation. The dominant frequency in this oscillation will be extracted then tracked over time for a larger data set around the event. In order to determine the dominant mode in the event, data from the corresponding time span is extracted. Once the necessary time span is determined the data for each measurement vector is filtered and extracted into a reduced data set. Doing this produces a set of signals that are zero centered and dominated by the oscillatory mode.

With the event data properly conditioned the extraction of the dominant modal frequency can begin. In order to extract this mode a Matrix Pencil [14-15] based analysis was employed across a sliding data window. As the window is progressed across the data an independent Matrix Pencil procedure is performed for each measurement vector and each data window. The results for one of the Matrix Pencil fits on the example data set are given in Figure 12.

Inspection of Figure 12 shows that one of the computed modes, MP result 3, closely matches the oscillation and captures the majority of the original signal. This result is typical of all the tested
cases and is primarily due to the data detrending of the input filter. In all of the observed cases one of the computed modes followed the original signal very closely, with the remaining Matrix Pencil results accounting for the difference. The expression for a modal component is given by Eq. [1] where $A$ is the amplitude, $\alpha$ is the damping, $f$ is the frequency and $\theta$ is the phase angle.

$$\ y = Ae^{\alpha t} \cos(2\pi ft + \theta) \quad [1]$$

The power of this signal is given by Eq. [2]. Note that this equation does not give a true electrical power as it is derived from a frequency signal instead of a current or voltage.

$$\ P_y = y^2 \quad [2]$$

The total energy of Eq. [1] can then be expressed as the summation over time of the power as stated in Eq. [3]. Once again this is not a true physical energy quantity, merely an analogous metric of the data signal.

$$\ E_y = \sum_t P_y \quad [3]$$

Whenever the Matrix Pencil method is performed on a measurement vector over a data window this energy metric is assessed for each resulting mode. The mode with the largest energy value is then selected as the representative mode for that data vector and data window. This process is repeated for each vector and for each data window within the span of the oscillation event. A representative mode is then determined for each iteration, giving many possible modal frequencies.

Each of these candidates represents one oscillatory frequency which prevailed during a specific data window on one of the measurement vectors. The set of candidate modal frequencies can thus be treated as a probability distribution describing the location of the dominant mode. The distribution of candidate modes for the example data set is given in Figure 13. This distribution represents an asymmetrical probability density function. From this data distribution it is wished to determine the most likely value.

![Distribution of Candidate Modal Frequencies for Example](image.png)

Figure 13: Distribution of Candidate Modal Frequencies from Example Data
Assessing the most commonly occurring value in the candidate mode distribution gives an estimate of the most often occurring frequency component. It is this estimate which will be taken as the dominant oscillatory mode during an event. Executing this procedure on the set of example data determined a modal frequency of 0.428 Hz. This value is within the desired inter-area band and when compared with the original data of Figure 7, it is consistent with the observed oscillation frequency.

3.3 Windowing Fit of Dominant Mode to Full Data Set

Now that the dominant oscillatory frequency has been determined it will be applied across the full length of the data set. Once again a moving window will be employed to fit the mode across time. For each measurement vector and window instance the damping, angle and phase are derived which provide a best fit to the given data. The moving window is sized such that it covers approximately one period of the dominant oscillation frequency. The data window then starts at the first data point to be included in the output. It then moves across the full data set shifting one time-step at a time until the end time is reach. For each instance of the window each measurement vector is filtered according to the process of Section 4.1. After filtering the data vectors are resampled to increase the discrete sampling rate. As before this resampling serves to increase the stability and accuracy of the final fit.

Once these conditioning steps have been performed the fit is ready to be executed. In this case it is wished to fit a damped sinusoid of a specified frequency. A least squares fit of a damped sinusoid function was performed. This function is of the form in Eq. [4]. Here the variable fit parameters are the amplitude, $A$, the damping factor, $\alpha$, the phase, $\theta$, and a DC offset, $C$. In Eq. [4], $f_0$ is the dominant modal frequency determined previously.

$$y = Ae^{\alpha t} \cos(2\pi f_0 t + \theta) + C$$  \[4\]

Performing this fit results in an optimal least squares solution to the four fit parameters. This solution gives the amplitude, damping and phase for the oscillatory mode of interest. As this filtering, resampling, and fitting procedure is iterated across time and measurement vectors an evolution of the dominant mode is captured. The amplitude, damping and phase angles are computed with respect to a constant mode frequency. These values may then be tracked both spatially and temporally to gain an understanding of the evolution of this particular mode.

