Real-Time Model-Free Detection of Low-Quality Synchrophasor Data

Meng Wu and Le Xie

Department of Electrical and Computer Engineering
Texas A&M University
College Station, TX

NASPI Work Group meeting
March 23, 2017

Presentation Outline

1 **Introduction**

2 **Technical Approach**

3 **Case Studies**

4 **Conclusions & Future Work**

TEXAS A&M UNIVERSITY

Motivation: PMU Data Quality Problems

Current Practice

- PMU-based decision making tools require accurate PMU data for reliable analysis.
- PMU data has **higher sampling rate** and accuracy requirement.
- Typical PMU bad data ratio in California ISO ranges from **10% to 17%** (in 2011) [1].

Critical Needs

- Urgent need to develop **scalable, real-time methods** to monitor and improve PMU data quality.
- **Conventional** bad data detection **algorithms** are rendered **ineffective**, novel algorithms are needed.

- [1] California ISO, "Five year synchrophasor plan," California ISO, *Tech. Rep*., Nov 2011.

Current Approaches for PMU Bad Data Detection

Model-Based Approach

- PMU-based state estimator [2].
- Kalman-filter-based approach [3].
- Require system parameter and topology information.
- Require converged state estimation results.

Data-Driven Approach

- Low-rank matrix factorization for PMU bad data detection [4].
- Pre-defined logics & thresholds for bad data detection [5].
- Matrix factorization involves high computational burden.
- Robustness of pre-defined logics under eventful conditions.

[2] S. Ghiocel, J. Chow, et al. "Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring".
[3] K. D. Jones, A. Pal, and J. S. Thorp, "Methodology for performing synchrophasor data conditioning and validation".
[4] M. Wang, J. Chow, et al. "A low-rank matrix approach for the analysis of large amounts of power system synchrophasor data".
[5] K. Martin, "Synchrophasor data diagnostics: detection & resolution of data problems for operations and analysis".

Overview of The Proposed Approach [6]

Problem Formulation

- Study *spatio-temporal correlations* among good / eventful / bad PMU data.
- Formulate bad PMU data as *spatio-temporal outliers* among other data.
- Apply *density-based outlier detection* technique to detect bad PMU data.

Online PMU Bad Data Detection Algorithm

Key Advantages:

- Online bad data detection.
- Fast without convergence issues.
- Data-driven algorithm.
- Operate under both normal and fault-on operating conditions.

Detect Various Types of Bad Data:

- High communication noise.
- Missing data (communication loss).
- Data spikes (gross error / GPS error).
- Un-updated data.
- Cyber attacks (false data injection).

TEXAS A&M UNIVERSITY

- [6] M. Wu, and Le Xie. Online Detection of Low-Quality Synchrophasor Measurements: A Data-Driven Approach. *IEEE Transactions on Power Systems*. Accepted, to appear.

