



U.S. DEPARTMENT OF
ENERGY

PNNL-23840

Prepared for the U.S. Department of Energy

2014 Eastern Interconnect Baselining and Analysis Report

BG Amidan
JE Dagle

November 2014



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <<http://www.ntis.gov/about/form.aspx>>
Online ordering: <http://www.ntis.gov>



This document was printed on recycled paper.

(8/2010)

2014 Eastern Interconnect Baselining and Analysis Report

BG Amidan
JE Dagle

November 2014

Prepared for the U.S. Department of Energy

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

This report looks at the application of situational awareness methodologies with respect to power grid data. These methodologies establish baselines that look for typical patterns and atypical behavior in the data. The objectives of the baselining analyses are to provide: real-time analytics, capability to look at historical trends and events, and reliable predictions of the near future state of the grid. This report focuses on the second objective of discovering historical trends and events.

Nine months of state estimator data was used to create and adapt these methodologies. The first methodology created used statistical algorithms to determine a subset of phase angle pairs that would represent the millions of possible phase angle pairs. This methodology relied on a cluster algorithm to group similar phase angle pairs together and then finding the best phase angle pair to represent each group. This resulted in only needing 104 phase angle pairs for analyses, instead of the over one million possible phase angle pairs. Further analysis determined that these groups can change slightly over time, meaning that the selected phase angle pairs should be updated periodically.

Once phase angle pairs were determined, operational limits could be established which defined the normal behavior for each given phase angle pair. A date/time model was created that used the past 4 weeks of data for a given phase angle pair to help predict the expected normal limits of operation for that particular phase angle pair for the near future. This model could be continuously updated, creating dynamic limits. Exceedances of these limits would indicate abnormal behavior and warrant further investigation.

Multivariate algorithms were created to establish normal baseline behavior and then score each moment in time according to its variance from the baseline. The most atypical times were determined to be atypical events. Atypical events were discovered across the 9 months of data, with many of these corresponding to an actual event on the grid. A list of actual events was collected from the TAG (Technical Advisory Group), with many of these events having high atypicality scores. Moments of time that were summaries of shorter time periods, like 5 minutes, tended to find quick changes in behavior; while moments of time that were summaries of longer time periods, like 1 hour, tended to focus on gradual changes in behavior. Baselines were established focused on all 9 months of data, as well as for each hour, 3 hour block, and season (3 month block). Many atypical events were discovered using any of the baselines; however a few atypical events were only identified when looking at a season, or each hour.

This preliminary work helped develop algorithms that could provide insight from power grid data, specifically state estimator data. The next step is to apply these methodologies to PMU data. These algorithms will first be used to provide insight into historical trends and events. As the algorithms are matured and confidence is gained in their insights, then they can be transitioned to a real-time analytic to monitor the current state of the grid. As these atypical events are better understood and classified into groups of events, then algorithms can be produced which look for precursors to events, providing reliable predictions concerning the near future conditions of the grid.

1.0 Introduction

As data become more available and more frequent across the power grid, the need for real-time situational awareness tools becomes more evident. Understanding typical behavior and atypical behavior is necessary to understand the current state of the grid. The “big picture” objective of this baselining and analysis study is to develop algorithms and tools that can ingest power grid related data from many different sources and provide:

- Real-time analytics, monitoring the state of the grid;
- capability to look at historical trends and events; and
- reliable predictions about the forthcoming state of the grid.

This paper contains the 2014 deliverables for the Eastern Interconnect Baselining and Analysis efforts done at PNNL (Pacific Northwest National Laboratory). This work focuses on analyses performed using state estimator data, which is recorded approximately every 5 minutes, except in the case of NY ISO which is approximately every 30 seconds. Methodologies and analysis techniques were developed using state estimator data in preparation for PMU data. PMU data consist of sub-second measurements. More frequent measurements should provide additional insight discovered at a faster pace. Future work will focus on applying these methodologies to PMU data with the intent to accomplish the “big picture” objectives, as listed above.

This paper is organized as follows –

- Section 2 introduces the methodologies developed to create baselines and perform analyses. These methodologies discuss how to select phase angle pairs to use within analyses, establish baseline limits for the phase angle pairs, and identifying anomalies in the measurements.
- Section 3 shows the results when applying these methodologies to 9 months of 2011 state estimator data from 4 Eastern Interconnect ISOs. This discussion will include the selection of phase angle pairs and the anomalies found during this time period. The anomaly results will be compared to the list of events as determined from the 4 ISOs. Results from sensitivity analyses varying certain aspects of the methodologies will also be discussed.
- Section 4 discusses the conclusions from the results and analyses. Next steps are presented to show the steps necessary to fulfill the “big picture” objectives.

2.0 Baseline and Analysis Methodology Overview

“Situational Awareness” is the process of understanding the elements in a complex system, discerning how they behave with changes to the system (i.e. over time), and projecting their status as these changes occur. Advanced statistical and mathematical algorithms can be applied to complex system data to help provide insight into the situational awareness of the system. This research looks at ways to build algorithms around power grid related data, to help provide the system engineers with an awareness of grid behavior. This research has been focused on two areas: 1) establishing a baseline of what is “normal” grid behavior, and 2) identifying unenvisioned anomalies within the power grid. Section 2.1 explains the methodology used to mathematically determine which phase angle pairs should be used for analyses. Section 2.2 explains the methodology used to establish the baselining limits for the selected phase angle pairs. Section 2.3 discusses the methodology and techniques used to identify patterns and detect anomalies in power grid data.

2.1 Phase Angle Pair Selection

With thousands of phase angles existing across the Eastern Interconnect, the number of possible phase angle pairs is overwhelmingly large. Because computing limitations make it impossible to follow millions of pairs in near real-time, it is necessary to pick which pairs should be computed and included in analyses. Domain experts can provide a list of pairs they deem important. In addition to this, mathematical methods can be employed to help pick pairs. This section discusses a mathematical method used to determine the best pairs and lists which pairs were selected.

When looking at many pairs of phase angle differences, it is easy to see that the values of many of the pairs are very similar across long periods of data. This means that a lot of redundant information is being included if all phase angle pairs are considered. This mathematical method is focused on reducing the redundant information, leaving unique phase angles pairs to be included in the analyses.

The first step in determining the best phase angle pairs to include in analyses is to calculate all possible phase angle pair differences for a period of time. The important aspect of the phase angle pair differences is how much that difference changes over time. The changes over time were best analyzed by centering each phase angle pair difference, by subtracting the first value from each subsequent difference value. This allows all phase angle pair differences to start from the same point (zero), so that the change of each pair difference can be investigated.

The second step was to perform a statistical clustering technique on all the centered pair differences. The k-means clustering algorithm was applied in this case; however any clustering algorithm could be applied. In order to cluster data using k-means, the number of clusters desired needs to be known. There are many different methods that can help determine the optimal number of clusters in a dataset. The method used in this case was selecting the minimum number of clusters that explain at least 90% of the total variance in the data.

Once the number of clusters is determined, the cluster algorithm can be applied and the phase angle pairs can be grouped into similar clusters. When all the phase angle pairs are assigned to the clusters, it is necessary to pick the best pair to represent that particular cluster. These pairs will be included in any future analyses. Pairs that have data quality issues are removed from this selection, so that only “trustworthy” pairs are included. The mathematical center of each cluster is established and the phase

angle pair difference that is closest to the center is chosen as the candidate representative. If domain expertise is available, the representative phase angle pair could be chosen from each cluster by the domain expert.

2.2 Date/Time Model to Establish Baseline of Normal Grid Behavior

Situational awareness of a system occurs as the behavior of elements within the system is better understood. Understanding the behavior of elements within a complex system is necessary to establish guidelines as to what is considered normal behavior. This section discusses the process of optimizing mathematical modeling techniques as they are applied to power grid variables to provide guidelines into the normal behavior of these power grid variables.

These efforts were performed by looking at phase angle differences between pairs of phase angles within the Eastern Interconnect. Nine months of 2011 state estimator data were available, for the purpose of establishing a baseline of what normal behavior is for each of the phase angle pairs.

Mathematical models based on time series analyses were studied to see what aspects of time and past data would help predict normal behavior for the near future for a given phase angle pair. These predictions consisted of the establishment of an upper and lower limit of where normal behavior was expected for each phase angle pair. The first baselining attempt was based on time related information like season, day of week, and time of data. Using cross validation techniques, it was found that this method did not provide very effective upper and lower limits of normal operations. There were too many false-positives, meaning that future data were outside of the limits of normal operation, triggering a warning that the data was not normal, when in fact the data was normal. Having less than a year of data did not provide enough information to be able to establish the seasonal effects. More years of data may improve the model, reducing the number of false-positives.

A second model was built, called the “Date/Time Model”, using the same data, but removing the season factor. To better account for changes in the seasons, the model was only based on the previous four weeks of data. This window of four weeks moved as time increased, such that the model was recalculated often, relying only on the previous four weeks. The following equation represents the model used:

$$\hat{A} = \mu + W_j + T_k + \varepsilon_{j,k}$$

where \hat{A} is the estimated angle pair difference value; W_j is the effect due to the day of week where $j = 1, 2, \dots, 7$; T_k is the effect due to the time of day (measured in hours) where $k = 1, 2, \dots, 24$; and $\varepsilon_{j,k}$ represents the expected error in the model (which should have a mean value of 0). This model produces a predicted estimate of the phase angle pair difference, based on the previous four weeks of data. The normal upper and lower limits are then established by calculating the prediction interval using the formula:

$$\hat{A} \pm t_{(v, 1-\frac{\alpha}{2})} SD(\hat{A})$$

where \hat{A} is the estimated angle value; $t_{(v, 1-\frac{\alpha}{2})}$ is from a t distribution with v degrees of freedom and α (type I error rate); and $SD(\hat{A})$ is the standard deviation of the estimate. Further information about how to calculate a prediction interval can be found in statistical textbooks like Draper and Smith (1998).

