FOA 1861 FINAL PROJECT BRIEFING
BIG DATA ANALYSIS OF SYNCHROPHASOR DATA

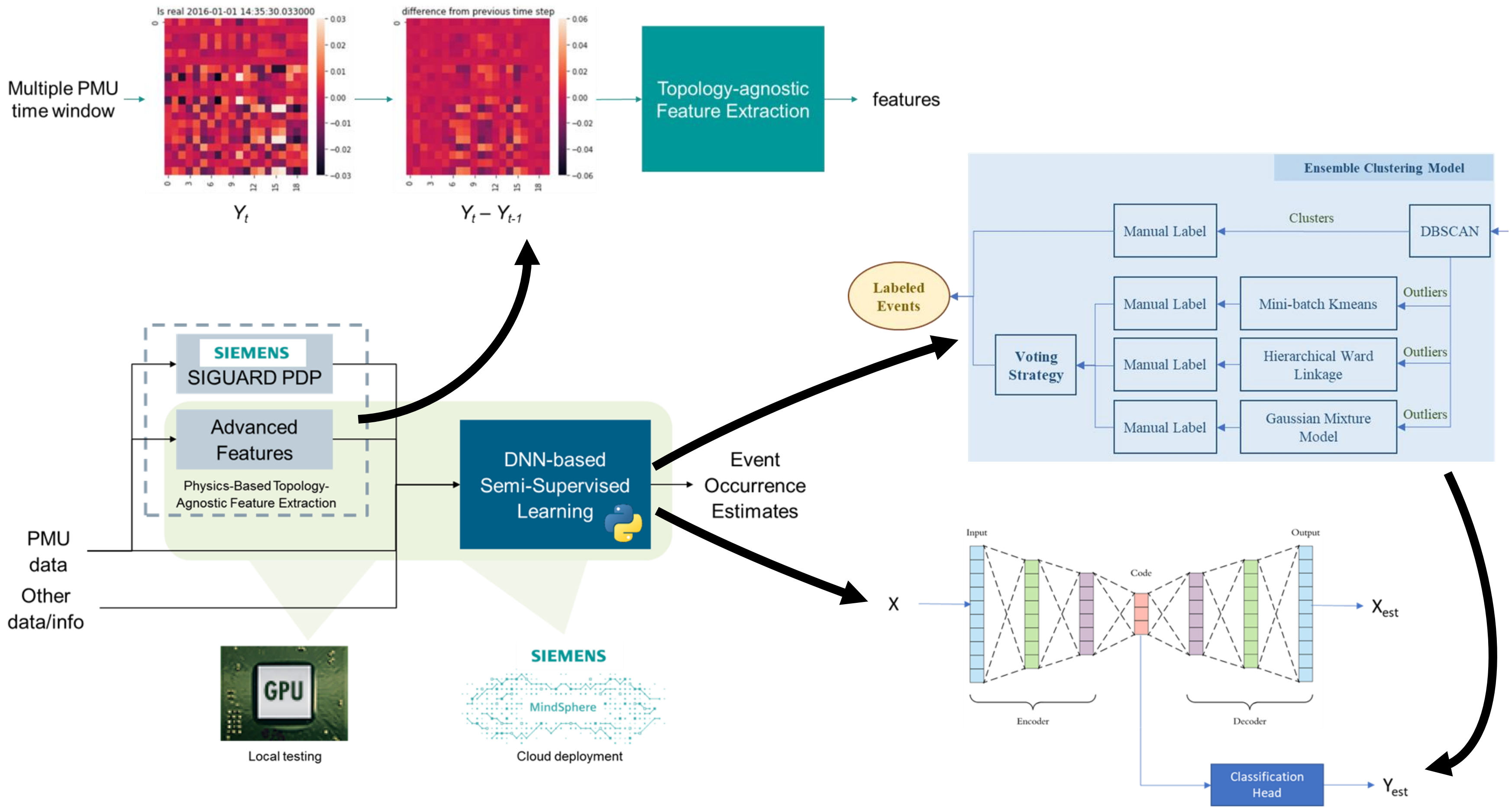
MindSynchro

Bruno Leao
Siemens Technology
bruno.leao@siemens.com
July 28, 2021

Project Partners: Siemens Digital Grid, Siemens Process Automation, Southern Methodist University, Temple University

