Robust Real-Time Modeling of Distribution Systems with Data-Driven Grid-Wise Observability

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

U.S. DEPARTMENT OF ENERGY

NATIONAL ENERGY TECHNOLOGY LABORATORY

Robust Real-Time Modeling of Distribution Systems with Data-Driven Grid-Wise Observability

DOE Funds: $1.41M / Share 80%

Applicant's Cost Share: $0.36M / Share 20%

Total Project Value: $1.77M

Team Members: Iowa State University (Lead), Maquoketa Valley Electric Cooperative, Argonne National Laboratory, SIEMENS, Alliant Energy, Cedar Falls Utilities.

Technology Summary

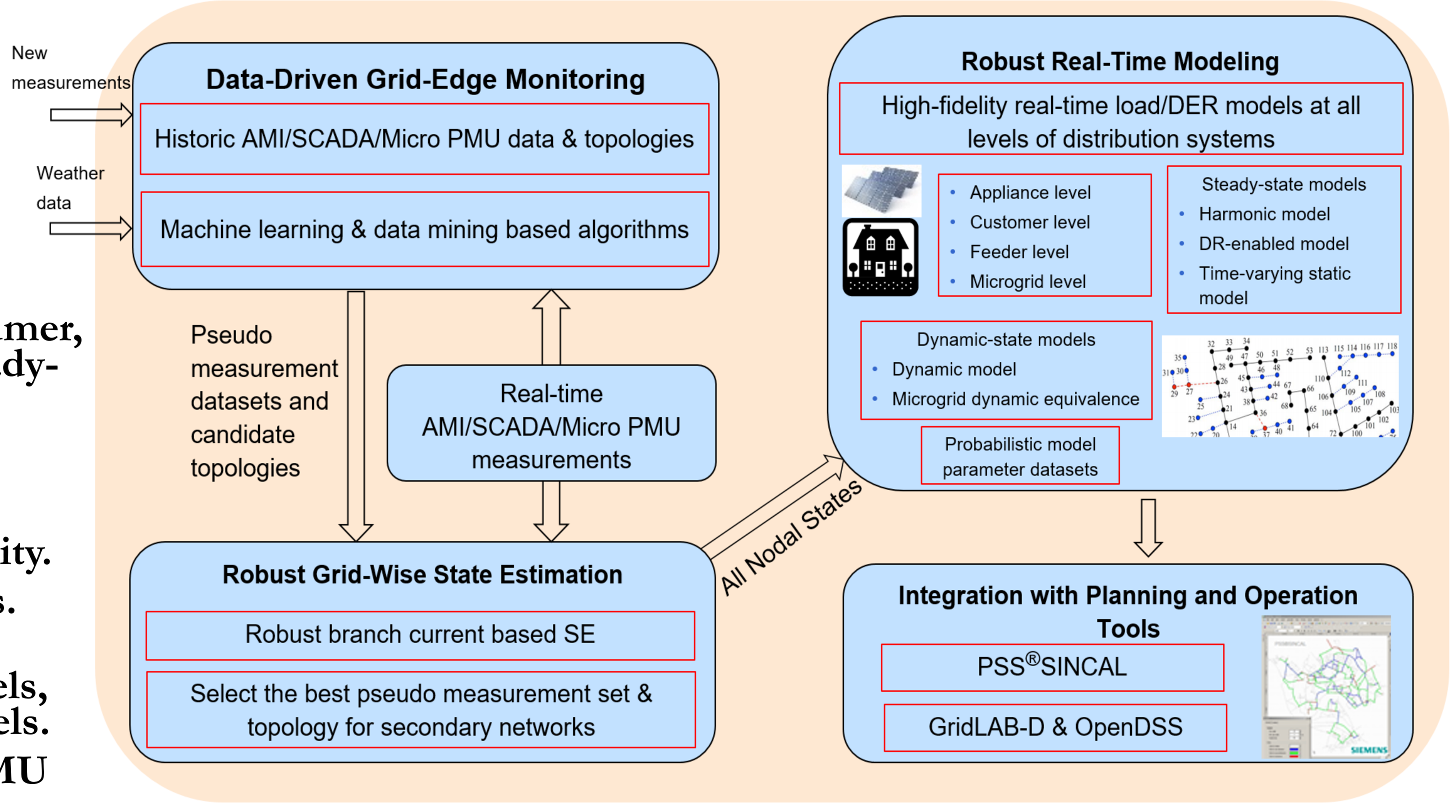
- A hybrid machine learning and branch current state estimation (BCSE) technique to enhance observability.
- Robust online modeling algorithms to develop real-time load/DER (distributed energy resource) models using practical data.
- Integration with SIEMENS software PSS®SINCAL.

Technology Impact

- Offer extended observability to DERs in secondary distribution systems.
- A set of real-time load/DER models at appliance, consumer, feeder and microgrid (MG) levels to support various steady-state and dynamic-state analyses of DERs' impacts on distribution system operation, control, and planning.

Proposed Project Objectives/Milestones

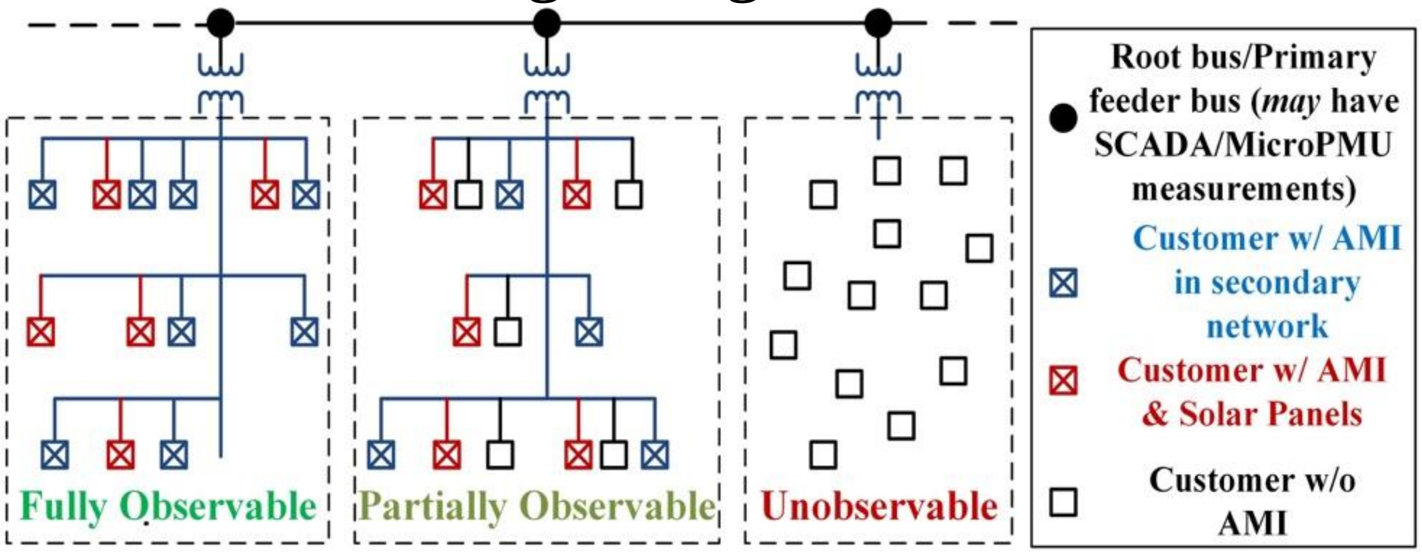
- Data-driven grid-edge monitoring to enhance observability.
- Robust grid-wise SE to provide states of all loads/DERs.
- Robust online modeling to develop real-time demand response-enabled models, static models, harmonic models, dynamic models and MG models at different voltage levels.
- Model validation using practical AMI/SCADA/MicroPMU data, and integration with PSS®SINCAL.



Leverage Voluminous Data to Enhance Observability and Develop Real-Time Load/DER Models

Project Objectives

- ***Project Definition***: Improving the observability of distribution systems for real-time monitoring, using data-driven methods.



Problem: How to Use the Data to Enhance System Observability?

- ***Project Goals:***

✓Developing machine learning models for estimating unobserved variables

✓Robust state estimation in distribution networks

✓Real-time load/DER modeling

Real Data from Utilities

- 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) with continuous updates
- **Measurement Type**: Smart Meters and SCADA
- Detailed circuit models of all feeders in Milsoft/OpenDSS and accurate smart meter locations
- **Data Time Resolution**: 15 Minutes - 1 Hour
- **Customer Type**:

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

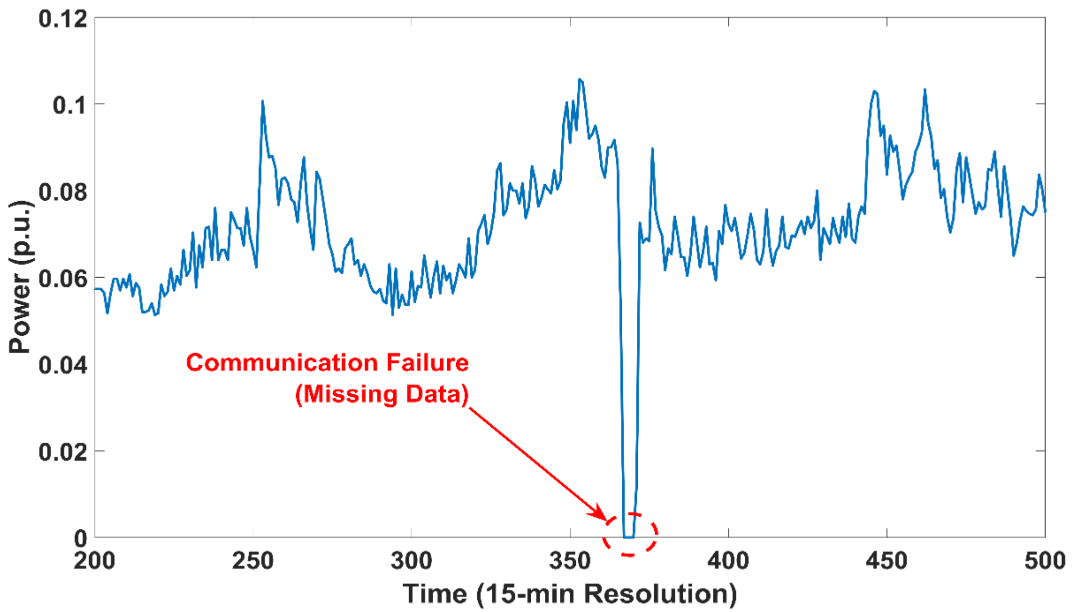
Smart Meter Data Pre-Processing

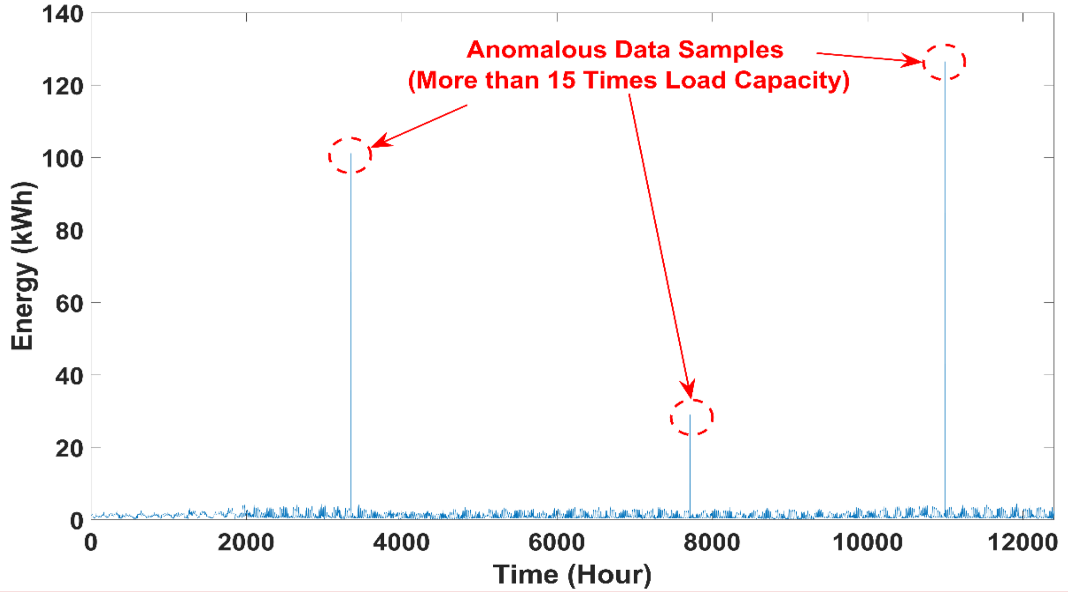
✓**Common Smart Meter Data Problems:**

- Outliers/Bad Data
- Communication Failure
- Missing Data

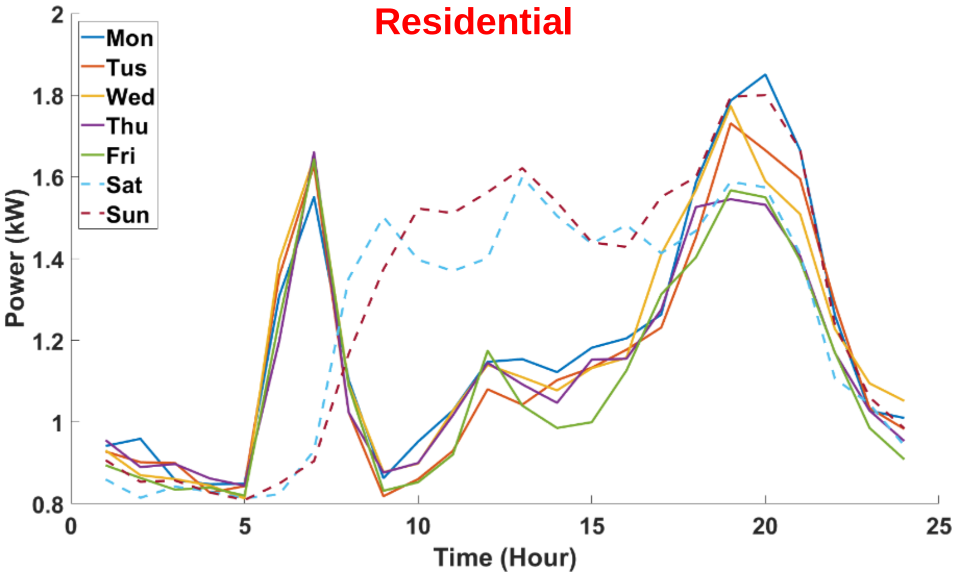
✓**Counter-Measures:**

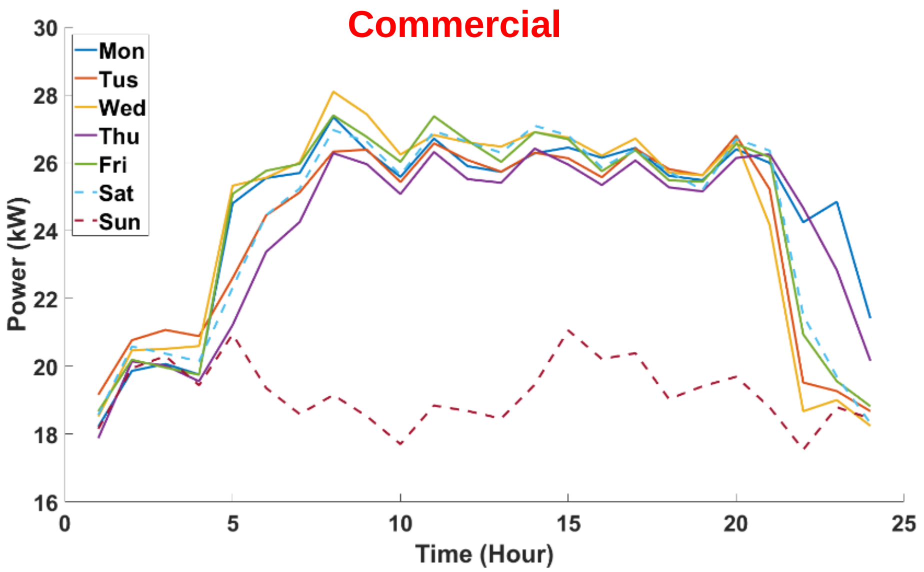
- Engineering intuition (data inconsistency)
- Conventional Statistical Tools (e.g. Z-score)
- Robust Computation (e.g. relevance vector machines)
- Anomaly Detection Algorithms