Prediction intervals were calculated across the nine months of state estimator data for each phase angle pair using 99% and 99.9% prediction intervals. This means that α was set to 0.01 and 0.001 when applying the prediction interval formula. The percentage of false-positives was greatly reduced when compared to the first model that was based on a seasonal effect, especially when $\alpha=0.001$.

When actual data values exceed the normal limits, the system engineers will be warned that something could be unusual or abnormal. Setting the value for α will influence how often the limits are exceeded. As α decreases, less actual values will fall outside of the limits, because the width of the limits will increase.

2.3 Multivariate Pattern Identification and Anomaly Detection

It is common within complex systems to establish rules to alert system engineers when certain behavior occurs. These rules are criteria that have been predetermined and envisioned by system engineers. An example rule would be providing an alert when power grid frequency exceeds a certain limit, like 61 Hz. These alerts are a simple way to provide a real-time check of a complex system for phenomena that has already been envisioned. While this approach has tremendous value, it also has the potential to miss abnormal phenomena or events that have not been envisioned. Advanced statistical and mathematical algorithms can provide additional situational awareness tools that look for patterns in the data and finds unenvisioned phenomena. This section details the algorithms used to discover patterns and find atypical events within the provided state estimator data.

The first step in analyzing data is to extract features from the data that will provide insight into the state of the system. These features are extracted from all relevant variables, like phase angle difference or voltage from each location. These features make up a mathematical signature which provides a summary of the important aspects of the system and these signatures are used in the analyses. The signature is determined by fitting the following regression equation across a moving window of data for each variable:

$$y = a + bx + cx^2 + \varepsilon$$

where y is the actual data within the window; x is time; a is the y -intercept, representing the mean across the window; b is the slope, or rate of change; c is the quadratic, or rate of rate of change; and ε is the error, or lack of fit of the data to the regression equation. This equation is fit for a window of data of a certain size (in this case one hour). The window is then moved a certain amount of time forward (in this case one minute) and the equation fit again. In each case, the a , b , c , and ε are calculated and stored. This continues across all the data for each variable.

This signature calculation results in the extraction of the following features for each given variable at a specific time –

- the magnitude of the actual data values, represented by a ;
- the rate of change of the data, or slope (first derivative), represented by b ;
- the rate of rate of change, or acceleration (second derivative), represented by c ; and

- the amount of error in the fit of the data to the equation, represented by ε .

Each of these signature elements provides insight into a different aspect of the data. These signature elements can then be *aggregated* (summarized) across a certain amount of time (in this case every 5 minutes). This aggregation is done by calculating the mean, minimum, maximum, and standard deviation of each signature element. This resulted in 16 signature elements calculated for each specified moment in time. In this case every 5 minutes was aggregated into one set of signature elements. These signature elements can then be included in all analyses concerning the data, especially those analyses looking for typical patterns in the data, or atypical events. Amidan and Ferryman (2005) provide more detailed instructions into the calculation of this signature.

Each signature element provides different insight into the state of the system at a given time. For example, the element a_{mean} provides a view of the average magnitude of the data values. The element a_{stdev} provides insight into the variability in the magnitude of the data values. The element b_{mean} provides the average rate of change, while the b_{stdev} provides the variability within the rate of change. When analyzing the data, it is important to note which signature elements are needed to provide the desired feedback. If an analyst is interested in exploring aspects about how the variables are changing over time, they may want to use the b_{mean} element and possibly the a_{stdev} element. If they are just interested in analyzing the actual values (magnitudes) of the data, they may want to use a_{mean} . If they want to explore all the aspects of the signature, they may include all 16 elements.

After the mathematical signatures are calculated, then analyses can be performed. The rest of this section is focused on a multivariate approach in determining patterns and looking for anomalies, or atypical events, in the data. The first step is to decide which data to perform the analysis on. This includes selecting the time period, the variables or variable types, and the signature elements.

Once the data is selected, then a data reduction algorithm is performed. A common method to do this is principal component analysis (Rencher, 1995). The purpose of this step is to reduce the number of variables in the analysis to a set of unique, uncorrelated variables. This is necessary because many of the variables are highly correlated. Including multiple variables that are related to a certain characteristic in the data will weight that characteristic too heavily in the analysis. Principal component analysis removes this issue by creating linear combinations of the variables that result in orthogonal variables, which are not correlated. The number of components that explained at least 90% of the total variation were retained.

A non-supervised clustering algorithm is then applied to the reduced data. There are many clustering algorithms to choose from. In this case, k-means is applied. Clustering uses multivariate distances within the data to determine which data points are similar. Similar data points (in this case a data point is each specific 5 minutes) form a cluster, or group. Each group represents a certain state that the system is in. These represent the common patterns that are in the selected data.

Figure 2.1 shows a 2-dimensional example of clustering. Cluster 1 is a typical (normal) cluster, because it contains many of the data points. Cluster 2 is also fairly typical. This example shows two data points labeled “inlier” and “outlier” that did not belong to each cluster. They are referred to as singletons because their clusters only contain one data point.

The clustering algorithm needs to know how many clusters should be produced. There are many algorithms that attempt to answer this question. In this case, the number of clusters is an input to the algorithm, meaning the decision of how many clusters should be picked is up to the user. In this case usually 50 to 100 clusters are requested.

The next step is the calculation of the global atypicality score (G). A large score indicates the data point (in this case each 5 minutes of time) is abnormal, or atypical. A score closer to zero indicates a typical or normal data point. Cluster membership is one of two parts included in the global atypicality score. The clustering results are used to calculate the cluster membership score, which is calculated by:

$$cms_i = \frac{n_i}{N}$$

where cms_i is the cluster member score for the i th data point; n_i is the number of data points in the cluster in which i belongs; and N is the total number of data points.

The second part included in the global atypicality score is the distance that the data point is from the center of all the data points. Distance is calculated by:

$$D_j = \sum_{j=1}^N \frac{PCA(j)^2}{\lambda_j}$$

where D_j is the distance score; j represents a row of data (a single time period in the data); $PCA(j)$ represents the j th row from the new data resulting from the principal component analysis; and λ_j represents the eigenvalue associated with the j th row of data from the principal component analysis.

Experience has shown that the distribution of the distance scores tends to have a skewed shape with a long tail to the right. The gamma distribution fits this distribution well, especially in the tails (which is the atypical area of interest). Therefore, the gamma distribution is used to estimate a p-value for each of the distance scores. This results in data points further from the center of the data points receiving p-values closer to 0, while points near the center receive p-values closer to 1.

The global atypicality scores (G) are then calculated using the p-values (p) and cluster membership scores (cms) using the equation:

$$G_i = -\log(p_i) - \log(cms_i)$$

This results in a global atypicality score for each data point (in our case, every 5 minutes) that is always positive, with larger scores meaning more atypical. Global atypicality scores usually range between 0 and 25. Further detail concerning the global atypicality score calculations can be found in Amidan and Ferryman (2005).

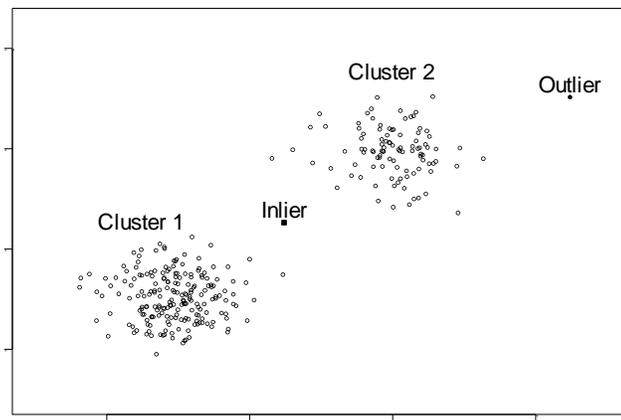


Figure 2.1. Example of Clusters and Singletons

3.0 Preliminary Baselining Results

Analyses were performed as discussed in Section 2. These analyses determined appropriate phase angle pairs to analysis, found typical baselining limits and patterns, and looked for anomalies (atypical events) in the 9 months of state estimator data. A comparison was made between the list of atypical events, as determined from this baselining effort, and a list of known events from the TAG organizations. Section 3.1 discusses the phase angle pairs that were selected. It also includes a sensitivity analysis to show how the phase angle pair selection changes over time. Section 3.2 shows an example of the date/time model that can be used to determine operational limits for phase angle pair differences. Section 3.3 discusses the results from the initial anomaly detection analyses performed on the mathematically selected phase angle pairs and PJM selected pairs. These results are compared to the known list of events. This section also discusses a sensitivity analysis that was performed using the TAG selected pairs (see Table 3.1) to determine how the analyses are affected when different aspects of the analyses are changed.

Table 3.1. TAG (Technical Advisory Group) Selected Phase Angle Pairs Included in the Baselining Analyses (pairs without enough data were not included).

Raun-Sub91	Labadie-Hanna	Labadie-Cumberland
Cumberland-Jackson's Ferry	Monroe-Canton Center	Canton Center-Alburtis
Jackson's Ferry-Alburtis	Alburtis-Ramapo	Niagara-Monroe
Niagara-Ramapo	Canton Center-Hanna	Monroe-Hanna

3.1 Phase Angle Pair Determination Results

In order to perform an analysis using phase angle, it is helpful to look at how a phase angle changes with respect to something else, like a reference angle. The phase angle difference between a pair of angles is informative to the analyst, as it may signal changes in the power grid as changes occur between angles. Section 3.1.1 discusses a mathematically based automated method to determine which phase angle pairs should be used for analysis. This methodology is explained in Section 2.1. Section 3.1.2 provides a sensitive analysis that discusses how the selection of phase angle pairs changes over time.

3.1.1 Selected Phase Angle Pairs

Table 3.2 shows the number of phase angles and the possible number of phase angle pairs that exist for each of three ISOs from the state estimator data that was provided to PNNL. There are well over 1 million phase angle pairs when looking only within each ISO. If phase angles across ISOs are considered, there are nearly 2 million pairs.

