PMU Data Anomaly Detection, Classification, and Prediction Using Machine Learning and Artificial Intelligence

Mohammadreza Maddipour Farrokhifard[1] (Presenter), Gang Zheng[1], Philip Hart[2], Weizhong Yan[2], Tianyi Wang[2], Vijay Kumar[2], Pengyuan Wang[2]

Mohammadreza.maddipourfarrokhifard@ge.com

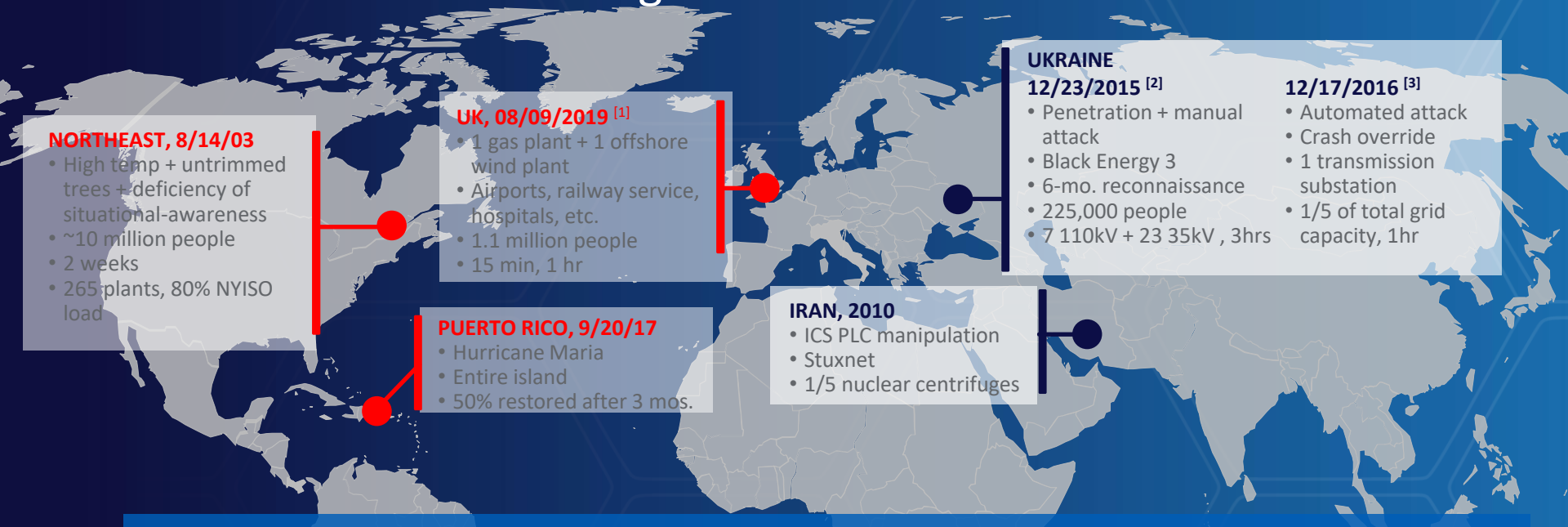
[1] GE Digital
[2] GE Research

Acknowledgment: "This material is based upon work supported by the Department of Energy under Award Number DE-OE0000915."

The 2nd IEEE International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA) *Virtual Event | May 24-27, 2021*

DIGITAL ENERGY

Power Grid Threats Growing



NORTHEAST, 8/14/03
- High temp + untrimmed trees + deficiency of situational-awareness
- ~10 million people
- 2 weeks
- 265 plants, 80% NYISO load

UK, 08/09/2019 [1]
- 1 gas plant + 1 offshore wind plant
- Airports, railway service, hospitals, etc.
- 1.1 million people
- 15 min, 1 hr

PUERTO RICO, 9/20/17
- Hurricane Maria
- Entire island
- 50% restored after 3 mos.

IRAN, 2010
- ICS PLC manipulation
- Stuxnet
- 1/5 nuclear centrifuges

UKRAINE

12/23/2015 [2]
- Penetration + manual attack
- Black Energy 3
- 6-mo. reconnaissance
- 225,000 people
- 7 110kV + 23 35kV , 3hrs

12/17/2016 [3]
- Automated attack
- Crash override
- 1 transmission substation
- 1/5 of total grid capacity, 1hr

"It's this modern electric power system that drives our digital economy and elevates our health, safety, and national security. Threats to our energy infrastructure, from extreme weather events to cyber and physical attacks, continue to grow and evolve."