3.3.1 Identification of Coherent Groups

With the mode angle and phase determined for each measurement point over time it is desired to estimate those units that form coherent groups. Given a set of phasors computed from a mode the coherent groups can be identified by clustering the phasors according to their phase angle. As each group will be oscillating against the other group(s) in the system the different groups should demonstrate differing phase angles. The desired cluster centroids in this case become phase angles defining the center of each coherent group.

Figure 14 gives the results of the clustering algorithm for a set of mode phasors drawn from the example data. Here each measurement device yields one phasor. The clustering algorithm proceeds to identify those forming coherent groups by classifying them according to the direction in which they point. In Figure 14, each group is identified by a separate color with the
group centroid given by the dashed line of the same color.

![Figure 14: Phasor Clustering Example](image)

### 3.4 Visualization of Modal Extraction Results

A visualization package was developed to display the results from the preceding modal identification procedure. This visualization is based around a geographic display which gives the mode amplitude and phase at each measurement point. These phasors are placed on the map so that the coherent groups can be observed spatially. This visualization architecture was used to generate movies which run across a given oscillation to present the graphics over time such that the temporal evolution can be observed in real-time. One such movie is represented below by a series of frame captures. This movie was constructed from a data set centered on the example case used throughout this section.

Figure 15 gives a movie frame from before the event and demonstrates a region with very little oscillatory content. The time range highlighted in red is the region for which this frame corresponds. This plot is intended to provide a timeline of the oscillation event under study as well as giving the time point currently being displayed. The main body of the frame of Figure 15 is made up of geographic region spanned by the power system with each mode phasor plotted at the location of its corresponding measurement point. The phasors in this frame are also color coded by the coherent group to which they belong. In the lower left corner of Figure 15 the modal frequency being extracted is given along with the average damping factor seen across the measurement points of the system. To the left of the average damping is a traffic light style indicator. When the damping is greater than zero this light turns red specifying an alarm condition as the oscillation is growing. If the oscillation is lightly damped (-0.1 < α ≤ 0) the light turns yellow. Finally this indicator is green when the oscillation is heavily damped (α < -0.1). The final feature of Figure 15 is a phasor diagram in the lower right-hand corner which plots all of the mode phasors on the same origin so that the clustering and relative magnitudes are better demonstrated. In this diagram the dashed lines represent the centroids of the computed clusters. As indicated by the location of highlighted time span (red moving window), Figure 15 is drawn from a time point before the inciting event. At this stage the system is in steady state and not experiencing an inter-area oscillation. Because of this all of the mode phasors have a very small magnitude and register as nearly zero in the diagrams.
As time progresses to Figure 16, the inter-area oscillation is fully developed. Higher amplitudes are seen by FDRs around the Great Lakes as well as the FDR located in Bangor. The average damping across the system is positive as the oscillation is growing. The clustering algorithm has identified two major coherent groups at this point, the first is the two New England FDRs in red and the second covers the Great Lakes regions in blue. For the most part these two groups are oscillating 180 degrees out of phase.

The inter-area oscillation has grown to its largest magnitude by Figure 17, here nearly all of the measurement points are registering significant amplitude, and the phasors are all observable on the map. Also at this point the oscillation magnitude has begun to decay, indicating damping across the system. The average damping factor is -0.902 indicating that the oscillation is decaying sharply throughout the system. Then, the two primary groups are defined by the Bangor, ME unit and the Bismarck, ND units which are oscillating nearly 180 degrees out of phase.
As time progresses through to Figure 18, the oscillation continues to decay. As a result the majority of the phasors in this capture are smaller than they were in Figure 17. The average system damping factor here has achieved -0.636. The New England units are still forming a coherent group with the Southern US as characterized by the blue and pink groups. This group is now oscillating against the rest of the system most notably the Northern Plains area as seen by the black group of units. A third group oscillating halfway between the other two has been identified by the clustering algorithm and is shown in red and green.