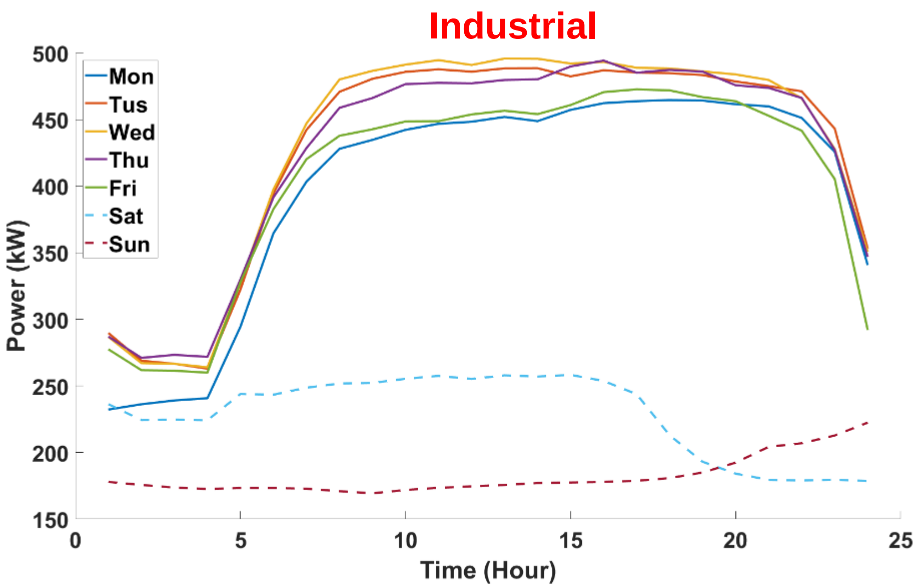




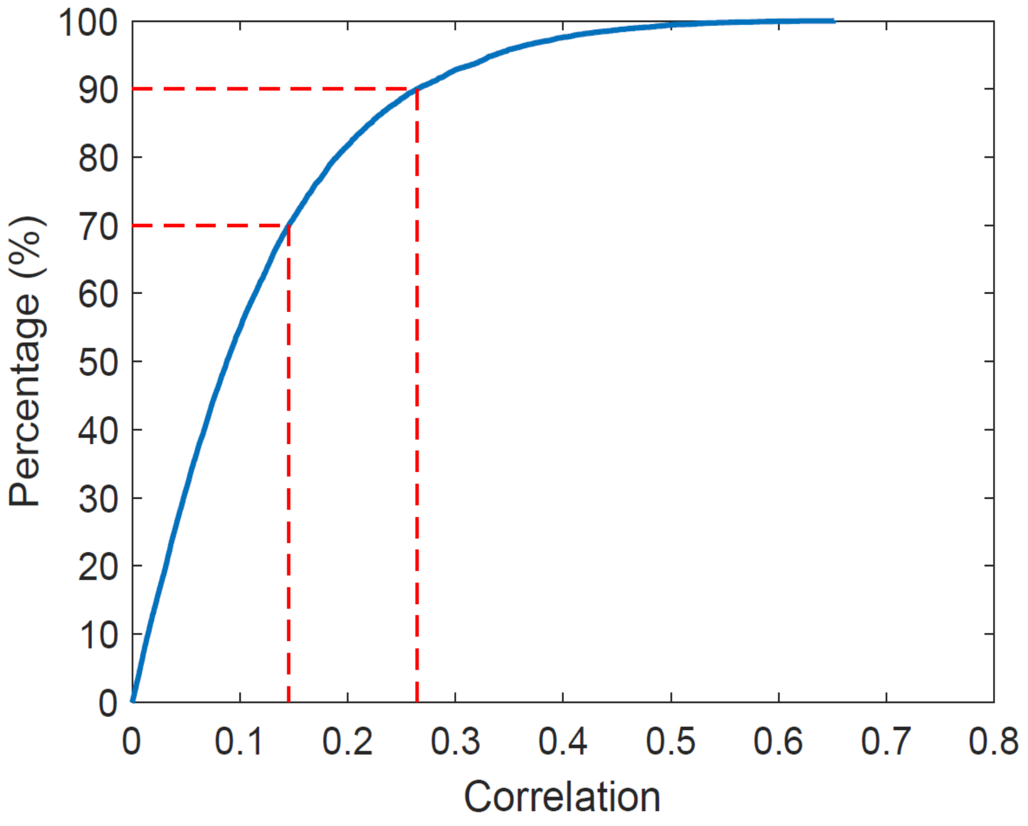
Daily Consumption of Sample Customers



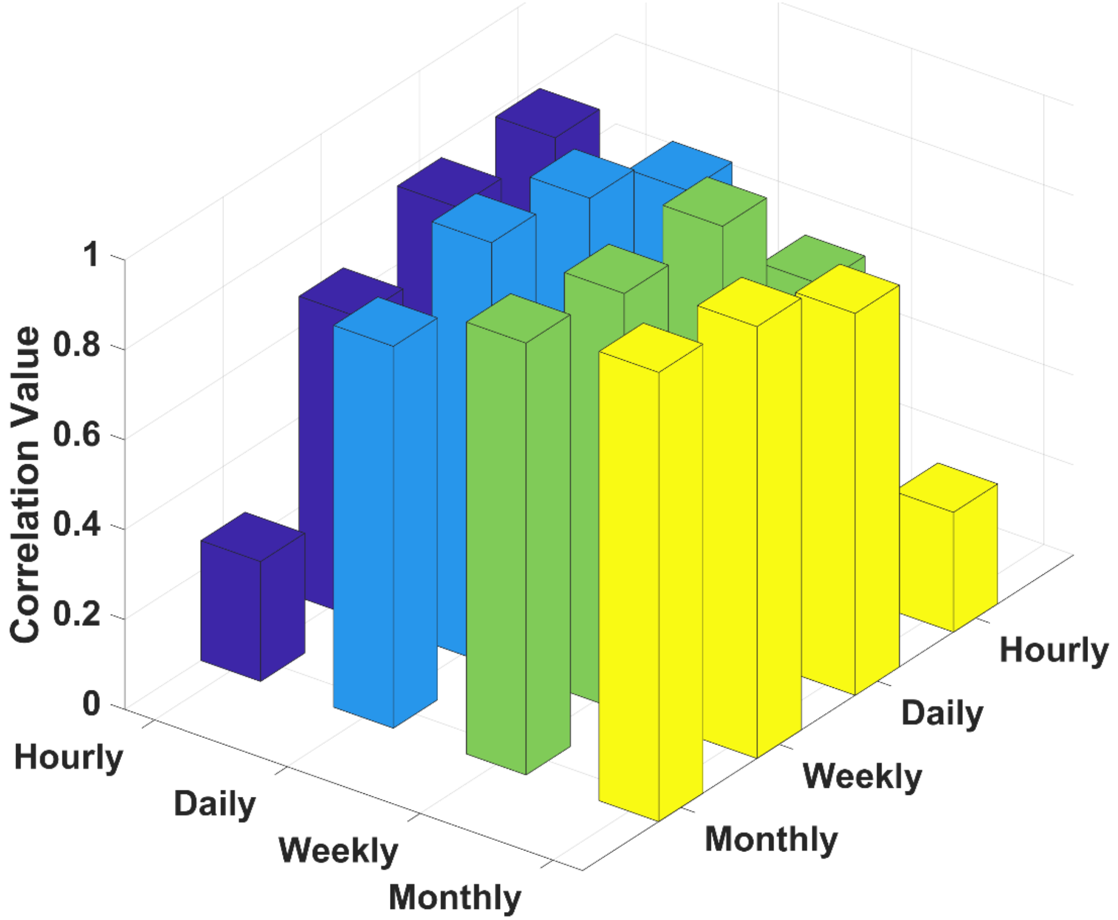




Evidence from Data: Loss of Correlation Problem



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

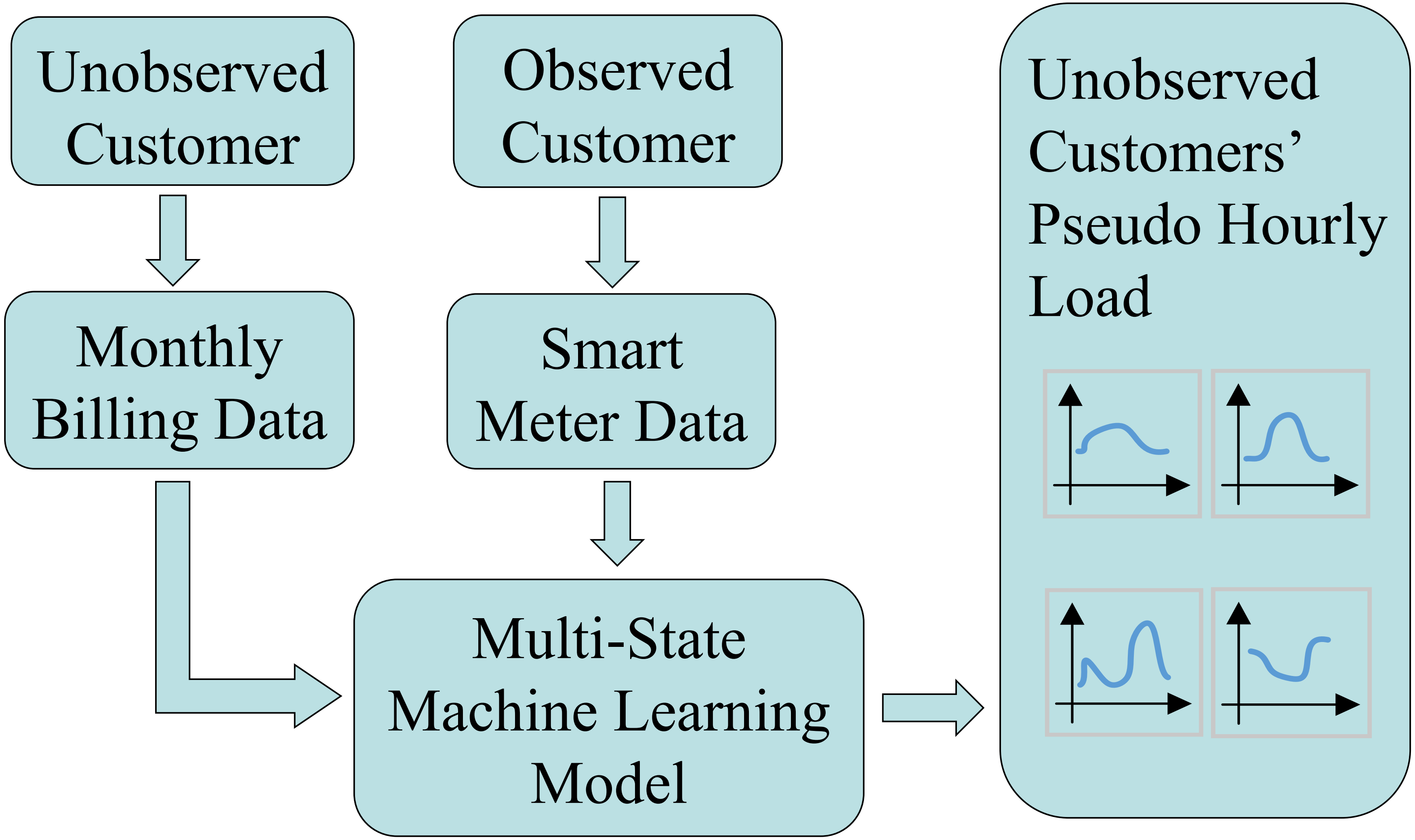


Average Correlation between Consumption of All Customers Decreases from Monthly to Hourly (Loss of Correlation Across Different Time-Scales)

Section I: Multi-timescale Data-Driven Observability Enhancement

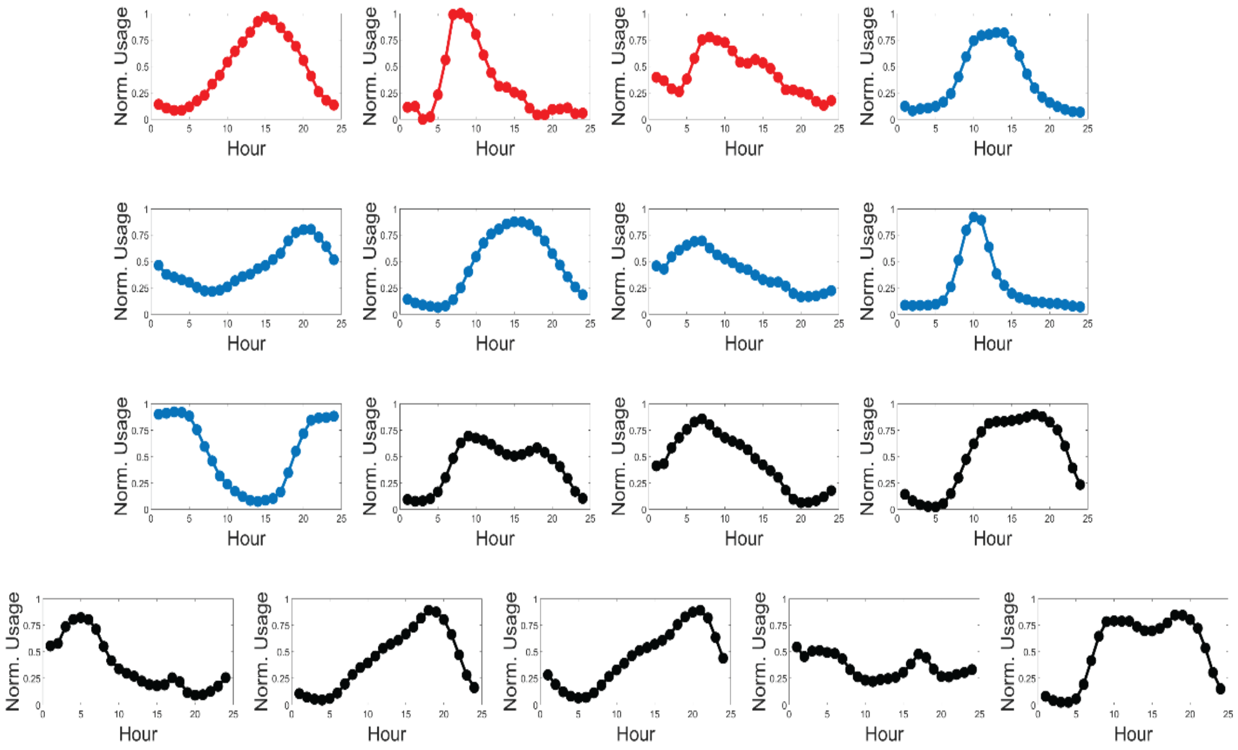
- **Problem Statement**: Inferring hourly consumption data from customer monthly billing information as pseudo-measurements
- **Challenges:**
  - ✓Loss of correlation between consumption time-series at different time-scales
  - ✓Unobserved customers’ unknown typical behaviors
- **Solution Strategy**: 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

