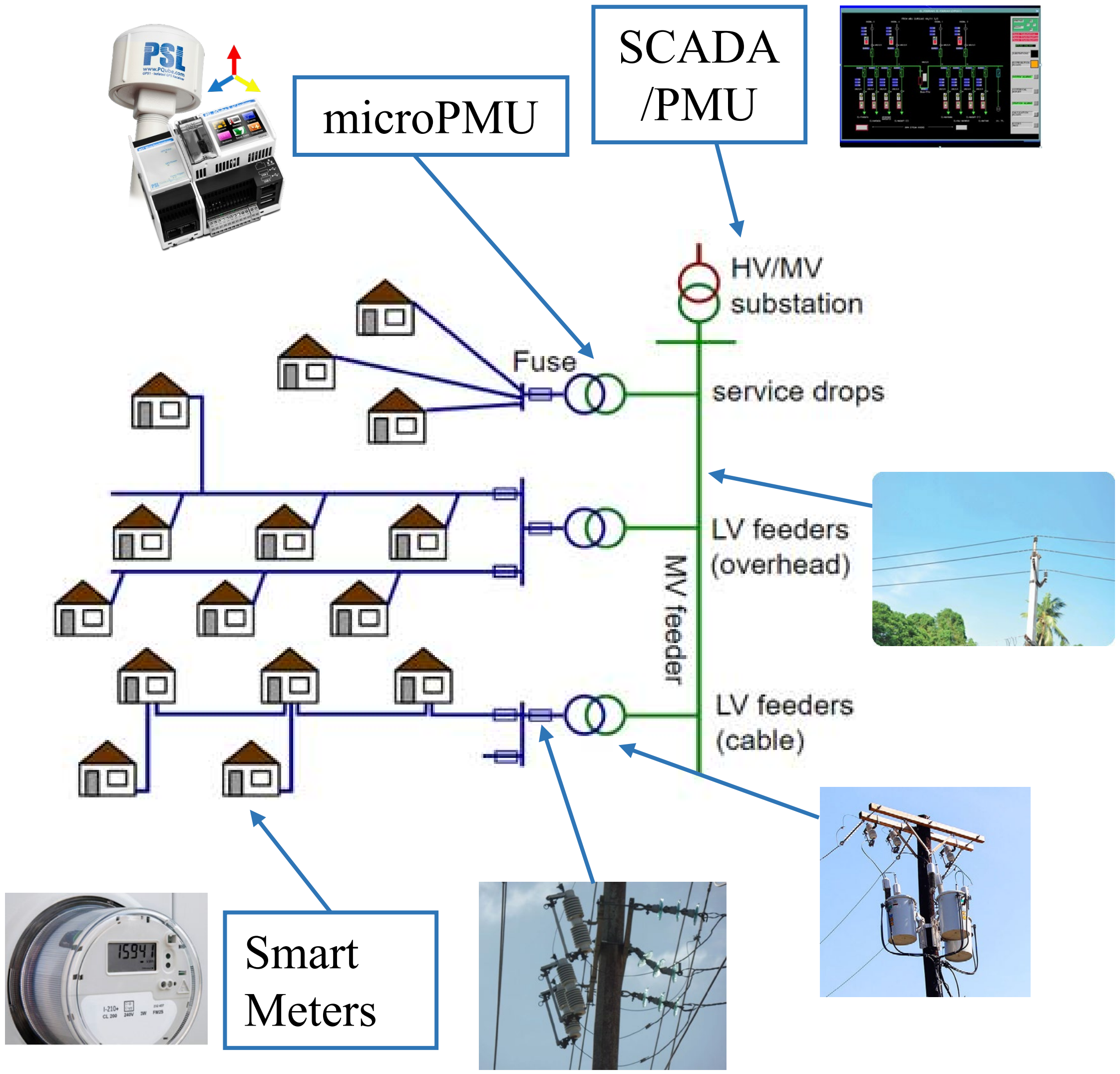
Mining Smart Meter Data for Improving Distribution Grid Operation and Resilience

Zhaoyu Wang
Harpole-Pentair Endowed Assistant Professor
Department of Electrical and Computer Engineering
Iowa State University
http://wzy.ece.iastate.edu

Content

- Real utility data and data sharing
- Multi-timescale load inference for observability enhancement
- Data-driven assessment of cold load pick-up demand
- Conclusions

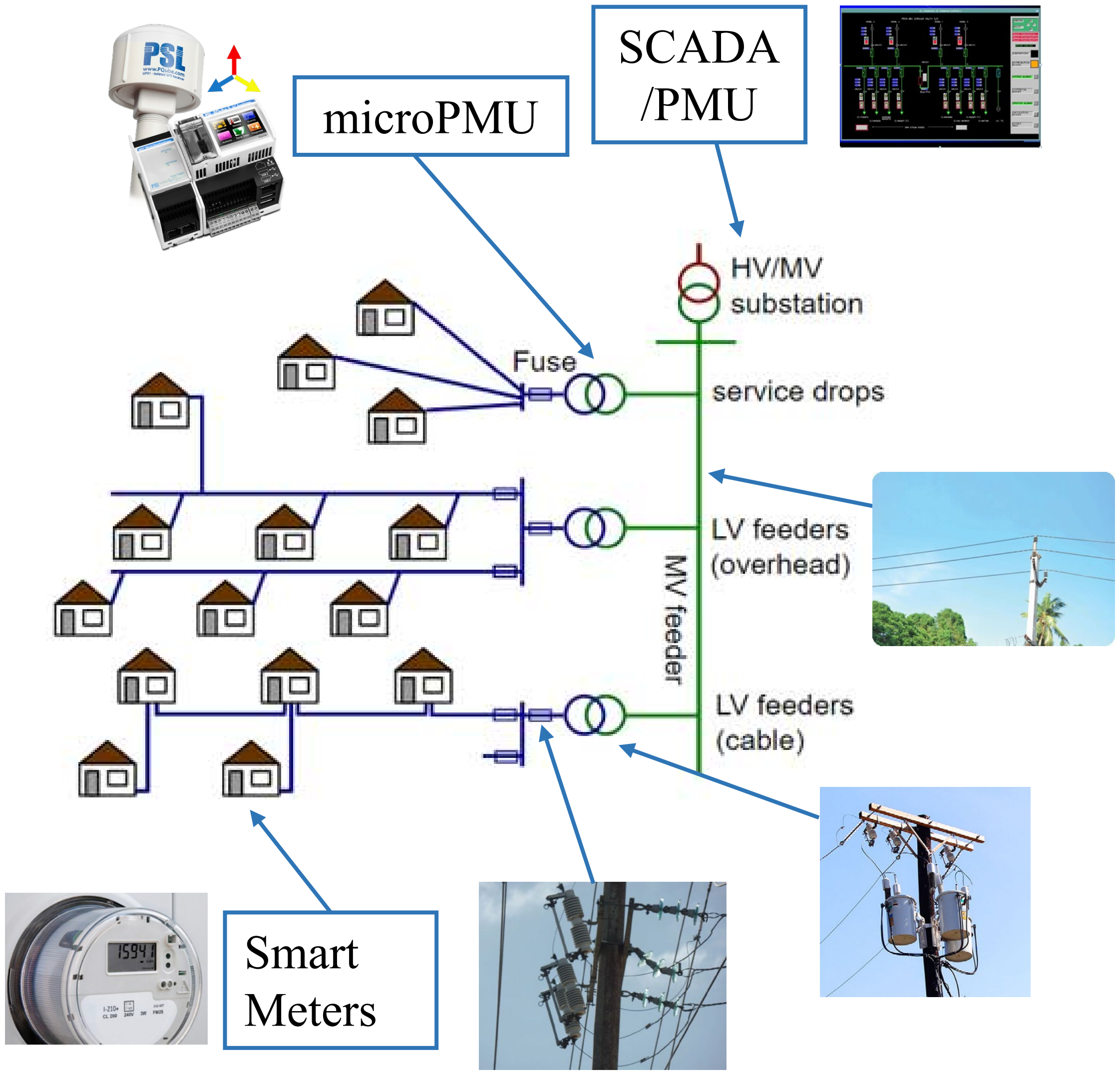
Power Distribution Grid Data



A Power distribution grid

- Where does the data come from?
  - SCADA (supervisory control and data acquisition); Smart Meters; Protection Devices; (micro)PMUs (phasor measurement units)
  - Measures voltage/current/frequency at different resolutions
- What are smart meters?
  - Different from conventional energy meters
  - Stay in your homes (not every home has it)
  - Measure energy and voltage
  - 15/30/60-minute resolution
- What are barriers to apply big data techniques in power industry?
  - Critical infrastructure
  - Conservative
  - Confidentiality

Power Distribution Grid Data



A Power distribution grid

- Where does the data come from?
  - SCADA (supervisory control and data acquisition); Smart Meters; Protection Devices; (micro)PMUs (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
- Features of smart meters
  - Very low resolution
  - Limited sensing capability
  - Severe data quality issues
  - **Is it a bad data source??**
  - **But, they are widely deployed!**

Real Data from Utilities

- We have NDAs with following utilities: MidAmerican Energy, Alliant Energy, Cedar Falls Utilities, Algona Municipal Utilities, Maquoketa Valley Electric Coop, Bloomfield, WAPA…
- We have multi-year PMU/SCADA/Smart Meter data from utility partners.

Data Type	Utilities	Measurement Locations	Data Length	Renewable Penetration	Historical Commands
AMI & SCADA	MVEC	140,000 customers	24 months with continuous updating	~45% relative to peak	Yes
AMI & SCADA	Alliant	10 substations	24 months with continuous updating	~35% relative to peak	Yes
AMI	CFU	2,500 customers	18 months with continuous updating	~10% relative to peak	Yes
PMU/SCADA	MidAmerican	3 Substations	24 months with continuous updating	~40% relative to peak	Yes
AMI&SCADA	Algona	3,000 customers	30 months	Unknown	N/A
SCADA	GPC	5 Substations	5 months	N/A	Yes
SCADA	Ameren	4 Substations	12 months	Unknown	Yes
SCADA	BGE	4 Substations	5 months	Unknown	Yes

Exemplary Real Data from Utilities

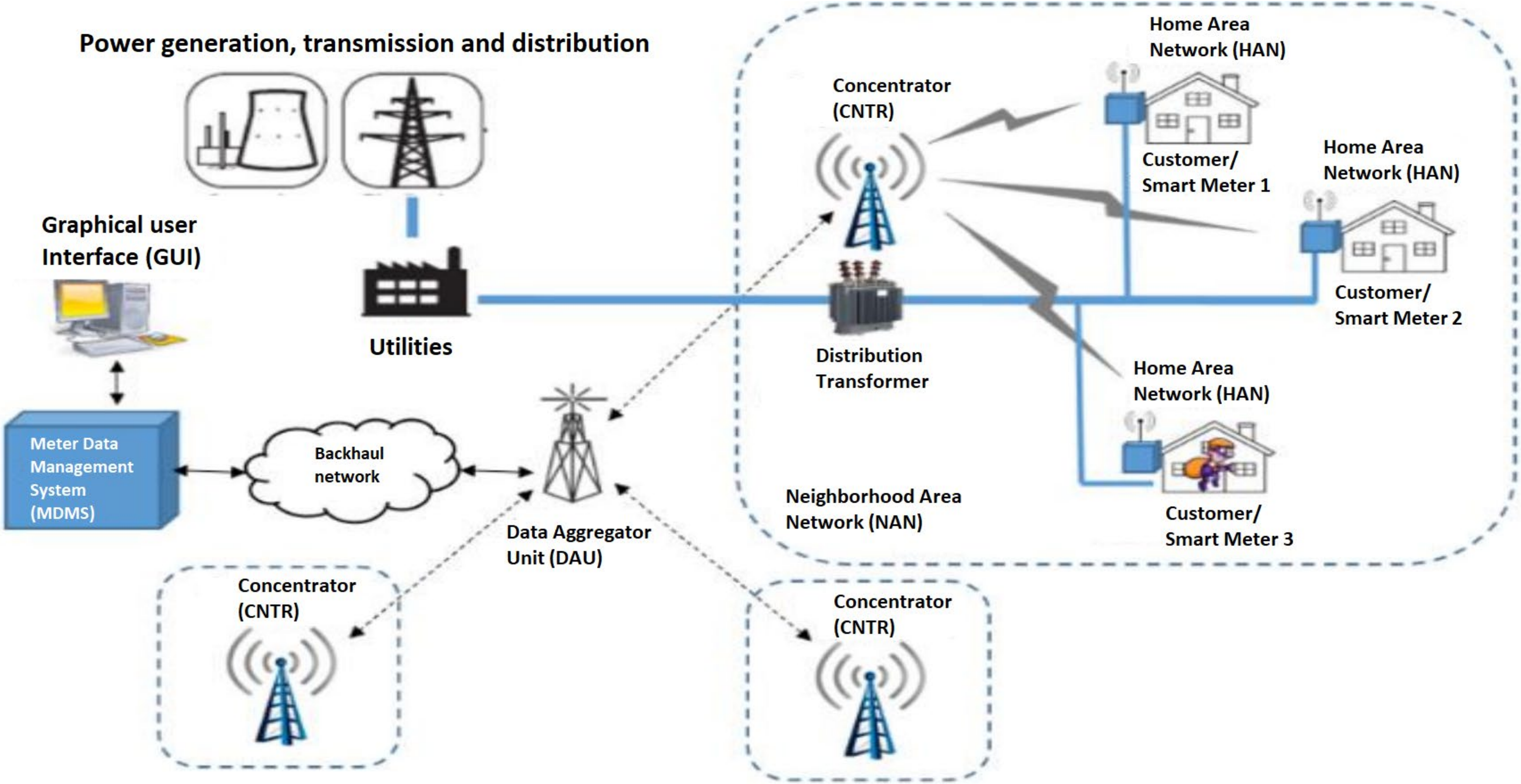
- More AMI data and circuit models:

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%

Smart Meter Data Collection



K. K. Kee, S. M. F. Shahab and C. J. Loh, “Design and development of an innovative smart metering system with GUI-based NTL detection platform”

Real Data from Utilities

An exemplary distribution system and associated SM data from our utility partner:

System Information

- 2 substations
- 4 load tap changing substation transformers (69/13.8 kV)
- 14 feeders (83 miles)
- 1489 overhead line sections
- 2582 underground cable sections
- 5 capacitor banks
- 361 switching devices
- >1000 distribution transformers
- 5212 customers

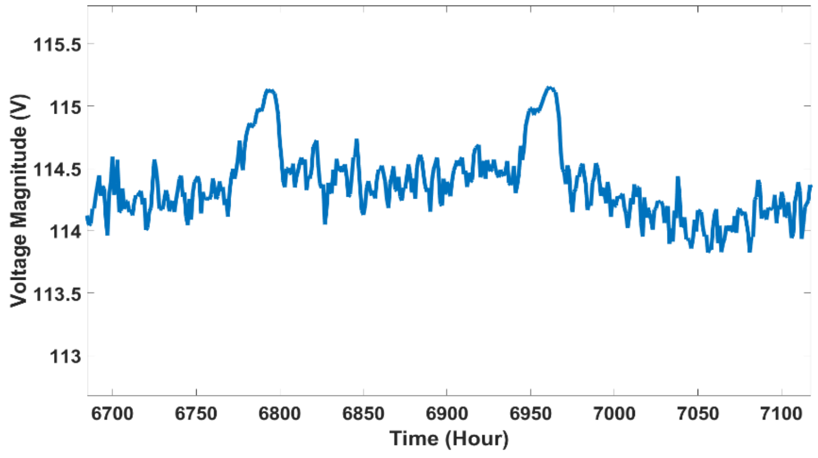
Smart Meter Data

- Time period: 4 years (2015-2018)
- 4321 residential customers
- 696 small commercial customers
- 146 large commercial customers
- 17 industrial customers
- 32 other customers
- Time resolution:
  - Hourly – residential, small commercial
  - 15-min – large commercial, industrial

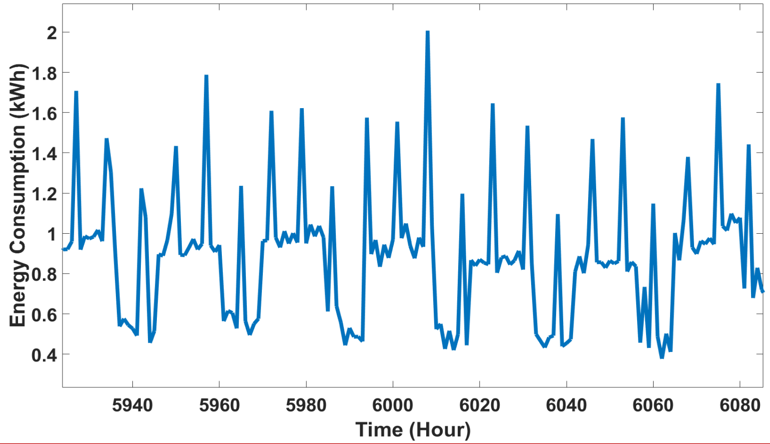
Real Data from Utilities

Smart Meter Measurement Data For Load Monitoring

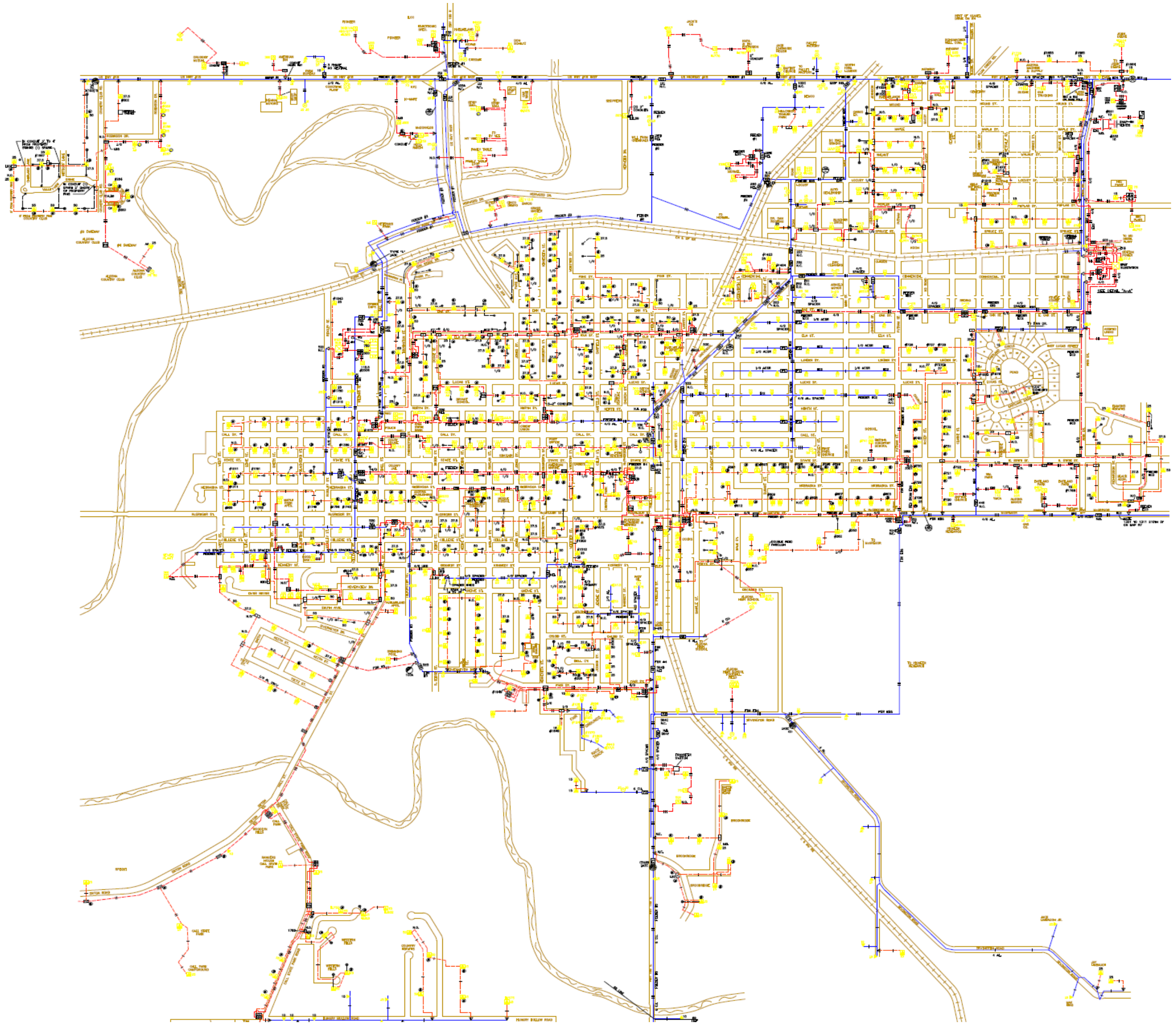
Sample Customer Voltage



Sample Customer Energy Consumption



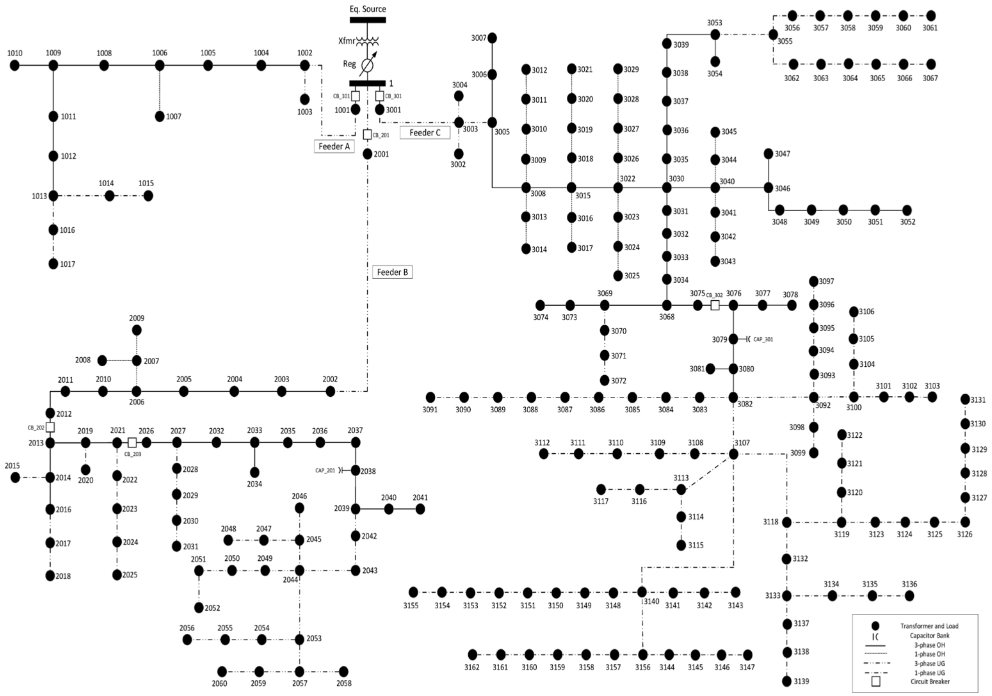
Network Topology/Model Information



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.
- 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.



