Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Discovery Through Situational Awareness

NASPI – Session 3: Data and Data Quality
March 2017

BRETT AMIDAN
JIM FOLLUM
CARLOS ORTIZ MARRERO
Pacific Northwest National Laboratory

March 29, 2017

DOE GMLC Award – Discovery Through Situational Awareness

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Project Goal

  Create a tool that applies statistical and machine-learning algorithms in context of big data analytics to investigate and implement anomaly and event detection algorithms in near real-time

- Current Focus
  - Working with the Eastern Interconnect
  - Initial focus on phase angle pair analyses
  - Provide the EI partners with a frequent (i.e. daily or weekly) report of the findings

NASPI Contribution

- EATT (Engineering Analysis Task Team) White Paper

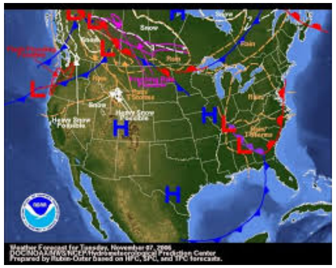
Data Mining Techniques and Tools for Synchrophasor Data

A high level view of data mining –

- What it is
- Why it is important
- How it -
  1) has been used in other industries,
  2) has been used in the power industry, and
  3) may be used in the future

“Big Picture” Vision

Power grid related data (PMUs, State Estimators, SCADA, etc)



Other data (i.e. weather [actual and forecasts], social media)

Data Driven Analytical Tool that provides:

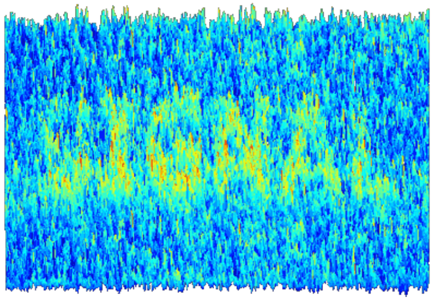
- Real time analytics, monitoring the state of the grid
- Capability to look at historical trends and events
- Reliable predictions about the forthcoming state of the grid

Is More Data Really Better?

Possibly, but beware of …

- Data quality

  Poor quality data will drive your analyses

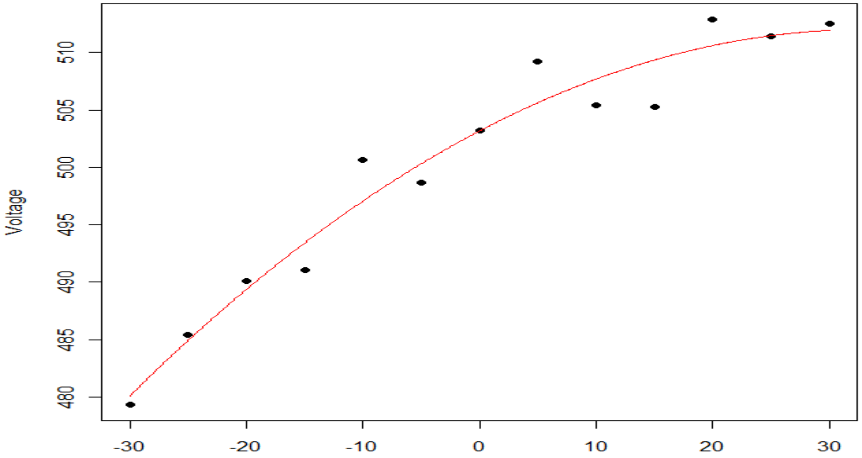


- Creating more noise, so that the signal is tougher to find

Feature extraction and feature selection are important steps that can really make a difference in the success of your algorithms

Feature Extraction (Data Signatures)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

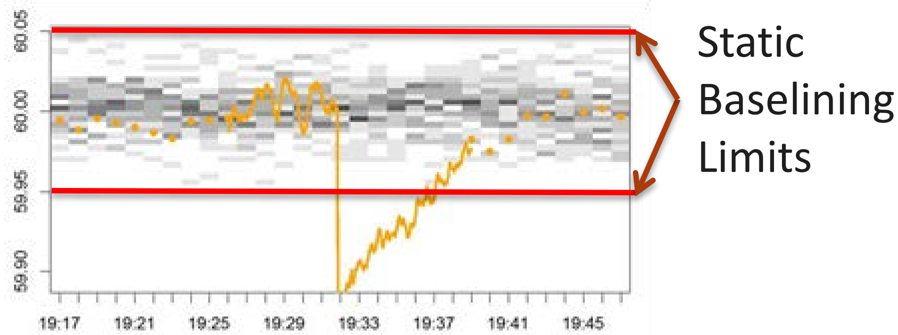


- Regression fits through the data calculate estimates of value, slope, curvature (acceleration), and noise.
- Can be calculated in the presence of missing or data quality flagged values.
- Summaries of these features are used in the analyses.