Presentation Outline

TEXAS A&M UNIVERSITY

Good Data VS Eventful Data VS Bad Data

Phase Angle Measured by A Western System PMU for A Recent Brake Test Event

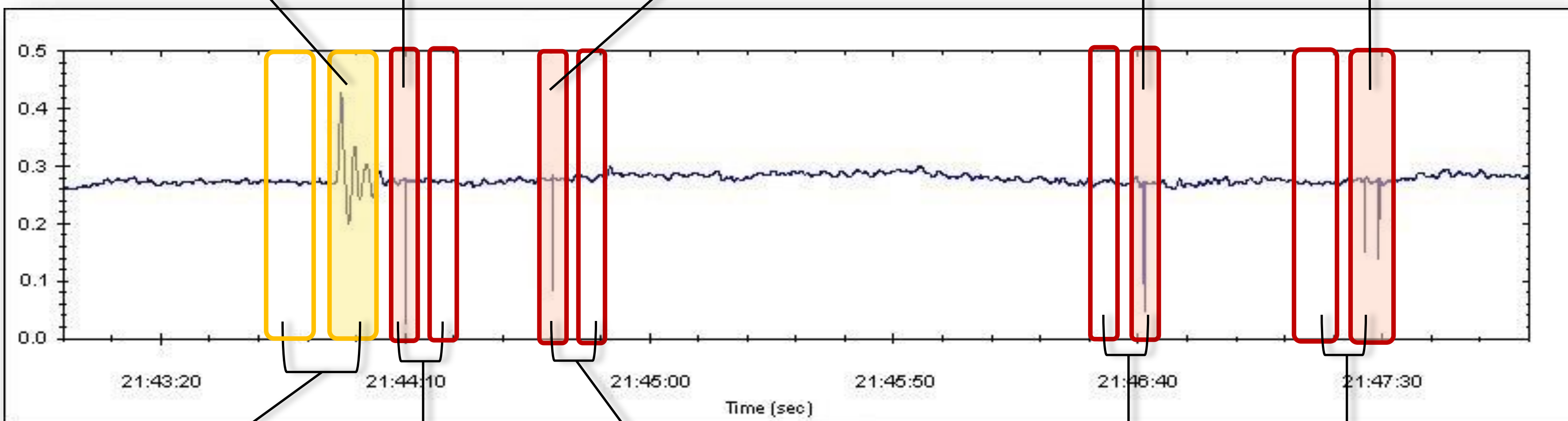
Event

Bad Data

Bad Data

Bad Data

Bad Data



Weak Temporal Correlation

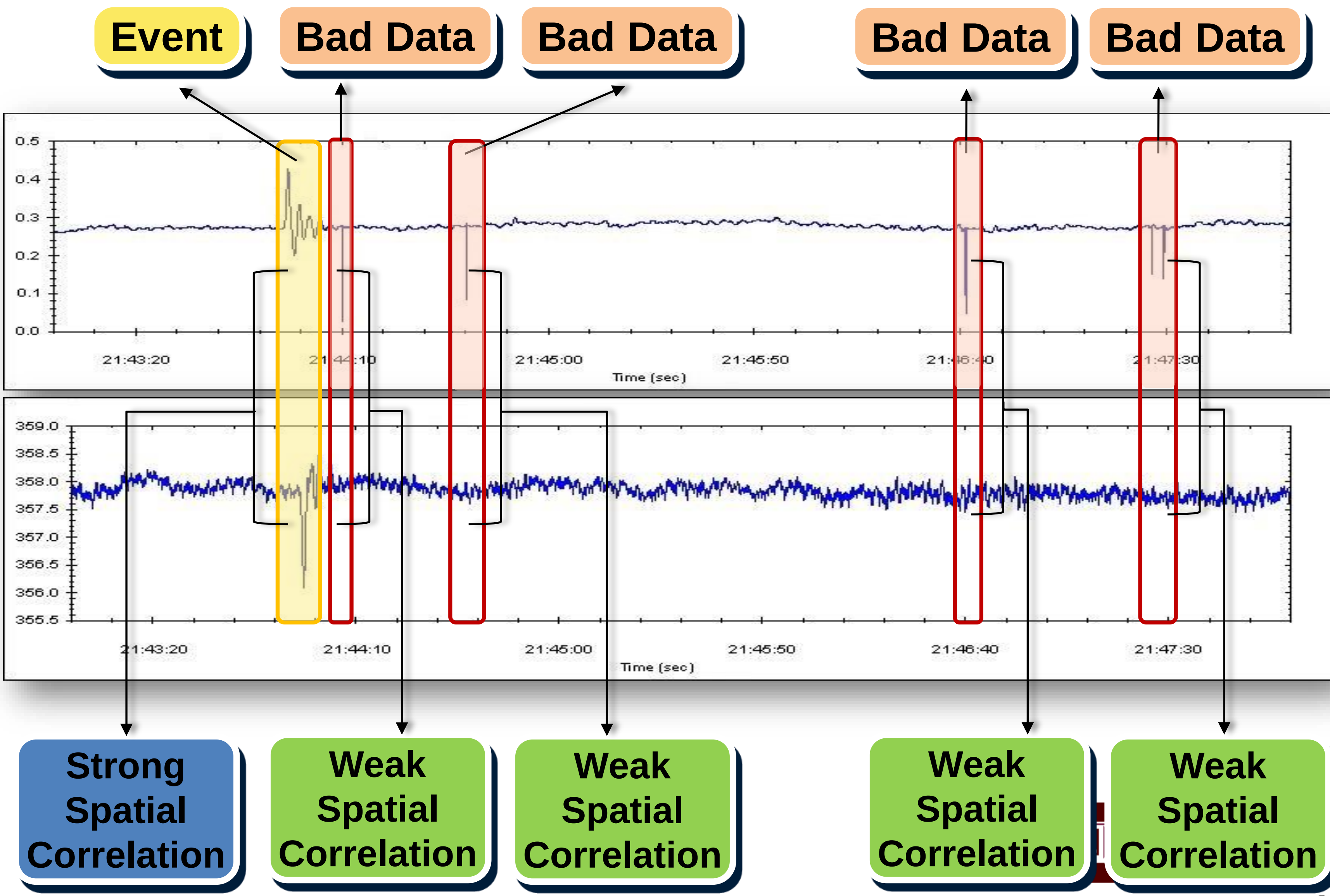
Weak Temporal Correlation

Weak Temporal Correlation

Weak Temporal Correlation

Weak Temporal Correlation

Good Data VS Eventful Data VS Bad Data

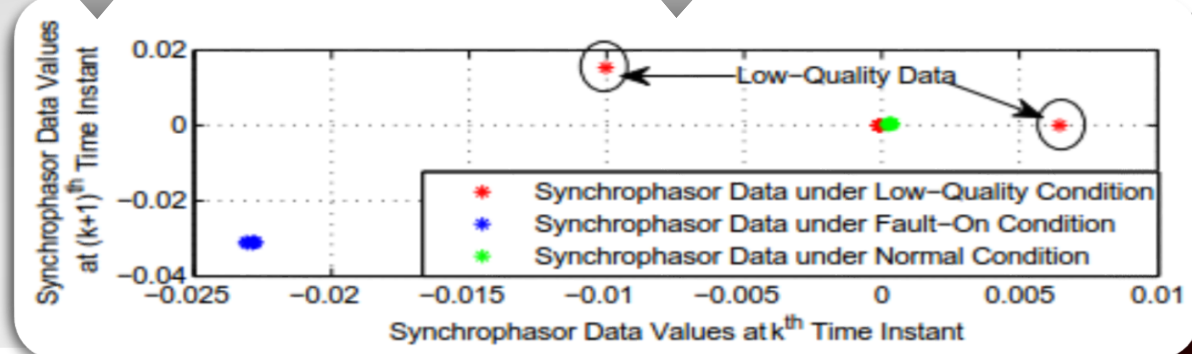


Features of Good / Eventful / Bad Data

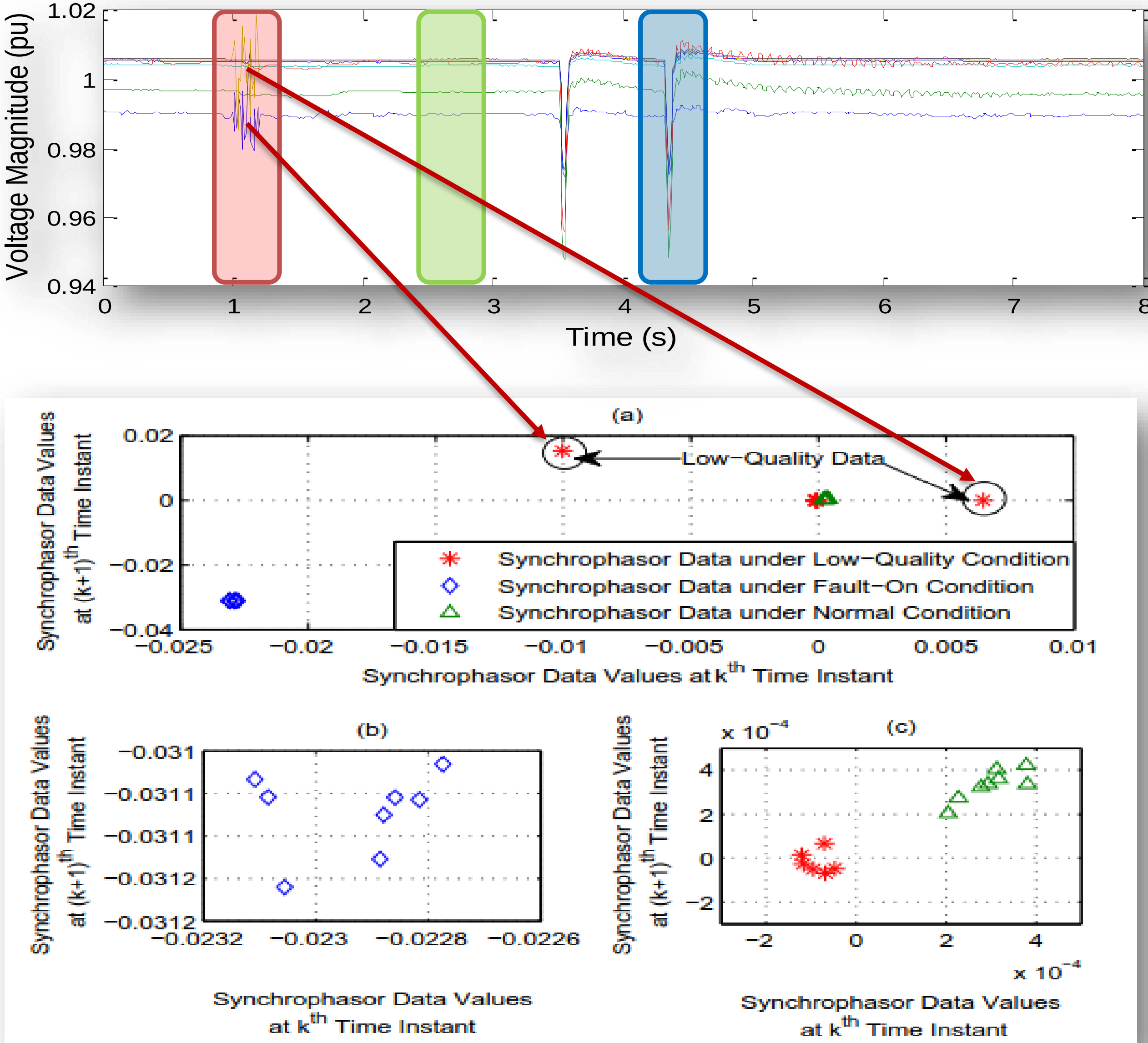
Criteria: Good Data VS Eventful Data VS Bad Data

- **Good Data**: **strong spatio-temporal correlations** with its neighbors.
- **Eventful Data**: **weak temporal but strong spatial correlations** with its neighbors.
- **Bad Data**: ***weak spatio-temporal correlations*** with its neighbors.

PMU Bad Data: Spatio-Temporal Outlier



Example of Spatio-Temporal Correlations



- ❖ 8 PMU curves
- ❖ 3 time windows