For this example, phase angle differences were calculated for one month of data for each ISO and then that data was clustered using k-means clustering and the methods described in Section 2.1. The 90% total variance explained rule was used to determine the number of clusters in the data. This resulted in the need for 104 clusters. The last column in Table 3.2 is the number of unique pairs (clusters) by ISO, which was determined by finding the optimal number of clusters for each ISO.

Using k-means clustering with the optimal number of clusters results in each of the phase angle pairs being assigned to the cluster (group) in which it is most similar to. Table 3.3 shows the number of pairs that were assigned to each cluster for the NY ISO. Figure 3.1 shows an example of the clustering results for four clusters, represented by the four colors. This plot represents many phase angle pair differences being plotted over one month of data and then colored according to which cluster they were assigned.

Figure 3.2 shows the nine candidate phase angle pair differences from the ISO=Ny phase angles. From this figure it is evident that these phase angle pair differences represent 9 unique pairs.

Table 3.4 shows the results from clustering the data, finding the optimal number of clusters, and then picking the candidate phase angle pair from each cluster. This exercise focused on finding phase angle difference pairs within each ISO. These methods could also be used to find the candidate phase angle pairs across multiple ISOs.

Table 3.2. Numbers of Phase Angles per ISO

ISO	# of Phase Angles	# of Possible Pairs	# of Mathematically Determined Unique Pairs
NE	136	9180	35
NY	36	630	9
PJM	1642	1,347,261	60

Table 3.3. The Number of Phase Angle Pairs in each Cluster for ISO=Ny

Cluster #	1	2	3	4	5	6	7	8	9
# of Pairs	32	64	34	31	35	28	138	79	189

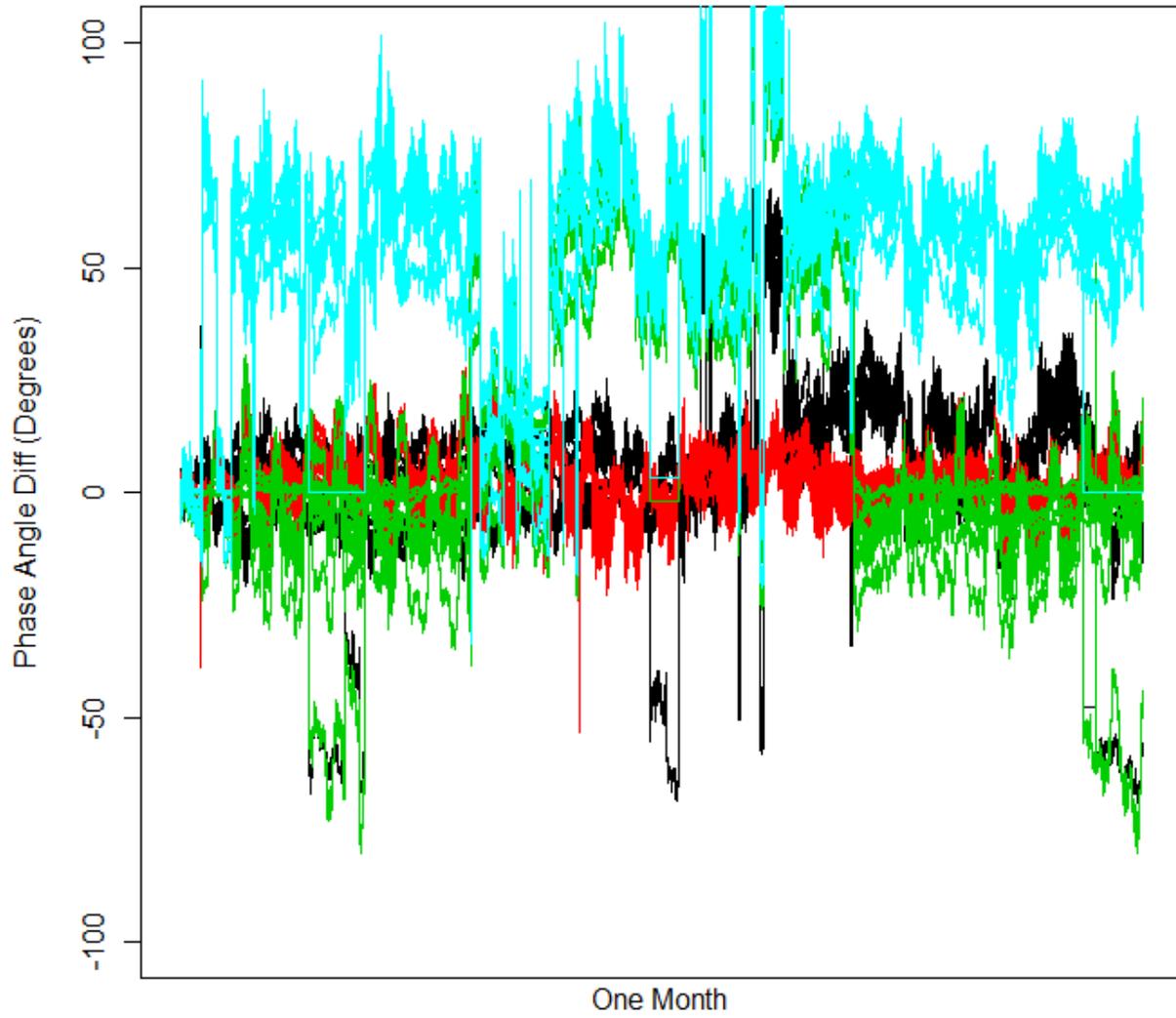


Figure 3.1. Example of Clustering results for 4 Clusters

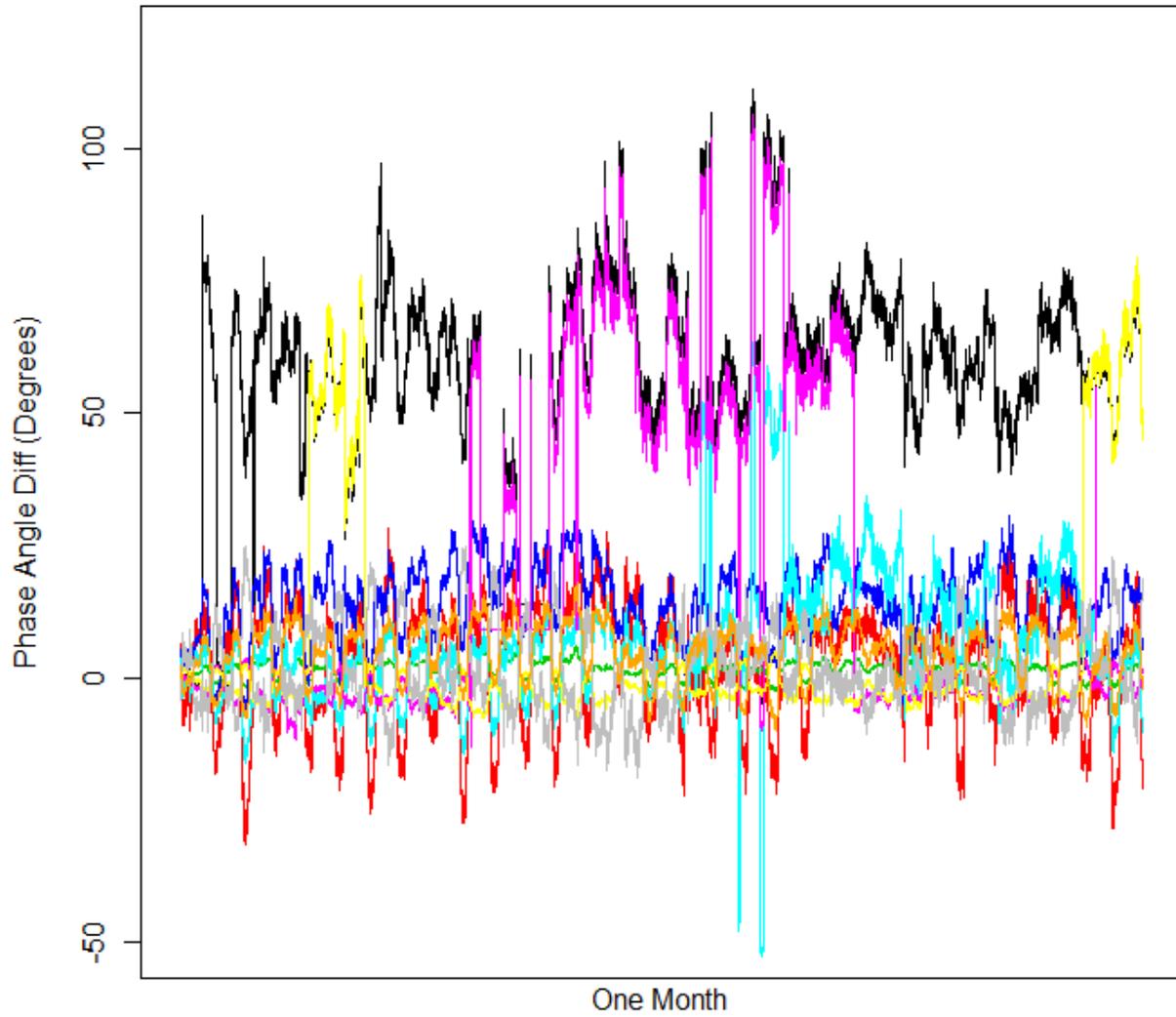


Figure 3.2. Plot of the Nine Candidate Phase Angle Pair Differences Over One Month for ISO = Ny

Table 3.4. Mathematically Determined Candidate Phase Angle Pair Differences (Angle 1 – Angle 2)

