Event Detection Using Correlation within Arrays of Streaming PMU Data

Rich Meier
Oregon State University

Ben McCamish
Oregon State University

Eduardo Cotilla-Sanchez
Oregon State University

Jordan Landford
Portland State University

Robert B. Bass
Portland State University, robert.bass@pdx.edu

See next page for additional authors

Follow this and additional works at: https://pdxscholar.library.pdx.edu/ece_fac

Part of the Electrical and Computer Engineering Commons

Let us know how access to this document benefits you.

Citation Details

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
Event Detection Using Correlation within Arrays of Streaming PMU Data

Rich Meier, Ben McCamish and Eduardo Cotilla-Sanchez
Oregon State University
Corvallis, OR, USA

Jordan Landford and Robert B. Bass
Portland State University
Portland, OR, USA
Email: robert.bass@pdx.edu

David Chiu
University of Puget Sound
Tacoma, WA, USA

Abstract—Phasor measurement units provide real-time power system monitoring. We present a data analysis method that leverages statistical correlation and analysis methods to identify power system events. This research uses archived phasor measurement unit data to show that the method is useful for detecting power system events. Results from a lighting strike case study are presented. A monitoring stratagem based on PMU clustering is discussed, and the viability of monitoring pertinent statistical parameters over various clustering schemes is demonstrated.

Keywords—phasor measurement unit, PMU, correlation, electrical distance, correlation, event detection

I. INTRODUCTION

Phasor measurement units (PMUs) provide time-synchronized, high-fidelity phasor measurements of electric power systems [1]. PMUs are time-synchronized via the Global Positioning System, thereby enabling phasor measurements to be timestamped. With measurement rates of 30, 60, or 120 readings per second, precise comparisons can be made between measurements from disparate points within a power system. These measurements are part of a hierarchical network that stream data from PMUs to higher-order systems, such as state estimators, that form a “big picture” overview of an entire system. This enables new insight into grid dynamics and allows for development of new applications [2].

In this paper, a monitoring stratagem based on a PMU site-clustering scheme is introduced, whereby data streams from a subset of PMUs within a balancing area are aggregated and then analyzed. This subset consists of both electrically near and electrically far PMUs. From this subset, a pairwise comparison between each PMU site’s positive sequence phase angle is made using a linear statistical method to create a vector of correlation values. When making this comparison, only a subset of PMUs are needed to provide beneficial information, avoiding redundant and/or unnecessary computation, thereby lowering computation costs.

In our previous work, the capability of this method was validated when used in analyzing pairwise positive sequence phase angles for event detection and its robustness was demonstrated when adjusting the window size (i.e., window of correlation) [3]. The correlation vectors are then used to instantiate a Rayleigh distribution in which statistical parameters can be quantified. We hypothesize that monitoring the statistical parameters that arise from correlation within small clusters of PMU sites may be used for identifying disturbances within the monitored transmission region.

Notable examples of methods for monitoring PMU data streams for system events have been developed. Researchers have developed algorithms for detecting periodic power system oscillations originating from large generation units. Zhou developed an algorithm that compares PMU measurements to an established, frequency-independent threshold, whereas Fullon and Pierre use a frequency-dependent threshold [4], [5]. Kumar, et al., build upon work by Zhou and Dagle, and demonstrate their technique using real-time PMU data from the North American Synchrophasor Initiative [6], [7].

These algorithms demonstrate the value of using PMU data to provide wide-area monitoring. They differ from the method presented within this paper in that our algorithm concurrently monitors multiple PMU data streams and uses correlation between these streams as a means for identifying system events. Other researchers are also investigating the value of PMU data stream correlation. Chakhchoukh, et al., enhance state estimation by supplementing some PMU data streams into the estimator, whereby the state estimation takes into account the time and spacial correlation known to exist between PMUs at adjacent substations [8].