- ❖ 2 instants / window

Map 3×8 Curves To 2D Euclidian Space

- ❖ 8 green points (normal window)
- ❖ 8 blue points (fault-on window)
- ❖ 8 red points (low-quality window)
- ❖ 2 red outliers (low-quality data)

Quantification of Spatio-Temporal Correlations [6]

Definition of Normalized Standard Deviation

- **Normalized standard deviation:**

$$\sigma_i^{Norm}(k) = \frac{\sigma_i(k)}{\frac{\sum_{t=1}^{t=k-1} \sigma_i(t)\chi_C(M_i(t))}{\sum_{t=1}^{t=k-1} \chi_C(M_i(t))}}$$

$$\chi_C(M_i(t)) = \begin{cases} 1 & (M_i(t) \in C) \\ 0 & (M_i(t) \notin C) \end{cases}$$

- **Explanation:**

✓ Standard deviation of PMU curve obtained from i^{th} PMU channel at k^{th} time window, normalized by the average standard deviation of the historical clean data of the same PMU channel.

Spatio-Temporal Correlation Metrics (Distance Function)

- **For high-variance bad data:**

$$f_H(i,j) = \left|\sigma_i^{Norm} - \sigma_j^{Norm}\right|$$

✓ High-variance bad data: data spikes, data loss, high noise, false data injections, etc.