The oscillation has almost completely died out by the time that the movie approaches the timespan of Figure 19. At all of the measurement points the mode phasors have little to no amplitude. In addition, the computed dampings do not expect them to be growing substantially over the next cycle. The average system damping is still negative but approaching zero, this is not necessarily an alarm but due mainly to the fact that the system is not demonstrating any appreciable oscillation. From this it is obvious that the oscillation has run its course and the system has achieved its new steady-state value.
The visualizations and movie creation described in this section was performed in MATLAB. The individual frames are rendered as MATLAB figures and then imported into an *.avi file to create the movie. The geographic mapping and coastlines of the displays were achieved with the use of M_Map [18], which is a set of MATLAB functions implementing a mapping toolbox. The approaches presented in this section can be used by engineers and control room operators to identify, extract, and visualize inter-area modes within the system in real-time.


This section describes a $K$-Median approach for clustering and identifying disruptive events in spatially-distributed data streams.

4.1 $K$-Medians Approach Detecting Disruptive Events

The frequency of the power system provides a great deal of information about the health of the system and the prevailing operating conditions. The frequency trend indicates the power balance of the system. When the amount of power generated is equivalent to the amount of power consumed the system is in steady state and the frequency remains constant. If there is an imbalance in generation/consumption the system responds by converting some of its kinetic energy to electrical energy to make up for the power imbalance, causing acceleration toward a new operating point. This acceleration can be seen as a change in frequency. An excess of generation causes the system to accelerate, resulting in an increase in frequency. An excess of load causes the system to decelerate, depressing the frequency. This change in frequency is proportional to the magnitude of power imbalance. This concept is demonstrated in Figure 1, which shows the system frequency as monitored by several FDRs over a span of 60 seconds. During this time-span at about $t = 32$ seconds a generator went offline, resulting in an instantaneous generation deficiency of about 1200 MW. This imbalance caused the frequency to drop abruptly. After the drop the system achieves a new operating point and settles into relatively stable frequency value.
The power imbalances resulting from these types of system events can have detrimental effects on the stability of a system as a whole thus it is desirable to monitor the incoming FDR data in real time to identify these sudden changes in frequency. When such an event is detected, further information can be extracted from the frequency data \[ [6,7] \] and the appropriate action can be taken by the system operator.

Streaming data mining is attracting increasing attention in view of its growing applications including electric power grid system. For streaming data mining, single-pass algorithms that use a small amount of memory are essential. Several data streaming algorithms have been proposed \[ [2,3,4,5] \]. Data sets shown in Figure 1 are representative of the type of feature that needs to be detected. An event is preceded by a relatively stable initial operating point followed by a rapid change in frequency and finally settling to a new operating point. This rapid change can either be a drop in frequency as shown in Figure 1. Fitting a step function to a data window should provide both the pre-event and the post-event operating points. Taking the difference of these two operating points yields the event size. A decision boundary can then be placed on this difference to decide between data sets containing events and those that do not.

### 4.2 Using K-Medians to Determine Pre-Event and Post-Event Operating Points

A data window containing an event will be dominated by two clusters of data points, the set of points belonging to the pre-event condition and the set of points belonging to the post-event condition. Each of these sets will also contain some points belonging to the event itself. If a window size is chosen such that it is sufficiently larger than the event region these points will be overshadowed by the pre-event and post-event clusters. Identification of these two clusters was achieved using the K-Medians algorithm with \( k = 2 \). As discussed in \[ [7] \], FDR data sets naturally contain a high number of outliers and as such the median becomes a more appropriate estimator of position than the mean, it is for this reason that K-Medians was chosen as opposed to K-Means. The K-Medians algorithm \[ [4] \] consists of the following steps:

1. Start by assuming 2 clusters (\( k = 2 \)) and selecting initial values for the centroid of each cluster, \( c_j \).
2. For each point in the data set, \( \pi_j \), find the closest cluster as measured using the 1-norm distance:

\[
Q \left( \{\pi_j\}_{j=1}^{k} \right) = \sum_{j=1}^{k} \sum_{x \in \pi_j} \| x - c_j \|_1
\]
3. Compute the new set of cluster centroids \( \{c_j^{(i)}\}_{j=1}^{q} \) by computing the median of the cluster. The median is used because it is the point that minimizes the total 1-norm distance from all points to it.