Phase Angle 1	Phase Angle 2
Ny:A:FRASER_345_1	Ny:A:NIAGARA_345_1
Ny:A:COOPERS_345_1	Ny:A:RAMAPO_345_13
Ny:A:COOPERS_345_1	Ny:A:MOSES_230_1
Ny:A:COOPERS_345_1	Ny:A:PANNELL_A006
Ny:A:FRASER_345_1	Ny:A:RAINEY_345_31
Ny:A:BUCHAN_S_345_7	Ny:A:ROCKTVRN_345_10
Ny:A:BUCHAN_S_345_7	Ny:A:NIAGARA_345_1
Ny:A:FARRAGUT_345A_5	Ny:A:ROCKTVRN_345_10
Ny:A:BOWLINE_345_2	Ny:A:COOPERS_345_1
Ne:A:PSNH.PSNH.AMHERST.345.9999	Ne:A:PSNH.PSNH.SCOBIE.345.9999
Ne:A:NEP.NEP.WACHUS.345.63	Ne:A:CON_ED.CON.MILLWOOD.345.1
Ne:A:NEP.NEP.SUROWIEC.345.9999	Ne:A:VELCO.VELC.W_RUTLND.345.9
Ne:A:NU.NU.MIDDLETN.345.9999	Ne:A:NBEP.NB.KESWICK.345.1
Ne:A:NU.NU.HADAMNK.345.9999	Ne:A:NBEP.NB.SALBRYNB.345.1
Ne:A:BE.BE.K_STREET.345.90	Ne:A:PSNH.PSNH.FITZWILL.345.7
Ne:A:NEP.NEP.WACHUS.345.63	Ne:A:NU.NU.PLUMTREE.345.4
Ne:A:EUA.EUA.W_FARNUM.345.9999	Ne:A:NEP.NEP.CARPTRHL.345.9999
Ne:A:NEP.NEP.WACHUS.345.63	Ne:A:NM_CNT.NM_C.EDIC.345.9
Ne:A:NEP.NEP.WACHUS.345.63	Ne:A:VELCO.VELC.ESSEX.115.9997
Ne:A:BHE.BHE.ORRINGTN.345.9999	Ne:A:NU.NU.HADAMNK.345.9999
Ne:A:NU.NU.NORWALK.345.44	Ne:A:NYSEG.NYSE.FRASER.345.1
Ne:A:BHE.BHE.CHSTRSVC.345.9999	Ne:A:PSNH.PSNH.SCOBIE.345.9999
Ne:A:NU.NU.BARBOURH.345.45	Ne:A:VELCO.VELC.ESSEX.115.9997
Ne:A:BE.BE.W_WALPOL.345.9999	Ne:A:NYSEG.NYSE.FRASER.345.1
Ne:A:NEP.NEP.SANDY_PD.345.9999	Ne:A:VELCO.VELC.VT_YK.345.9999
Ne:A:NBEP.NB.EDMUNSTN.345.9999	Ne:A:NEP.NEP.WACHUS.345.63
Ne:A:NEP.NEP.SUROWIEC.345.9999	Ne:A:NU.NU.FRSTBRDG.345.9999
Ne:A:NU.NU.HADAMNK.345.9999	Ne:A:NU.NU.LUDLOW.345.9999
Ne:A:NEP.NEP.COMERFRD.230.9998	Ne:A:NBEP.NB.KESWICK.345.1
Ne:A:NU.NU.FRSTBRDG.345.9999	Ne:A:NEP.NEP.DUNBAR_T.230.9998
Ne:A:NU.NU.HADDAM.345.18	Ne:A:NU.NU.NORWLK_J.345.5
Ne:A:NEP.NEP.GOLDN_HL.345.1	Ne:A:PSNH.PSNH.FITZWILL.345.7
Ne:A:NEP.NEP.BELL_303.345.9999	Ne:A:NEP.NEP.W_AMESBY.345.5
Ne:A:NU.NU.SCOVLRK.345.9999	Ne:A:NEP.NEP.WACHUS.345.63
Ne:A:NEP.NEP.WACHUS.345.63	Ne:A:NBEP.NB.SALBRYNB.345.1
Ne:A:NEP.NEP.WACHUS.345.63	Ne:A:NEP.NEP.DUNBAR_T.230.9998
Ne:A:NU.NU.SCOVLRK.345.9999	Ne:A:PSNH.PSNH.LITTLETN.230.2
Ne:A:NEP.NEP.TEWKSBRY.345.9999	Ne:A:NEP.NEP.COMERFRD.230.9998
Ne:A:NU.NU.FRSTBRDG.345.9999	Ne:A:NBEP.NB.BELLDUNE.345.7
Ne:A:NU.NU.FRSTBRDG.345.9999	Ne:A:NBEP.NB.ST_ANDRE.345.1
Ne:A:BE.BE.W_WALPOL.345.9999	Ne:A:NEP.NEP.SUROWIEC.345.9999
Ne:A:VELCO.VELC.W_RUTLND.345.9	Ne:A:PSNH.PSNH.MERIMACK.230.9998
Ne:A:NU.NU.MANCHSTR.345.9999	Ne:A:VELCO.VELC.W_RUTLND.345.9
Ne:A:NU.NU.MIDDLETN.345.9999	Ne:A:PSNH.PSNH.LAWRENCE.345.9999

Phase Angle 1	Phase Angle 2
Pj:A:WAYNEJCT_0230	Pj:A:LOUISAPP_0230
Pj:A:BOSWELL6_0230	Pj:A:LUMBERTO_0230
Pj:A:EDIC~~~~_0230	Pj:A:FRANKFRT_0230
Pj:A:WHITESTO_0345	Pj:A:STERLIN2_0230
Pj:A:BRISTERS_0500	Pj:A:OTTUMW16_0345
Pj:A:KANSAS~3_0345	Pj:A:LOUISAPP_0230
Pj:A:EDIC~~~~_0230	Pj:A:EDANVILL_0230
Pj:A:JUNIPER~_0345	Pj:A:EDANVILL_0230
Pj:A:LEESVIL~_0230	Pj:A:BISMARCK_0345
Pj:A:CHATEAUG_0765	Pj:A:ROCKSPRI_0500
Pj:A:NJTMEADO_0230	Pj:A:OTTUMW16_0345
Pj:A:GWDCO~~~~_0230	Pj:A:LENAPE~~_0230
Pj:A:ANNANDAL_0230	Pj:A:BRISTERS_0230
Pj:A:BETZWOOD_0230	Pj:A:WHAVERTS_0345
Pj:A:STEELCTY_0500	Pj:A:FORTSON~_0500
Pj:A:WBELLAII2_0345	Pj:A:BUSHRV~~_0230
Pj:A:PORTLAND_0230	Pj:A:WILSNTVA_0500
Pj:A:ELROY~~~~_0500	Pj:A:STOLLERD_0230
Pj:A:LEESVIL~_0230	Pj:A:LOUISAPP_0230
Pj:A:STOLLERD_0230	Pj:A:AXTONAEP_0765
Pj:A:POHATCON_0230	Pj:A:MANNING2_0345
Pj:A:CANTONCE_0345	Pj:A:BRKL2007_0345
Pj:A:BEDFORD2_0345	Pj:A:HORTONVI_0345
Pj:A:AXTONAEP_0765	Pj:A:RBRAS~~~~_0500
Pj:A:WBELLAII2_0345	Pj:A:MORGAN~3_0345
Pj:A:CLINTON4_0345	Pj:A:NEWPORT4_0500
Pj:A:GLENGARD_0230	Pj:A:BRKL2007_0345
Pj:A:LENAPE~~_0230	Pj:A:BRISTERS_0230
Pj:A:FRANKFRT_0230	Pj:A:LOUISAPP_0230
Pj:A:HUBBARD6_0230	Pj:A:COXSCORN_0230
Pj:A:POHATCON_0230	Pj:A:REDROCK6_0230
Pj:A:CLOVERD2_0765	Pj:A:LOUISAPP_0230
Pj:A:MARQUIS2_0345	Pj:A:BRKL2007_0345
Pj:A:WARMINST_0230	Pj:A:WIXOM~~~~_0230
Pj:A:NJTMEADO_0230	Pj:A:BLURIG~~_0230
Pj:A:POHATCON_0230	Pj:A:HOLLYMD~_0230
Pj:A:SIBLEY~5_0345	Pj:A:TERMINL6_0345
Pj:A:GWDCO~~~~_0230	Pj:A:LOUISAPP_0230
Pj:A:GREENBRO_0230	Pj:A:HUBBARD6_0230
Pj:A:STEELCTY_0500	Pj:A:RBRAS~~~~_0500
Pj:A:ESSEX~~~~_0230	Pj:A:BOSWELL6_0230
Pj:A:DARWIN~~_0345	Pj:A:DARLCO~~_0230
Pj:A:ELROY~~~~_0500	Pj:A:CANTONCE_0345
Pj:A:MAURY~~~~_0500	Pj:A:AXTONAEP_0765
Pj:A:NE~~~~~_0230	Pj:A:BRISTERS_0500
Pj:A:BYRON~~~6_0345	Pj:A:WHITESTO_0345
Pj:A:BOWEN4~~_0500	Pj:A:EDANVILL_0230
Pj:A:KANSAS~3_0345	Pj:A:WAYNEJCT_0230
Pj:A:EDIC~~~~_0230	Pj:A:STEELCTY_0500
Pj:A:NTIFTO~~_0500	Pj:A:WIXOM~~~~_0230
Pj:A:WATERMAN_0230	Pj:A:DUMONT2~_0345
Pj:A:BYRON~~~6_0345	Pj:A:WIXOM~~~~_0230
Pj:A:OTTUMW16_0345	Pj:A:PERKIOME_0230
Pj:A:FALLCREE_0345	Pj:A:STEELCTY_0500
Pj:A:MILAN~~~~_0345	Pj:A:INDEPEND_0345
Pj:A:DUNEACRE_0345	Pj:A:SCHUYLKI_0230
Pj:A:BETZWOOD_0230	Pj:A:LEESVIL~_0230
Pj:A:CARSON4~_0230	Pj:A:CHATEAUG_0765
Pj:A:TERMINL6_0345	Pj:A:EDANVILL_0230
Pj:A:EDANVILL_0230	Pj:A:STON~NB~_0345

3.1.2 Phase Angle Pair Sensitivity Analysis

The mathematical method to select phase angle pairs, as discussed in the previous section, looked at phase angle pairs for one month of data. There is concern that the nature of the phase angle pairs may change dramatically over time. This section discusses how much change occurs and how this might influence the phase angle selection process.