Baselining Grid Behavior

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

- Univariate Approach
  - Create a baseline of typical behavior for each individual variable
  - Determine abnormal behavior based on the baseline



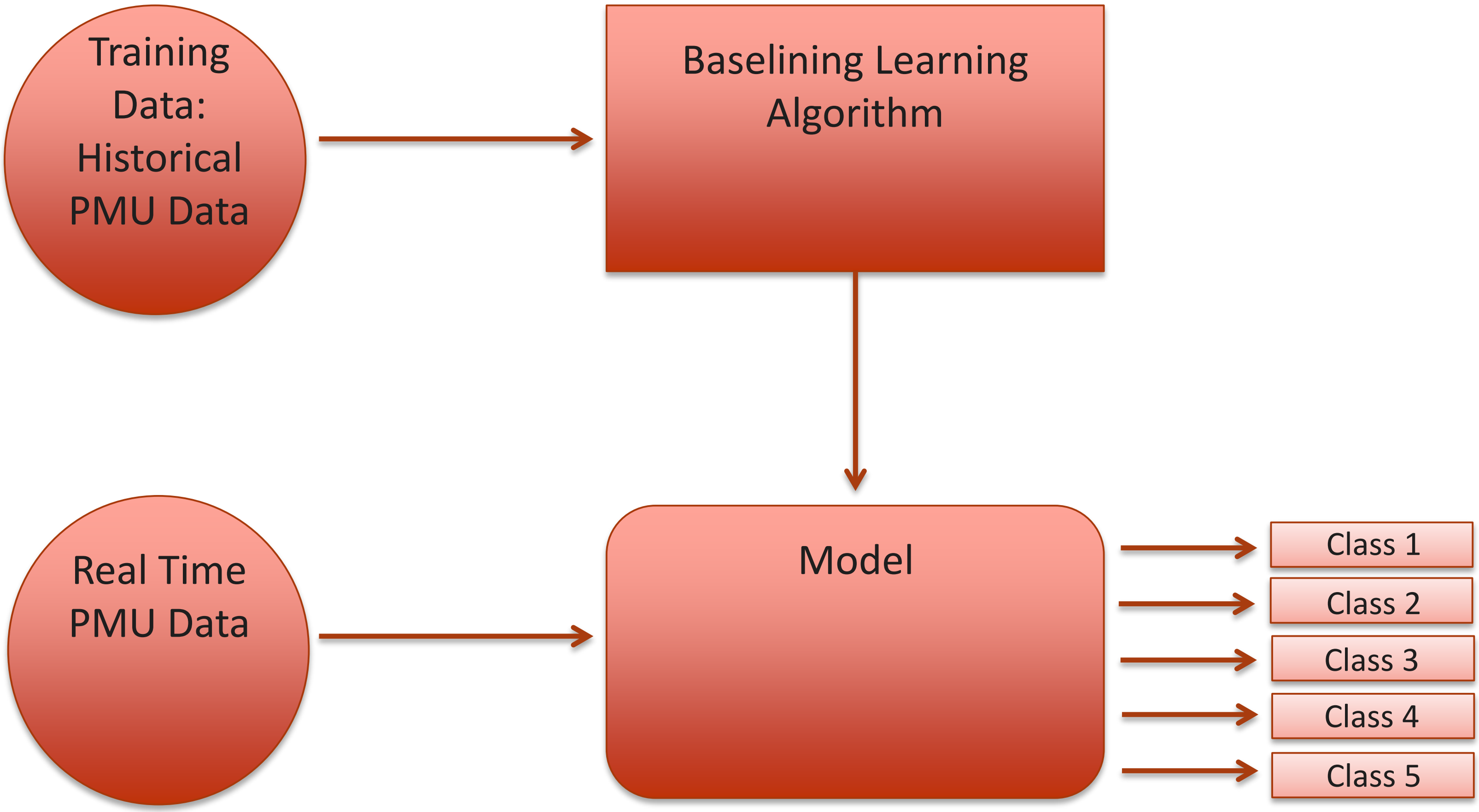
- Multivariate Approach
  - Create a baseline across many (hundreds or even thousands) of variables
  - Relationship between variables is considered when determining abnormal behavior

Multivariate Baselining

Baseline captures what normal behavior is expected to be

- Group similar behavior
  - **Time periods** that group together indicate normal grid behavior
  - **Variables** that group together indicate highly correlated variables and may be candidates for feature reduction
- Identify data that does not belong with the normal behavior
  - **Time period** contains data that is **unusual** (possible abnormal grid behavior)
  - **Variable** is unlike other variables, or something has happened to indicate a **behavioral change in the variable**

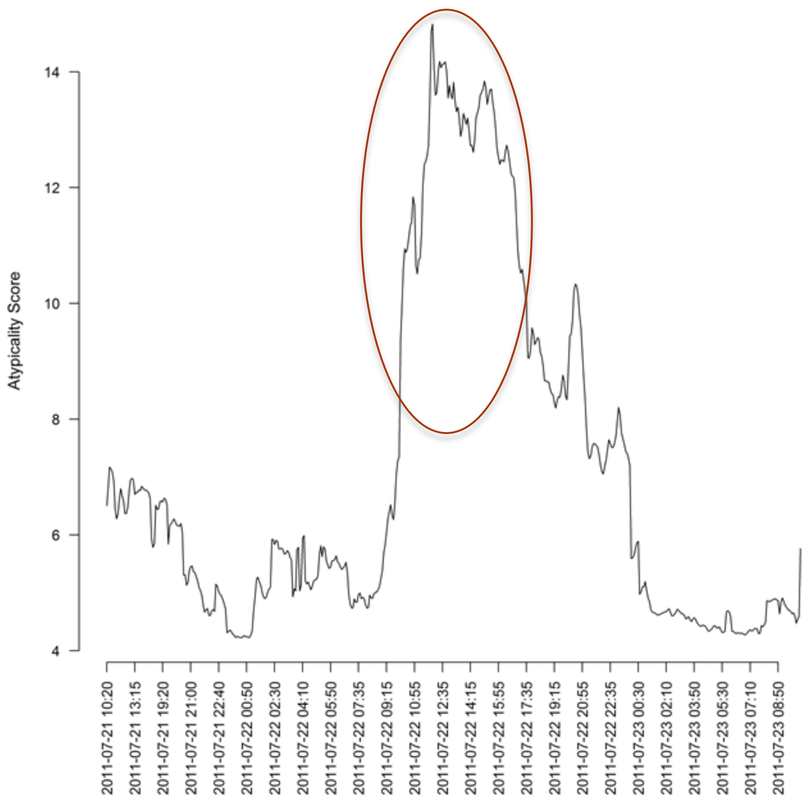
Creating a Baseline – Unsupervised Learning



