IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Learning Smart Meter Data for Distribution Grid Modeling and Observability Enhancement

Zhaoyu Wang

Northrop Grumman Endowed Associate Professor

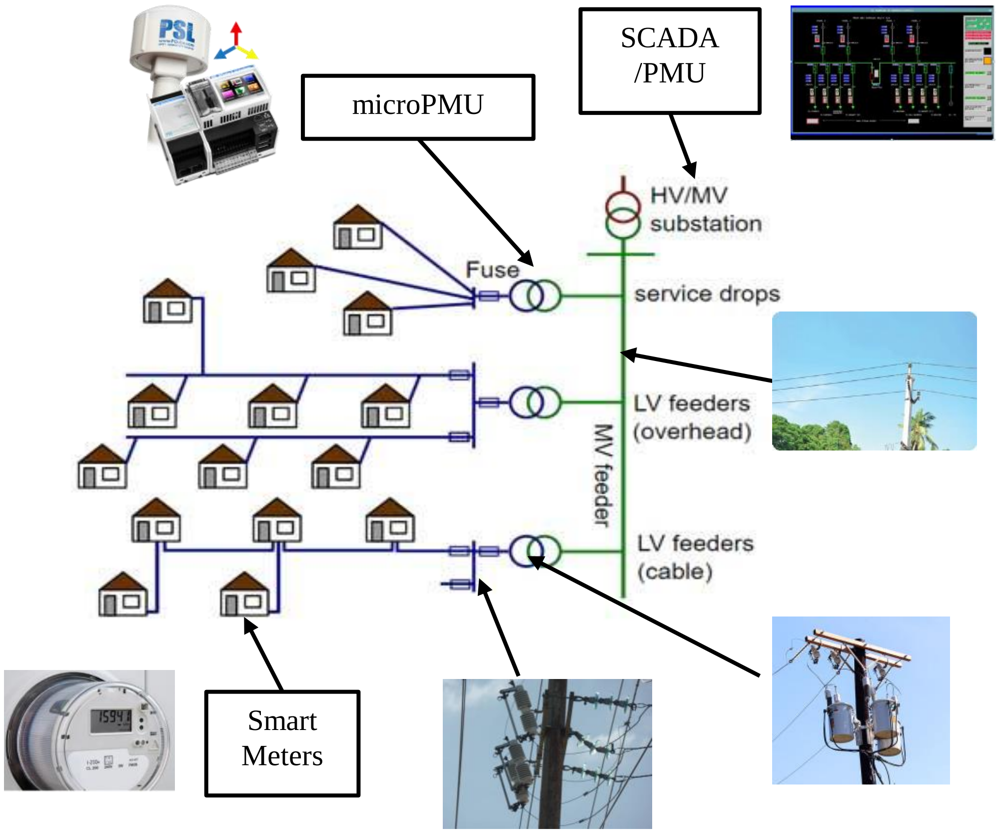
Iowa State University

(wzy@iastate.edu)

Presentation Outline

- **Introduction to Smart Meter Data**
- **Distribution Grid Topology and Line Parameter Identification**
  - **Existing Work and Challenges**
  - **Robust Network Topology Identification via Weighted Laplacian Matrix**
  - **A Bottom-Up Sweep Approach for Line Impedance Estimation**
- **Brief Introduction to Other SM Data Applications**
  - **Outage Detection**
  - **Customer Peak Load Estimation**
- **Conclusion and Future Work**

Data in Power Distribution Grids



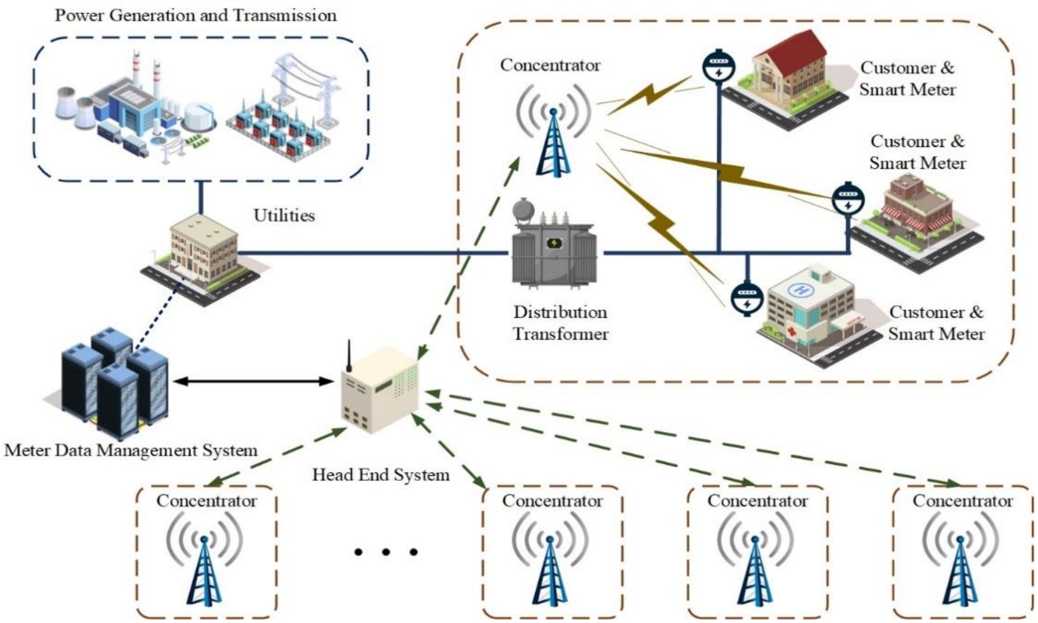
A Power distribution grid

- **Where does the data come from?**
  - SCADA (supervisory control and data acquisition); Smart Meters; Protection Devices; microPMUs (phasor measurement units)
  - Measures voltage/current/frequency at different resolutions

- **What are smart meters?**
  - Stay in your homes
  - Measure energy and voltage
  - 15/30/60-minute resolution

Smart Meter Data

- With the increasing integration of DERs in power distribution systems, utilities need to improve systematic situational awareness in order to execute behind-the-meter (BTM) load control strategies.
- In recent decades, the deployment of advanced metering infrastructure (AMI) in distribution systems has extended monitoring capability to grid edges.
- The core element of AMI is smart meter (SM) which is a device installed at customer house or facility.



SM data is a good resource for enhancing distribution grid monitoring and control thanks to extensive customer-side installations!

Y. Yuan and Z. Wang, "Mining Smart Meter Data to Enhance Distribution Grid Observability for Behind-the-Meter Load Control," *IEEE Electrification Magazine*, vol. 9, no. 3, PP. 92-103, September 2021.

Available Utility Data

- General Description:

Utilities	Substations	Feeders	Transformers	Total Customer	Customers with Meters
3	5	27	1726	9118	6631

- Duration: 4 years (2014 - 2018)
- Measurement Type: Smart Meters and SCADA
- **Detailed circuit models of all feeders in Milsoft/OpenDSS and exact smart meter locations**
- Data Time Resolution: 5 Minutes – 1 Hour
- Customer Type:

Residential	Commercial	Industrial	Other
84.67%	14.11%	0.67%	0.55%

Exemplary Real Data from Utilities

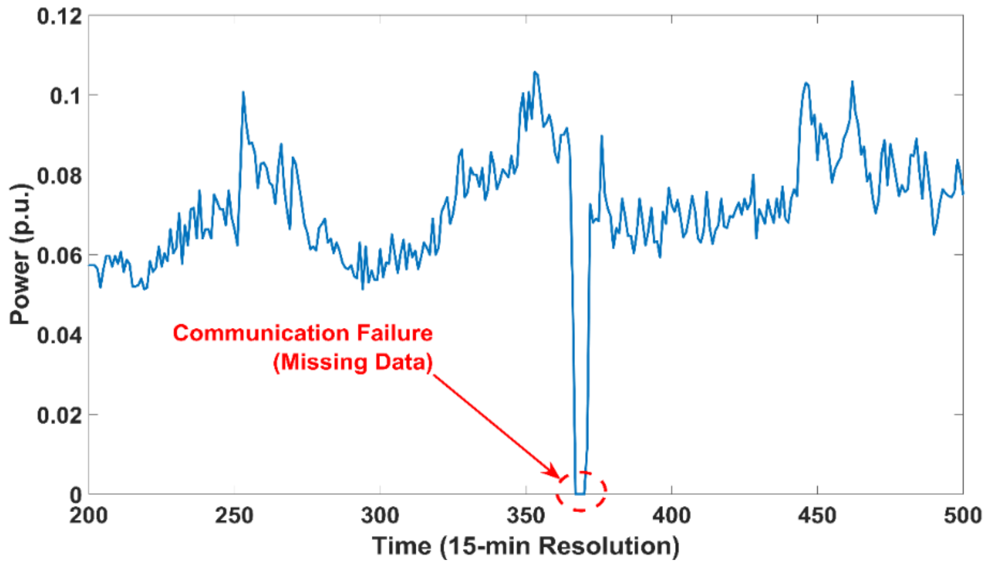
Time — Hourly energy & instantaneous voltage