- **For low-variance bad data:**

$$f_L(i,j) = max\left(\left|\frac{\sigma_i^{Norm}}{\sigma_j^{Norm}}\right|, \left|\frac{\sigma_j^{Norm}}{\sigma_i^{Norm}}\right|\right)$$

✓ Low-variance bad data: un-updated data, etc.

- [6] M. Wu, and Le Xie. Online Detection of Low-Quality Synchrophasor Measurements: A Data-Driven Approach. *IEEE Transactions on Power Systems*. Accepted, to appear.

Online Detection of Low-Quality PMU Data [7]

Spatio-Temporal Correlation Metrics (Distance Function)

- For high-variance bad data:

$$f_H(i,j) = \left|\sigma_i^{Norm} - \sigma_j^{Norm}\right|$$

✓ High-variance bad data: data spikes, data loss, high noise, false data injections, etc.

- For low-variance bad data:

$$f_L(i,j) = max\left(\left|\frac{\sigma_i^{Norm}}{\sigma_j^{Norm}}\right|, \left|\frac{\sigma_j^{Norm}}{\sigma_i^{Norm}}\right|\right)$$

✓ Low-variance bad data: un-updated data, etc.

Density-Based Local Outlier Detection

- Local Reachability Density:

$$lrd_{MinPts}(p) = 1/\left(\frac{\sum_{o \in N_{MinPts}(p)} reach-dist_{MinPts}(p,o)}{|N_{MinPts}(p)|}\right)$$

- Local Outlier Factor [12]:

$$LOF_{MinPts}(p) = \frac{\sum_{o \in N_{MinPts}(p)} \frac{lrd_{MinPts}(o)}{lrd_{MinPts}(p)}}{|N_{MinPts}(p)|}$$

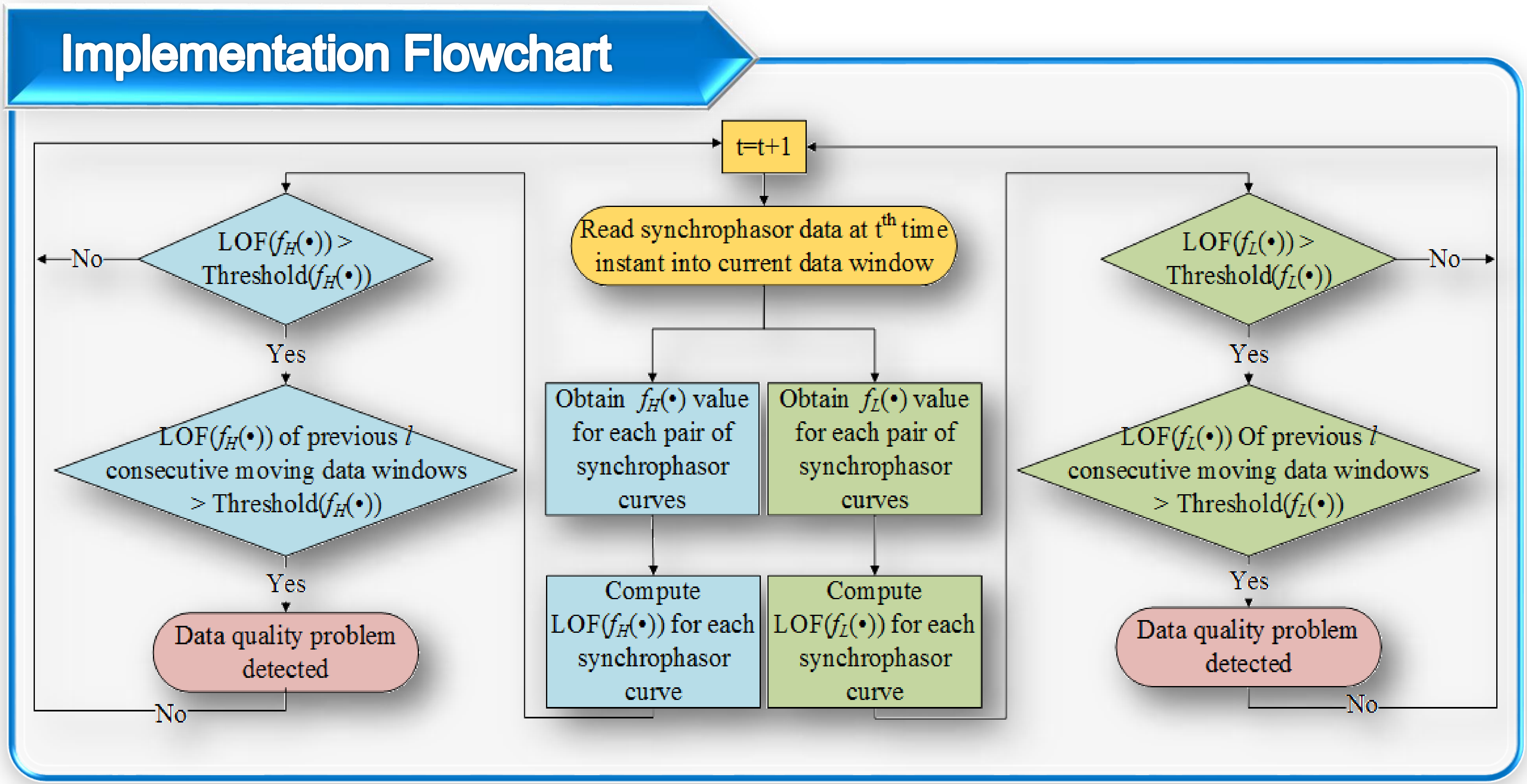
- Bad Data Detection:

✓ LOF(p) >> 1: p contains bad data.