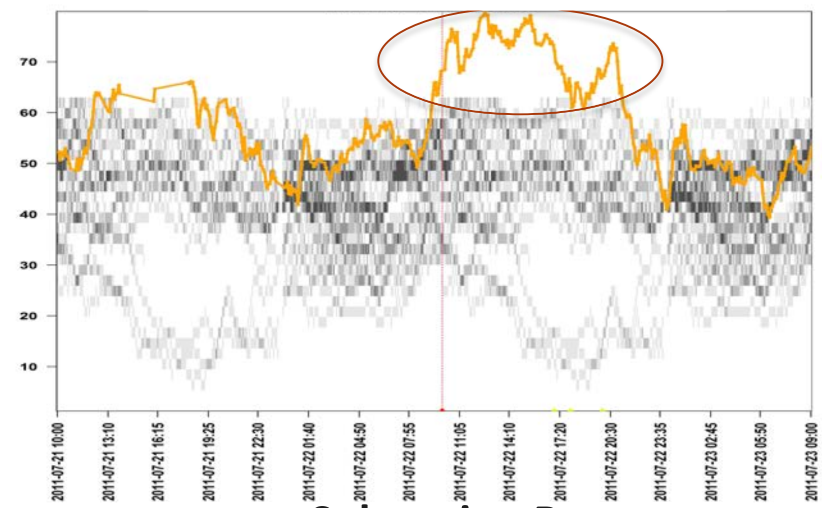
Identifying Data Driven Atypical Events

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

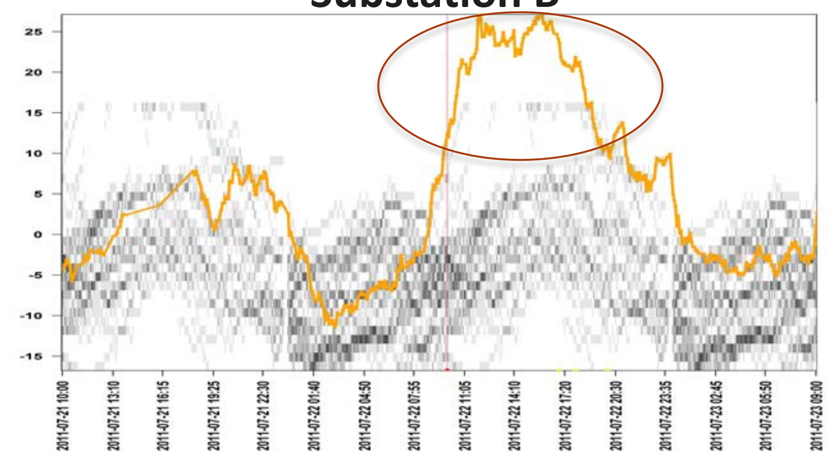
Using multivariate statistical techniques to establish baselines of typical behavior, atypical moments in time can be discovered and the responsible variables can be identified.



