Energy Systems Innovation Center

Ensemble Based Technique for Synchrophasor Data Quality and Analyzing its Impact on Applications

A. Srivastava, S. Pandey, M. Zhou, P. Banerjee, Y. Wu

Smart Grid Demonstration and Research Investigation Lab (SGDRIL)

Energy System Innovation center, Washington State University

Contact: anurag.k.srivastava@wsu.edu

NASPI MARCH 2017

Washington State University
World Class. Face to Face.

SGDRIL
Smart Grid Demonstration & Research Investigation Lab, Washington State University, Pullman

Outline

Synchrophasor Data Quality (SDQ)

Ensemble based Methodology for SDQ Detection

- Architecture of the developed technique
- Base Detectors and Normalization
- The proposed Maximum Likelihood Estimator (MLE)
- Inference and Outlier Detection

Effect of Data Quality on Power System Applications

- Assessment Methodology
- Effect of Data on Event Detection Algorithm
- Effect of Data on Load Modelling Application
- Effect of Bad data on Stability Application

Results

Summary

Outline

Synchrophasor Data Quality (SDQ)

Ensemble based Methodology for SDQ Detection

- Architecture of the developed technique
- Base Detectors and Normalization
- The proposed Maximum Likelihood Estimator (MLE)
- Inference and Outlier Detection

Effect of Data Quality on Power System Applications

- Assessment Methodology
- Effect of Data on Event Detection Algorithm
- Effect of Data on Load Modelling Application
- Effect of Bad data on Stability Application

Results

Summary

Synchrophasor Data Quality (SDQ)

<u>Factors that cause SDQ issues</u>

- The synchrophasor provides high sampling data which are prone to anomalies.
- These measurement anomalies can be outlier or a missing data.
- These results due to loss of data, GPS sync problem, incorrect measurements etc.

<u>Impact of SDQ on Applications</u>

- The anomalous data can lead to performance deterioration of various application using PMU measurements.
- Once enough PMUs are deployed, it is very important to ensure the quality of PMU data as control actions might depend upon it.

<u>Solutions</u>

- The traditional Bad Data detection algorithm do not perform satisfactorily for data driven applications as real time requirement is a priority.
- An Ensemble Based Technique, with less or no parameter tuning, improves performance using unsupervised learning method to detect data anomalies.

Outline

Synchrophasor Data Quality (SDQ)

Ensemble based Methodology for SDQ Detection

- Architecture of the developed technique
- Base Detectors and Normalization
- The proposed Maximum Likelihood Estimator (MLE)
- Inference and Outlier Detection

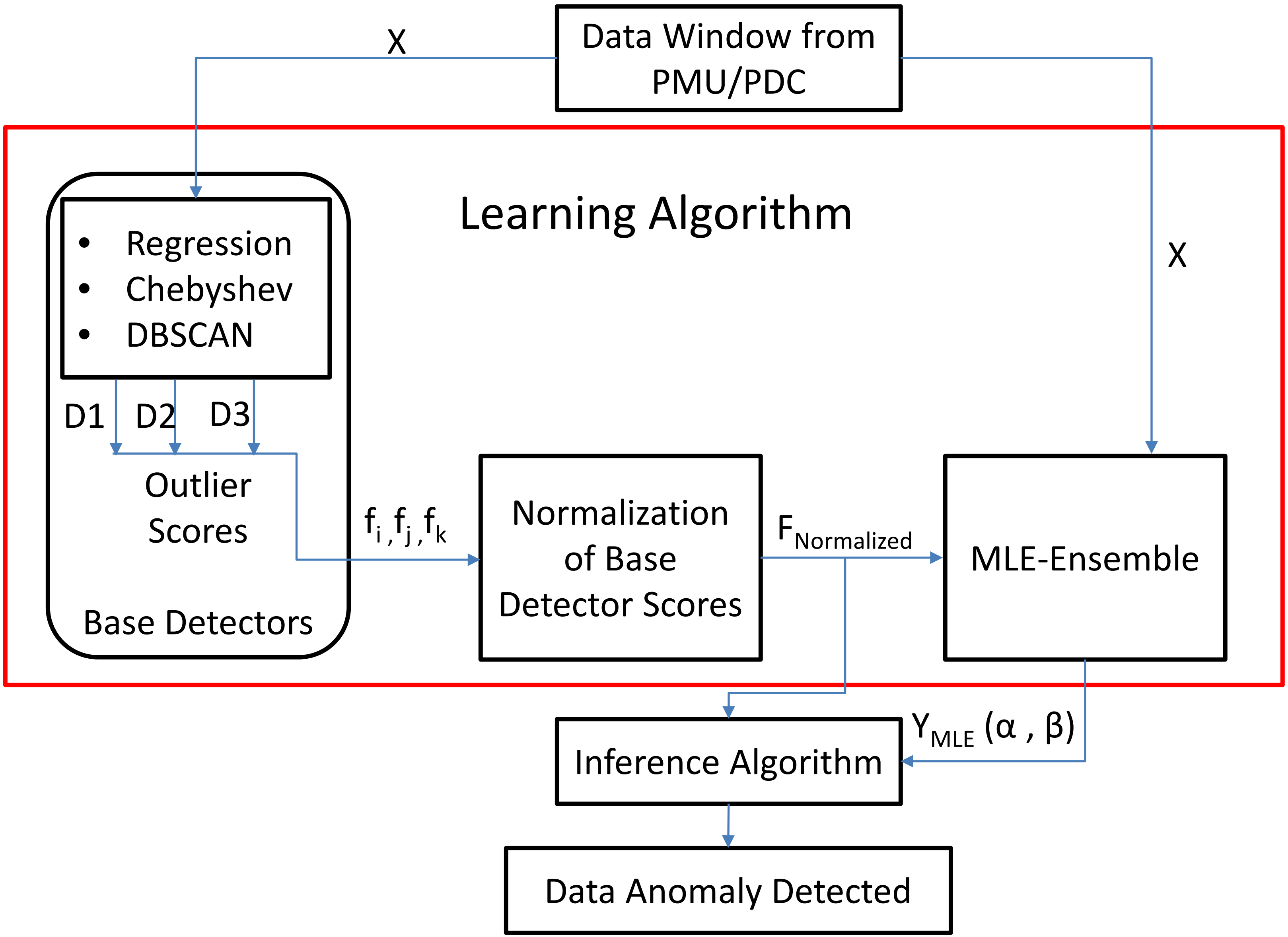
Effect of Data Quality on Power System Applications