✓ LOF(p) ≈ 1: p contains good data only.

- [7] Breunig, Markus M., et al. "LOF: identifying density-based local outliers." *ACM sigmod record*. Vol. 29. No. 2. ACM, 2000.

Online Detection of Low-Quality PMU Data [6]



- [6] M. Wu, and Le Xie. Online Detection of Low-Quality Synchrophasor Measurements: A Data-Driven Approach. *IEEE Transactions on Power Systems*. Accepted, to appear.

Presentation Outline

TEXAS A&M UNIVERSITY

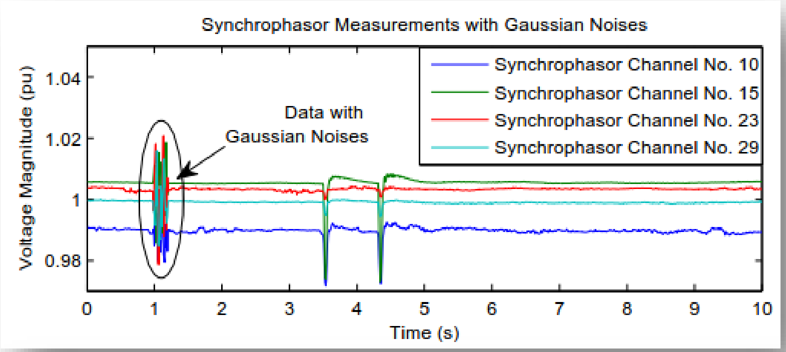
Numerical Results – High Sensing Noise

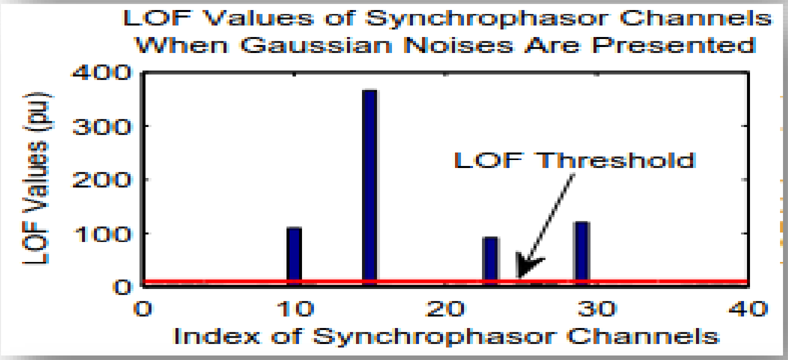
Test Case Description

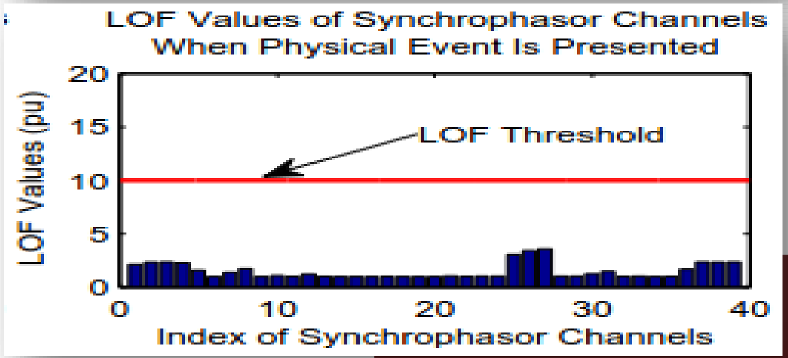
- **39 real-world PMU voltage magnitude data curves.**
- **PMU No. 10, 15, 23, 29 contain Gaussian noises (SNR = 40 db) lasting from 1s to 1.2s.**
- **Line tripping fault is presented around 4s.**

Numerical Results Description

- **All the 4 bad data segments are detected.**
- **System event does not cause false alarms.**
- **Detection delay is less than 0.19s.**
- **Computation time for each data window is 0.0376s.**