- DOE OE RFI

PMU-Based Data Analytics Using Digital Twin and PhasorAnalytics Software

- **DOE Award Number: *DE-OE0000915***
- **FOA 1861**
- **Lead: GE Global Research**
- **Partner(s): GE Grid Software Solutions**

Context:

DOE has provided exclusive access to 2 years' worth of raw synchrophasor data from 100s of PMUs in the Eastern, Western, and Texas interconnections, as well as event logs with 1000s of labeled events.

Objectives:

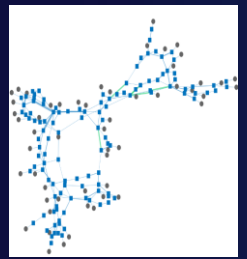
- Identifying signatures for events and anomalies.
- Identifying precursor conditions to equipment failure.
- Comprehensive data quality assessment.

DIGITAL ENERGY

WAMS Anomaly Analytics Framework

Holistic Anomaly Management

Transmission System



Data

- PMU
  - Voltage
  - Current
  - Angle
  - Frequency
- SCADA data
- Event logs
- State estimates

(System Info)

- **PMU Locations**
- **Weather Data**
  - Radiation
  - Air temperature
  - Relative humidity
  - Wind speed
  - precipitation
- **Others**

Data Quality Handling

- SVD
- PCA
- Hankel Structure

Data Preprocess

- Normalization
- Filtering
- Mapping to image

Unsupervised Learning using GE's signature base

Disturbance Detection

- SVD-Based
- Hankel structured

Semi-Supervised Learning and Root Cause Analysis

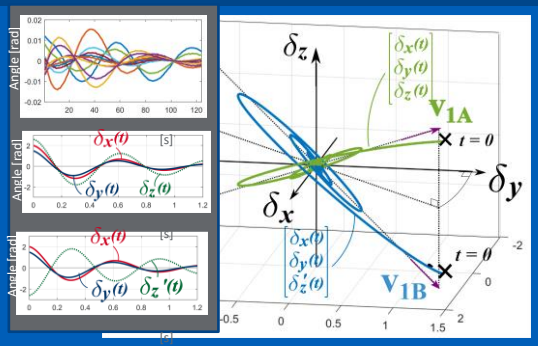
Disturbance Classification

- RBF
- SVM
- DT
- CNN
- RNN
- LSTM

Causal/correlated factors analysis

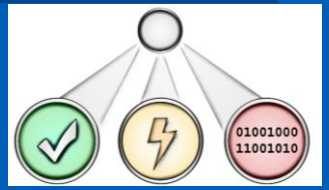
- Prevent wind Oscillation
- Cyber/Fault warning
- Gen. performance prediction
- Load Model Calibration
- Xformer, insulator failure prediction
- Gen. model Val. & Cal.
- Parameter Identifiability
- Failure signature & event correlation

SVD + Power Systems Analysis



Rapid Detection and Classification:

Feature



Normal | Physical Anomaly | Cyber Anomaly

Examples of physical anomalies:

- Generator
- Transformer
- Line fault
- Line Lightening
- Oscillation

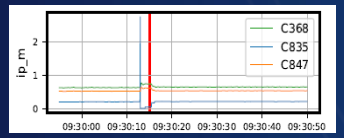
Examples of cyber anomalies:

- Data spoofing
- Replay Attack
- Denial of Service
- Packet Analysis

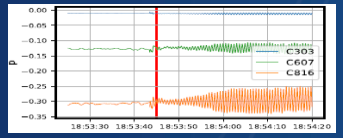
Feature-based event detection and classification