Test system diagram

The dataset has been viewed/downloaded more than 10,000 times since June 12, 2019

Distribution Course Material Sharing

EE653: Power distribution system modeling, optimization and simulation

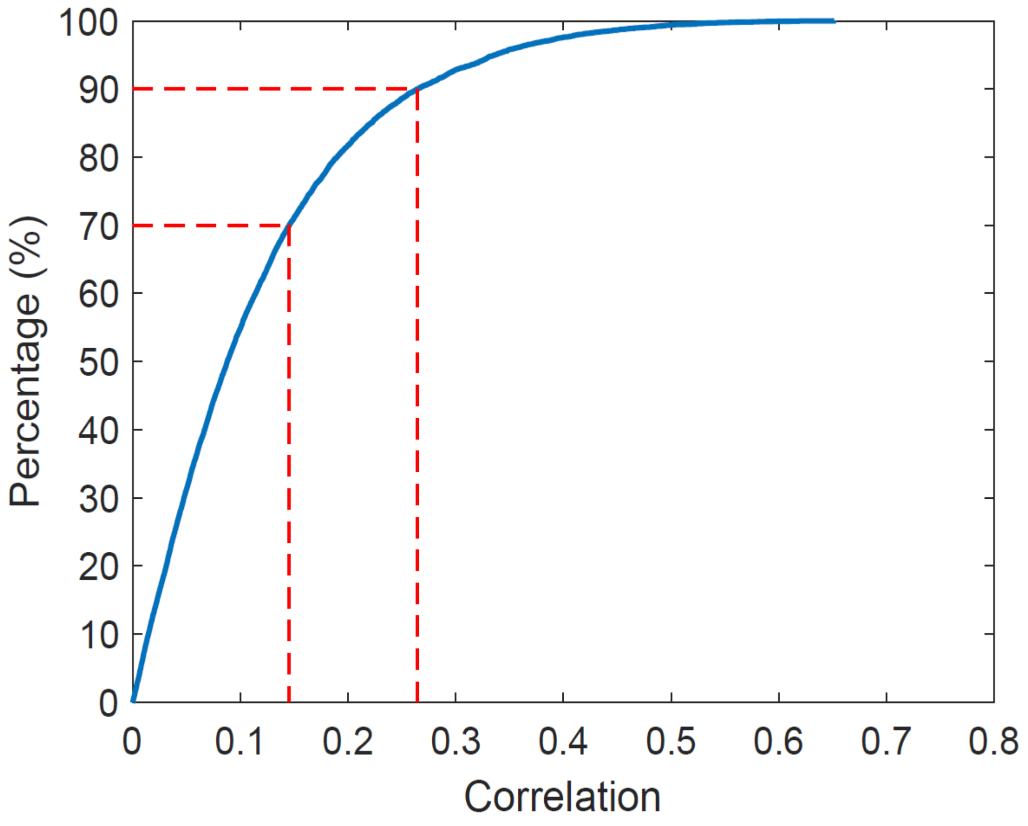
- Introduction to Distribution Systems
- Modeling Series Components – Distribution Lines
- Modeling Series Impedance of Overhead and Underground Lines
- Modeling Shunt Admittance of Overhead and Underground Lines
- Modeling Shunt Components – Loads and Caps
- Distribution Feeder Modeling and Analysis 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!
- The slides have been downloaded more than 5,000 times since Dec. 25, 2019

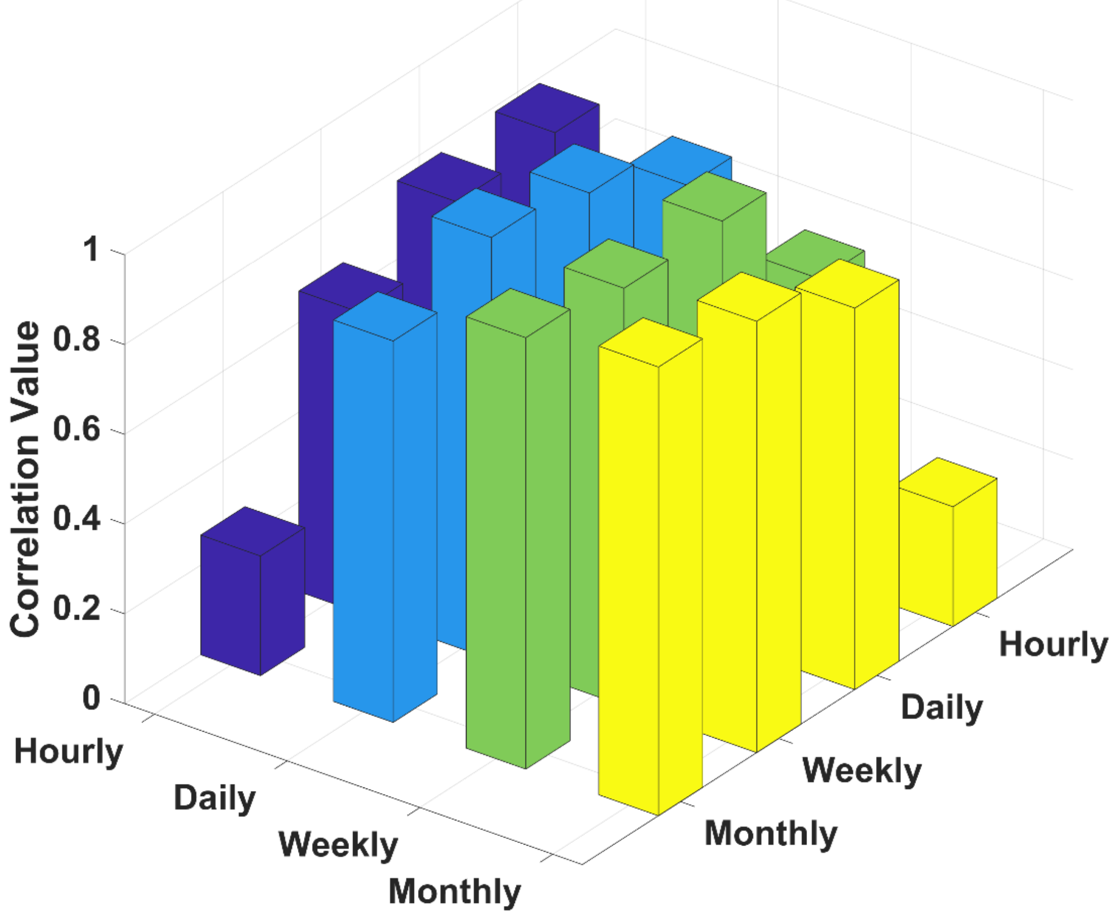
Multi-timescale Load Inference

- **Problem Statement**: Inferring hourly consumption data from customer monthly billing information as pseudo-measurements in partially observable systems
- **Challenges:**
  - ✓Loss of correlation between consumption time-series at different time-scales
  - ✓Unobserved customers' unknown typical behaviors
- **Solution Strategy**: Extending observability from observed customers to unobserved customers
- **Proposed Solution:**
  - ✓Multi-timescale load inference (stage by stage inference chain)
  - ✓Using data clustering for capturing customer typical behaviors
  - ✓Using state-estimation-based Bayesian learning for inferring unobserved customers' typical behaviors

Evidence from Data: How to Maintain Correlation



Very Small Correlation Between Different Customers' Smart Meter Time-Series: 90% below 0.27 (Loss of Correlation Across Customers)



Average Correlation between Consumption of All Customers Decreases from Monthly to Hourly (Loss of Correlation Across Different Time-Scales)

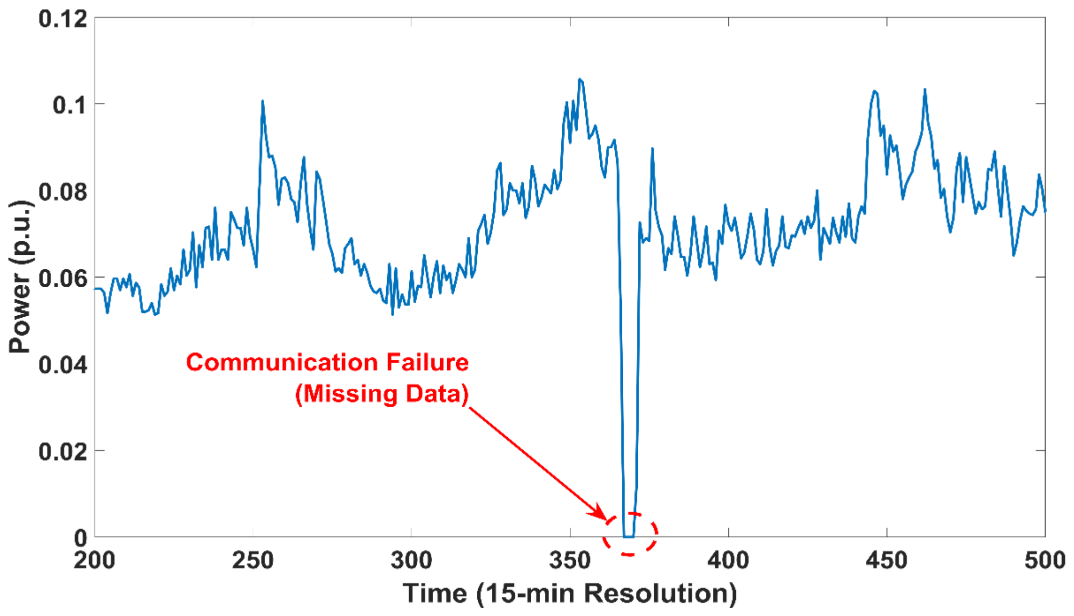
Solution Step I: Smart Meter Data Pre-Processing

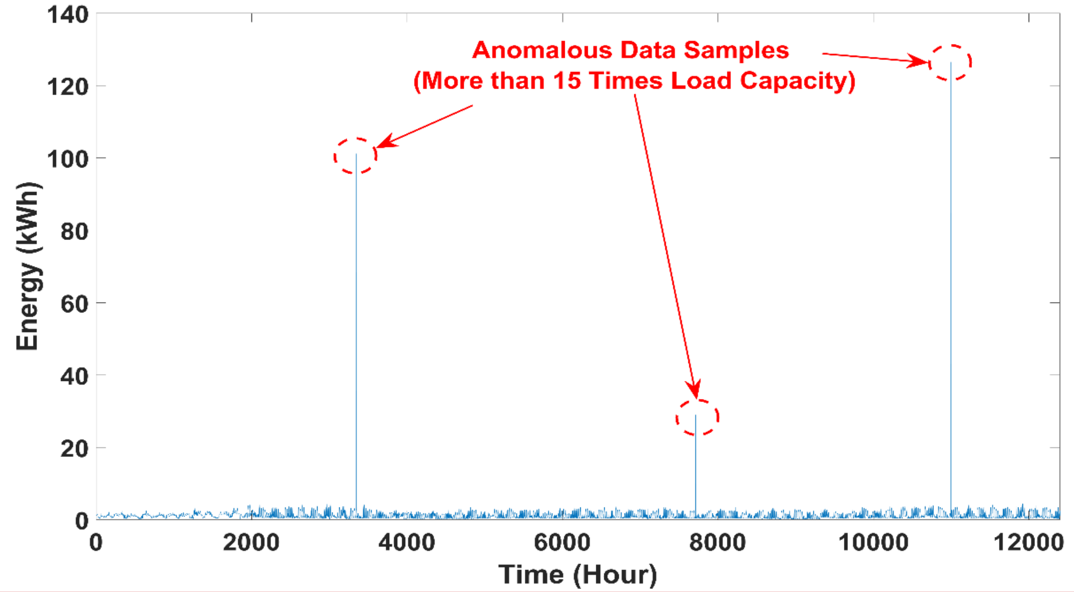
✓**Smart Meter Data Problems:**

- Outliers/Bad Data
- Communication Failure
- Missing Data

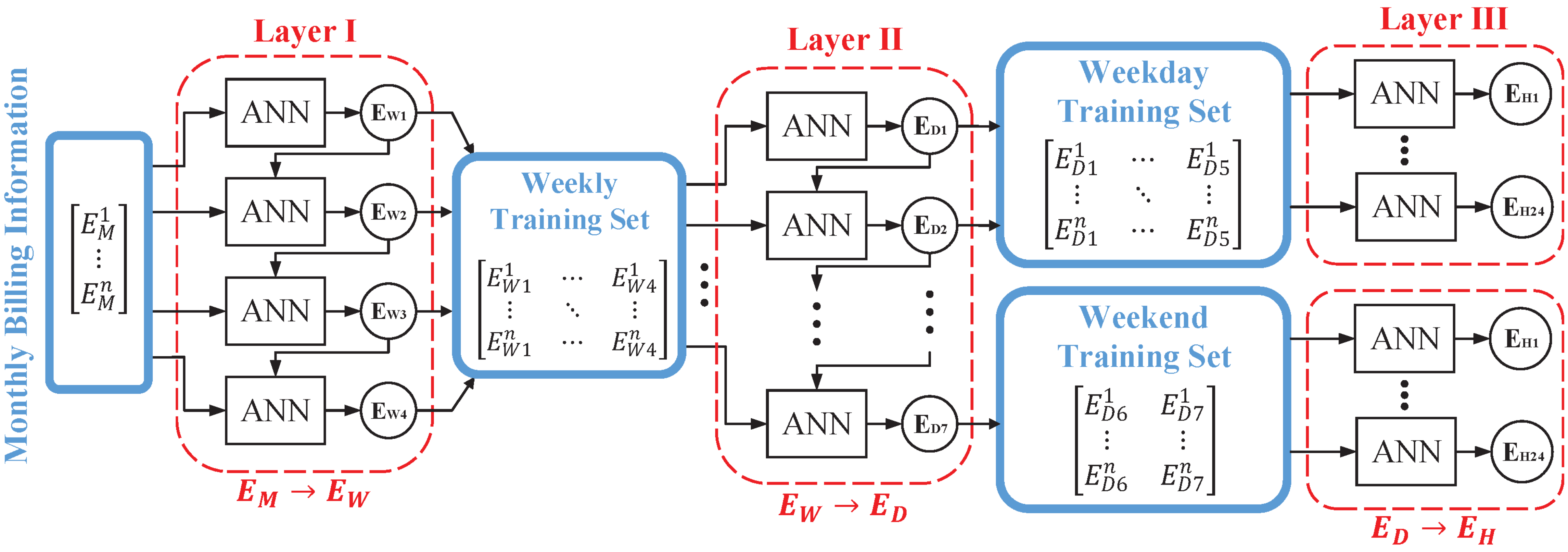
✓**Counter-Measures:**

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





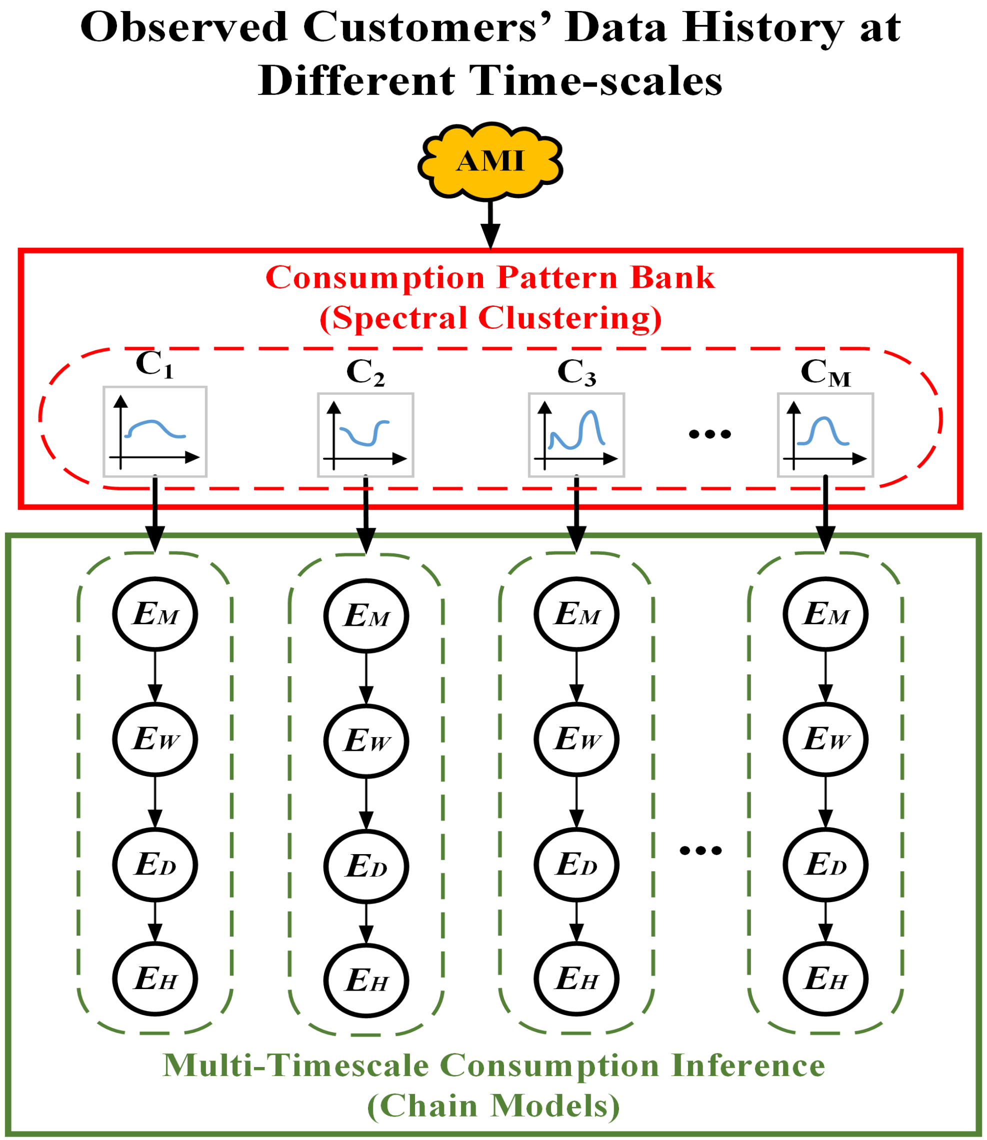
Solution Step II: Using Observed Customers' Data for Training Multi-Timescale Load Inference Chain Models



E_M – Monthly Consumption
E_W – Weakly Consumption
E_D – Daily Consumption
E_H – Hourly Consumption

✓Extends observability using data of customers with smart meters to obtain a stage-by-stage consumption transition process (Maintains High Correlation!)

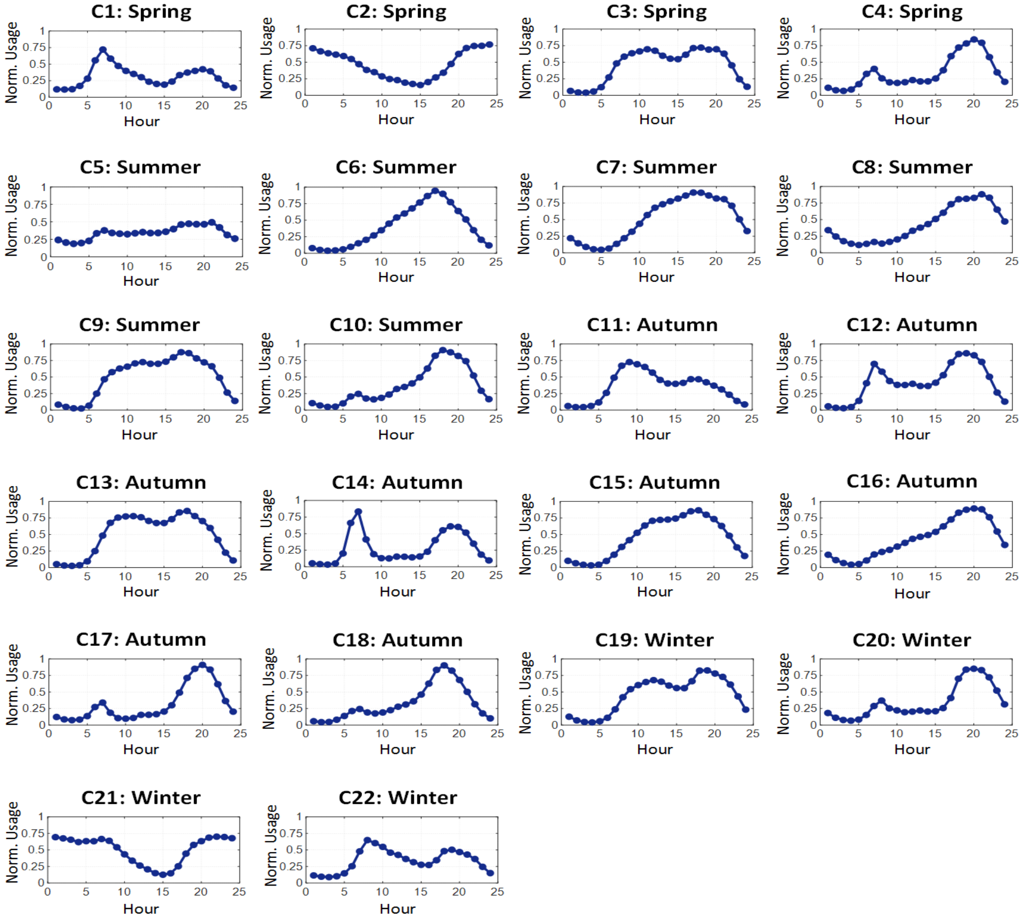
Solution Step III: Observed Customer Daily Load Pattern Bank Formation and Training Multi-Timescale Models



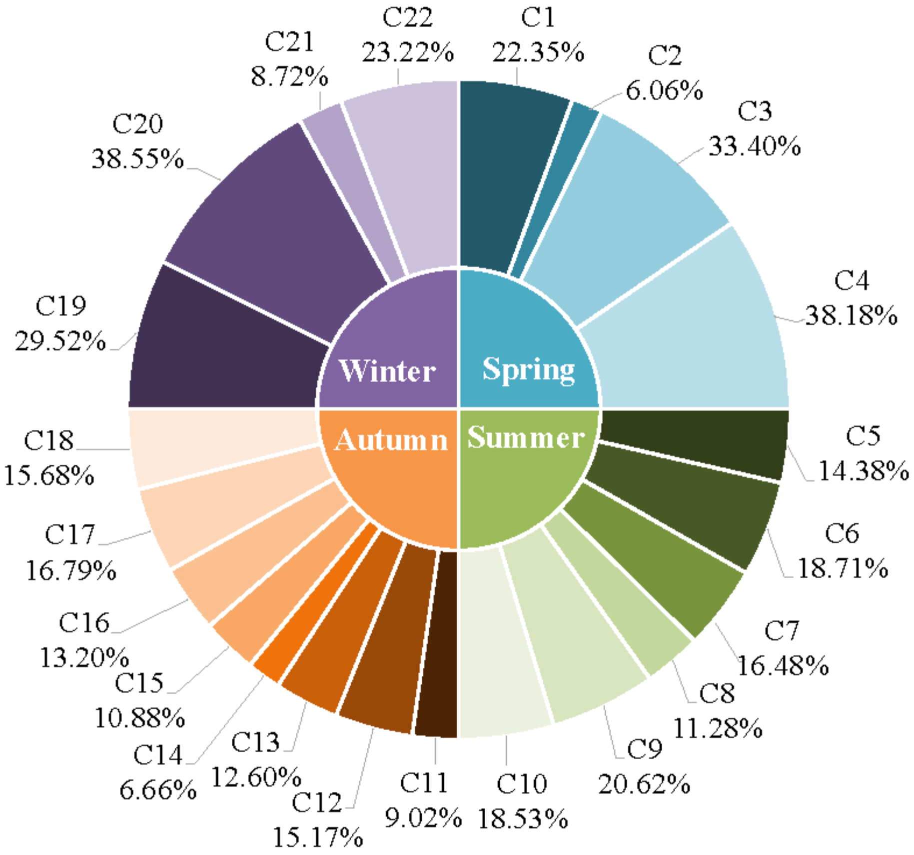
- Problem: Performance of Multi-timescale Chain Models Highly Depend on Typical Daily Consumption Patterns of Different Customers
- Solution: Assign a Multi-Timescale Model to Each Typical Load Behavior Pattern Discovered From Observed Loads (Method: Data Clustering)
- Train Load Inference Chain Models Using the Data of Observed Customers Belonging to Each Cluster (C_i)

Customer Behavior Pattern Bank: Sensitivity to Time of Day and Load Type

Typical discovered load profiles in different seasons from smart meter data

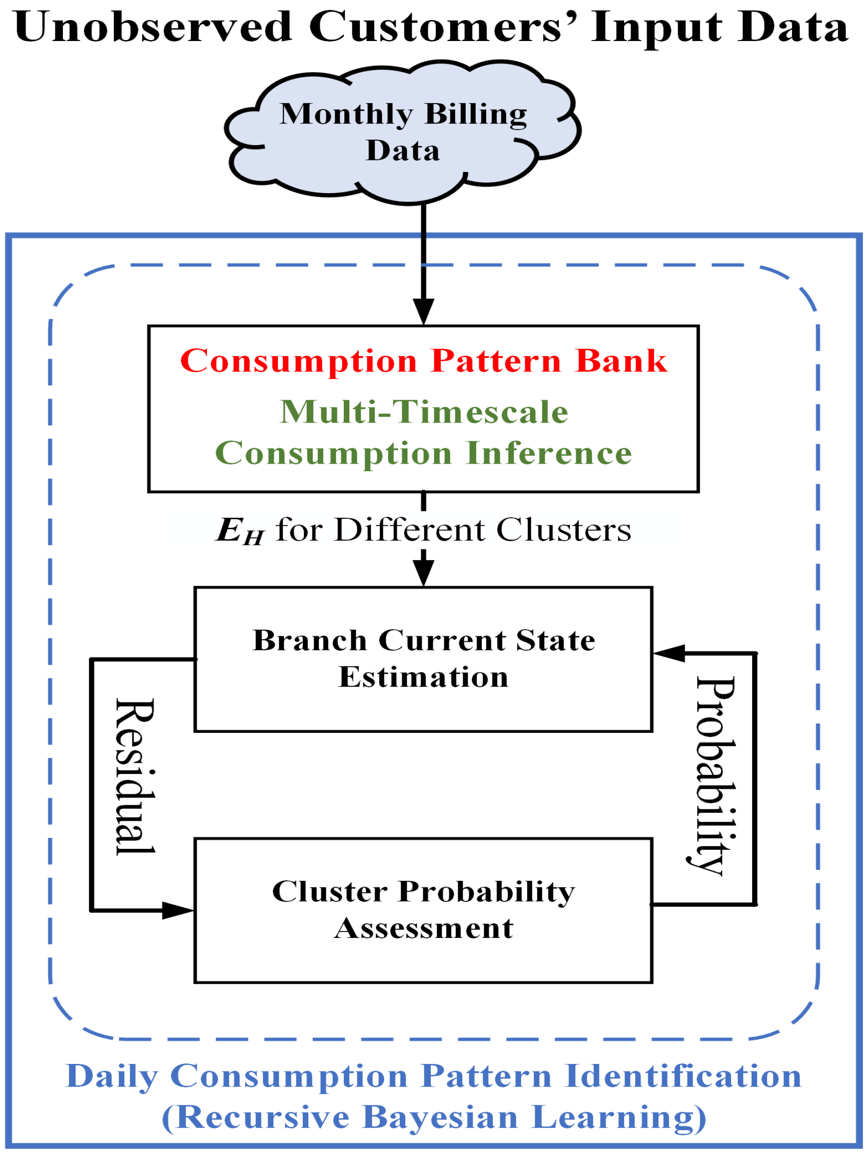


The percentage of customers belonging to each typical load profile in different seasons



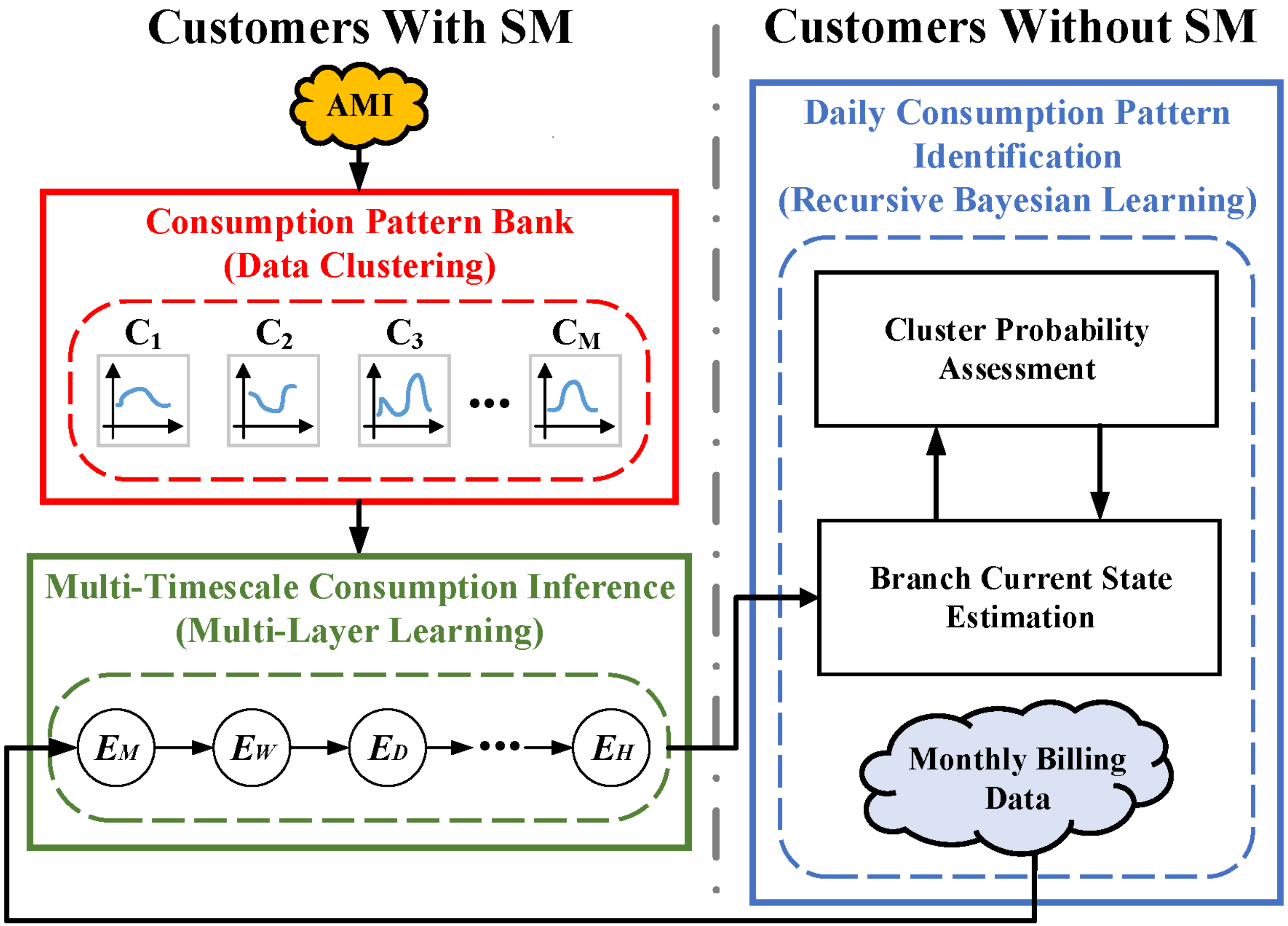
✓ Methodology: Data Clustering (Unsupervised Learning – Spectral Clustering Algorithm)