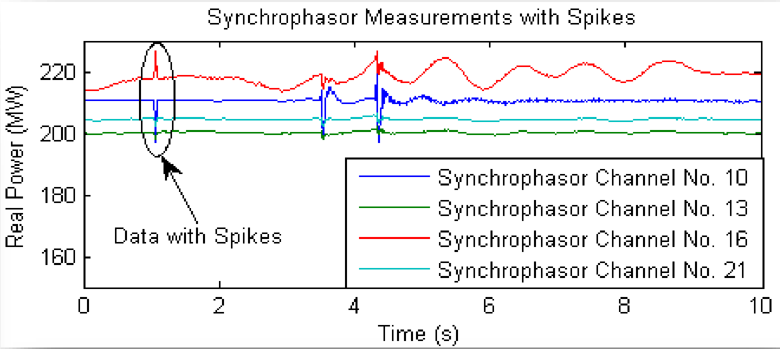
Numerical Results – Data Spikes

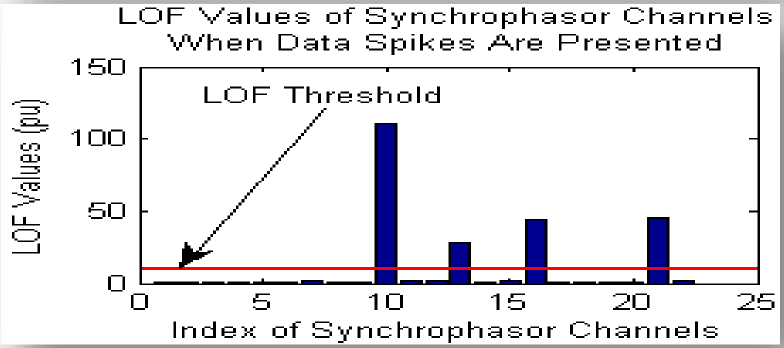
Test Case Description

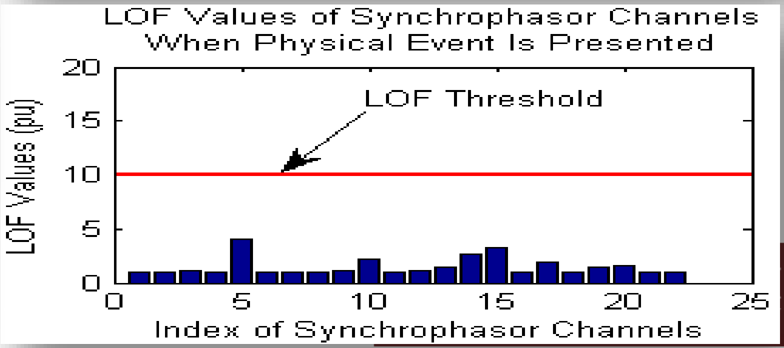
- 22 real-world PMU real power data curves.
- PMU No. 10, 13, 16, 21 contain data spikes lasting from 1.05s to 1.1s.
- Line tripping fault is presented around 4s.

Numerical Results Description

- All the 4 bad data segments are detected.
- System event does not cause false alarms.
- Detection delay is less than 0.18s.
- Computation time for each data window is 0.0197s.







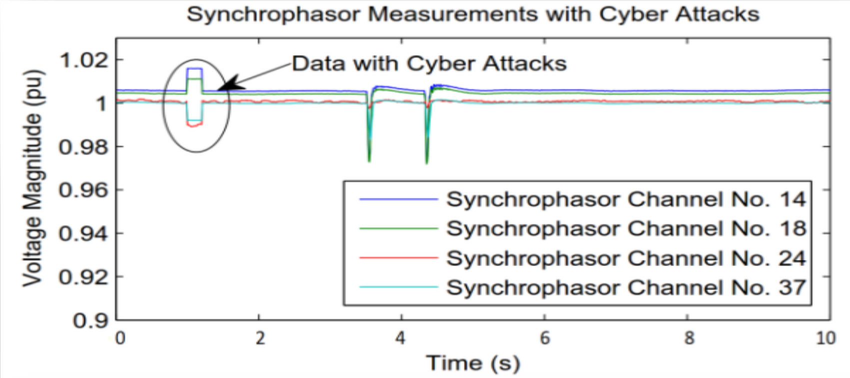
Numerical Results – False Data Injection Attacks

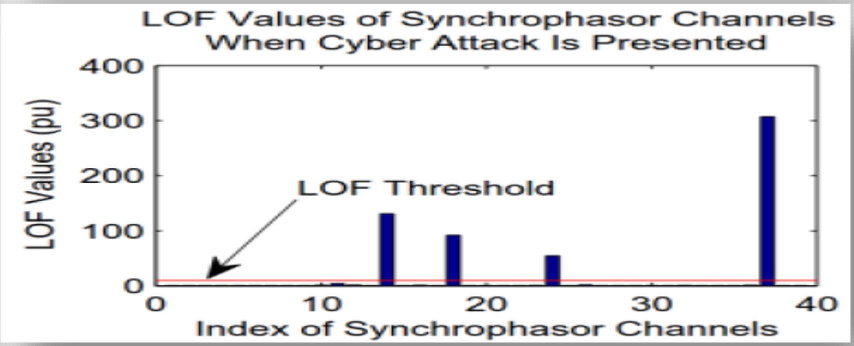
Test Case Description

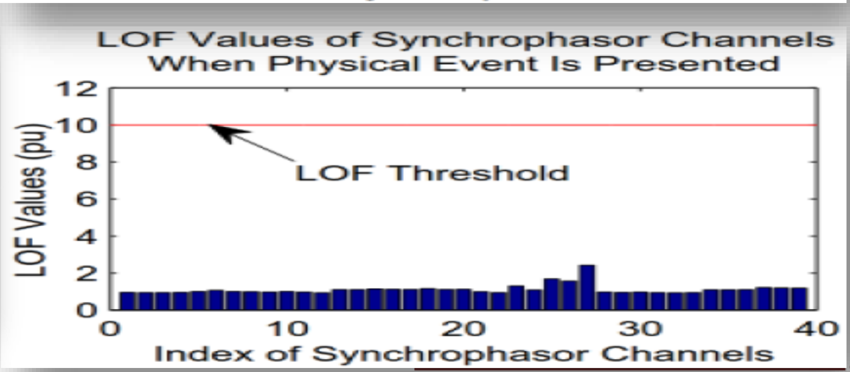
- **39 real-world PMU voltage magnitude data curves.**
- **PMU No. 14, 18, 24, 37 contain false data injections lasting from 1s to 1.2s.**
- **Line tripping fault is presented around 4s.**

Numerical Results Description

- **All the 4 false data injections are detected.**
- **System event does not cause false alarms.**
- **Detection delay is less than 0.19s.**
- **Computation time for each data window is 0.040s.**