Substation A

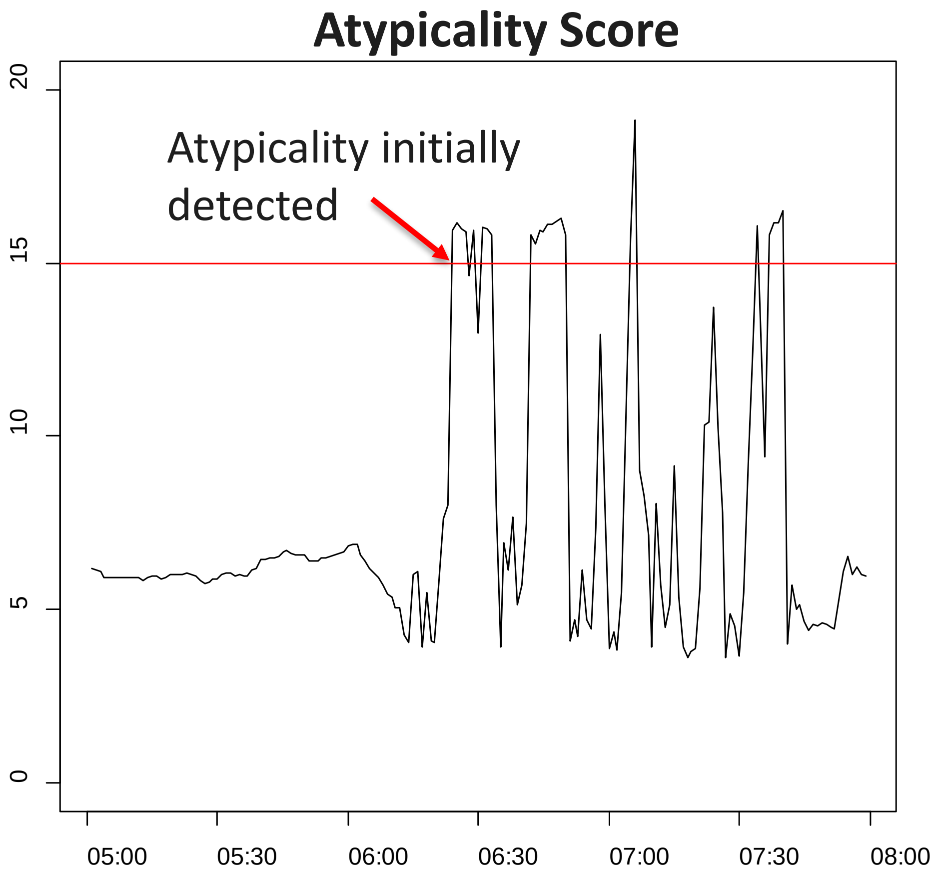


Substation B



Recent Atypicality Example

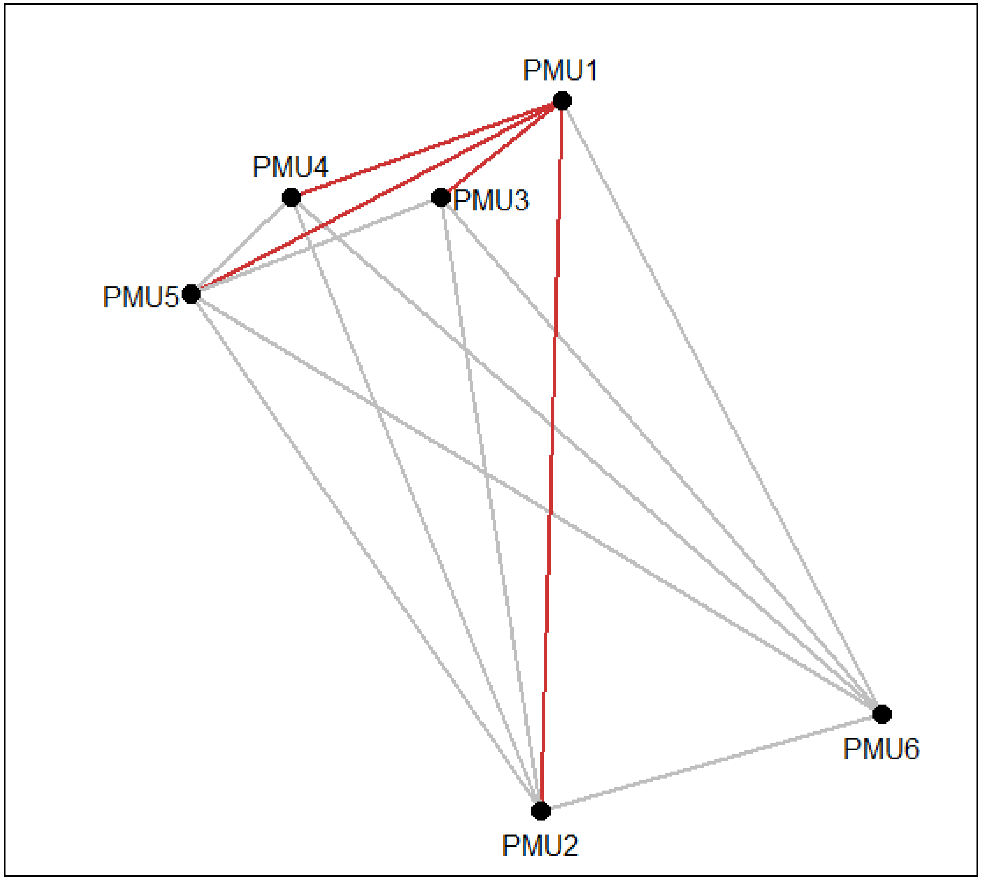
- Processed 5 months of 60 frames/sec PMU data (15 PMUs)
- This analysis only focused on phase angle pairs
- Atypicality first detected at 6:23 and then atypicality occurred off and on for the next hour or so.



All results have been de-identified

Atypicality Displays

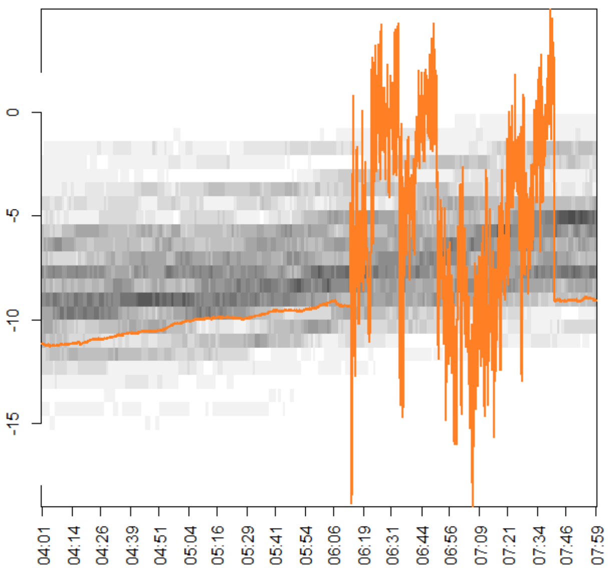
- Spatial plot showing which phase angle pairs were contributing to the atypicality (red indicates atypical)