Section I: Multi-timescale Data-Driven Observability Enhancement

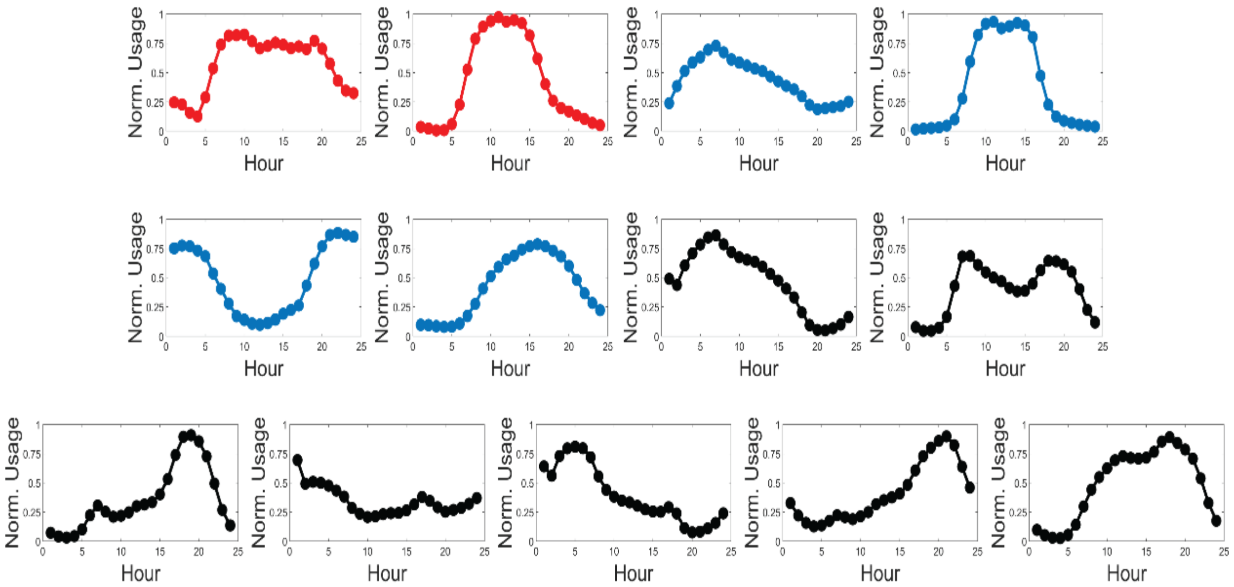


Section I: Customer Behavior Visualization: Typical Daily Demand Profile Construction from Smart Meter Data

Typical Load Patterns on Weekdays



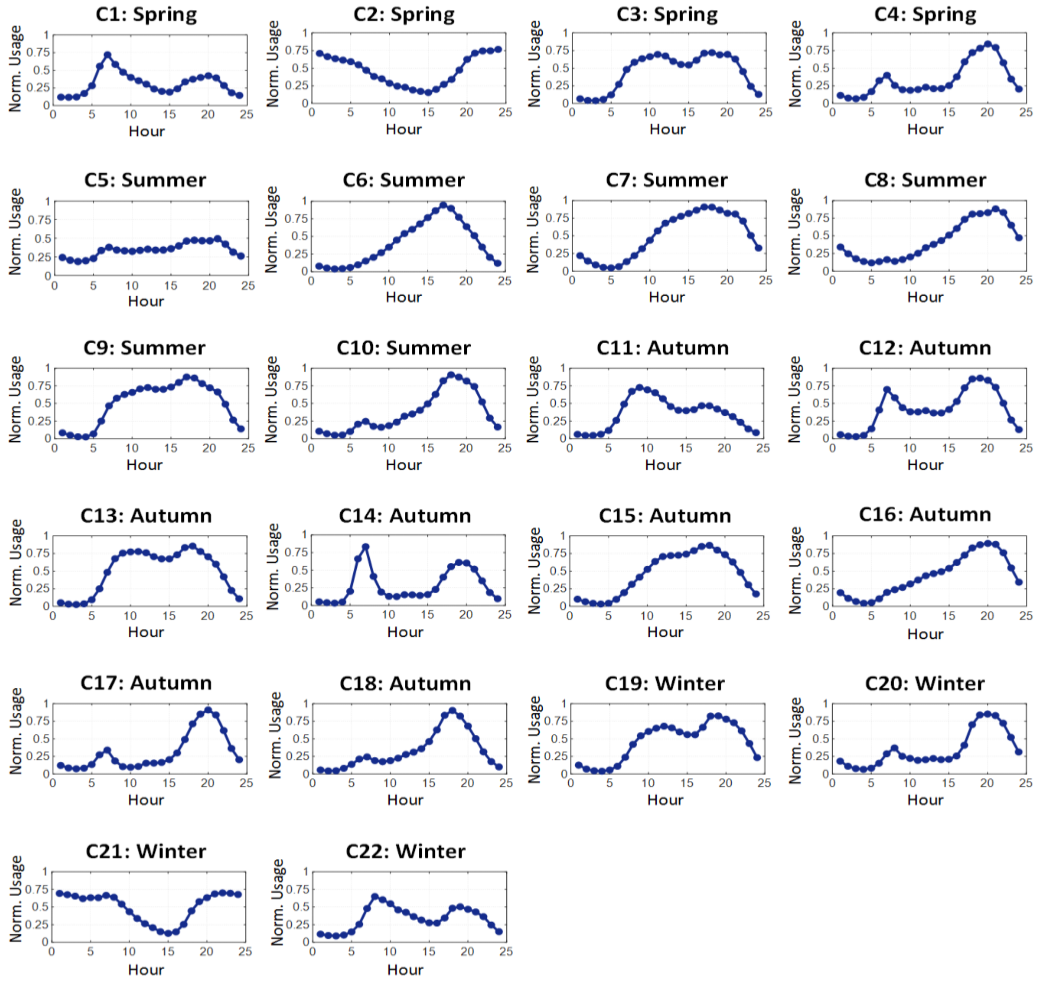
Typical Load Patterns on Weekends

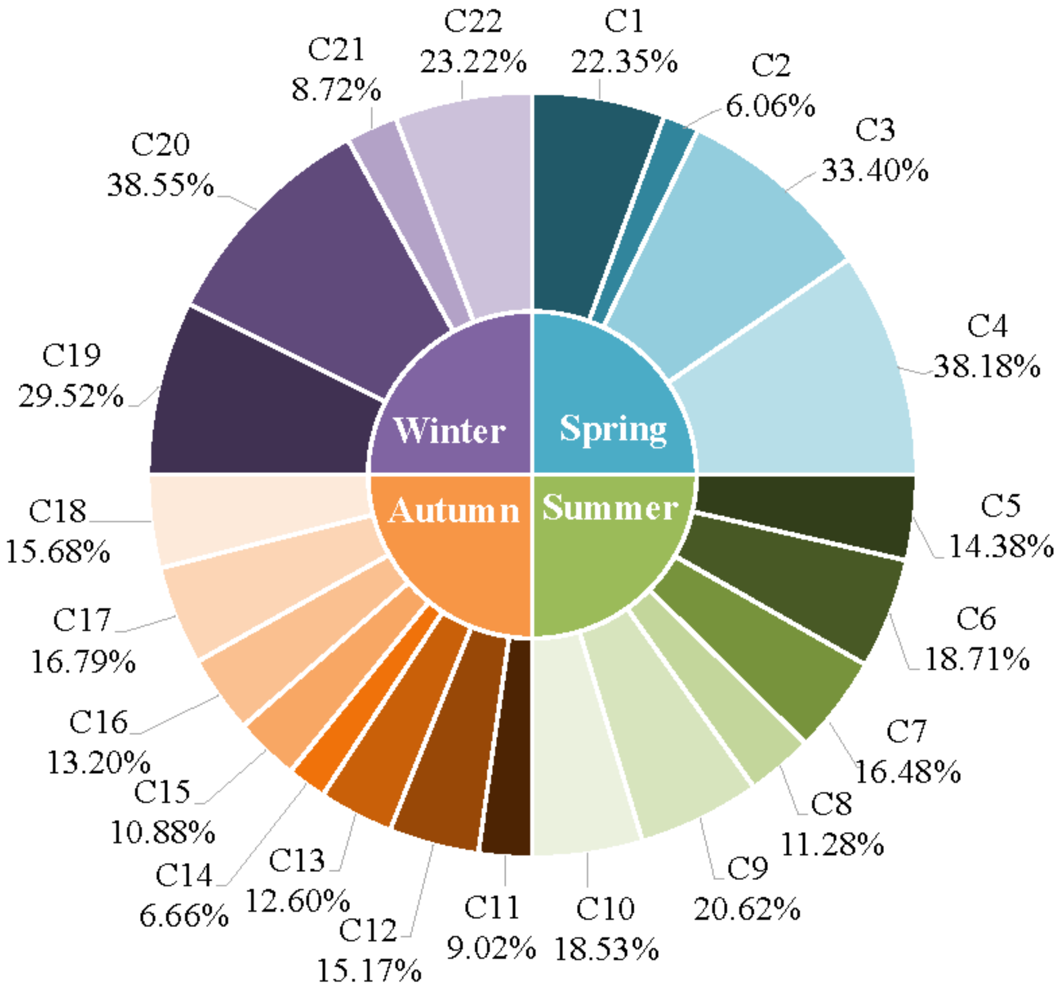


- Industrial **- Commercial** **- Residential**

✓Methodology: Data Clustering (Unsupervised Learning)

Section I: Customer Behavior Visualization: Typical Daily Demand Profile Construction in Different Seasons

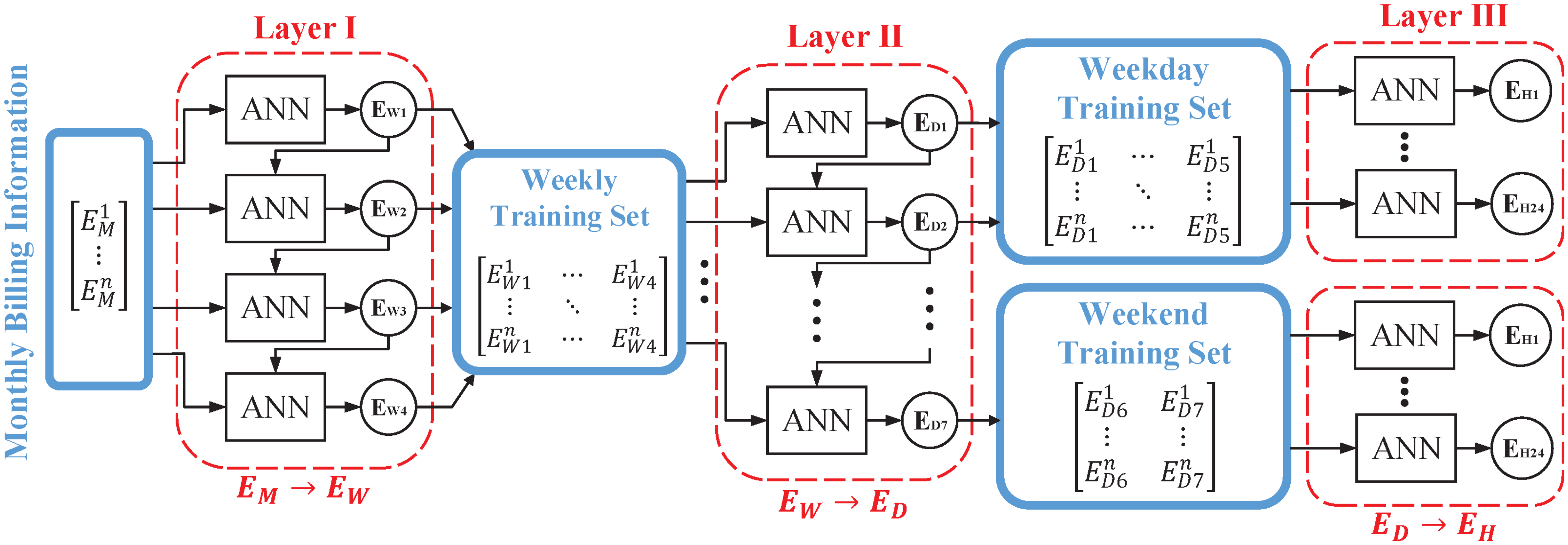




✓Typical discovered load profiles in different seasons from smart meter data

✓The percentage of customers belonging to each typical load profile

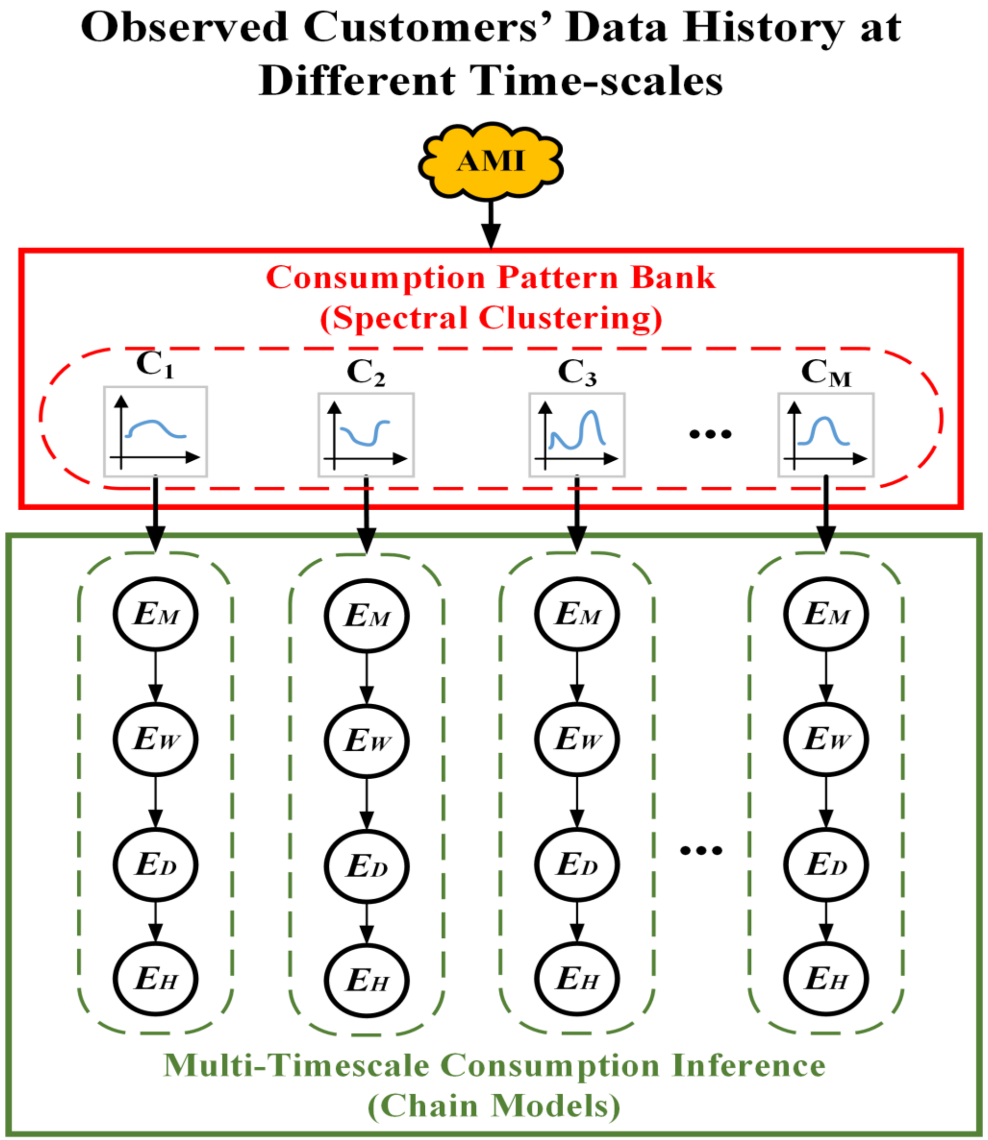
Section I: Multi-Timescale Load Inference Chain Models



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!)