Generator Event

Line Fault Event

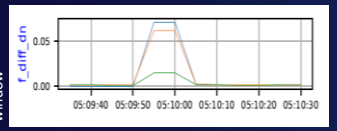


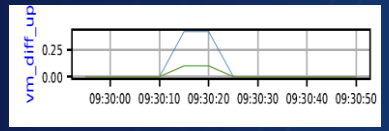
Oscillation Event

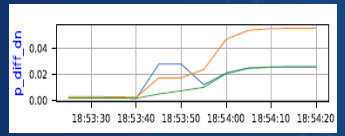


Two Dominant Features:

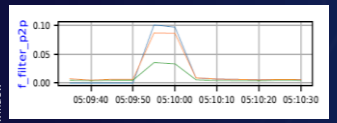
Max step-down Freq in a rolling window

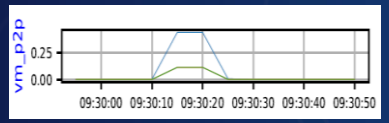


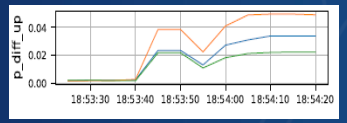




Diff max and min within a pre-processed rolling window







Some of the most dominant features for different type of anomalies:

Generator events:
- im_std
- f_diff_dn
- im_RP
- im_diff_dn
- f_filter_p2p

Oscillation events:
- p_diff_dn
- vm_diff_dn
- vm_p2p
- im_diff_dn
- vm_step_mag

Transformer events:
- vm_diff_up
- vm_diff_dn
- p_diff_up
- vm_p2p
- im_diff_dn

Line events:
- vm_diff_dn
- vm_diff_up
- vm_p2p
- vm_step_mag_1
- vm_step_mag _2

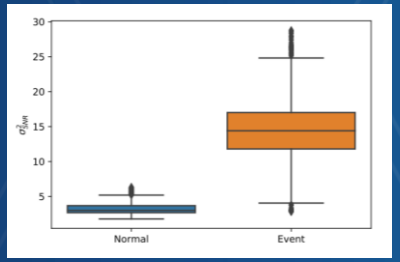
- **Useful signatures have been obtained for key grid events by generating and ranking an overabundance of (70+) features across two years of data from hundreds of PMUs.**
- **Using Eastern interconnection historical WAMS data for anomaly detection**

Feature function category	Description of feature function category	Example feature functions
Fundamental feature functions	A set of basic statistics & physics-based features that may be considered applicable to most or all events and all PMU variables (frequency, ROCOF, positive sequence voltage/current mag. and angle, etc.)	Deviation from nominal value, accumulated deviation, max change, max rate of change, peak-to-peak, slope
Event-specific feature functions	Physics-based features expected to be most impactful/insightful when applied to specific events such as a generator trip, line fault, equipment failure or FIDVR; applicable to specific PMU variables.	Bump/dip area, ‘n’ largest step-changes, SNR or NSR
Frequency-domain feature functions	A set of features that involve modal and/or spectral analysis.	Ring-down damping and mode, energy-based ringdown event detection, zero crossing count, FFT

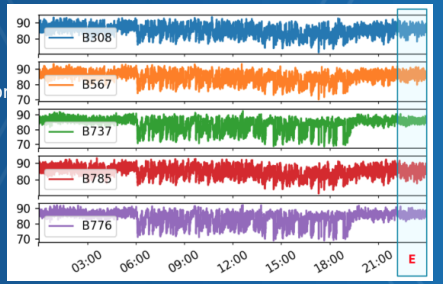
Causal Sequences:

Signal to Noise Ratio (SNR) of voltage signals:

- Transformers failure
- Useful for understanding the network structure
- Deriving causality model



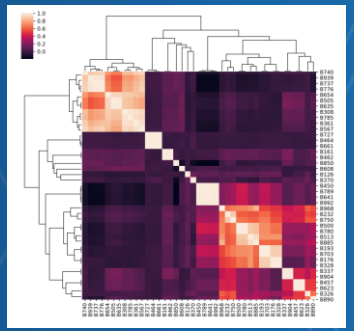
Distribution of variance of SNR over a 20-minute window for event data and normal data