Solution Step IV: Unobserved Customers' Pattern Identification and Hourly Consumption Inference

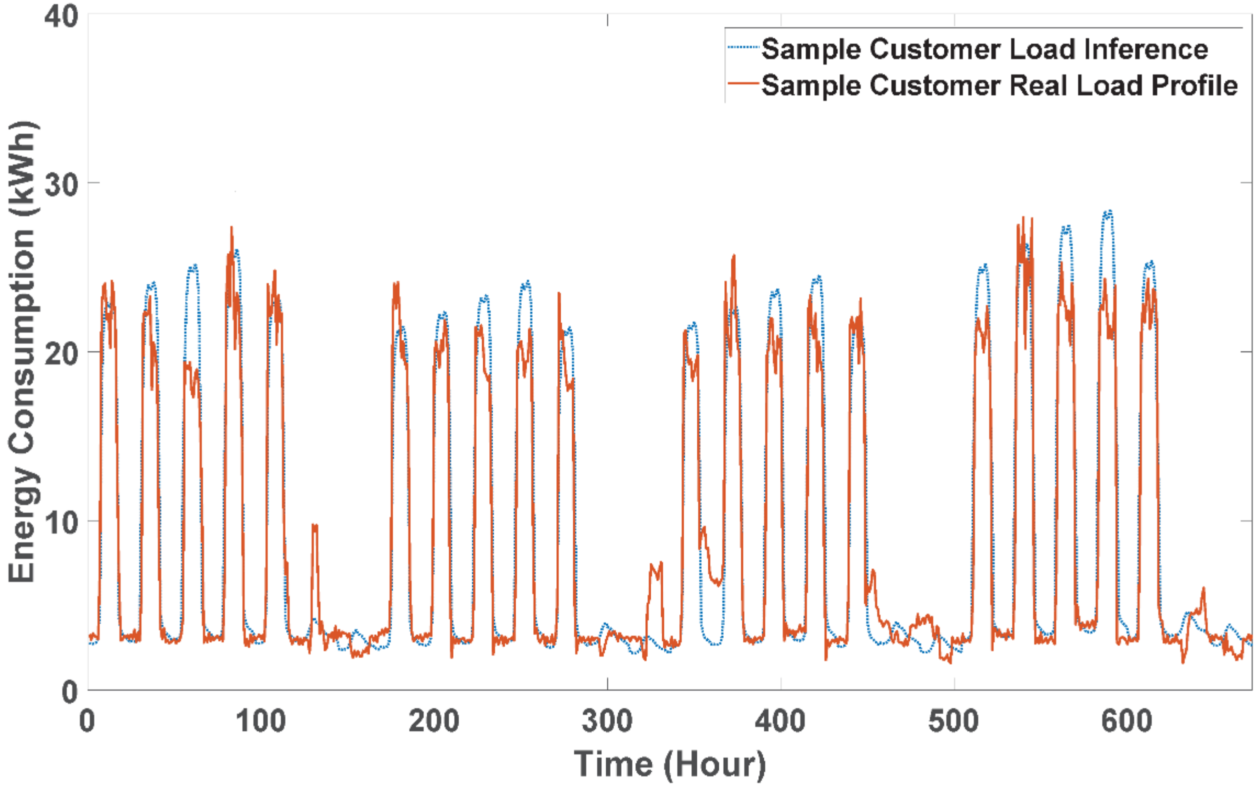


- Basic Idea: Pick the Cluster that has the Best State Estimation Performance for Each Customer
- Methodology: Assign and Update Probability Values to Different Clusters Based on State Estimation Residuals (Recursive Bayesian Learning)
- Outcome: Pick the Most Probable Cluster for Each Unobserved Customer and Use its Corresponding Chain Model for Hourly Load Inference

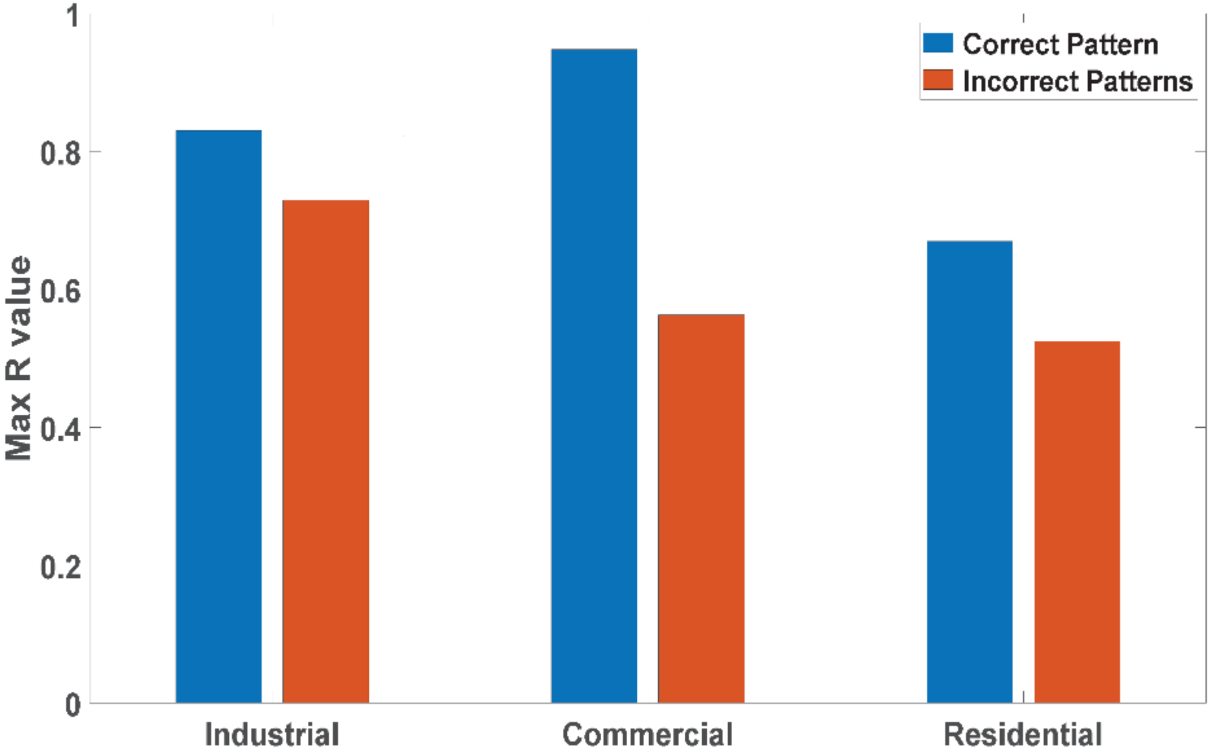
Overall Structure of the Proposed Solution



Numerical Results: Unobserved Individual Customer Hourly Load and Pattern Inference

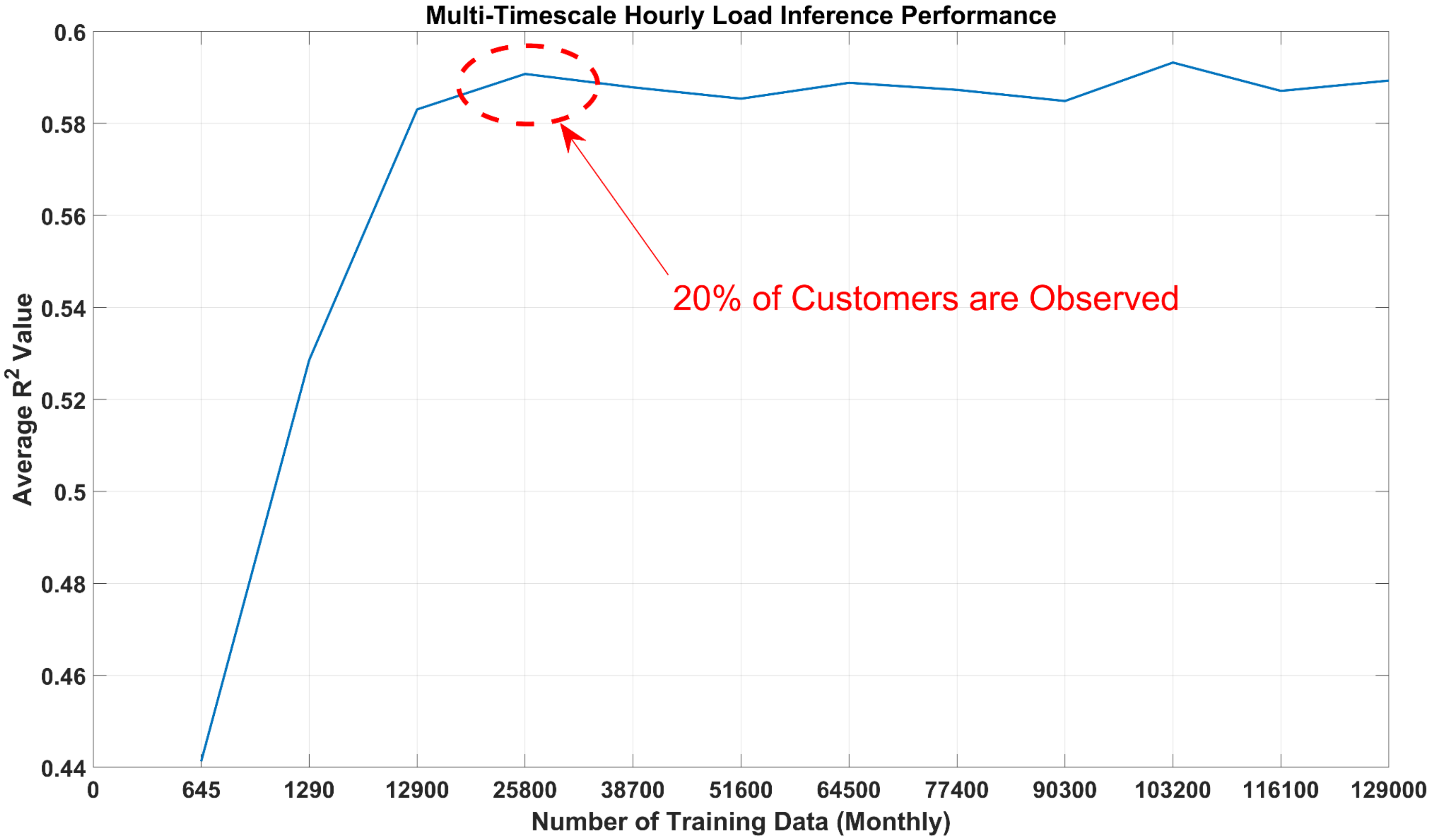


Inferring the hourly demand of an unobserved residential load in one month (average estimation error ≈ 8.5% of total energy)



Impact of accurate consumption pattern identification on the accuracy of the inference (industrial load patterns are close and stable)

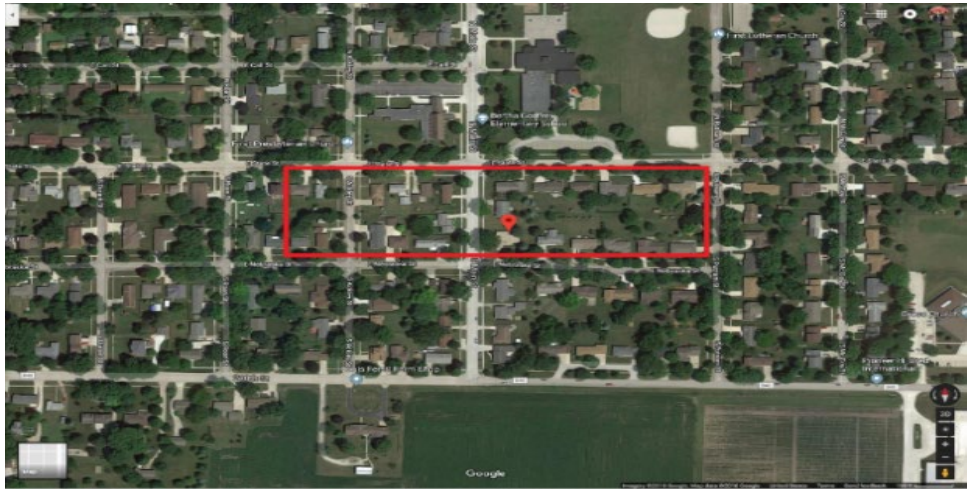
Sensitivity Analysis of Observability

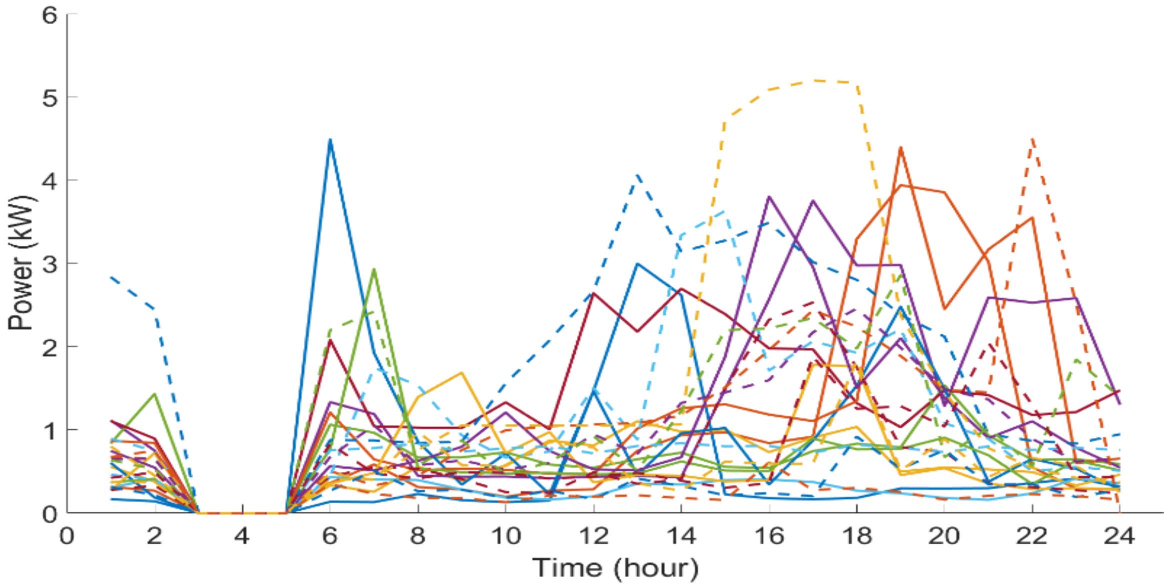


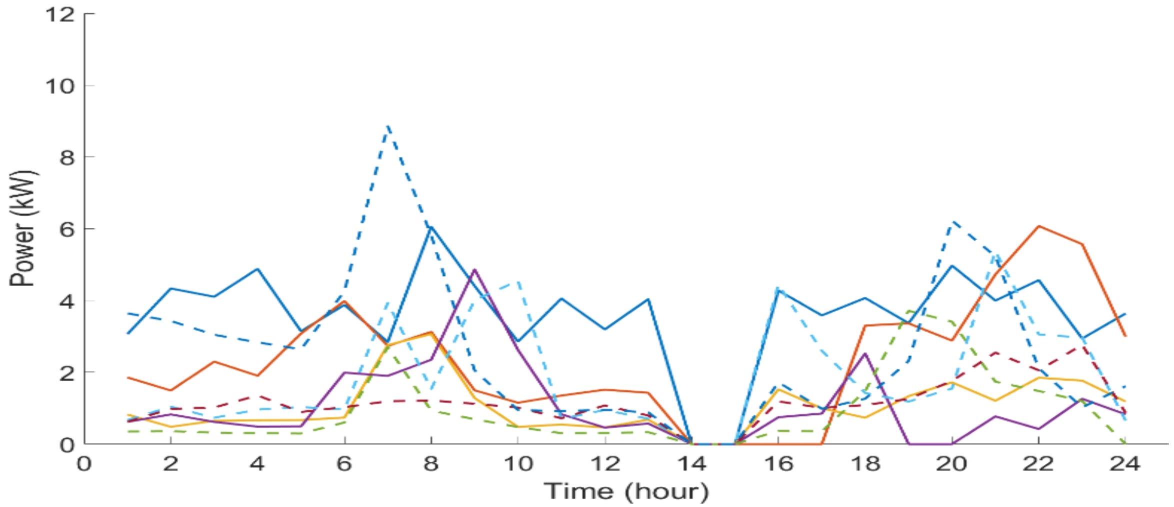
Assessing Cold Load Pick up Demands Using Smart Meter Data

- Problem Statement: Estimating post-outage cold load pick up (CLPU) demand at feeder-level and customer contribution to CLPU overshoot using smart meter data. This overshoot is critical in designing restoration plan since it may overload transformers and DERs.
- Challenges:
  - ✓Customer behavior volatility
  - ✓Lack of behind-the-meter information on customer thermostatically controlled loads
- Solution Strategy: Develop a data-driven "model-free" framework to estimate CLPU demand at both feeder-level and customer-level using only smart meter data
- Proposed Solution Components:
  - ✓Machine learning-based diversified load predictor at feeder-level
  - ✓Probabilistic reasoning at customer-level to model behavioral uncertainty

Impact of Outage on Customer Behavior

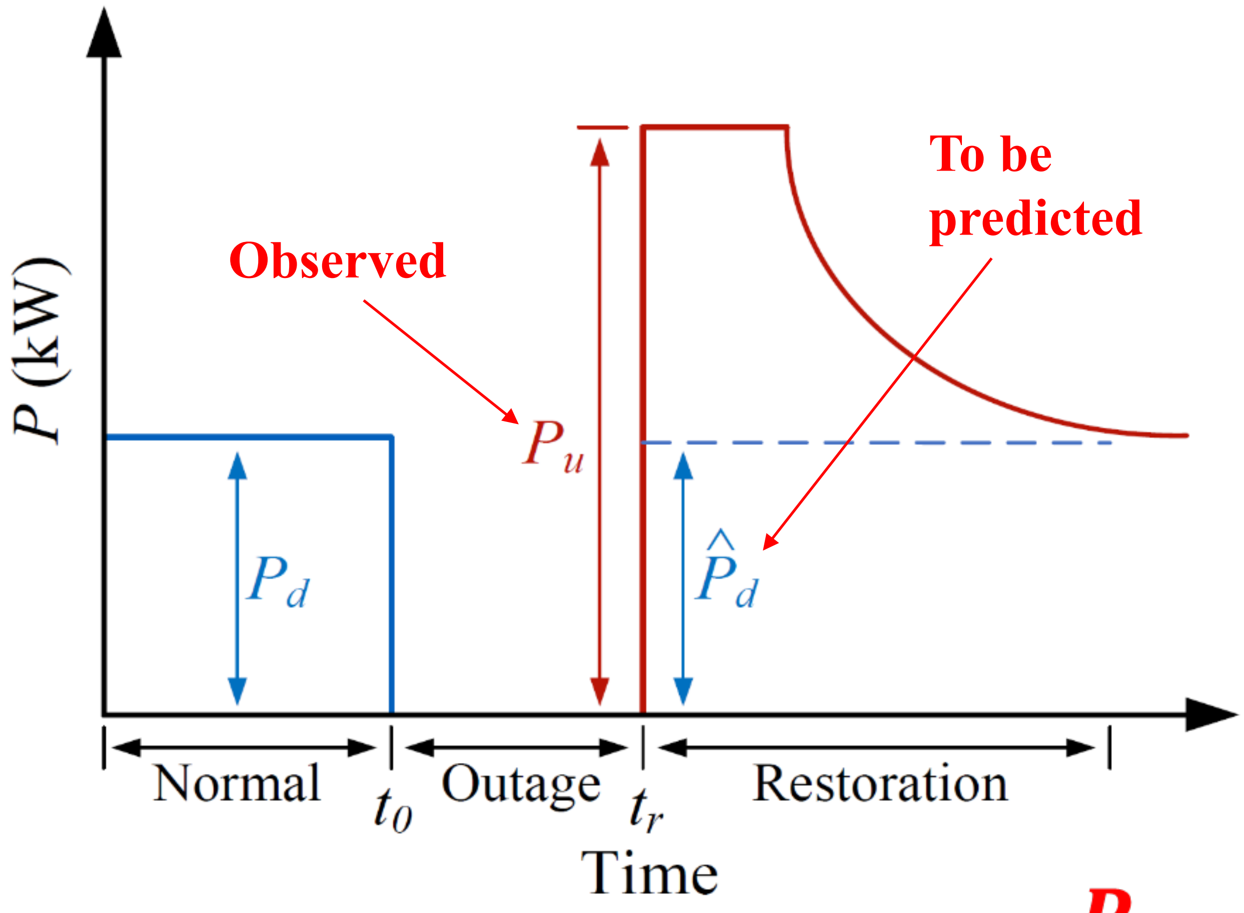




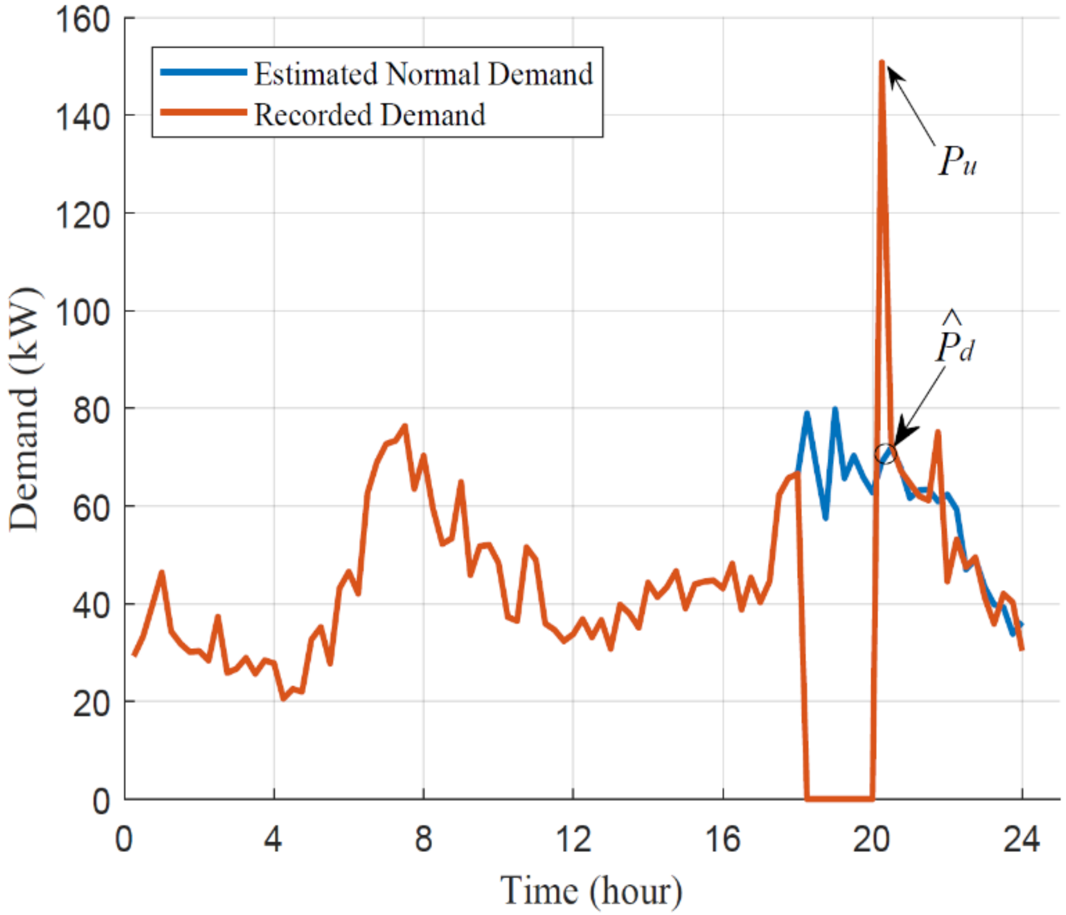


Abnormal Post-Outage Demand Increase: Cold Load Pick-up

Post-Outage Cold Load Pick-up (CLPU): Loss of Diversity



$$\text{CLPU Ratio} = \frac{P_u}{\hat{P}_d}$$



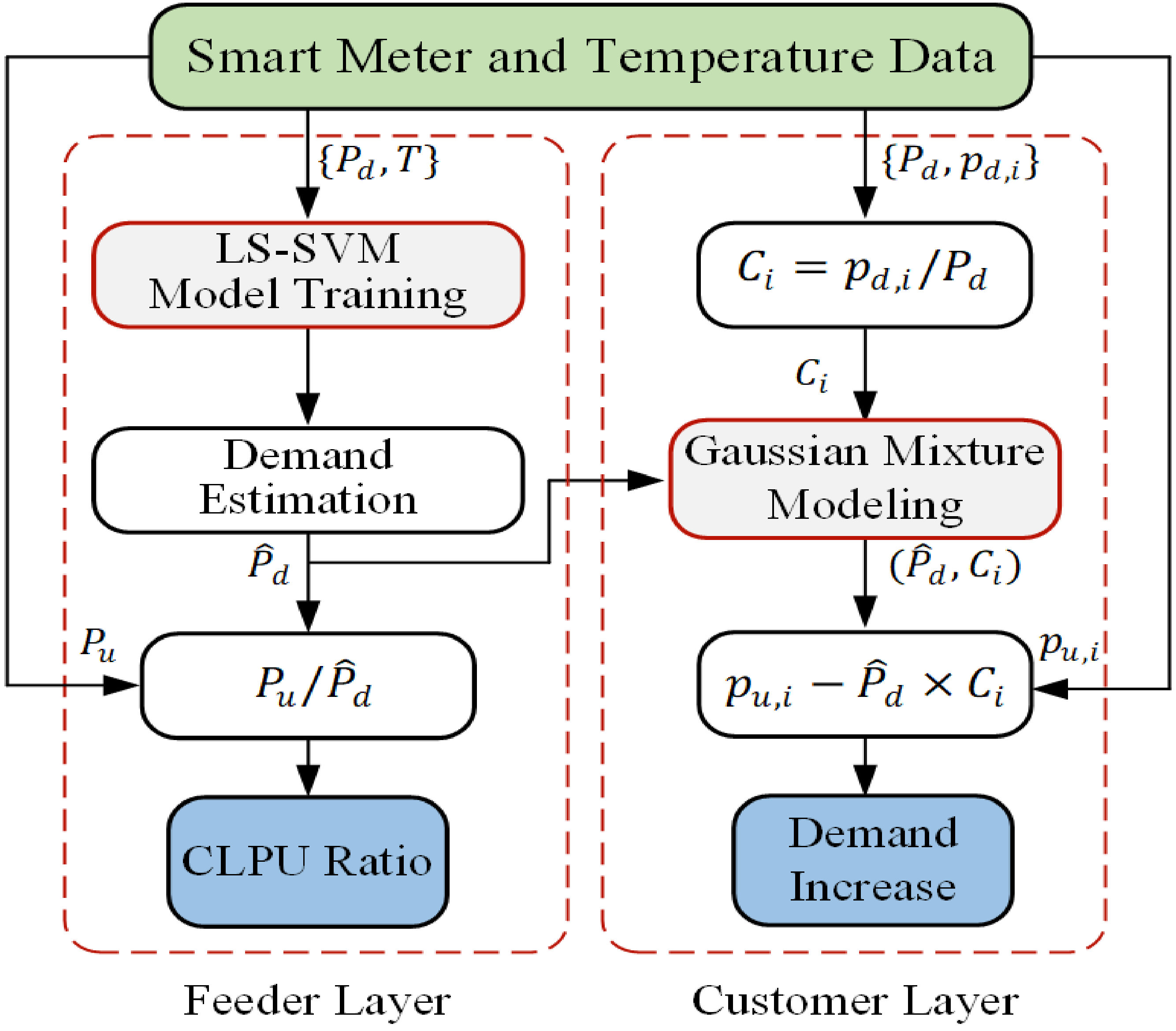
Real CLPU at the Feeder-Level

Literature Review

- Previous works have mainly focused on ***model-driven*** methods to obtain CLPU ratios [1-3]
  - ➢Use thermostatically controlled load models and thermal parameters to model houses
- ***Comments***:
  - ➢Need to collect detailed house-level thermal ***parameters***
  - ➢Need to model ***individual*** thermostatically controlled load

[1] K. P. Schneider, E. Sortomme, S. S. Venkata, M. T. Miller, and L. Ponder, "Evaluating the magnitude and duration of cold load pick-up on residential distribution using multi-state load models," IEEE Trans. Power Syst., vol. 31, no. 5, pp. 3765–3774, Sep. 2016.
[2] D. Athow and J. Law, "Development and applications of a random variable model for cold load pickup," IEEE Trans. Power Del., vol. 9, no. 3, pp. 1647–1653, Jul. 1994.
[3] E. Agneholm and J. Daalder, "Cold load pick-up of residential load," IEE Proceedings - Generation, Transmission and Distribution, vol. 147, no. 1, pp. 44–50, Jan. 2000.