- Assessment Methodology
- Effect of Data on Event Detection Algorithm
- Effect of Data on Load Modelling Application
- Effect of Bad data on Stability Application

Results

Summary

Architecture of the developed technique



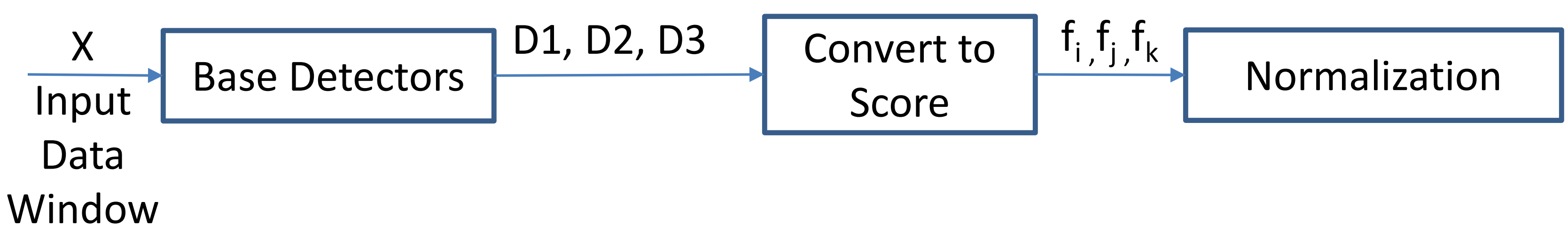
Base detectors

In this work three base detectors have been considered:

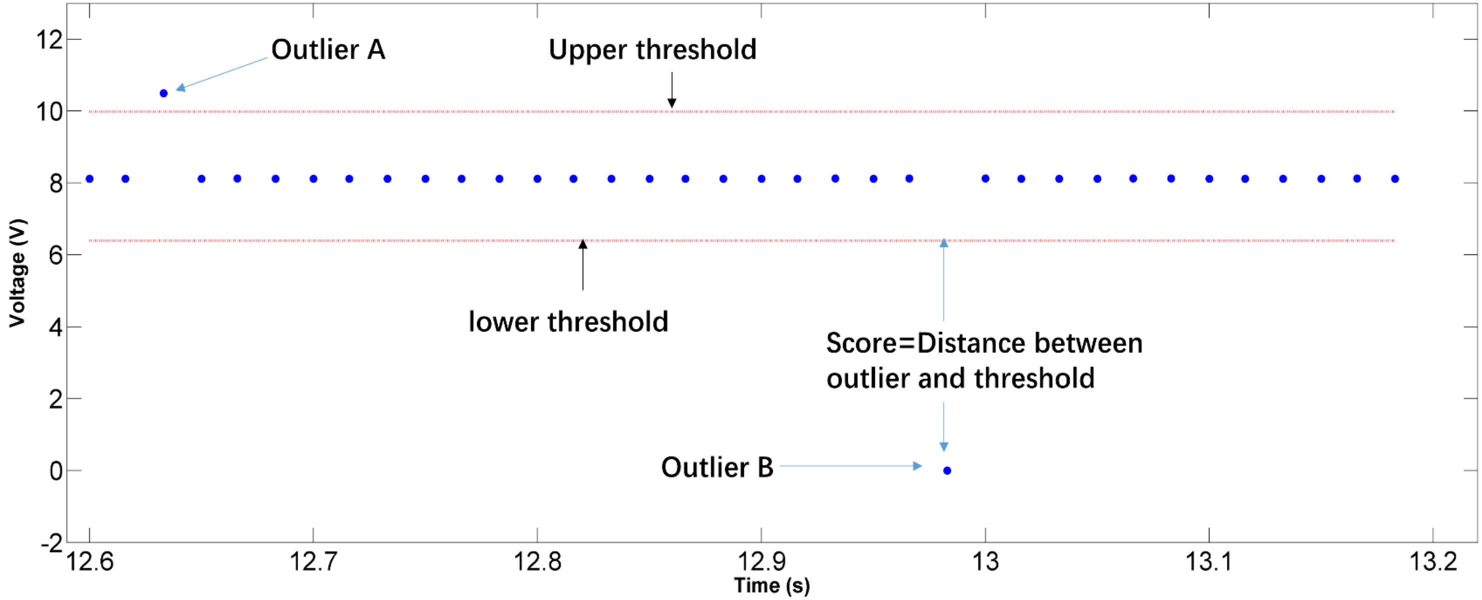
1. Chebyshev
2. Regression
3. DBSCAN

"No single winner" i.e. if all the above mentioned base detectors are applied individually desired level of accuracy could not be acheived.

Each Base Detectors give their output of bad data as D1, D2, D3. Which is then converted into score f_i, f_j, f_k.



Converting Detected Points into Numbers (Scores)



Example: Score definition for Chebyshev

- A and B are the two outlier points in the present window.
- In this case, the score is determined by the distance between outlier and the threshold. The longer the distance, the larger the score, so the point is more likely to be an anomaly data.
- The "score" for each base detector are calculated using the distance criteria .

Normalization

The score for each base detector is converted to probability estimates [1]:

1. Probability of outlier for point 'i' using Chebyshev : $P_{chebyshev}(O \mid f_i)$
where, f_i is the corresponding outlier score.

Similarly probability using linear regression and DBSCAN is given by,
2. Linear Regression : $P_{regression}(O \mid f_i)$

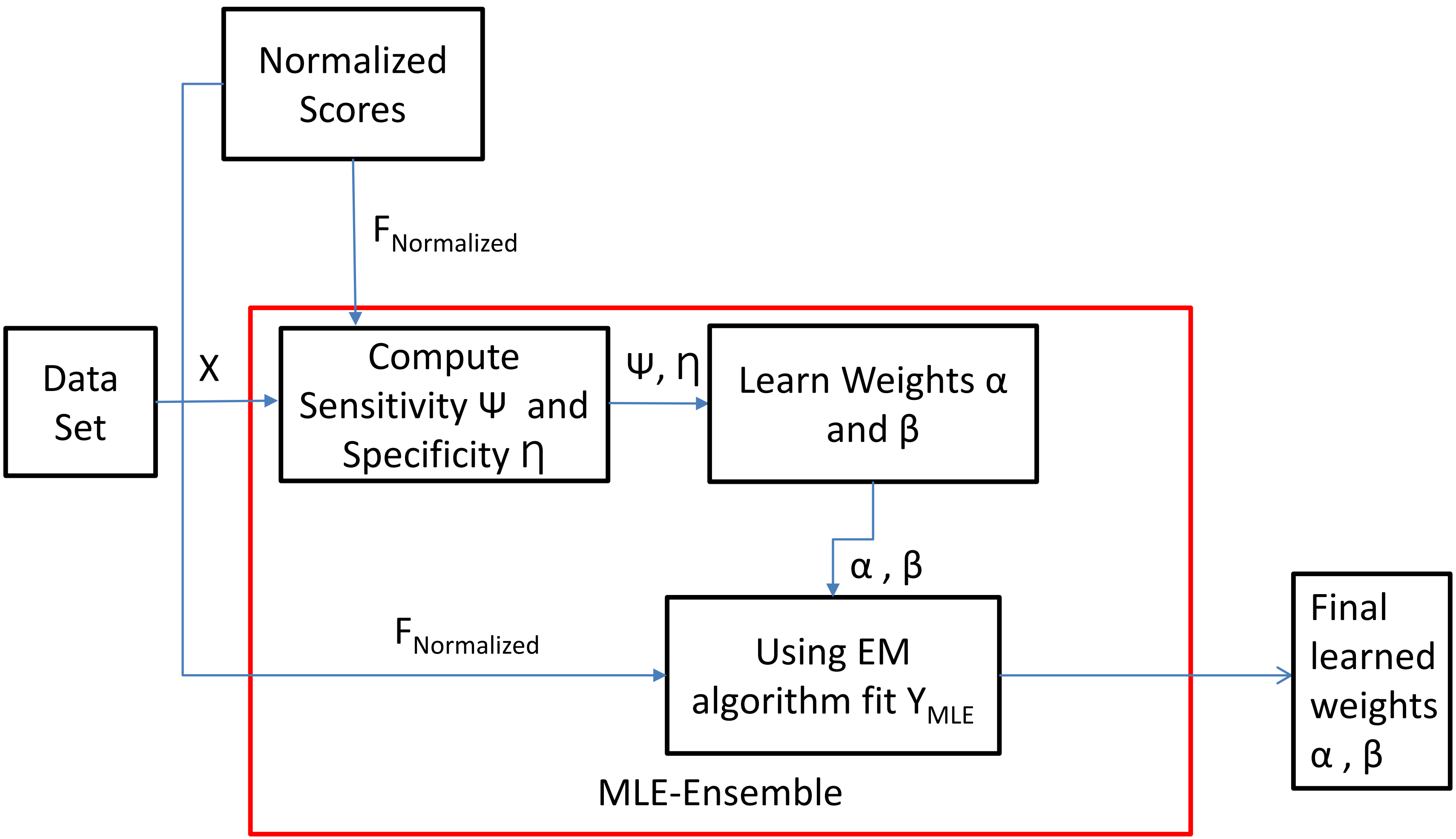
3. DBSCAN: $P_{DBSCAN}(O \mid f_i)$

The vector of probabilities for three detector is denoted as $F_{Normalized}$

$$f_i, f_j, f_k \rightarrow \boxed{\text{Normalization}} \rightarrow F_{Normalized}$$