SIEMENS
Ingenuity for life

Project Overview/Background

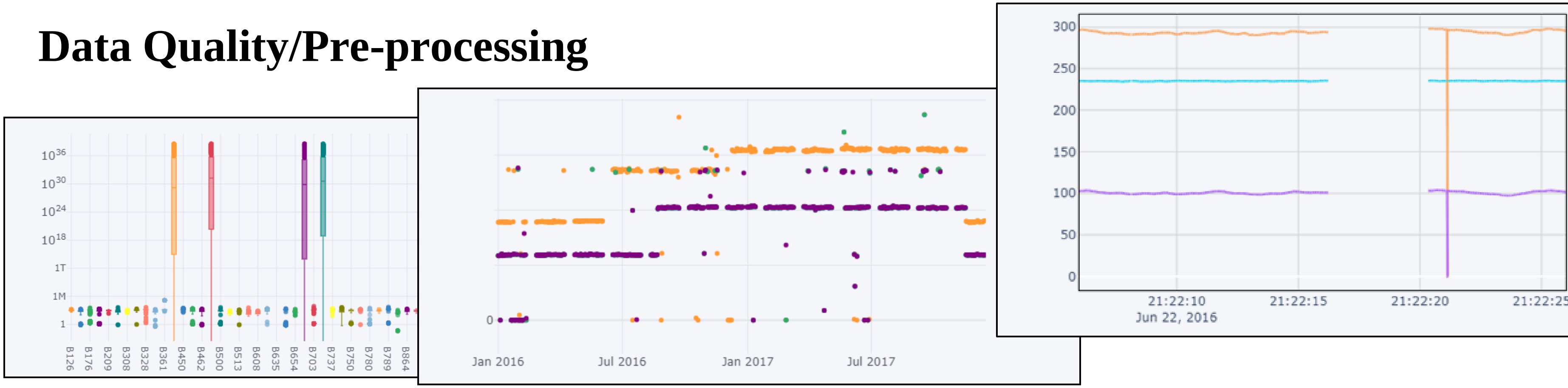


SIEMENS

Ingenuity for life

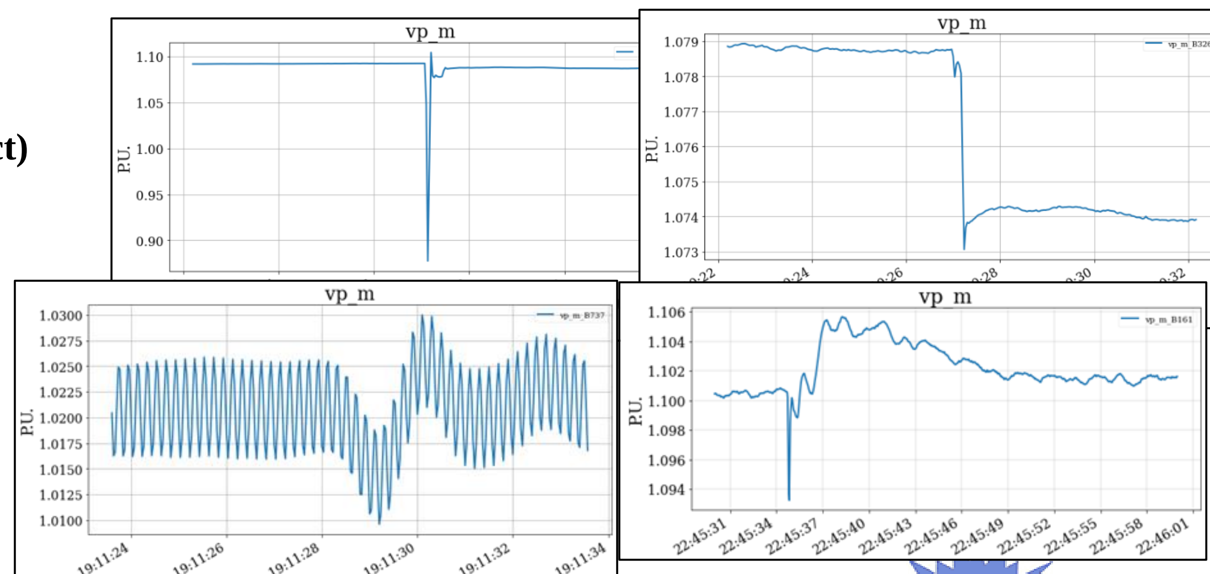
Experimental Results

Data Quality/Pre-processing



Final Patterns from UL

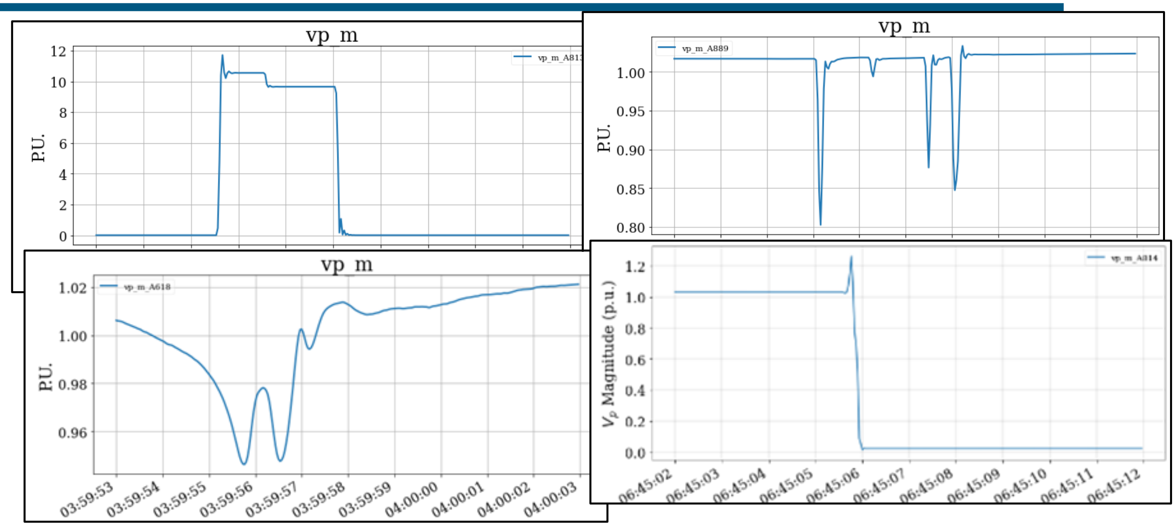
- Line trip (direct/indirect)
- Short circuit (direct/indirect)
- Auto-reclosure
- Loss of generation
- Load disconnection
- Sustained oscillation
- Damped oscillation
- Transient
- Noise
- Normal
- Unknown