Overall Structure of Data-Driven CLPU Estimation Method



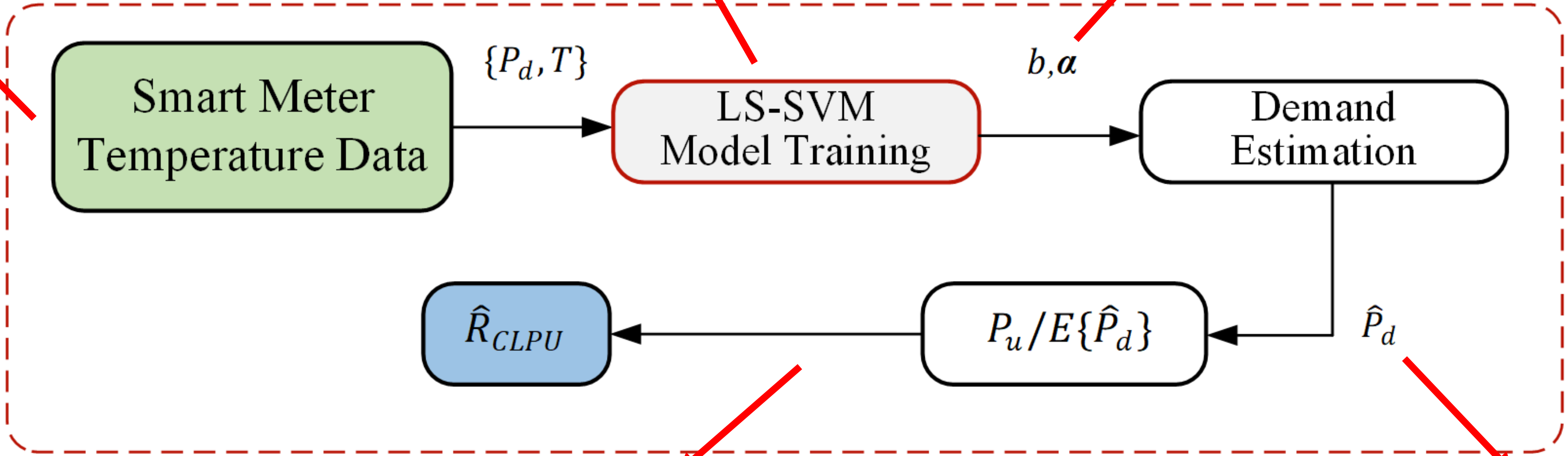
- ✓ **Characterizes CLPU at Feeder-level Using Learning-Based Demand Prediction**
- ✓ **Determine Customer Contribution to CLPU Demand Increase Using Probabilistic Reasoning (GMM)**
- ✓ **Obtain Useful Statistics at Feeder- and Customer-Level to Fully Quantify CLPU**

Feeder-Level Data-Driven CLPU Ratio Estimation

Trained to Predict Feeder Load Under Normal Operation (Least-squares support-vector machines)

Machine Learning Model Parameters

T – Ambient temperature
P_d – Normal feeder demand

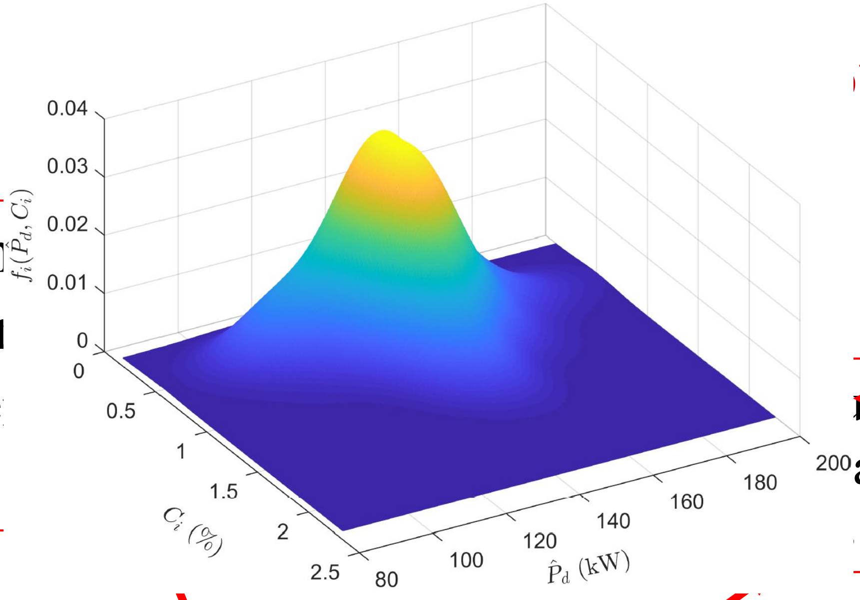


Estimates the CLPU Overshoot (R_{CLPU}) by Dividing the Observed Feeder Demand at Time of Restoration (P_u) by the Estimated "Expected" Hypothetical Predicted Normal Demand, $E\{P_d\}$

Estimated *Diversified* Demand (what would happen if there was no outage)

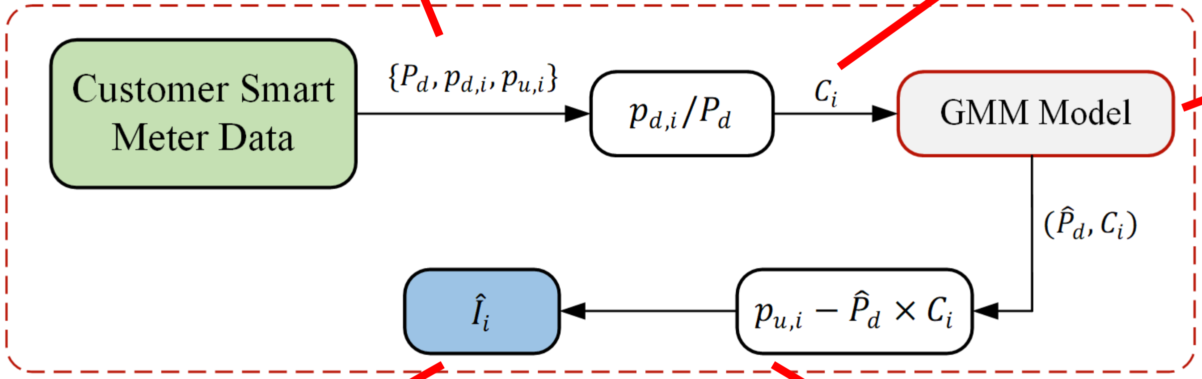
Customer-Level ... U Estimation

P_d – Normal feeder deman...
$p_{d,i}$ – Normal customer de...
$p_{u,i}$ – Post-outage custome...
time of restoration



...nate customer contribution to
...al feeder demand (C_i) at different

Given the time-variability and uncertainty of customer behavior Gaussian Mixture Modeling (GMM) has been used to model the probability distribution of C_i and P_d in normal operation



$p_{u,i}$ – Post-outage customer demand at the time of restoration
I_i – Customer contribution to CLPU demand

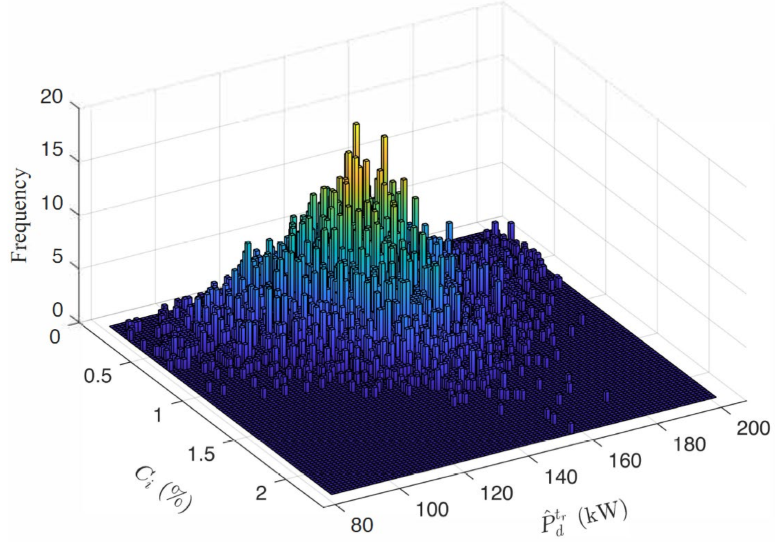
At restoration the learned GMM-based joint distribution of C_i and P_d (quantifying customer's normal behavior) is used to identify customer contribution to CLPU by estimating customer deviation from its expected normal load

Gaussian Mixture Model (GMM)

Why do we need GMM?

- The estimated feeder-level normal demand, $\hat{P}_d^{tr}$, follows a distribution due to ***regression residuals***.
- The historical customer contribution factor, C_i , also follows a distribution due to the ***uncertainty*** of customer demand. Note that historical C_i is calculated by $C_i = p_{d,i} / P_d$.
- The bivariate pair, $\{\hat{P}_d^{tr}, C_i\}$, forms a ***2-dimensional*** empirical histogram.

- This 2-dimensional histogram does not strictly fit a ***single*** distribution model. Therefore, a ***mixture*** model should be used to represent the empirical histogram.
- In our problem, we used ***Gaussian mixture models*** (GMM).

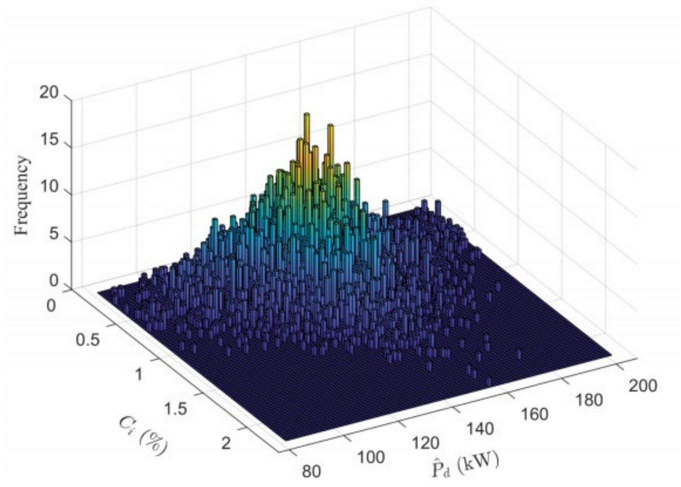


Gaussian Mixture Model (GMM)

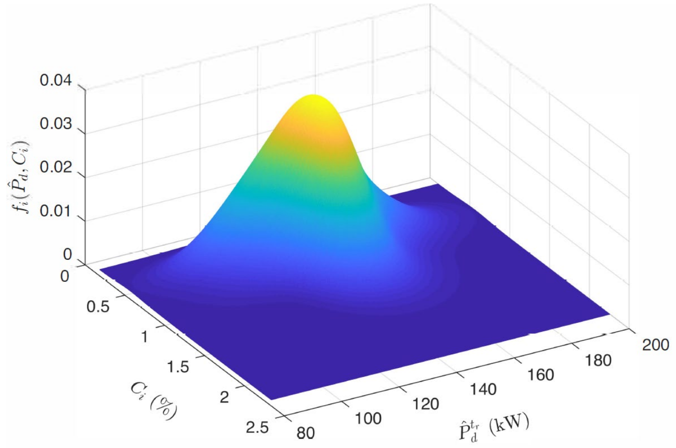
- For our problem, we approximate the joint 2-dimensional PDF of $\hat{P}_d^{tr}$ and C_i, using multiple Gaussian functions

$$f(\hat{P}_d^{t_r}, C_i) = \sum_{j=1}^{S_i} \omega_j g_j(\hat{P}_d^{t_r}, C_i)$$

where, $g_j(\cdot)$ denotes a bi-variate Gaussian function, w_j is the weight corresponding to each $g_j(\cdot)$, S_i is the total number of Gaussian functions. Note that w_j and the parameters in $g_j(\cdot)$ are determined by the maximum likelihood (ML) estimation, using the empirical histogram.



Empirical histogram

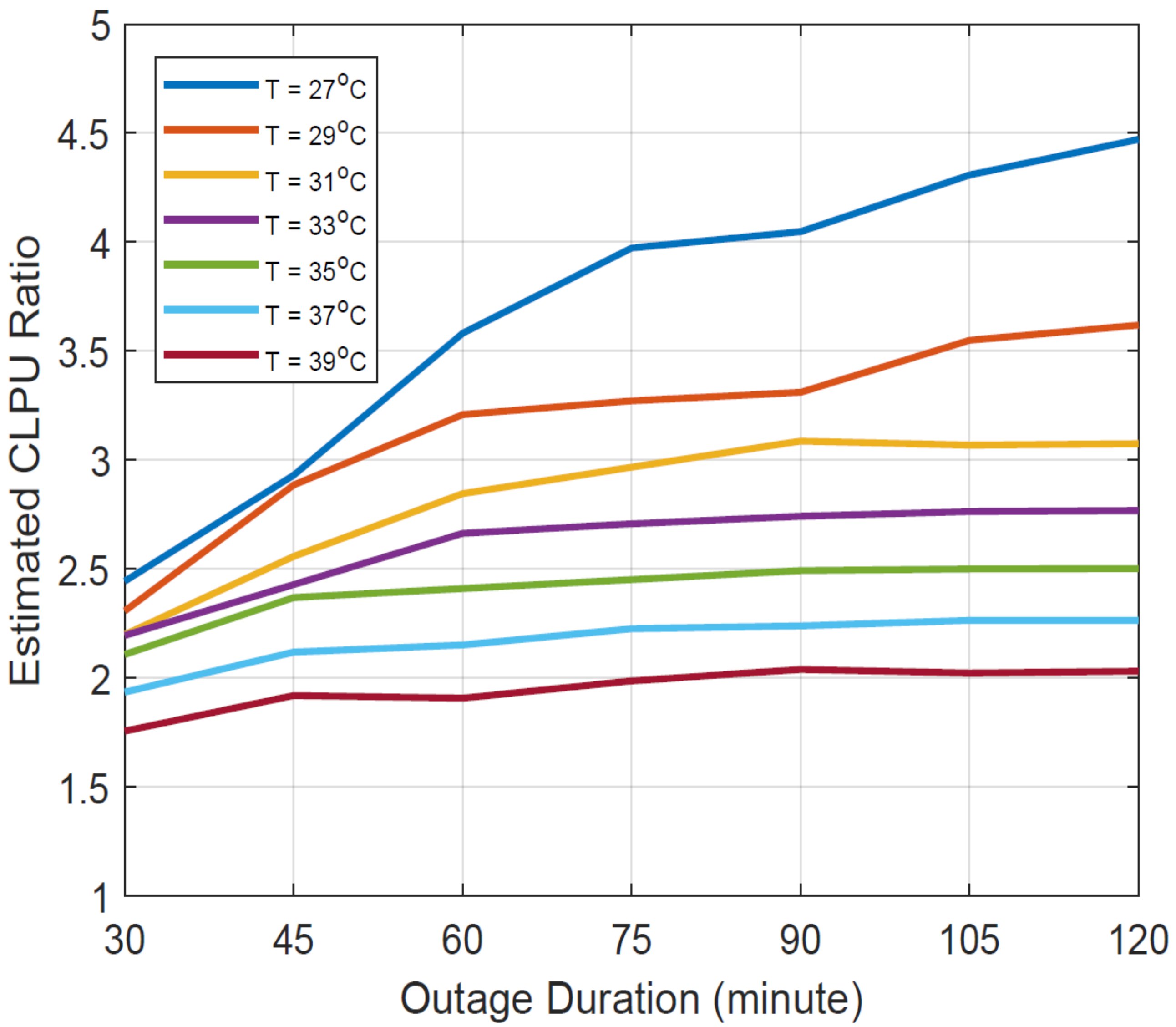


GMM-based estimation

$\hat{P}_d^{t_r}$ -- Estimated feeder-level normal demand at t_r

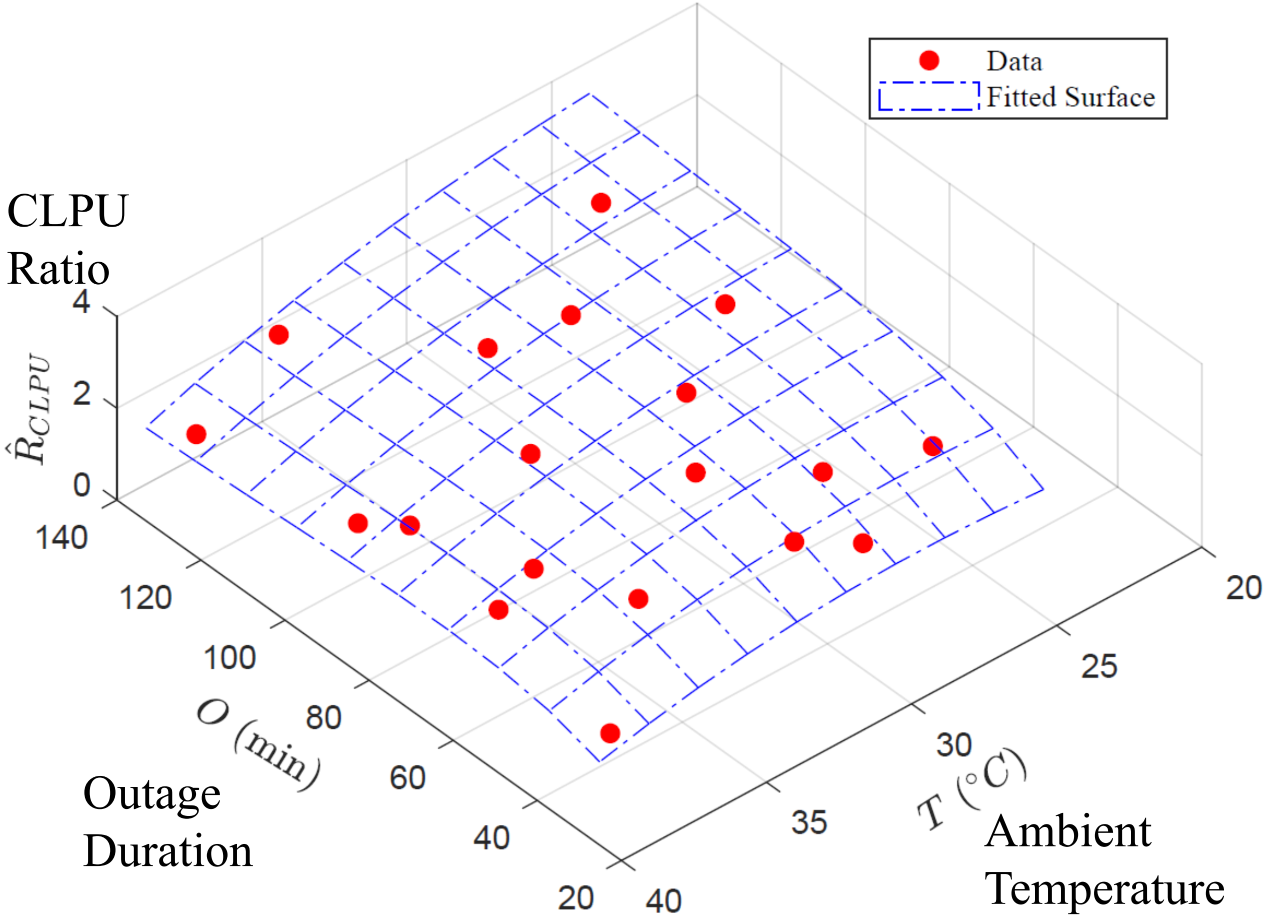
C_i -- Historical customer contribution factor

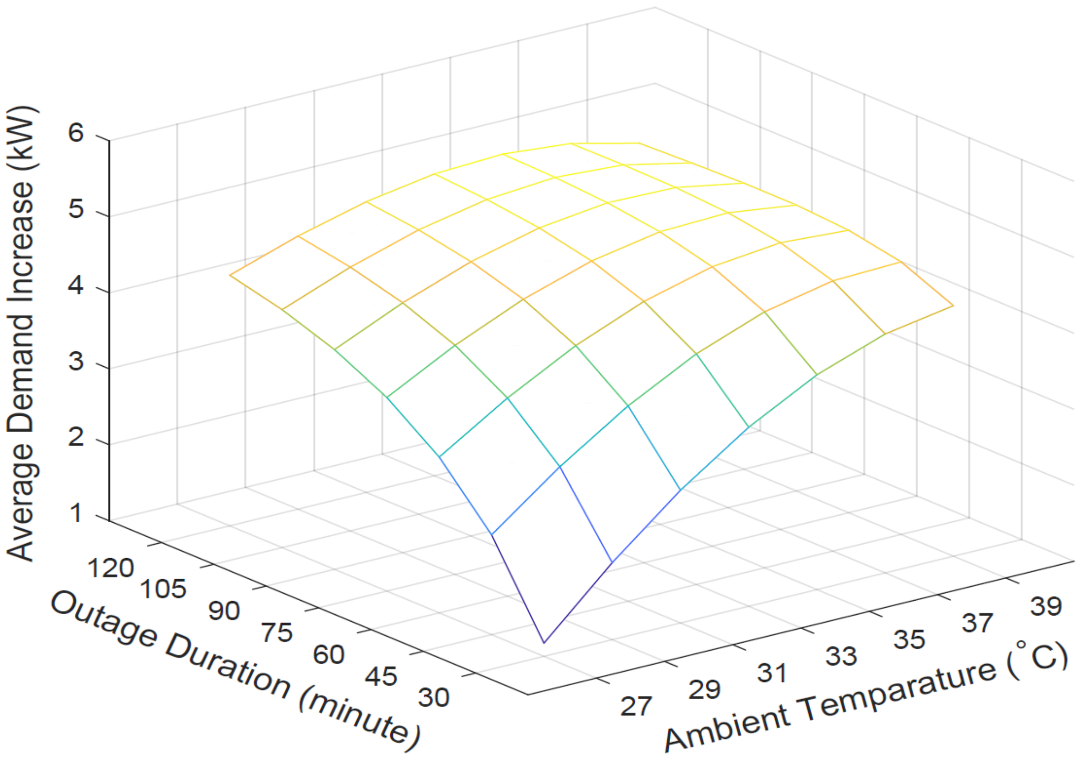
Feeder-Level CLPU Characteristics



- ✓ CLPU ratio increases and saturates with outage duration
- ✓ CLPU ratio is sensitive to ambient temperature

CLPU Characteristics: Feeder- and Customer-Level





Feeder-Level CLPU ratio characterization through regression as a function of outage duration and ambient temperature in summer

Expected customer contribution to CPLU demand as a function of outage duration and ambient temperature in summer

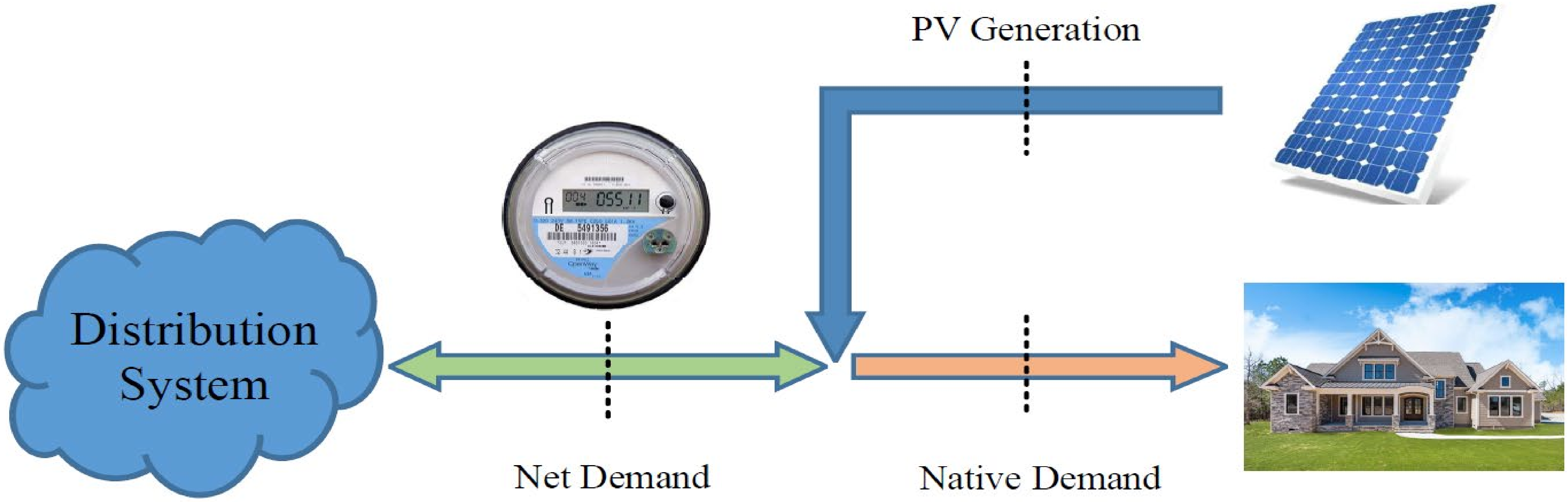
Conclusions

- We have archived a large amount of real data from utility partners, including smart meters, SCADA, PMUs and circuit models.
- We have shared one real distribution grid model with one-year smart meter data.
- A Data-Driven Load Inference method is developed for Monitoring Distribution Systems:
  - ✓ Identifying Temporal Correlations for Load Estimation
  - ✓ Exploiting Latent Trends in Customer Behavior at Different Time-Scales for Enhancing Inference Accuracy
- We have used smart meter data to model the cold load pick up, which would be useful to utilities in designing restoration plan.