In order to look at how phase angles pairs change in relation to other phase angle pairs over time, 152 phase angle pairs were studied. This consisted of 26 pairs selected by PJM, 12 pairs selected by the EI TAG (Technical Advisory Group) (some of the other pairs selected by the TAG were not included due to lack of data), and the 104 mathematically selected pairs. Clustering algorithms were applied to this data for the months of January, March, June, and September to identify groupings in the data. The transitions from month to month were studied to see if angle pairs generally stay together.

Tables 3.5 to 3.7 show the cluster transitions from January to March to June to September. These tables show how the cluster membership during one month can change to the next month. For example, in Table 3.5 there were 43 phase angle pairs that were in the first cluster January (sum of the first row) and 40 of them were in the first cluster during March (the 5 in parentheses shows the number of TAG phase angle pairs), while 1 was in cluster “2” and 3 were in cluster “3”. This shows that those 3 phase angle pairs that clustered in groups 2 and 3 were not consistent in the transition. It is important to note that the actual cluster number is not important, clusters are arbitrarily numbered; however it is important to see where those in one cluster one month go the next month. Do they all go to the same 1 or 2 clusters, or does their transition seem to be random? From January to March there were 11 phase angle pairs (indicated in red font) that did not transition into similar clusters. From March to June, there were only 7 phase angle pairs that did not transition into similar clusters. From June to September, all phase angle pairs transitioned into similar clusters. These findings indicate that phase angle pairs that are similar in one month are mostly similar in future months, meaning that the relationship between phase angle pairs is not changing dramatically over time. This means that using one month of data to cluster phase angle pairs and determine the most distinct representative phase angle pairs is reasonable, and the results shouldn’t drift much over a 9 month period. However, it may be reasonable to study these relationships over a longer time period, and as more drift occurs, re-cluster and update the representative phase angle pairs.

Table 3.5. Phase Angle Pair Cluster Transitions between January and March (red indicates phase angle pairs that did not consistently transition; TAG pairs are in parentheses).

		March Clusters					
		1	2	3	4	5	6
January Clusters	1	40 (5)	1	2	0	0	0
	2	14 (2)	18	0	0	0	0
	3	1	16 (3)	1	6 (2)	0	0
	4	3	0	15	0	0	0
	5	2	0	10	0	1	0
	6	0	0	0	0	6	0
	7	0	0	0	4	0	0
	8	0	0	0	0	0	1

Table 3.6. Phase Angle Pair Cluster Transitions between March and June (red indicates phase angle pairs that did not consistently transition; TAG pairs are in parentheses).

		June Clusters						
		1	2	3	4	5	6	7
March Clusters	1	41 (5)	10 (1)	8 (1)	0	1	0	0
	2	2	24	0	9 (3)	0	0	0
	3	2	0	15	0	11	0	0
	4	0	0	0	6 (2)	0	0	4
	5	0	0	0	0	2	5	0
	6	0	0	0	0	0	1	0

Table 3.7. Phase Angle Pair Cluster Transitions between June and September (TAG pairs are in parentheses).

		September Clusters					
		1	2	3	4	5	6
June Clusters	1	37 (4)	8 (1)	0	0	0	0
	2	0	22 (1)	0	12	0	0
	3	14 (1)	0	9	0	0	0
	4	0	0	0	9 (3)	6 (2)	0
	5	0	0	14	0	0	0
	6	0	0	0	0	0	6
	7	0	0	0	0	4	0

3.2 Date/Time Model Results

Section 2.2 discusses a methodology to determine operational limits of where normal behavior should be expected for a phase angle pair difference. This methodology uses the past 4 weeks of data to determine the phase angle pair difference limits to be used until the limits are updated. These updates could occur daily. Because this calculation is dynamic it isn't feasible to show all phase angle pairs across all 9 months of state estimator data. Only one phase angle pair is shown here for demonstrative purposes. Figure 3.3 shows a de-identified phase angle pair over a 3 month time period. The blue line is the actual data, while the black line is the predicted value (\hat{A}). The orange lines are the upper and lower limits when $\alpha = 0.01$ and the red lines are the upper and lower limits when $\alpha = 0.001$. The instances where the blue line extends past the orange or red intervals could be determined to be unusual and worthy of further investigation. This example is very similar to what the other phase angle pairs looked like.

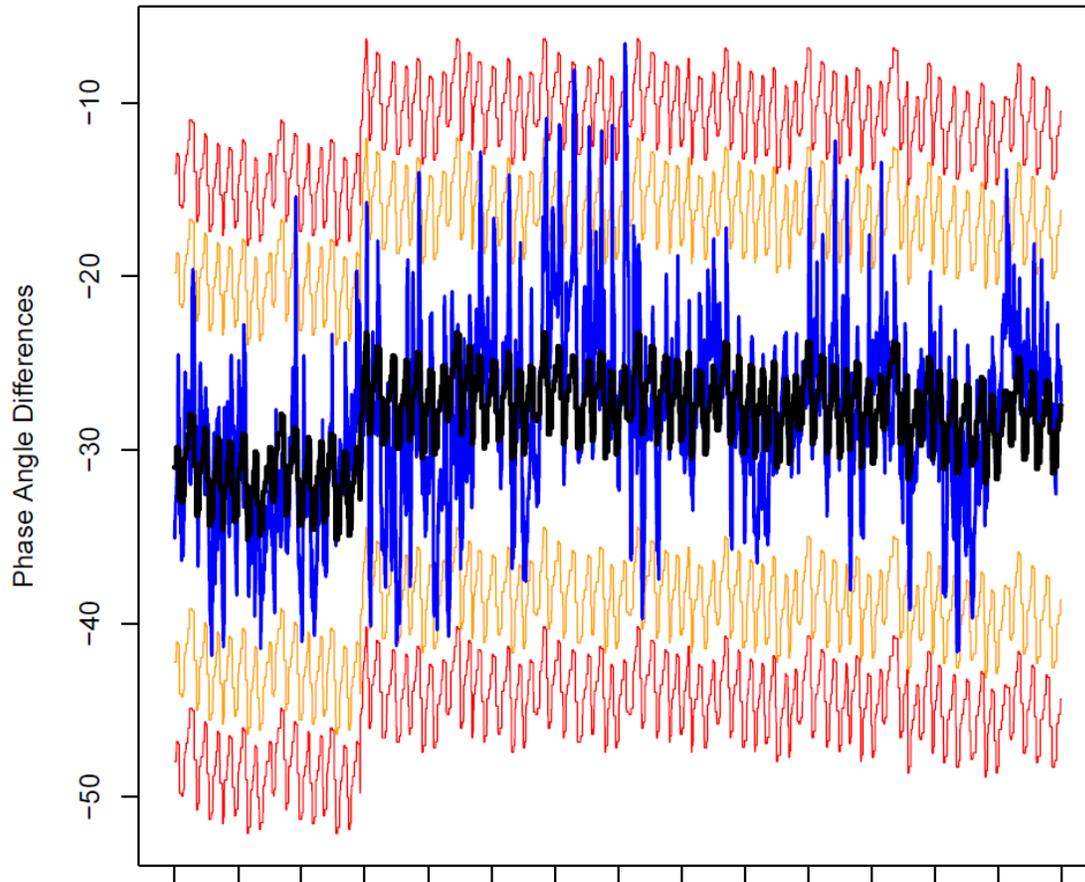


Figure 3.3. Baselining Limits of Normal Operation for an Example Phase Angle Pair Across Three Months of Data. The black line represents the estimated phase angle pair difference; the blue line represents the actual phase angle pair difference; the orange lines represent the 99% prediction limits; and the red lines represent the 99.9% prediction limits.

3.3 Situational Awareness Baselining Results

The baselining anomaly detection methodology explained in Section 2.3 was used to investigate the 9 months of state estimator data. Section 3.3.1 discusses the atypical events discovered using this methodology and compares these events to the actual events listed by the TAG (Technical Advisory Group). Section 3.3.2 shows a sensitivity analysis concerning some of the inputs into the anomaly detection methodology. It shows how the results change as certain inputs are allowed to vary.

3.3.1 Anomaly Detection Results

For this investigation the methodology in Section 2.3 was used, with signatures being calculated using aggregates made every 5 minutes. This means that the quadratic fit equation was calculated at every minute within a window of an hour's worth of data. The coefficients from five of the 1 minute quadratic

fit equations were then aggregated into one multivariate data point. This was done for all of the phase angle pairs that were calculated for analysis purposes.

These analyses resulted in 8 atypical events, as listed in Table 3.8. These events were sent to the TAG. The TAG went through their records to determine if anything of interest happened around those times. The “Possible Reason” column in Table 3.8 lists the responses from the TAG.

Table 3.8. Baselining Determined Atypical Events and Possible Reasons for Atypicality Based on Feedback from the TAG.

Time Period (2011)	Possible Reason
1/11 (16:00)	NYISO TLR: ONT-Frontier earlier in the day at 05:00 hr
1/14 (18:55)	At 01:04 hr a 115 kV line in Connecticut tripped & reclosed due to lightning
1/22 (14:20)	NYISO TLR: Central East and DYSINGER
6/10 (14:20)	NYISO TLR: New Scotland-Leeds beginning 16:00 hr
6/12 (21:50)	At 21:05 hr a 115 kV line in Maine tripped due to lightning
7/11 (6:40)	
7/28 (17:25)	
8/26 (16:40)	

Table 3.9. TAG Identified Events and the Baselining Atypicality Results.