4. Repeat steps 2 and 3 until there is no change in the cluster centroids.

Given a window of consecutive measurement points and using the frequency value of each point and a \( k \) value of 2 identifies 2 clusters within the input data, one being the measurements with higher frequency values and the other being those measurements with lower frequency values. Taking the difference of the two centroids of these clusters gives a robust indication of the size of a possible event within the data window. This procedure is demonstrated on two sets of data in Figures 2 and 3.

These two figures demonstrate how the \( K \)-median detection algorithm responds to an event. When no event is present within the data, the frequency data is relatively stable and the medians of the two clusters are close together producing a small difference. When the source data contains an event the data before the event is placed into one cluster with the data after the event being placed into the other cluster, as the size of the event increases the medians of these groups will spread apart and their difference will increase. The magnitude of this difference will form
the metric that determines whether an event occurred with the source data.

4.3 Selection of Window Data
Since the FDR measurements are sent continuously in real time it is desirable to implement the event detector in such a way that it can be fed continuously. This was accomplished by using a sliding data window approach. The $k$-Medians algorithm would be performed on a window of data to determine whether an event had occurred within that time-span. This window would be moved across the data as it arrives from the measurement devices and re-evaluated. An optimal window size needs to be determined such that the decision metric provides a clear definition for event and non-event cases. In order to evaluate this several data windows of differing sizes were tested on events from the training set. A selection of these results is presented in Figure 4. The decision metric value is plotted as a function of the data window size. The blue lines are derived from a few typical events and the red lines are derived from data sets without events.

It is seen that as the window size increases past 50 points there is a separation between the event and non-event cases. When the window size is small the event drop occupies most of the data window, this pulls the group centroids together, reducing the magnitude of the decision metric. As the window size increases more data from the stable regions is included and each centroid will be pulled toward the desired position. Increasing the size of the data window past a certain point produces diminishing returns, with the cluster centroids settling to the desired stable points. This effect is seen in the right-hand portion of Figure 4 as the traces approach horizontal. Additionally it was observed that an excessively large window may capture slower data trends that we are not concerned with here, corrupting the results. Finally a smaller window size also means that less data needs to be buffered in a real time application, decreasing the delay between an event occurrence and its identification.

![Response of Decision Metric to Window Size for Selected Events](image)

Figure 4: Effect of Window Size on k-Medians Metric for Select Training Events

Considering all of these factors it was determined that a 150 point (15 sec) data window produced the best results for the training cases. In the final implementation the data window shifts by a quarter of its width during each successive evaluation. Shifting in this manner reduced the amount of times the metric needed to be computed while ensuring that any events would be properly positioned within at least one data window.

4.4 Determination of Decision Boundary
A training set was compiled from 23 data sets containing events; in additional, several hours of data that contained no events was included in the training set. The 150 point sliding window was
passed through this data and the metric was evaluated for each window. These results were then referenced against the known event times to evaluate the performance of the $k$-Medians difference metric. These results are given in Figure 5.

![Figure 5: k-Medians Metric Evaluated on the Training Data](image)

Each point in Figure 5 represents one window within the training set data. The blue crosses correspond to windows which contain no event. The black circles represent windows which contain a full event signature and the red circles are windows that contain a portion of an event, but not the entire drop. The windows containing incomplete events are scattered throughout both of the other groups. These points represent a transition from a non-event case to an event case and as such their appearance within each group is expected. Due to the sliding nature of the data window and the width of an event signature every event is guaranteed to appear fully within at least one window. Because of this the red points in Figure 5 become “don’t care” situations at they represent events which have either already been detected or that will be detected by an upcoming window. An evaluation of Figure 5 to determine a decision boundary between the blue and black groups showed that the optimal decision boundary occurs when the absolute value of the $k$-Medians difference metric is 0.0165 Hz. Values less than this indicate no event while values greater indicate an event occurred. The decision boundary determined here is specifically designed for the Eastern Interconnection in the US power grid system. Other interconnections within the power grid systems will have a different decision boundary.