[1] Jing Gao; Pang-ning Tan, " Converting Output Scores from Outlier Detection Algorithms into Probability Estimates"
..\..\PMU bad date detection_paper\coding\J.C.platt. probabilistic outputs (two parameter optimization problem)\converting scores to probability estimates.pdf

Maximum Likelihood Estimator (MLE)



Maximum Likelihood Estimator (MLE)

Compute:
Sensitivity ψ_i which is the fraction of correctly identified outliers

And Specificity η_i which is the fraction of correctly identified non-outliers

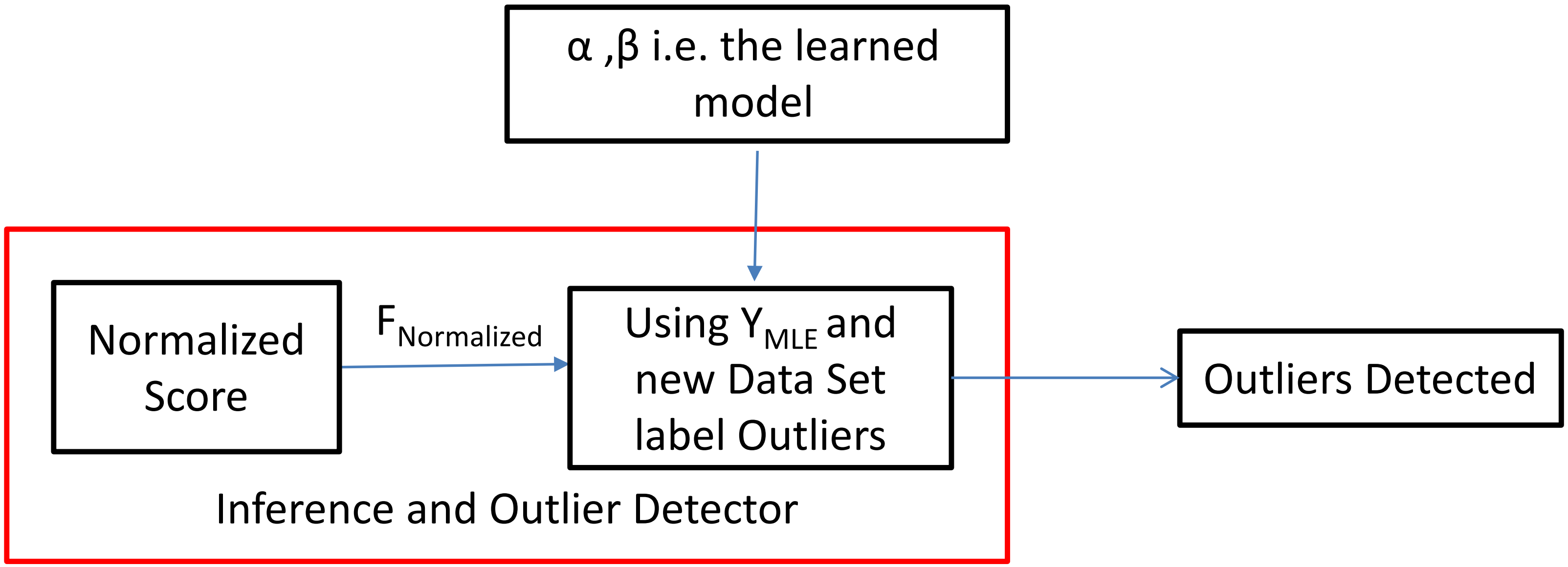
Calculate Weights: $$\alpha_i = \frac{\psi_i \eta_i}{(1-\psi_i)(1-\eta_i)}, \quad \beta_i = \frac{\psi_i(1-\psi_i)}{\eta_i(1-\eta_i)}.$$

Calculate weighted score: $$\begin{aligned} \hat{y}^{(\mathrm{ML})} &= \operatorname*{argmax}_y \mathfrak{L}(f_1(x), \ldots, f_M(x); y) \\ &= \operatorname{sign}\Big(\sum_{i=1}^{M} f_i(x) \log \alpha_i + \log \beta_i\Big) \end{aligned}$$

Inference and Outlier Detection

- After the MLE-Ensemble step, weights of each base detector is learned which is Y_{MLE} .

- Now on the new data set using these weights and the Normalized scores of the base detectors the inference algorithm makes decision on bad data.