Section I: Observed Customer Daily Load Pattern Bank Formation and Training Multi-Timescale Models

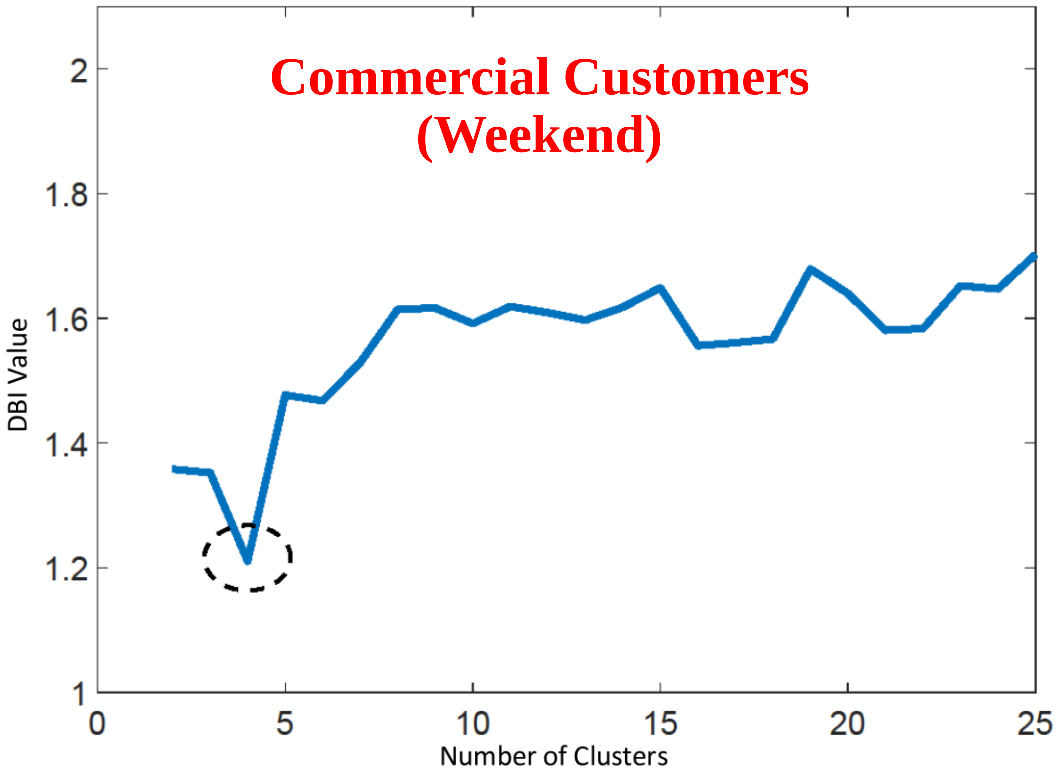


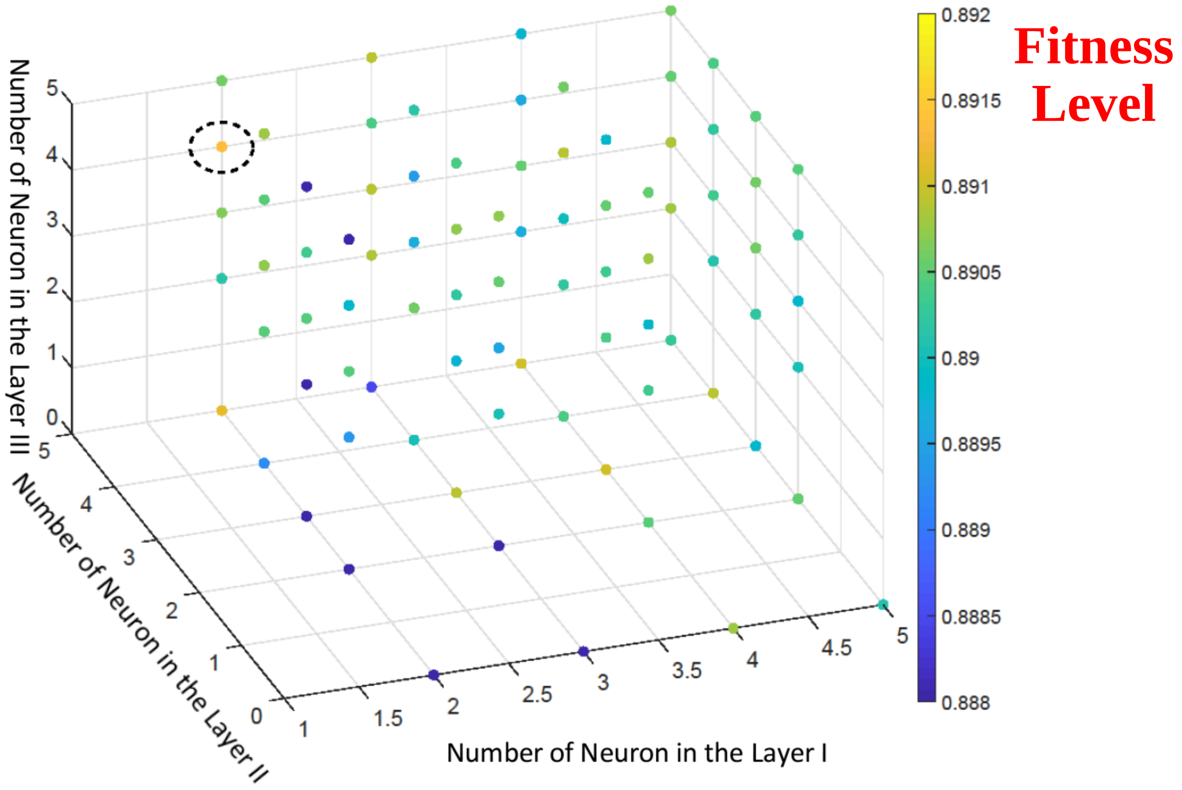
- Problem: Performance of Multi-timescale Chain Models Highly Depend on Typical Daily Consumption Patterns of Different Customers

- 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)

Section I: Learning Component Calibration

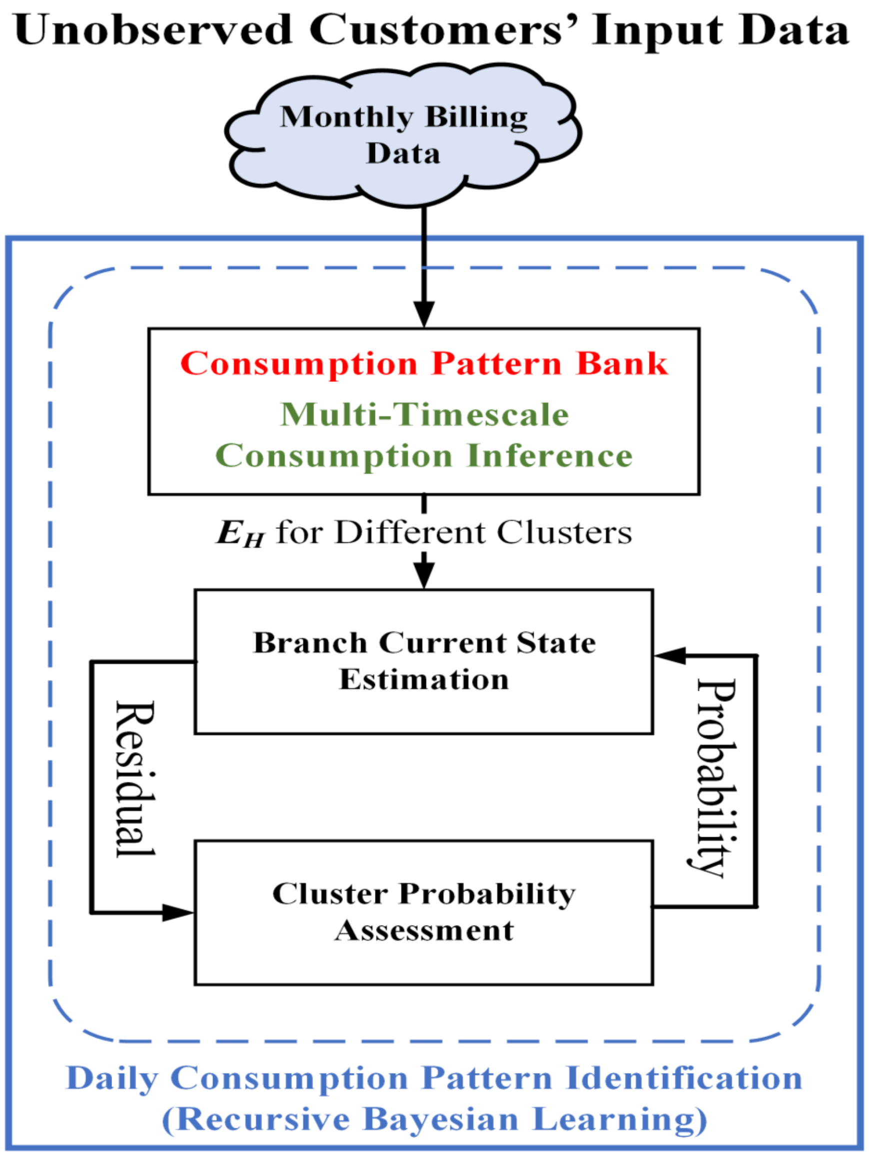




✓**Finding the optimal number of clusters for the consumption pattern bank by minimizing the Davies Bouldin Index (DBI), which measures the quality of the clustering algorithm.**

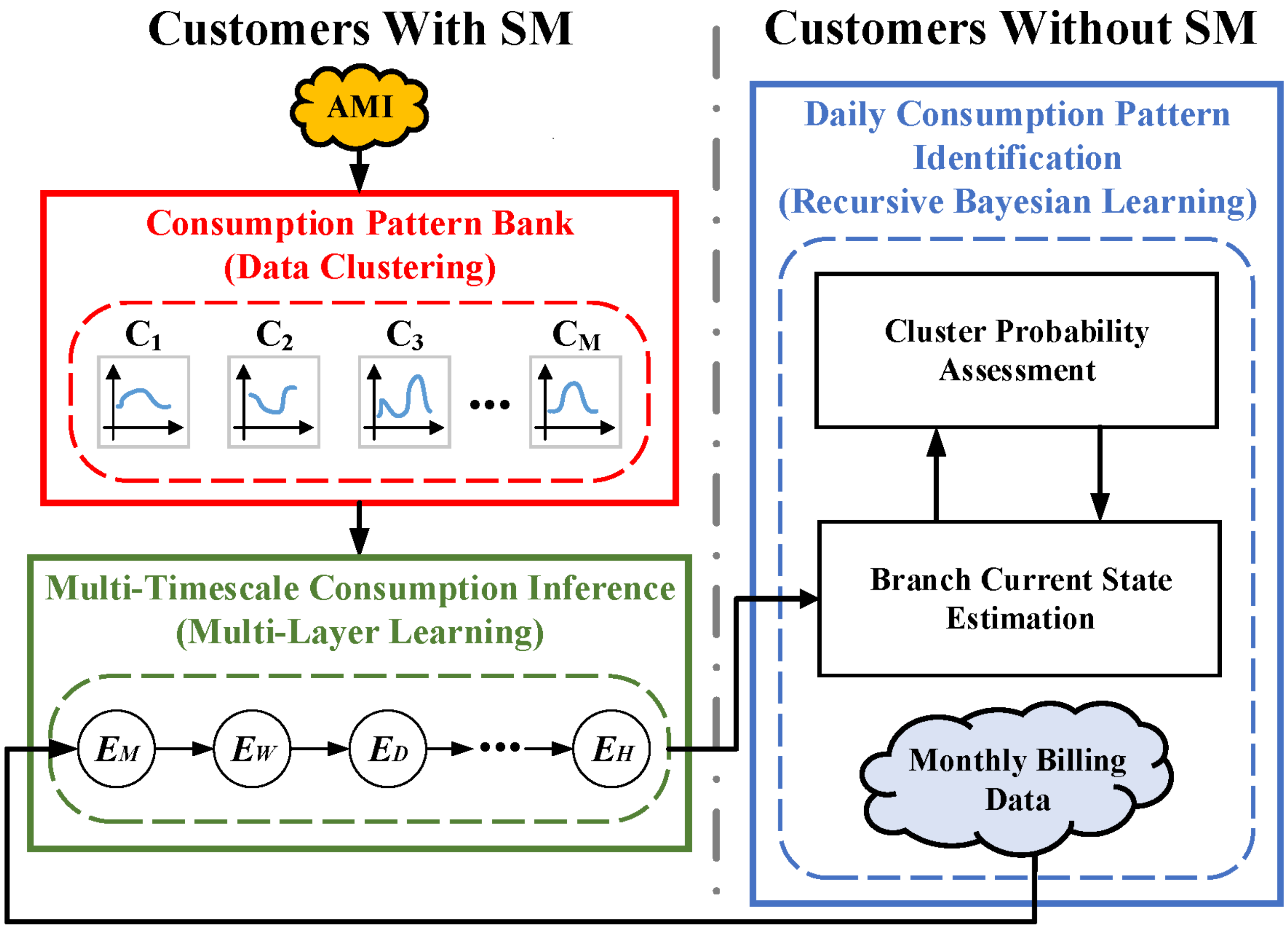
✓**Finding the optimal structure of ANNs by maximizing the performance of load inference using 10-fold cross-validation.**

