Iowa State University

ECpE Department

Mining Smart Meter Data for Improving Distribution Grid Operation and Resilience

Zhaoyu Wang

Harpole-Pentair Assistant Professor

Iowa State University

IEEE PES Big Data Tutorial Series

Contents

Contents

Power Distribution Grid Data

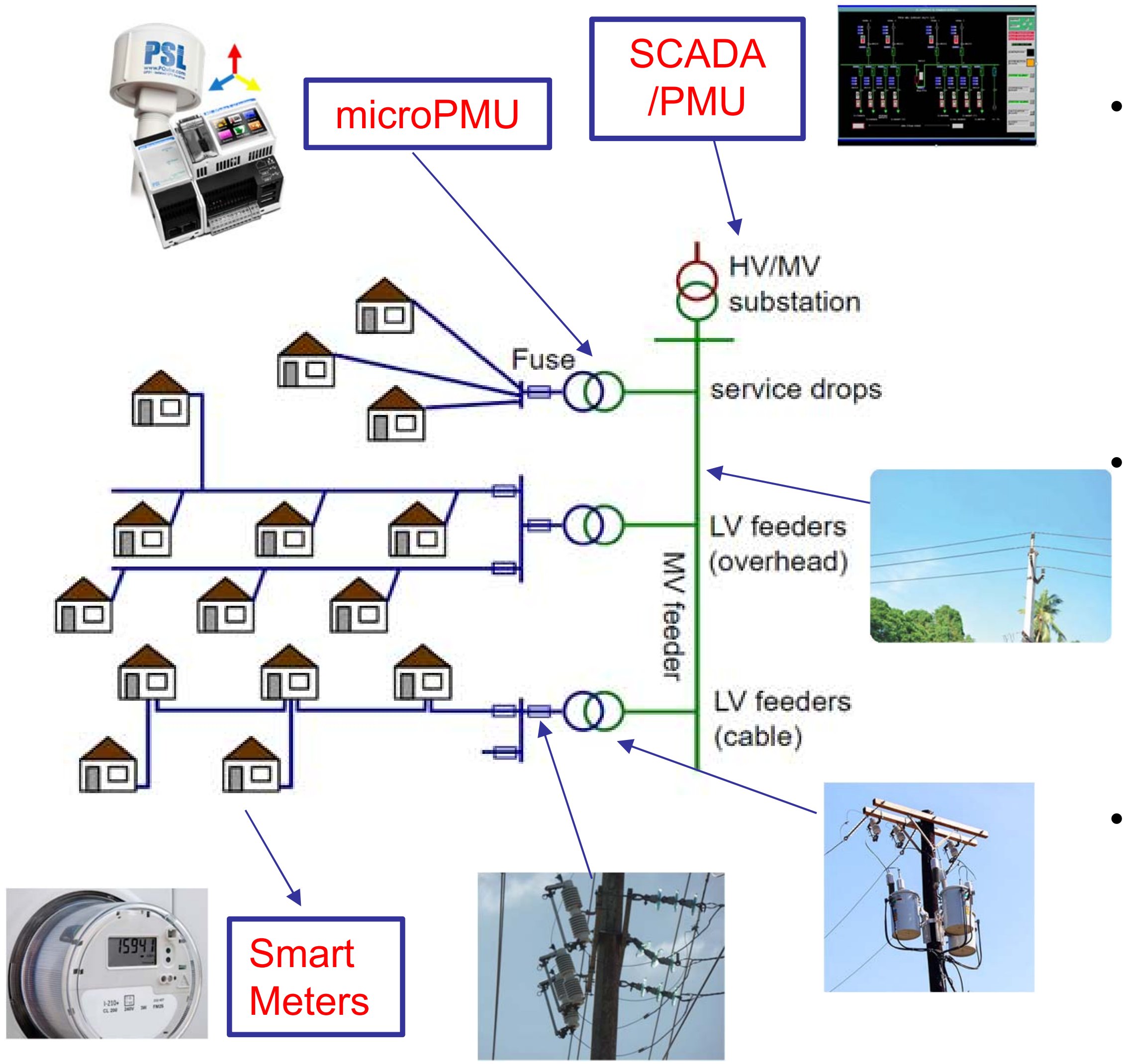


Fig. 1 Power distribution grid.

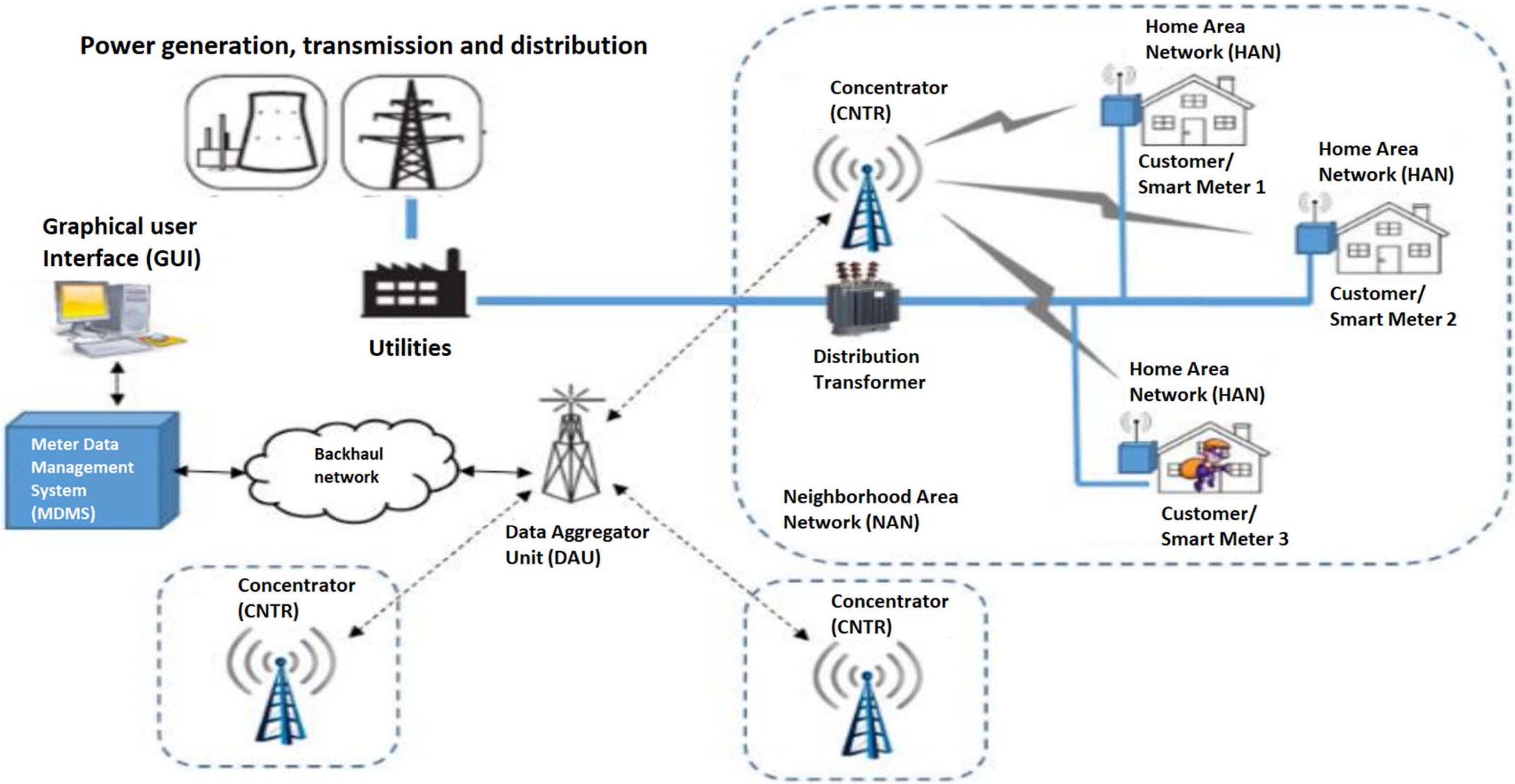
- Where does the data come from?
  - SCADA (supervisory control and data acquisition); Smart Meters; Protection Devices; (micro)PMUs (phasor measurement units)
  - Measures voltage/current/frequency at different resolutions
- What are smart meters?
  - Different from conventional energy meters
  - Stay 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

Our 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, …
- 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	14,000 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
AMR & SCADA	Bloomfield	1,329 customers	18 months	10%	N/A

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"

Exemplary Smart Meter Data from Utilities

- More AMI data and circuit models:

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

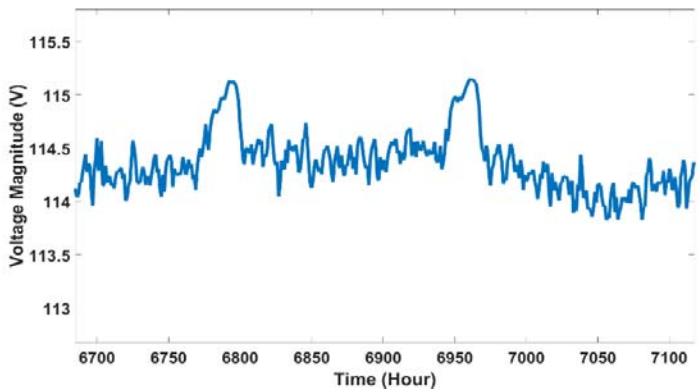
- Duration: 4 years (2014 - 2018)
- Measurement Type: Smart Meters and SCADA
- **Detailed circuit models of all feeders in Milsoft and exact smart meter locations**
- Data Time Resolution: 15 Minutes – 1 Hour
- Customer Type:

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

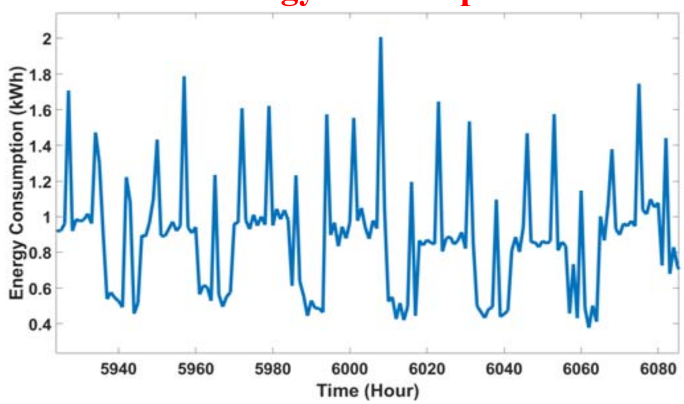
Exemplary Smart Meter Data from Utilities

Smart Meter Measurement Data For Load Monitoring

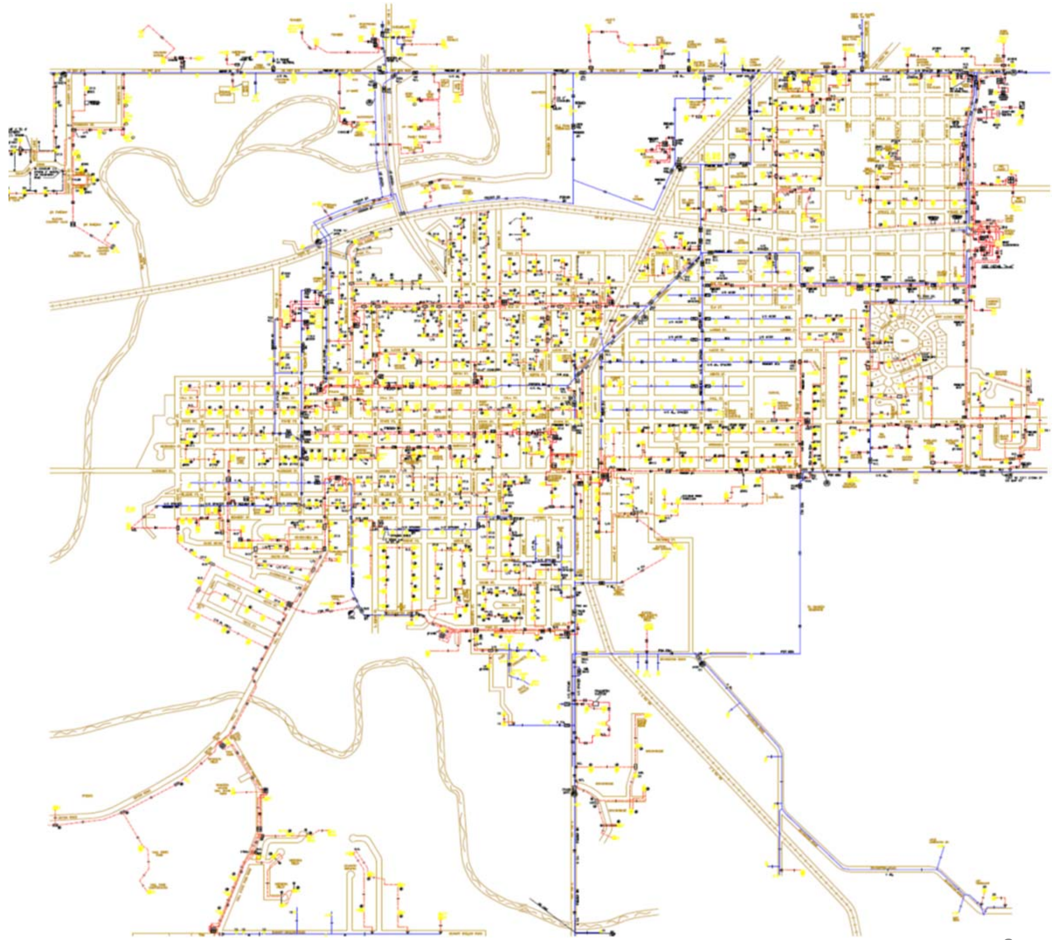
Sample Customer Voltage



Sample Customer Energy Consumption



Network Topology/Model Information



Exemplary Smart Meter 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

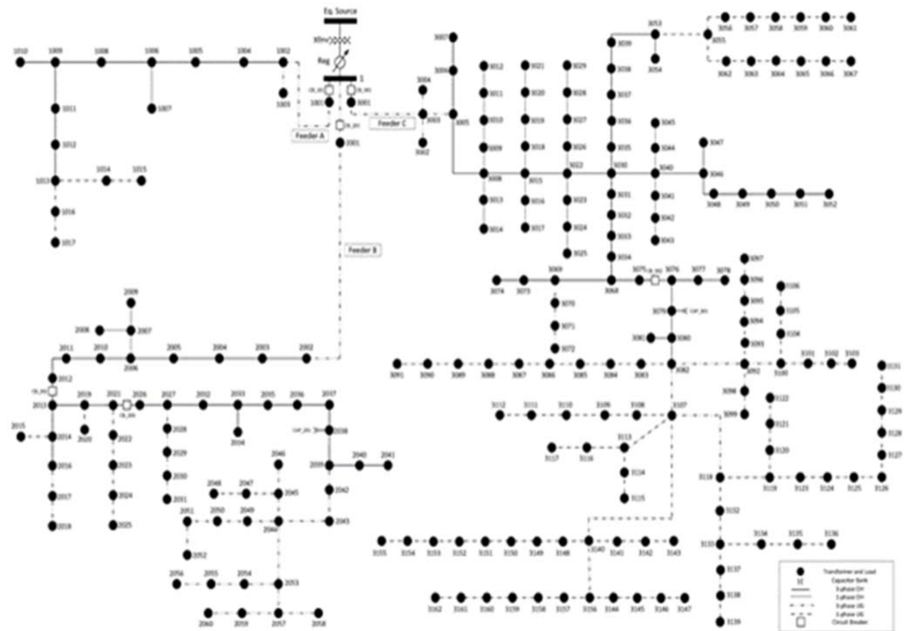
SM Data

- 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

Data Sharing

With permission from our utility partner, we share ***a real distribution grid model with one-year smart meter measurements*** [1]. 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

Smart Meter Data Quality Problem

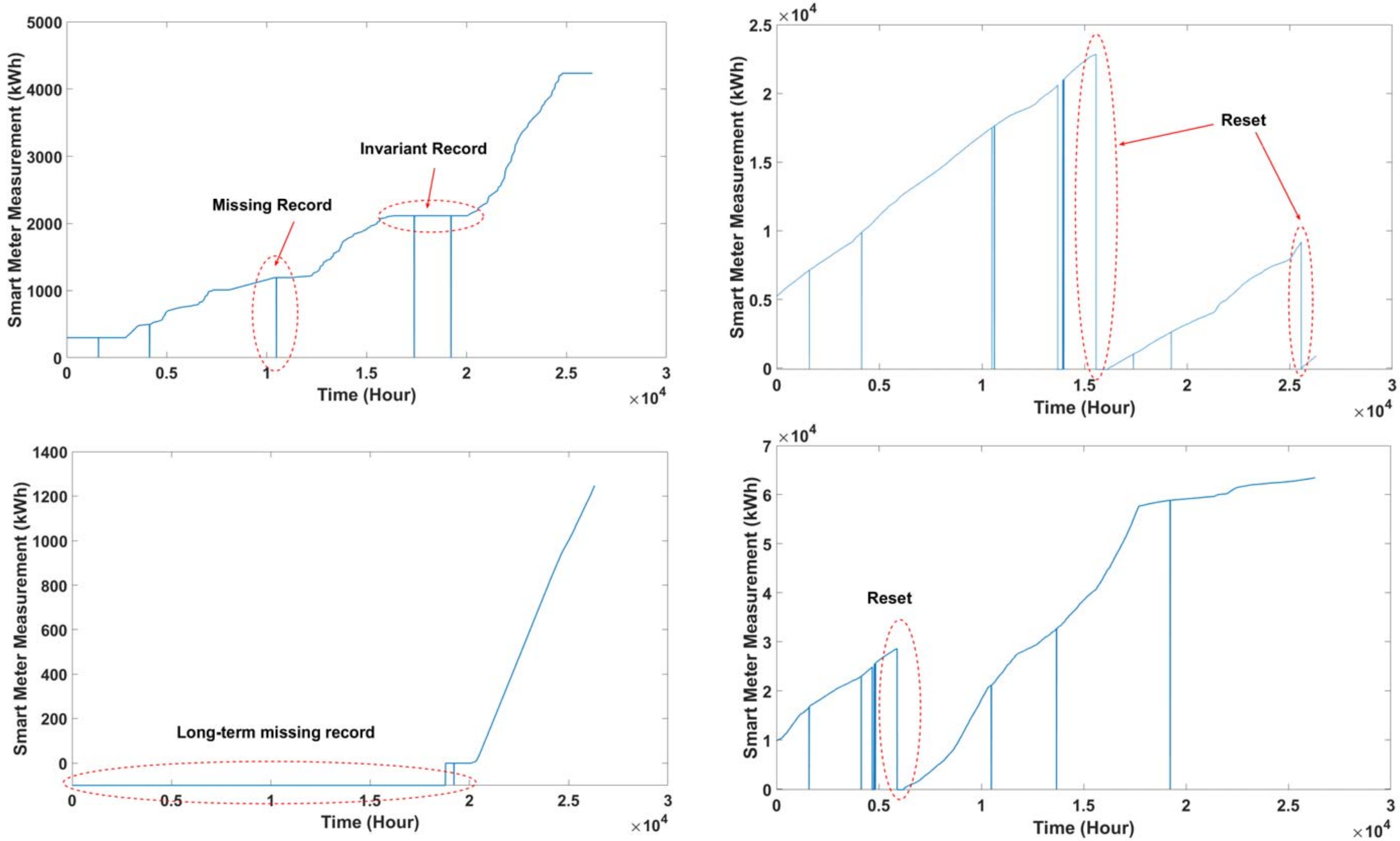
SM data quality problems:

- Outliers/Bad Data
- Missing Data
- Data Reset
- Repetitive Data

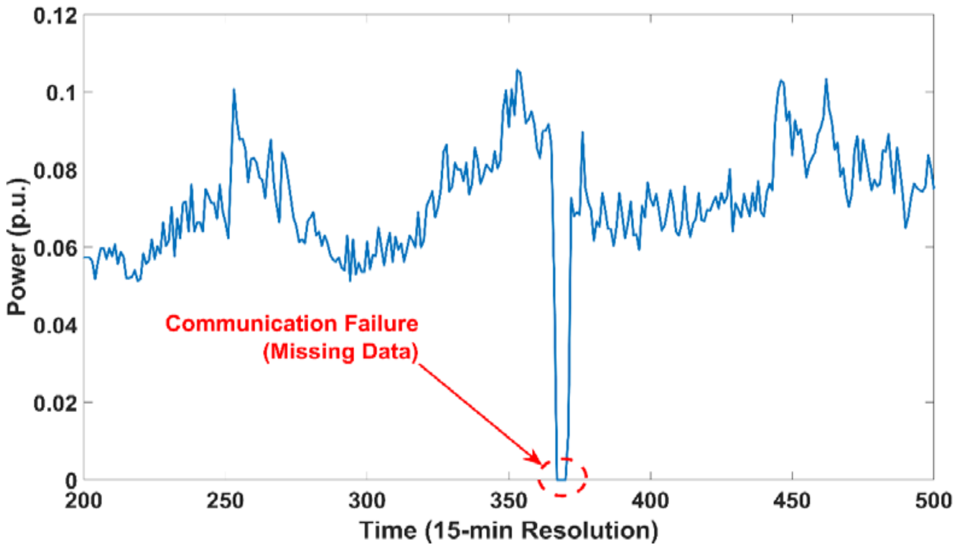
Reasons that cause SM data quality problems:

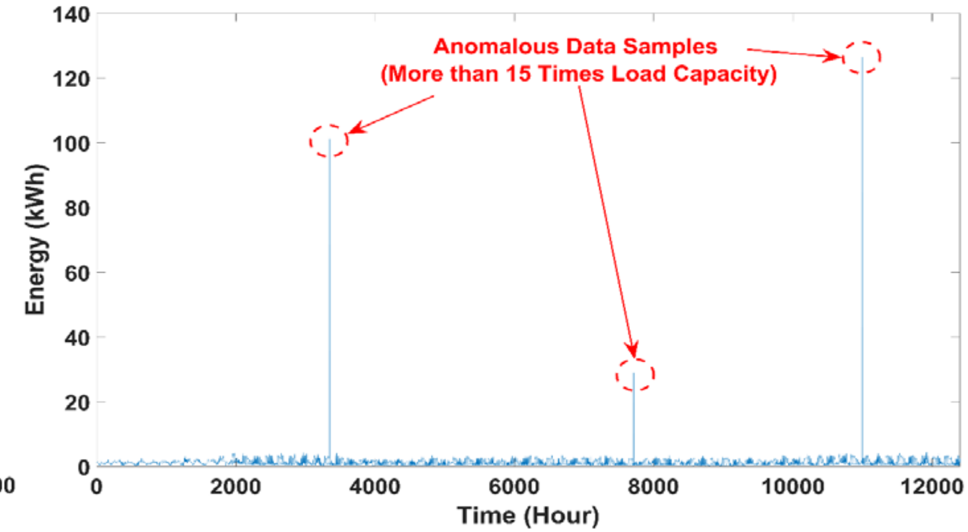
- **Intermittent errors** – large noise or temporary failures due to the communication failure or meter malfunction.
- **Systematic errors** – deterioration of measurements due to age, temperature, weather, and other environmental effects [2].