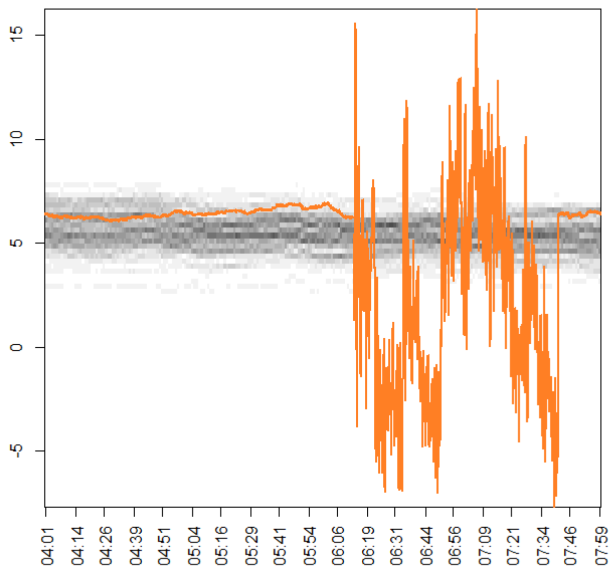
Atypicality Displays

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

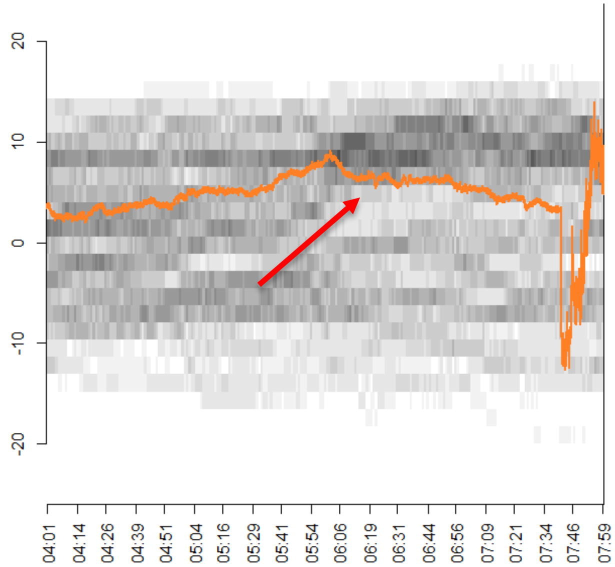
PMU1-PMU2 Pair



PMU1-PMU3 Pair



PMU1-PMU6 Pair

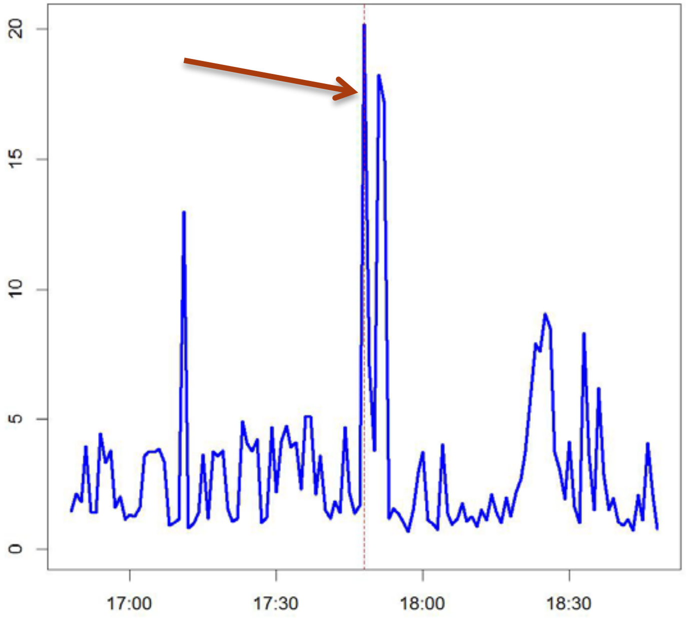


2 of the 4 angle pairs that were significant.

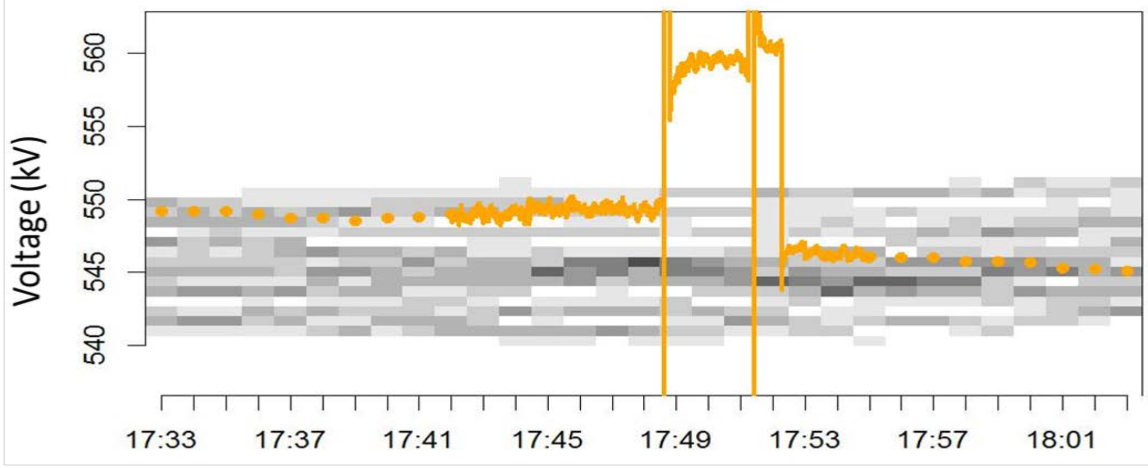
Angle pair was not significant, although it also contained PMU1.