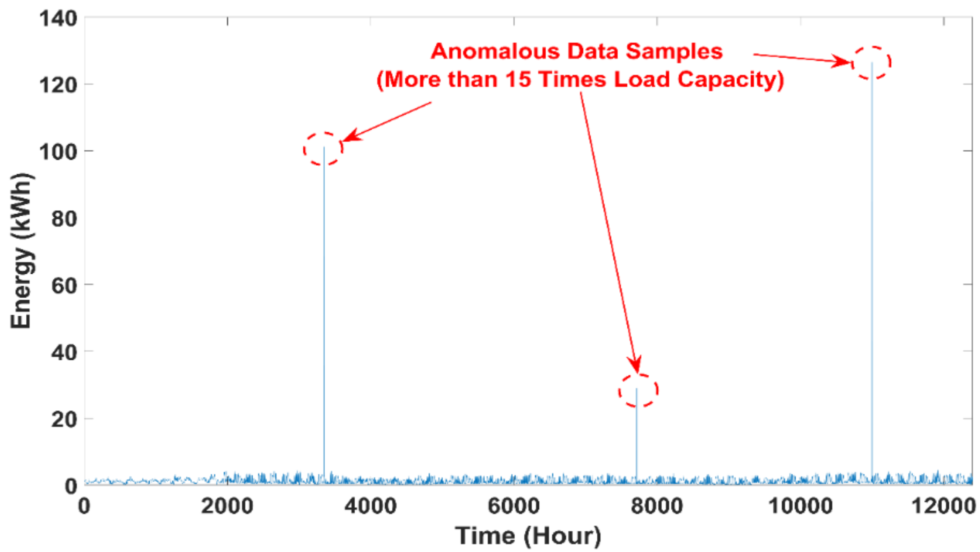
Account		time	kWH or V	time	kWH or V	time	kWH or V	time	kWH or V
100000001	KWH	201704010100	0.392	201704010200	0.257	201704010300	0.215	201704010400	0.239
100000001	VOLTS	201704010100	239	201704010200	239	201704010300	238	201704010400	240
100000002	KWH	201704010100	0.245	201704010200	0.204	201704010300	0.252	201704010400	0.342
100000002	VOLTS	201704010100	241	201704010200	240	201704010300	240	201704010400	240
100000003	KWH	201704010100	1.479	201704010200	0.417	201704010300	0.816	201704010400	0.414
100000003	VOLTS	201704010100	240	201704010200	239	201704010300	239	201704010400	240
100000004	KWH	201704010100	1.009	201704010200	0.555	201704010300	0.39	201704010400	0.382
100000004	VOLTS	201704010100	241	201704010200	237	201704010300	237	201704010400	239
100000005	KWH	201704010100	0.798	201704010200	0.809	201704010300	0.87	201704010400	0.692
100000005	VOLTS	201704010100	239	201704010200	238	201704010300	238	201704010400	240
100000006	KWH	201704010100	0.109	201704010200	0.188	201704010300	0.205	201704010400	0.148
100000006	VOLTS	201704010100	241	201704010200	240	201704010300	240	201704010400	242
100000007	KWH	201704010100	1.199	201704010200	1.512	201704010300	1.759	201704010400	1.474
100000007	VOLTS	201704010100	241	201704010200	240	201704010300	239	201704010400	241
100000008	KWH	201704010100	0.422	201704010200	0.419	201704010300	0.43	201704010400	0.537
100000008	VOLTS	201704010100	239	201704010200	239	201704010300	238	201704010400	240
100000009	KWH	201704010100	2.288	201704010200	2.278	201704010300	2.335	201704010400	2.297
100000009	VOLTS	201704010100	243	201704010200	242	201704010300	242	201704010400	242
100000010	KWH	201704010100	0.223	201704010200	0.257	201704010300	0.292	201704010400	0.25
100000010	VOLTS	201704010100	242	201704010200	241	201704010300	241	201704010400	241

Smart Meter Data Pre-Processing

- **Smart Meter Data Problems:**
  - Outliers/Bad Data
  - Communication Failure
  - Missing Data

- **Counter-Measures:**
  - ✓ Engineering intuition (data inconsistency)
  - ✓ Conventional Statistical Tools (e.g. Z-score)
  - ✓ Robust Computation (e.g. relevance vector machines)
  - ✓ Anomaly Detection Algorithms

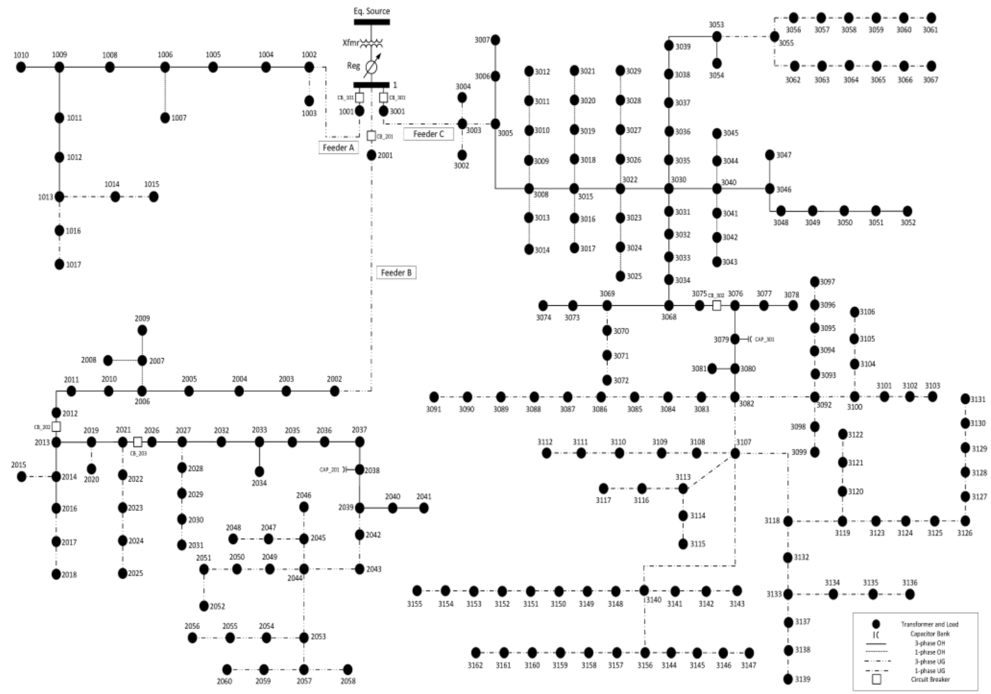




Data Sharing

With permission from our utility partner, we share a real distribution grid model with one-year smart meter measurements. This dataset provides an opportunity for researchers and engineers to perform validation and demonstration using real utility grid models and field measurements.

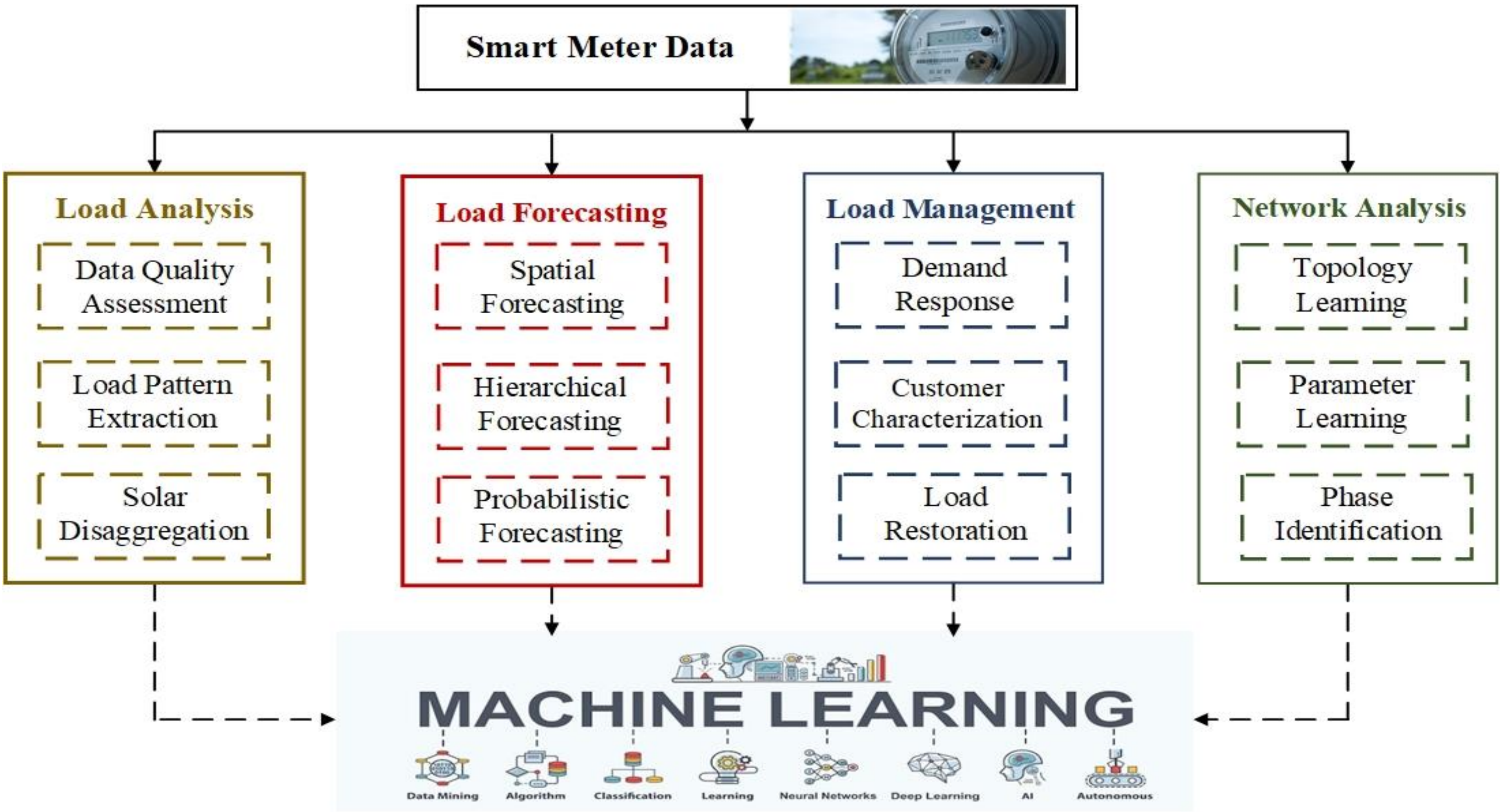
- The system consists of 3 feeders and 240 nodes and is located in Midwest U.S.
- The system has 1120 customers and all of them are equipped with smart meters. These smart meters measure hourly energy consumption (kWh). We share the one-year real smart meter measurements for 2017.
- The system has standard electric components such as overhead lines, underground cables, substation transformers with LTC, line switches, capacitor banks, and secondary distribution transformers. The real system topology and component parameters are included.



Test system diagram

You may download the dataset at: http://wzy.ece.iastate.edu/Testsystem.html , including system description (in .doc and .xlsx), smart meter data (in .xlsx), OpenDSS model, and Matlab code for quasi-static time-series simulation!

What can be learned from smart meter data to improve distribution system operation?



Recent Publications in Machine Learning and SM Data Analytics

- Y. Guo, Y. Yuan, and Z. Wang, "Distribution Grid Modeling Using Smart Meter Data," *IEEE Transactions on Power Systems*, accepted for publication.
- Y. Yuan, K. Dehghanpour, Z. Wang, and F. Bu, "Multi-Source Data Fusion Outage Location in Distribution Systems via Probabilistic Graphical Models," *IEEE Transactions on Smart Grid*, accepted for publication.
- Y. Yuan and Z. Wang, "Mining Smart Meter Data to Enhance Distribution Grid Observability for Behind-the-Meter Load Control," *IEEE Electrification Magazine*, vol. 9, no. 3, PP. 92-103, September 2021.
- F. Bu, K. Dehghanpour, and Z. Wang, "Enriching Load Data Using Micro-PMUs and Smart Meters," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, PP. 5084-5094, November 2021.
- F. Bu, K. Dehghanpour, Y. Yuan, Z. Wang, and Y. Guo, "Disaggregating Customer-level Behind-the-Meter PV Generation Using Smart Meter Data and Solar Exemplars," *IEEE Transactions on Power Systems*, vol. 36, no. 6, PP. 5417-5427, November 2021.
- Y. Yuan, K. Dehghanpour, and Z. Wang, "Mitigating Smart Meter Asynchrony Error Via Multi-Objective Low Rank Matrix Recovery," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4308-4317, September 2021.
- Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "Outage Detection in Partially Observable Distribution Systems using Smart Meters and Generative Adversarial Networks, " *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5418-5430, November 2020.
- Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A Data-Driven Customer Segmentation Strategy Based on Contribution to System Peak Demand," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 4026-4035, September 2020.
- F. Bu, K. Dehghanpour, Y. Yuan, Z. Wang, and Y. Zhang, "A Data-Driven Game-Theoretic Approach for Behind-the-Meter PV Generation Disaggregation," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3133-3144, July 2020.
- F. Bu, K. Dehghanpour, Z. Wang, and Y. Yuan, "A Data-Driven Framework for Assessing Cold Load Pick-up Demand in Service Restoration," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4739-4750, November 2019.
- K. Dehghanpour, Y. Yuan, Z. Wang, and F. Bu, "A Game-Theoretic Data-Driven Approach for Pseudo-Measurement Generation in Distribution System State Estimation," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 5942-5951, November 2019.
- Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A Multi-Timescale Data-Driven Approach to Enhance Distribution System Observability," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3168-3177, July 2019.
- Q. Zhang, K. Dehghanpour, Z. Wang, F. Qiu, and D. Zhao, "Multi-Agent Safe Policy Learning for Power Management of Networked Microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1048-1062, March 2021.
- Q. Zhang, K. Dehghanpour, Z. Wang, and Q. Huang, "A Learning-based Power Management Method for Networked Microgrids Under Incomplete Information," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1193-1204, March 2020.