Time Period (2011)	ISO	Reason	Atypicality Results
1/11 (10:14)	NE	Loss of Mystic 8 & 9	5.1 (53 rd percentile)
1/12 (23:12)	NY	Athens 1 & 3 trip due to Xformer explosion	4.8 (44 th)
1/23 to 1/24 (19:00)	NE	Severe cold weather in NE	8.1 (97 th)*
2/1	MISO	MW ice storm	7.4 (95 th)
2/6	MISO	Super Bowl XLV	6.3 (85 th)*
4/22 (11:55)	NE/NY	Loss of Mystic 8 & 9	4.6 (34 th)
4/27	MISO	Brown’s Ferry lost to tornado	7.0 (92 nd) (15:30)
5/10	MISO	Upper Michigan blackout	9.4 (99 th)*
5/26 (3:08)	NY	Sprainbrook Dunwoodie exceeds LTE	Large Atyp. 8 hrs earlier
6/21 (20:18)	NY	Neptune tripped	Not enough data for calc
7/22 (15:00)	NE	Heavy load summer day	High atypicalities in afternoon*
7/28 (1:30)	NY	SAR for ONT loss of Darlington 3	6.6 (89 th) large Atyp. 4 hrs earlier
8/14 (15:20)	NY	SAR for the L/O 7040 NYISO 1054	8.1 (97 th)
8/23 (14:10)	NE	2000 PJM MW lost due to earthquake	>10 (99 th) 12:05 to 16:00*
8/28	NE	Hurricane Irene	>10 (99 th) Throughout day*

*These atypicalities are further discussed in this report.

The TAG also sent a list of identified events that happened during this same time frame. This list was used to see how the atypicality scores behaved during these events. Table 3.9 contains the list of events, time period, and ISO that identified it. The last column contains the atypicality score results, with the atypicality score and the atypicality percentile. Six events with larger atypicality scores are marked with an * and further analyses will be discussed.

Figure 3.4 shows the atypicality scores during a severe cold spell in New England. The atypicality scores do increase; however the increase is not very dramatic. Figure 3.5 shows the atypicality scores around the time of the Super Bowl. The atypicality scores do increase during the time of the game; however the increase is not very dramatic. In both cases the atypicality scores were not very large and neither would have triggered any issues.

Figure 3.6 shows the atypicality scores before a blackout in upper Michigan. The atypicality score is gradually increasing until the point where the data drops out. This may indicate an increasing stress to the system that was occurring before the blackout.



Figure 3.4. Atypicality Score During Severe Cold Spell.

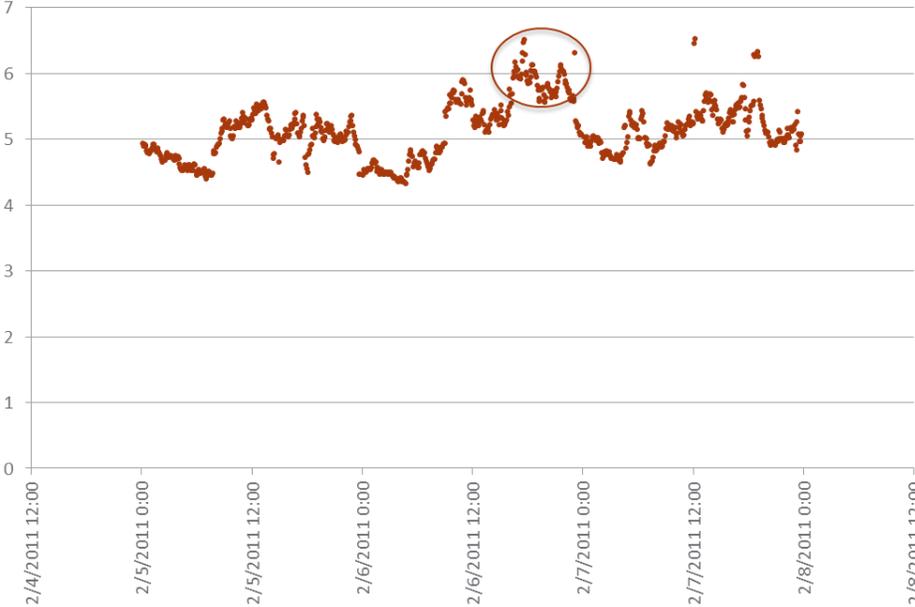


Figure 3.5. Atypicality Score During the Super Bowl.

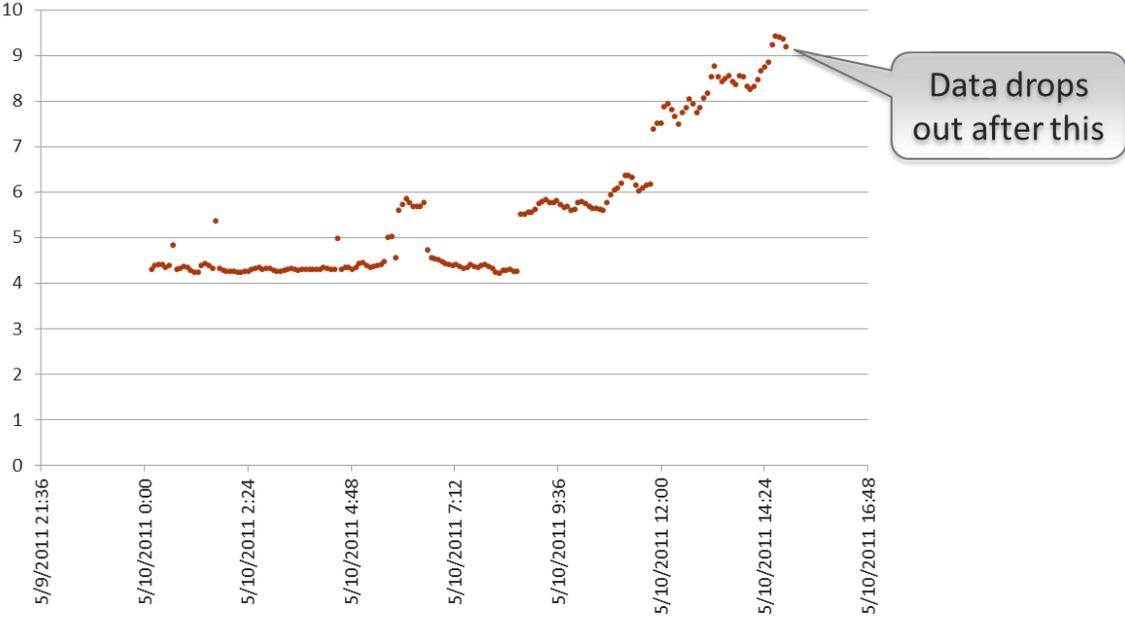


Figure 3.6. Atypicality Score before a Blackout in Upper Michigan.

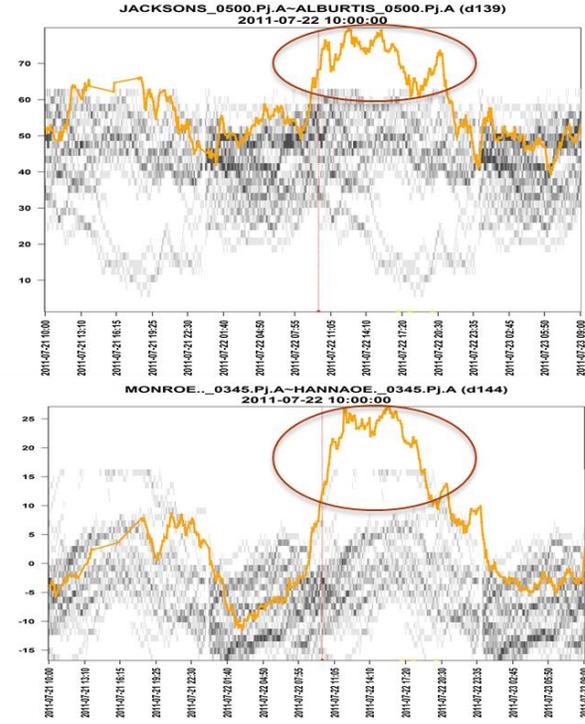
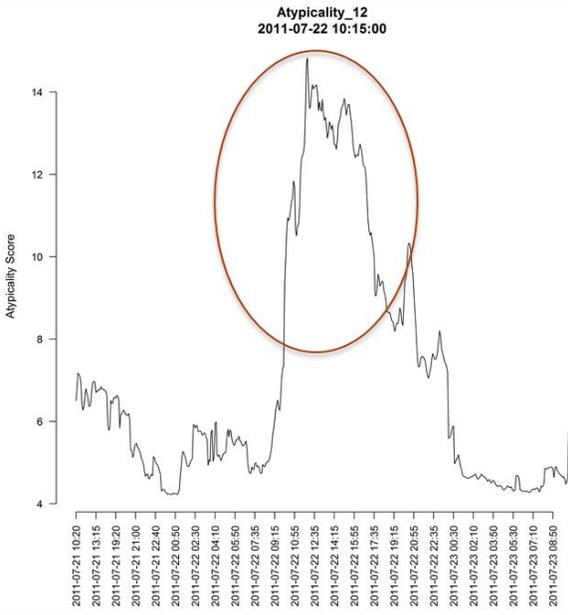


Figure 3.7. Atypicality Score during a Heavy Load Summer Day and Plots of Atypical Phase Angle Pairs.