Outline

Synchrophasor Data Quality (SDQ)

Ensemble based Methodology for SDQ Detection

- Architecture of the developed technique
- Base Detectors and Normalization
- The proposed Maximum Likelihood Estimator (MLE)
- Inference and Outlier Detection

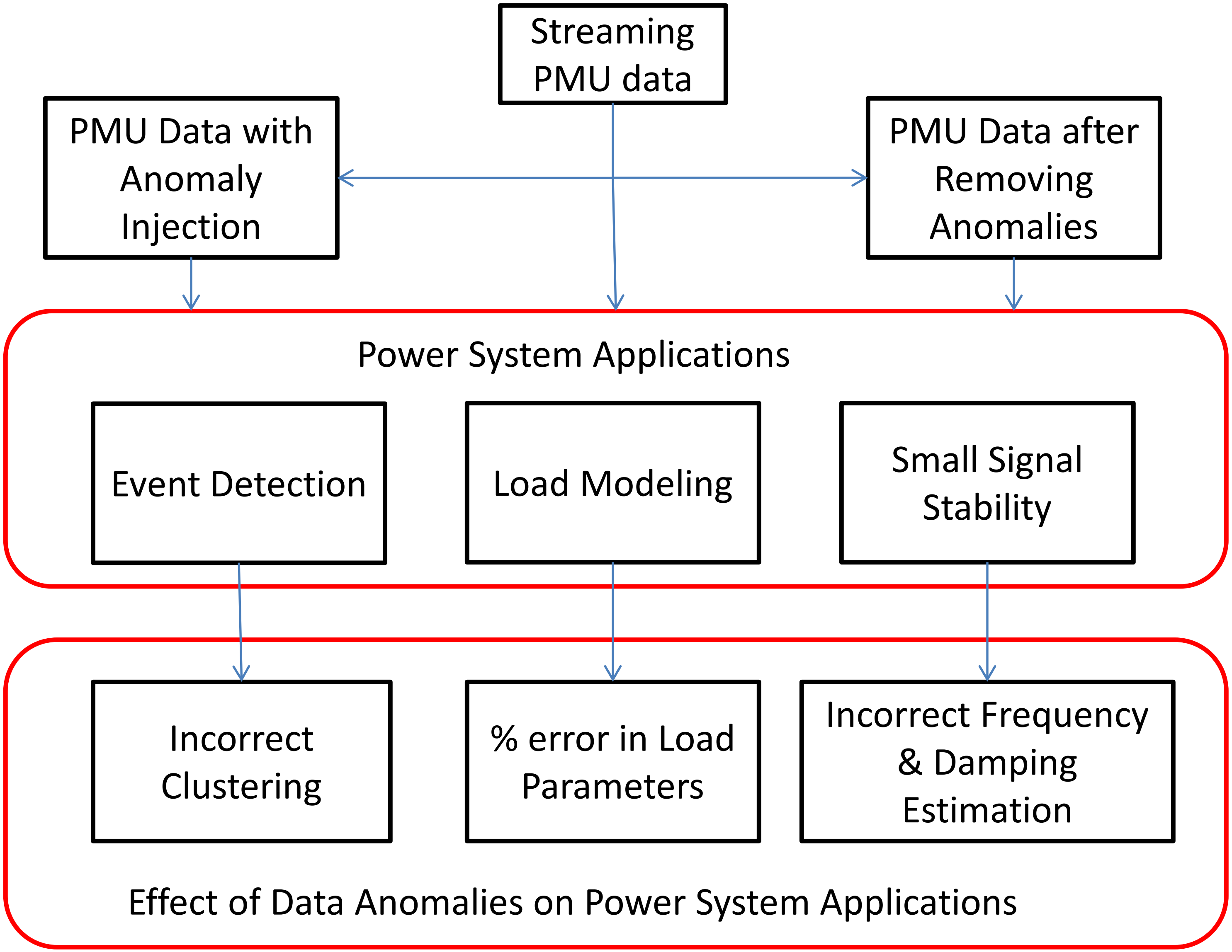
Effect of Data Quality on Power System Applications

- Assessment Methodology
- Effect of Data on Event Detection Algorithm
- Effect of Data on Load Modelling Application
- Effect of Bad data on Stability Application