Recent Publications in Data Analytics

- 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*, accepted for publication.
- 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*, accepted for publication.
- 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.
- H. Sun, Z. Wang, J.Wang, Z.Huang, N. Carrington, and J. Liao, "Data-Driven Power Outage Detection by Social Sensors," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2516-2524, September 2016.
- L. Fang, K. Ma, R. Li, and Z. Wang, "A Statistical Approach to Estimate Imbalance-Induced Energy Losses for Data-Scarce Low Voltage Networks," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2825-2835, July 2019.
- 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*, accepted for publication.
- C. Wang, Z. Wang, J. Wang, and D. Zhao, "Robust Time-Varying Parameter Identification for Composite Load Modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 967-979, January 2019.
- C. Wang, Z. Wang, J. Wang, and D. Zhao, "SVM-Based Parameter Identification for Composite ZIP and Electronic Load Modeling," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 182-193, January 2019.
- T. Lu, Z. Wang, J. Wang, Q. Ai, and C. Wang, "A Data-Driven Stackelberg Market Strategy for Demand Response-Enabled Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2345-2357, May 2019.

BTM PV Disaggregation -- Background



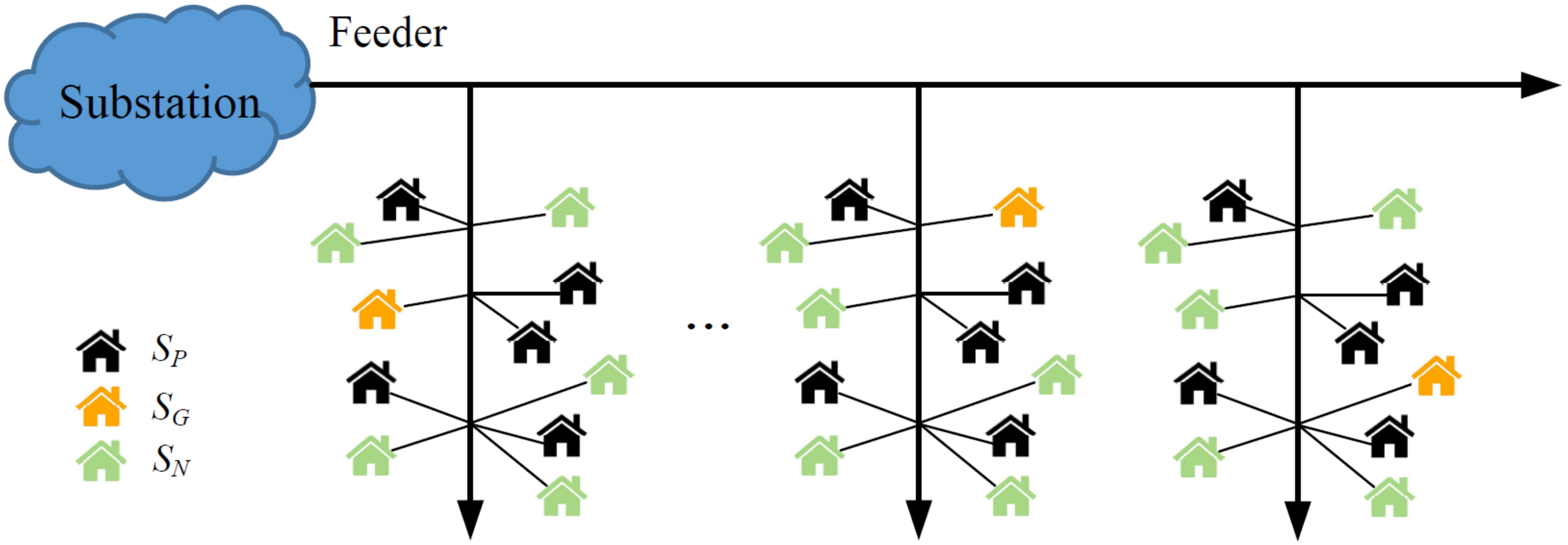
$$\text{Net Demand} = \text{Native Demand} - \text{PV Generation}$$

- In most cases, utilities only measures net demand.
- PV generation and native demand are usually invisible to utilities.
- Posing challenges in load forecasting, outage load pickup, grid expansion planning and grid control.

Problem Statement

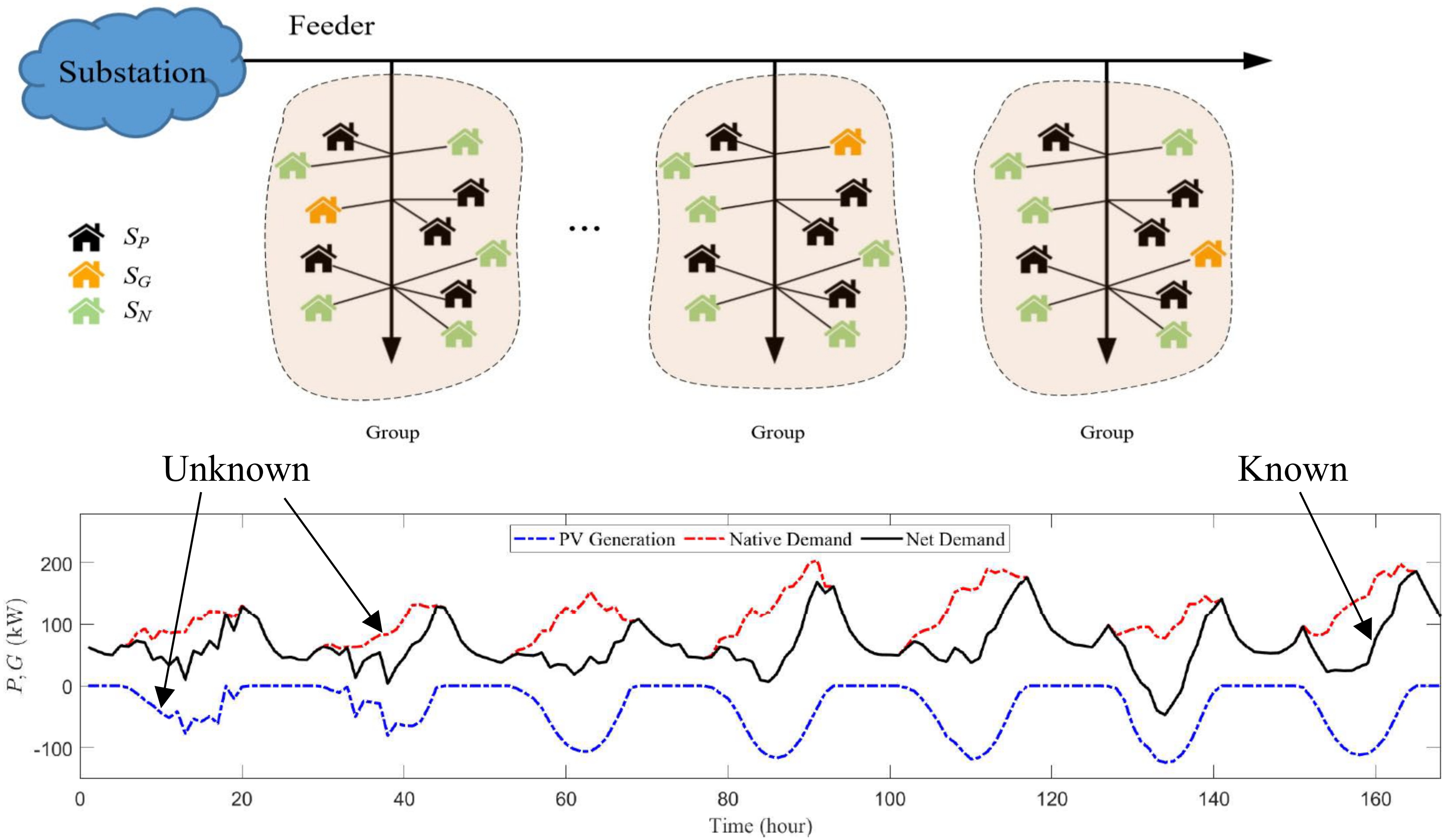
- Types of customers

S_P: Customers without PVs, whose *native demand* is recorded.
S_G: Fully observable customers with PVs, whose *native demand* and *PV generation* are recorded.
S_N: Customers with PVs, whose *net demand* is recorded.



Problem Statement

- Problem formulation: Separate aggregate BTM PV generation of groups of customers



State-of-the-Art and Challenges

	Approaches	Comments
Model-based methods [1-3]	PV performance models and weather information are used to estimate solar generation	Require detailed PV array parameters Unadaptable to changing conditions
Non-Intrusive Load Monitoring (NILM) methods [4]	Decomposing solar generation and demands of different appliances	Require high-resolution data (1-second)
Our approach	Leveraging low-resolution but widely-available smart meter data	No prior knowledge of PV array models and parameters. Adaptive to changing conditions such as PV disconnection and new installation.

[1] D. L. King, W. E. Boyson, and J. A. Kratochvil, Photovoltaic Array Performance Model. Albuquerque, NM: Sandia National Labs., 2004

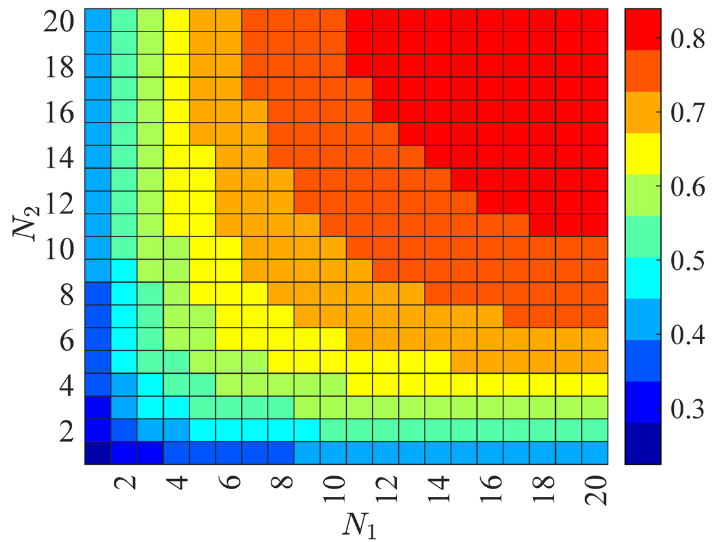
[2] Q. Zhang, J. Zhang, and C. Guo, "Photovoltaic plant metering monitoring model and its calibration and parameter assessment," in Proc. IEEE PES General Meeting, pp. 1–7, Jul. 2012.

[3] D. Chen and D. Irwin, "Sundance: Black-box behind-the-meter solar disaggregation," in e-Energy, pp. 16–19, May. 2017.

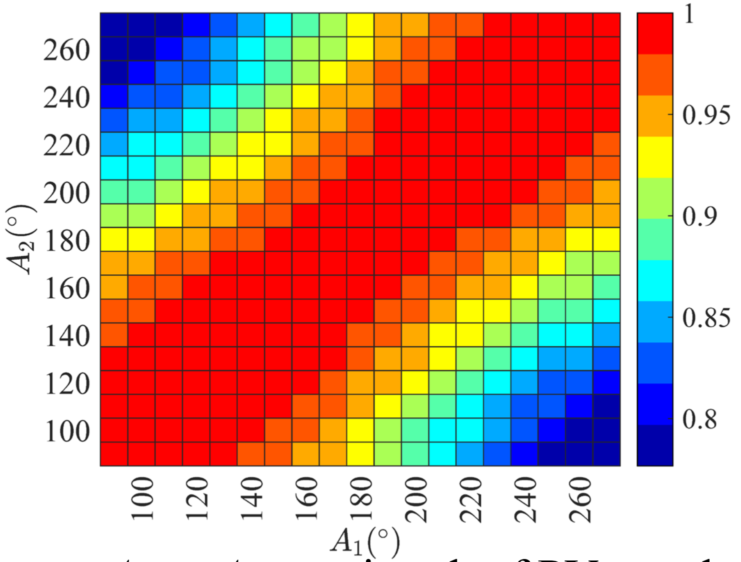
[4] C. Dinesh, S. Welikala, Y. Liyanage, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Non-intrusive load monitoring under residential solar power influx," Appl. Energy, vol. 205, pp. 1068–1080, Aug. 2017.

Observations

- Observed correlations from real smart meter data



N_1 or N_2 – size of customer group



A_1 or A_2 – azimuth of PV panel

The correlation between

- native demands of two sizable groups of customers → *high*
- generations of two PVs with similar orientation → *high*
- native demand and PV generation → *small*

For a group of customers with BTM solar

- The number of customers with a particular demand pattern is unknown
- The number of PVs with a particular orientation is unknown

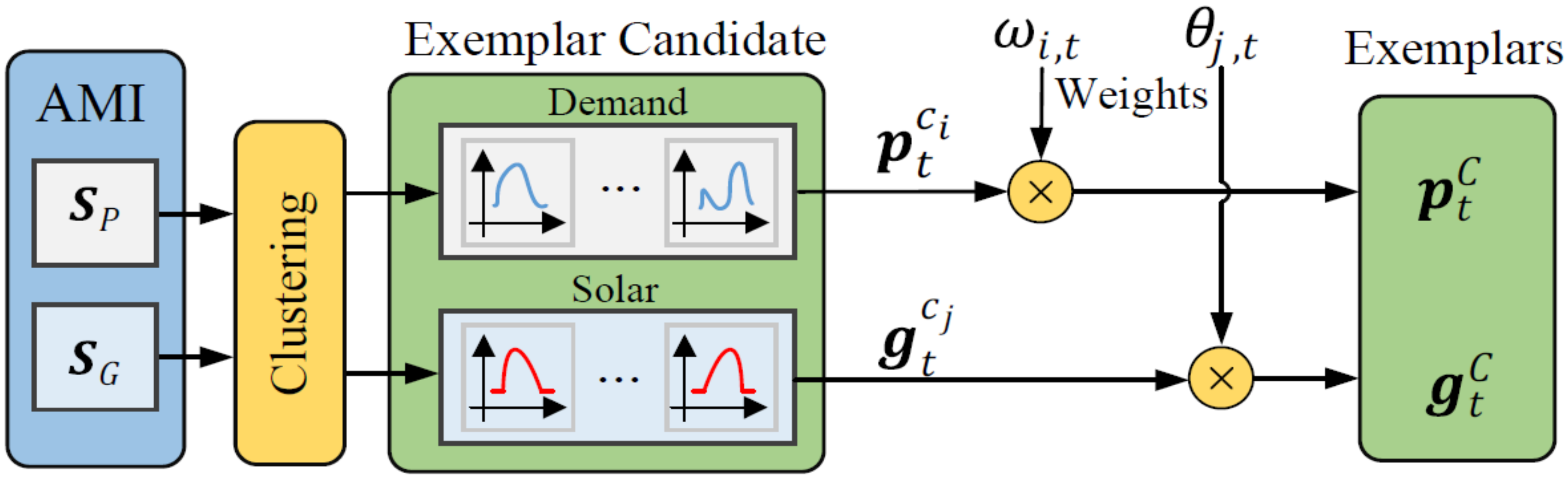
Solution

- Step I: Building native demand and PV generation exemplars
  - How to represent the exemplar of unknown native demand, using known typical demand patterns
  - How to represent the exemplar of unknown PV generation, using known typical solar generation patterns

$$\boldsymbol{p}_t^C = \sum_{i=1}^{M} \boldsymbol{p}_t^{c_i} \omega_{i,t} \qquad \boldsymbol{g}_t^C = \sum_{j=1}^{N} \boldsymbol{g}_t^{c_j} \theta_{j,t}$$

$\boldsymbol{p}_t^{c_i}$ – candidate native demand exemplars corresponding to typical load patterns

$\boldsymbol{g}_t^{c_j}$ – candidate PV generation exemplars corresponding to typical orientations



Solution

- Step II: Disaggregating PV generation

The disaggregation is formulated as finding coefficients to rescale exemplars, by minimizing disaggregation residuals.

Native demand exemplar

PV generation exemplar

$$\min_{\boldsymbol{p}_t, \boldsymbol{g}_t, \alpha_t, \beta_t} \frac{1}{2}(||\boldsymbol{p}_t - \boldsymbol{p}_t^C \alpha_t||_2^2 + ||\boldsymbol{g}_t - \boldsymbol{g}_t^C \beta_t||_2^2)$$

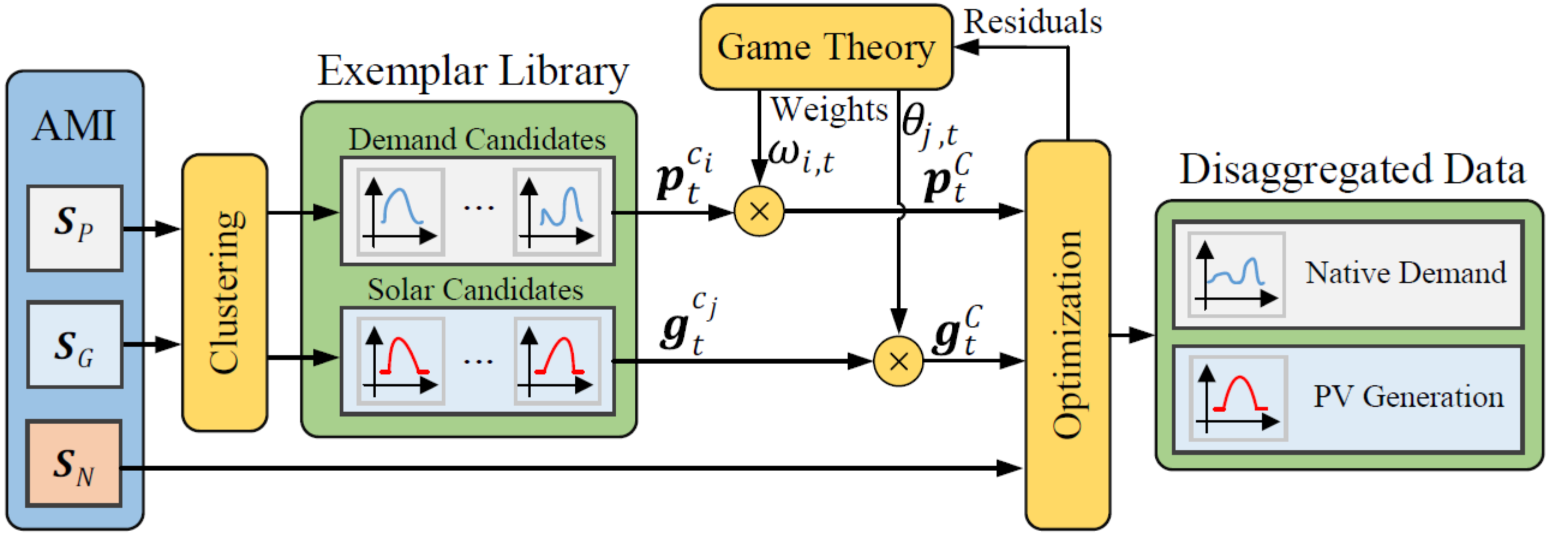
$$s.t. \quad \boldsymbol{p}_t + \boldsymbol{g}_t = \boldsymbol{p}_t^n$$

Recorded net demand

Unknown native demand

Unknown PV generation

For a group of customers, we do not know the number of customers with a particular load pattern. Therefore, the weights, $\omega_{i,t}$ and $\theta_{j,t}$, should be iteratively updated based on the disaggregation residuals.



Solution

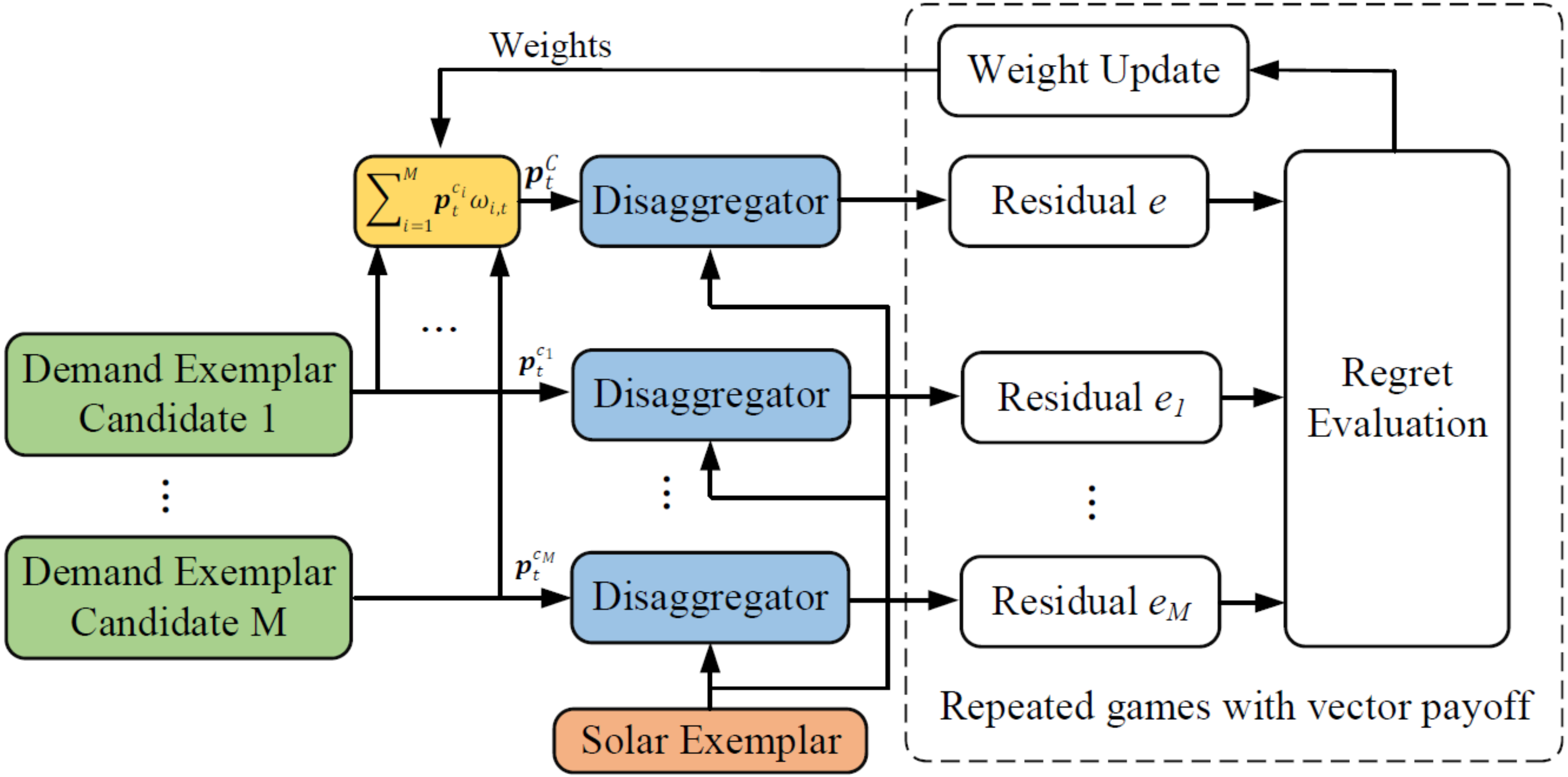
- Step III: Updating weights

Regret:

$$r_i = e - e_i$$

Weight:

$$\omega_i = \frac{e^{\lambda r_i}}{\sum_{j=1}^{M} e^{\lambda r_j}}$$



* The PV generation candidates have a similar weights updating mechanism.