Smart Meter Data Quality Problem



Smart Meter Data Quality Problem





Smart Meter Data Pre-processing

How to identify and correct the missing/bad data?

Solution for Missing/Bad Data identification:

- Exclude the outage cases, the missing data is identified based on invariant record.

- A statistical outlier detector (e.g., z-score) is used to identify the bad data.

$$z_i = \frac{x_i - \mu}{\sigma}$$

The samples with z-score outside a range of ± 5 are identified as the bad data.

Solution for Missing/Bad Data Correction:

- The discontinuously missing/bad data is filled/corrected by taking an average of the two neighboring measurements.

- The continuously missing/bad data is filled/corrected by taking an average of the 1000 random samples generated from historical demand probability distributions. These probability distributions were developed for each hour using probability density estimators (e.g., kernel density estimation (24 PDFs in total)).

Contents

Smart Meter Data Analytics

How Smart Meter Data Benefits Utility Operations?

- The high-resolution data from SMs provides rich information to model consumption behaviors of consumers.
  - Load profiling
  - Load forecasting
  - Peak contribution estimation
  - Behind-the-metering generation modeling

- The large amount of customer metering data provides an unique opportunity for using data-driven techniques to improve distribution grid real-time monitoring and control.
  - Distribution system state estimation
  - VVC/CVR
  - Outage detection
  - Demand response implementation

Smart Meter Data Analytics

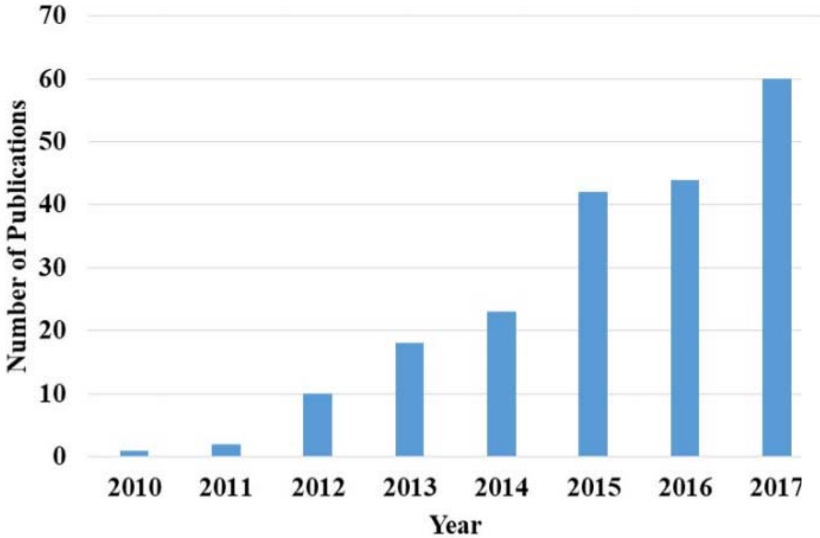


Fig. 2 Number of publication about SM data analytics [3].

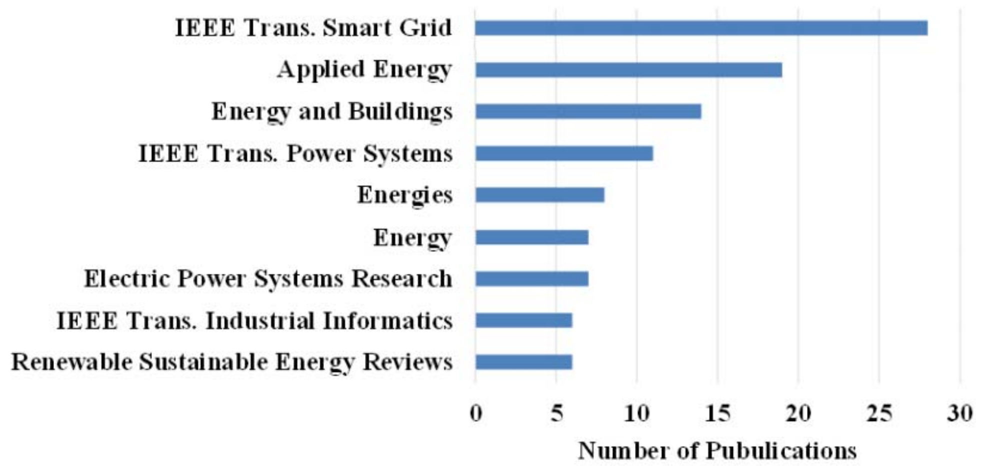


Fig. 3 Number of publication in nine most popular journals in 2017 [3].

- In recent year, researchers have put considerable efforts in SM data analytics.
- The number of publications increases rapidly from 2012.

Smart Meter Data Analytics

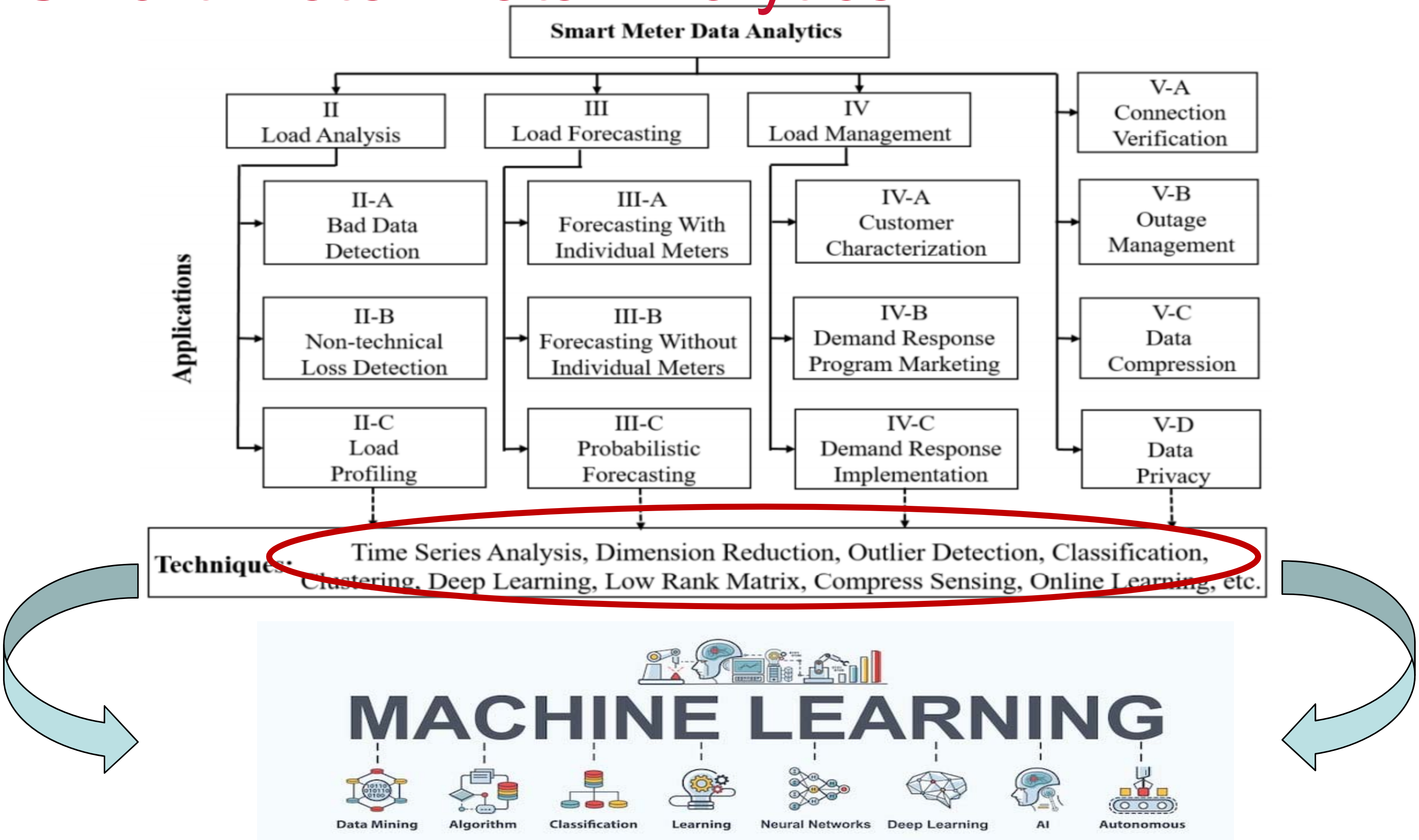


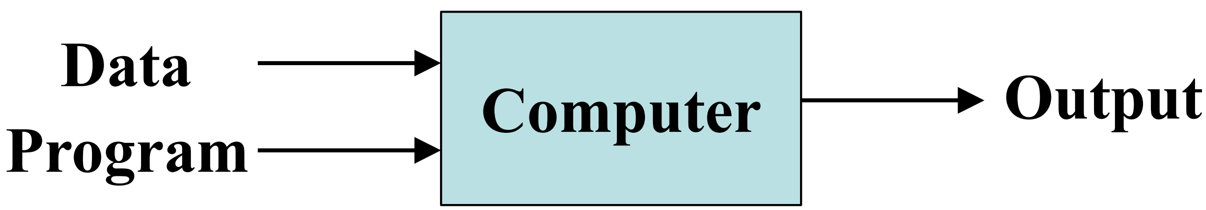
Fig. 4 Taxonomy of SM data analytics [3].

What is Machine Learning?

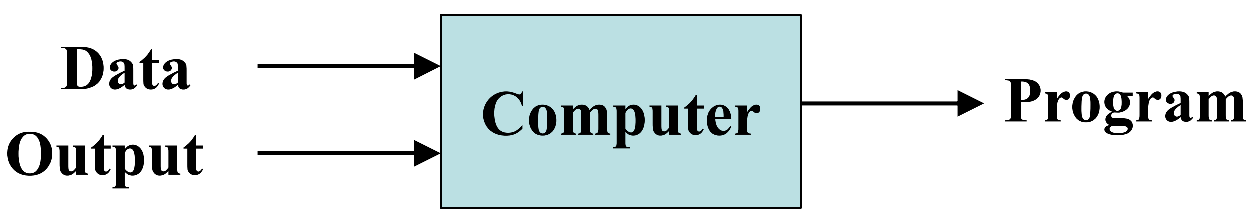
"Learning is any process by which a system improves performance from experience."

- Herbert Simon

Traditional Programming



Machine Learning



Types of Machine Learning

Supervised learning

- Given training data with desired labels
- Classification (categorical labels)
- Regression (real-valued labels)

Unsupervised learning

- Given training data without desired labels
- Data clustering
- Independent component analysis
- Probability distribution estimator
- Dimension reduction

Reinforcement learning

- Given a sequence of states and actions with rewar
- Output is a policy
- Policy is a mapping from states to actions
- Decision making

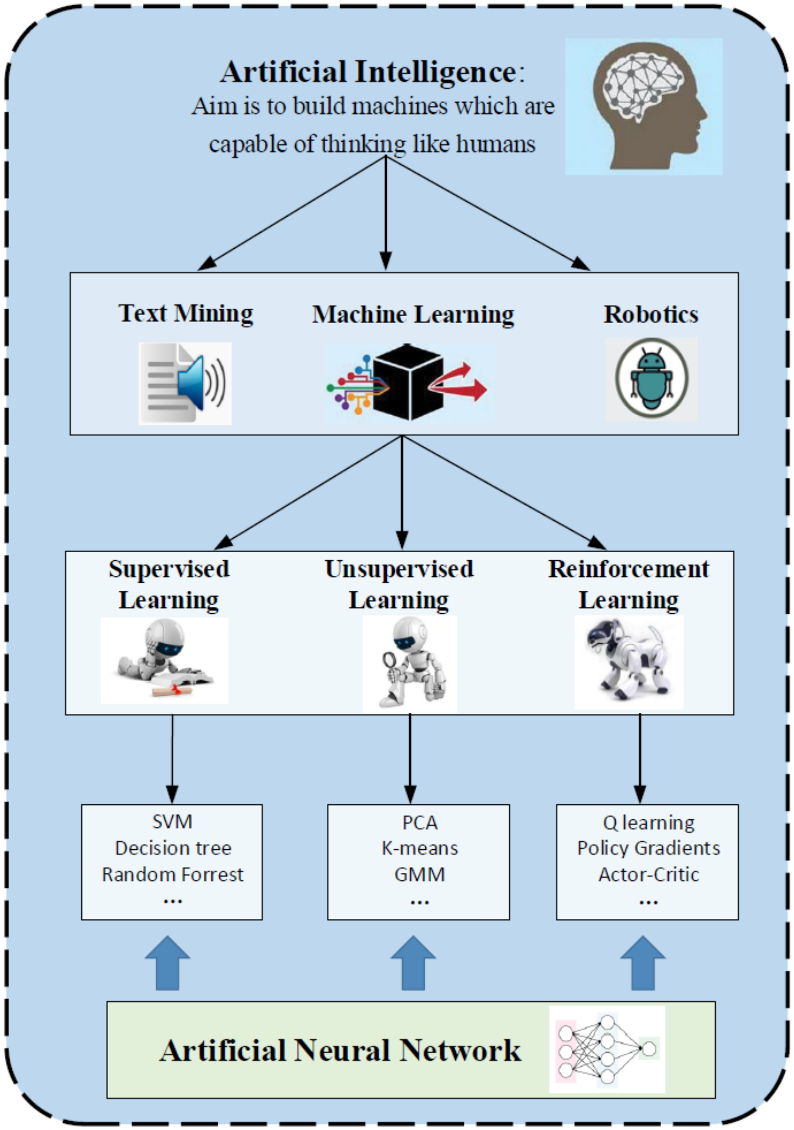
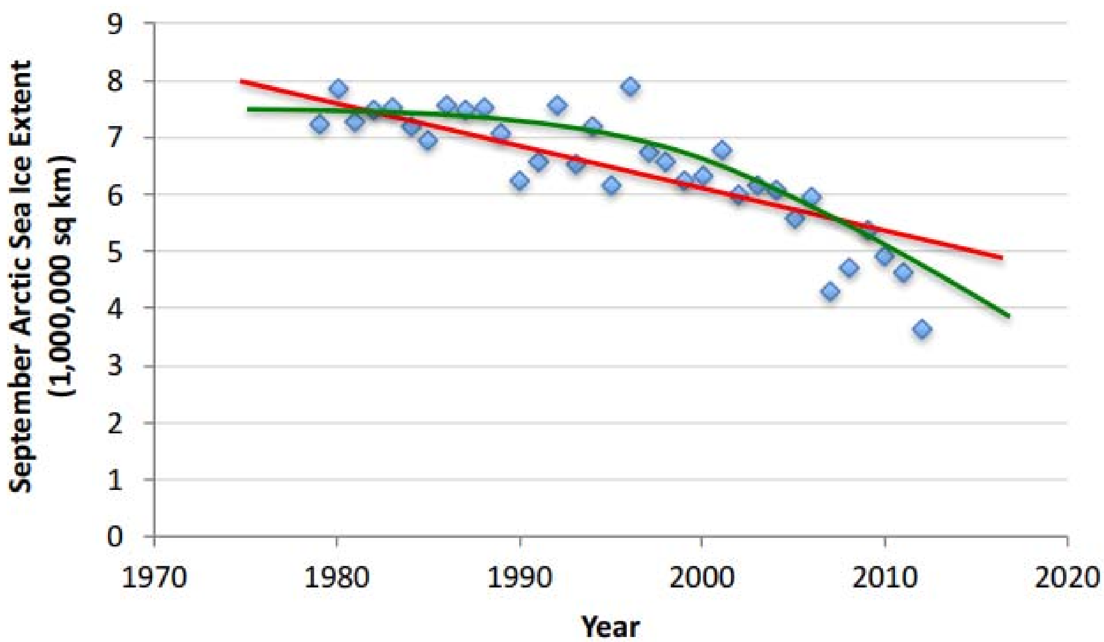


Fig. 5 Artificial intelligence methods.

Supervised Learning: Regression

Regression Model:

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
    - When y is real-valued, this learning process is called regression.



Based on slide by Eric Eaton

Supervised Learning: Classification

Classification Model:

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
    - When y is categorical, this learning process is called classification.