II. BACKGROUND

This section provides discussions on the characteristics of our dataset and the concept of electrical distance. The PMU-generated dataset consists of a single year’s worth of data from twenty 500 kV substations within the Pacific Northwest of the United States. These are a subset of Bonneville Power Administration’s PMU network. The dataset size is 950 GB and includes positive sequence voltage magnitude ($V_+$) and positive sequence voltage phase angle ($\phi_+$), among other parameters. Each measurement is given in phasor representation with its associated timestamp. The discretization interval between measurements is 16.7 milliseconds (60 Hz). $\phi_+$ is a time-varying real number within the range of $\pm 180^\circ$ and $V_+$ is a non-negative real number.

The concept of electrical distance is applied to decide which PMUs to include in the clustering scheme [9], [10]. Electrical distance, which is substantially different than topological distance, quantifies the electrical proximity between substations [11]–[13]. Two electrically-near substations will experience similar responses to a nearby system event, whereas two substations that are electrically far from each other will not exhibit similar responses to an event. For any power system network, an electrical distance matrix $\mathbf{E}$ may be derived from the $J_{PB}$ quadrant of the network’s Jacobian. Elements of $\mathbf{E}$
quantify the incremental change in phase angle between two nodes given an incremental change in active power between those nodes. Electrical distance has been applied to different scenarios, such as multi-objective power network partitioning, identifying structural vulnerabilities, and evaluating marginal loss factors [14], [15].

Within the spatial and temporal extent of our data set, the power system network experienced several lightning strikes near the PMU site ‘Monrovia’ sufficient to trigger power system protection responses. Figure 1 focuses on this subregion. All substations that have PMUs in this region are labeled and enumerated in ascending order in terms of electrical distance with respect to site Monrovia. These sites represent a PMU cluster for a case study in Section IV.

III. Methodology

This section presents a statistically-based framework for detecting power-system contingencies. This entails the manipulation of correlation vectors obtained through Pearson’s coefficients to conform to a Rayleigh distribution. Statistical parameters are quantified and used to derive an event-detection metric. Using this metric, a clustering scheme is devised to provide sufficient monitoring capabilities while minimizing computation costs.

A. Pearson Correlation

Correlation is a common statistical method used to determine whether relationships exist between two continuous variables. The approach presented here uses a linear statistical method known as Pearson correlation. To assess the linear relationship between variables, the Pearson correlation coefficient \( r \) is determined based on the following equation:

\[
r = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}
\]

(1)

where \( \text{cov}(X,Y) \) is the covariance between \( X \) and \( Y \), \( \sigma_X \) is the standard deviation of \( X \), \( \sigma_Y \) is the standard deviation of \( Y \), and \( X \) and \( Y \) are two independent continuous variables of size \( N \). The output of \( r \) is such that \( r \in [-1, 1] \) where \( r = -1 \) represents a perfectly negative linear relationship, \( r = 1 \) represents a perfectly positive linear relationship, and \( r = 0 \) represents an inconclusive relationship.

The independent continuous variables \( X \) and \( Y \) represent vectors of data from two PMU sites. These vectors can be either positive sequence voltage magnitude, \( V_{+} \), or positive sequence phase angle, \( \phi_{+} \). Since \( X \) and \( Y \) may be of any size \( N \), we vary \( N \) over a range of sizes. We use \( N \) to represent the ‘window size’ of correlation and will refer to it as such from this point forward. Varying the window size increases robustness at detecting different types of dynamics that may occur within power systems.

B. Rayleigh Distribution

The distribution of Pearson correlation coefficient values between PMUs can be characterized using a Rayleigh distribution, though not directly. The probability density function of the Rayleigh distribution is given by

\[
f(x;\sigma) = \frac{x}{\sigma^2} e^{-(x^2/\sigma^2)}
\]

(2)

where \( x \geq 0 \) is the Rayleigh distribution parameter and \( \sigma > 0 \) is the scale parameter, with:

\[
\sigma = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} x_i^2}
\]

(3)

This distribution requires a two-dimensional random vector whose coordinates are independent, identically distributed random variables. This distribution was selected for our analysis because \( r \) meets these criteria when analyzing \( \phi_{+} \) data. The mean and variance of Rayleigh-distributed \( \phi_{+} \) correlation coefficients change during power system events, thereby providing metrics for event detection.