Section I: Unobserved Customers' Pattern Identification and Hourly Consumption Inference

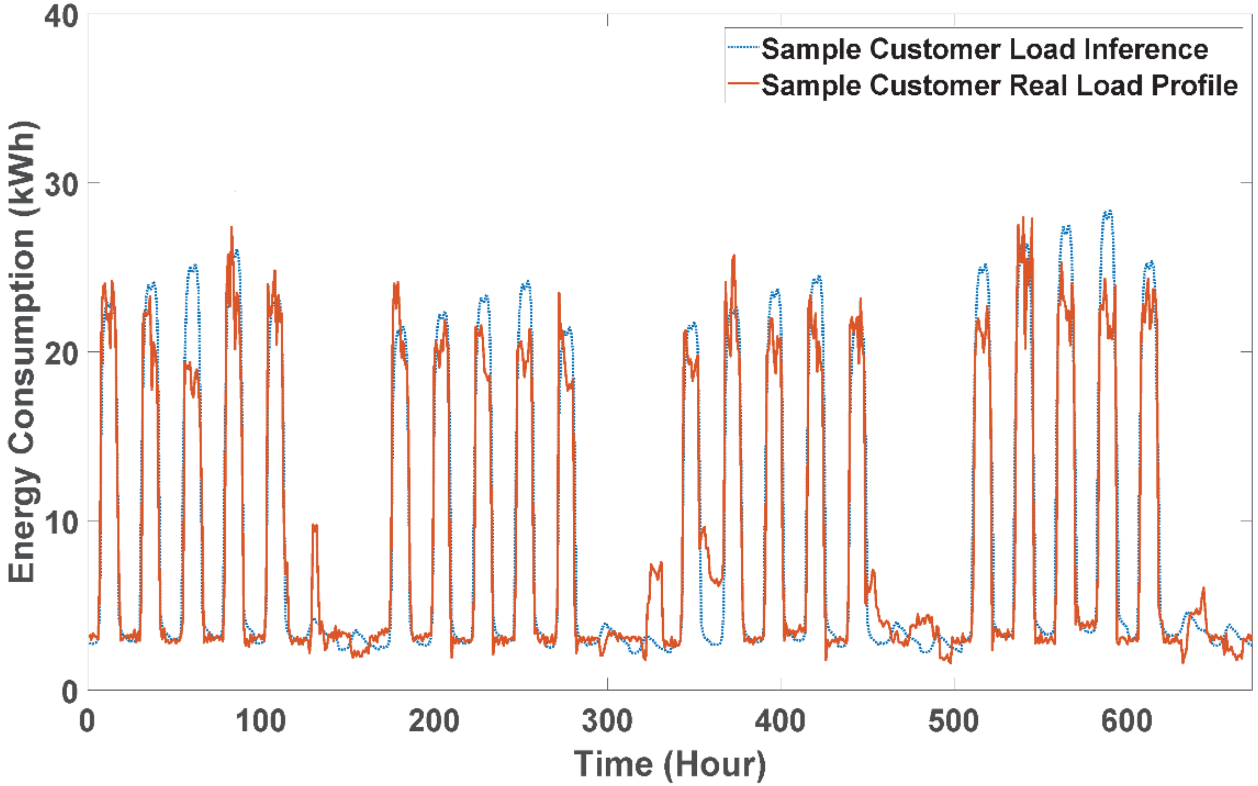


- Basic Idea: Pick the Cluster that has the Best State Estimation Performance for Each Customer
- Methodology: Assign and Update Probability Values to Different Clusters Based on State Estimation Residuals (Recursive Bayesian Learning)
- Outcome: Pick the Most Probable Cluster for Each Unobserved Customer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

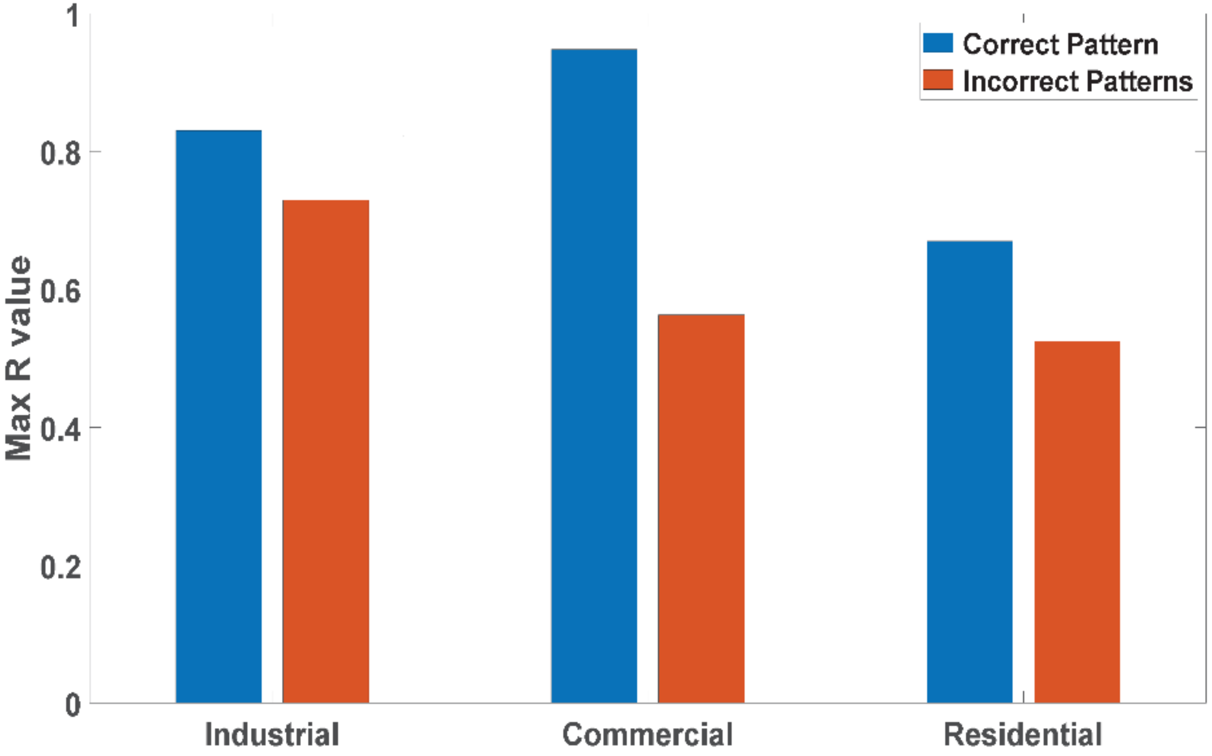
Section I: Overall Structure of the Proposed Solution



Section I: Unobserved Individual Customer Hourly Load and Pattern Inference

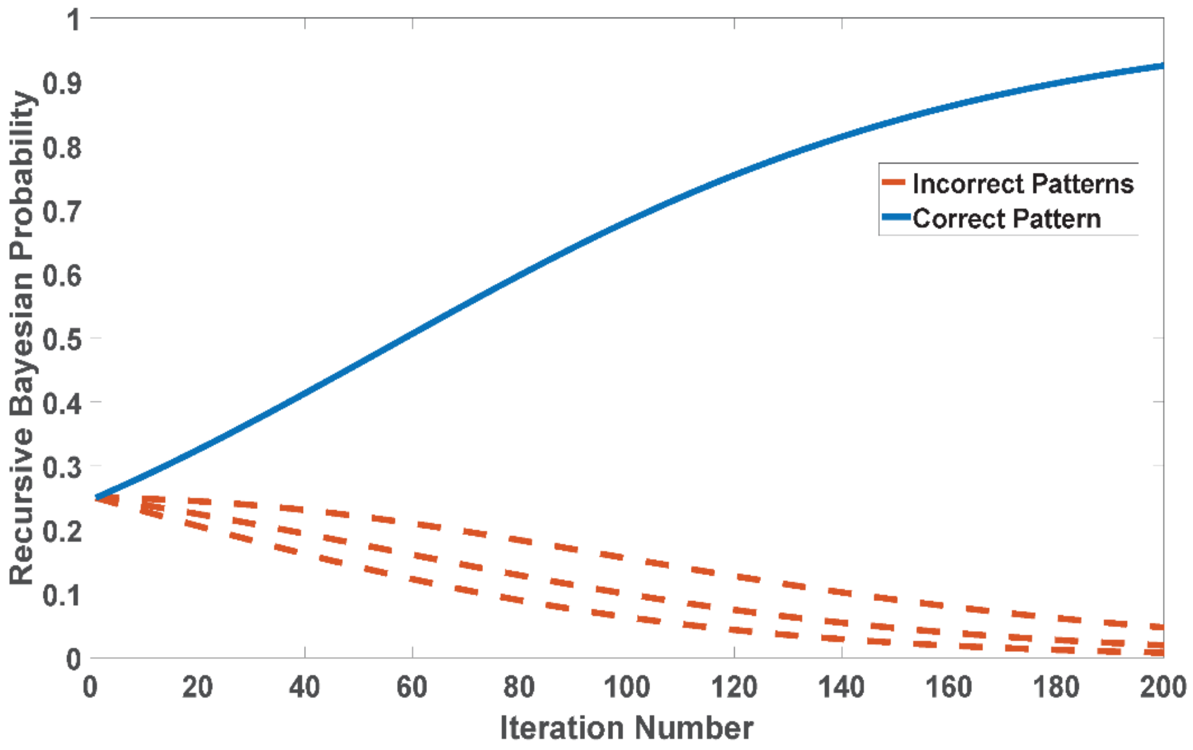


Inferring the hourly demand of an unobserved residential load in one month (average estimation error ≈ 8.5% of total energy)

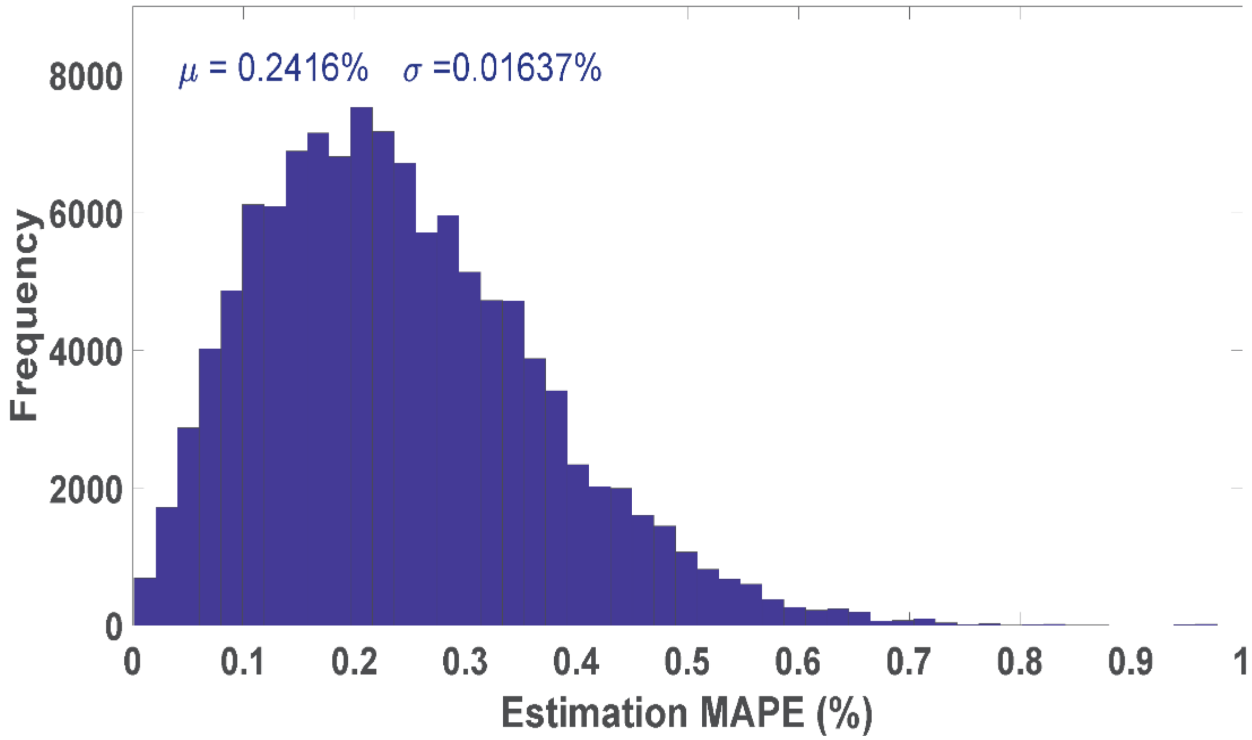


Impact of accurate consumption pattern identification on the accuracy of the inference (industrial load patterns are close and stable)

Section I: Unobserved Individual Customer Pattern Identification Process, State Estimation Performance



Tracking the typical daily consumption pattern of unobserved customers using a Bayesian learning approach



Using inferred load for accurate system monitoring (branch current state estimation)