Experimental Results

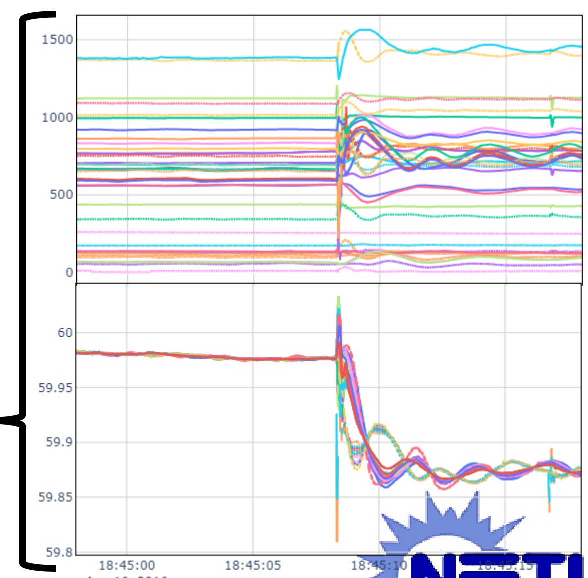
Event Identification

Test Dataset	Number of events detected (UL general)
Texas IC (A)	3,661
Western IC (B)	2,583
Eastern IC (C)	9,476



DSSL Model Results

Model	#Labeled (P/N)	Accuracy	Precision	Recall	FAR
Short circuit (SC)	1878 / 296898	99.83%	94.46%	77.13%	0.03%
SC with features	1983 / 284876	99.87%	96.28%	84.63%	0.02%
Trip no short circuit	6258 / 271117	99.12%	88.03%	71.09%	0.22%
Loss of generation	11226 / 302824	>99.99%	99.87%	99.96%	0.005%