During nominal operations, \( \phi_{+} \) varies slowly in contrast to \( V_{+} \), which can experience more rapid changes. Furthermore, variations in \( \phi_{+} \) correlate strongly between adjacent PMU sites, whereas \( V_{+} \) between sites tends to be less correlated. For these reasons, the methods presented are shown to be highly sensitive to sudden changes in \( \phi_{+} \), making it the desirable parameter for use in detecting power-system events.

Table I shows, for the case study PMU cluster, the amount of negative \( r \) entries as a percentage of total \( \phi_{+} \) entries over a single minute, 1-hour, and 24-hours of nominal data (void of any data/event-related issues) with 15, 30, and 60 cycle window sizes. One can observe that there are very few occurrences of negative \( r \) values of \( \phi_{+} \), less than 0.33% at worst (in bold). Since the amount of negative \( r \) entries of \( \phi_{+} \) is disproportionately low, these entries are discarded when analyzing \( r \) of \( \phi_{+} \) and fitting them to a Rayleigh distribution.

A Rayleigh distribution requires \( x \in [0, \infty) \). However, the range of the Pearson correlation coefficient, \( r \), is \( r \in [-1, 1] \). Noting that very few of the PMU \( r \) values are negatively correlated, we consider only the positive values of \( r \), \( r_+ \in [0, 1] \). We then take its inverse, \( y = 1/r_+ \), for which \( y \in [1, \infty) \). Noting that \( y = x + 1 \), we use \( x = y - 1 = 1/r_+ - 1 \) as the Rayleigh distribution variable, for which \( x \in [0, \infty) \).
For electrically-near PMUs in steady-state, the expected value of $r$ is very close to 1, since under steady-state conditions the voltage magnitude and phase-angle profiles trend approximately the same. Therefore, deviations from $r = 1$, are very rare. As such, the expected value for $x$ is near zero. This is, in fact, what we observe. Observing $\mu(x)$ and $\text{var}(x)$ for PMU pairs included in the case study, it can be seen that there is little variation from steady-state values. Tables II and III show the calculated $\mu(x)$ and $\text{var}(x)$ over a single minute, 1-hour, and 24-hour-periods of time with 15, 30, and 60 cycle window sizes.

A disproportionate number of $x = \infty$ ($r = 0$) values from our PMU dataset do not conform within a Rayleigh distribution. These are related to data drops, data drifts and other data-related decorrelation events. As such, these values are ignored when fitting the Rayleigh $\sigma$ to the Rayleigh distribution. In our previous work, we identified and characterized these types of data-related events. The focus of this work is on events related to power system abnormalities.

### C. Event Detection Metric

$x$ values that conform to the distribution may be used as metrics for power-system event detection. Specifically, monitoring $x$ for deviations outside a multiple, $K$, of the variance can be used as a metric for detecting events, defined as

$$\frac{x}{\mu(x)} > 1 + K\frac{\text{var}(x)}{\mu(x)}$$

where $K$ is an integer multiple. Equation 5 is normalized by the mean, $\mu(x)$. From Table III, $\mu(x)$ and $\text{var}(x)$ are on the order of $1 \times 10^{-3}$ and $1 \times 10^{-4}$, respectively. During steady-state, $x/\mu(x)$ will be approximately equal to 1, and $\text{var}(x)/\mu(x)$ will be approximately 0.1 (from Table II and III). When normalized, $1 + K\frac{\text{var}(x)}{\mu(x)}$ is used as the event detection threshold value and Equation 5 is the event detection threshold inequality.