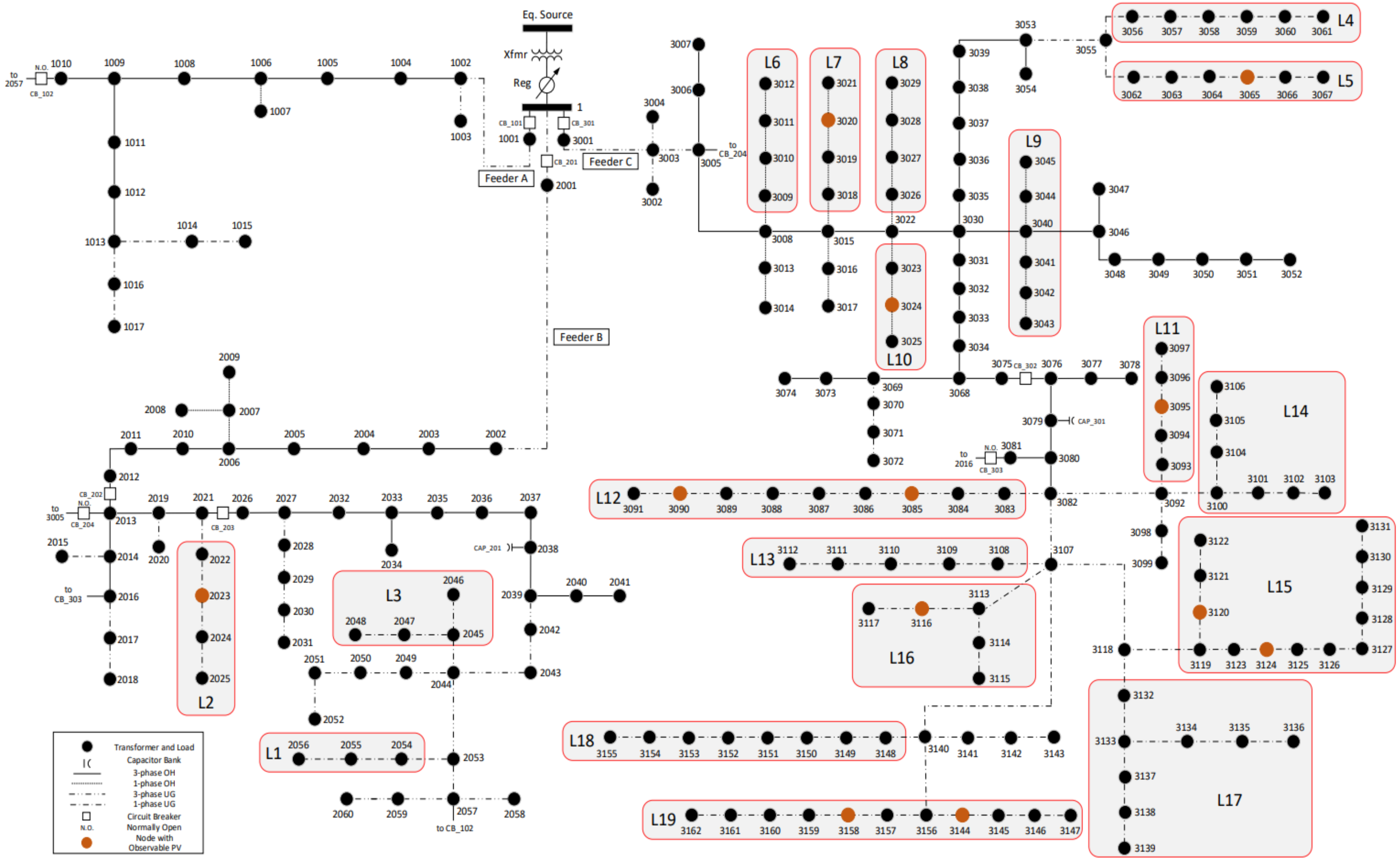
Case Study

- Disaggregate PV generation on 19 laterals of a real distribution system in the Midwest U.S.
  - Time resolution: hourly
  - Customer number: 1120
  - PV number: 337
  - Percentage of observable PV: 5%

-- Laterals with residential customers

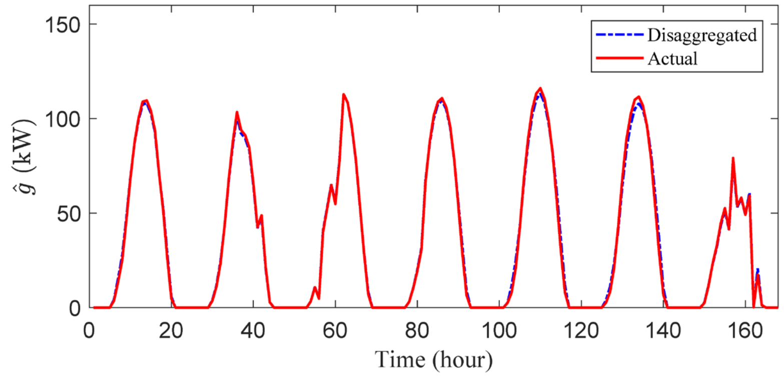
-- Transformers with observable PVs

-- Transformers with BTM solar or no solar

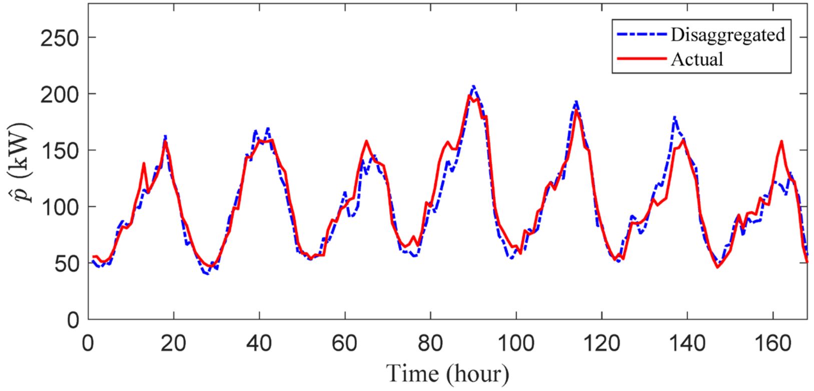


Case Study

- Disaggregated PV generation and native demand profiles

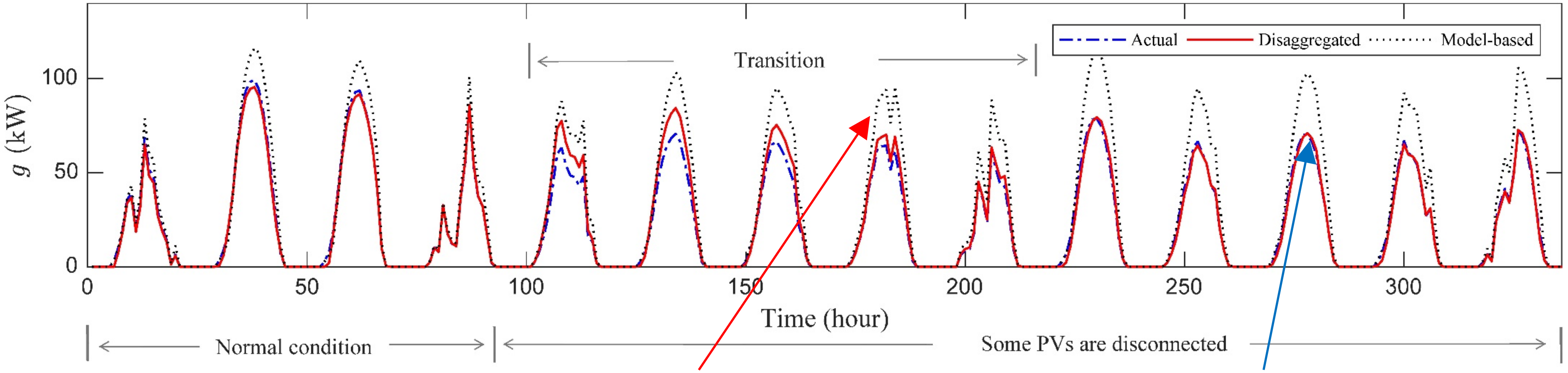


(a) PV generation



(b) Native demand

- 20% PVs are suddenly disconnected



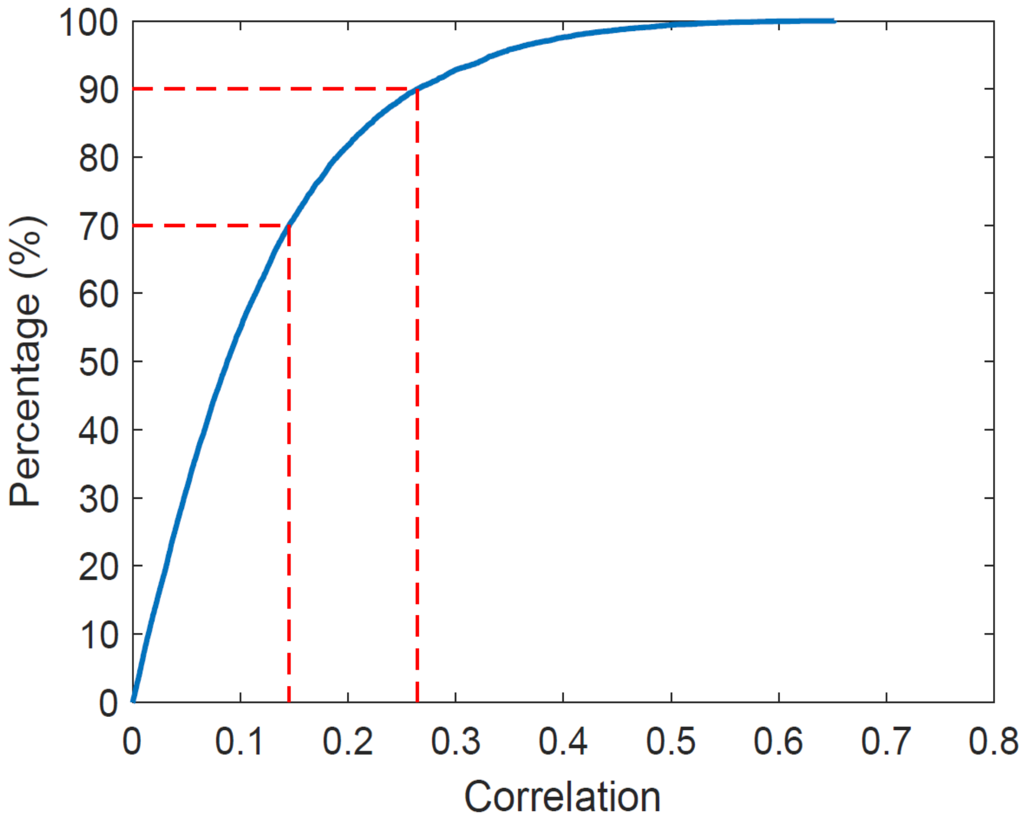
Since we do not know which PVs are disconnected, model-based method suffers from overestimation.

Our data-driven approach is adaptive.

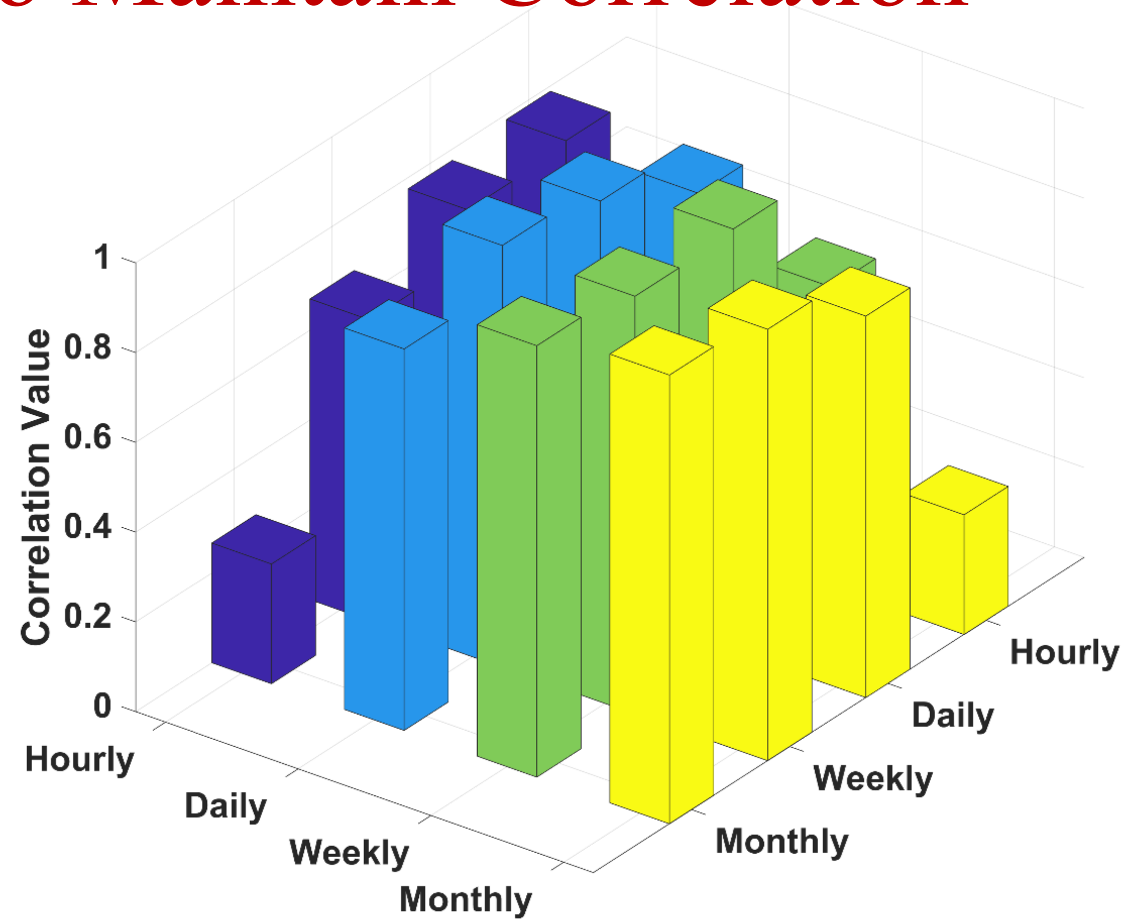
Multi-timescale Load Inference

- **Problem Statement**: Inferring hourly consumption data from customer monthly billing information as pseudo-measurements in partially observable systems
- **Challenges:**
  - ✓Loss of correlation between consumption time-series at different time-scales
  - ✓Unobserved customers' unknown typical behaviors
- **Solution Strategy**: Extending observability from observed customers to unobserved customers
- **Proposed Solution:**
  - ✓Multi-timescale load inference (stage by stage inference chain)
  - ✓Using data clustering for capturing customer typical behaviors
  - ✓Using state-estimation-based Bayesian learning for inferring unobserved customers' typical behaviors

Evidence from Data: How to Maintain Correlation



Very Small Correlation Between Different Customers' Smart Meter Time-Series: 90% below 0.27 (Loss of Correlation Across Customers)



Average Correlation between Consumption of All Customers Decreases from Monthly to Hourly (Loss of Correlation Across Different Time-Scales)

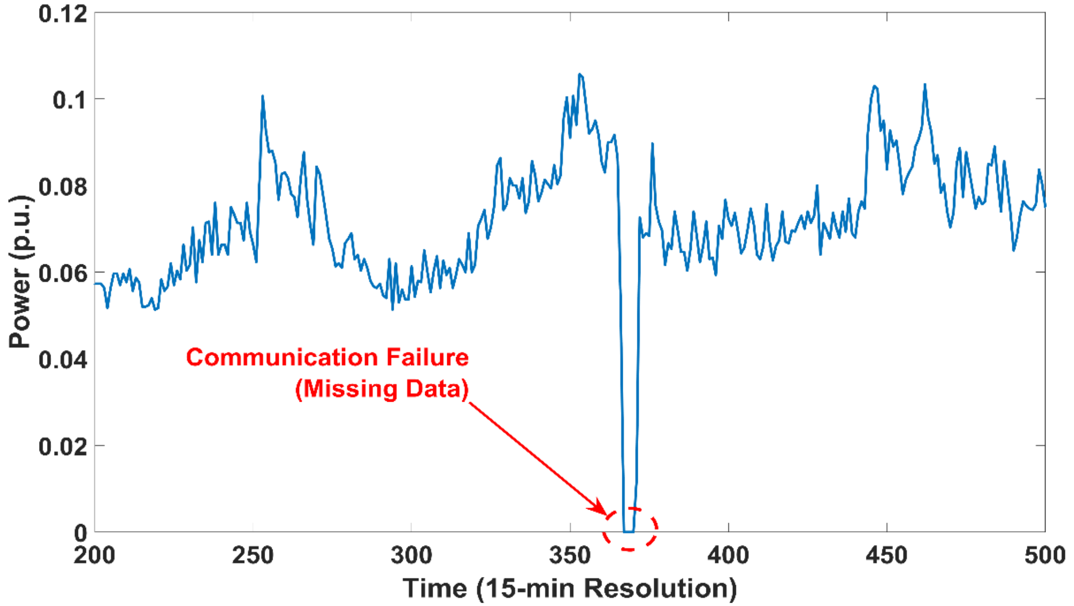
Solution Step I: Smart Meter Data Pre-Processing

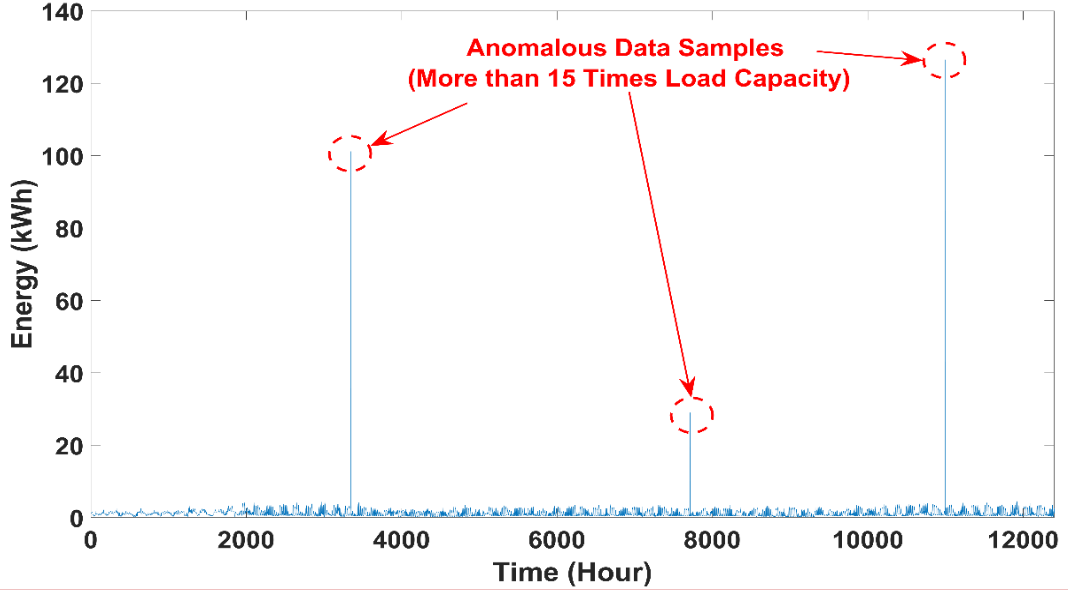
✓**Smart Meter Data Problems:**

- Outliers/Bad Data
- Communication Failure
- Missing Data

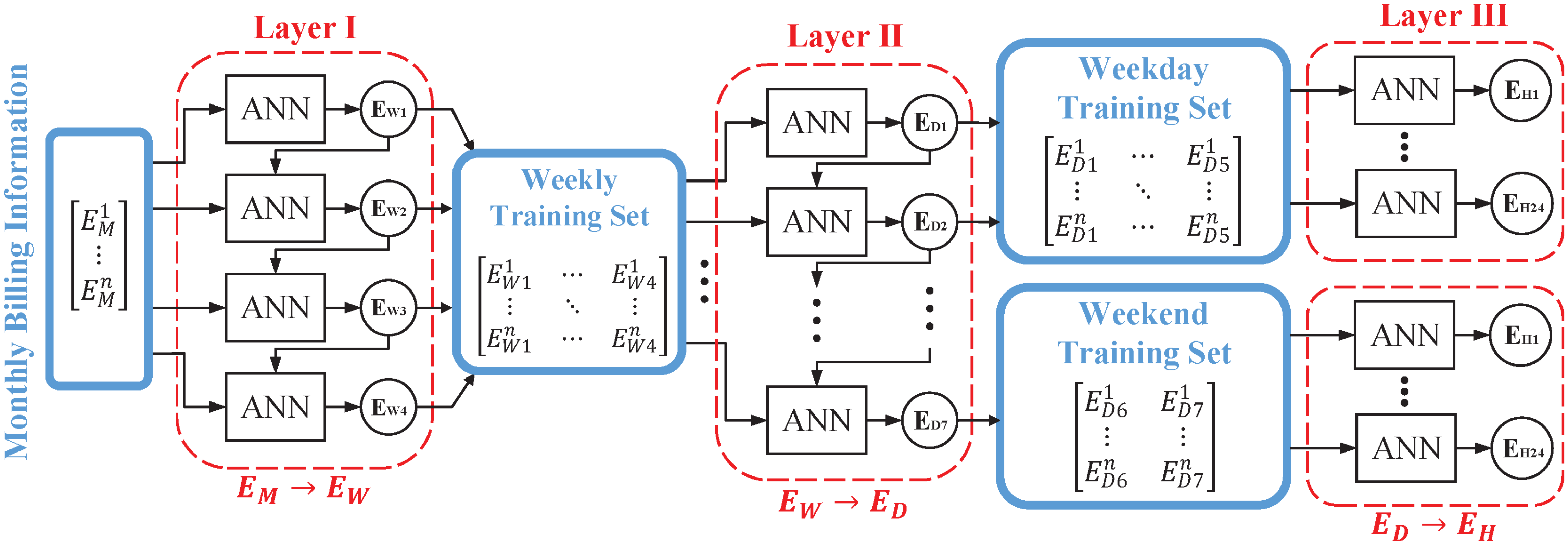
✓**Counter-Measures:**

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





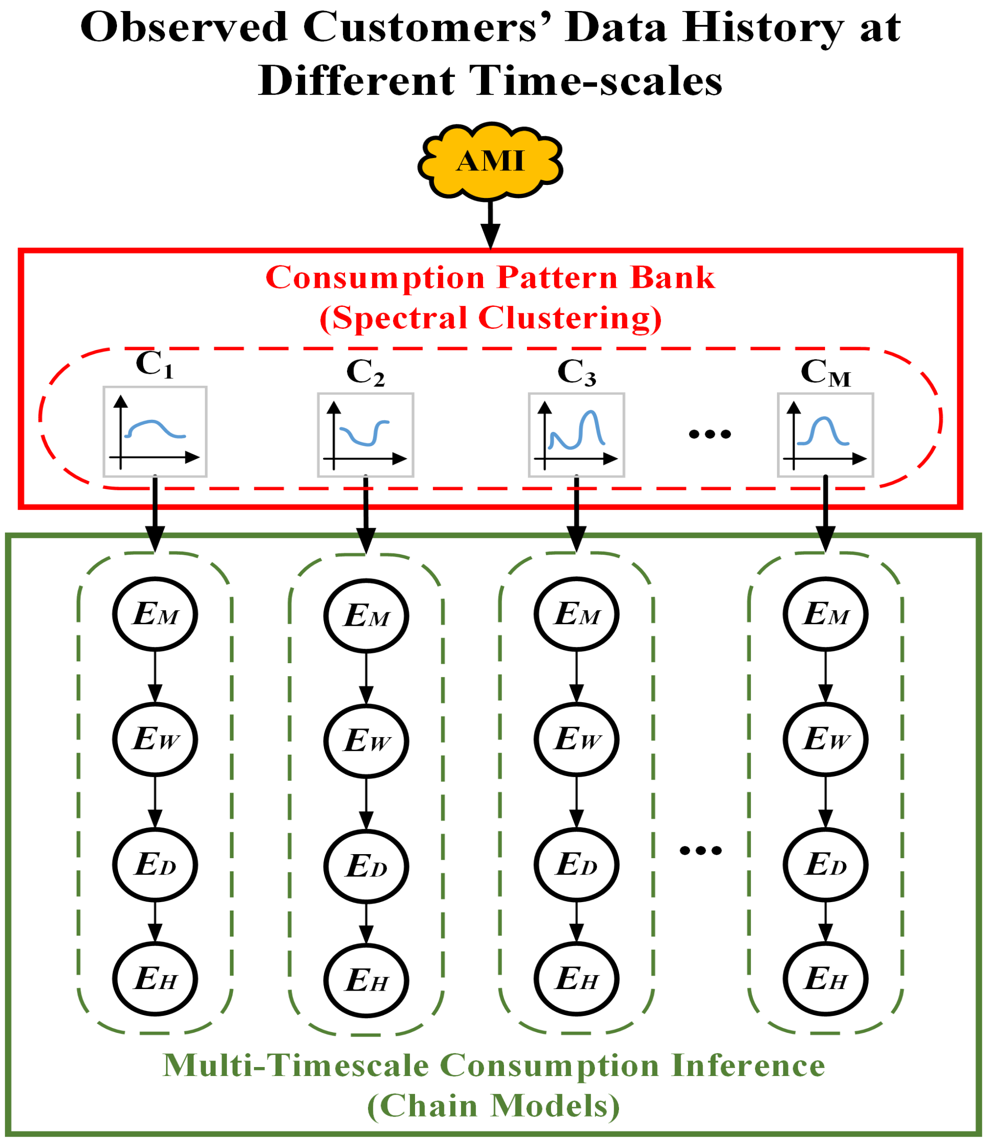
Solution Step II: Using Observed Customers' Data for Training Multi-Timescale Load Inference Chain Models



E_M – Monthly Consumption
E_W – Weakly Consumption
E_D – Daily Consumption
E_H – Hourly Consumption

✓Extends observability using data of customers with smart meters to obtain a stage-by-stage consumption transition process (Maintains High Correlation!)