Atypicality Detection
Lightning Related Anomaly

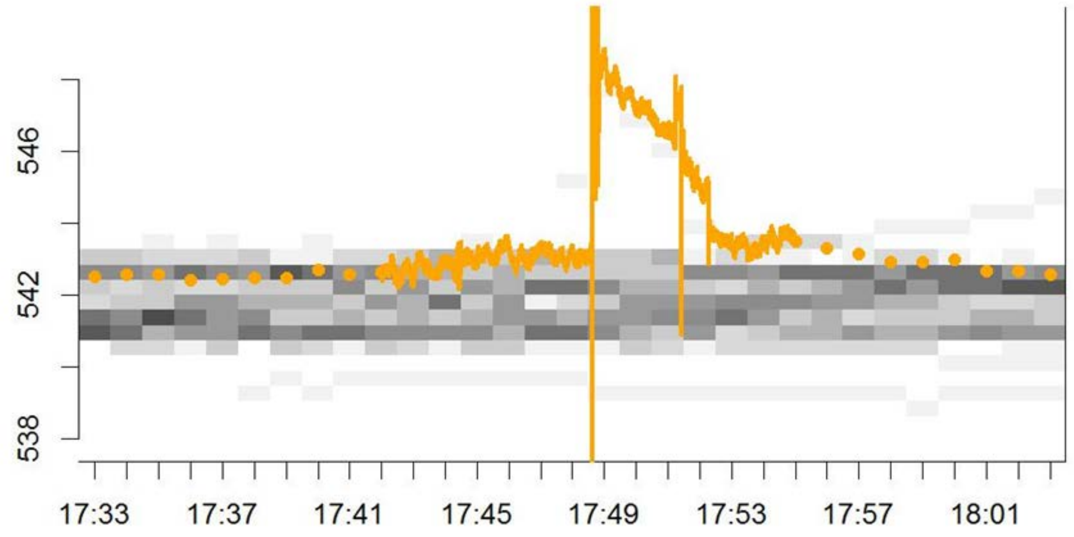
Atypicality Score



Substation A



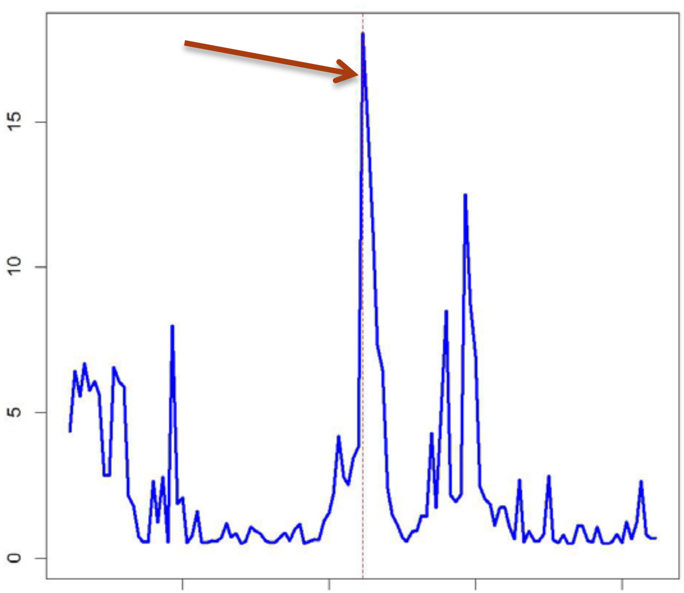
Substation B

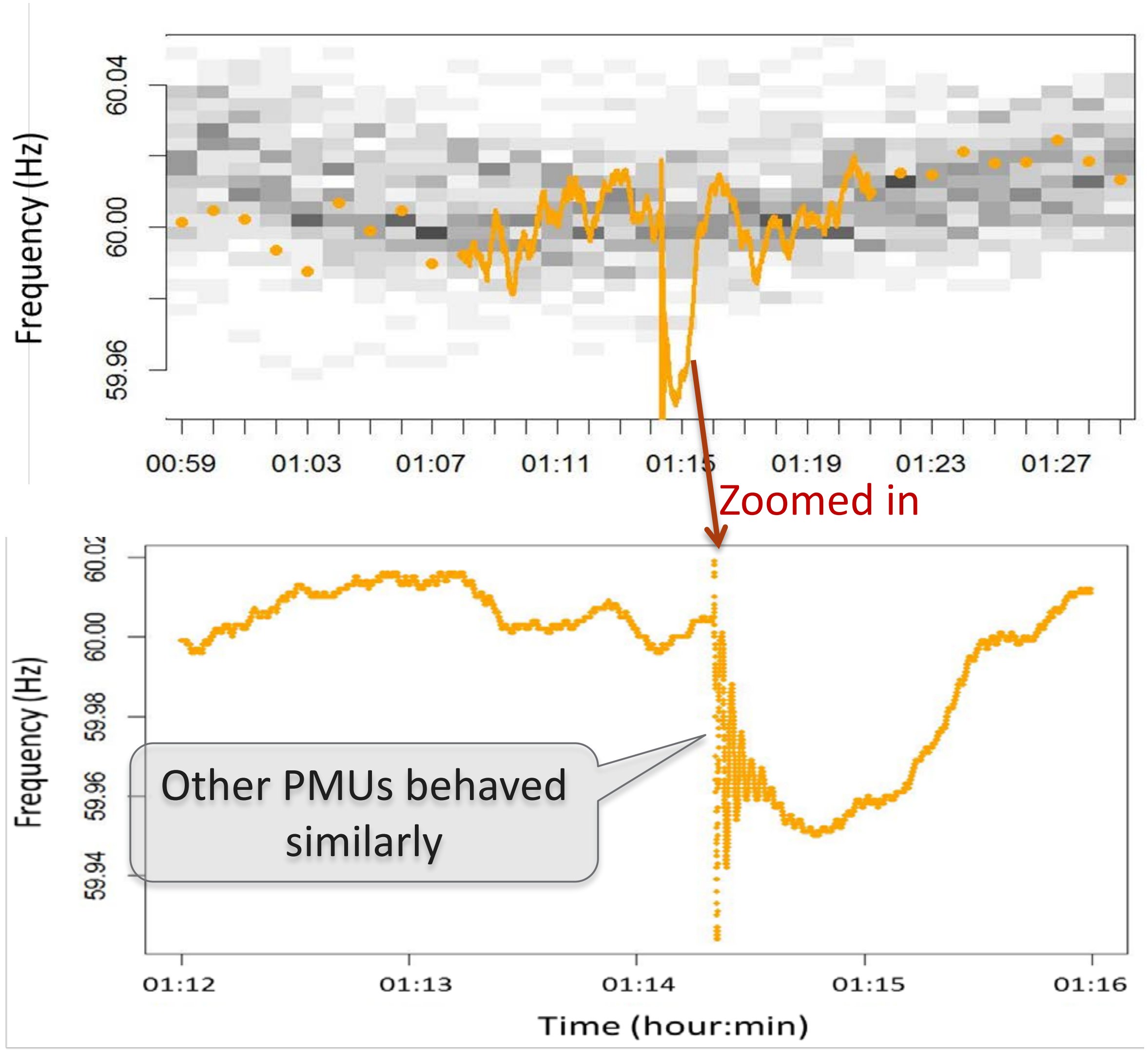


Atypicality Detection
Anomaly Related to Loss of Generation

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

Atypicality Score