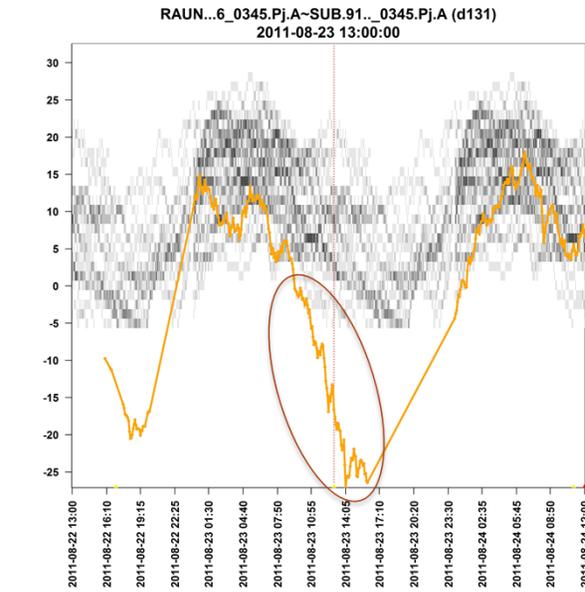
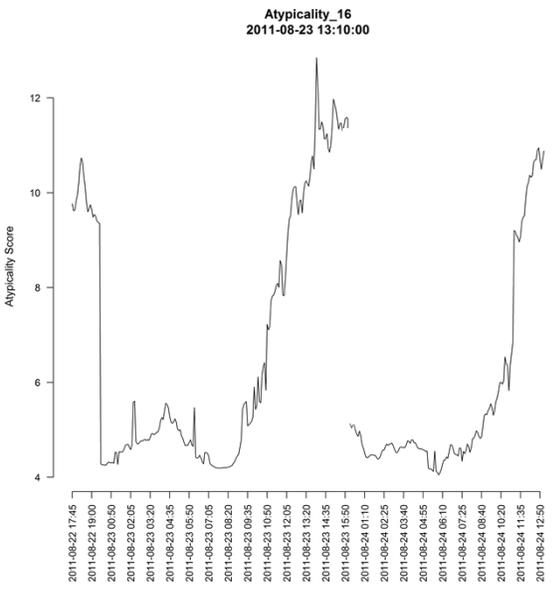


Figure 3.8. Atypicality Score during a Loss Due to an Earthquake and a Plot of an Atypical Phase Angle Pair.

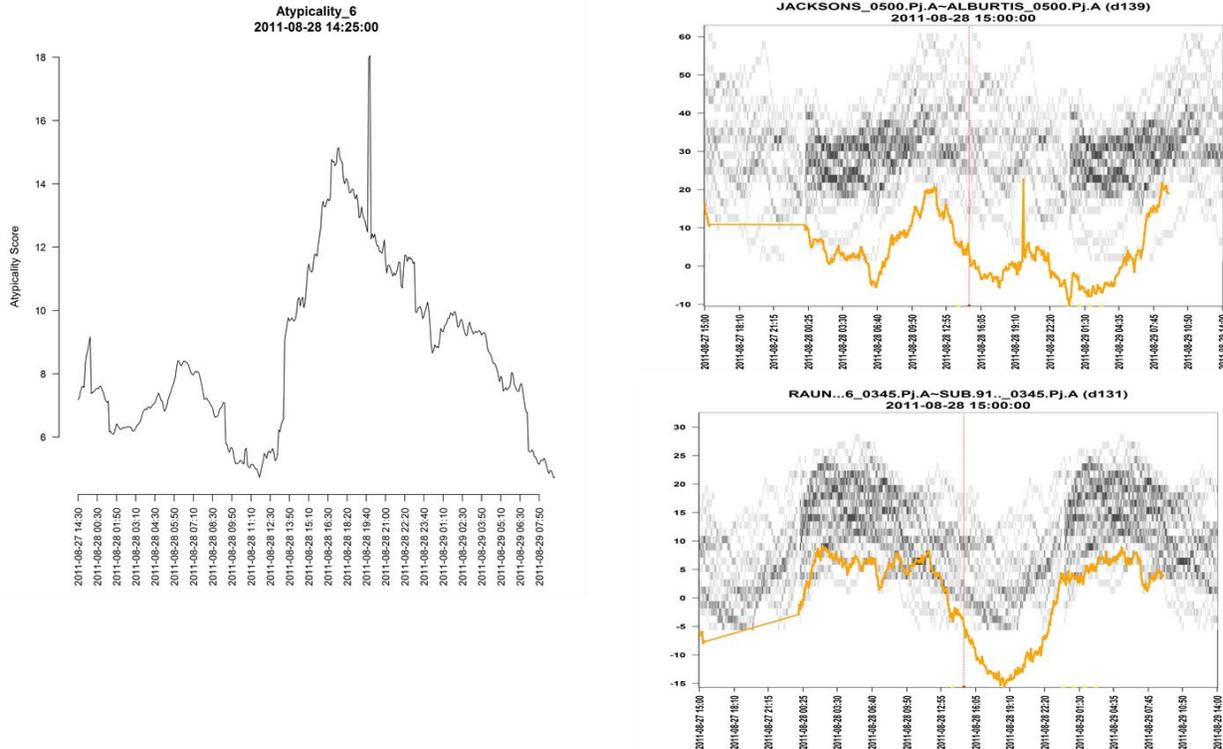


Figure 3.9. Atypicality Score during Hurricane Irene and Plots of Atypical Phase Angle Pairs.

Figure 3.7 shows plots consisting of the atypicality scores (on the left) and a couple of phase angle pairs (on the right) during a heavy load summer day. The atypicality scores are quite large during this time and likely would have triggered a warning that something in the data was atypical. The two plots on the right show the two most atypical phase angle pairs during this time. The background shaded areas within each plot show the typical values of each phase angle pair during that month and during that time of day. Both phase angle pairs have difference values that are well outside of the typical behavior.

Figure 3.8 shows plots consisting of the atypicality scores and the most atypical phase angle pair leading up to and during a loss due to earthquake. The atypicality scores increased until there was a data drop. The atypical phase angle pair plot on the right shows its values were well below the typical values until the data feed returned and the phase angle difference went back to typical.

Figure 3.9 shows how the atypicality scores increased during Hurricane Irene, with a spike occurring. This spike in atypicality score would have likely triggered a warning that something in the data was atypical. The atypical phase angle pair plots on the right show how the phase angle difference was much lower than typical.

3.3.2 Baselining Sensitivity Analysis

The previous analyses were performed by aggregating the 1 minute quadratic fit data into 5 minute aggregates. A sensitivity analysis was performed to study how varying the aggregation size would affect

the atypicality results. Aggregations were calculated at the 5 minute, 30 minute, and 60 minute intervals. The atypical events were determined using each aggregation. Figure 3.10 shows how these aggregations affect the atypical events. The first row of plots is for the most atypical when using the 5 minute aggregation. This same moment in time was only the 8th most atypical when using the 60 minute aggregation. The second row of plots is for the most atypical when using the 60 minute aggregation. This same moment in time was only the 8th most atypical when using the 5 minute aggregation.

From this part of the sensitivity analysis, it can be concluded that 5 minute aggregation tends to focus more on quick changes in trends (spikes). This will find outlier type of behavior when the variables change dramatically and quickly, and then often change back. This type of aggregation should identify data quality issues better.

The 60 minute aggregation tends to focus more on continuing trends. Spikes tend to be more ignored. It was also noticed that 30 minute aggregation also tends to ignore spikes in the data.

Another area in which a sensitivity analysis was performed was in the selection of the “block” of time to include in the analyses. The previous analyses in Section 3.3.1 were performed on all 9 months of data at one time. This results in the typical behavior being defined by 9 months of data. Other blocks of time were studied to see how this changed the atypical events that were identified. The following three blocks were used:

- hour of day – each hour of the day (00 to 23) was analyzed separately. This means that all the data for a given hour (i.e. 12:00 to 12:59) was analyzed for atypical events across all 9 months. This means that each data point (5 minute aggregate) was compared to only data points recorded during that same hour.
- 3 hour blocks of the day – this was done similar to the hour of day block, except that the hours were combined into the groups: 00-02, 03-05, 06-08, 09-11, 12-14, 15-17, 18-20, & 21-23.
- 3 month blocks (season) – this means that each set of 3 months was analyzed separately. The 3 month sets were: January-March, April-June, & July-September.

Most of the atypical events found with no blocking (analyses discussed in Section 4.1) were similar to the atypical events found when blocking each hour, 3 hours, or 3 months. Figure 3.11 shows an exception that was deemed atypical when 3 hour blocking was used, but wasn't significant without blocking. The plots are 2 of 5 phase angle pairs that looked atypical at that time. This moment in time was the 5th most atypical time. The top four atypical events were at the same time as the top four atypical events determined without blocking.

Another exception is shown in Figure 3.12. This moment in time was the 4th most atypical event when blocking using 3 month blocks but was not as atypical without blocking. Figure 4.8 shows that two of the phase angle pairs were outside of typical behavior at this time. All other atypical events using the 3 month blocks were also atypical when not using blocks.