Solution Step III: Observed Customer Daily Load Pattern Bank Formation and Training Multi-Timescale Models



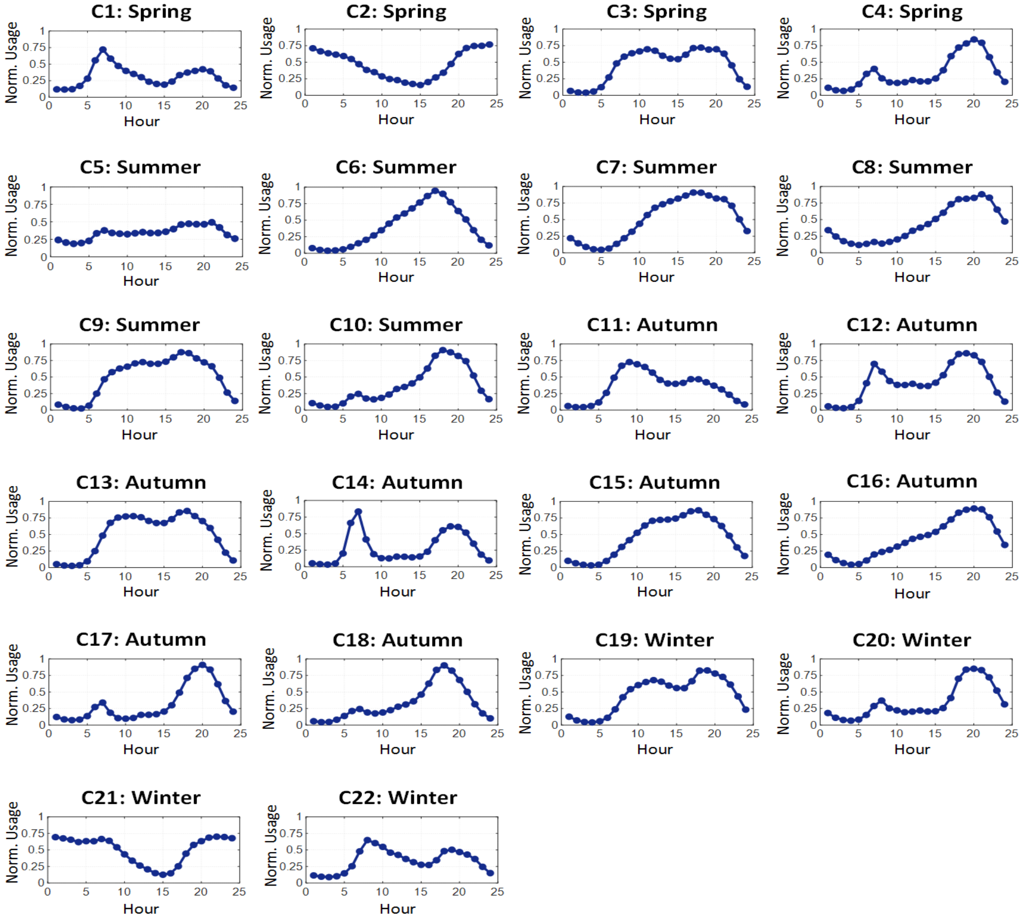
- Problem: Performance of Multi-timescale Chain Models Highly Depend on Typical Daily Consumption Patterns of Different Customers

- Solution: Assign a Multi-Timescale Model to Each Typical Load Behavior Pattern Discovered From Observed Loads (Method: Data Clustering)

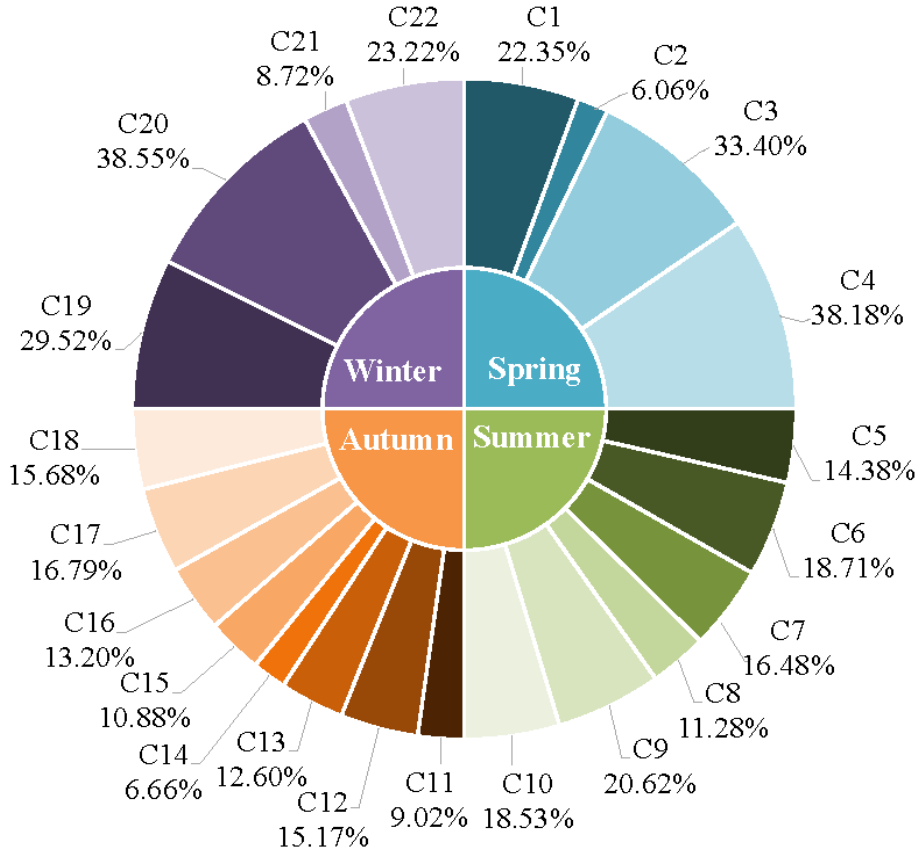
- Train Load Inference Chain Models Using the Data of Observed Customers Belonging to Each Cluster (C_i)

Customer Behavior Pattern Bank: Sensitivity to Time of Day and Load Type

Typical discovered load profiles in different seasons from smart meter data

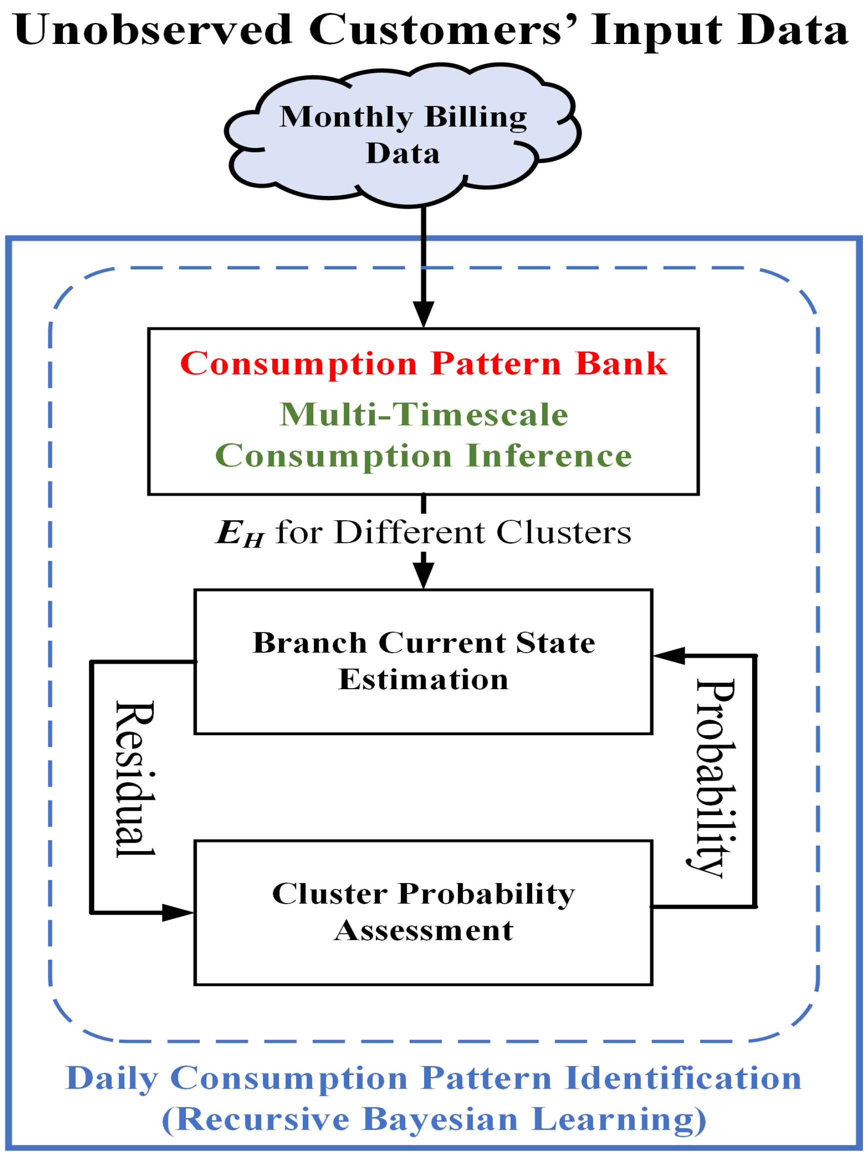


The percentage of customers belonging to each typical load profile in different seasons



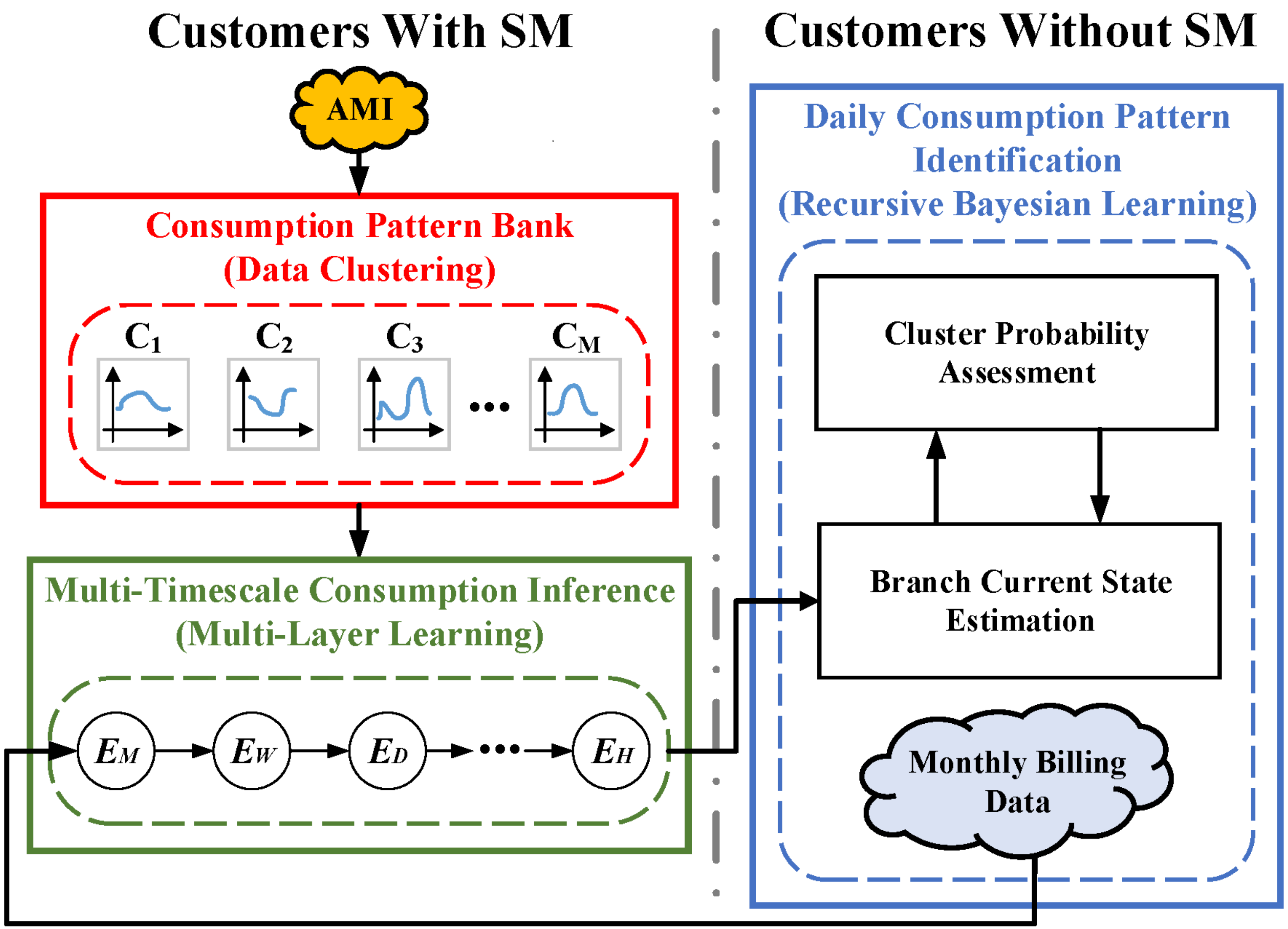
✓ **Methodology: Data Clustering (Unsupervised Learning – Spectral Clustering Algorithm)**

Solution Step IV: Unobserved Customers' Pattern Identification and Hourly Consumption Inference

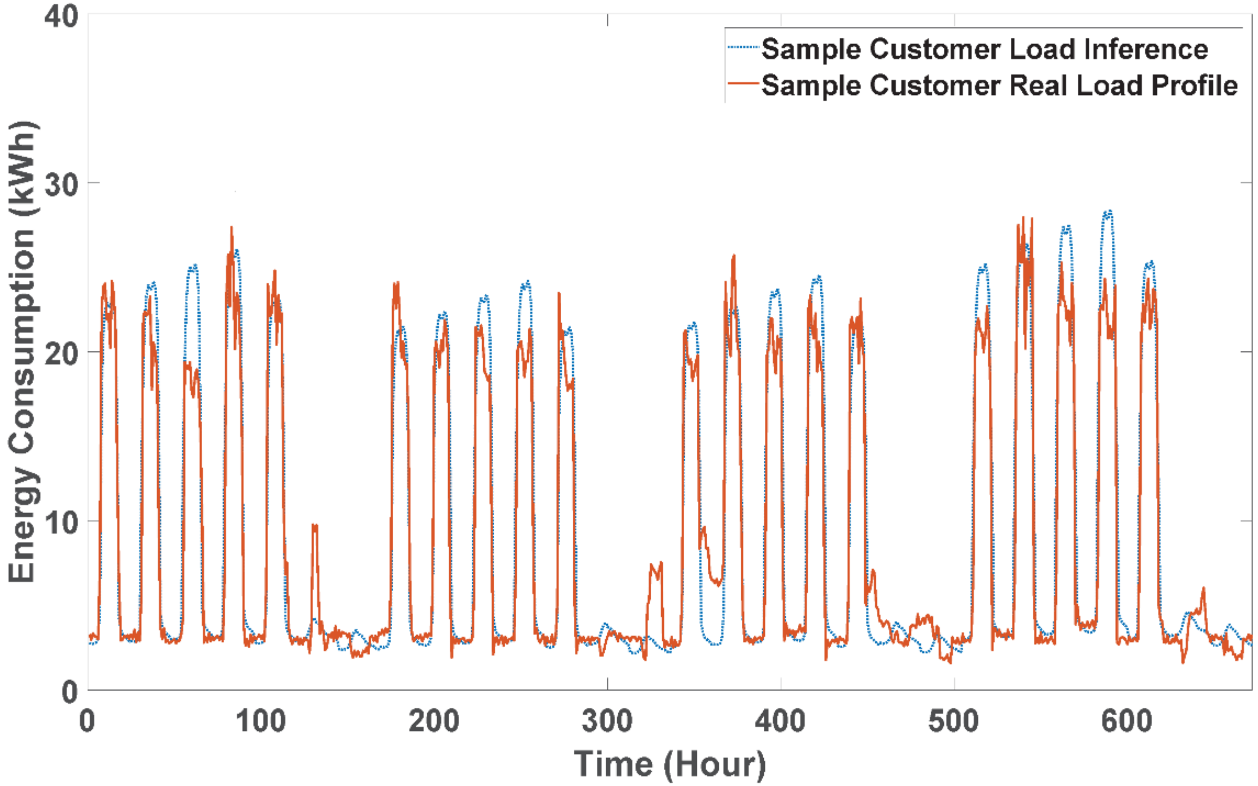


- Basic Idea: Pick the Cluster that has the Best State Estimation Performance for Each Customer
- Methodology: Assign and Update Probability Values to Different Clusters Based on State Estimation Residuals (Recursive Bayesian Learning)
- Outcome: Pick the Most Probable Cluster for Each Unobserved Customer and Use its Corresponding Chain Model for Hourly Load Inference

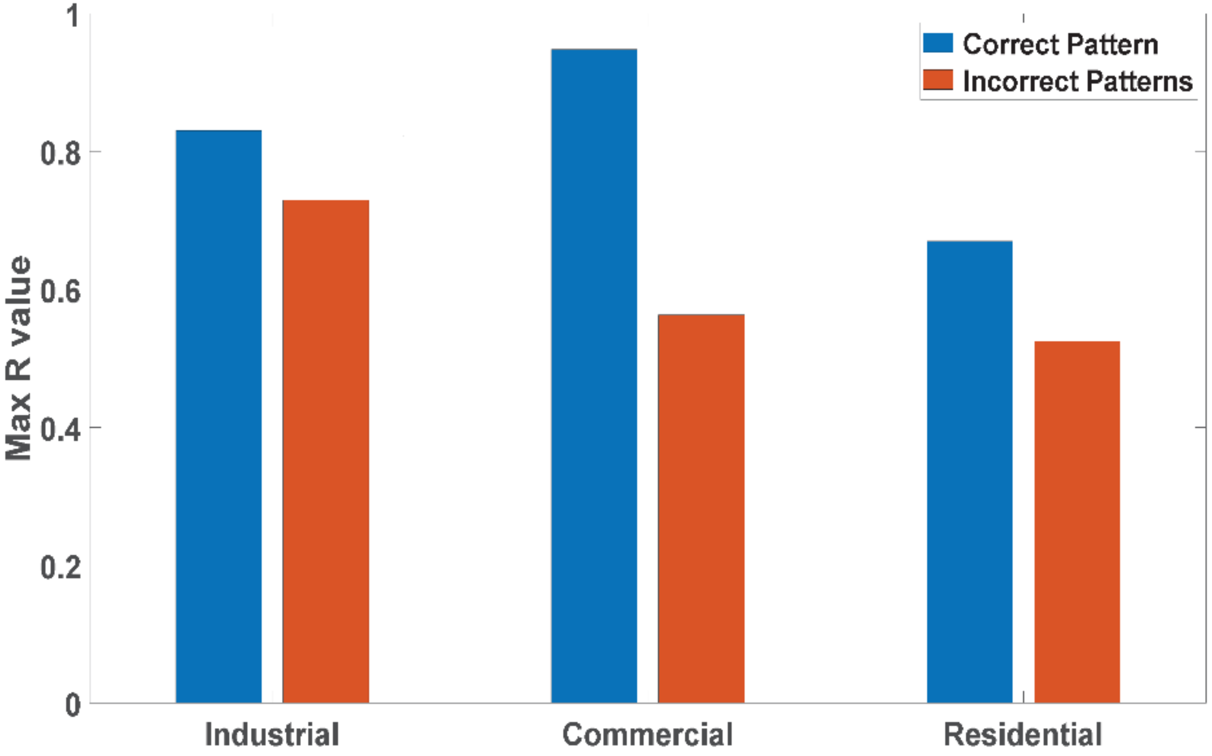
Overall Structure of the Proposed Solution



Numerical Results: Unobserved Individual Customer Hourly Load and Pattern Inference

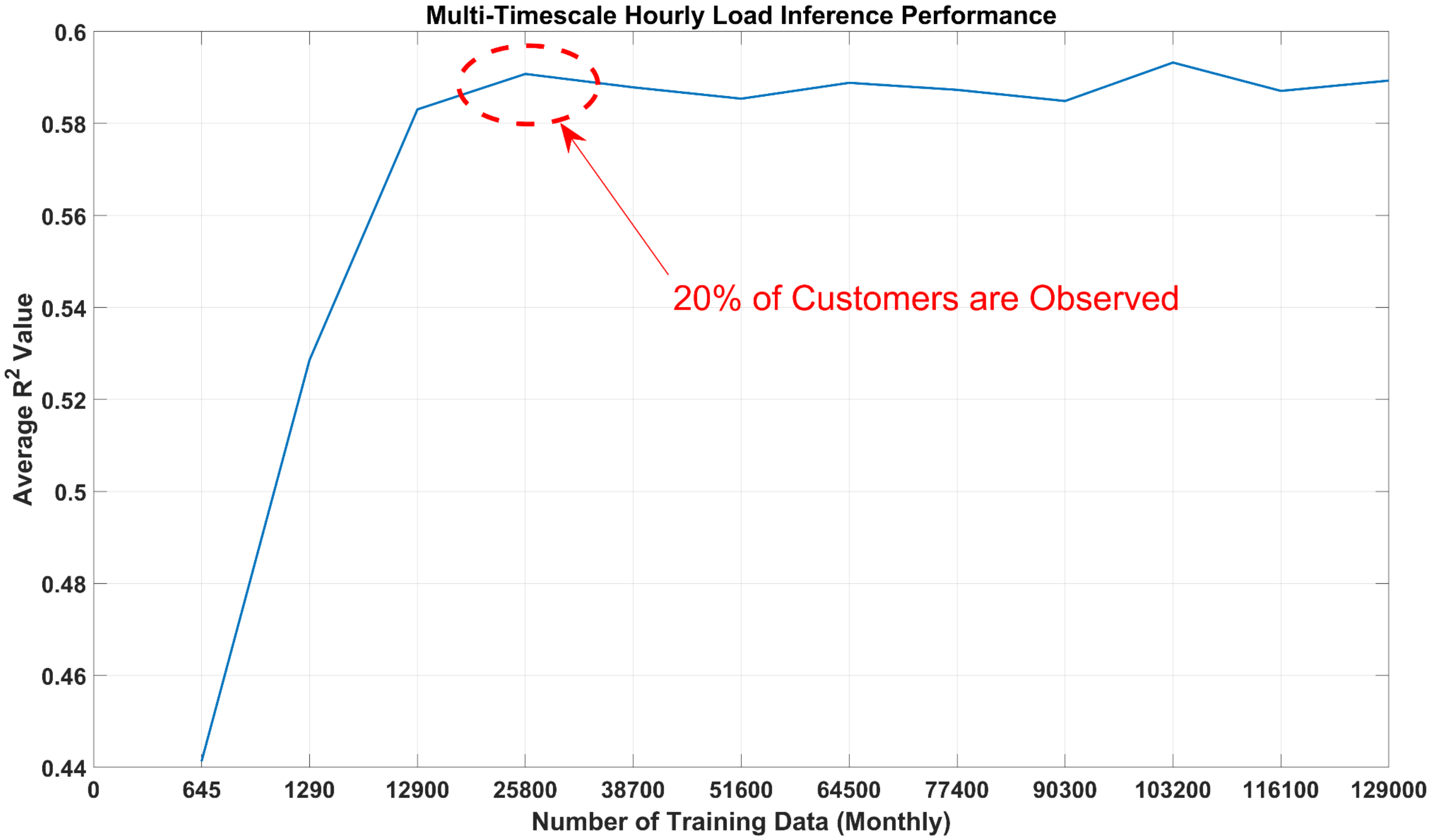


Inferring the hourly demand of an unobserved residential load in one month (average estimation error ≈ 8.5% of total energy)



Impact of accurate consumption pattern identification on the accuracy of the inference (industrial load patterns are close and stable)

Sensitivity Analysis of Observability



Conclusions

- We have archived a large amount of real data from utility partners, including smart meters, SCADA, PMUs and circuit models.
- We have shared one real distribution grid model with one-year smart meter data.
- A Data-Driven Load Inference method is developed for Monitoring Distribution Systems:
  - ✓ Identifying Temporal Correlations for Load Estimation
  - ✓ Exploiting Latent Trends in Customer Behavior at Different Time-Scales for Enhancing Inference Accuracy
- We have developed a data-driven method to take advantage of low-resolution but widely available smart meter data to disaggregate BTM solar generation.

Recent Publications in Data Analytics

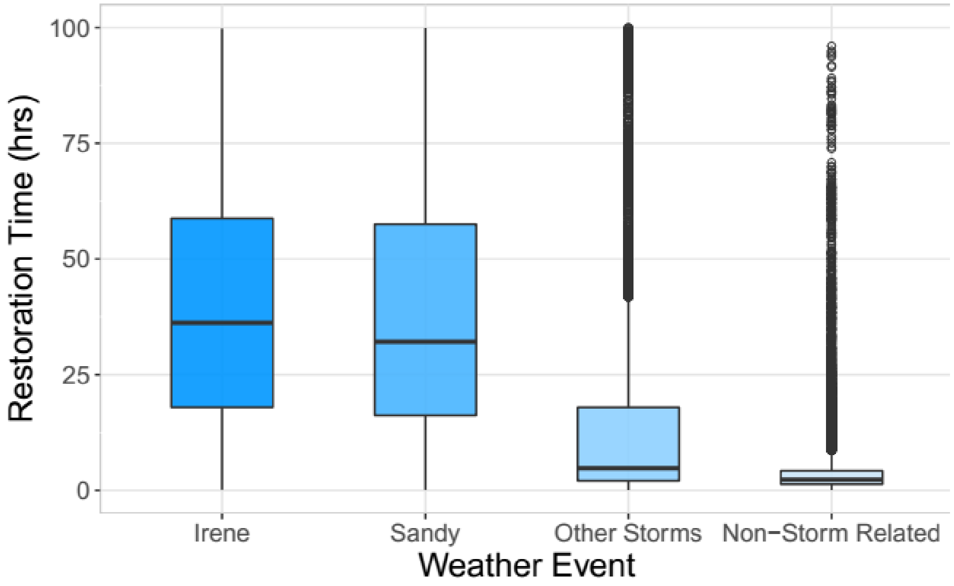
- F. Bu, K. Dehghanpour, Y. Yuan, Z. Wang, and Y. Zhang, "A Data-Driven Game-Theoretic Approach for Behind-the-Meter PV Generation Disaggregation," IEEE Transactions on Power Systems, accepted for publication.
- Q. Zhang, K. Dehghanpour, Z. Wang, and Q. Huang, "A Learning-based Power Management Method for Networked Microgrids Under Incomplete Information," *IEEE Transactions on Smart Grid*, accepted for publication.
- 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*, accepted for publication.
- 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.
- H. Sun, Z. Wang, J.Wang, Z.Huang, N. Carrington, and J. Liao, "Data-Driven Power Outage Detection by Social Sensors," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2516-2524, September 2016.
- L. Fang, K. Ma, R. Li, and Z. Wang, "A Statistical Approach to Estimate Imbalance-Induced Energy Losses for Data-Scarce Low Voltage Networks," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2825-2835, July 2019.
- 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*, accepted for publication.
- C. Wang, Z. Wang, J. Wang, and D. Zhao, "Robust Time-Varying Parameter Identification for Composite Load Modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 967-979, January 2019.
- C. Wang, Z. Wang, J. Wang, and D. Zhao, "SVM-Based Parameter Identification for Composite ZIP and Electronic Load Modeling," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 182-193, January 2019.
- T. Lu, Z. Wang, J. Wang, Q. Ai, and C. Wang, "A Data-Driven Stackelberg Market Strategy for Demand Response-Enabled Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2345-2357, May 2019.