To show the viability of this detection metric over a window of data containing an event, the normalized $x$ is plotted and the threshold value imposed for a particular $K$ value and site pair. Prior-event and post-event data are indicated with a dotted blue marker. A dotted red marker indicates the event occurrence (starting at cycle 1100 with a duration of approximately 10 seconds), and the established threshold value is indicated with a solid black horizontal line. Figure 3 shows an example. Normalized $x$ that exceed the threshold value are noted with a red ‘o’ marker. Normalized $x$ that exceed the threshold value prior and post event lightning strike are classified as false-positives, differentiated with a blue ‘+’ marker.

In order to eliminate the event detection threshold inequality from evaluating true due to false-positives, the threshold must be set higher than the maximum normalized $x$ value pre-event or post-event. This value, referred to as the maximum false-positive threshold value (MFPTV), is determined over data void of any data-related events, for all PMU pairs over a 60-cycle window size. For each PMU pair, taking the ceiling of the MFPTV signifies the lowest allowable value the threshold should be set to in order avoid erroneously flagging normalized $x$ values as indicative of a power-system event occurrence.

In a similar fashion, in order to preserve event detection, the event detection threshold inequality must evaluate true at least once during the event. This value, referred to as the maximum true-negative threshold value (MTNTV), is determined for all PMUs paired with Monrovia over a 60 cycle window size. For each PMU pair, the MTNTV signifies the largest value the threshold can be set to in order to detect normalized $x$ values during the event occurrence. The MFPTV and MTNTV are used to justify which PMUs to cluster.

### D. Cluster Size

To choose which PMUs are ideal for monitoring, the MFPTV and MTNTV are used to eliminate unnecessary PMUs and select those that are least likely to flag false-positives. It can be observed that there are some MFPTV > MTNTV. This implies that this PMU pair will always report an event due to a false-positive. As such, these PMU pairs should be excluded from being clustered. These tend to be electrically-near PMUs that are highly correlated with the monitored site.
Certain PMU pairs exhibit small MFPTV. These are most suitable for monitoring as they decrease the probability of flagging for false-positives and better preserve event detection. Since large window sizes will be used to detect events and small window sizes to confirm they are not data-related, this approach to determining which PMUs to cluster is only performed for a large window size. It should be noted that these PMUs represent the case-study cluster and, due to the reasons above, justify their inclusion into the cluster.

E. Cluster Content

A cluster consists of four or five electrically-near and far PMU sites. Clusters containing only ‘near’ PMU sites are not sufficient for properly detecting an event, since \( r \) will remain highly correlated given the similar responses that adjacent sites will experience. Conversely, clusters possessing all ‘far’ PMU sites lead to inadequate event detection as correlation will vary, having some strongly and weakly correlated \( r \), thereby leading to more false positives. Cluster composition of near and far sites is discussed in more detail in Section IV-B.

IV. RESULTS

Facilitating development of our event detection algorithm, our dataset includes an event log summarizing locations, times, and causes of all known events that occurred during the covered time frame. Five events were captured during this time, and of these five events, three were lightning-strike occurrences at or near site Monrovia. The other two events, one due to technical difficulty and the other, a lightning-strike occurrence, occur at PMUs that were not included in the dataset. This section presents one of the three lightning-strike occurrences near PMU site Monrovia.

A. Rayleigh Characteristics

When observing the Rayleigh distribution of \( \phi_+ \) over nominal data, the distribution (Figure 2, top) resembles a Rayleigh probability density function. When observing 13 cycles into an occurrence of a lightning strike, \( x \) deviates from its previous distribution profile, Figure 2 (bottom, red ‘x’ markers).