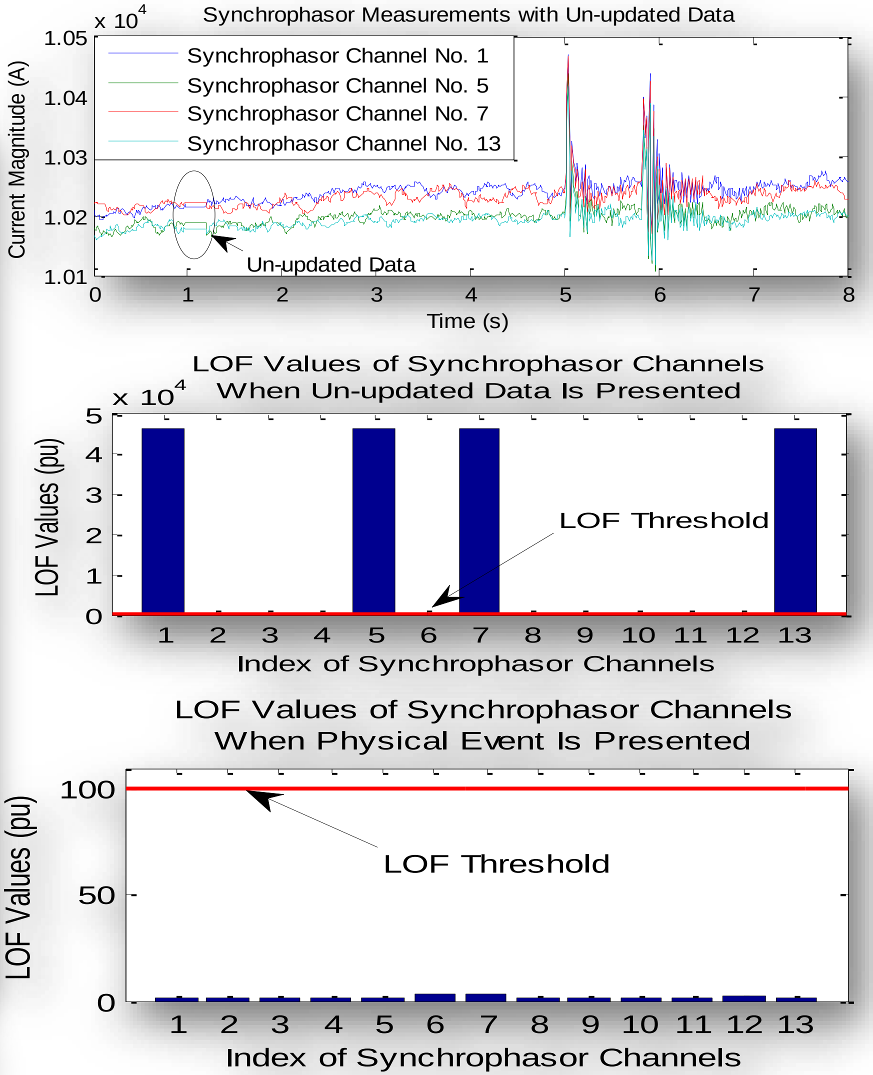
Numerical Results – Un-updated Data

Test Case Description

- **13 real-world PMU current magnitude data curves.**
- **PMU No. 1, 5, 7, 13 contain un-updated data lasting from 1s to 1.2s.**
- **Line tripping fault is presented around 4s.**

Numerical Results Description

- **All the 4 bad data segments are detected.**
- **System event does not cause false alarms.**
- **Detection delay is less than 0.18s.**
- **Computation time for each data window is 0.0115s.**



Numerical Results – Different Similarity Metrics

Similarity Metric for High-Variance Bad Data

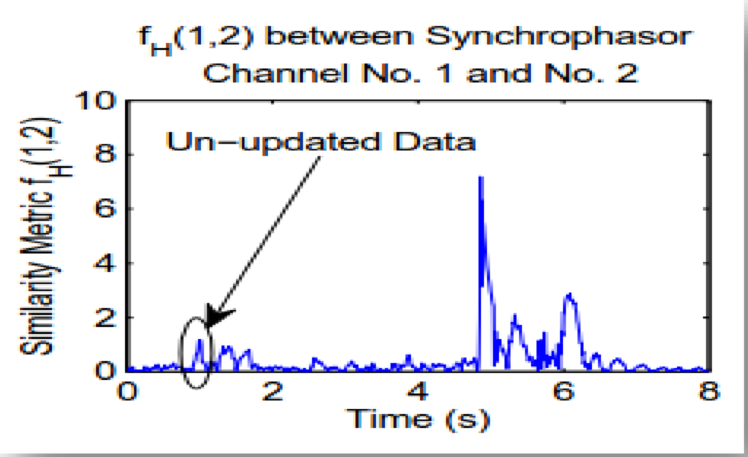
$$f_H(i,j) = \left|\sigma_i^{Norm} - \sigma_j^{Norm}\right|$$

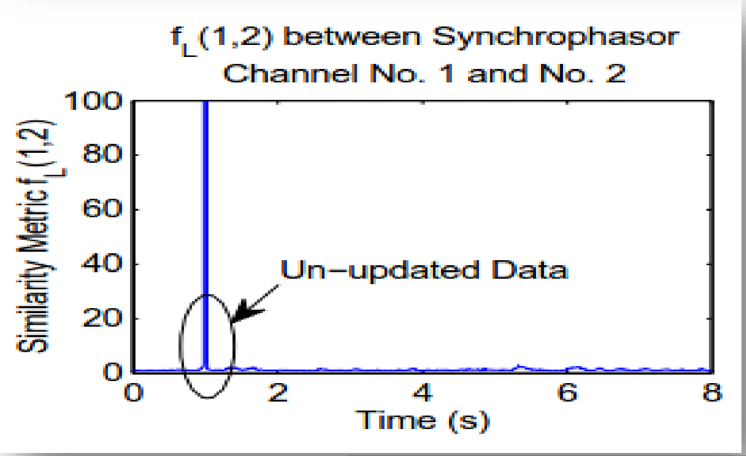
Similarity Metric for Low-Variance Bad Data

$$f_L(i,j) = max\left(\left|\frac{\sigma_i^{Norm}}{\sigma_j^{Norm}}\right|, \left|\frac{\sigma_j^{Norm}}{\sigma_i^{Norm}}\right|\right)$$