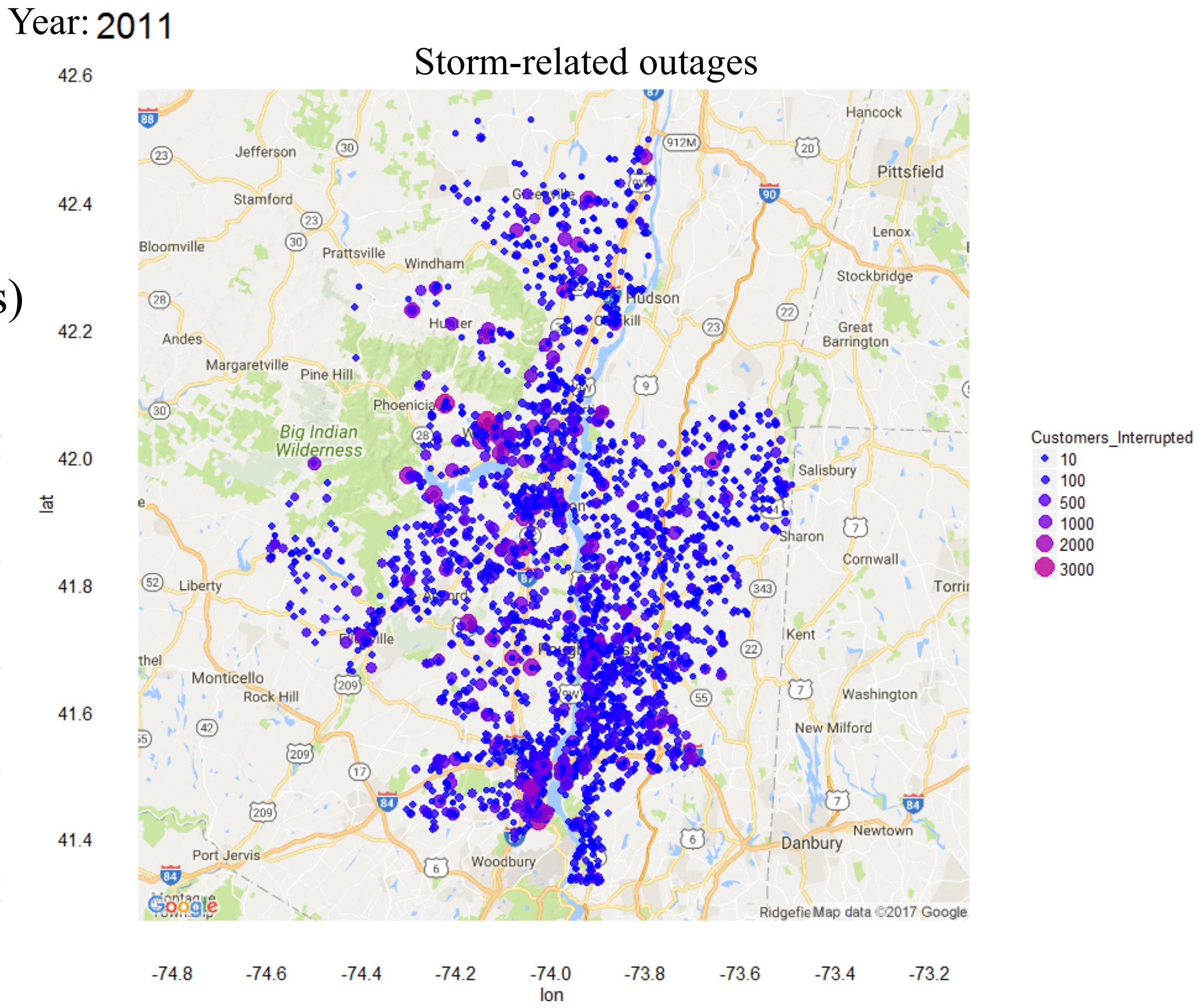
Thank You!
Q & A

Zhaoyu Wang
http://wzy.ece.iastate.edu

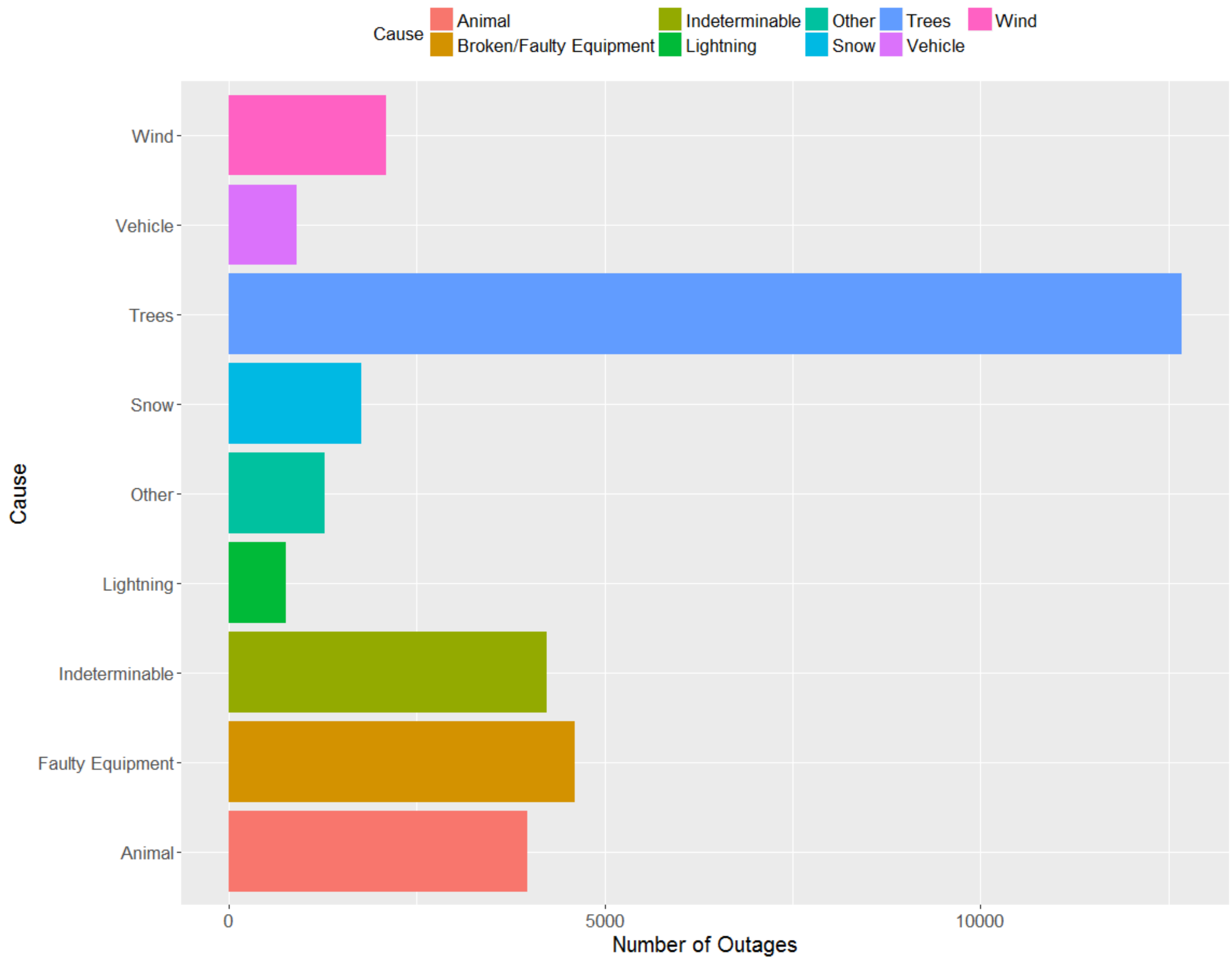
Real Data from Outage Management Systems

- Outages from 2011-2016
- 32,291 power outages
- 63 cause codes
- 253 Circuits
- 32 weather events (including 19 storm events)
- Hurricane Irene (2011) and Sandy (2012)



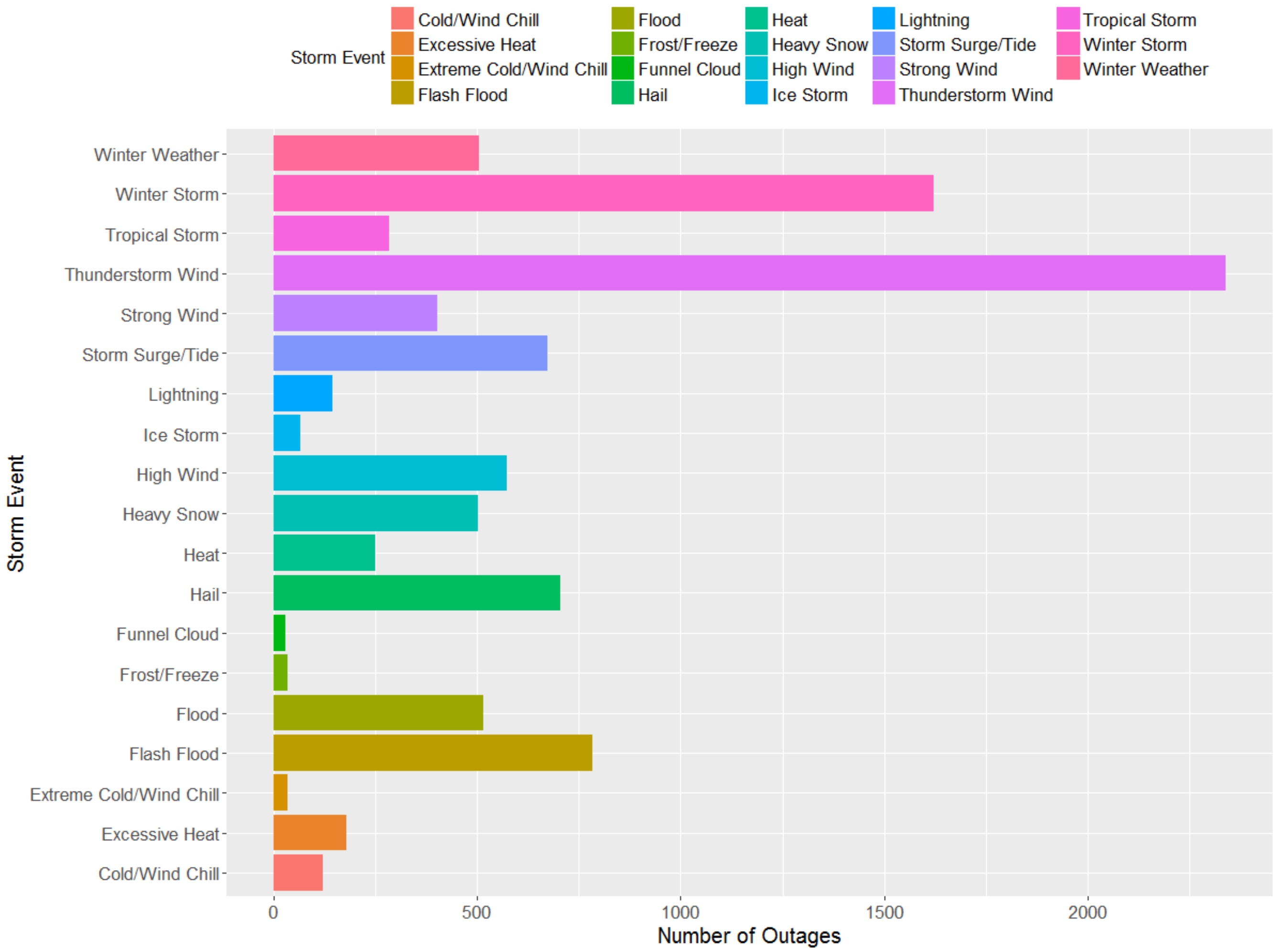


Outage Causes



Cause	Restoration Time (hrs)		Repair Time (hrs)	
	mean	standard deviation	mean	standard deviation
Animal	1.79	1.76	0.41	0.63
Broken/Faulty Equipment	4.43	5.07	1.58	2.38
Indeterminable	3.32	4.22	0.67	1.29
Lightning	5.26	5.46	0.82	2.04
Other	3.41	13.11	1.46	13.26
Snow	37.37	29.04	0.89	4.87
Trees	10.48	17.59	0.97	2.07
Vehicle	2.82	2.50	1.91	6.11
Wind	25.65	24.73	0.84	3.78

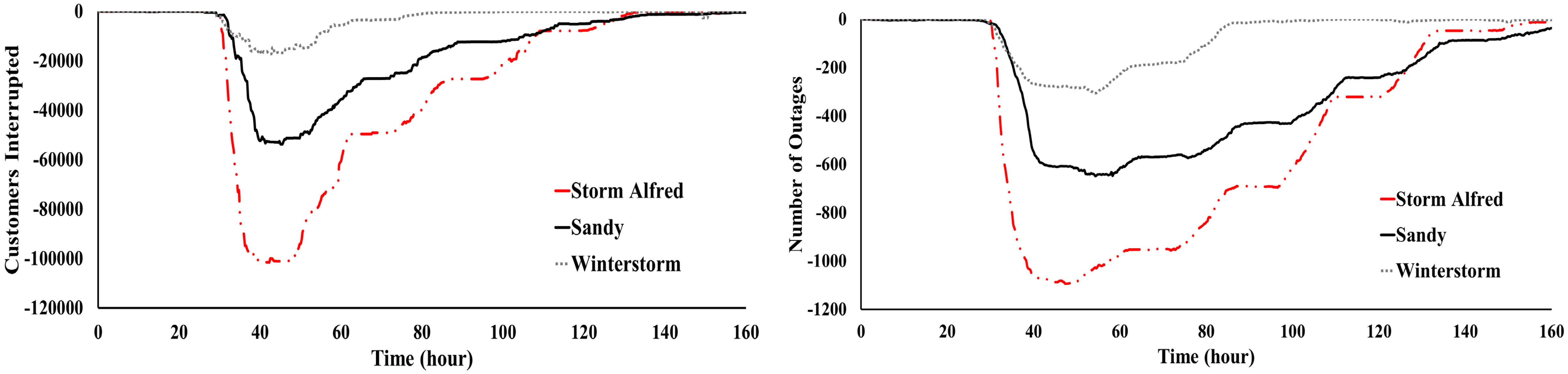
Outages and Weather Events



Weather Event	Restoration Time (hrs)		Repair Time (hrs)	
	mean	standard deviation	mean	standard deviation
Cold/Wind Chill	2.73	1.79	1.09	1.39
Excessive Heat	6.80	5.84	1.25	1.32
Extreme Cold/Wind Chill	2.73	2.07	1.74	2.31
Flash Flood	26.07	28.68	0.76	3.07
Flood	30.55	25.63	0.91	4.46
Frost/Freeze	2.07	1.55	0.67	1.14
Funnel Cloud	9.23	6.58	0.82	0.77
Hail	8.14	7.60	0.95	1.65
Heat	3.65	2.96	1.18	1.45
Heavy Snow	32.34	27.75	0.59	2.21
High Wind	26.46	26.35	0.78	4.34
Ice Storm	50.29	18.89	1.22	4.34
Lightning	4.57	4.07	1.12	1.35
Storm Surge/Tide	38.03	28.45	1.03	5.19
Strong Wind	6.08	6.26	1.41	2.37
Thunderstorm Wind	7.39	7.15	1.02	1.89
Tropical Storm	35.50	22.37	1.73	5.52
Winter Storm	37.55	29.83	0.79	4.63
Winter Weather	4.47	5.12	1.17	1.42

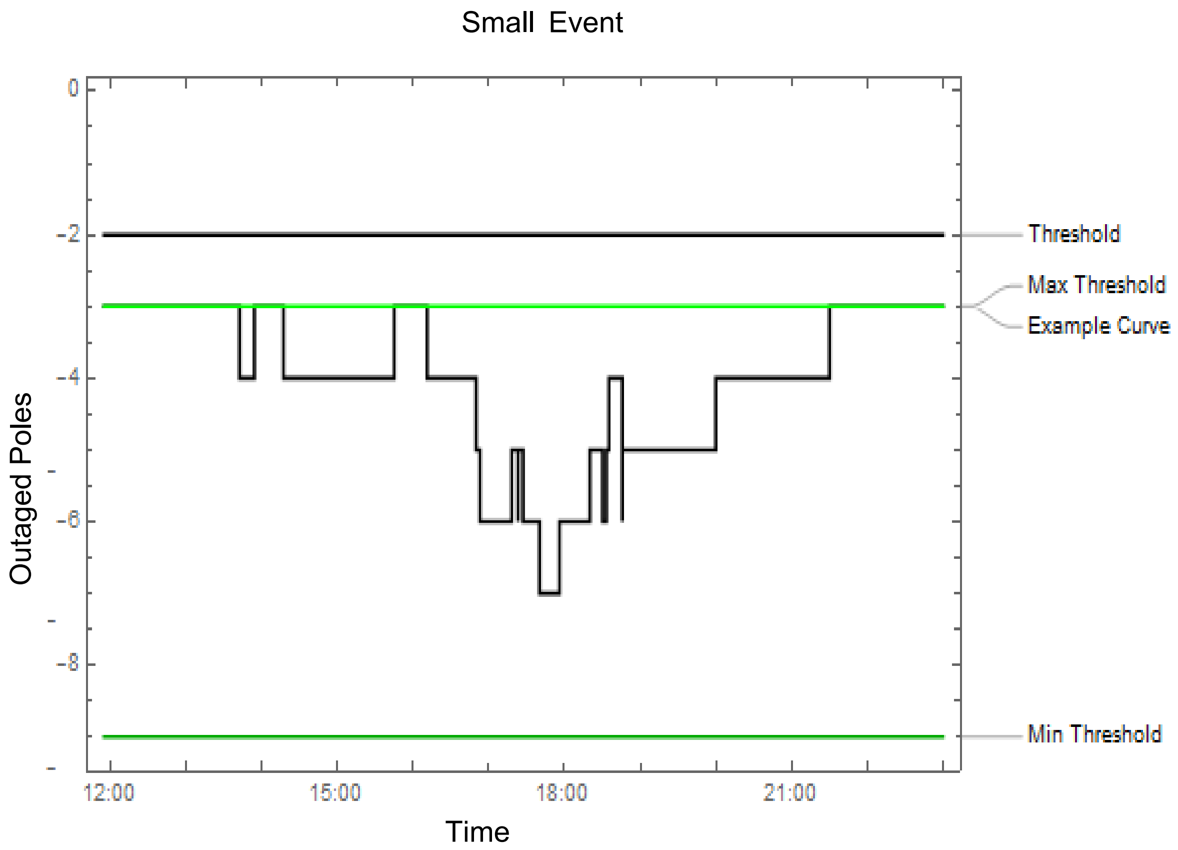
Resilience Curves – Real Data

- The figures show the number of interrupted customers and outages for three different events
  - Storm Alfred (October 2011)
  - Hurricane Sandy (October 2012)
  - Winter storm (November 2014)
- Storm Alfred occurred two months after Hurricane Irene



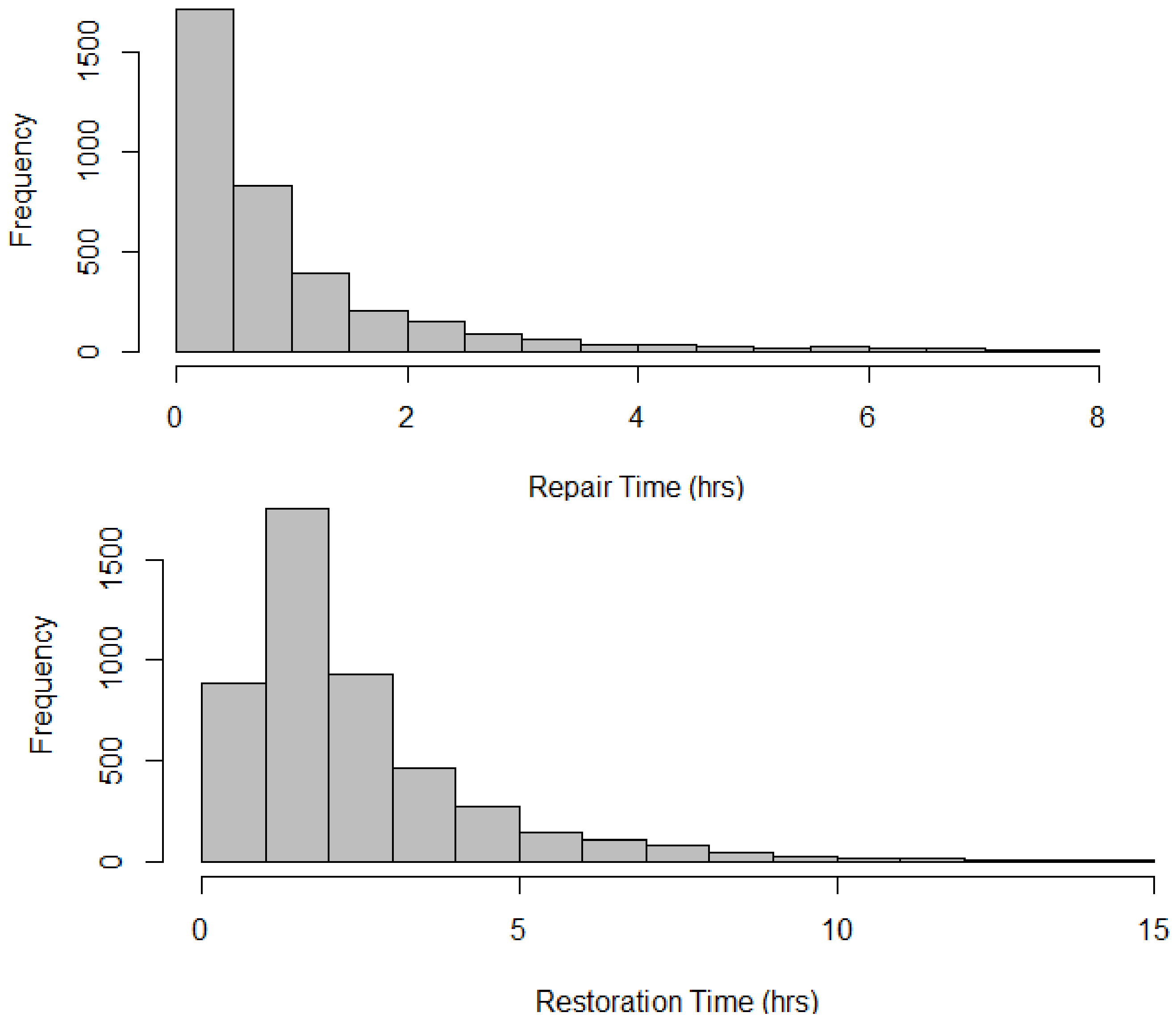
Resilience Curves – Real Data (cont.)

- There are 1637 outage events.
- We classify them into three categories:
  - 1493 small: number of outaged poles ≤ 9
  - 87 medium: 10 ≤ number of outaged poles ≤ 19
  - 57 large: 20 ≤ number of outaged poles



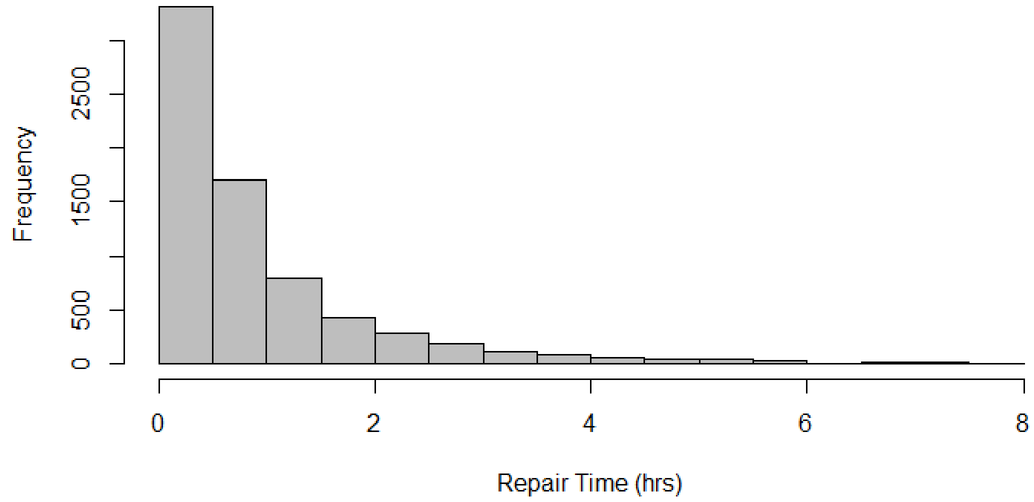
Data Statistics – Small Events

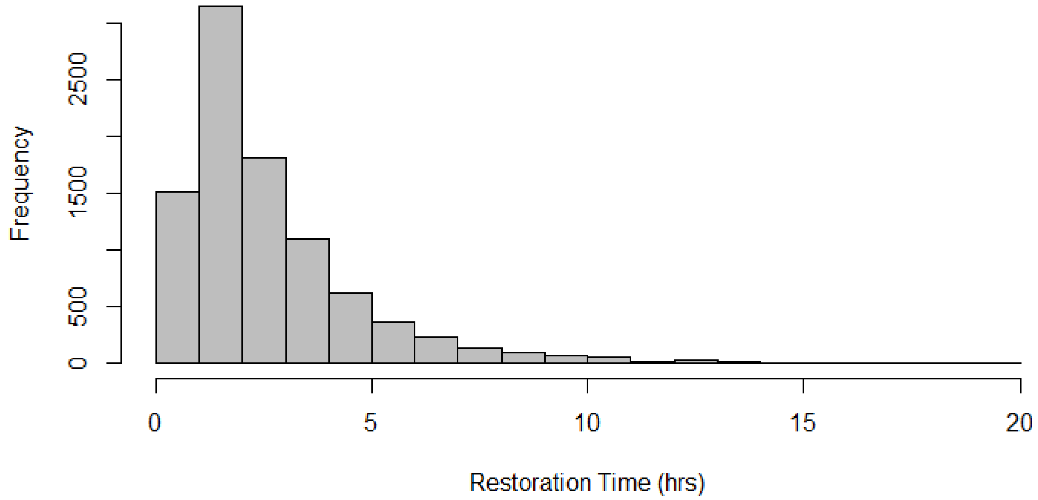
- Repair time
  - Average = 1.08 hours
  - Standard deviation = 1.69 hours
  - Maximum = 24.6 hours
- Restoration time
  - Average = 2.53 hours
  - Standard deviation = 2.68 hours
  - Maximum =64.78 hours



Data Statistics – Medium Events

- Repair time
  - Average = 1.01 hours
  - Standard deviation = 1.5 hours
  - Maximum = 36.92 hours
- Restoration time
  - Average = 2.71 hours
  - Standard deviation = 2.4 hours
  - Maximum =39.73 hours





Data Statistics – Large Events

- Repair time
  - Average = 1.35 hours
  - Standard deviation = 3.33 hours
  - Maximum = 91.5 hours
- Restoration time
  - Average = 14.7 hours
  - Standard deviation = 21.32 hours
  - Maximum =134 hours
- The data show that large events have higher repair and restoration times