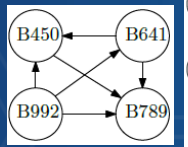


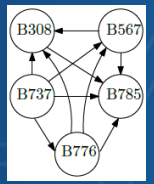
Early Warning Prediction…

PMU Voltage Channels → ***Calculating SNR***

- ***Distribution of SNR variance during non-event periods*** → χ^2 ***Test over a sliding window*** → ***Event Detection***
- ***Pearson correlation coefficients*** → ***Clustering*** → ***determining the inter–node connectivity***
- ***Granger causality model*** → ***determining the causality model***

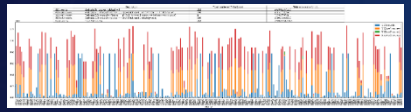


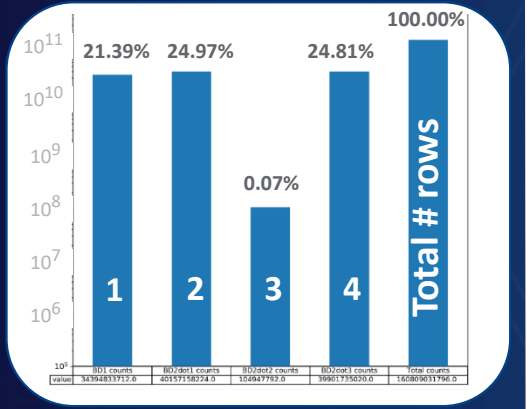




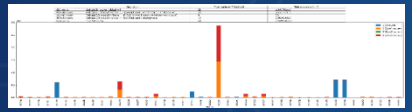
Comprehensive Data Quality Assessment

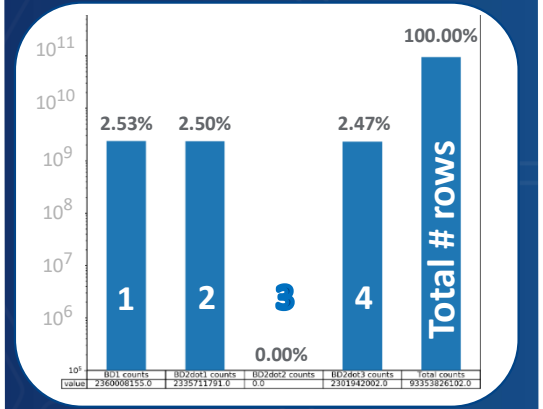
Interconnect A:



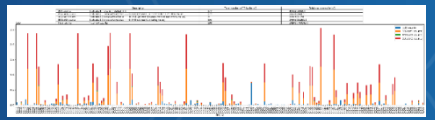


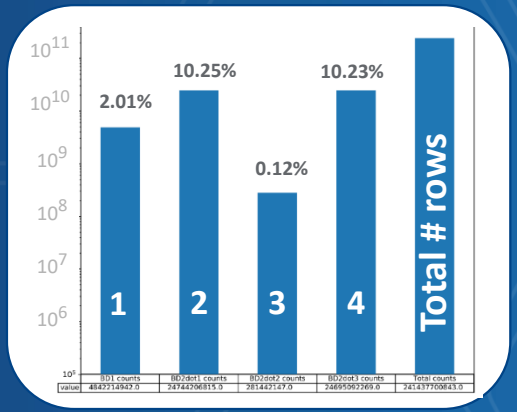
Interconnect B:





Interconnect C:





1: # of rows with 'status' !=0
2: # of rows with >0 'unreasonable' values*
3: # of rows with >0 non-numerical values
4: # of rows with >0 missing values*

*Rows containing missing & 'unreasonable' values overlap.

Easy to detect/ mitigate

25%-50% of each dataset is bad data of type 1-4; also encountered other examples of bad data (e.g. intermittent 1000x drops in voltage magnitude) that are much harder to detect and mitigate.

Software Implementation

- Study of features from archived data:
  - Python-based custom calculations module in offline tool
  - Plan to add more ready-to-use signatures for study of archived data
- Real-time and offline Event detection and classification :
  - Currently implemented: Limited number of features are already implemented.
  - Future plan: Covering more features for different type of event

Challenges:

- Inaccurately or inadequately labeled abnormal data
- Computational burden of pre-processing and feature generation
- Insufficient good quality data or event logs
- Anonymization of dataset to protect the identity of the data providers (no network topology, PMU location, or event location information given by the utilities, ISOs, etc.)
- …