Tiger

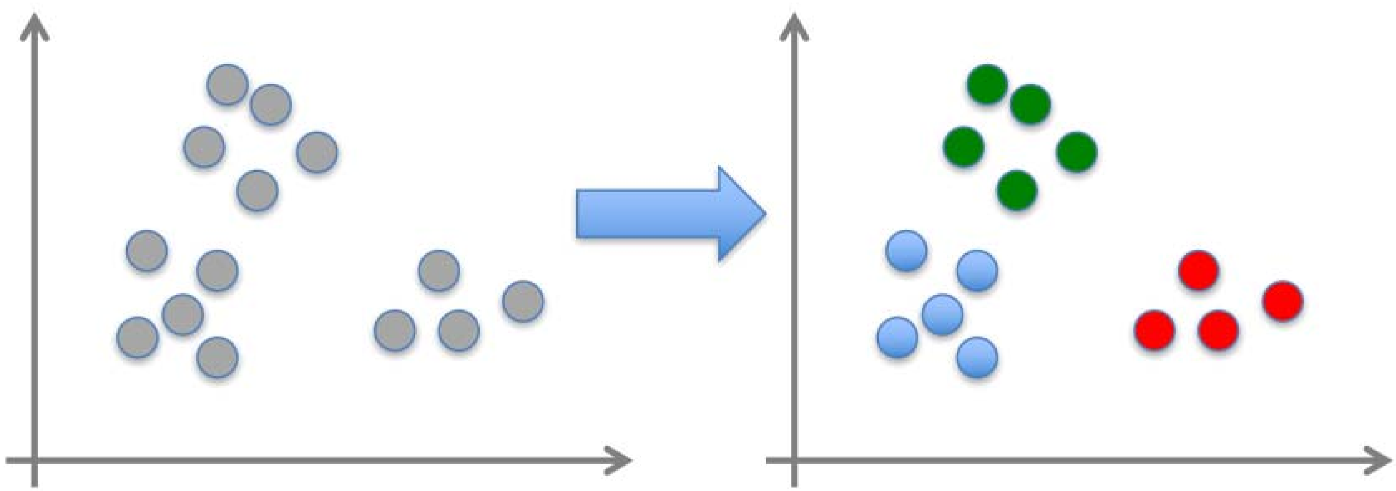
Cat

Based on slide by Eric Eaton

Unsupervised Learning: Clustering

Clustering Model:

- Given $x_1, x_2, \ldots, x_n$
- Represent a form of data summarization
- Discover hidden structures and patterns behind the x's
- Group data with similar patterns together



Based on slide by Eric Eaton

Unsupervised Learning: Clustering

- ***Probabilistic techniques***
  - Identify the cluster with a ***certain probabilistic model*** whose unknown parameters have to be found.
  - Assume the data were generated from a mixture of k probability distributions.
  - One advantage of probabilistic-based models is that they can be easily generalized to different types of data.
    - For numerical data, we may use a Gaussian mixture model.
    - For categorical data, we may use a Bernoulli model.
    - For sequence data, we may use a hidden Markov model.

Unsupervised Learning: Clustering

- ***Distance-based techniques***
  - Optimize global criteria based on the distance between data.
  - Distance-based algorithms can be generally divided into two types:
    - Flat: e.g., k-Means, k-Medians.
    - Hierarchical: e.g., agglomerative, divisive.
  - Distance-based algorithms can be used with almost any data type when an appropriate distance function is created for that data type.
  - For ***high-dimensional data***, the quality of the distance functions reduces because of many irrelevant dimensions [5].

Unsupervised Learning: Clustering

- ***Density-based techniques***
  - Optimize local criteria based on density distribution of data.
  - Classical density-based clustering methods: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Statistical Information Grid (STING)
  - One advantage is that these methods can handle ***outliers*** within the dataset.
  - Hard to use in a discrete or non-Euclidean space and the density computations becomes significantly difficult to define with greater dimensionality.

Reinforcement learning

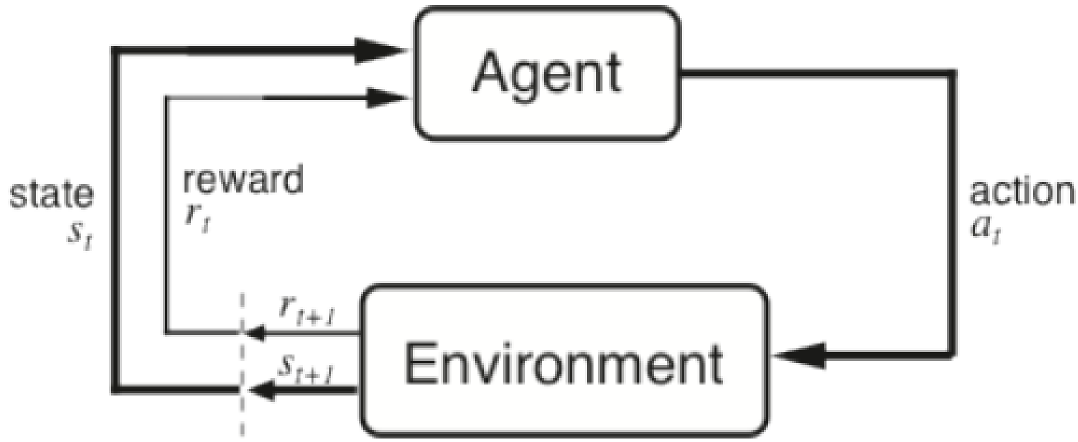


Fig. 6 Artificial intelligence methods [6].

- Agent and environment interact at discrete time step: $t = 0,1 \ldots K$
- Agent observes state at step t: $s_t \in S$
- Produces action at step t: $a_t \in A(s_t)$
- Gets resulting reward: $r_{t+1} \in R$
- Gets resulting next state: s_{t+1}

Based on slide by Sutton & Barto

Reinforcement learning

Objective: Win the game!

State: Position of all pieces

Action: Where to put the next piece down

Reward: 1 if win at the end of the game, 0 otherwise

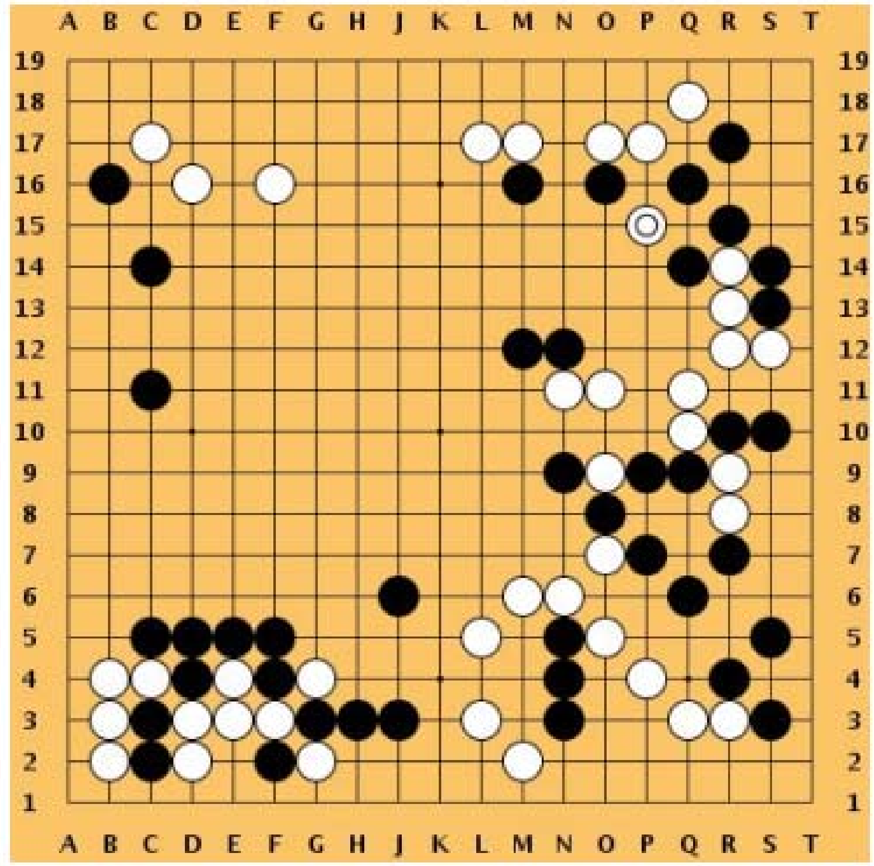
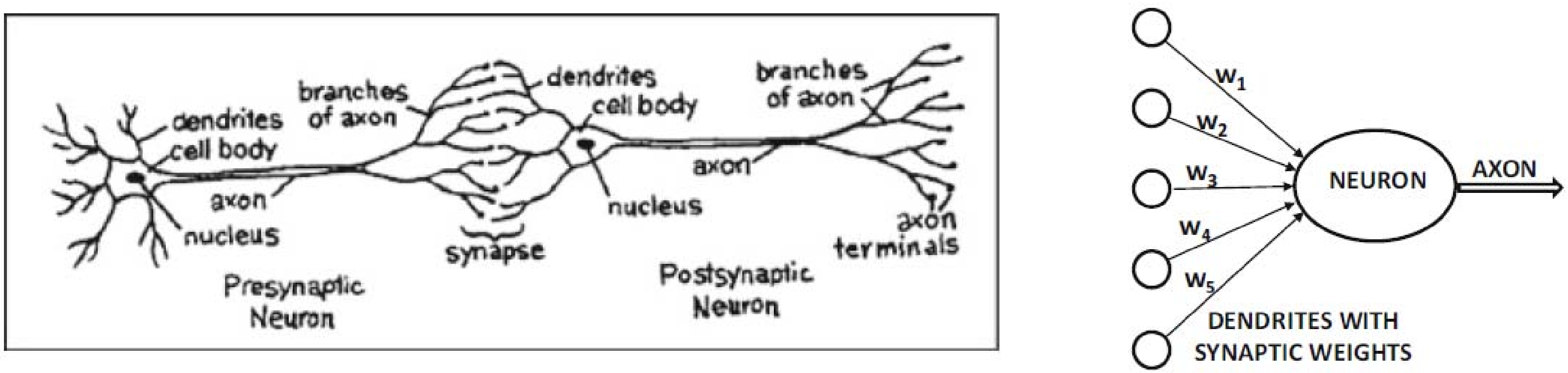


Fig. 7 The game of Go.

Based on slide by Fei-Fei Li & Justin Johnson & Serena Yeung

One Commonly-used Machine Learning technique: Artificial Neural Network



(a) Biological neural network (b) Artificial neural network

Fig. 8: The synaptic connections between neurons. The image in (a) is from "*The Brain: Understanding Neurobiology Through the Study of Addiction* [7]. The image in (b) is from "*Neural Networks And Deep Learning* [8]."

- Artificial neural network (ANN) is one of the most important machine learning techniques that simulates the learning mechanism of biological neural network.
- Computation units in ANN are defined as neurons which are connected to one another through weights.
- An ANN computes a function of the inputs by propagating the computed values from the input neurons to the output neurons and using the weight as intermediate parameters [8].

One Commonly-used Machine Learning technique: Artificial Neural Network

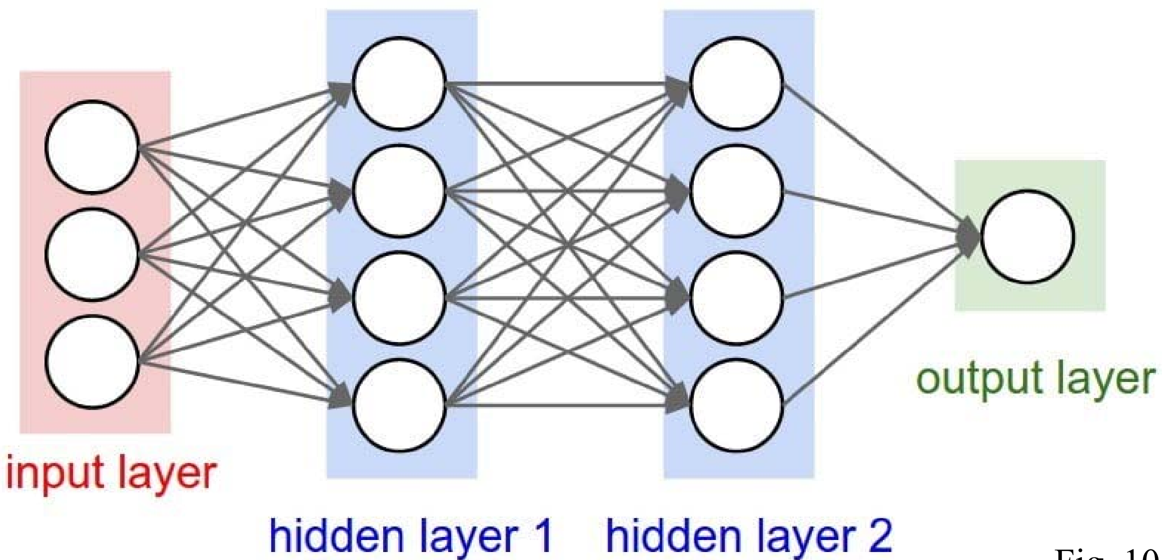


Fig. 9 Multi-layer ANN structure.

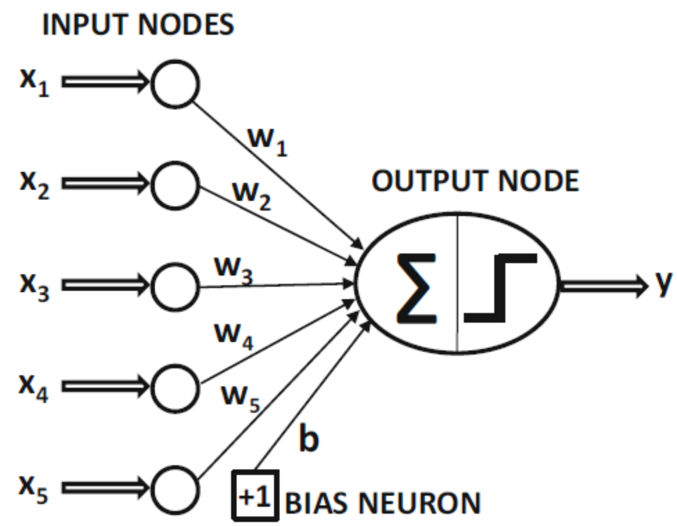


Fig. 10 Single layer ANN structure (perceptron) [8].

Input to Hidden Layer ⟹ $\bar{h}_1 = \sigma(W_1^T \bar{x} + b_1)$

Hidden to Hidden Layer ⟹ $\bar{h}_{n+1} = \sigma\left(W_{n+1}^T \bar{h}_n + b_{n+1}\right) \forall n \in \{1, \dots, m-1\}$

Hidden to Output Layer ⟹ $\bar{o} = \sigma(W_{m+1}^T \bar{h}_m + b_{m+1})$

where, σ is the nonlinear activation function (i.e., sigmoid, hyperbolic tangent, or rectified linear unit.)

One Commonly-used Machine Learning technique: Artificial Neural Network

How can we determine the hyperparameters of ANNs (i.e., number of layers, the number of neurons, and the types of activation.) ?

- Manually or automatically select optimum hyperparameter combinations by using high computation complexity strategies, such as grid search and random search [9].

How can we ensure the generalization ability of ANNs ?

- In order to ensure the generalization ability of ANNs, different strategies are developed to prevent overfitting problem which occurs when the model shows low bias but high variance, such as early stopping, regularization, cross-validation, dropout, etc.

ANNs can be utilized to perform the supervised learning (regression/classification), the unsupervised learning (clustering, generative model), and the reinforcement learning (deep reinforcement learning).

A deep neural network (DNN) is an ANN with multiple layers between the input and output layer.

Contents

Case Study I – A Multi-Timescale Data-Driven Approach to Enhance Distribution System Observability

Case Study I: Background

Distribution system state estimation (DSSE) is the process of inferring the values of distribution system's state variables using a limited number of measured data at certain locations in the system [10].

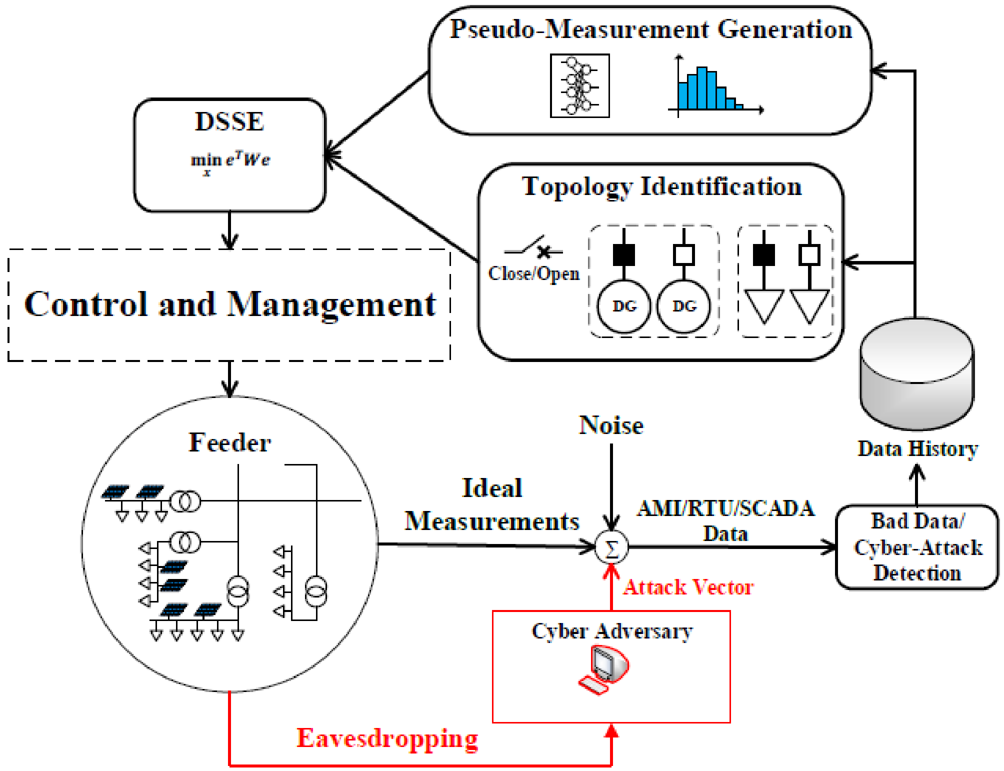


Fig. 11 DSSE function in smart grid environment [10].

Case Study I: Background

At presents, distribution systems can be divided into three groups according to the observability: fully observable systems, partially observable systems and fully unobservable systems.

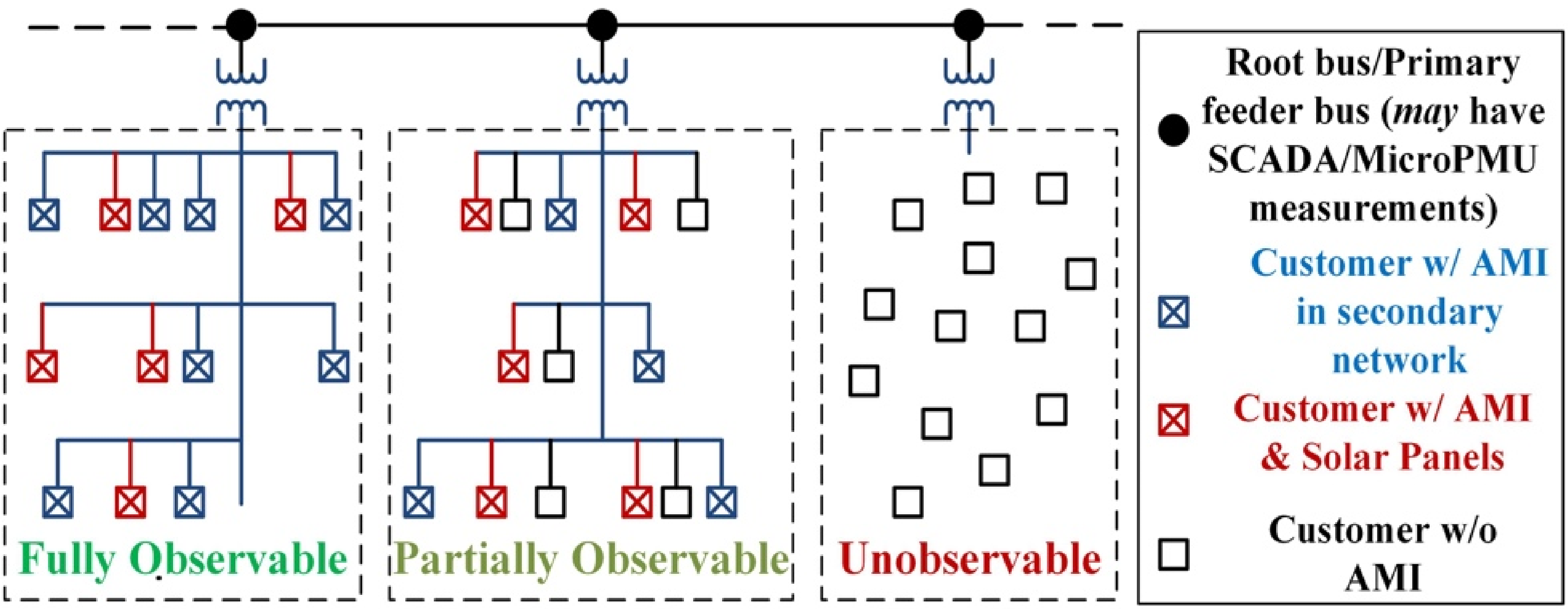


Fig. 12 Distribution systems with different observability.