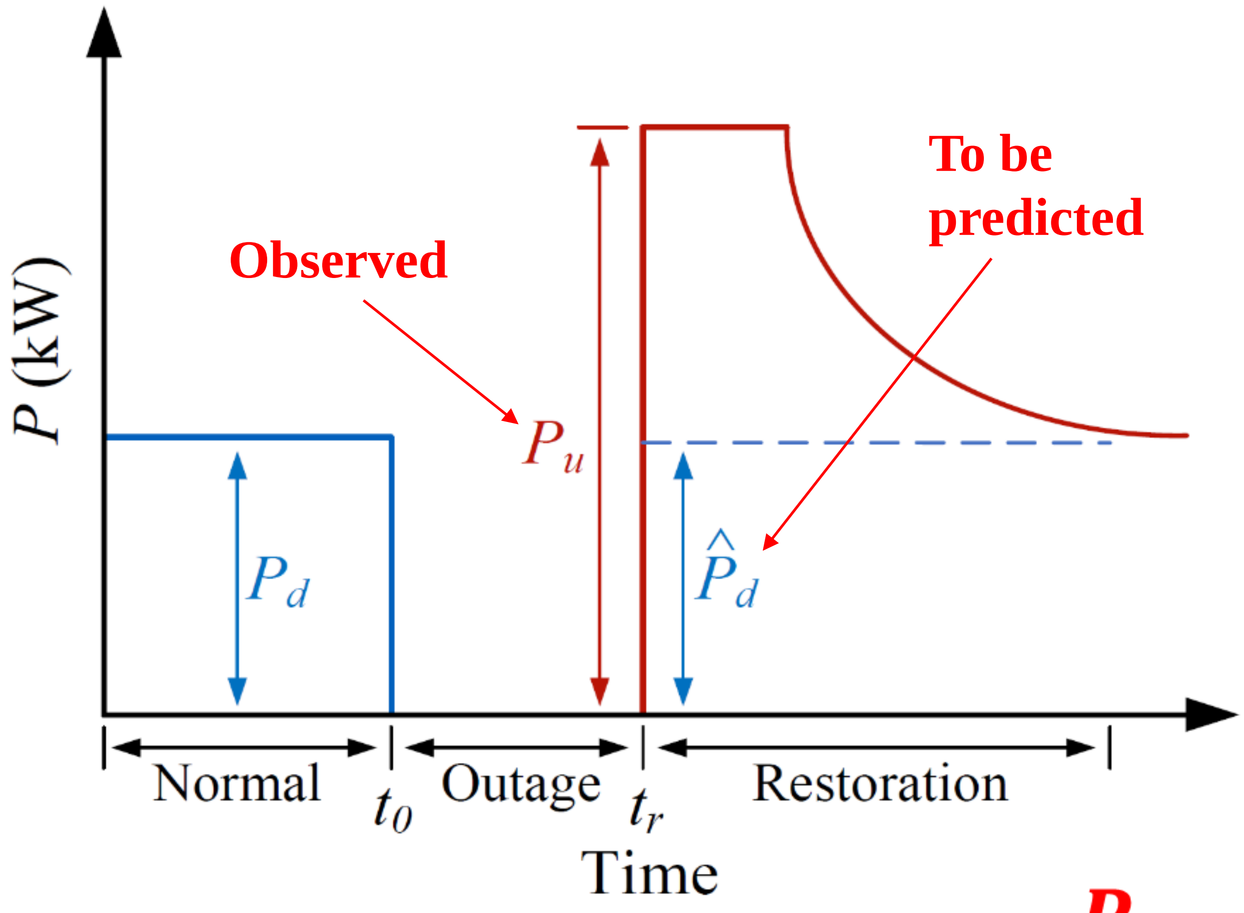
Section II: Assessing Cold Load Pick up Demands Using Smart Meter Data

- Problem Statement: Estimating post-outage cold load pick up (CLPU) demand at feeder-level and customer contribution to CLPU overshoot using smart meter data.
- Challenges:
  - ✓ Customer behavior volatility
  - ✓ Lack of behind-the-meter information on customer thermostatically controlled loads

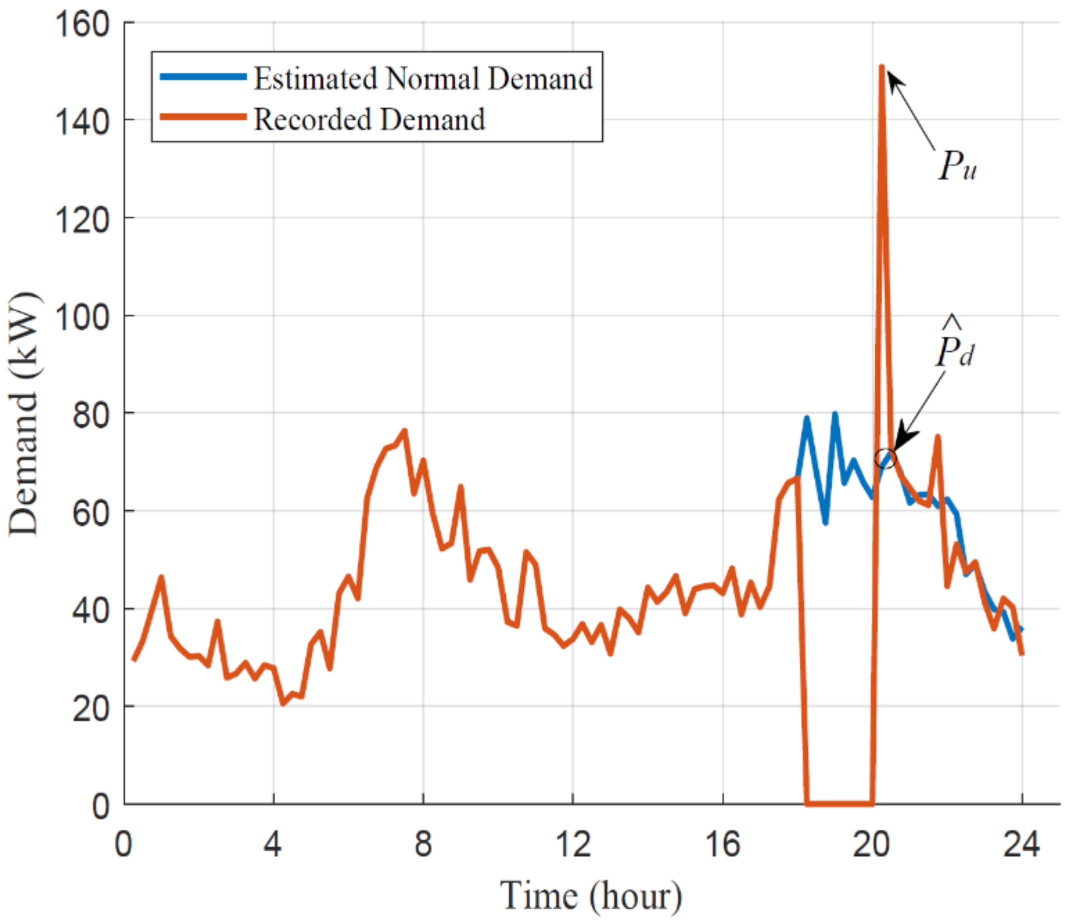
- Solution Strategy: Develop a data-driven “model-free” framework to estimate CLPU demand at both feeder-level and customer-level using only smart meter data

- Proposed Solution Components:
  - ✓ Machine learning-based diversified load predictor at feeder-level
  - ✓ Probabilistic reasoning at customer-level to model behavioral uncertainty

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

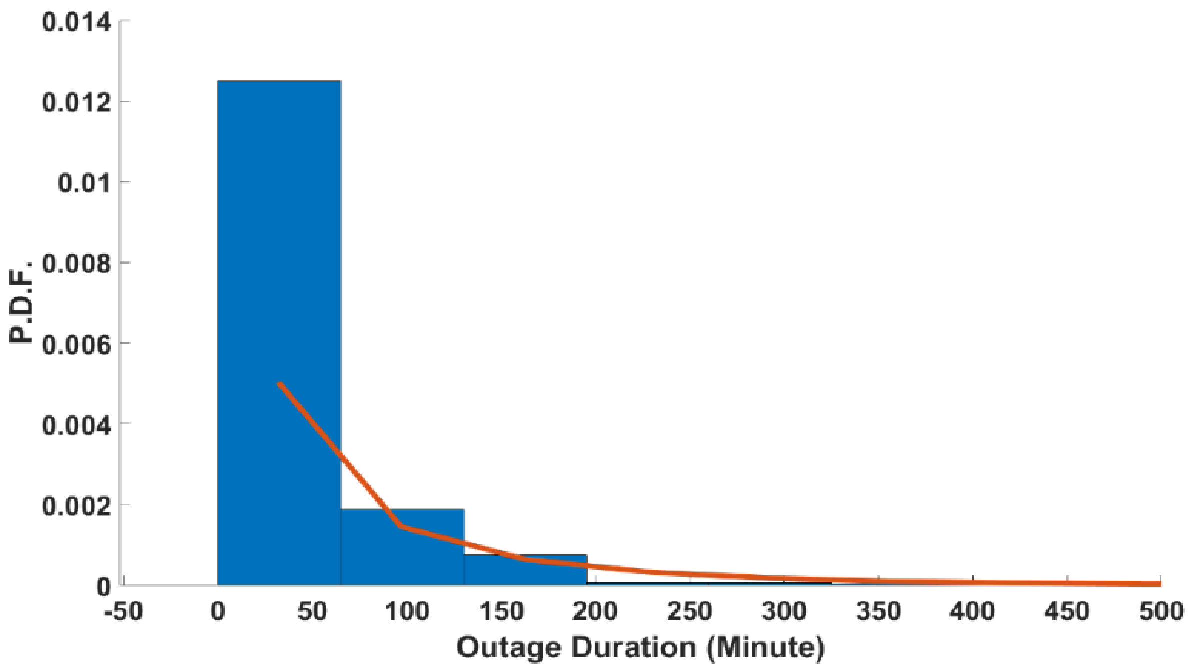


$$CLPU\ Ratio = \frac{P_u}{\hat{P}_d}$$

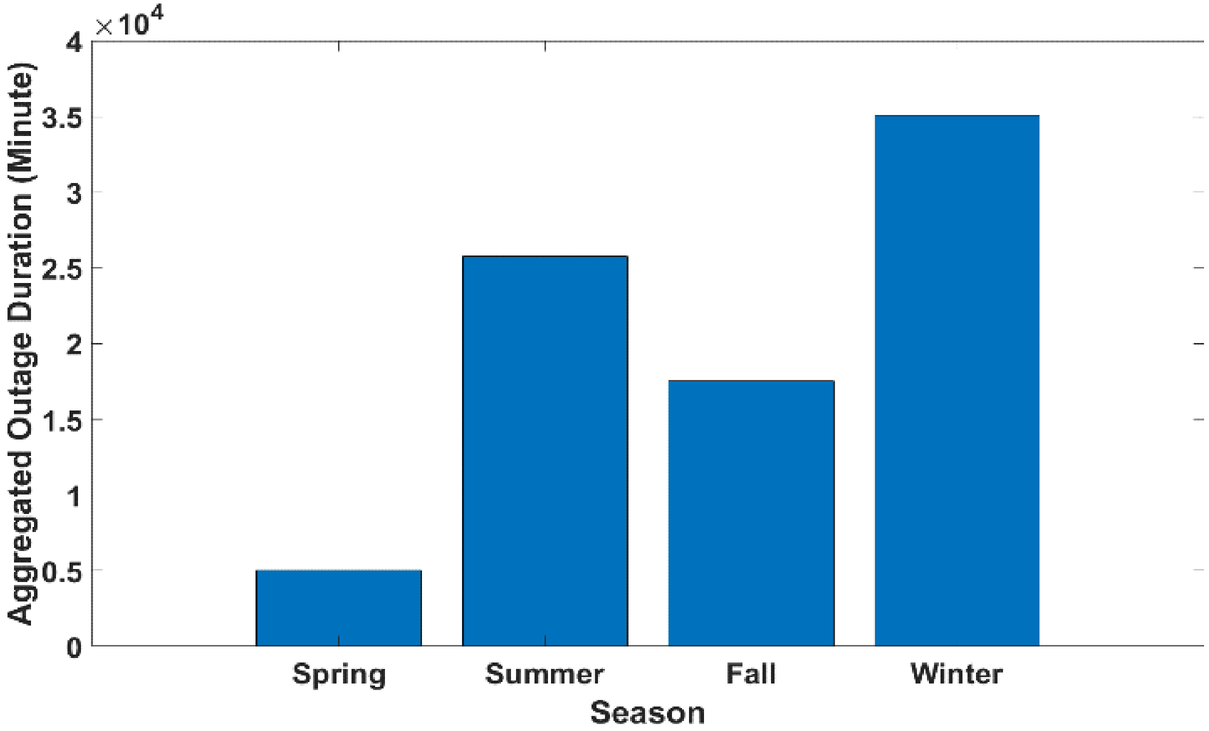


Real CLPU at the Feeder-Level

Section II: Power Outage Statistics Using Smart Meter Data

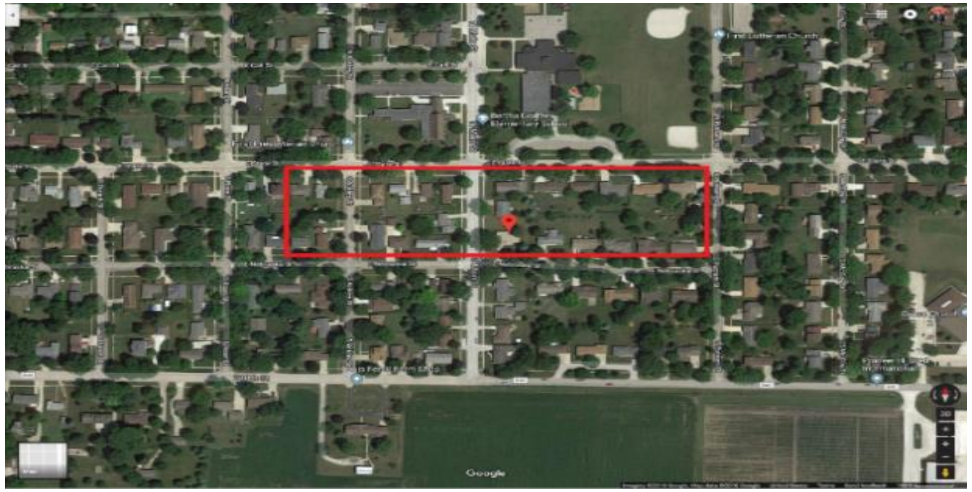


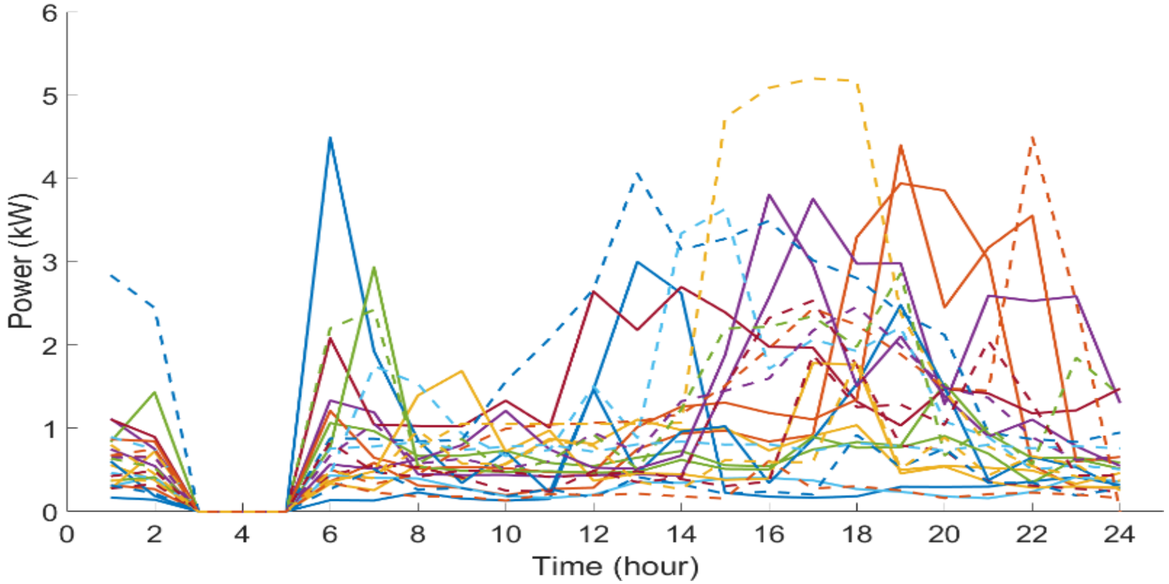
Outage Duration Distribution Follows a Gamma Density Function
(Mean value = 41 minutes)