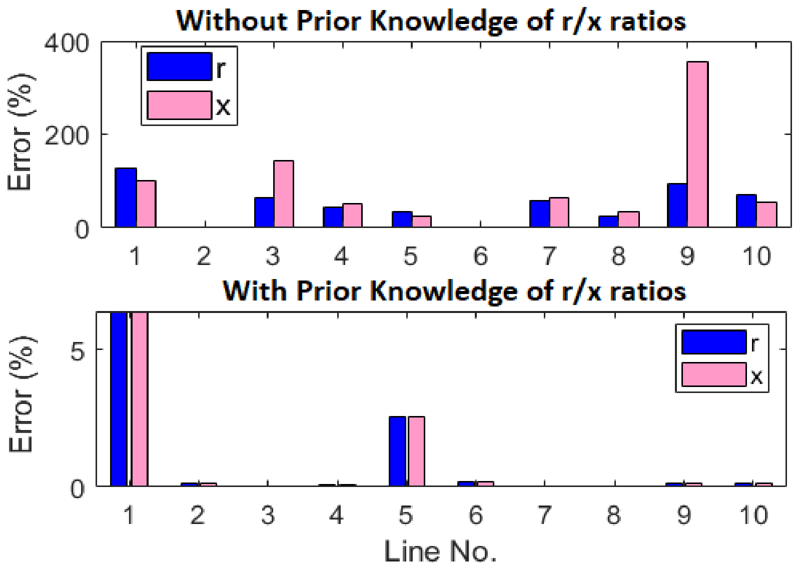
Distribution Grid Topology & Parameter Identification

- Complete and accurate distribution grid models are essential to system monitoring and control.
- Many small and medium utilities only have simple one-line diagrams of their systems without any detailed information.
- System models are often incomplete or outdated due to the frequent system expansion and reconfiguration.
- Conventional field inspection is laborious, costly, and time-consuming, especially for large-scale systems.

Existing Work and Challenges

- Using Y-bus injection model and phasor information (J. Yu 19, O. Ardakanian 19, Y. Yuan 20)
  - **Limitation**: require full coverage of µPMUs (cost-prohibitive).

- Using Branch flow model and smart meter data (A. M. Prostejovsky 16, H. Xu 18, W. Wang 20)
  - **Limitation**: require prior knowledge (i.e., R/X ratios of ***all*** line sections and network connectivity).
  - **Reason for this requirement**: searching space of the optimization (***ill-conditioned***).
  - **Another challenge**: scalability and computational complexity.



How to perform real-time topology and parameter identification using very limited yet readily available SM data?

Our Solution

- ✓ **Topology Identification**: Modeling the distribution network as a graph and identifying its ***weighted Laplacian matrix*** using SM data streams, where the matrix has a special structure that reveals the network connectivity.
  - ✓ High computational efficiency.
  - ✓ Robustness with respect to heterogeneous R/X ratios and model/measurement errors.

- ✓ **Parameter Estimation**: designing a ***bottom-up sweep algorithm*** to identify line impedances.
  - ✓ Based on the full nonlinear power flow, a least absolute deviations (LAD) with mixed-integer semidefinite programming (MISDP) model, and a least square (LS) model with mixed-integer second-order cone programming (MISOCP) model are developed.
  - ✓ Adding a library of R/X ratios (rather than exact R/X of all line sections) as a constraint to narrow down the search space.
  - ✓ Dividing the network into multiple layers. Parameter identification and power flow calculations are performed layer-by-layer in an alternate manner from bottom to top layers.

Y. Guo, Y. Yuan, and Z. Wang, "Distribution Grid Modeling Using Smart Meter Data," IEEE Transactions on Power Systems, accepted for publication.

Distribution Grid Topology Identification

- Our topology identification approach builds on the linear approximation of the branch flow model.

$$\mathbf{v} \cong 2\mathbf{A}^{-T}\mathbf{R}\mathbf{A}^{-1}\mathbf{p} + 2\mathbf{A}^{-T}\mathbf{X}\mathbf{A}^{-1}\mathbf{q} - v_0\mathbf{A}^{-T}\mathbf{a}_0 \quad (1)$$

- For a radial distribution network, A is non-singular and $\mathbf{A}^{-T}\mathbf{a}_0 = 1_n$

$$\frac{1}{2}\mathbf{A}\mathbf{X}^{-1}\mathbf{A}^T(\mathbf{v} - v_0 1_n) = \mathbf{A}\mathbf{X}^{-1}\mathbf{R}\mathbf{A}^{-1}\mathbf{p} + \mathbf{q} \quad (2)$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X}^{-1}\mathbf{A}^T;\ \mathbf{\Phi} = \mathbf{A}\mathbf{X}^{-1}\mathbf{R}\mathbf{A}^{-1} \quad (3)$$

where **v**, **p**, **q** denote the vectors collecting squared bus voltage magnitudes, real power, and reactive power injections; $[a_0, A^T]^T \in \{0, \pm 1\}^{(n+1)\times n}$ is the incidence matrix of the radial-topology graph; **R** and **X** are diagonal resistance and reactance matrices; **Y** is a *weighted Laplacian* matrix of the network with a sparse structure.

Weighted Laplacian Matrix of the Network

Proposition 1: $\mathbf{Y} \coloneqq [y_{ij}]_{n\times n}$ is a sparse symmetric matrix :

$$\begin{bmatrix} 3 & -1 & 0 & 0 & -2 & 0 \\ -1 & 5 & -4 & 0 & 0 & 0 \\ 0 & -4 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & -1 & -5 \\ -2 & 0 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -5 & 0 & 5 \end{bmatrix} \tag{4}$$

- $\mathbf{Y}$ holds a salient feature: for any connected nodes *i* and *j*, $y_{ij} < 0$ and for any non-connected nodes $y_{ij} = 0$.
- If one can approximately identify $\mathbf{Y}$, the topology can be extracted by observing the unique features of $\mathbf{Y}$.

Weighted Laplacian Matrix Identification

- Assume the network has a homogeneous R/X ratio, $\frac{r_1}{x_1} = \cdots = \frac{r_n}{x_n} = \lambda$, $\mathbf{\Phi}$ reduces to $\mathbf{\Phi} = \mathbf{A}\mathrm{diag}\left(\frac{r_1}{x_1}, \dots, \frac{r_n}{x_n}\right)\mathbf{A}^{-1} = \lambda 1_n$.
- For ***heterogeneous*** networks, our method still works because we do not require accurate estimation of **Y** and only need to distinguish zero and negative non-diagonal entries to identify connectivity, which will be proved in case study.
- The error vector regarding *k*-th measurement can be defined based on (2).

$$\frac{1}{2}\mathbf{A}\mathbf{X}^{-1}\mathbf{A}^T(\mathbf{v} - v_0 1_n) = \mathbf{A}\mathbf{X}^{-1}\mathbf{R}\mathbf{A}^{-1}\mathbf{p} + \mathbf{q} \Rightarrow \frac{1}{2}Y(\mathbf{v} - v_0 1_n) = \lambda\mathbf{p} + \mathbf{q}$$

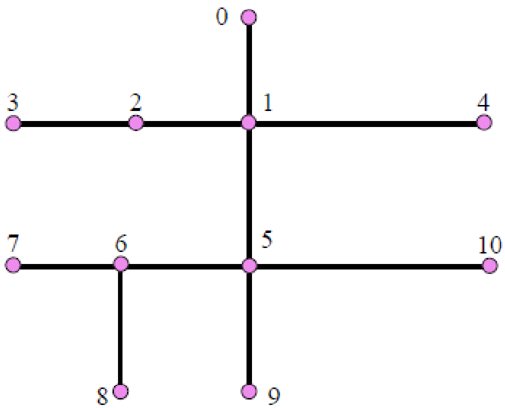
$$e^k \coloneqq \mathbf{Y}\left(\mathbf{v}^k - v_0^k 1_n\right) - 2\lambda\mathbf{p}^k - 2\mathbf{q}^k \qquad (5)$$

- **Our Model**: Based on (5) and a time window of length *K*, we develop a linear LS regression mode to estimate Y.