4.5 Evaluation of Decision Boundary

After the construction of the event detection algorithm and setting the decision boundary for its metric it was tested against several data sets outside of the training set. For all cases it correctly differentiates between windows that contained event signatures and those which did not. A time based comparison of the event detection metric against the source data is presented in Figure 6. Here the top chart gives measured data values. The bottom chart gives the corresponding values of the $k$-Medians difference decision metric during the event. The selected decision boundary appears as a horizontal dotted line on the decision metric plot. As demonstrated by these plots the metric is below the decision boundary before the event, it exceeds the value during the time frame of the event signature and finally returns below the boundary after the event has passed.
5. Identifying Grid Disruptions Using Time Series Change Detection

A challenge associated with identification of the grid disruptions is to detect the events as early as possible. The K-Median approach discussed in Section 3 requires data before and after the actual event to reliably identify an event. This constraint might render the approach undesirable for real time decision making applications. This section describes a faster method to identify grid disruptions in individual data streams which applies a time series change detection algorithm to identify significant changes.

Time series change detection has been a much researched area in statistics, especially in the context of statistical quality control [19] [20]. Recently, several data mining approaches have been proposed to identify changes in time series [21] [22]. But many of these approaches are either not online or are not scalable to the high throughput data, as encountered in this domain.

An approach based on the widely used quality control method called cumulative sum control chart (CUSUM) [23] for time series change detection is discussed here. Change detection has been extensively studied in the context of time series analysis and forecasting. The standard approaches include various smoothing techniques, the Box-Jenkins ARIMA modeling, innovation and outlier analysis, and more recently wavelet-based methods.

5.1 A CUSUM Based Fast Time Series Anomaly Detection Method

In this subsection a CUSUM based methodology to identify anomalies in a univariate time series is discussed. The exact problem can be defined as:

Given a regularly sampled time series $T$, such that, each value $T_t$, is the measurement at time $t$, for a given sensor, identify spans of the form $[t_1, t_2]$, such that the underlying system is in an anomalous state from time $t_1$ to $t_2$.

While traditionally, CUSUM is used to identify changes in a time series, it can be adapted for anomaly detection in time series. The core idea behind CUSUM is that one can use it to identify the start and the end of an anomalous regime in a given time series by identifying a change point.
The CUSUM approach is a sequential method which involves calculation of a cumulative sum. For a given time series, at a given time $t (> 1)$, the following two quantities are computed:

\[
S^+_t = \max(0, S^+_{t-1} + (T_t - \omega_t))
\]

\[
S^-_t = \max(0, S^-_{t-1} + (\omega_t - T_t))
\]

The quantity $\omega_t$ is the weight assigned at each time instance. While this can be dependent on $t$, it is set to:

\[
\omega_t = \mu_0 + k|\mu_0 - \mu|
\]

$\mu_0$ denotes the expected mean of the in-control distribution of $T$, while $\mu$ denotes the minimum shift that needs to be detected. The method has one parameter, the *allowance value* $k$, and is typically set to 0.5 or 1. The allowance value $k$ governs how much deviation from the expected mean is allowed. If a process is naturally noisy, $k$ is typically set higher while for a relatively stable process $k$ is set to a low value.

The process is initialized with:

\[
S^+_1 = \max(0, T_1 - \omega_1)
\]

\[
S^-_1 = \max(0, \omega_1 - T_1)
\]

The $S^+$ statistic monitors the changes in the positive direction (also sometimes referred to as “high-side” CUSUM) and the $S^-$ statistic monitors the changes in the negative direction (also sometimes referred to as “low-side” CUSUM). For monitoring grid-disruptions, the low-side CUSUM is relevant since one is interested in detecting events in which the frequency of the power falls down.

Figure 20 denotes a simple example that illustrates the CUSUM based anomaly detection on a synthetically generated time series data. The first 100 and last 100 points in the time series are generated from a normal distribution with mean 0 and standard deviation 1. The points from time 101 to 200 are generated from a normal distribution with mean -0.25 and standard deviation 1. It can be seen that, even though the anomalous region is indistinguishable to the naked eye, the CUSUM based approach can still identify the anomalous region. During the anomalous state, the CUSUM score increases and starts falling down once the time series is in the normal state.
Computationally, this approach is fast, since it requires a constant time operation at every time second, and is also memory efficient, since we only need to maintain the value of the CUSUM statistic of the previous time instance.