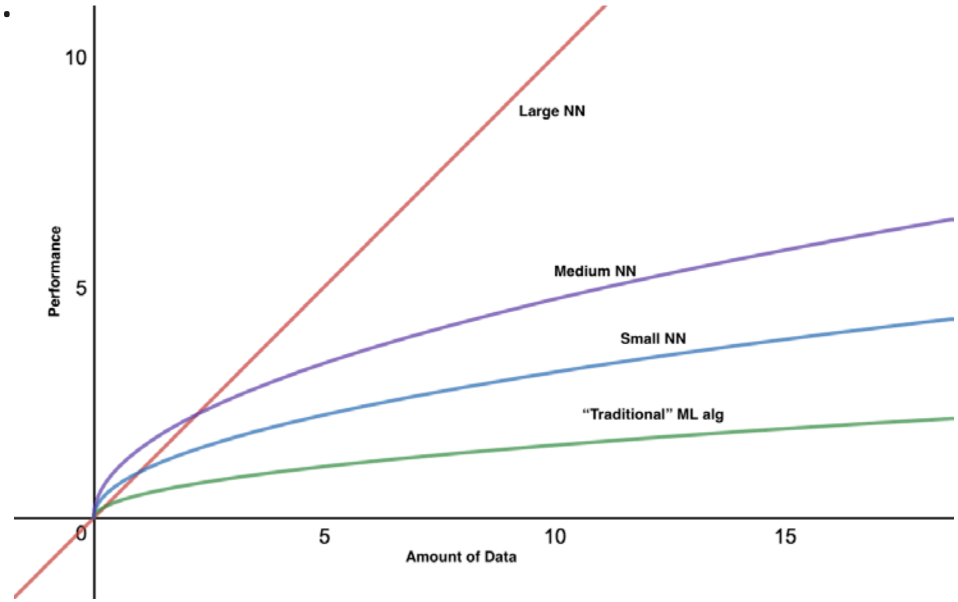




Unsupervised Learning
Neural Networks

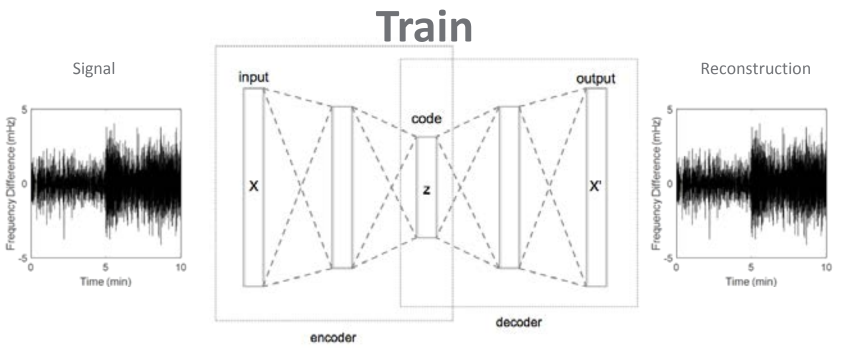
- **What** are neural networks?
  - Machine learning models that can learn highly non-linear behavior.

- **Why** we need neural networks?
  - For sufficiently large networks, performance becomes a function of the amount of data.