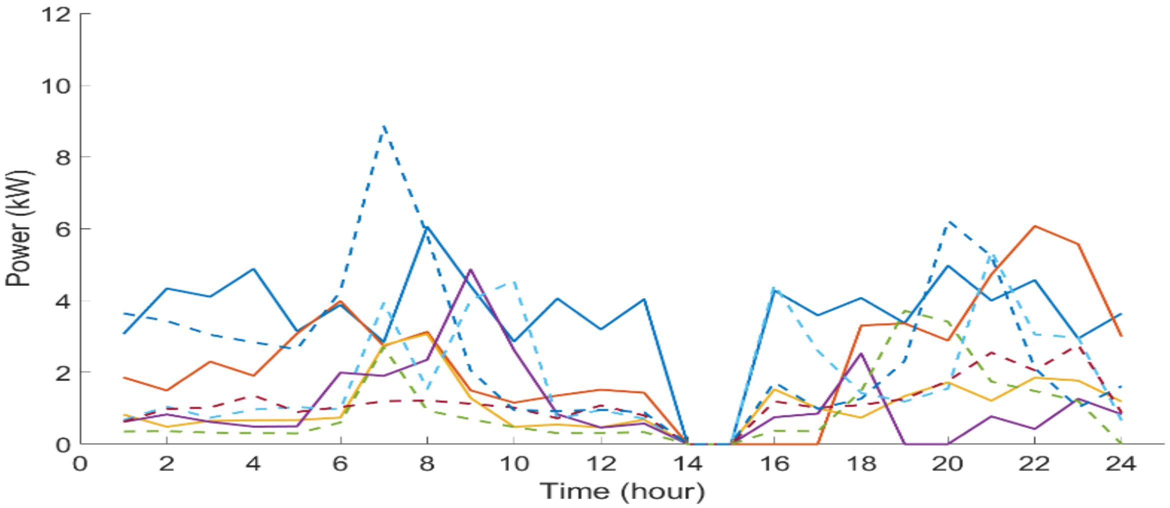


Total Service Lost Time Due to Outages in Different Seasons at a Mid-West Utility

Section II: Impact of Outage on Customer Behavior







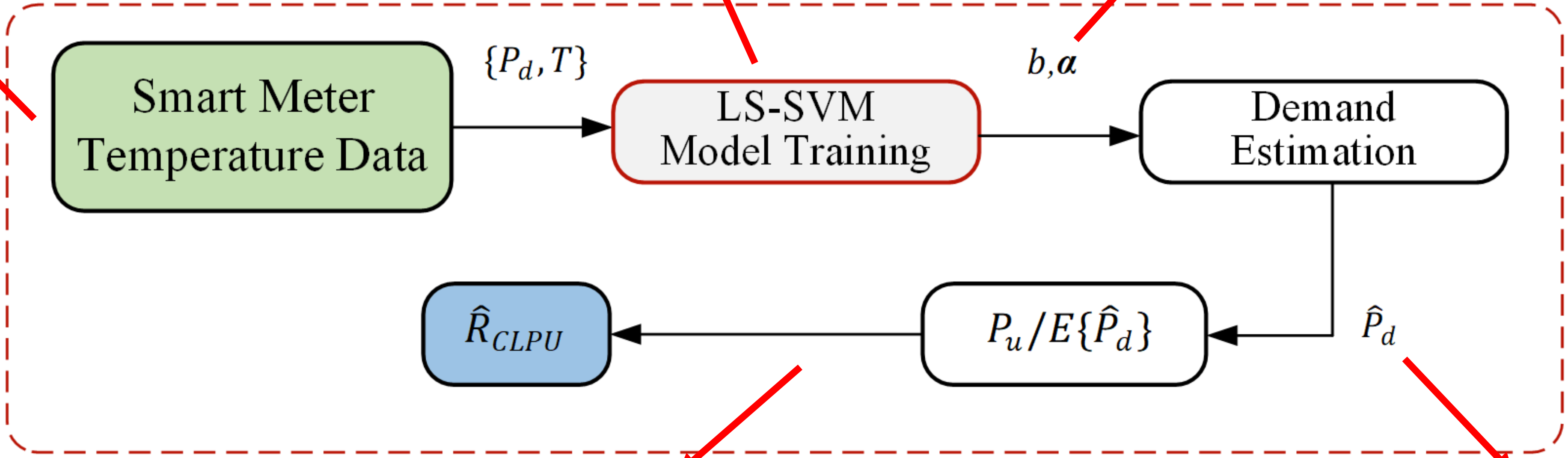
Abnormal Post-Outage Demand Increase: Cold Load Pick-up

Section II: Feeder-Level Data-Driven CLPU Ratio Estimation

Trained to Predict Feeder Load Under Normal Operation

Machine Learning Model Parameters

T – Ambient temperature
P_d – Normal feeder demand

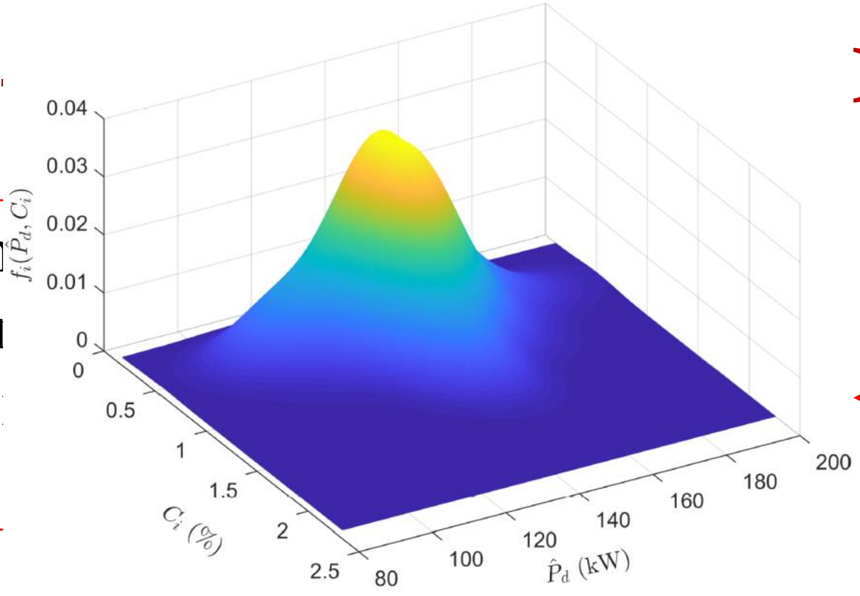


Estimates the CLPU Overshoot (R_{CLPU}) by Dividing the Observed Feeder Demand at Time of Restoration (P_u) by the Estimated Expected Hypothetical Predicted Normal Demand, $E\{P_d\}$

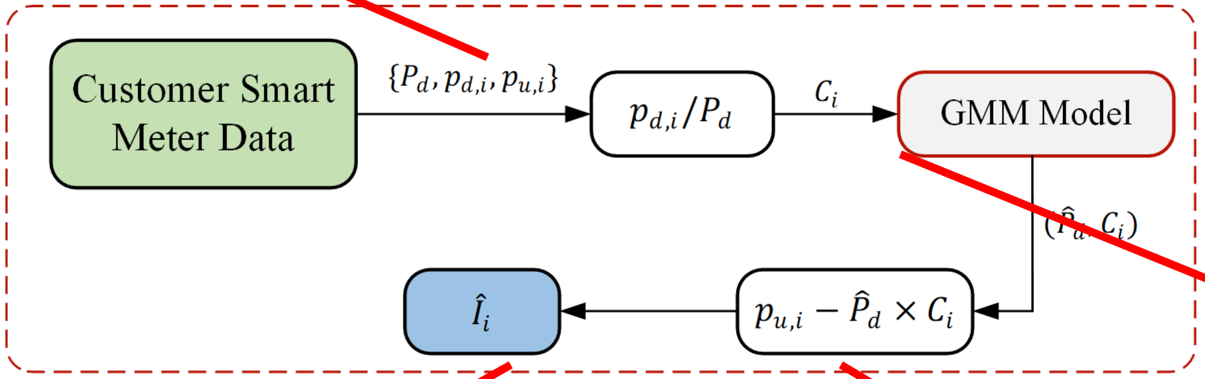
Estimated *Diversified* Demand (what would happen if there was no outage)

Section II: Customer CLPU Estimation

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



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



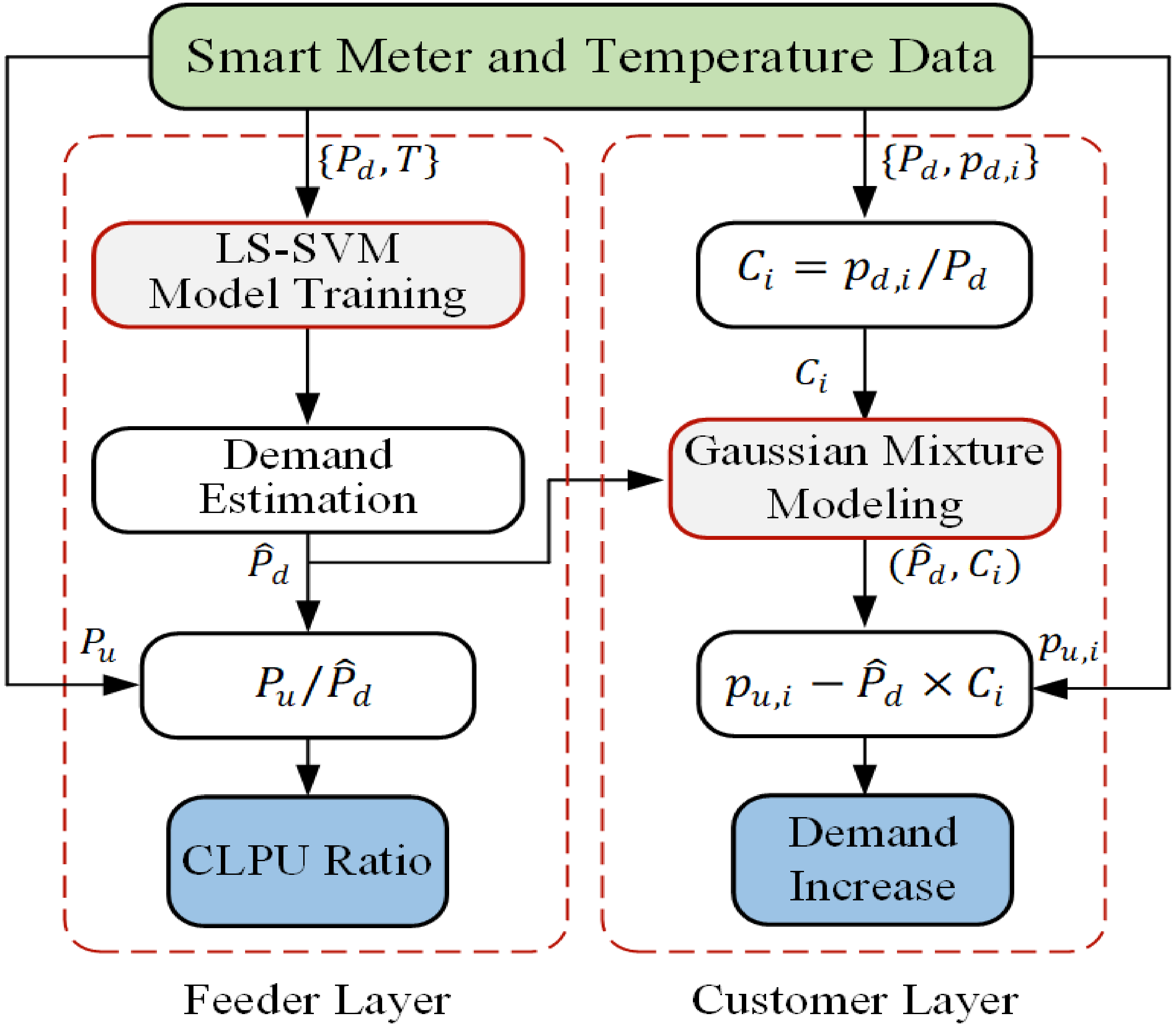
Calculate customer contribution to normal feeder demand (C_i) at different times

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

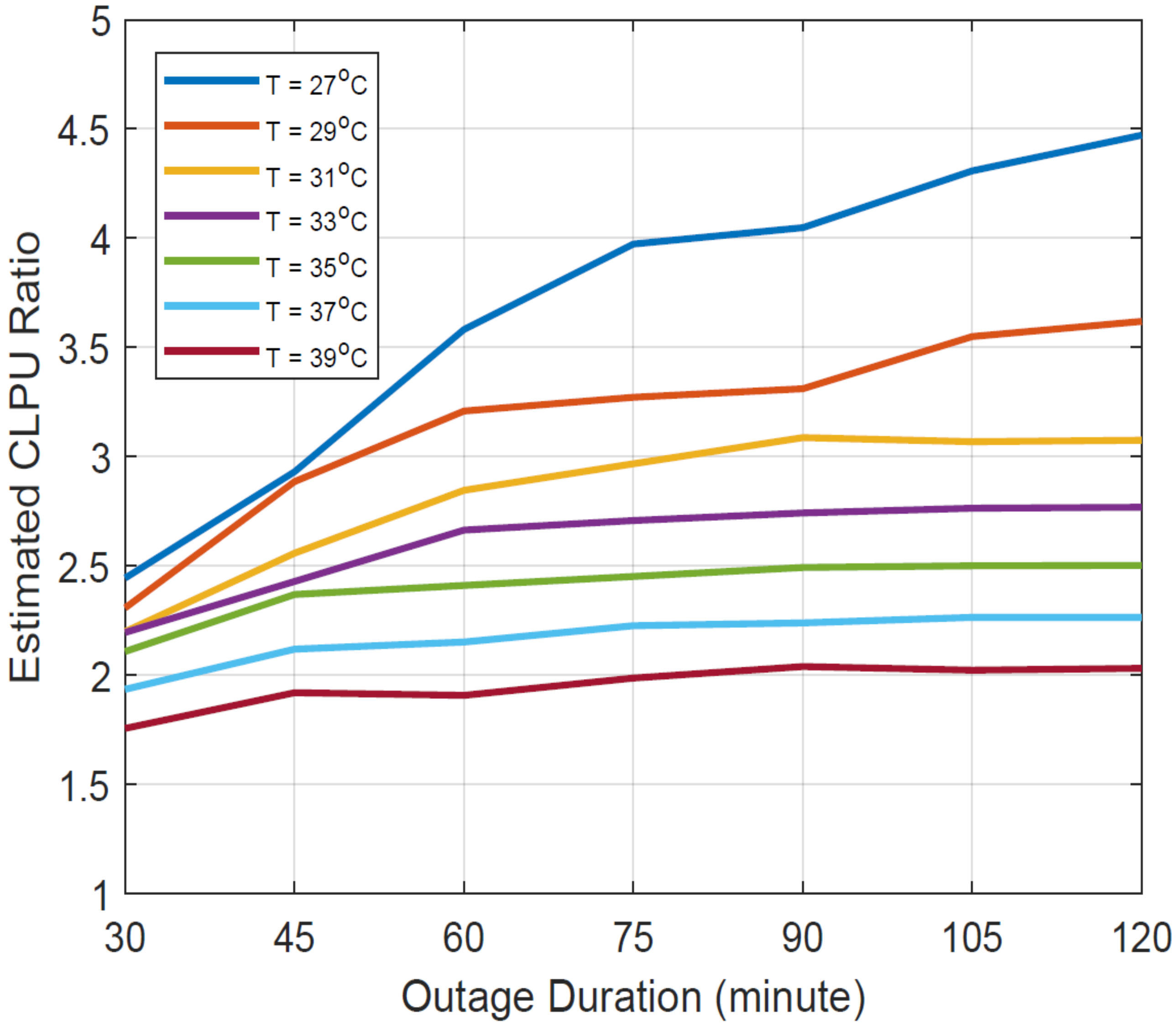
At restoration the learned GMM-based joint distribution of C_i and P_d (quantifying customer's normal behavior) is used to identify customer contribution to CLPU by estimating customer deviation from its expected normal load