The key issue with this approach is that the output of CUSUM requires a threshold to declare when the system is in anomalous state. If the threshold is set very low, the false positive rate is high, and when the threshold is set high, it might result in a delay in identifying an event. To alleviate this problem, the quality control literature provides a way to set the threshold based on the Average Run Length (ARL) metric. Typically, the threshold is set based on the number of expected events in an in-control process.

5.2 Results on Grid Data

This section summarizes the performance of the CUSUM based anomaly detector, as described in the previous section, on frequency data collected from the FDR sensors described in Section 3 of this chapter. The objectives of these experiments are:

1. Can the CUSUM based approach identify useful events?
2. What is the typical false alarm rate?

5.2.1 Data Description and Algorithm Parameters

For the results shown in this section, data from 21 FDRs located within the Eastern Interconnect (EI) of United States. Data for two months (May 2008 and June 2008, total 61 days) is used. The time series are sampled at the rate 10 Hz. The length of the time series for each day is 864000. The data was analyzed separately for each month. The frequency data is preprocessed using the
K-median approach with k set to 5.

The allowance value \((k)\) for the CUSUM algorithm is set to 1, the minimum shift \((\mu)\) to be detected is set to 0.05 and the in-control distribution mean \((\mu_0)\) is 0\(^1\). An anomalous event is defined as a subsequence of a month long time series in which the CUSUM statistic is greater than zero. Based on understanding of the domain, an anomalous event is considered significant if it lasts for at least 2 seconds (20 observations).

### 5.2.2 Raw Results

Table 1 Summary statistics for CUSUM based anomaly detection output for different FDRs summarizes the number of significant anomalous events identified for each of the sensors for May 2008 and June 2008. The results show that for all of the FDRs, the fraction of the time in which the system is in anomalous state is a small fraction of the total time, but that fraction itself can be a large number. For example, FDR 11 in Grand Rapids, MI, was in anomalous state, 0.0082 fraction of the total time, approximately 200,000 observations. But the number of significant anomalous events for each sensor for a month is not more than 119.

Evaluation of the output of the anomaly detector is a challenge, given the lack of ground truth data about the grid related events for that time period. A possibility is to examine the local news sources for the specific days of the events, though that process can be expensive as well as not guaranteed to cover every grid event. To further consolidate the output, a spatial co-location constraint can be applied, as discussed below.

<table>
<thead>
<tr>
<th>FDR #</th>
<th>Location</th>
<th>State</th>
<th>May 2008</th>
<th>June 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fraction Alarms</td>
<td># Anomalous Events</td>
</tr>
<tr>
<td>11</td>
<td>Grand Rapids</td>
<td>MI</td>
<td>0.0082</td>
<td>117</td>
</tr>
<tr>
<td>13</td>
<td>Carmel</td>
<td>IN</td>
<td>0.0081</td>
<td>115</td>
</tr>
<tr>
<td>38</td>
<td>Holyoke</td>
<td>MA</td>
<td>0.007</td>
<td>97</td>
</tr>
<tr>
<td>40</td>
<td>St Paul</td>
<td>MN</td>
<td>0.008</td>
<td>116</td>
</tr>
<tr>
<td>42</td>
<td>Tallahassee</td>
<td>FL</td>
<td>0.0081</td>
<td>116</td>
</tr>
<tr>
<td>510</td>
<td>Blacksburg</td>
<td>VA</td>
<td>0.0038</td>
<td>64</td>
</tr>
<tr>
<td>513</td>
<td>State College</td>
<td>PA</td>
<td>0.0082</td>
<td>119</td>
</tr>
<tr>
<td>514</td>
<td>Simpsonville</td>
<td>SC</td>
<td>0.0009</td>
<td>16</td>
</tr>
<tr>
<td>516</td>
<td>Rochester</td>
<td>NY</td>
<td>0.0056</td>
<td>88</td>
</tr>
<tr>
<td>519</td>
<td>Newport News</td>
<td>VA</td>
<td>0.0034</td>
<td>53</td>
</tr>
</tbody>
</table>