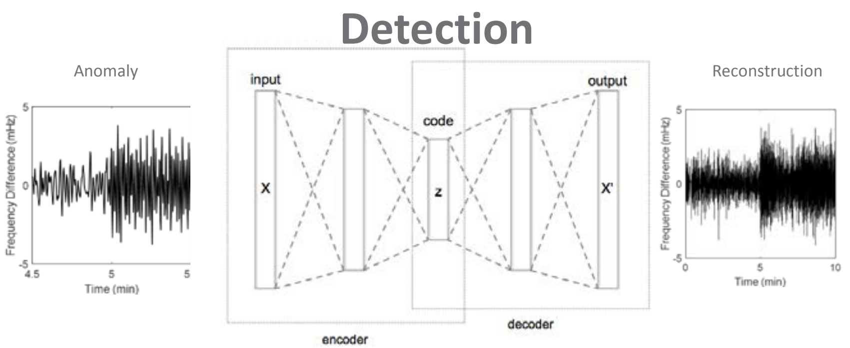
Neural Networks for Power Grid Data

- Neural networks a perfect fit for **power grid** applications because:
  - We have access to a lot of data (volume, high frequency).
  - Power grid behavior is highly non-linear.

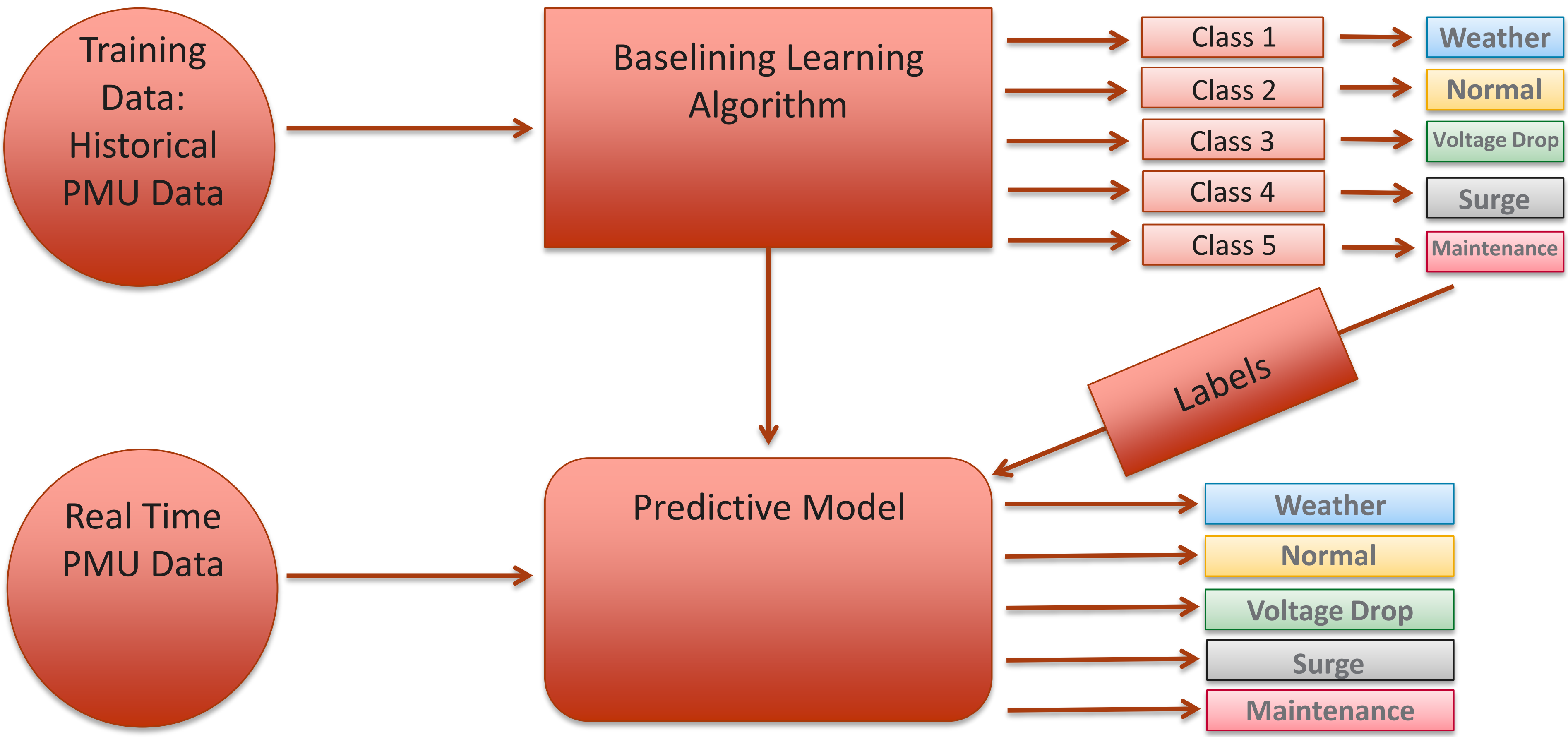


Example: Detecting Oscillations in Power System

- Oscillation events are rare and can be difficult to detect.
- **Approach**: Train a neural network (called **autoencoder**) to learn when the grid is stable.



Supervised Learning



Conclusions

- The power grid community is in the infancy stages of applying statistical and machine learning algorithms.
- Care must be taken in determining which data should be used, how features can be extracted from the data, and selecting which features will provide insight.
- Initial results show that data driven anomalies can be identified using multivariate analyses techniques. Some of these anomalies correspond to actual events, but some do not.
- The potential is great in bringing valuable insight into the power grid community by applying multivariate statistics and machine learning techniques.