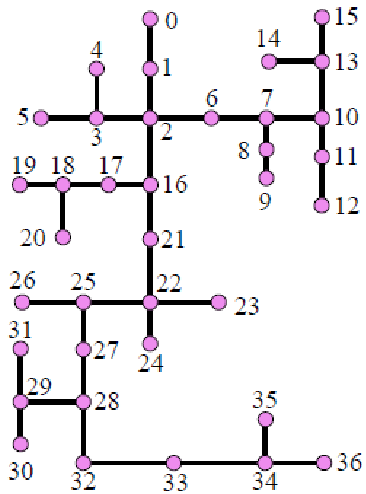
$$\min_{Y,\lambda} |[e^1, \dots, e^K]|_2^2 \qquad (6)$$

Estimated Weighted Laplacian Matrix for IEEE 13-, 37-, 69-Bus Test Feeder

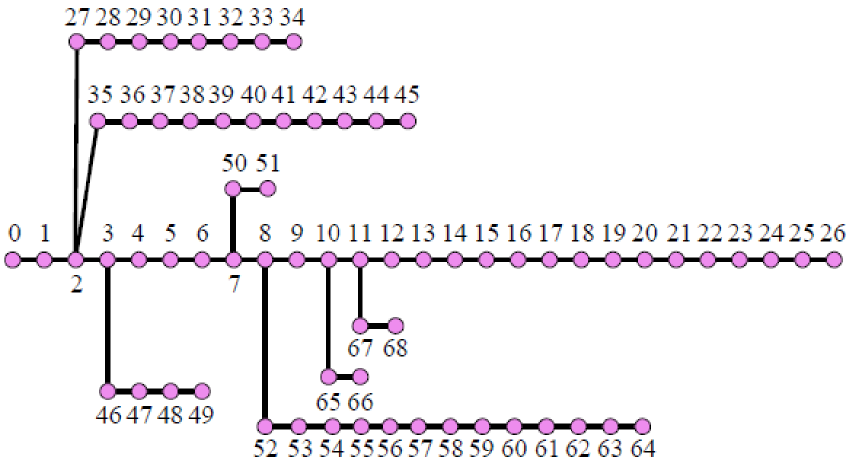
13-Bus Test Feeder



37-Bus Test Feeder



69-Bus Test Feeder

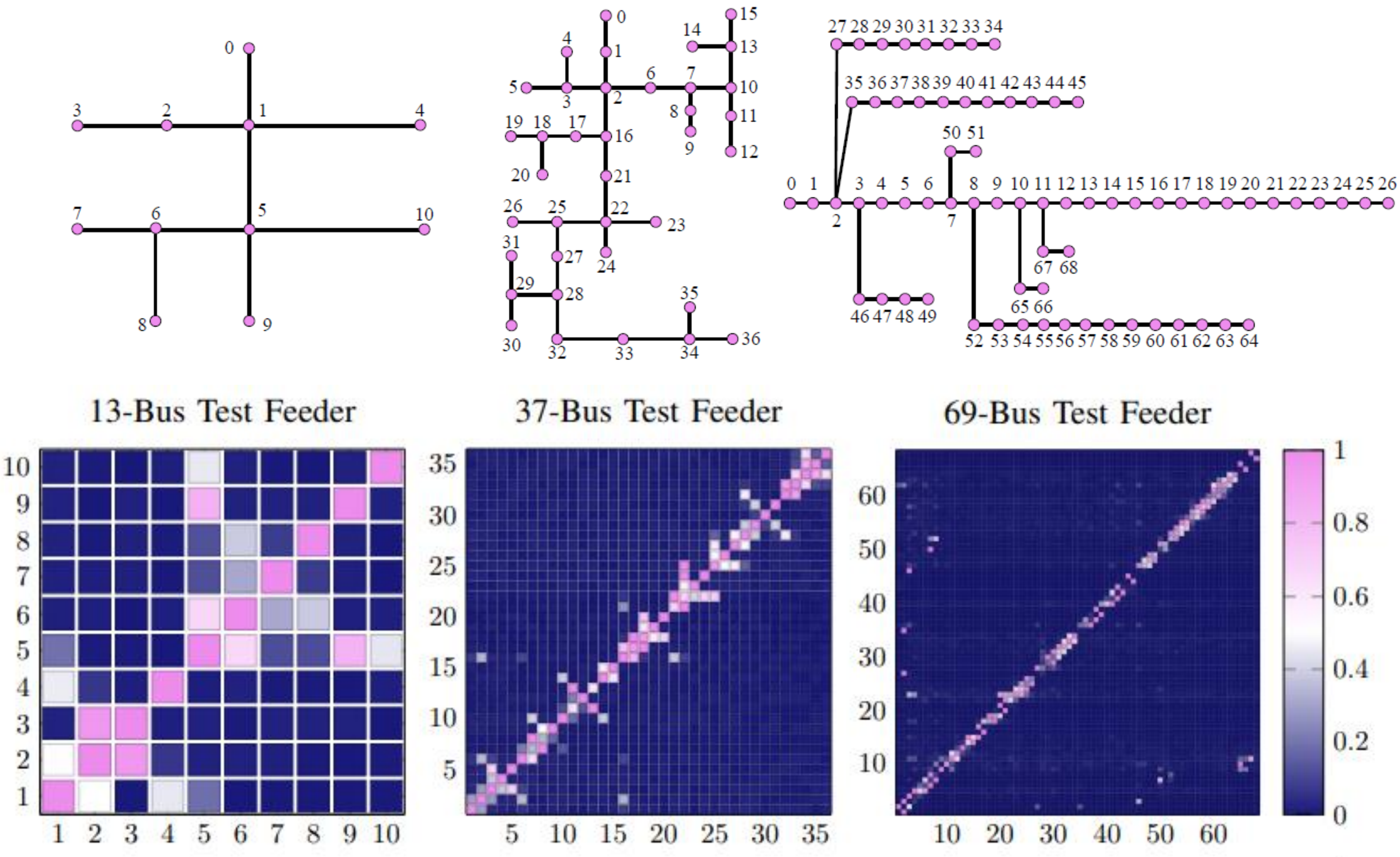


- The nodal load demand is calculated based on our real smart meter data with 1-h resolution. The length of window is selected as 200.

- Even though our method is derived on the assumption of a homogeneous R/X ratio, it shows the robustness to the systems with heterogeneous R/X ratios.

- The minimum and maximum R/X values of the three feeders are {0.5153, 2.0655}, {1.4536, 2.7482} and {0.4, 3.4}, respectively (***three heterogeneous feeders***).

Estimated Weighted Laplacian Matrix for IEEE 13-, 37-, 69-Bus Test Feeder



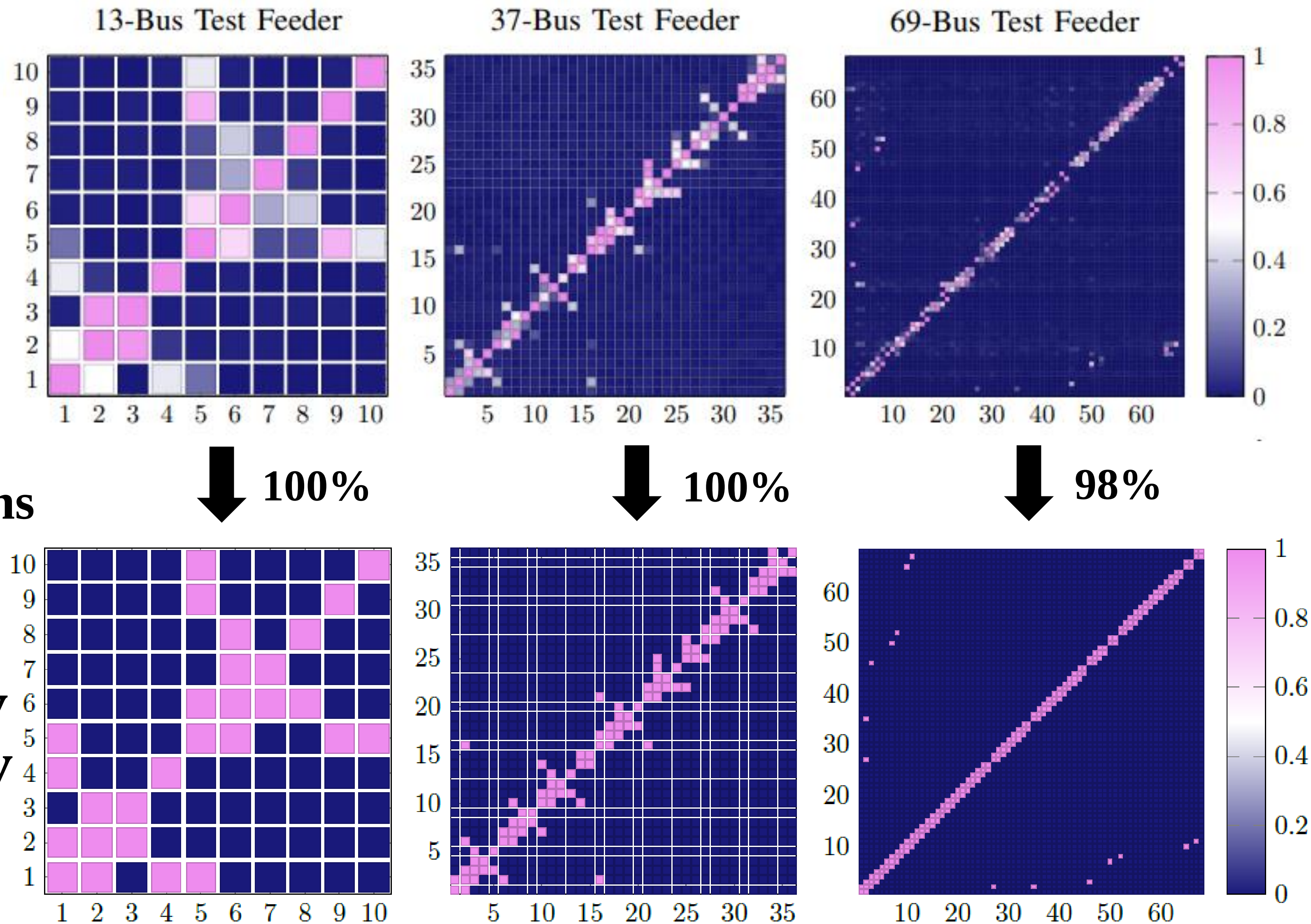
$\mathbf{Y}^*$ from (6)

Estimated Topology for IEEE 13-, 37-, 69-Bus Test Feeder

Y^* from (6)

