

Data Considerations for ML in Distribution System Model Calibration





Logan Blakely

PI - Matthew Reno



This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 34226. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. SAND2024-10030PE How do we leverage data processing and physical system characteristics to improve the machine learning results, while navigating common data quality issues? Using distribution system model calibration applications as an motivating example.

Motivating Point: Even after a ML model was selected, performance and robustness in real-world usage depended on data processing improvements. Without the problem-specific data pre-processing steps, the algorithm would not have been accurate or reliable enough to use outside of the lab

Presentation Agenda:

- Application Overview What is distribution system model calibration, and why do we care?
- Phase Identification
 - Problem Characteristics
 - Data Considerations Application Specific
 - Data Consideration Data Quality
- Key Takeaways

Application Overview

3

- The distribution system is the low voltage section of the power grid which serves electricity to homes and business
- Utilities often do not have accurate models of the distribution system due to maintenance, storm restoration, and legacy system challenges
- Available data is Advanced Metering Infrastructure (AMI) data from smart meters on homes/business. These meters record timeseries data, often at 15-min intervals

PV Detection Detect PV configuration (size, tilt, and azimuth) and settings **Phase Identification** Identify the phase of laterals and phase of single-phase transformers

Parameter Estimation Estimate cable length and topology of the lowvoltage system **Customer Transformers** Identify which transformer each meter is connected to

ħ

Phase Identification Problem Characteristics

- Customers connected to the same phase are electrically connected to all other customers on that phase. Thus, their voltages will be more highly correlated than customers on other phases
- Inherently lends itself to a clustering approach

4

• Customers will each have distinct energy usage characteristics





Even after deciding on an appropriate ML approach (Spectral Clustering), many of the improvements which made the overall algorithm work in the real world came from data processing changes

Phase Identification - Data Processing Solutions

5

Application specific considerations

- We care mainly about correlations in voltage
- Customer usage patterns and magnitudes can be very different
 - Voltage profiles can be dominated by trends (seasonal and daily)
- Voltage 'Snapshots' can be misleading due to variation in customer behavior, season, etc.
- Many issues found in utility data can cause unknown problems

Convert voltage profiles into 'difference' (change in voltage) and per-unit form

- Analyze windows/snapshots of voltage data
- Use an ensemble approach to leverage many windows
- Use ensemble information to provide a per customer confidence score

Phase Identification - Data Processing Solutions

Data quality considerations

6



Key Takeaways

ħ

• Choosing a ML approach is just one step in the process

Questions to ask:

- What data processing can I apply based on the characteristics on my specific application?
- What data processing can I apply to resolve, or mitigate, data quality issues?
- Is there data processing I can use to increase algorithm robustness for unknown issues?