Case Study I: Overview Framework

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:
- Using data clustering for capturing customer typical behaviors
- Multi-timescale load inference (stage by stage inference chain)
- Using state-estimation-based Bayesian learning for inferring unobserved customers' typical behaviors

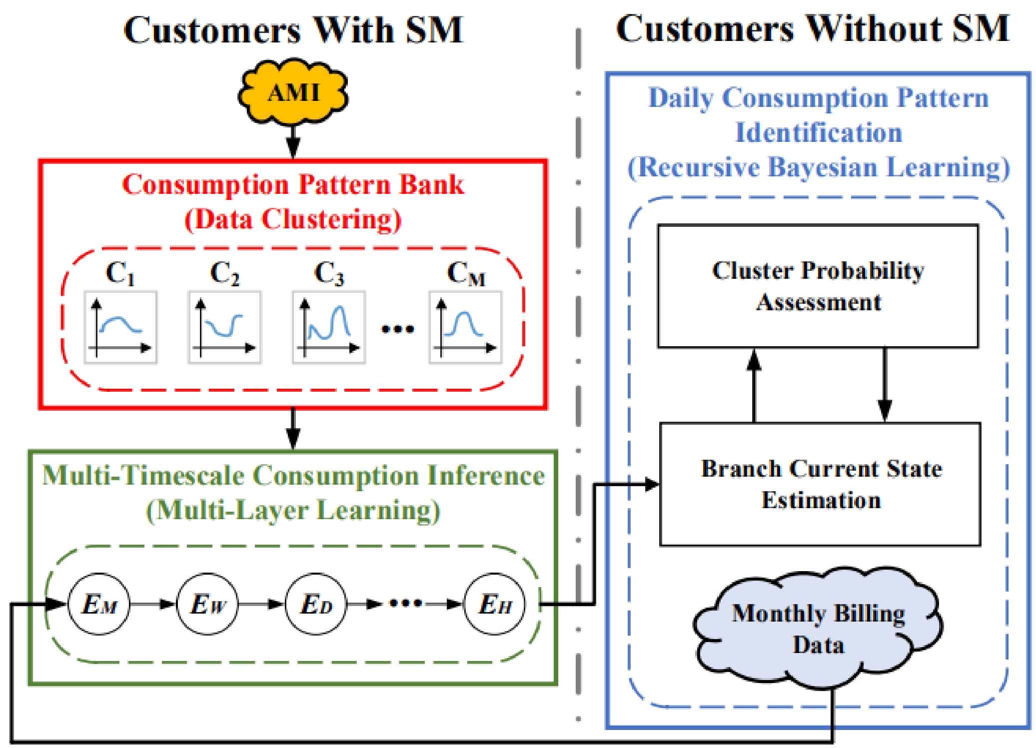
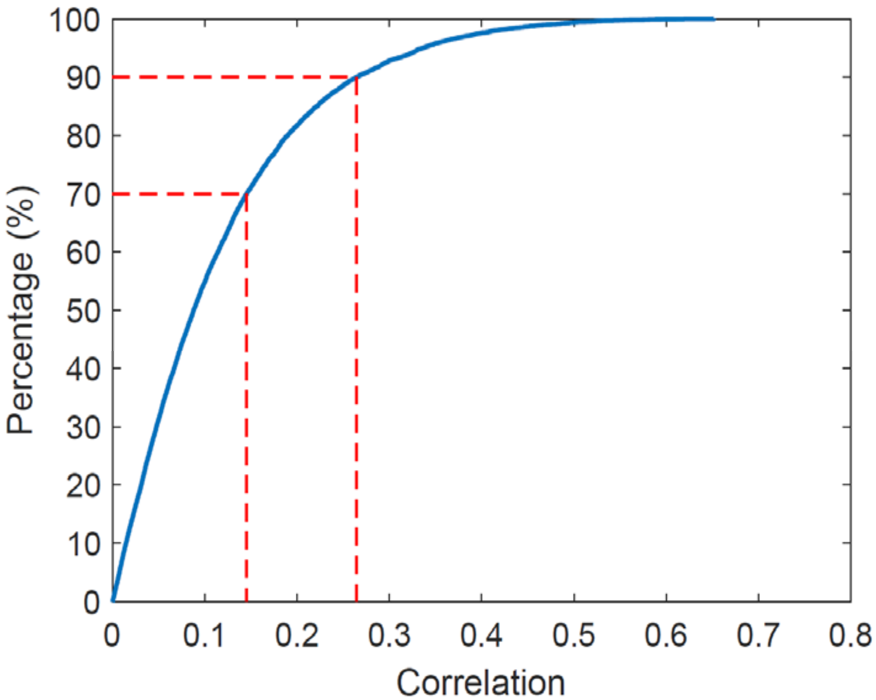


Fig. 13 Overall structure of the proposed method.

Evidence from Data: How to maintain high correlation



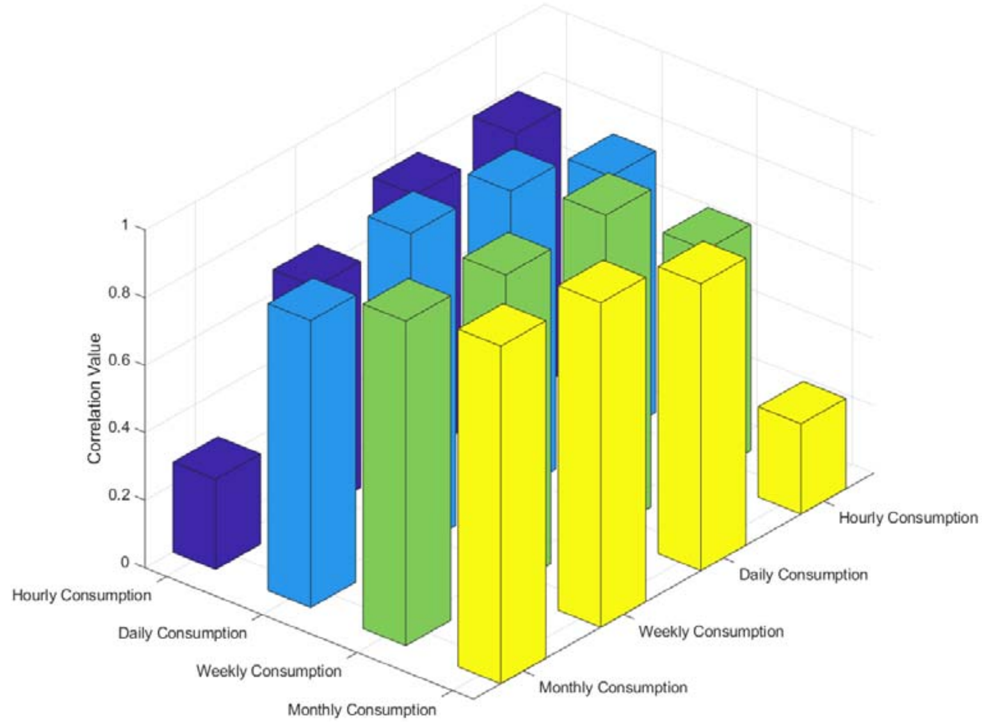


Fig. 14 Correlation between different customers' smart meter and correlation between Consumption at Different Time-Scales.

- Hourly load variations cannot be directly observed at the monthly level.

- An efficient inference model needs to keep the high correlation level between different time resolution data.

Using Observed Customers' Data for Training Multi-Timescale Load Inference Chain Models

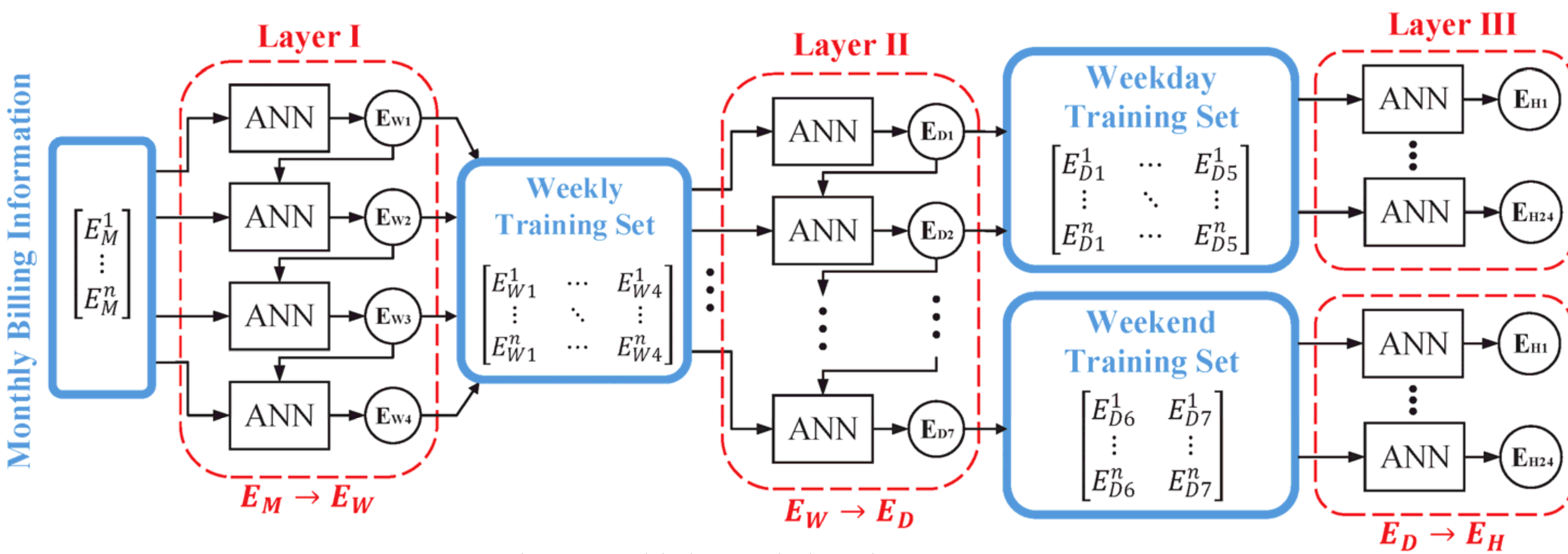


Fig. 15 Multi-timescale learning structure.

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!)

Multi-Timescale Load Inference Chain Model Description

- Layer I maps total monthly consumption, E_M, to the set of weekly consumption values $E_W = \{E_{W1}, \dots, E_{W4}\}$ using ANNs connected in series.

- To capture the temporal correlation between consumption behaviors, each week's estimated consumption is also fed to the next ANN corresponding to the following week's consumption.

- Above idea is generalized to all the layers of the proposed multi-timescale load inference method.

How to Update the ANN Weight and Bias

- The ***Levenberg-Marquardt*** (LM) backpropagation method is used to update the ANN weight and bias variables [24].

- The training objective function and the update equation of LM algorithm can be written as:

$$\min_{b} F(b) = \sum_{i=1}^{Q} v_i^2(b) = v^T(b)v(b)$$

$$\Delta b_l = -[J^T(b_l)J(b_l) + \mu_l I]^{-1} J^T(b_l)v(b_l)$$

where, μ_l is the combination parameter at iteration l, b is the set of learning parameters, J is the training objective function's Jacobian, I is the identify matrix, v is the error vector, T is the matrix transposition operation, and Δb_l defines the learning parameter updates at each iteration.

- In each iteration, the value of μ_l is updated based on the change of approximated performance index $F(b)$. If a smaller value is obtained, the μ_l is divided by some factor $\vartheta > 1$. Otherwise, μ_l is multiplied by ϑ for the next iteration.

Hyperparameter Calibration Results

- For each ANN, the dataset is randomly divided into three separate subsets for training (70% of the total data), validation (15% of the total data), and testing (15% of the total data).

- ***Early stopping mechanism*** and ***noise injection strategy*** are utilized to reduce the overfitting in the training process.

- Several hyper-parameters are calibrated using the grid search methods.
  - The number of hidden layer.
  - The number of neurons.
  - The value of learning rate.

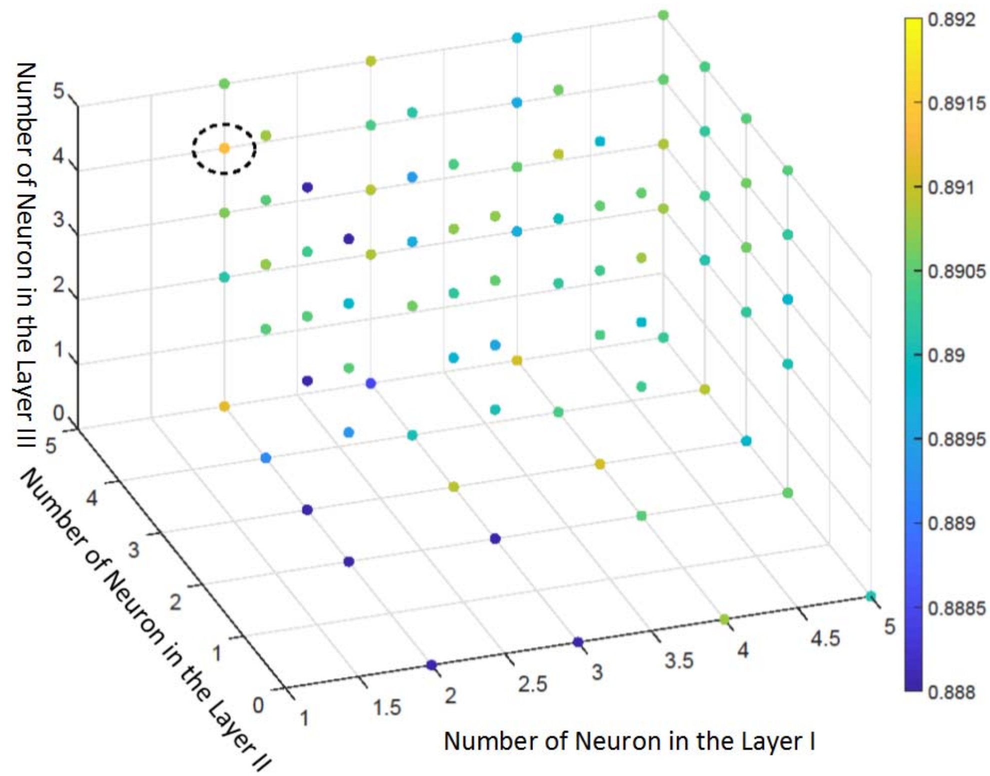


Fig. 16 Calibration result for ANN.

Sensitivity Analysis of Observability

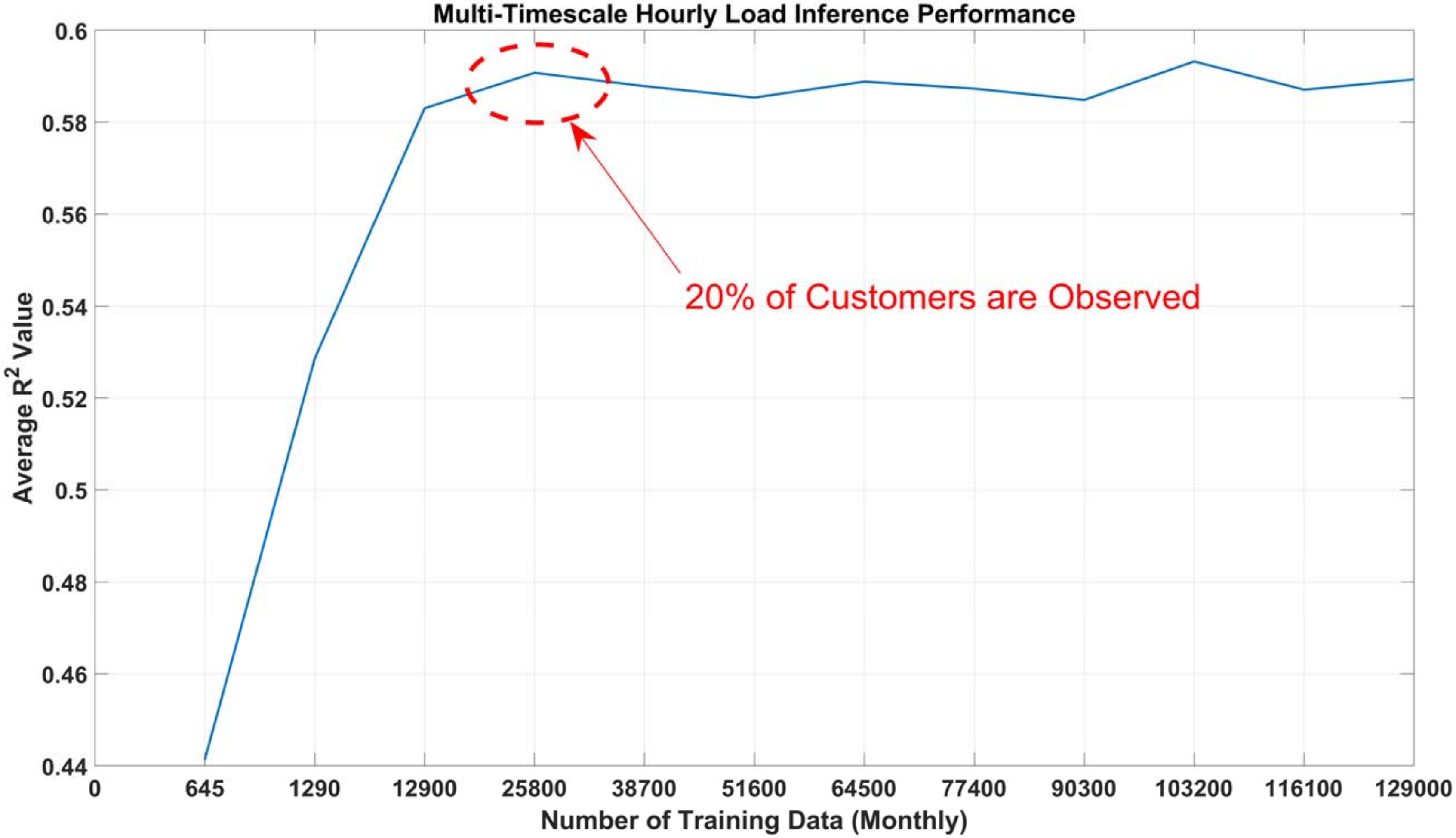
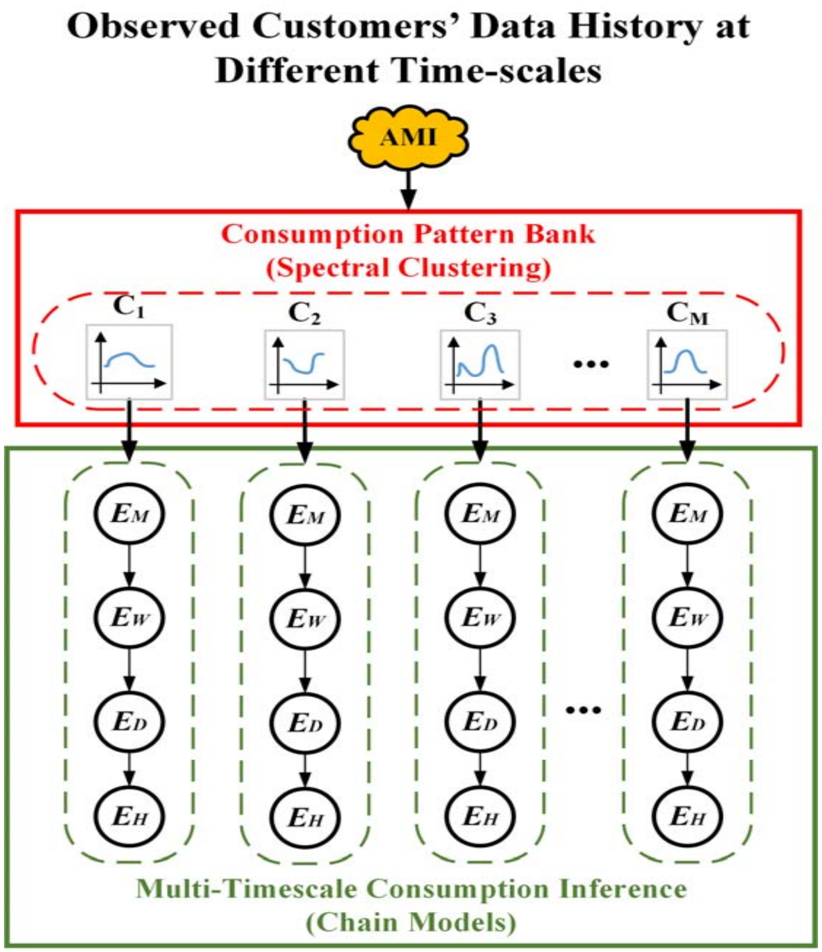


Fig. 17 Performance of quantifying learning reliability.