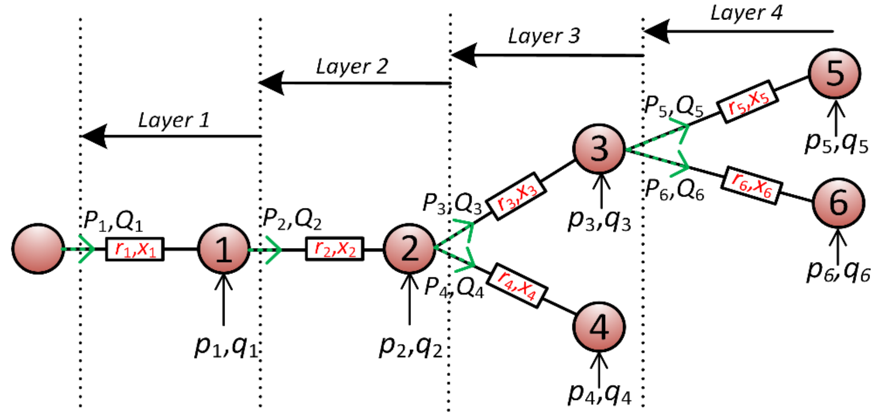
500 Simulations

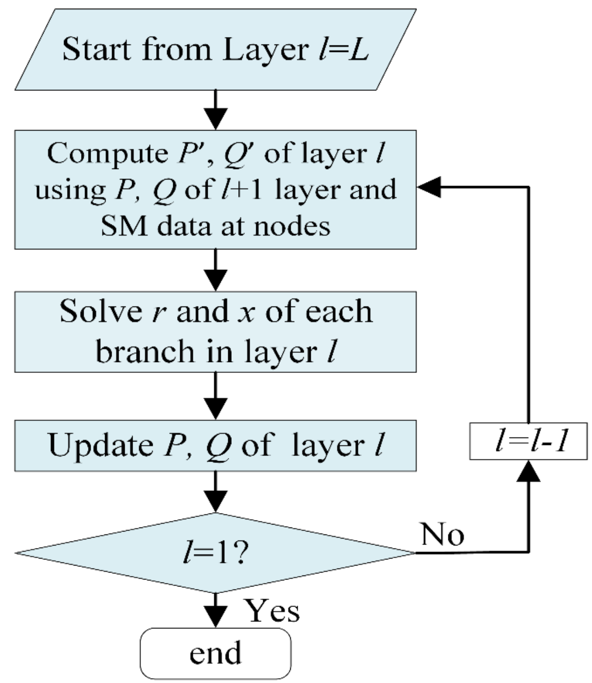
Topology Recovery



Bottom-Up Sweep Parameter Identification

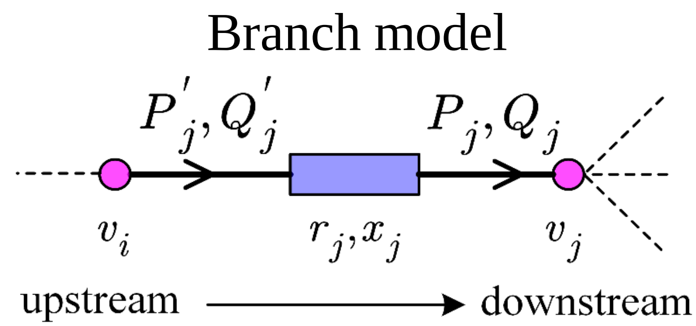
- Decomposing a radial distribution network into multiple layers labeled $1, \dots, L$ (where L is the bottom layer).
- Our bottom-up sweep algorithm performs the line flow and line parameter estimation in an alternating way based on the layers of the network.
- Addressing the dimensionality issue and enabling parallel computation of all line sections within the same layer.





Branch Flow Model

- The proposed line impedance estimation establishes on the voltage drop relationship over a branch that can be modeled as:



$$\text{Branch Flow: } P'_j = P_j + \frac{r_j\left(P_j^2+Q_j^2\right)}{v_j},\ \ Q'_j = Q_j + \frac{x_j\left(P_j^2+Q_j^2\right)}{v_j}$$

$$\text{Voltage: } v_i - v_j = 2\left(r_j P_j + x_j Q_j\right) + \frac{\left(r_j^2+x_j^2\right)\left(P_j^2+Q_j^2\right)}{v_j}$$

- P'_j and Q'_j denote power flows out of the upstream node i; P_j and Q_j denote power flows into the downstream node j.

- "Upstream" and "downstream" represent the relative positions of the nodes and power could flow in either direction.

Network Parameter Estimation Model

- The parameter estimation establishes on the voltage drop over a branch that is defined as follows:

$$e_j^k := v_i^k - v_j^k - 2\left(r_i P_j^k + x_j Q_j^k\right) - \left(R_j + X_j\right) \cdot \frac{\left[\left(P_j^k\right)^2 + \left(Q_j^k\right)^2\right]}{v_j^k} \quad (7)$$

- Based on (7) and the R/X ratio library, the line parameter estimation is cast as a mixed-integer ***nonlinear*** programming model.

$$\min_{\alpha_z, r_j, x_j, R_j, X_j} \left|\left|e_j\right|\right|_1 \quad (8)$$

subject to $R_j = r_j^2$

$$X_j = x_j^2$$

$$r_j = \sum_{z=1}^{Z} \lambda_z \alpha_z x_j$$

$$\sum_{z=1}^{Z} \alpha_z = 1, \alpha_z \in \{0,1\}, \forall z.$$

LAD Parameter Estimation Model

- The Big-M technique is exploited to linearize the bilinear term $\alpha_z x_j$.
- We rewrite (8) without L1-norm operator by introducing the auxiliary variables. The SDP relaxation is used to tackle the non-convex quadratic equalities.

$$\min_{\alpha_z, r_j, x_j, R_j, X_j, \theta_j^k} \sum_{k=1}^{K} \theta_j^k \tag{9}$$

$$\text{subject to} \quad \theta_j^k \geq e_j^k, \forall k$$

$$-\theta_j^k \leq e_j^k, \forall k$$

$$\mathbf{W}_j^r = \begin{bmatrix} 1 & r_j \\ r_j & R_j \end{bmatrix} \succcurlyeq 0, \text{rank}\{\mathbf{W}_j^r\} = 1, \forall j$$

$$\mathbf{W}_j^x = \begin{bmatrix} 1 & x_j \\ x_j & X_j \end{bmatrix} \succcurlyeq 0, \text{rank}\{\mathbf{W}_j^x\} = 1, \forall j$$

$$\sum_{z=1}^{Z} \alpha_z = 1, \alpha_z \in \{0,1\}, \forall z.$$

$$-M_j(1-\alpha_z) \leq r_j - \lambda_z x_j \leq M_j(1-\alpha_z), \forall z$$

LS Parameter Estimation Model

- The Big-M technique is exploited to linearize the bilinear term $\alpha_z x_j$.
- We rewrite (8) by introducing the auxiliary variable μ_j, and additionally imposing the constraints.
- Relaxing the quadratic equalities, we obtain a MISOCP model.

$$\min_{\alpha_z, r_j, x_j, R_j, X_j, \mu_j} \quad \mu_j \tag{10}$$

$$\text{subject to} \quad \left\| \begin{matrix} \frac{\mu_j - 1}{2} \\ e_j \end{matrix} \right\|_2 \leq \frac{\mu_j + 1}{2}$$

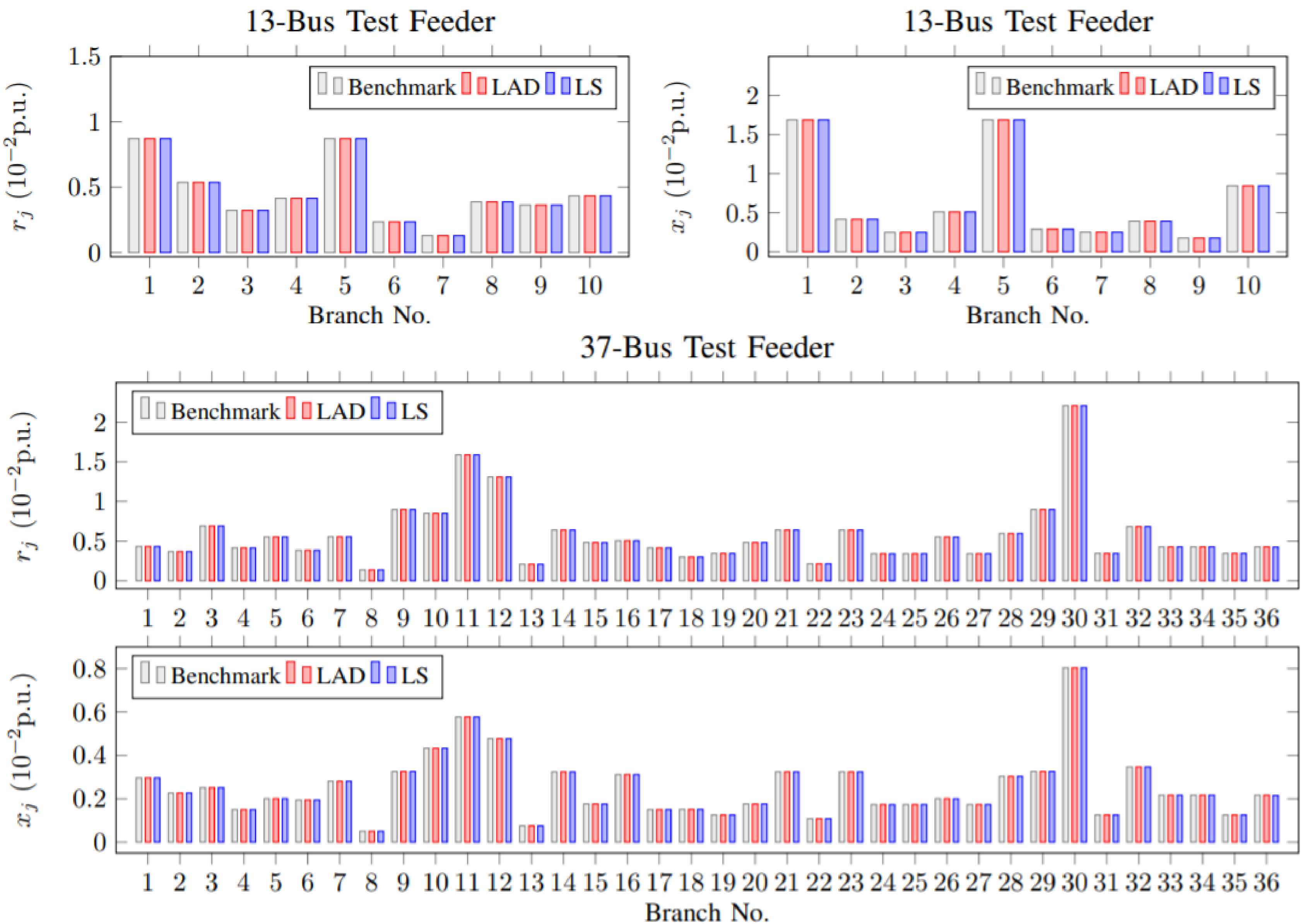
$$\left\| \begin{matrix} \frac{R_j - 1}{2} \\ r_j \end{matrix} \right\|_2 \leq \frac{R_j + 1}{2}$$

$$\left\| \begin{matrix} \frac{X_j - 1}{2} \\ x_j \end{matrix} \right\|_2 \leq \frac{X_j + 1}{2}$$