Results

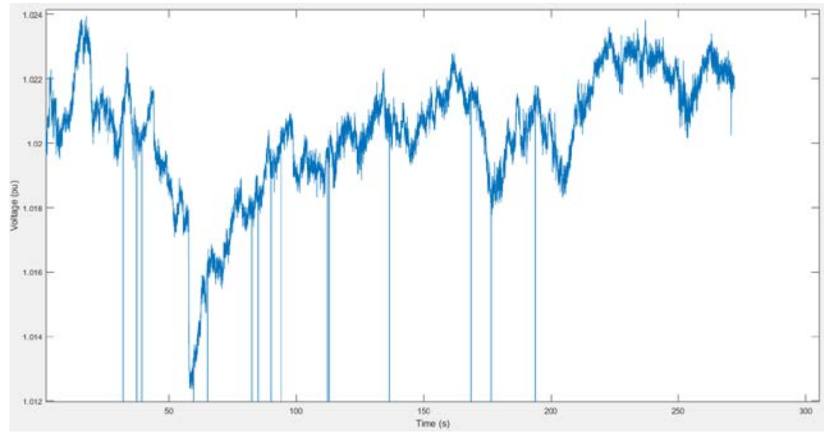
Summary

Assessment Methodology

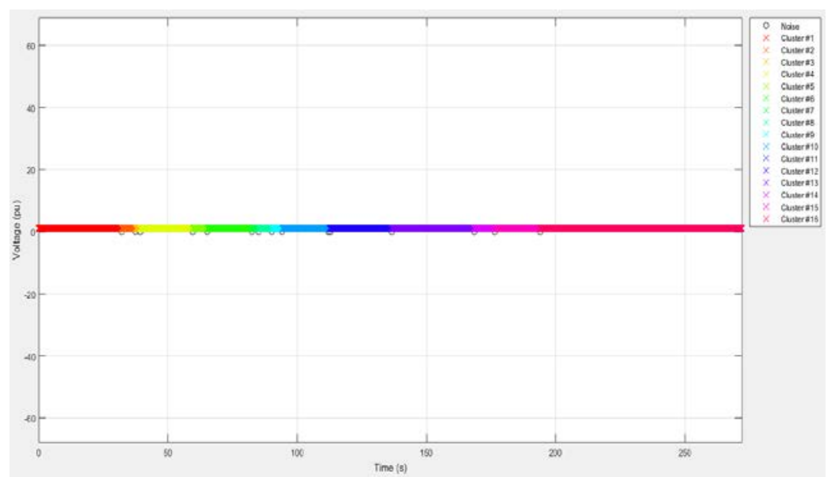


Effect of Bad Data on Event Detection Algorithm

- There are several missing and bad data in the voltage measurement as observed in this figure.
- The instances of missing and bad data results in incorrect clustering (16 clusters Detected) and hence badly prepared for event detection.
- These cluster in voltages suggest that an event capable of causing voltage change has occurred.



Raw Industry Voltage Data

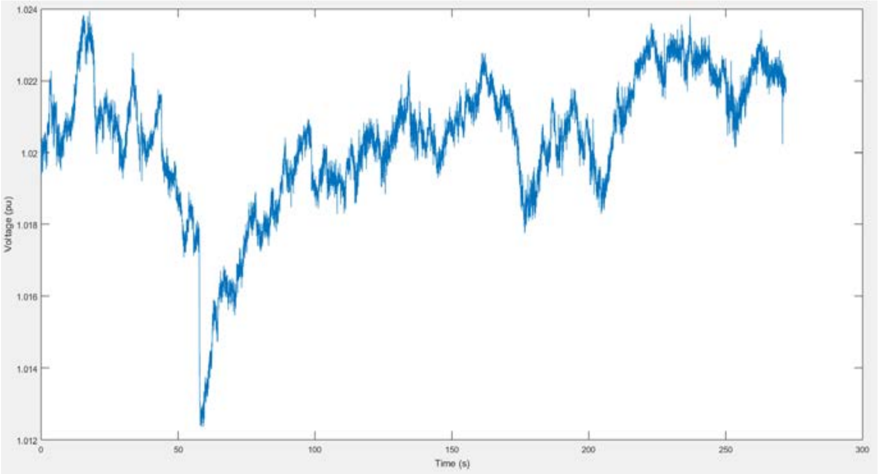


Clustering on Industry Raw Voltage Data

*Thanks to EPRI

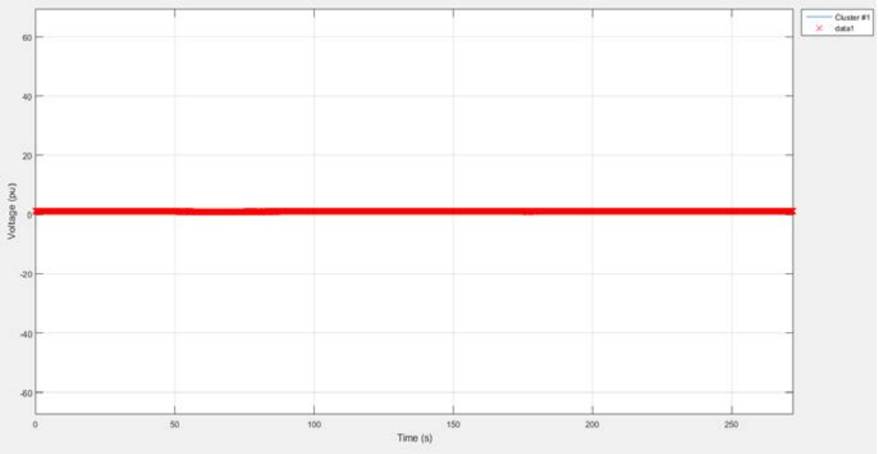
Effect of Bad Data on Event Detection Algorithm

- The anomaly detection algorithm filters missing data and replaces bad data.