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

Spectral Clustering

Advantages of the spectral clustering:

- Spectral clustering can work with ***arbitrary complex*** data, because it does not make assumptions on the shapes of clusters.

- Spectral clustering is able to handle the ***high-dimensional time-series load data***, because it leverages the inherent redundancy of the *entire* distance matrix and enhances distance representations of newly embedded data in a lower dimensional space.

- Unlike the standard spectral clustering that relies on a ***scaling parameter*** $\boldsymbol{\alpha}$ to measure the similarity between two data samples, a local parameter strategy is applied in our work. This strategy not only provides a way for automatic pick the α but also effectively exploits the local statistics for handling multi-scale data.

Spectral Clustering

Using the spectral clustering to develop the consumption pattern bank can be summarized as a ***three-step*** algorithm:

1) The first step is to transform SM dataset into an ***undirected similarity graph G***.

2) The data points are embedded in a space, in which the clusters are more ***observable***, with the use of the ***eigenvectors of the graph Laplacian***.

3) A classical clustering algorithm (e.g., k-mean) is applied to ***partition the embedding***.

Spectral Clustering

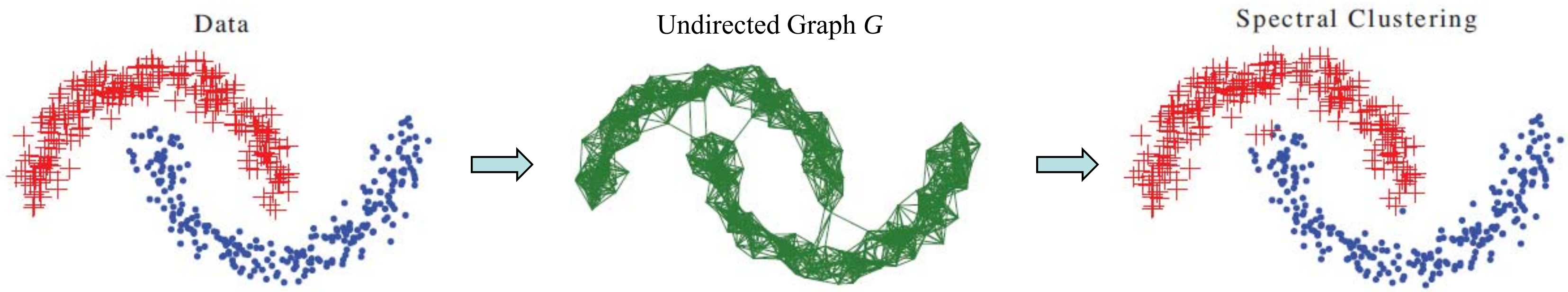


Fig. 18 Example illustrating **three steps** of spectral clustering [21].

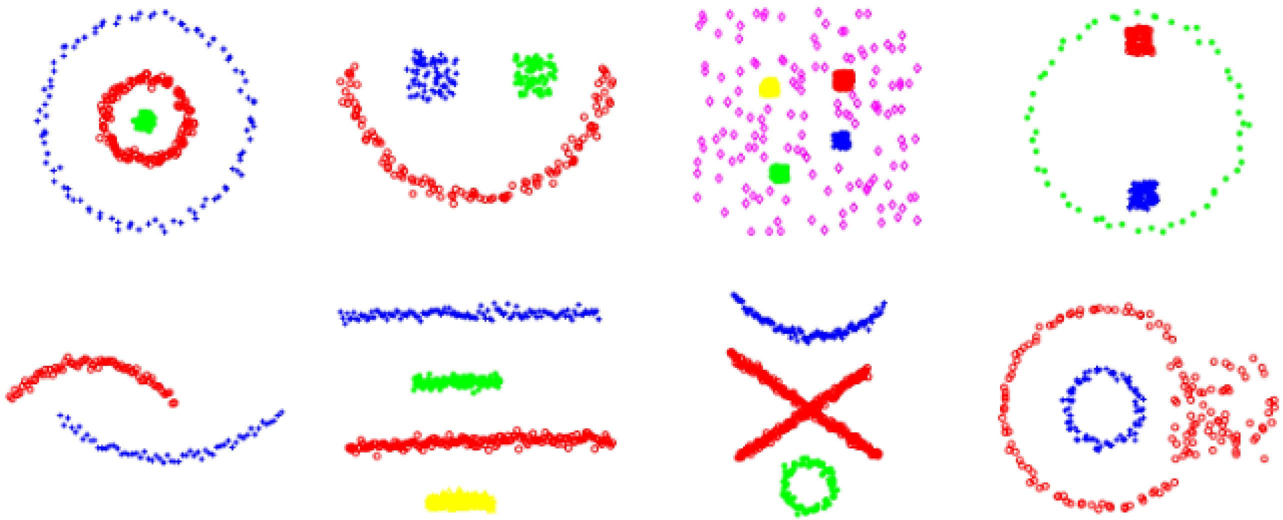


Fig. 19 Spectral clustering on toy datasets [22].

Numerical Results: Customer Behavior Pattern Bank

- The spectral clustering is utilized to classify different load shapes and to create the consumption pattern banks.
- Fig. 17(a) shows the typical load patterns for different types of customers for ***weekends***.
- Fig. 17(b) shows the typical load patterns for different types of customers for ***weekdays***.
- Red represents ***Industrial***, blue represents ***commercial***, and black represents ***residential*** customers.

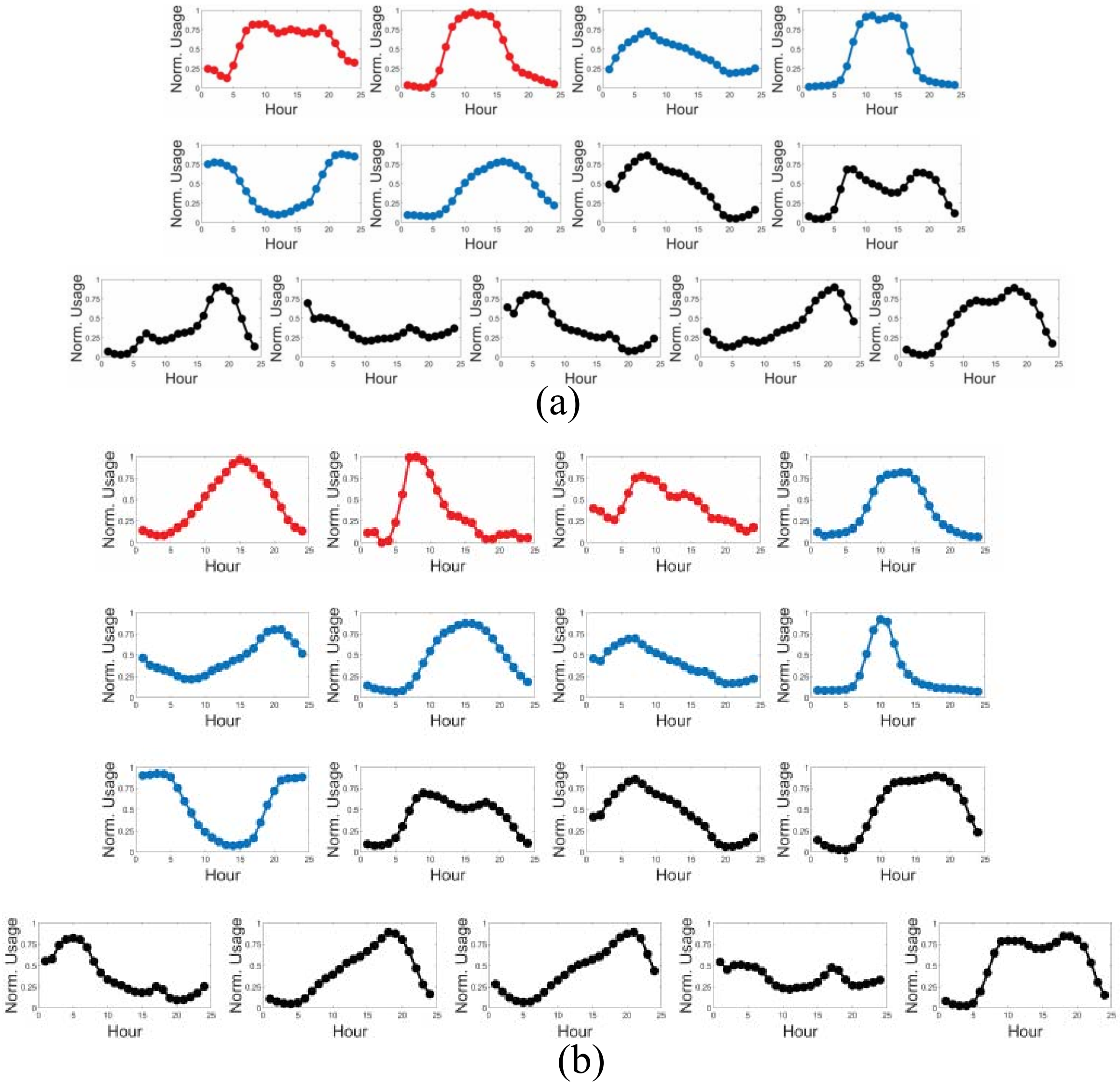
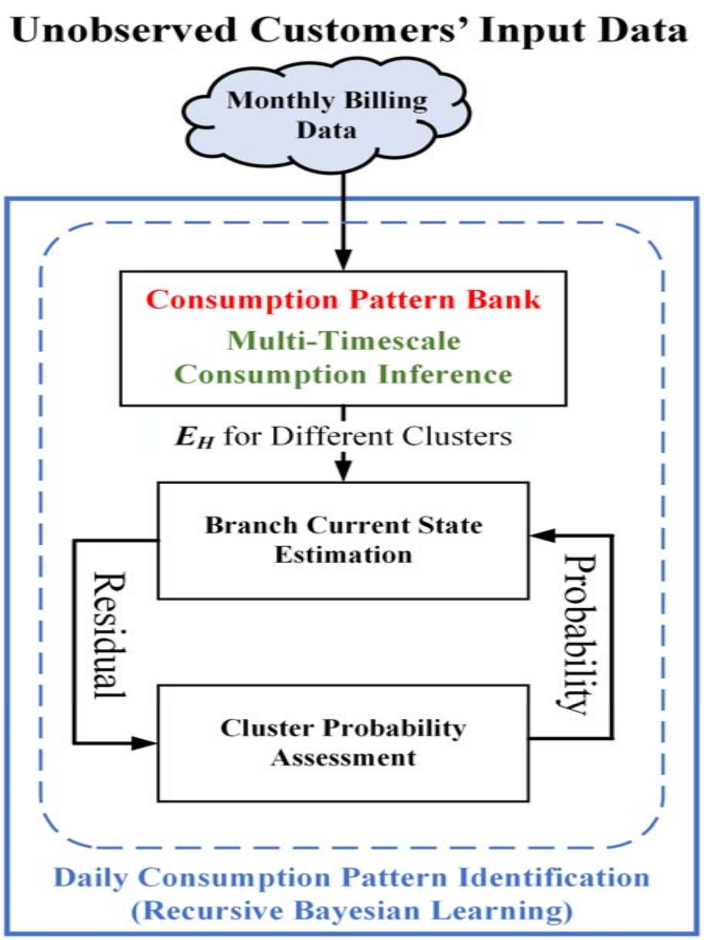


Fig. 20 Typical pattern banks for weekday and weekend.

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

Branch Current State Estimation

BCSE algorithm:

- Three-phase feeder model
- Use ***branch current magnitudes and phase angles*** as state variables
- Insensitive for branch impedance
- Better performance in computation speed and memory usage
- Weighted Least Square (WLS) estimator

$$\min_{x} J = \left(z - h(x)\right)^T \Sigma (z - h(x))$$

$$G(x) = H^T(x) \Sigma H(x)$$

$$[G(x^m)]\Delta x^m = H^T(x^m) \Sigma (z - h(x^m))$$

$$x^{m+1} = x^m + \Delta x^m$$

Note: z is the real measurement, x is the state variable, h(.) is the nonlinear measurement function, Σ is the weight matrix that represents the users confidence in the measured data

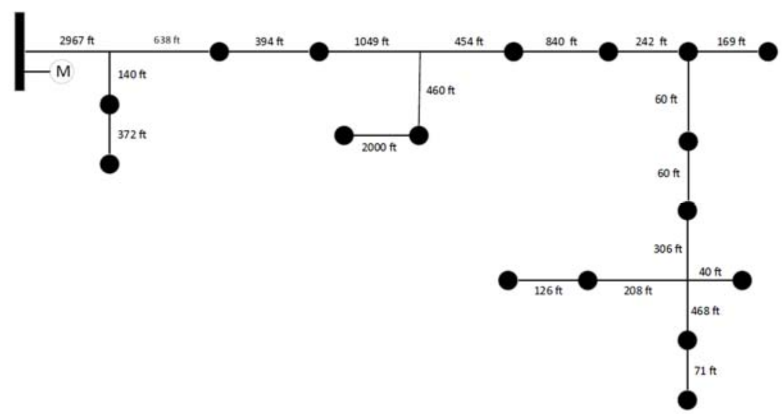


Fig. 21 A 18-node real utility feeder case.

BCSE-aided Recursive Bayesian Learning Model

- The probability of each typical load pattern is computed using the ***residuals of the BCSE algorithm*** and the ***recursive Bayesian learning model***.
- The residuals are calculated by comparing the real measurements of several feeder-level devices and the estimated values from the BCSE.
- Applying the Bayes theorem and assuming a Gaussian distribution for measurement error, a recursive expression for updating the probability of each load pattern over time can be written as [25]:

$$p_{i,j}^{o} = \frac{\exp\left(-\frac{1}{2} {r_{i,j}^{o}}^{T} * \Phi * r_{i,j}^{o}\right) * p_{i,j}^{o-1}}{\sum_{t-1}^{N} \exp\left(-\frac{1}{2} {r_{i,j}^{o}}^{T} * \Phi * r_{i,j}^{o}\right) * p_{i,j}^{o-1}}$$

Where, o is the iteration count, $r_{i,j}^{o}$ is the residual vector of the i'th class with respect to j'th customer and is computed by the corresponding state and real measurement vectors, Φ is a diagonal matrix that represents the variances corresponding to the branch current real/imaginary part residuals.

Numerical Results: Unobserved Individual Customer Hourly Load and Pattern Inference

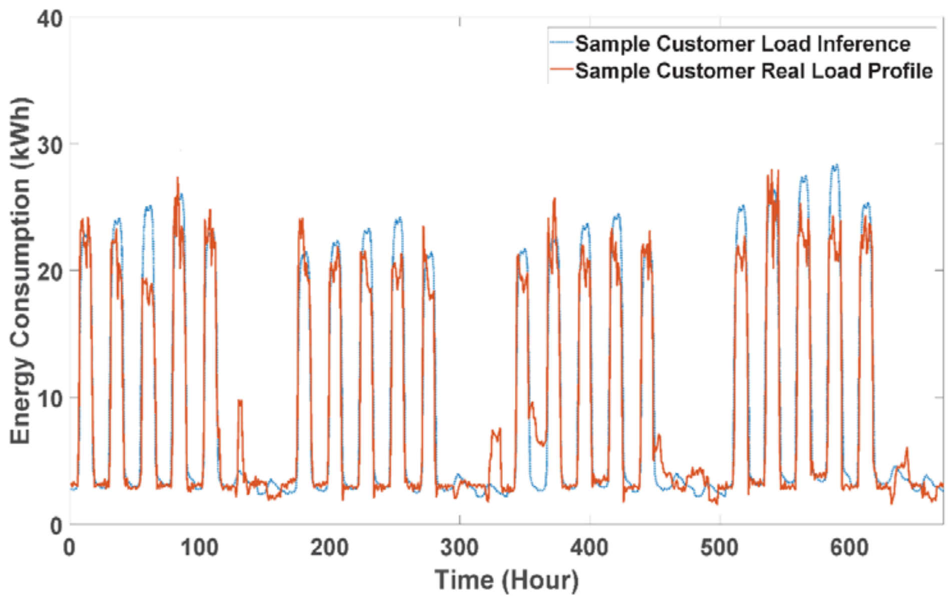


Fig. 22 Inferring the hourly demand of an unobserved residential load in one month (average estimation error ≈ 8.5% of total energy)

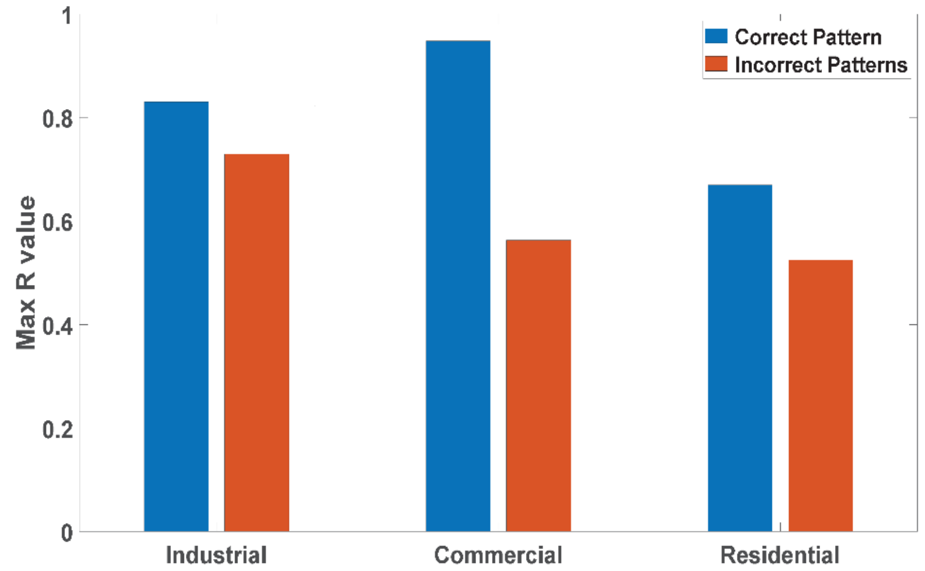


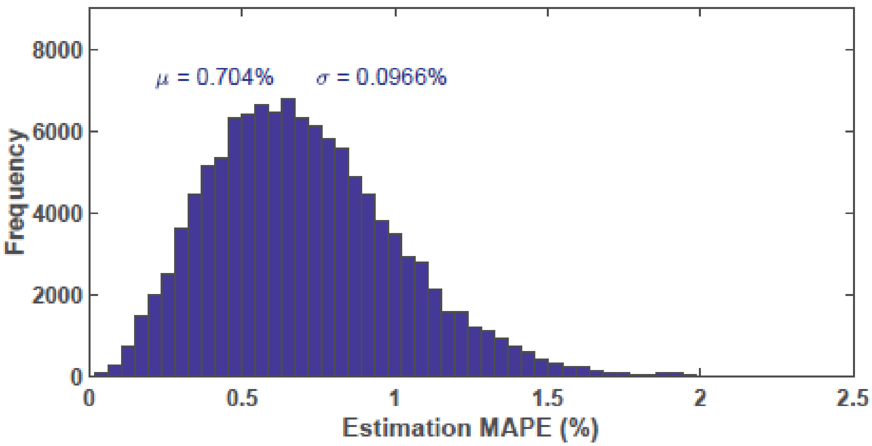
Fig. 23 Impact of accurate consumption pattern identification on the accuracy of the inference (industrial load patterns are close and stable)

Numerical Results: DSSE Performance Using the Proposed Method

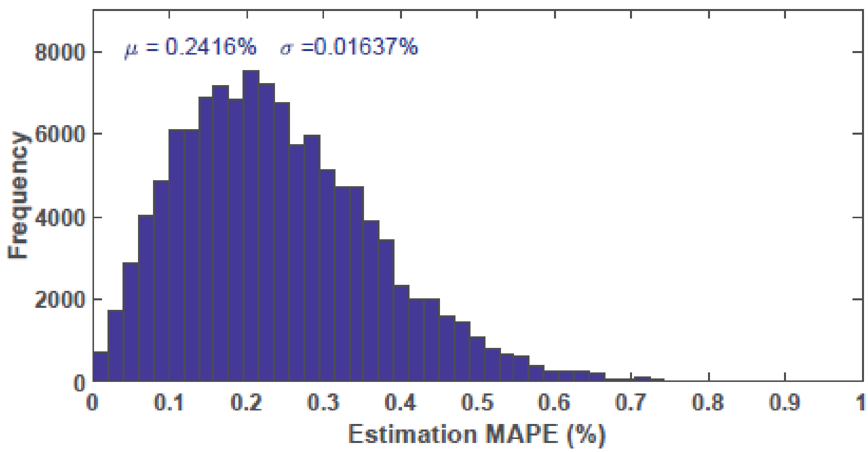
- Apply the MAPE criterion to evaluate the accuracy of BCSE based on the proposed pseudo load estimation.