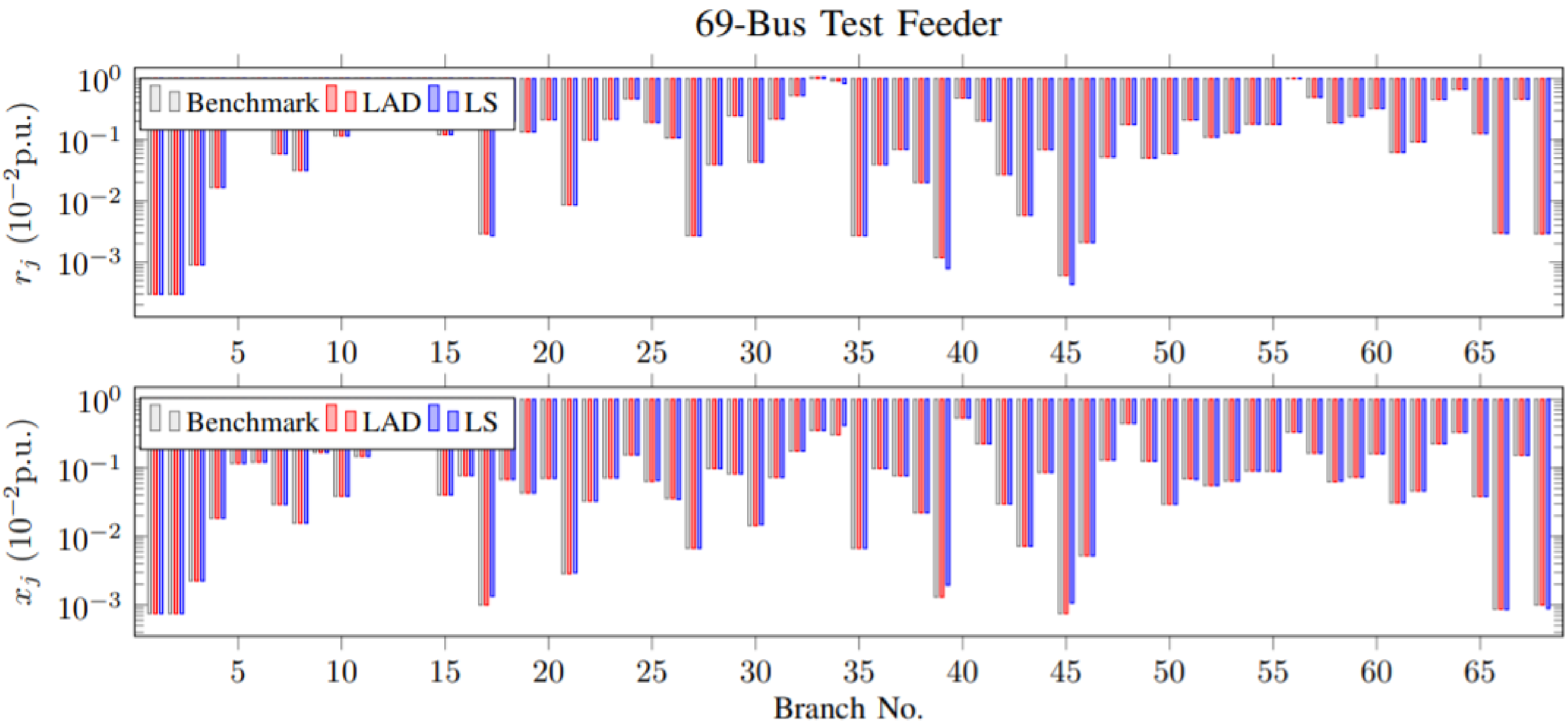
$$\sum_{z=1}^{Z} \alpha_z = 1, \alpha_z \in \{0,1\}, \forall z.$$

$$-M_j(1 - \alpha_z) \leq r_j - \lambda_z x_j \leq M_j(1 - \alpha_z), \forall z$$

Estimated Line Parameters for IEEE 13- and 37-Bus Test Feeder



Estimated Line Parameters for IEEE 69-Bus Test Feeder



Estimated Line Parameters for IEEE 69-Bus Test Feeder - Summary

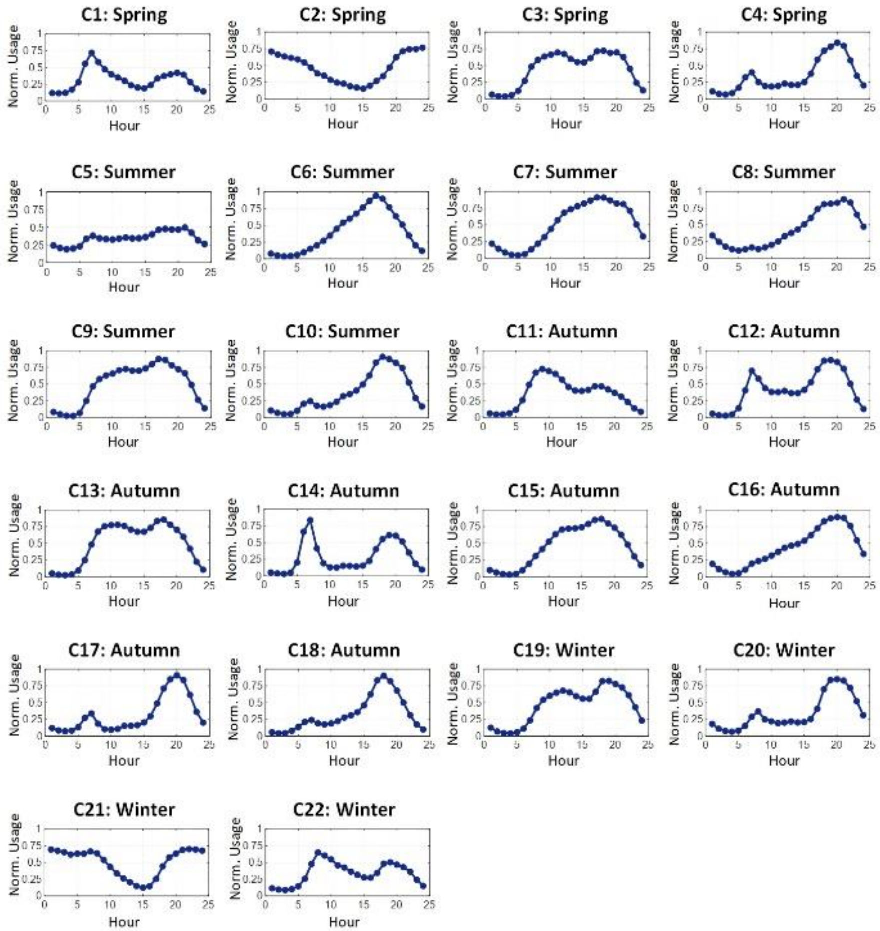
SDP-based LAD model:

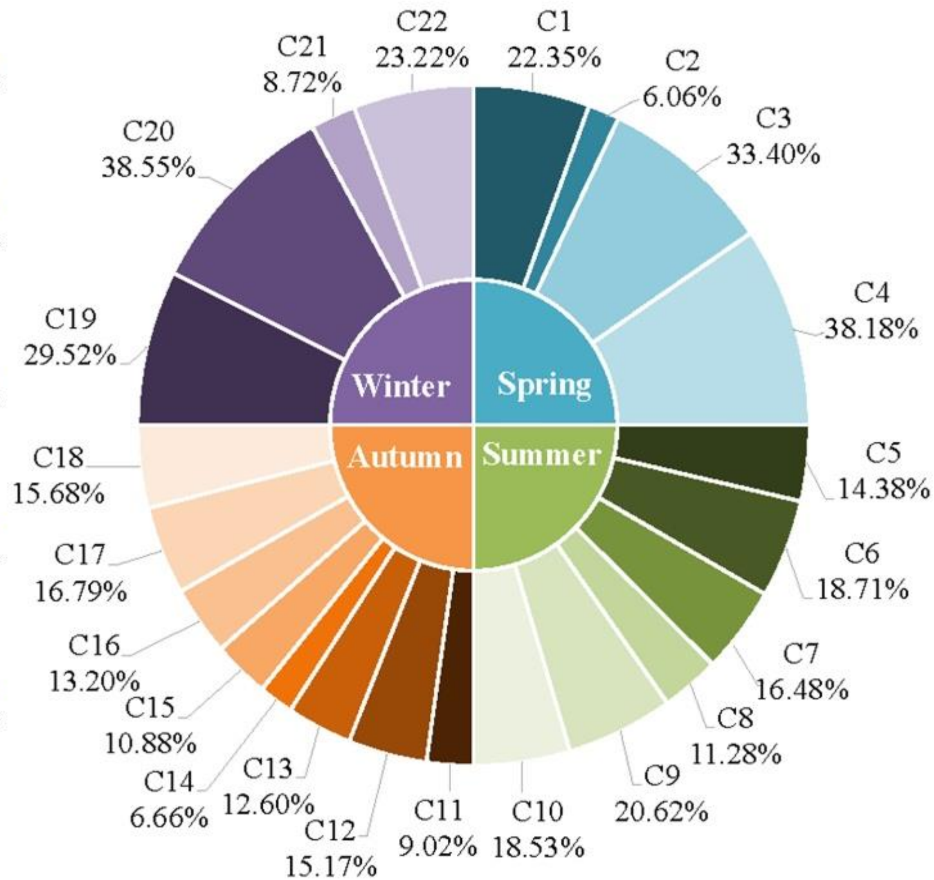
- For the IEEE 13-bus feeder, the largest relative errors (among all branches) for r_j and x_j are $3.331 \times 10^{-5}\%$ and $3.335 \times 10^{-5}\%$.
- For the IEEE 37-bus feeder, the largest relative errors (among all branches) for r_j and x_j are $3.402 \times 10^{-4}\%$ and $3.403 \times 10^{-4}\%$.
- For the IEEE 69-bus feeder, the largest relative errors (among all branches) for r_j and x_j are $1.444 \times 10^{-4}\%$ and $7.061 \times 10^{-5}\%$.

SOCP-based LS model:

- For the IEEE 13-bus feeder, the largest relative errors (among all branches) for r_j and x_j are 0.256% and 0.952%.
- For the IEEE 37-bus feeder, the largest relative errors (among all branches) for r_j and x_j are 0.251% and 0.958%.
- For the IEEE 69-bus feeder, the largest relative errors (among all branches) for r_j and x_j are 33.95% and 46.81%. But these large errors ($\geq 5\%$) only occur in a few branches with high R/X ratios (17,34,39,45, and 68).

Customer Load Profiling





- Customer typical load profiles are valuable for utilities to understand customer consumption behaviors.
- By using machine learning techniques, load profiling can be cast as an unsupervised clustering problem.

➢ **Curse of Dimensionality** ➢ **Algorithm-specific Limitations** ➢ **Hyperparameter Calibration**

Customer Peak Contribution Estimation

Problem Statement: Inferring residential customer peak contribution using customer monthly energy billing in *partially* observable distribution systems.