Clean Industry Voltage Data

- Different clusters which are detected previously, are grouped as single cluster and hence no wrong event detection



Clustering on Clean Industry Voltage Data

Effect of Bad Data on Load Modeling

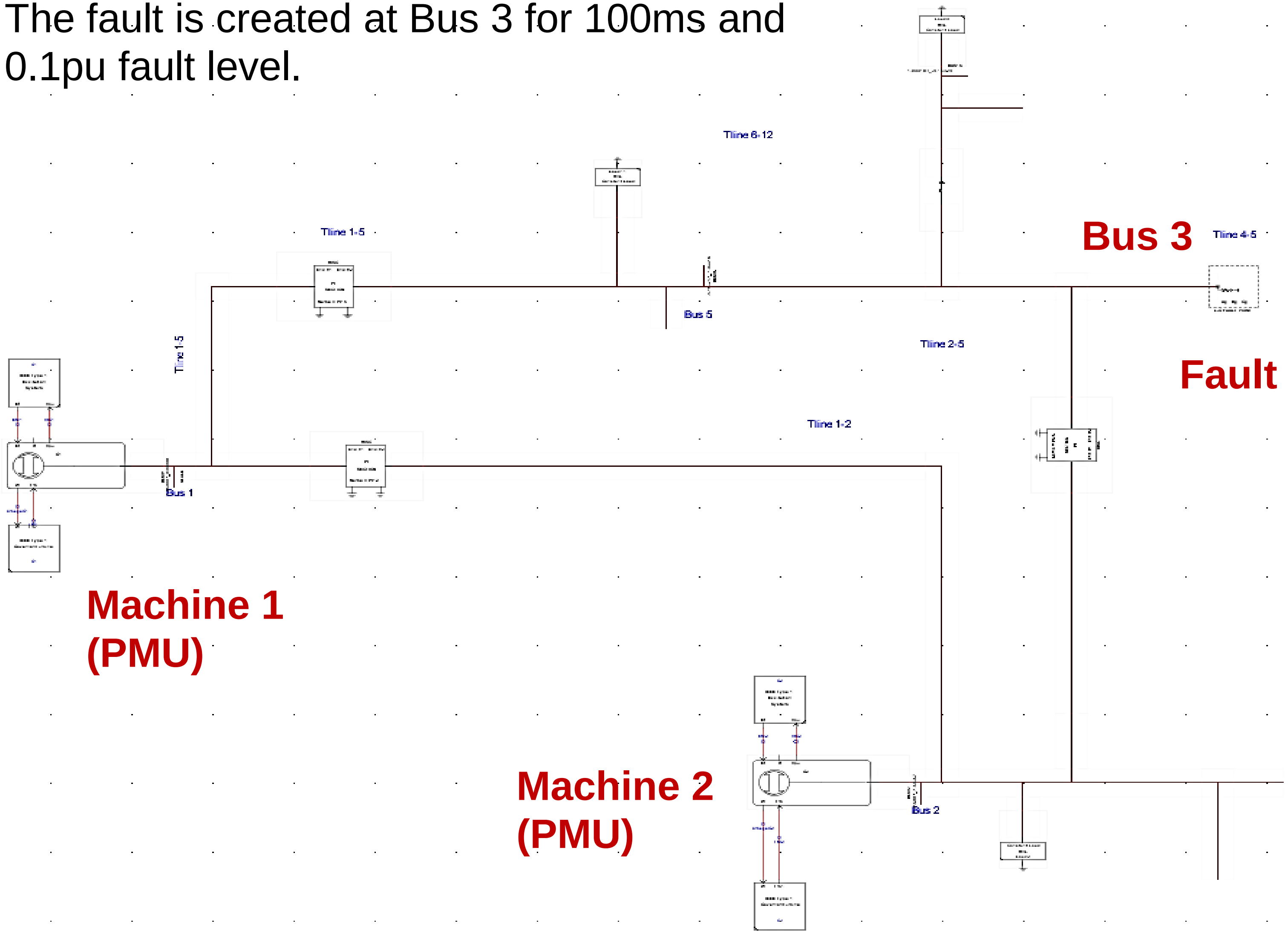
- ZIP estimation error is less for PMU data without anomaly for both Least Square and Constrained Least Square methods.
- Maximum PMU data anomaly of 4% magnitude is incorporated with 0.4% anomaly density.
- PMU data anomaly results in the error of ZIP estimation to increase 3 folds for certain cases.
- Detected anomalies are removed and replaced by interpolated value.

Estimation Method	**Error**	**PMU data without anomaly**	**4% anomaly magnitude & 0.4% density**	**Ensemble based cleaned data**
Constrained least square	% Error in Z	1.53	27.05	1.51
	% Error in I	2.39	40.9	2.37
	% Error in P	0.73	12.07	0.72