- The MAPE values for the voltage magnitude and phase angle are around 0.704% and 0.24%, respectively.

- In the previous work [11], the MAPE values are around 0.73% and 0.36%, respectively with 20% maximum Gaussian error for pseudo measurements.



(a) Voltage magnitude component error



(b) Voltage phase component error

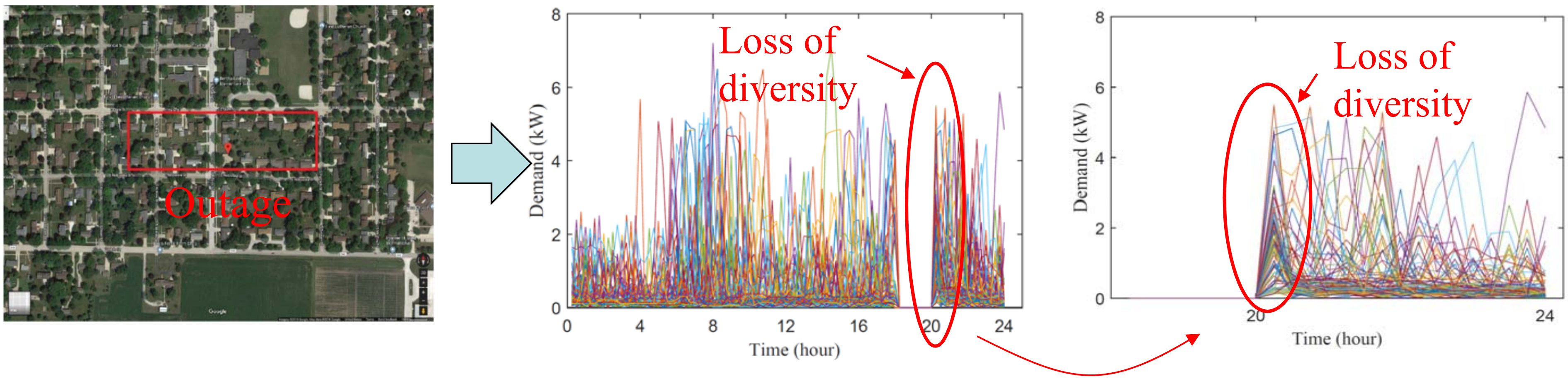
Fig. 24 BCSE-based state estimation performance using the proposed load inference model.

Contents

Case Study II - A Data-Driven Framework for Assessing Cold Load Pick-up Demand in Service Restoration

What is CLPU?

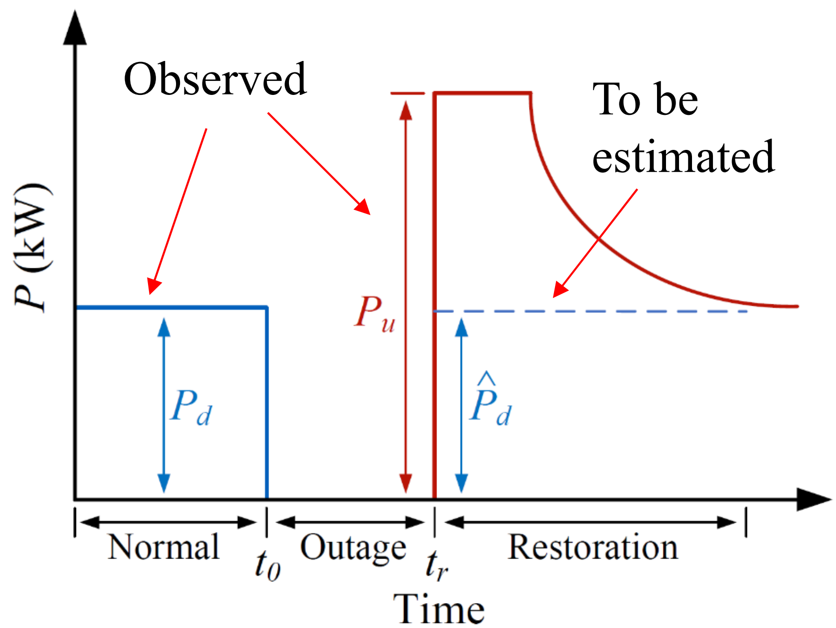
- In distribution systems, thermostatically controlled loads (TCLs) are ***diversified*** during normal operation, and ***undiversified*** in restoration

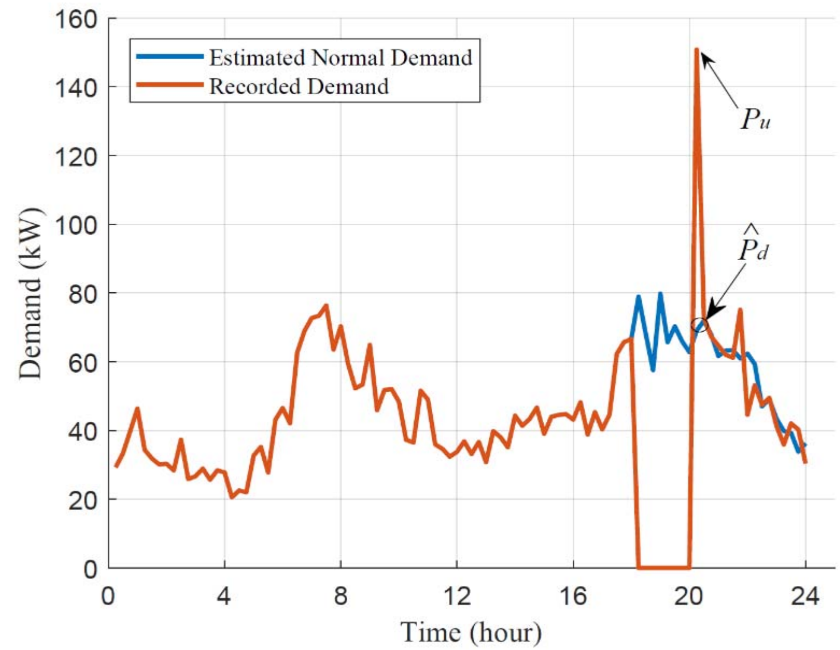


- The phenomenon of losing load diversity is called Cold load pick-up (CLPU) and can cause considerable demand ***increase***
- Serious ***consequences***: restoration failure, transformer aging and overloading, and unacceptable voltage drops
- It is necessary to ***assess*** historical CLPU demand, and ***extract*** useful information to predict CLPU demand in future restorations

Background

- ***Typical exponential CLPU model***





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

Real CLPU at the feeder level

P_d -- Normal demand $\hat{P}_d$ -- Estimated normal demand

P_u -- Undiversified demand at the time of restoration

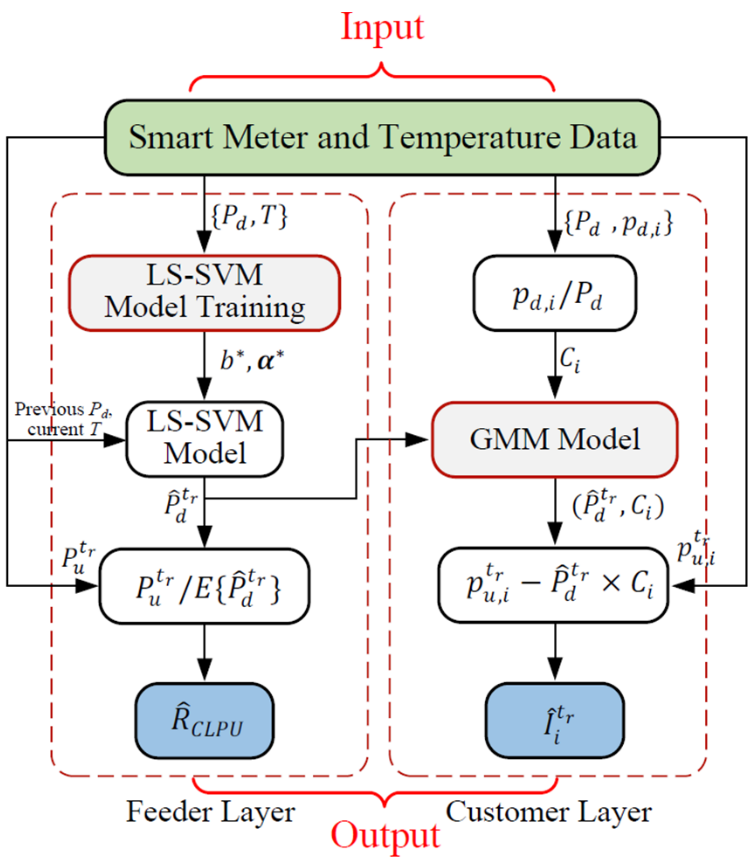
- CLPU ratio is used to predict CLPU demand in future restorations

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
- ***Drawbacks***:
  - ➢ 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 Framework



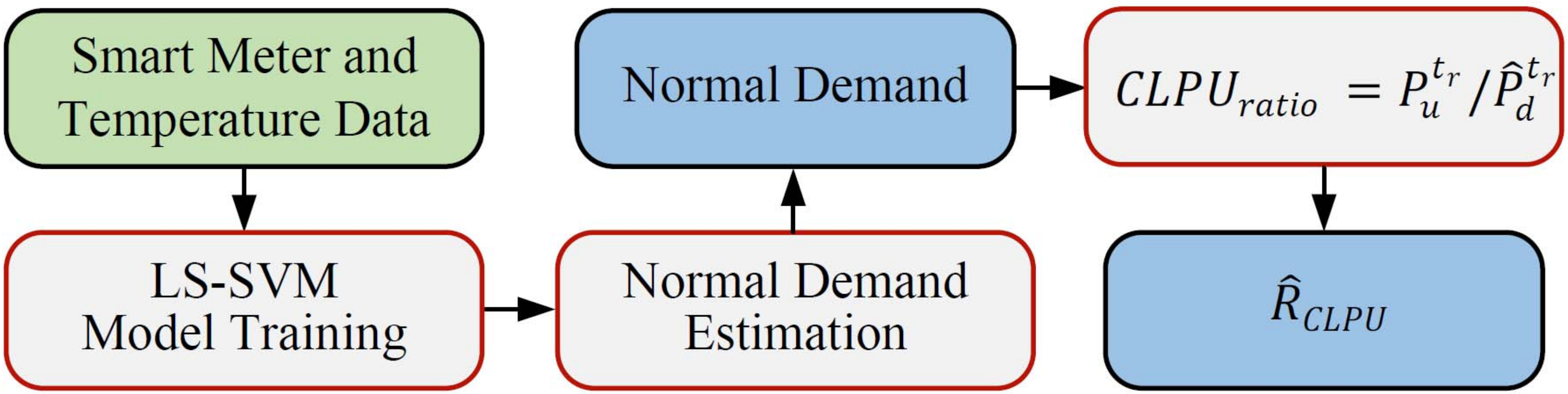
- ✓ Obtain CLPU ratio at feeder-level, using ***LS-SVM*** prediction approach
- ✓ Determine customer demand increase, using ***GMM*** and probabilistic reasoning
- ✓ Obtain useful statistics at feeder- and customer-level to fully quantify CLPU demand

P_d -- Feeder-level normal demand
T -- Temperature
$b^*, \boldsymbol{\alpha}^*$ -- Trained model Parameters
$\hat{P}_d^{t_r}$ -- Estimated feeder-level normal demand at t_r
$P_u^{t_r}$ -- Undiversified feeder-level demand at t_r

$p_{d,i}$ -- Historical customer-level normal demand
C_i -- Historical customer contribution factor
$(\hat{P}_d^{t_r}, C_i)$ -- Joint distribution of $\hat{P}_d^{t_r}$ and C_i
$p_{u,i}^{t_r}$ -- Customer demand at the time of restoration
$\hat{I}_i^{t_r}$ -- Estimated customer demand increase due to CLPU

Feeder-Level CLPU Demand Assessment

- ***Objective***: Obtain CLPU Ratios from historical outage events, and develop a CLPU ratio regression model
- ***Methodology***: Estimate the feeder-level normal demand assuming *the outage did not happen*
- ***Algorithm***: Least-squares support-vector machines (LS-SVM)



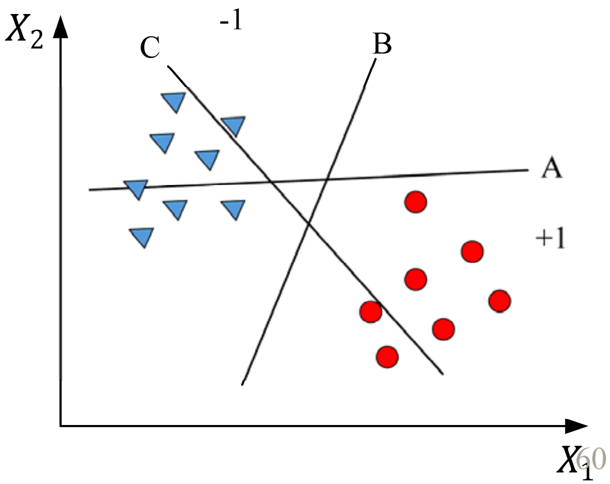
LS-SVM

- ***Why do we choose LS-SVM?***
  - Advantages: good generalization capability, and low susceptibility to local minima, etc.

Let's start with SVM.

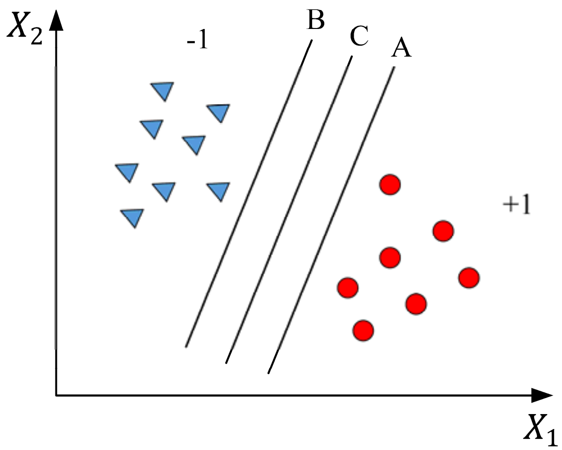
- ***What is Support Vector Machine (SVM)?***
  - It is a ***supervised*** machine learning algorithm
  - It can be used for ***classification*** or ***regression***
- ***How does SVM work?***

  Let's start with classification.
  In this scenario,
  - hyper-plane "B" ***can*** successfully segregate the two classes.
  - hyper-plane "A" and "C" ***cannot***.



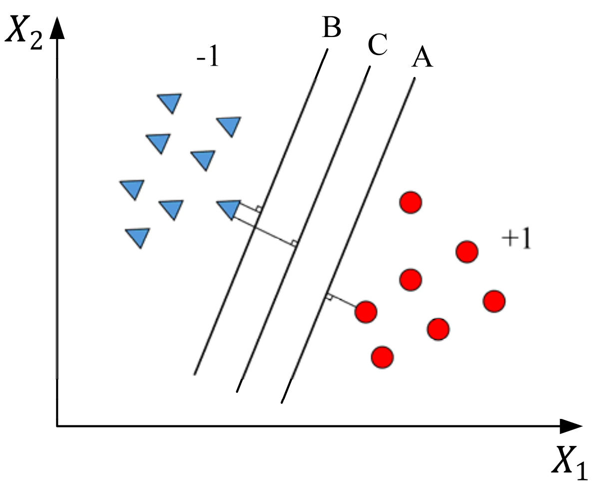
LS-SVM

- ***How does SVM work?***



In this scenario, hyper-planes A, B and C all can segregate the two classes.

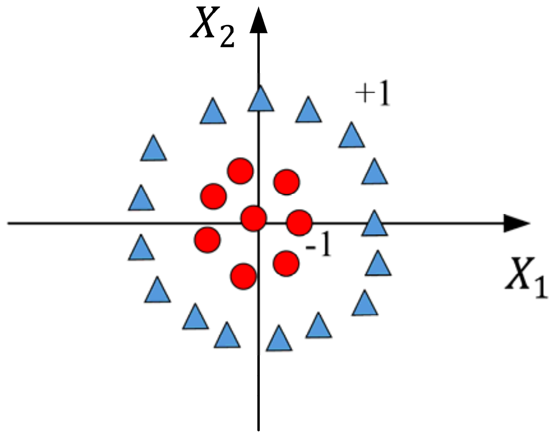
Which hyper-plane is ***best***?



- By drawing perpendicular lines, we can know that maximizing the ***distances*** between nearest point and hyper-plane can determine the optimal hyper-plane.
- This distance is called ***margin***, the nearest points are called ***support vectors***.
- A large margin means less chance of ***misclassification***, and this is what we want.

LS-SVM

- ***How does SVM work?***



In this scenario, the previous linear hyper-plane cannot work.

What can we do?

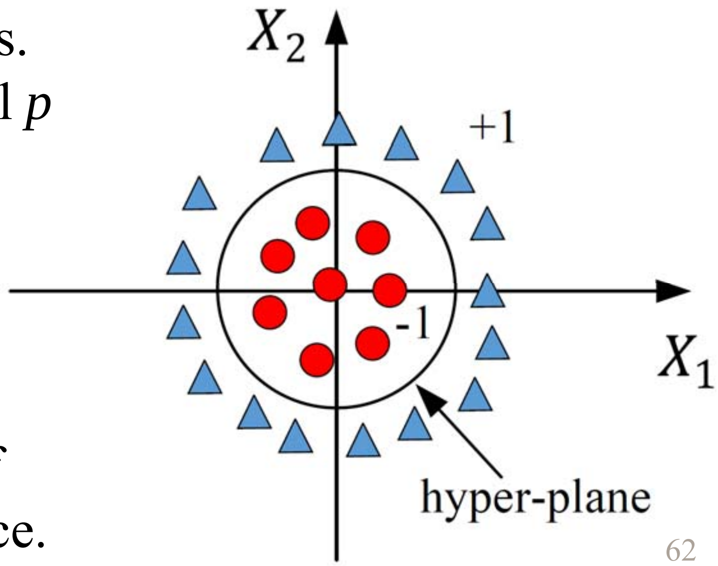
Solution: Introduce *higher-order* features. For instance, instead of using the original *p* features:

$$X_1, X_2, \dots, X_p,$$

we enlarge the feature space to 2*p*:

$$X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2.$$

In general, a ***kernel***, which has a form of $K(\boldsymbol{x}_i, \boldsymbol{x}_i')$, is used to enlarge feature space.