Challenges:

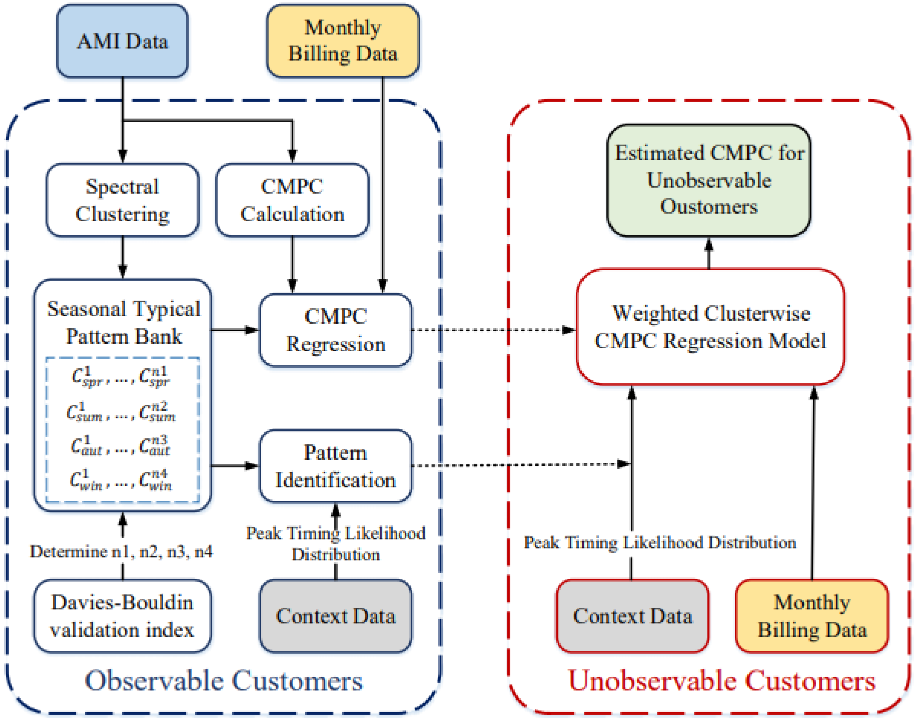
- System is partially observable – no meter for unobservable customers.
- Customers with high monthly billing do not necessarily have high customer peak load.
- Customers with high peak load do not necessarily have high customer peak contribution due to the noncoincidence between customers and system peak time.

Our Solution:

✓ Coincident monthly peak contribution (CMPC): ratios of individual customers' demands during daily peak load times of the system to the daily system peak demand.

✓ For unobservable customers without SMs, a weighted cluster-wise regression method can be used to estimate CMPC using their monthly billing information.

✓ The basic idea is to exploits the strong correlation between CMPC and monthly energy consumption when the customers' load profiles are similar.



CMPC Regression: Mapping CMPC to customer billing data for each typical pattern

Pattern Identification: Mapping context data to typical patterns

Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A Data-Driven Customer Segmentation Strategy Based on Contribution to System Peak Demand," IEEE Transactions on Power Systems, vol. 35, no. 5, pp. 4026-4035, September 2020.

Probabilistic Graphical Learning for Outage Detection and Location

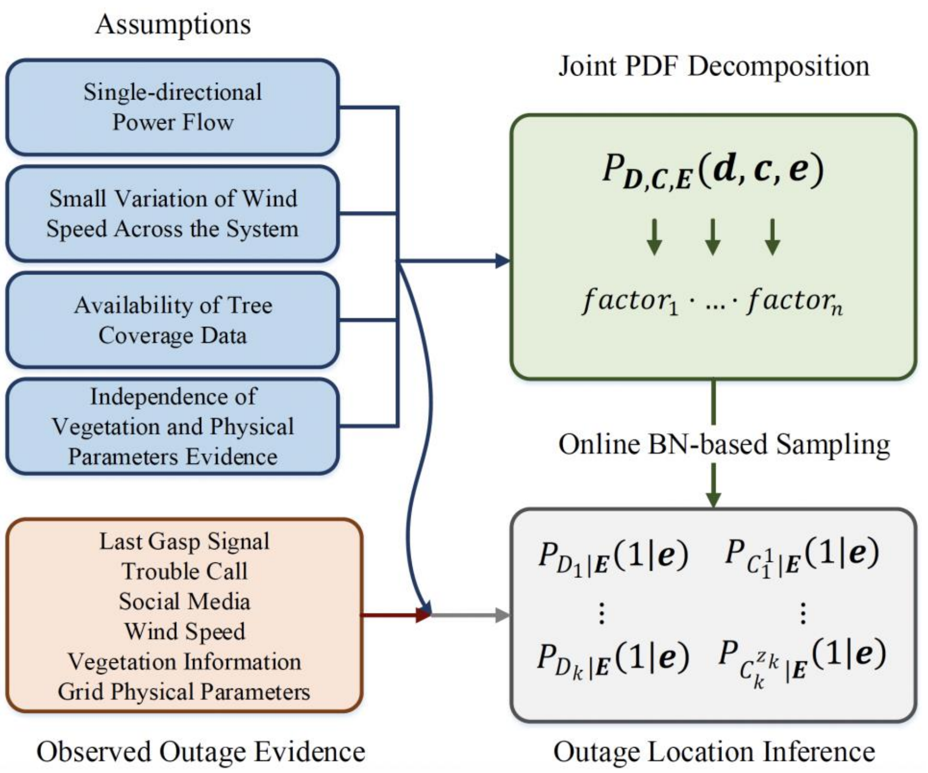
Problem Statement: SMs have capacitors that can generate last-gasp signals to report outages down to laterals. However, some networks may not have full SM coverage, and SMs are subject to misreports and communication failures. A solution is to fuse SMs with trouble calls, social media reports, and weather/fragility information.

Challenges:

- Heterogeneity of diverse outage data sources, including availabilities, confidence levels, and contradiction.
- Lacking scalability: Huge joint PDFs for large system & Large computational burden in calculating the probabilities of post-event topology candidates.

Our Solution:

- ✓ Propose a probabilistic graphical learning approach to encode distribution grids and heterogeneous outage data sources into probabilistic graphs
- ✓ Leverage the conditional independencies inherent in the grid and data and fragility model to simplify the probabilistic graphical modeling to improve scalability
- ✓ Use a Gibbs sampling method to overcome the scalability issue in online probabilistic outage location inference.

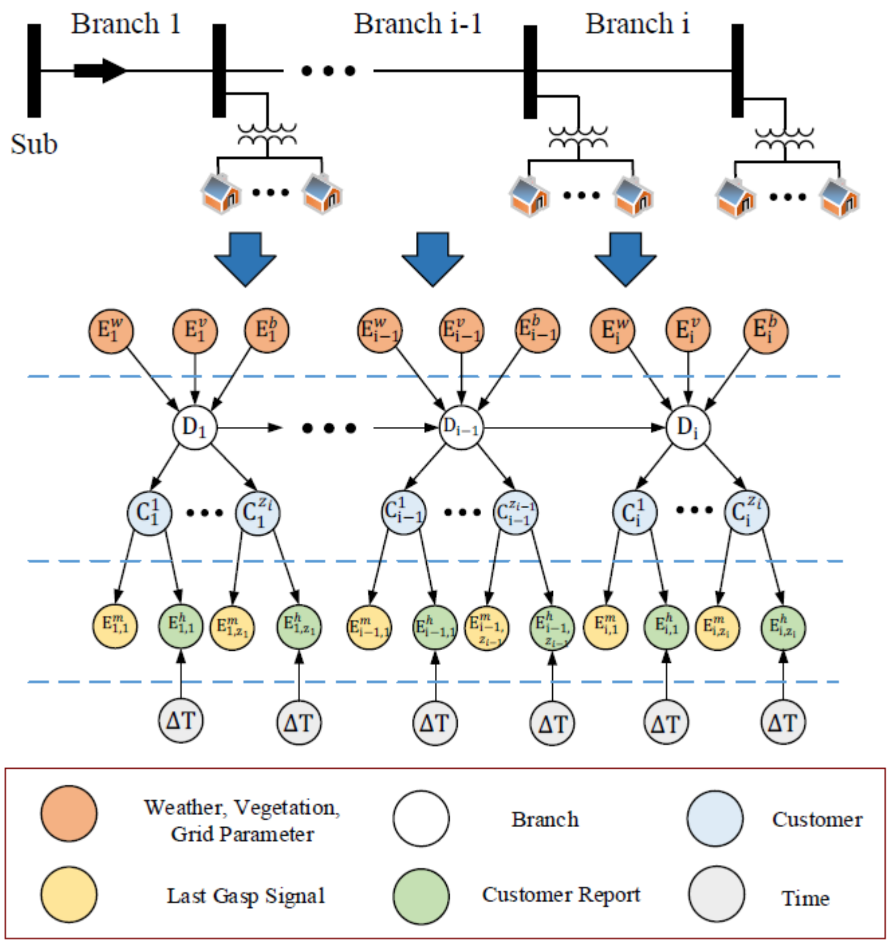


Y. Yuan, K. Dehghanpour, Z. Wang, and F. Bu, "Multi-Source Data Fusion Outage Location in Distribution Systems via Probabilistic Graphical Models," IEEE Transactions on Smart Grid, accepted for publication.

Probabilistic Graphical Learning for Outage Detection and Location

Our idea:

- An inherent feature of radial grids is their tree-like structure, resulting in a unique, one-directional path between all nodes. If this path is disrupted at any branch, the states of all downstream branches can be inferred as de-energized without the need for further search. This feature can be represented using a Bayesian Network (BN).
- The nodes in the graph represent states of branches/customers and outage data sources. The edges represent probabilistic influence of one node on another.
- Each parent variable in the BN is the immediate causal source of influence for its child node; i.e., knowing the parent variable is sufficient to determine the probability distribution of the child.
- The underlying principle is that any node in the graph is conditionally independent of its upstream nodes if the values of its parents are known or inferred.
- These conditional independencies enable a compact and scalable graphical representation of different data and significantly accelerate outage detection.



Y. Yuan, K. Dehghanpour, Z. Wang, and F. Bu, "Multi-Source Data Fusion Outage Location in Distribution Systems via Probabilistic Graphical Models," IEEE Transactions on Smart Grid, accepted for publication.

Conclusion and Future Work

- Smart meter data, although may be of low resolution and limited measurement variables, can still be used to greatly help distribution system monitoring and operation. There are many applications such as network modeling, outage detection and behind-the-meter solar disaggregation.

- We demonstrated how to use smart meter data together with optimization and machine learning to estimate topology and line parameters in radial distribution systems.

- In the future, we will focus on using smart meter data to identify/calibrate network models in unbalanced mesh distribution systems.

Distribution Course Material Sharing

EE653: Power distribution system modeling, optimization and simulation

- Introduction to Distribution Systems
- Modeling Series Components – Distribution Lines
- Modeling Series Impedance of Overhead and Underground Lines
- Modeling Shunt Admittance of Overhead and Underground Lines
- Modeling Shunt Components – Loads and Caps
- Distribution Feeder Modeling and Analysis Part I
- Modeling Voltage Regulators
- Modeling Three-Phase Transformers
- Distribution Feeder Modeling and Analysis Part II
- Various Power Flow Calculation Methods in Distribution Systems
- Optimal Power Flow in Distribution Systems
- Voltage/VAR Optimization and Conservation Voltage Reduction
- Distribution System State Estimation and Smart Meter Data Analytics
- Microgrids – Introduction and Energy Management
- Microgrids – Dynamic Modeling and Control
- OpenDSS Tutorial
- Real Distribution System Modeling and Analysis using OpenDSS
- Introduction to GridLAB-D
- Distribution System Resilience: Hardening, Preparation and Restoration
- Energy Storage

- You may download the course material at: http://wzy.ece.iastate.edu
- All slides are editable, feel free to use.
- Comments are very welcome!