In NERC report

SIEMENS

Ingenuity for life

Technical Accomplishments

- ✓ Findings
  - ✓ Data quality
  - ✓ Patterns
  - ✓ Events
- ✓ Pre-processing
- ✓ Physics-based Features
  - ✓ Admittance matrix-based
  - ✓ Oscillation
  - ✓ Topology (PMU clustering)
- ✓ Labeling for ML training
  - ✓ Categories definition
  - ✓ UL-based scalable label definition
- ✓ DSSL model
  - ✓ Investigation of GAN and AE-based models
  - ✓ Training of final models for multiple event types
- ✓ Deployment to industrial cloud environment (MindSphere)
- ✓ Paper publications / Panel participations

Value of Work

- Benefit to the utilities: **situational awareness**
  - **Online information** that would require hours of analysis from SME
  - **Info for acting** in response to events to minimize their consequences
  - Long term: integration into **protection and control**
- Tools developed in the project provide:
  - Identification of specific types of **power grid events** for each PMU **individually** (focus on real-world use)
  - **Scalable data labeling**: enable utilities to define their **requirements**
- Dataset
  - Broad applicability and validation can only happen because of **data shared from multiple utilities**.
  - **Anonymized data** provides big challenges for integrating data from multiple PMUs but not a big issue for individual PMU analysis.
  - **Many opportunities** for improvement, especially on **labels** (golden dataset)

SIEMENS
Ingenuity for life

Readiness for Commercialization

- Next step: work in **partnership with utilities**
  - Work is currently in TRL 6
  - Need to integrate **utility specific requirements**
    - Labels
    - Performance metrics
    - Integration with existing tools
  - **Pilot project (< 1 year duration)**
    - Refining and validating models in partnership with utilities
    - Use historical data from utilities or from FOA 1861
    - Integration to relevant tools
    - Online monitoring once performance and integration goals are achieved

Being Ready for ML & BD Analytics

- SOTA in ML not an important limitation for PMU data analytics
  - Performance can potentially be improved , for instance, by integrating physics
- Relevant challenges/limitations:
  - Adequate labels ☺: validated and associated to clear data patterns
  - Data quality
  - Topology / location for system level and improved local awareness
- Opportunities
  - Golden datasets
  - Standardization/Best practices
    - event categories
    - pre-processing best practices to overcome data quality issues
  - Topology: maybe it is possible to share an adjacency matrix for a subset of PMUs?
  - Utilities working together with vendors, academia/labs:
    - Include requirements and guide effort to solution of most relevant problems
    - Familiarize with potential and limitation
    - Define best way to integrate to operation
    - Demonstrate the technologies in real world

Lessons Learned and Next Steps

- Importance of **adequate labels** ☺☺
  - **Golden datasets** would unleash a diversity of opportunities
  - Analogous to https://image-net.org and other repos for the (very successful) image analytics field
  - Currently, even if we have perfect models they may not be useful as they may **not be solving the right problem**
- **Data quality** is a big challenge, very relevant for results
- Adding **physics** can help, also **topology** (even if simplified)
- Next steps – suggested **roadmap** for the industry

(Preliminary) definition of categories → Creation of golden datasets → "Kaggle" like competitions → Integration of operation requirements → Integration into protection and control

Publications

- B. P. Leao et al. “Big Data Processing for Power Grid Event Detection”, IEEE Big Data 2020
- Y. Du et al. “Physics-Based Feature Extraction from Bulk Time-Series PMU Datasets for Event Detection”, IEEE PES GM 2021 (Best Paper)
- T. Lan et al. “Unsupervised Power System Event Detection and Classification Using Unlabeled PMU Data”, IEEE ISGT Europe 2021 (Submitted)
- Additional papers in preparation
- Panel participations:
  - NASPI Work Group Virtual Meeting
  - IEEE SGSMA 2021
  - PES GM 2021

SIEMENS

Ingenuity for life

NETL
NATIONAL ENERGY TECHNOLOGY LABORATORY