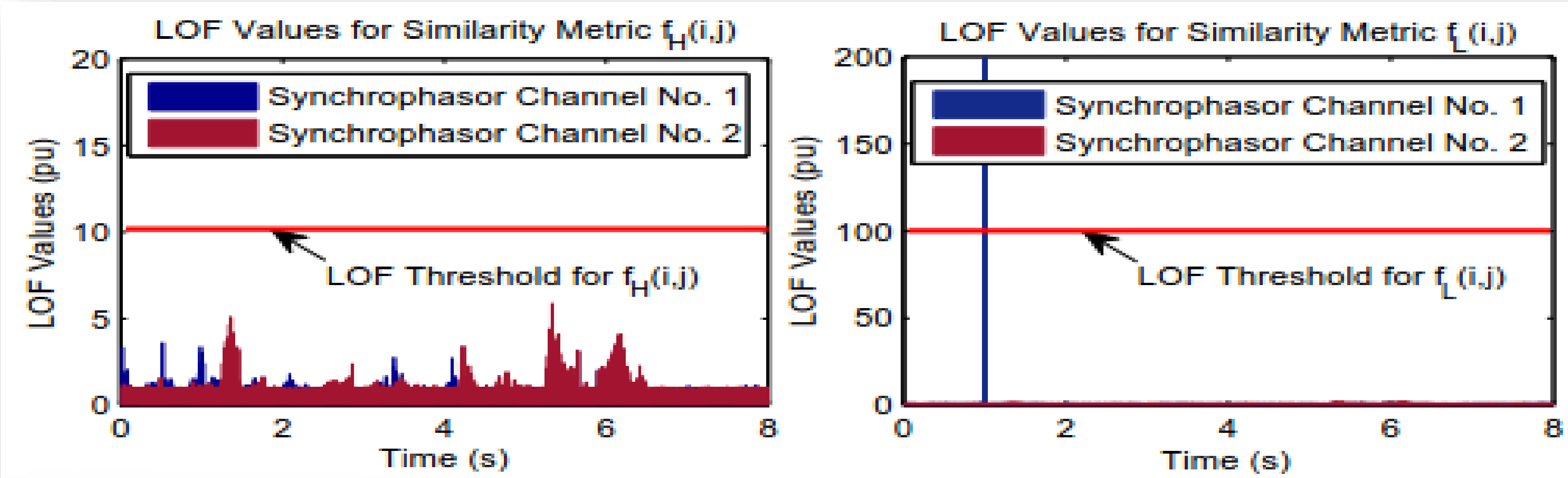
Performance Difference

- $f_H(i,j)$ **is more sensitive to high-variance bad data.**
- $f_L(i,j)$ **is more sensitive to low-variance bad data.**





Numerical Results – Different Similarity Metrics



Observations

- LOF indicator based on $f_H(i,j)$ is more sensitive to ***high-variance*** bad data.
- LOF indicator based on $f_L(i,j)$ is more sensitive to ***low-variance*** bad data.

Presentation Outline

1 Introduction

2 Technical Approach

3 Case Studies

4 **Conclusions & Future Work**

TEXAS A&M UNIVERSITY

Real-Time Detection of Low-Quality PMU Data

Conclusions

<u>An approach for PMU low-quality data detection is proposed:</u>

- It is purely data-driven, without involving any knowledge on network parameters or topology, which avoids the impact of incorrect parameter/topology information on the identification results.
- It encounters no convergence issues and has fast computation performance, which is desirable for online application.
- It is suitable for identifying low-quality data in PMU outputs under both normal and eventful operating conditions.

Future Work

- Identify the root cause of the low-quality PMU data.
- Propose correction mechanism for the low-quality PMU data.

TEXAS A&M UNIVERSITY

References

References

- [1] California ISO, “Five year synchrophasor plan,” California ISO, *Tech. Rep.*, Nov 2011.
- [2] S. Ghiocel, J. Chow, et al. "Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring," *IEEE Tran. Power Systems*, 2014.
- [3] K. D. Jones, A. Pal, and J. S. Thorp, “Methodology for performing synchrophasor data conditioning and validation,” *IEEE Tran. Power Systems*, May 2015.
- [4] M. Wang, J. Chow, P. Gao, X. Jiang, Y. Xia, S. Ghiocel, B. Fardanesh, G. Stefopolous, Y. Kokai, N. Saito, and M. Razanousky, “A low-rank matrix approach for the analysis of large amounts of power system synchrophasor data,” in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, Jan 2015, pp. 2637–2644.
- [5] K. Martin, “Synchrophasor data diagnostics: detection & resolution of data problems for operations and analysis”, in *Electric Power Group Webinar Series*, Jan 2014.
- [6] M. Wu, and Le Xie. Online Detection of Low-Quality Synchrophasor Measurements: A Data-Driven Approach. *IEEE Transactions on Power Systems*. Accepted, to appear.
- [7] Breunig, Markus M., et al. "LOF: identifying density-based local outliers." *ACM sigmod record*. Vol. 29. No. 2. ACM, 2000.

TEXAS A&M UNIVERSITY




TEXAS A&M UNIVERSITY