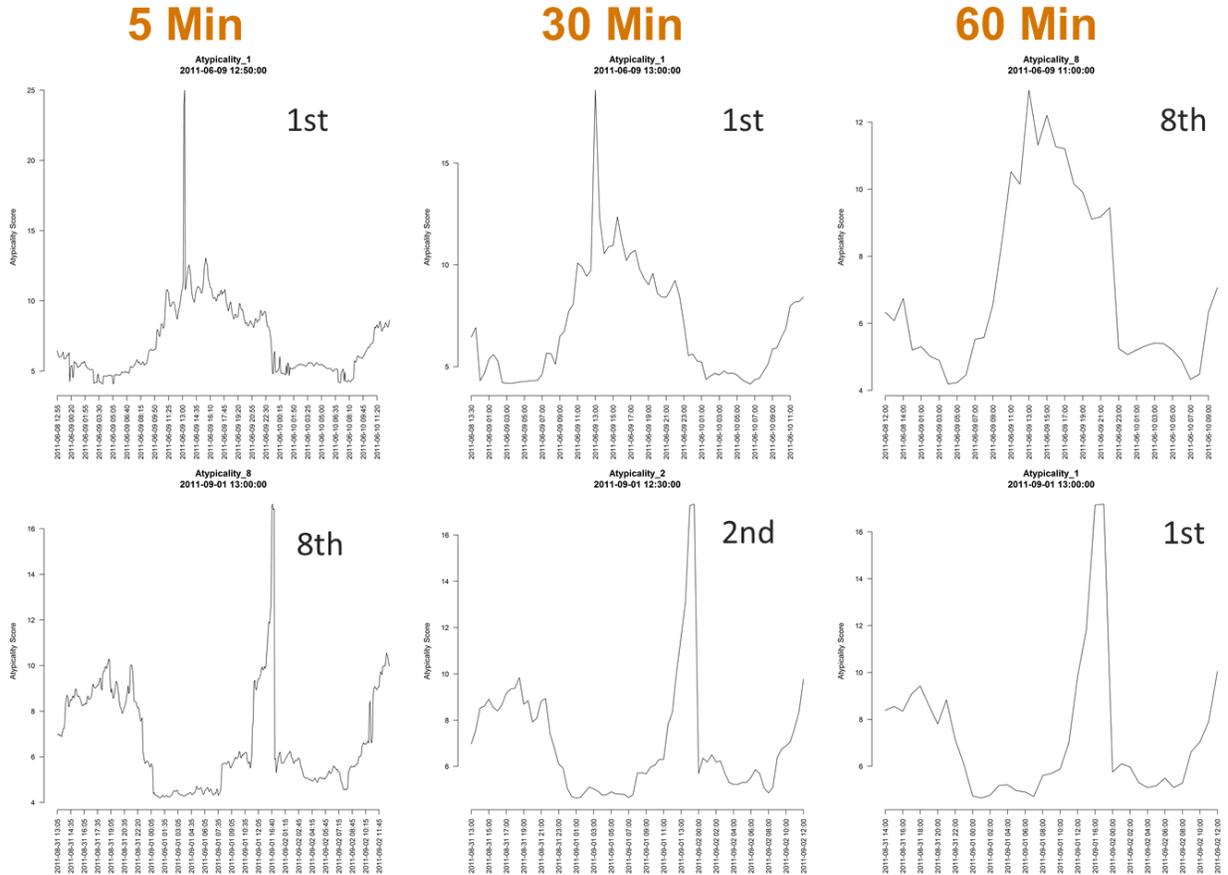


Figure 3.10. Plots of the Atypicality Scores for the Most Atypical Time Period when Aggregating the Signature over 5 Minutes (Top Row of Plots) and the Most Atypical Time Period when Aggregating the Signature over 60 Minutes (Bottom Row of Plots).

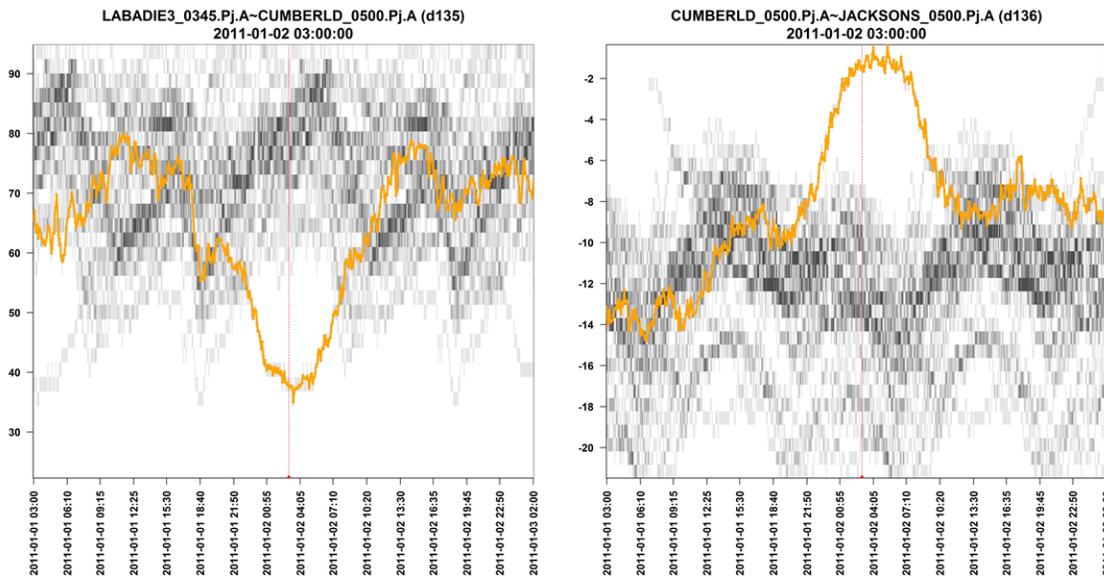


Figure 3.11. Phase Angle Pair Plots for a Case that Was Atypical when Using 3 Hr Blocking for the Analyses, but Not Atypical When No Blocking Was Used.

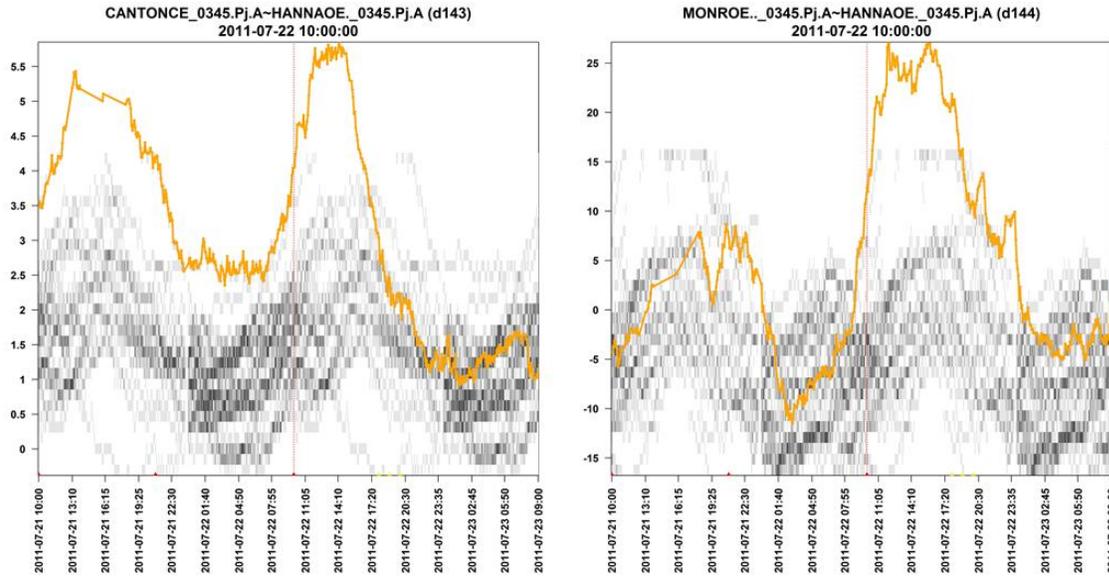


Figure 3.12. Phase Angle Pair Plots for a Case that Was Atypical when Using 3 Month Blocking for the Analyses, but Not Atypical When No Blocking Was Used.

4.0 Conclusions and Future Direction

The purpose of this investigation was to establish statistically based methodologies that perform baselining analyses on power grid data. These analyses have been focused on discovering historical trends and events in past state estimator data. From these analyses the following conclusions can be made –

- A statistically based method was created that can automatically select phase angle pairs that represent all the possible phase angle pairs. The selected pairs are fairly stable over time, although it makes sense to update the selected phase angle pairs periodically.
- Dynamic limits can be made using a date/time model, which uses the last 4 weeks of data to predict the expected range of near future phase angle difference values. This method tends to produce a number of false positives. These limits should be tuned such that the amount of false positives is minimized and acceptable.
- The anomaly detection methodology was applied to 9 months of state estimator data, resulting in further investigation into the top 8 atypical events (Table 3.7). Many, but not all of these events, could be mapped to an actual event that occurred on the grid (as reported by the TAG).
- A list of 15 events during this same time period was collected from the TAG (Table 3.8). Atypicality scores were investigated for each event to determine if they would be detected. Six of these events resulted in high atypicality scores and were further investigated in Section 3.3.1.
- A sensitivity analysis concerning the aggregation of signature elements was performed. This was done to see the differences in summarizing data every 5 minutes, 30 minutes, and 60 minutes. The 5 minute summary data was effective in finding abrupt changes in the data. These changes generally occurred over a short amount of time and returned to normal. These changes could include bad data. The 30 minute and 60 minute summaries were more effective in finding gradual changes in the data.
- Another sensitivity analysis was performed to see if analyzes should be done by combining all the data together (all 9 months), or only looking within a certain block of time. The blocks looked at include: hour of day, 3 hour blocks of the day, and 3 month blocks (seasonal). The blocks usually found the same atypical events as were found when looking at all the data. There were a few cases where additional atypical events were found using the blocks, which were not found when looking at all the data. There were not significant differences in the results for the hour of day and 3 hour blocks of day time period.

This preliminary work helped develop algorithms that could provide insight from power grid data, specifically state estimator data. The next step is to apply these methodologies to PMU data. These algorithms will first be used to provide insight into historical trends and events. As the algorithms are matured and confidence is gained in their insights, then they can be transitioned to a real-time analytic to monitor the current state of the grid. As these atypical events are better understood and classified into groups of events, then algorithms can be produced which look for precursors to events, providing reliable predictions concerning the near future conditions of the grid.

5.0 References

Amidan BG, and TA Ferryman. 2005. "Atypical Event and Typical Pattern Detection within Complex Systems." *IEEE Aerospace Conference Proceedings, March 2005*.

Draper NR and H Smith. 1998. *Applied Regression Analysis*. John Wiley and Sons, NY.

Rencher AC. 1995. *Methods of Multivariate Analysis*. John Wiley and Sons, NY.



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)
www.pnnl.gov



U.S. DEPARTMENT OF
ENERGY