\(^1\) The data has already been centered using the k-Medians approach by using the medians as centers to shift the subset of data.
<table>
<thead>
<tr>
<th>FDR #</th>
<th>Location</th>
<th>State</th>
<th>May 2008</th>
<th>June 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Grand Rapids</td>
<td>MI</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>Carmel</td>
<td>IN</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>38</td>
<td>Holyoke</td>
<td>MA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>St Paul</td>
<td>MN</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>42</td>
<td>Tallahassee</td>
<td>FL</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>510</td>
<td>Blacksburg</td>
<td>VA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>513</td>
<td>State College</td>
<td>PA</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1 Summary statistics for CUSUM based anomaly detection output for different FDRs

5.2.3 Incorporating Spatial Information

Electric grid events (such as outages, trips) typically have a cascading effect. Hence, a grid event must manifest itself in data collected at spatially co-located sensors. Table 2 Number of significant events for spatially aware CUSUM based anomaly detection output for different FDRs shows the number of significant events identified by the CUSUM based anomaly detection system for each sensor, while taking the spatial information into consideration. Thus an alarm raised at a sensor is considered “true”, if it was also raised by the neighboring sensors (8 for this study). The number of significant events is reduced by incorporating the spatial information. While one cannot infer that the events which are thus ignored are not useful, but this approach does allow a reduced set for further analysis.

As shown in Table 2 Number of significant events for spatially aware CUSUM based anomaly detection output for different FDRs the number of significant events is sharply reduced to 143 by taking the spatial context into account. While the real benefit of this approach can be validated only by confirming that the detected events are indeed beneficial in terms of identifying the actual events, the lack of availability of reliable ground truth regarding grid events makes the validation challenging. At the time of writing of this chapter, the detected events were being validated through a manual process, the results of which shall be documented in future publications.
Table 2 Number of significant events for spatially aware CUSUM based anomaly detection output for different FDRs

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simpsonville</td>
<td>SC</td>
<td>0</td>
</tr>
<tr>
<td>Rochester</td>
<td>NY</td>
<td>0</td>
</tr>
<tr>
<td>Newport</td>
<td>VA</td>
<td>0</td>
</tr>
<tr>
<td>Chillicothe</td>
<td>OH</td>
<td>0</td>
</tr>
<tr>
<td>Oak Ridge</td>
<td>TN</td>
<td>0</td>
</tr>
<tr>
<td>Birmingham</td>
<td>AL</td>
<td>9</td>
</tr>
<tr>
<td>Duluth</td>
<td>MN</td>
<td>8</td>
</tr>
<tr>
<td>Madison</td>
<td>WI</td>
<td>8</td>
</tr>
<tr>
<td>Gulfport</td>
<td>MS</td>
<td>9</td>
</tr>
<tr>
<td>Montgomery</td>
<td>AL</td>
<td>9</td>
</tr>
<tr>
<td>Atlanta</td>
<td>GA</td>
<td>0</td>
</tr>
<tr>
<td>Pensacola</td>
<td>FL</td>
<td>9</td>
</tr>
<tr>
<td>Cookeville</td>
<td>TN</td>
<td>0</td>
</tr>
<tr>
<td>Cookeville</td>
<td>TN</td>
<td>0</td>
</tr>
</tbody>
</table>

6. Conclusions

Data mining has immense significance in terms of addressing several key power grid problems, specifically in the arena of rapid event detection, as discussed in this chapter, which have the potential of going a long way in terms of realizing the promised benefits of the smart grid. The key challenges associated with this domain, in terms of data analysis, is the massive nature of the data and the short reaction time allowed (of the order of a few seconds) for allowing adequate response. The analytic solutions proposed in this chapter focus primarily on simple analysis which can be scaled to the data sizes and the high sampling rate of the incoming power signal. In future, as the synchrophasors become more and more advanced (both in terms of sampling rate as well as the number of deployed sensors across the country), more research will be required to make the data analysis solutions scalable.

References