Figure 2 shows how instances of \( x \) outliers in the Rayleigh distribution emerge. When analyzing \( \phi_+ \) between Monrovia and its clustered sites prior to the event, correlation is quantified as very high (\( r = 1 \)) since the system is operating in steady-state. At the instant the lightning strike occurs, a line-to-ground fault takes place near Monrovia, causing a sudden change in real power. Note the large change of scales for both the \( x \) and \( y \) axes between the steady-state data (top four plots) and the event-related data (bottom four plots). \( \phi_+ \) is selected for analysis because it varies slowly (in contrast to \( V_+ \)) and trends similarly between adjacent PMU sites, causing \( r(\phi_+) \) to be highly sensitive to impulse events that alter real power. From Equation 5, \( x \) should remain approximately zero during steady-state. With the sensitive nature of \( r(\phi_+) \), a lightning event causes an abrupt change in real power, and therefore \( r(\phi_+) \), which is captured and signified by \( x \) outliers of the Rayleigh distribution, as shown in the bottom plots of Figure 2.

B. Clustering Scheme

To optimize event detection while preventing unnecessary computation, the algorithm need only monitor a subset of PMUs, so long as that cluster contains both electrically-near and -far PMUs. A near-far cluster offers advantages over all-near or all-far clustering schemes. Using an all-near cluster provides no reference to what constitutes an event, while an all-far cluster produces too many false positives. By clustering a combination of both near and far PMUs, and observing at the correlation layer, an event occurrence can be detected via far sites since correlation becomes decoupled with electrical distance. Furthermore, the event can be confirmed via near sites since adjacent sites experience events in a similar fashion and so their \( \phi_+ \) will trend similarly.

To show the viability of this clustering scheme, the case study cluster was analyzed using one small and one large window size. Figure 3 shows the normalized Rayleigh distribution variable, \( x/\mu(x) \), with blue ‘.’ markers. Each PMU pair has its own predetermined threshold value. Observing 13 cycles into
the event, Figure 3 shows the threshold values as a solid black horizontal line. The Figure shows event-related data that do not exceed the threshold value as red ‘.’ markers, and normalized x values that exceed the threshold as red ‘x’ markers.

From Figure 3, observing the 15-cycle window size scenario, normalized x values will begin to exceed the threshold value a few cycles into the event. About five cycles in, however, a majority of the normalized x values go below the threshold value. Some sites, namely Cully, will experience normalized x values oscillating about the threshold value. Declaring an event based solely on this window size would be inadequate given the limited amount of time to declare the event and the uncertainty caused by the oscillation at Cully. As such, only a small number of PMUs need to be clustered. A simple voting scheme may be used in which a majority of the cluster can vote to declare an event. For instance, for an individual PMU in the cluster, a certain amount of cycles need to consecutively exceed the threshold before declaring an event occurrence. Once an event has been declared, a flag is set and held for a certain amount of cycles in order to wait for other clustered sites to declare an event. If a majority of clusters declare an event during this period, higher-level applications can then be informed of its occurrence.

V. CONCLUSION

This work demonstrates a monitoring framework that aims to facilitate the use of real-time PMU data streams for decision-making and improved control over modern power systems. Given the granularity offered by PMUs, the positive sequence voltage phase angles between PMUs can be analyzed to build a correlation layer. By mathematically manipulating the correlation vectors to conform to a Rayleigh distribution, pertinent statistical parameters can be quantified and used for detecting both power system and data events. The monitoring scheme uses small cluster sizes, thereby minimizing computation costs.

The monitoring scheme considers data streams from a PMU cluster consisting of four or five electrically-near and electrically-far PMUs. The technique may be used to analyze archived PMU data, or to monitor for events in real-time. However, we must emphasize the methods presented here are too slow to actuate protection equipment; PMU sampling and reporting latency alone preclude such application.

The method has potential to detect slowly-devolving power system events, which could instantitate remedial action schemes to prevent actuation of protection systems. A slow decay in system frequency preceding a generator drop is one such example. Such detection is possible because of the high data reporting rate of PMUs, as opposed to SCADA where data reporting is too infrequent to reliably detect such events and where data are not time-stamped, which precludes correlation. Due to the ease of computing both correlation and Rayleigh parameters, and coupled with the computational savings gained by monitoring only select PMUs rather than the entire network, results suggest this framework to be highly viable for so-called slow-fast cascading events detection.

REFERENCES