Section II: 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**

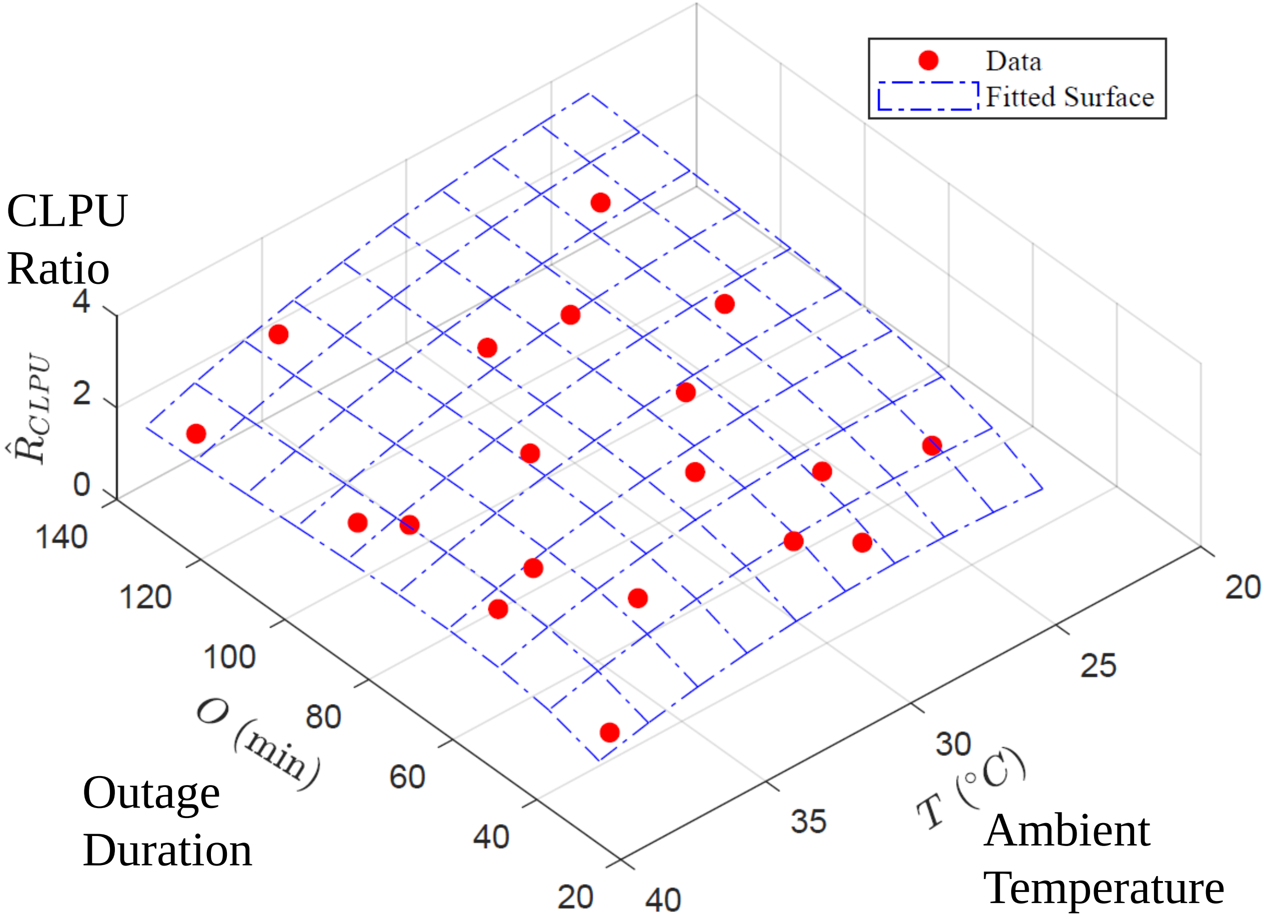


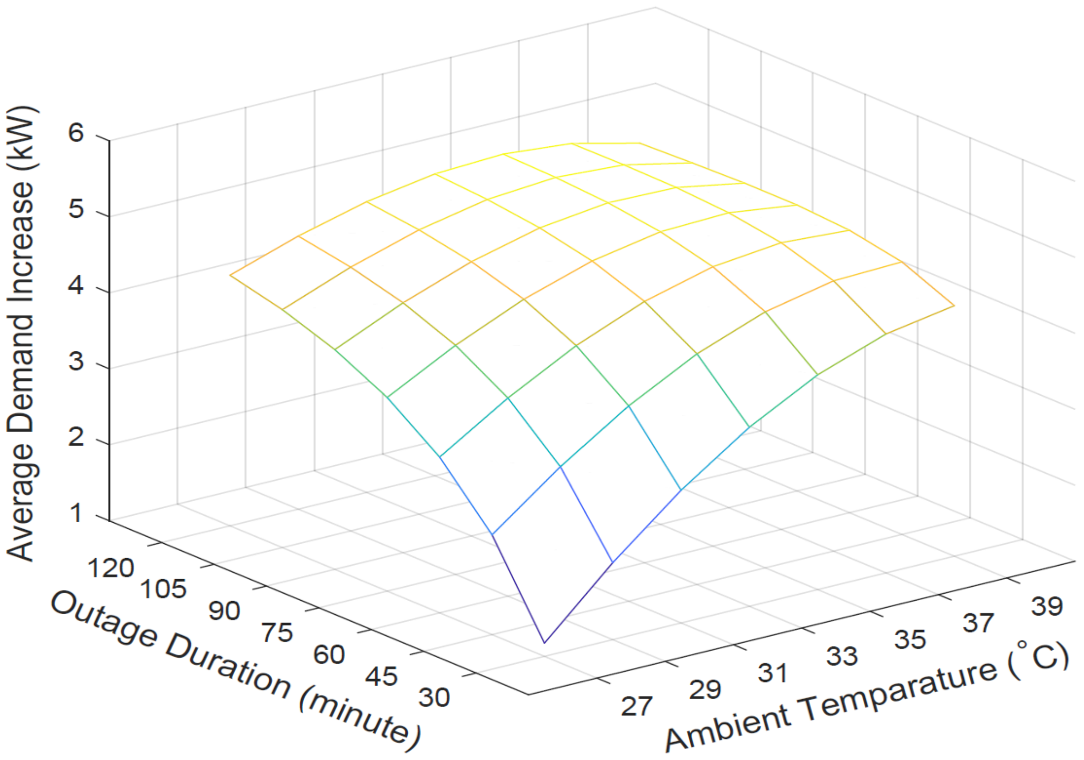
Section II: Feeder-Level CLPU Characteristics



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

Section II: 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

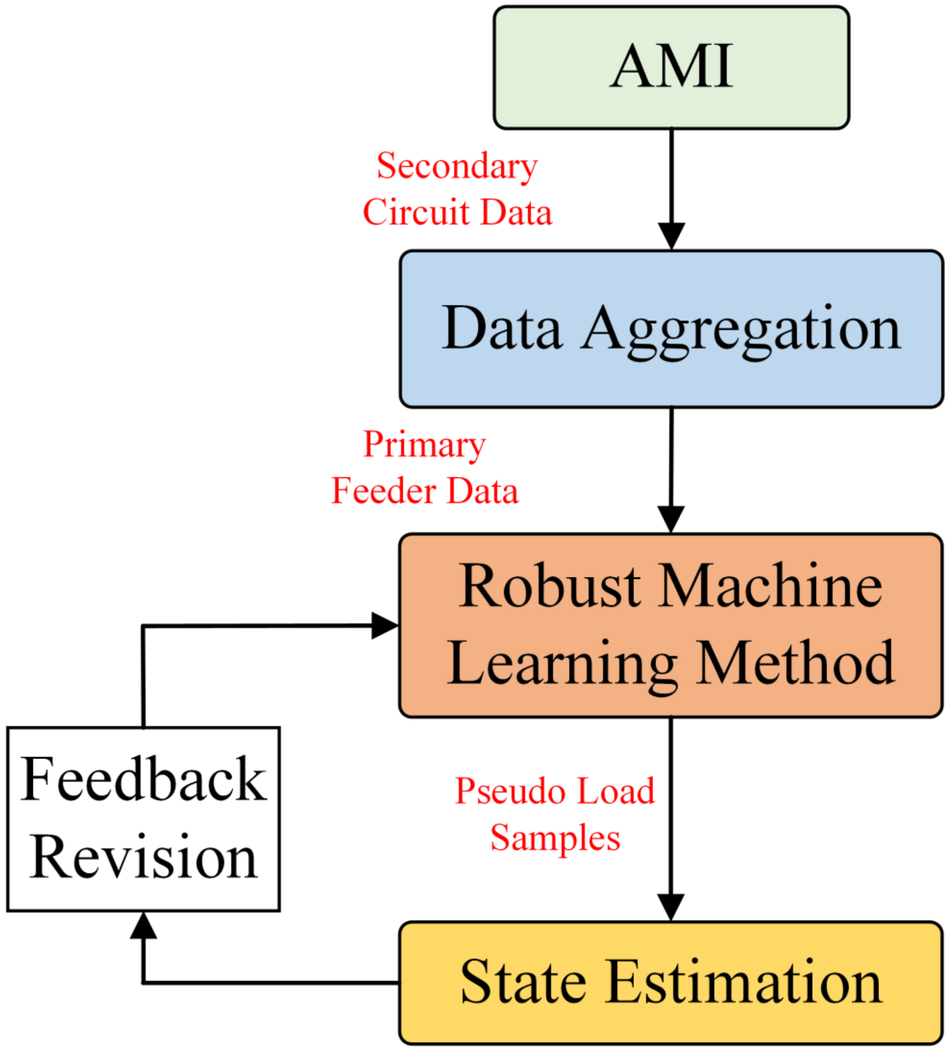
Section III: A Game-Theoretic Data-Driven Approach for Pseudo-Measurement Generation in Distribution System State Estimation

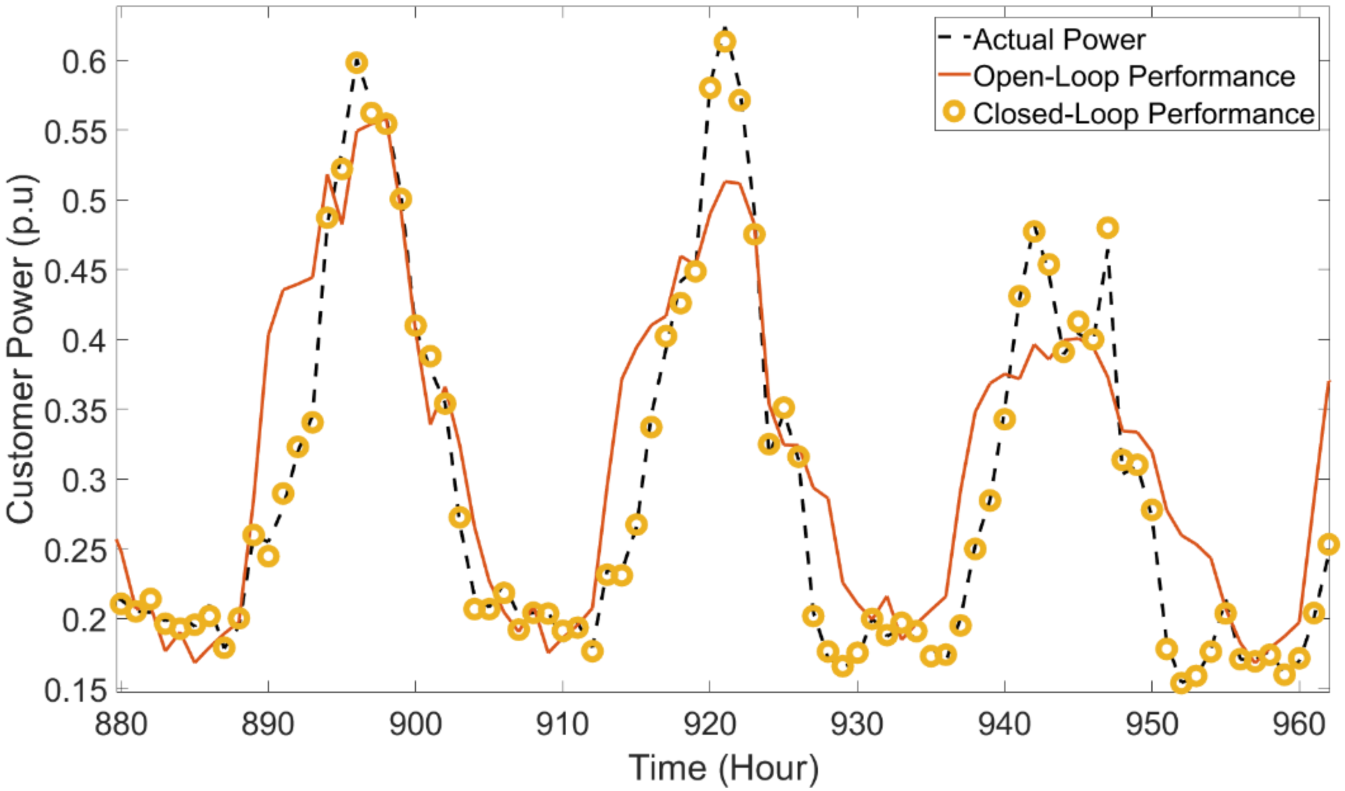
- **Problem Statement**: **A robust closed-loop state estimation method with machine learning components that are trained using real utility data**
- **Challenges:**
  - ✓High computation burden of data-driven approach
  - ✓Unobserved customers’ unknown typical behaviors

- **Solution Strategy**: Take advantage of a branch current state estimator and machine learning technology to further improve the performance of the designed machine learning framework.

- **Proposed Solution:**
  - ✓Game-theoretic expansion of relevance vector machine
  - ✓Using parallel training of multiple machine learning units to exploit the seasonal patterns in load
  - ✓Using a closed-loop information system to improve the accuracy of pseudo measurements

Section III: Solution and Numerical Results





Estimating the Behavior of Unobserved Customers Using Available AMI Dataset

Thank You!
Q & A

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