Reference

[1] J. Yu, Y. Weng, and R. Rajagopal, "PaToPaEM: A data-driven parameter and topology joint estimation framework for time-varying system in distribution grids," IEEE Transactions on Power Systems, vol. 34, no. 3, pp. 1682-1692, May 2019.

[2] O. Ardakanian, Y. Yuan, V. Wong, R. Dobbe, S. Low, A. von Meier, and C. J. Tomlin, "On identification of distribution grids," IEEE Transactions on Control of Network Systems, vol. 6, no. 3, pp. 950-960, Sep. 2019.

[3] Y. Yuan, S. H. Low, O. Ardakanian, and C. J. Tomlin, "Inverse Power Flow Problem," Online Available: https://arxiv.org/pdf/1610.06631.pdf.

[4] A. M. Prostejovsky, O. Gehrke, A. M. Kosek, T. Strasser, and H. W. Bindner, "Distribution line parameter estimation under consideration of measurement tolerances," IEEE Transactions Industrial Informatics, vol.12, no.2, pp. 726–735, 2016.

[5] H. Xu, A. D. Domínguez-García, and P. W. Sauer, "A data-driven voltage control framework for power distribution systems," In Proceedings of IEEE Power and Energy Society General Meeting, pp. 1-5, Aug. 2018.

[6] W. Wang and N. Yu, "Parameter estimation in three-phase power distribution networks using smart meter data,"2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–6, 2020.

Thank You!
Q & A

Fully funded and immediately open PhD and Postdoc positions are available!

http://wzy.ece.iastate.edu

Recovering Topology From Estimated $\mathbf{Y}^*$

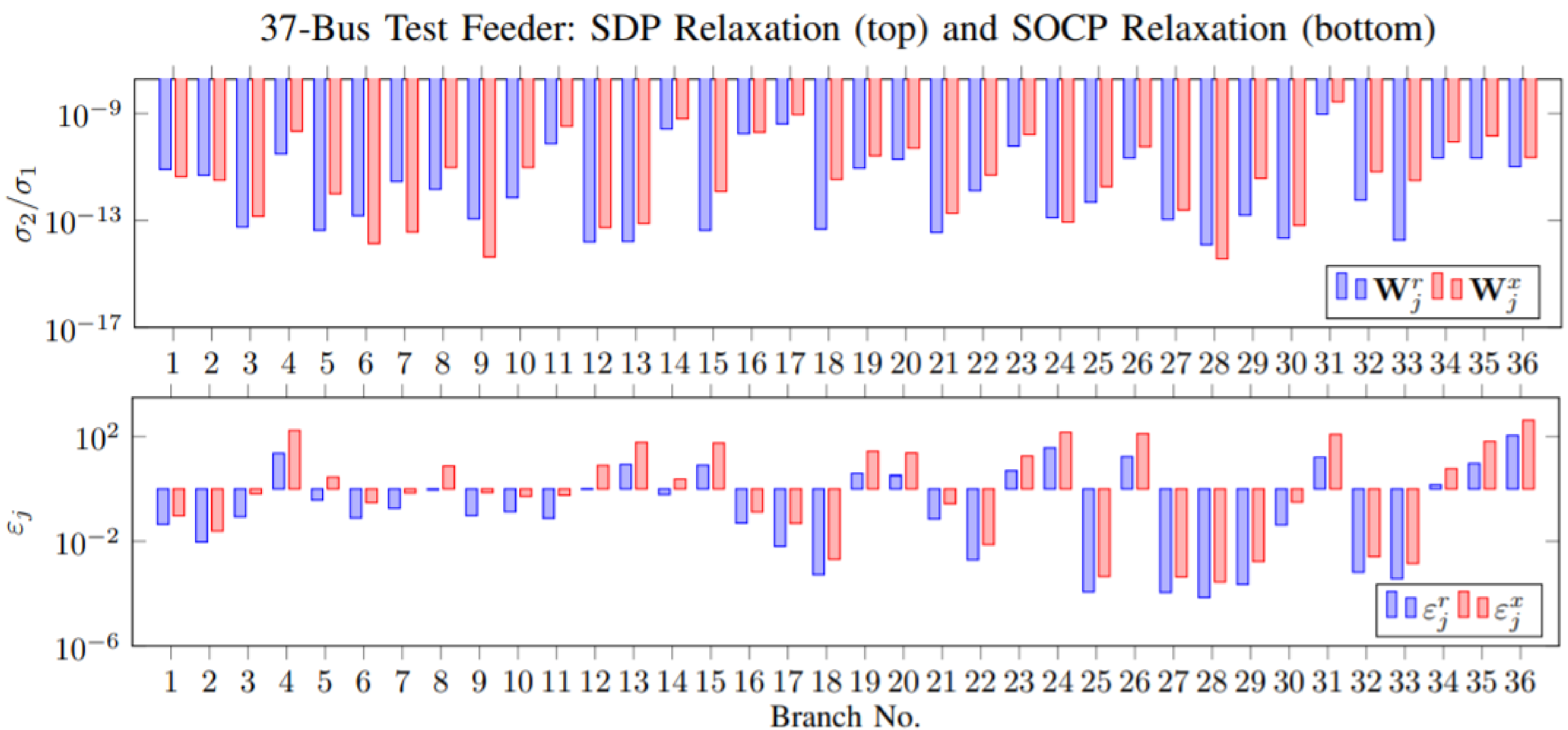
- Recovering the topology from $\mathbf{Y}^*$ is cast as an *anomaly detection* problem.
- **Our Solution**: We have utilized a density based spatial clustering of applications with noise (DBSCAN) method to extract the topology from $\mathbf{Y}^*$.
- DBSCAN can marking as anomaly points that lie alone in low-density regions
- **Advantage**: DBSCAN can discover clusters with arbitrary shapes.
- DBSCAN does not require a prior specification on the number of clusters.

Algorithm 1 Recovering Topology From $\mathbf{Y}^\star$ by Clustering

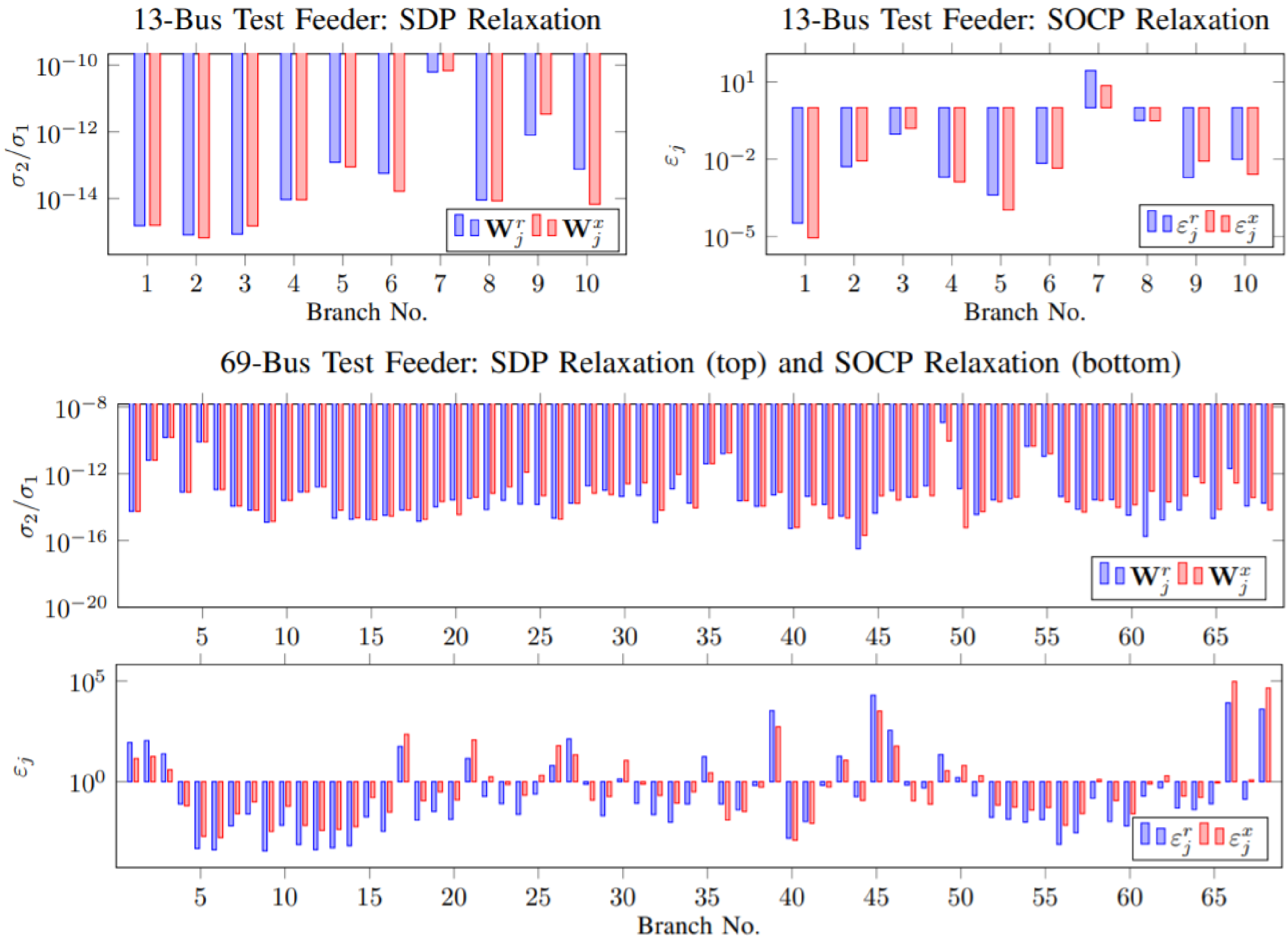
Initialization: Initialize $i \leftarrow 1, j \leftarrow 1$, γ, ξ
repeat
  [**S1**]: Select the ith row of $\mathbf{Y}^\star$.
  repeat
    [**S2**]: Pick $y^\star_{ij}$ and retrieve all direct density-reachable points using ξ.
    [**S3**]: Based on γ, if $y^\star_{ij}$ is a core point, a cluster is formed; otherwise, update $j \leftarrow j + 1$.
  until $j = n$ or no new point can be added to any cluster
  [**S4**]: Update $i \leftarrow i + 1$.
until $i = n$.

Exactness of SDP and SOCP Relaxation of IEEE 37-Bus Test Feeders

- To quantify the exactness of SDP and SOCP relaxation in (9) and (10), we compute the ration between the largest two eigenvalues of $\mathbf{W}_j^r$ and $\mathbf{W}_j^x$ and the resultant errors ε_j^r and ε_j^x, respectively.



Exactness of SDP and SOCP Relaxation of IEEE 13- and 69-Bus Test Feeders



✓ **The SDP relaxation is exact on all branches in three cases.**