Effect of Bad Data on Load Modeling

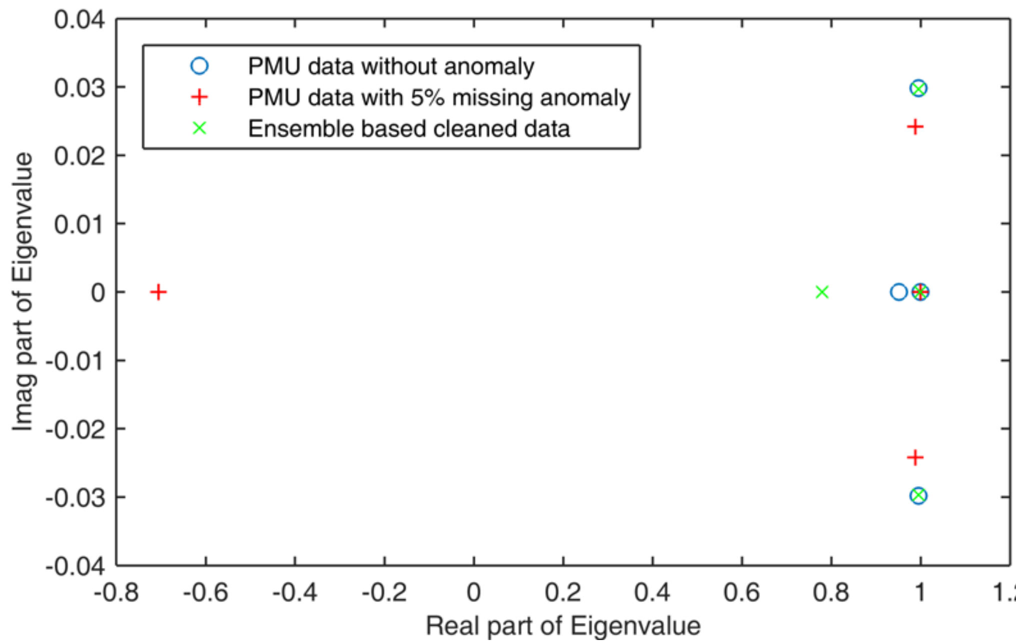
Effect of Bad Data on online Small Signal Stabilty Analysis

The fault is created at Bus 3 for 100ms and 0.1pu fault level.



Effect of Anomalous Data on online Small Signal Stabilty Analysis

- The eigen values without anomalies are in the right side of the plane.
- It can be seen that the missing data changes the location of the eigen values with one in left part.
- The ensemble based clean data brings the eigenvalue close to the original eigenvalues.



Effect of bad data on Small Signal Stability

Outline

Synchrophasor Data Quality (SDQ)

Ensemble based Methodology for SDQ Detection

- Architecture of the developed technique
- Base Detectors and Normalization
- The proposed Maximum Likelihood Estimator (MLE)
- Inference and Outlier Detection

Effect of Data Quality on Power System Applications

- Assessment Methodology
- Effect of Data on Event Detection Algorithm
- Effect of Data on Load Modelling Application
- Effect of Bad data on Stability Application

Results

Summary

Simulation Results

Tests on the RTDS simulated PMU data (1.5 hours, 5% bad data points, 5%-10% range)

	Recall	Precision	False positive
Linear Regression	0.9021	0.8565	0.1435
DBSCAN	0.8821	0.8821	0.1179
Chebyshev	0.9154	0.8754	0.1246
MLE ensemble	0.9251	0.8913	0.1087

Tests on the RTDS simulated PMU data (1.5 hours, 10% bad data points, 10%-20% range)

	Recall	Precision	False positive
Linear Regression	0.7854	0.7655	0.2345
DBSCAN	0.7216	0.7015	0.2985
Chebyshev	0.8125	0.7542	0.2458
MLE ensemble	0.8912	0.9021	0.0979

- Recall = *Detected Bad Data ∩ Actual Bad Data/Actual Bad Data*
- Precision= *Detected Bad Data ∩ Actual Bad Data/Detected Bad Data*
- False Positive = 1 - *Detected Bad Data ∩ Actual Bad Data/Detected*

Outline

Synchrophasor Data Quality (SDQ)

Ensemble based Methodology for SDQ Detection

- Architecture of the developed technique
- Base Detectors and Normalization
- The proposed Maximum Likelihood Estimator (MLE)
- Inference and Outlier Detection

Effect of Data Quality on Power System Applications

- Assessment Methodology
- Effect of Data on Event Detection Algorithm
- Effect of Data on Load Modelling Application
- Effect of Bad data on Stability Application

Results

Summary

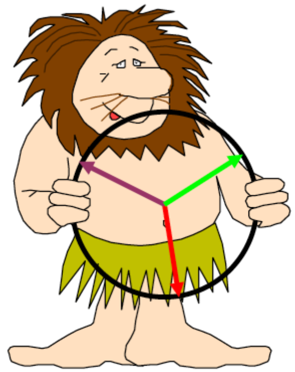
Summary

- PMU data with bad measurements can result in failure of Power System Applications.
- A single method is not sufficient to solve the bad data problem.
- An integrated methods called as base detectors is required.
- The MLE-Emsemble produces a result equivalent to or better than all the base detectors.

Advantage of the ensemble based method:-

- Plug in more base detectors as needed, which will continuously improve the method.
- Little or no effort in parameter tuning
- Unsupervised, but works well with more training data
- Established and well-supported data mining and machine learning.

- This algorithm can work for real time streaming PMU, and can help in realizing close-loop Control Systems.
- Integrating bad data detection into applications can result into ‘quality-aware applications’




SGDRIL
Smart Grid Demonstration & Research Investigation Lab, Washington State University, Pullman