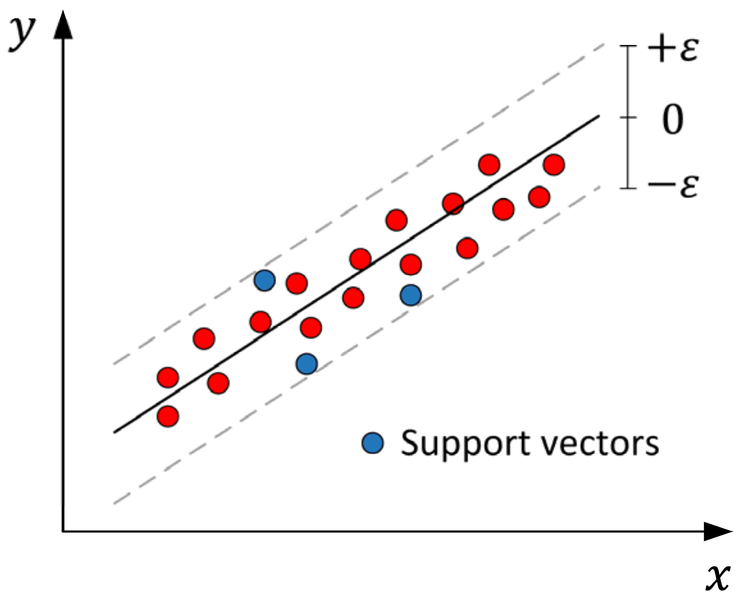


LS-SVM

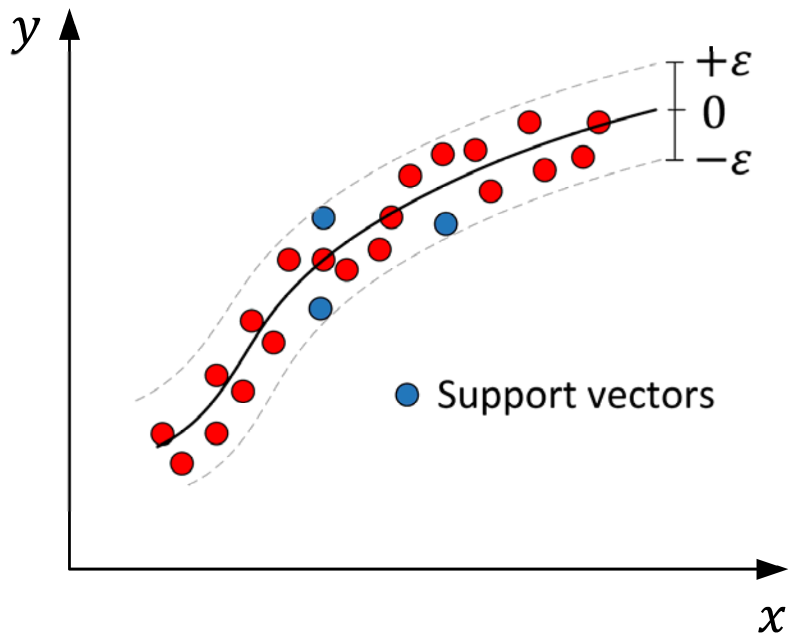
- ***How does SVM work?***

Now we extend the classification to regression.

- Similarly, for regression, we use support vectors to determine an optimal margin.



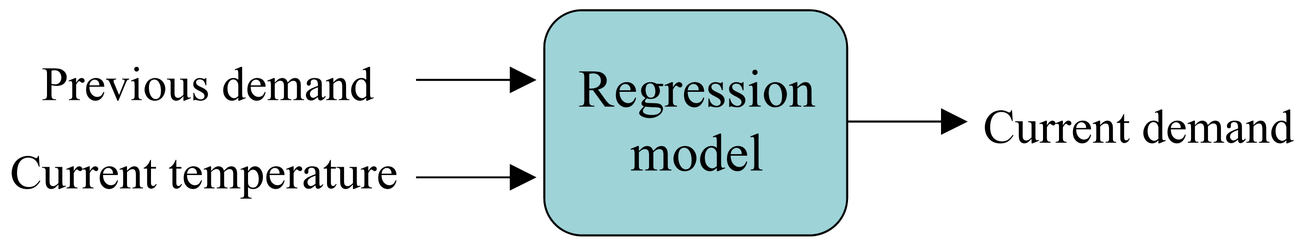
Liner regression



Nonliner regression

LS-SVM

- ***How does SVM work?***
  - In our problem, we have



  - To train the ***regression model***, first, we define:
    - ➢ A regression model: $P_d^t = \boldsymbol{w}^T \phi(\boldsymbol{x}_t) + b$
    - ➢ P_d^t denotes the normal demand at time t
    - ➢ $\boldsymbol{x}_t = [P_d^{t-1}, \dots, P_d^{t-n_{lag}}, T_t]^{\mathrm{T}}$
      where, P_d^{t-1} denotes the normal demand at time $t-1$, n_{lag} is the time lag, T_t is the temperature at time t.
    - ➢ $\phi(\boldsymbol{x})$ denotes the nonlinear transformation.
    - ➢ $\boldsymbol{w}$ and b are regression parameters, where $\boldsymbol{w} = [w_1, \dots, w_n]^{\mathrm{T}}$

LS-SVM

- Then, maximizing the ***margin*** is formulated as

$$\min_{\boldsymbol{w},b} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$\text{s.t.}\ \ P_d^t = \boldsymbol{w}^T\phi(\boldsymbol{x}_t) + b, \qquad t = 1, \dots, N$$

where, $\boldsymbol{x}_t = [P_d^{t-1}, \dots, P_d^{t-n_{lag}}, T_t]^{\mathrm{T}}$.

- By introducing a "***Least Square***" term, we obtain LS-SVM

$$\min_{\boldsymbol{w},b,e} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \gamma\frac{1}{2}\sum_{t=1}^{N} e_t^2$$

$$s.t.\ \ P_d^t = \boldsymbol{w}^T\phi(\boldsymbol{x}_t) + b + e_t, \qquad t = 1, \dots, N$$

where, e_t denotes the estimation residual, γ is a regularization constant.

LS-SVM

- To solve this optimization problem, the Lagrangian, $\mathcal{L}$, is constructed as a function of regression parameters:

$$\mathcal{L}\,(\boldsymbol{w}, b, e; \alpha) = J(\boldsymbol{w}, e) - \sum_{t=1}^{N} \alpha_t \{\boldsymbol{w}^T \phi(\boldsymbol{x}_t) + b + e_t - P_d^t\}$$

where, α_t are Lagrange multipliers.

- The ***solution*** is given by

$$\begin{bmatrix} b^* \\ \boldsymbol{\alpha}^* \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1}^{\mathrm{T}} \\ \mathbf{1} & \Omega + \frac{1}{\gamma}\mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \boldsymbol{P}_d \end{bmatrix}$$

where,

$\boldsymbol{P}_d = \left[P_d^1, \ldots, P_d^N\right]^{\mathrm{T}}, \mathbf{1} = [1, \ldots, 1]^{\mathrm{T}}, \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^{\mathrm{T}},$

$\Omega_{t't} = K(\boldsymbol{x}_{t'}, \boldsymbol{x}_t) = exp(-\frac{||\boldsymbol{x}_{t'} - \boldsymbol{x}_t||^2}{\sigma^2}), \quad t', t = 1, \ldots, N$, σ is the width of Gaussian Kernel.

LS-SVM

- Then, the ***normal demand at the time of restoration*** is estimated as

$$\hat{P}_d^{t_r} = \sum_{t=1}^{N} \alpha_t^* K\left(\boldsymbol{x}_t, \boldsymbol{x}_{t_r}\right) + b^*$$

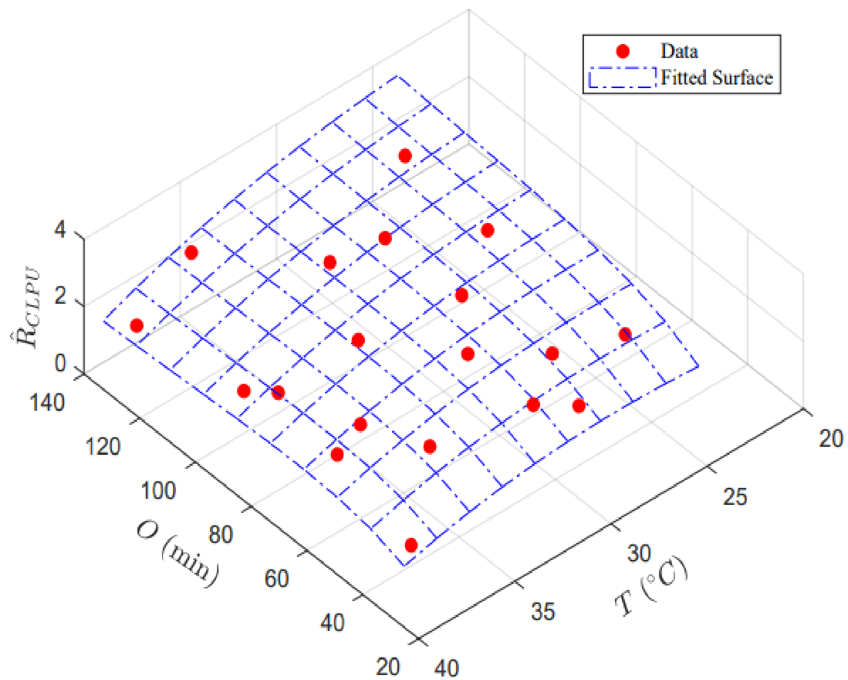
where, $\boldsymbol{x}_t = [P_d^{t-1}, \ldots, P_d^{t-n_{lag}}, T_t]^{\mathrm{T}}$, $\boldsymbol{x}_{t_r}^* = [P_d^{t_r-1}, \ldots, P_d^{t_r-n_{lag}^*}, T_{t_r}]^{\mathrm{T}}$, t_r is the time of restoration, n_{lag}^* is the optimal time lag which is obtained by grid search.

- Finally, we calculate the ***CLPU ratio*** via

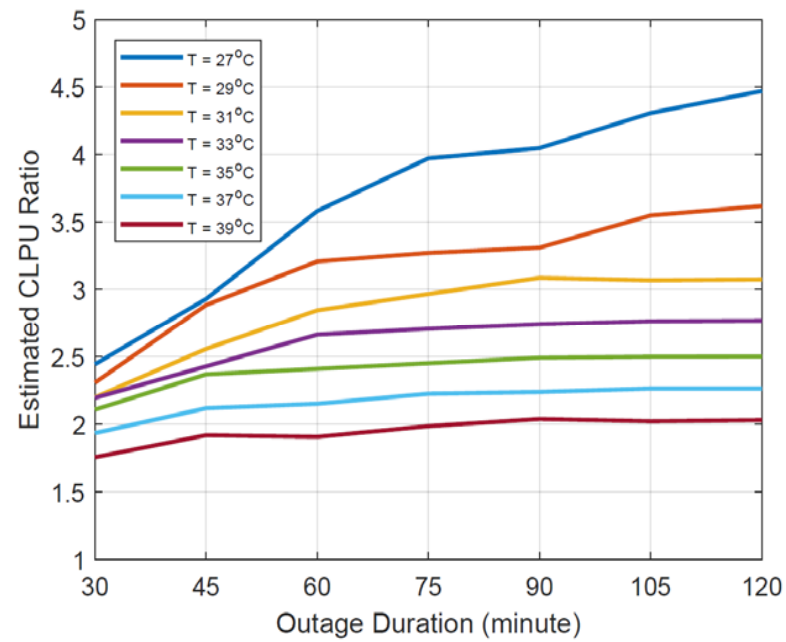
$$\hat{R}_{CLPU} = \frac{P_u^{t_r}}{E\{\hat{P}_d^{t_r}\}}$$

when, $P_u^{t_r}$ is the recorded *undiversified* feeder demand at t_r. $E\{\cdot\}$ denotes the *empirical averaging operator*, this is based on the consideration that $\hat{P}_d^{t_r}$ follows a distribution due to estimation errors.

Feeder-Level CLPU Ratio Result



Developed CLPU ratio regression model

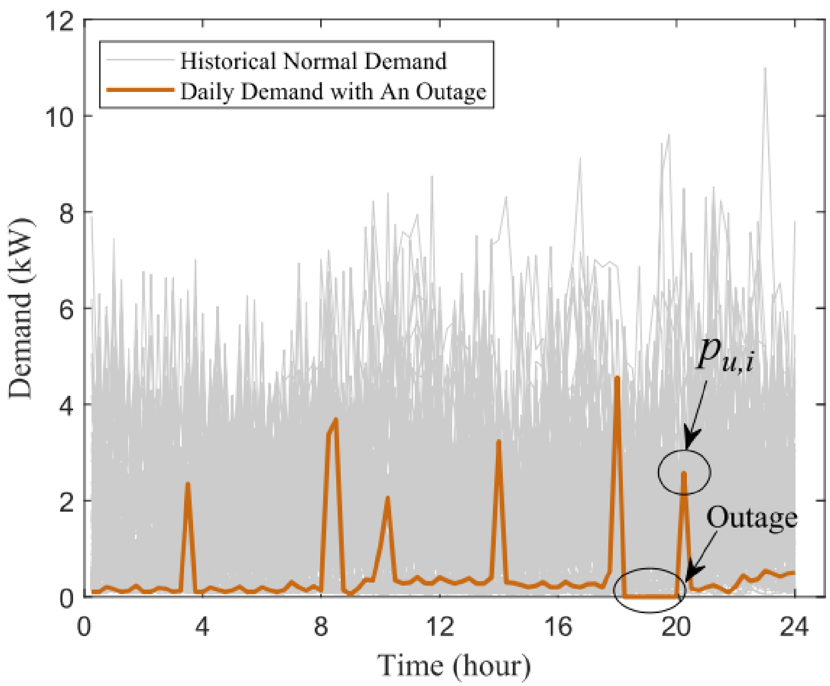


Expected CLPU ratio with respect to temperature and outage duration

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

Customer-Level CLPU

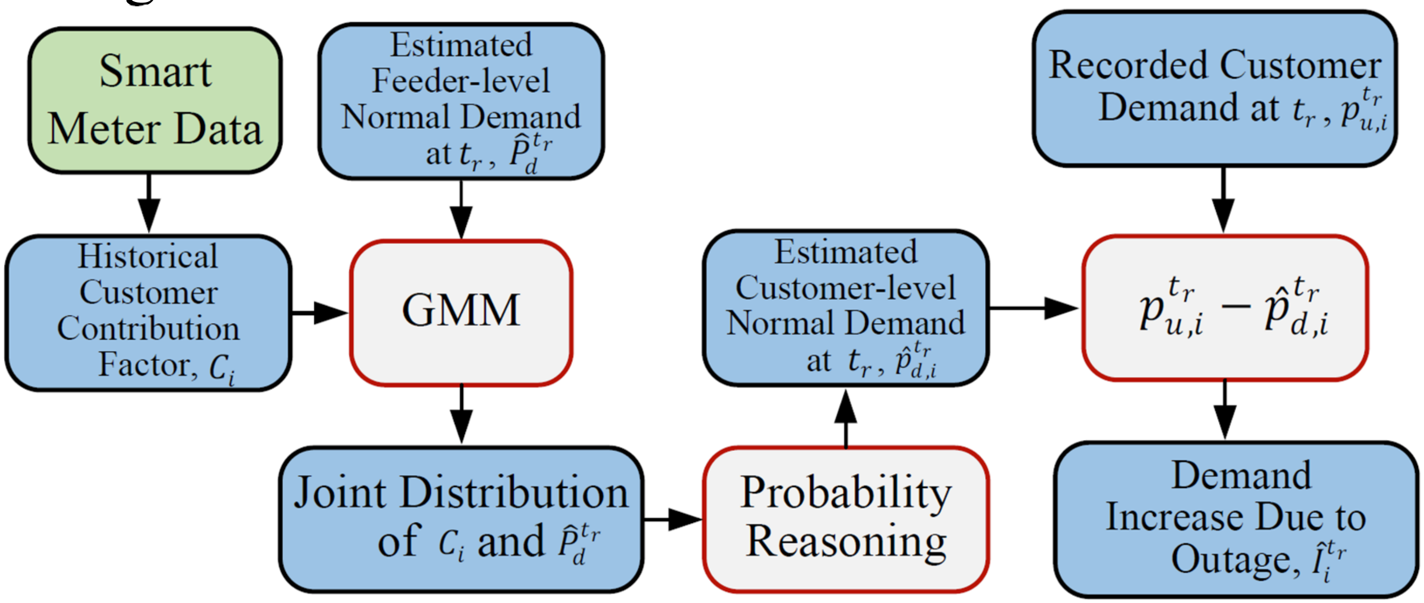
At the feeder level, we obtained historical CLPU ratios. Can we apply the same approach to the customer level?



- Individual customer consumption can be quite ***stochastic***.
- Considering this uncertainty, the feeder-level approach cannot be directly applied to customer-level demand estimation.
- We seek to use ***GMM*** and ***probabilistic reasoning*** to estimate the distribution of normal customer demand at the time of restoration.

Customer-Level CLPU Demand Assessment

- ***Objective***: Quantify the CLPU demand increase caused by outage.
- ***Methodology***: Estimate the distribution of customer-level normal demand assuming *the outage did not happen*
- ***Algorithm***: Gaussian mixture model (GMM) and probabilistic reasoning

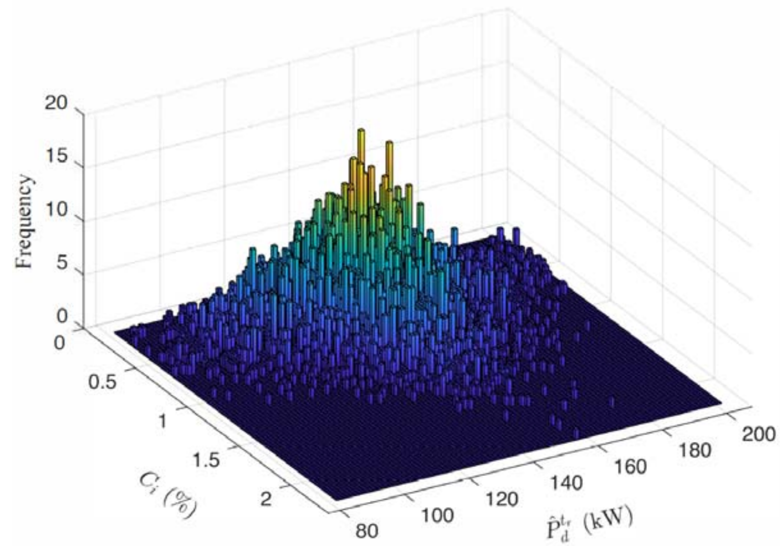


GMM

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.

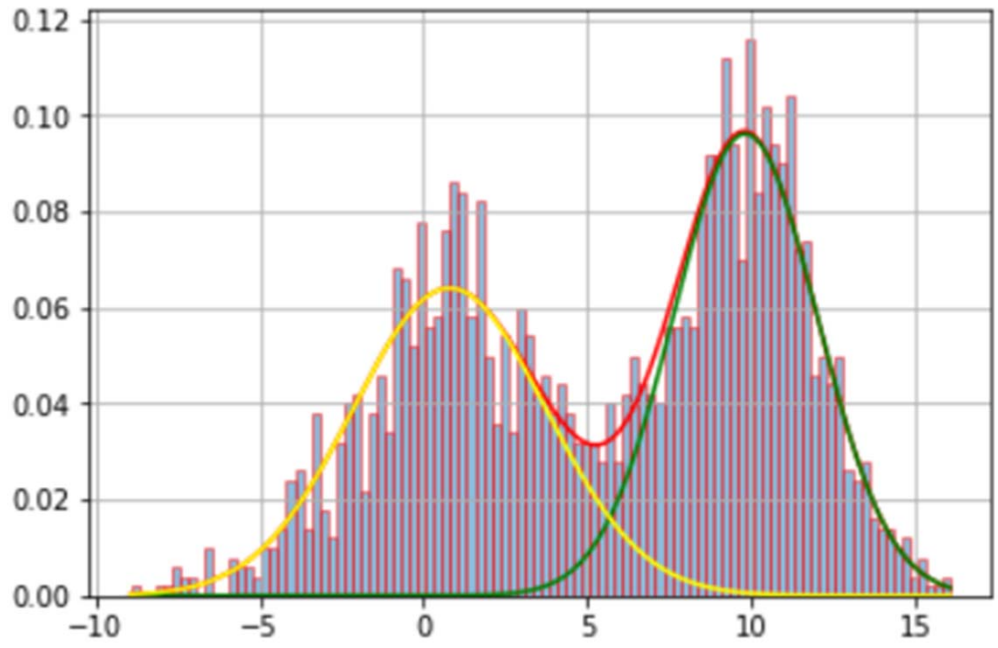
- This 2-dimensional histogram does not strictly fit a ***single*** distribution model. Therefore, a ***mixture*** model should be used to represent the empirical histogram.
- In our problem, we used ***Gaussian mixture models*** (GMM).



GMM

How does GMM work?

- It approximates ***arbitrary*** probability density functions (PDFs) using a weighted summation of Gaussian density components



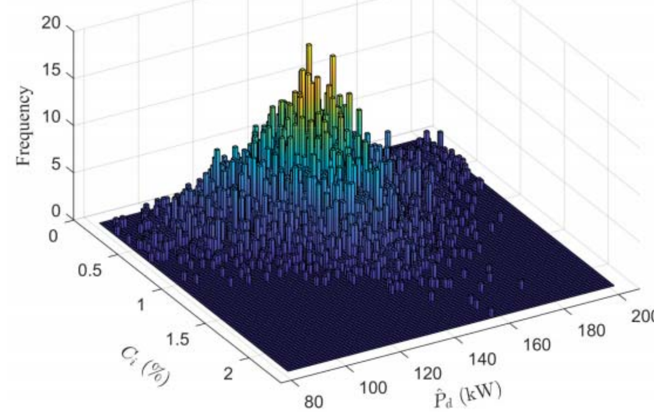
The 1-dimension empirical histogram (bars) is approximated by two Gaussian density components (yellow and green lines)

GMM

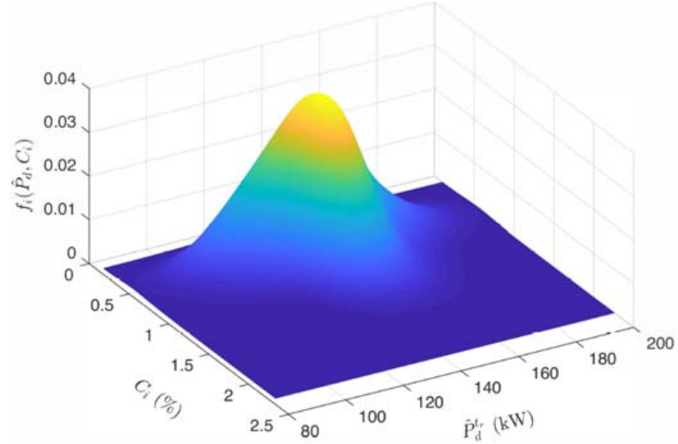
- For our problem, we approximate the joint 2-dimensional PDF of $\hat{P}_d^{tr}$ and C_i, using multiple Gaussian functions

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

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

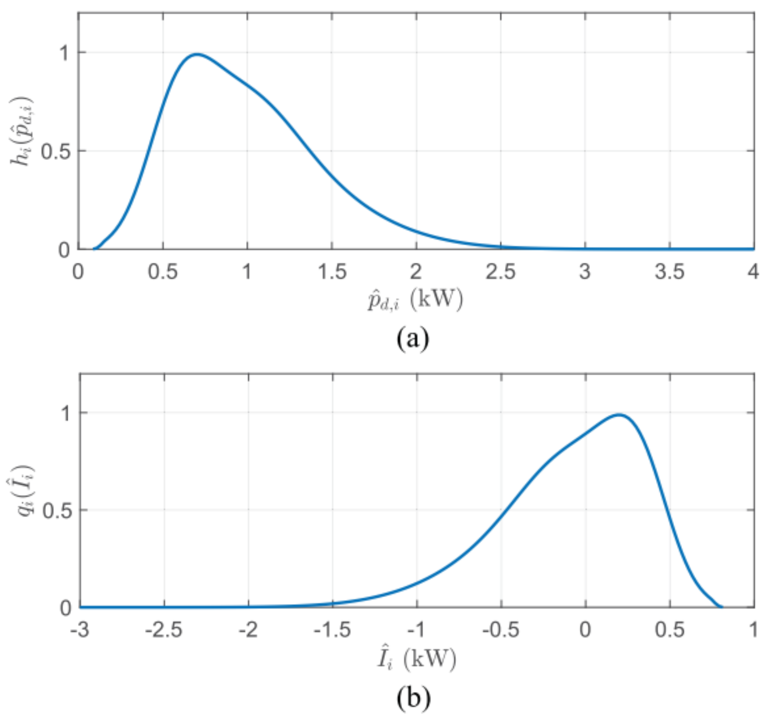
Customer-level Demand Increase Estimation

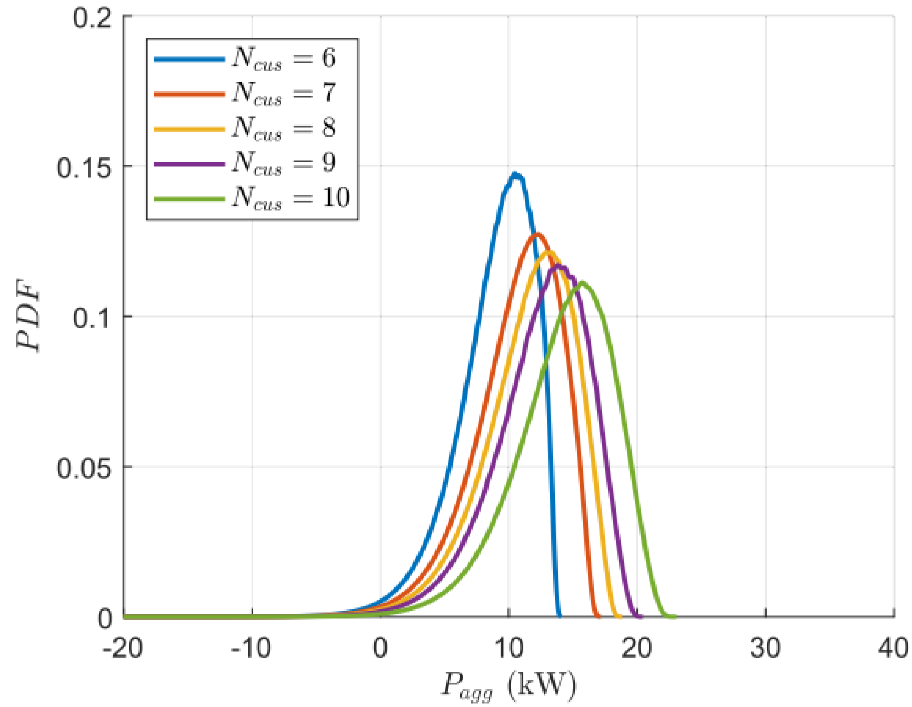
Estimate customer-level normal demand and demand increase?

- Note that the estimated customer-level normal demand, $\hat{p}_d^{t_r}$, is determined by
$$\hat{p}_d^{t_r} = \hat{P}_d^{t_r} \times C_i$$
- Until now, we know the joint distribution of $\hat{P}_d^{t_r}$ and C_i, that is, $f(\hat{P}_d^{t_r}, C_i)$, therefore, we can compute the distribution of the ***estimated customer-level normal demand*** using
$$h(\hat{p}_{d,i}^{t_r}) = \int_{0_+}^{1} f^*\left(\frac{\hat{p}_{d,i}^{t_r}}{C_i}, C_i\right)\frac{1}{C_i}\,dC_i$$
- Finally, the distribution of ***demand increase*** of the *i*th customer is calculated as
$$q(\hat{I}_i^{t_r}) = h\left(p_{u,i}^{t_r} - \hat{I}_i^{t_r}\right)$$

where, $p_{u,i}^{t_r}$ is the actual customer demand at the time of restoration, which is recorded by smart meter.

Customer-level Result

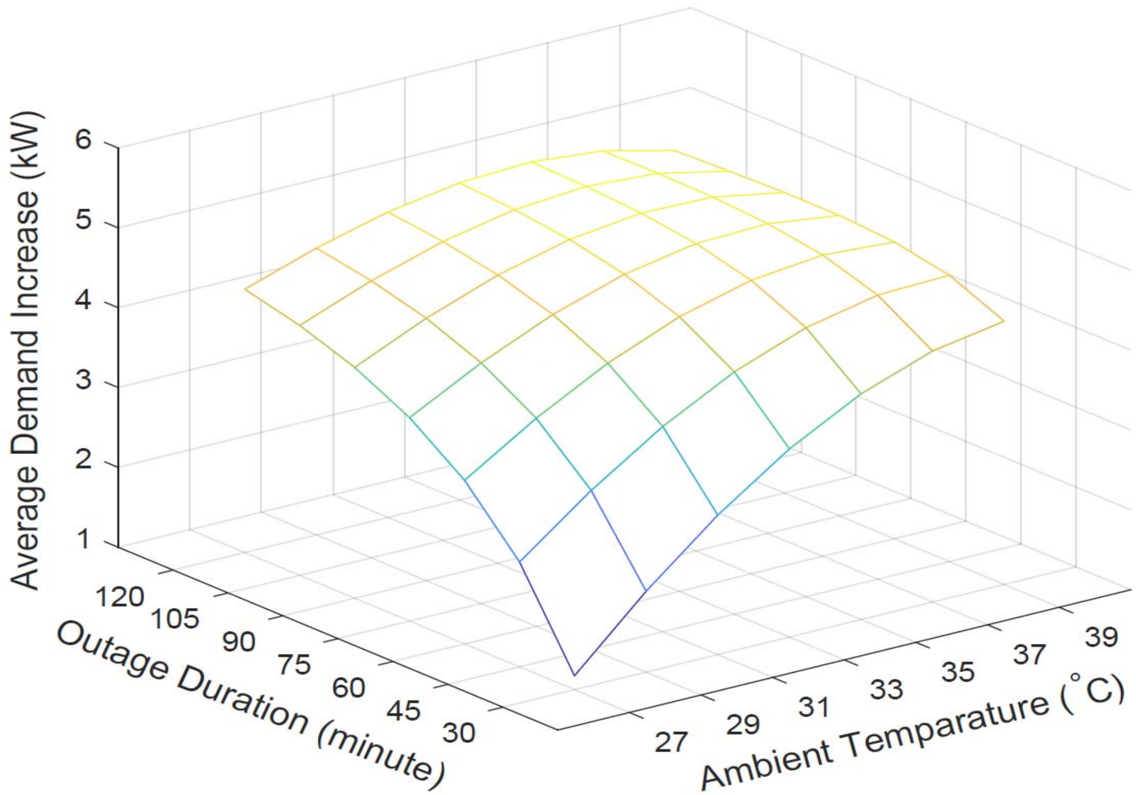




(a) Distribution of estimated customer normal demand. (b) Distribution of customer demand increase.

Distributions of aggregate demand increase of a group with different number of customers

Customer-level Result



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

Contents

Conclusion

- We have archived a large amount of real smart meter data from our utility partners, which provides an unique opportunity to improve distribution grid modeling.

- The development of machine learning techniques provides a useful tool to process and analyze smart meter data.

- A data-driven load inference method is developed for enhancing distribution system observability.

- We have used smart meter data to model the cold load pick up, which can help utilities design restoration plans.

Reference:

[1] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A time-series distribution test system based on real utility data," North American Power Symposium, pp. 1–1, Oct. 2019.
[2] Zhong, Shan, and Ali Abur, "Combined state estimation and measurement calibration," in IEEE Trans. Power Syst., vol. 20, no. 1, pp. 458–465, Feb., 2005.
[3] Y. Wang, Q. Chen, T. Hong and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," in IEEE Trans. on Smart Grid, vol. 10, no. 3, pp. 3125-3148, May 2019.
[4] Zaharchuk G, Gong E, Wintermark M, et al. "Deep learning in neuroradiology," American Journal of Neuroradiology, 39(10), pp.1776-1784, 2018.
[5] Charu C. Aggarwal and Chandan K. Reddy. " Data Clustering: Algorithms and Applications, " 2013.
[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press,2016. http://www.deeplearningbook.org.
[7] National Institutes of Health. "The brain: Understanding neurobiology through the study of addiction," 2007.
[8] Aggarwal, Charu C. "Neural networks and deep learning," Cham: Springer International Publishing, 2018.
[9] J. Bergstra and Y. Bengio, "Random search for hyper-parameter opti-mization,"Journal of Machine Learning Research, vol. 13, pp. 281–305,Feb. 2012.
[10] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan and F. Bu, "A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems," in IEEE Transactions on Smart Grid, vol. 10, no. 2, pp. 2312-2322, March 2019.
[11] H. Wang and N. N. Schulz, " A revised branch current-based distribution system state estimation algorithm and meter placement impact, " IEEE Trans. Power Syst., vol. 19, no. 1, pp. 207–213, Feb. 2004.
[12] Energy Information Administration. (2017) Annual Electric Power Industry Report. [Online]. Available:https://www.eia.gov/electricity/data/eia861/
[13] A. Angioni, T. Schlosser, F. Ponci, and A. Monti, "Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids," IEEE Trans. Instrum. Meas., vol. 65, no. 1, pp. 70–77, Jan. 2016.
[14] Y. Li and P.J. Wolfs,"A hybrid model for residential loads in a distribution system with high PV penetration," IEEE Trans. Power Syst., vol. 28 ,no. 3, pp. 3372–3379, Aug. 2013.
[15] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, and P. Jrventausta,"Enhanced load profiling for residential network customers," IEEE Trans. Power Del., vol. 29, no. 1, pp. 88–96, Feb. 2014
[16] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks," IEEE Trans. Power Syst., vol. 20, no. 2, pp. 548–555, May 2005.
[17] E. Manitsas, R. Singh, B.C. Pal, and G. Strbac, "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," IEEE Trans. Power Syst., vol. 27, no. 4, pp. 1888–1896, Nov. 2012
[18] Y. R. Gahrooei, A. Khodabakhshian, and R. A. Hooshmand, "A new pseudo load profile determination approach in low voltage distribution networks," IEEE Trans. Power Syst., vol. 33, no. 1, pp. 463–472, Jan. 2018.

Reference:

[19] J. A. Jardini, C. M. V. Tahan, M. R. Gouvea, S. U. Ahn, and F. M. Figueiredo, “Daily load profiles for residential, commercial and industrial low voltage consumers,” IEEE Trans. Power Del., vol. 15, no. 1, pp. 375–380, Jan. 2000.
[20] D. T. Nguyen, “Modeling load uncertainty in distribution network monitoring,” IEEE Trans. Power Syst., vol. 30, no. 5, pp. 2321–2328, Sep. 2015.
[21] Tenenbaum, Joshua B., Vin De Silva, and John C. Langford. "A global geometric framework for nonlinear dimensionality reduction." science 290.5500 (2000): 2319-2323.
[22] L. Zelnik-Manor and P. Perona, “Self-tuning spectral clustering,” Proceedings of the 17th International Conference on Neural Information Processing System, pp. 1601–1608, 2004.
[23] U.Luxburg, “A tutorial on spectral clustering,”Statistics and Computing,vol. 17, no. 4, pp. 395–416, Mar. 2007.
[24] C. L. et al., “Levenbergmarquardt backpropagation training of multilayerneural networks for state estimation of a safety-critical cyber-physicalsystem,” IEEE Trans. Ind. Inf, vol. 14, no. 8, pp. 3436–3446, Aug. 2018.
[25] R. Singh, E. Manitsas, B. C. Pal, and G. Strbac, “A recursive Bayesian approach for identification of network configuration changes in distribution system state estimation,” IEEE Trans. Power Syst., vol. 25, no. 3,pp. 1329–1336, Aug. 2010.
[26] 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.
[27] W. W. Lang, M. D. Anderson, and D. R. Fannin, “An analytical method for quantifying the electrical space heating component of a cold load pick up,” IEEE Trans. Power App. Syst., vol. PAS-101, no. 4, pp. 924–932, Apr. 1982.
[28] 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.
[29] V. Kumar, R. Kumar H. C., I. Gupta, and H. O. Gupta, “DG integrated approach for service restoration under cold load pickup,” IEEE Trans. Power Del., vol. 25, no. 1, pp. 398–406, Jan. 2010.
[30] A. Al-Nujaimi, M. A. Abido, and M. Al-Muhaini, “Distribution power system reliability assessment considering cold load pickup events,” IEEE Trans. Power Syst., vol. 33, no. 4, pp. 4197–4206, Jul. 2018.
[31] B. Chen, C. Chen, J. Wang, and K. L. Butler-Purry, “Multi-time step service restoration for advanced distribution systems and microgrids,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 6793–6805, Nov. 2018.
[32] A. Arif et al., “Optimizing service restoration in distribution systems with uncertain repair time and demand,” IEEE Trans. Power Syst., vol. 33, no. 6, pp. 6828–6838, Nov. 2018.
[33] J. C. Lpez, J. F. Franco, M. J. Rider, and R. Romero, “Optimal restoration/maintenance switching sequence of unbalanced three-phase distribution systems,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 6058–6068, Nov. 2018.
[34] J. Aubin, R. Bergeron, and R. Morin, “Distribution transformer overloading capability under cold-load pickup conditions,” IEEE Trans. Power Del., vol. 5, no. 4, pp. 1883–1891, Oct. 1990.
[35] F. Edstrom, J. Rosenlind, K. Alvehag, P. Hilber, and L. Soder, “Influence of ambient temperature on transformer overloading during cold load pickup,” IEEE Trans. Power Del., vol. 28, no. 1, pp. 153–161, Jan. 2013.

Reference:

[36] V. Gupta and A. Pahwa, “A voltage drop-based approach to include cold load pickup in design of distribution systems,” IEEE Trans. Power Syst., vol. 19, no. 2, pp. 957–963, May 2004.
[37] J. J. Wakileh and A. Pahwa, “Optimization of distribution system design to accommodate cold load pickup,” IEEE Trans. Power Del., vol. 12, no. 1, pp. 339–345, Jan. 1997.
[38] R. E. Mortensen and K. P. Haggerty, “A stochastic computer model for heating and cooling loads,” IEEE Trans. Power Syst., vol. 3, no. 3, pp. 1213–1219, Aug. 1988.
[39] C. Chong and R. P. Malhami, “Statistical synthesis of physically based load models with applications to cold load pickup,” IEEE Power Eng. Rev., vol. PER-4, no. 7, pp. 33–33, Jul. 1984.
[40] K. P. Schneider, J. C. Fuller, and D. P. Chassin, “Multi-state load models for distribution system analysis,” IEEE Trans. Power Syst., vol. 26, no. 4, pp. 2425–2433, Nov. 2011.
[41] E. Agneholm and J. Daalder, “Cold load pick-up of residential load,” IEE Proc.—Gener., Transmiss. Distrib., vol. 147, no. 1, pp. 44–50, Jan. 2000.
[42] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, “A survey on state estimation techniques and challenges in smart distribution systems,” IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 2312–2322, Mar. 2019.
[43] M. Espinoza, J. A. K. Suykens, R. Belmans, and B. D. Moor, “Electric load forecasting: Using kernel based modeling for nonlinear system identification,” IEEE Control Syst. Mag., vol. 27, no. 5, pp. 43–57, Oct. 2007.
[44] National Oceanic and Atmospheric Administration, 2018. [Online]. Available: https://www.noaa.gov/
[45] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, Least Squares Support Vector Machines. Singapore: World Sci., 2002.
[46] Z. S. Hosseini,M.Mahoor, and A. Khodaei, “Ami-enabled distribution network line outage identification via multi-label SVM,” IEEE Trans. Smart Grid, vol. 9, no. 5, pp. 5470–5472, Sep. 2018.
[47] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, “A shortterm and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization,” IEEE Trans. Smart Grid, vol. 9, no. 4, pp. 3341–3350, Jul. 2018.
[48] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of smart meter data analytics: Applications, methodologies, and challenges,” IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 3125–3148, May 2019.
[49] A. G. Glen, L. M. Leemis, and J. H. Drew, “Computing the distribution of the product of two continuous random variables,” Comput. Statist. Data Anal., vol. 44, no. 3, pp. 1–14, Jul. 2002.
[50] R. Singh, B. C. Pal, and R. A. Jabr, “Statistical representation of distribution system loads using Gaussian mixture model,” IEEE Trans. Power Syst., vol. 25, no. 1, pp. 29–37, Feb. 2010.
[51] G. Valverde, A. T. Saric, and V. Terzija, “Stochastic monitoring of distribution networks including correlated input variables,” IEEE Trans. Power Syst., vol. 28, no. 1, pp. 246–255, Feb. 2013.
[52] D. A. Reynolds, “Gaussian mixture models,” in Encyclopedia of Biometrics, 2nd ed. New York, NY, USA: Springer, 2015, pp. 827–832.
[53] S. A. Billings, Nonlinear System Identification: NARMAX Methods in the Time, Frequency, Spatio-Temporal Domains. Hoboken, NJ, USA: Wiley, 2013.

Reference:

[54] J. Suykens, J. D. Brabanter, L. Lukas, and J. Vandewalle, “Weighted least squares support vector machines: Robustness and sparse approximation,” Neurocomputing, vol. 48, no. 1, pp. 85–105, Jun. 2002.
[55] S. Bashash and H. K. Fathy, “Modeling and control of aggregate air conditioning loads for robust renewable power management,” IEEE Trans. Control Syst. Technol., vol. 21, no. 4, pp. 1318–1327, Jan. 2013.
[56] V. V. Petrov, Sums of Independent Random Variables. New York, NY, USA: Springer-Verlag, 1975.
[57] J. Beirlant, E. Dudewicz, L. Gyorf, and E. van der Meulen, “Nonparametric entropy estimation: An overview,” Int. J. Math. Statist. Sci., vol. 6, pp. 17–39, Jul. 2001.

Thanks!

Q & A