By Yuxuan Yuan and Zhaoyu Wang

Mining Smart Meter Data to Enhance Distribution Grid Observability for Behind-the-Meter Load Control

Significantly improving system situational awareness and providing valuable insights.

DISTRIBUTED ENERGY RESOURCES (DERs) are playing an increasingly important role in power systems. In 2023, five categories of DERs—distributed solar, electric vehicles (EVs), energy storage, residential smart thermostats, and small-scale combined heat and power—are expected to contribute about 104 GW to the U.S. summer peak (see GTM, 2018). With the increasing integration of DERs in power distribution systems, distributed load control is imperative to smooth the fluctuations that they introduce. However, a main challenge is that distribution systems lack systematic situational awareness because of their limited sensors. Furthermore, most customer-level behind-the-meter (BTM) DERs, such as rooftop photovoltaics (PVs), are being integrated into distribution systems, which complicates the system monitoring and control. Enhanced electric grid monitoring is needed to promote renewable integration while ensuring reliability, but current approaches rely on expensive sensors.

BACKGROUND—©SHUTTERSTOCK.COM/OMELCHENKO, METER—©SHUTTERSTOCK.COM/BET_NOIRE

Digital Object Identifier 10.1109/MELE.2021.3093636
Date of current version: 2 September 2021

2325-5897/21©2021IEEE

In recent decades, the deployment of advanced metering infrastructure (AMI) in distribution systems provides a unique opportunity to extend the monitoring capability to the grid edges. AMI is a configured infrastructure of smart meters (SMs), meter data management systems (MDMSs), and communication networks, which enables two-way communication between customer meters and utilities. The introduction of AMI provides utilities with many features that were previously unreachable or had to be accomplished manually, thus significantly reducing labor costs. Based on an analysis from the U.S. Department of Energy, AMI can help medium-to-large-sized utilities save an average of US$16.6 million in operation and maintenance costs over three years (see Office of Electricity Delivery and Energy Reliability, 2016). Moreover, widespread AMI has enabled utilities to collect an unprecedented amount of demand-side data that facilitate the transition to a data-enabled modernized power system. However, most electric utilities use AMI data only for billing. The challenge is that, without new computing innovations, SMs can provide only limited insights into grid performance.

In this article, an overview of AMI is presented, including concept, communications, and current applications. Then we introduce several advanced applications that allow unlocking the untapped potential of AMI data using machine learning techniques. The proposed solutions can significantly improve system situational awareness and provide valuable insights to better control BTM loads and DERs.

An Overview of AMI

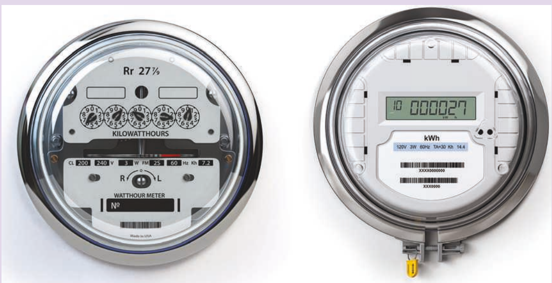
AMI and SM Data Introduction

The Federal Energy Regulatory Commission (FERC) defines AMI as "a metering system that records customer consumption and possibly other parameters hourly or more frequently and that provides for daily or more frequent transmittal of measurements over a communication network to a central collection point" (see FERC, 2020). AMI is developed on the basis of automatic meter reading (AMR). AMR is an older technology and can avoid the need for staff to manually record monthly energy consumption data. Compared to AMR, AMI is more expensive, but it offers more benefits. The core element of AMI is the SM, which is a device installed at a customer's house or facility. As shown in Figure 1, unlike conventional electromechanical meters that rely on a series of dials to record the total energy consumption, SMs use an LCD screen to show customer usage. The energy consumption reading of the SM is accumulative, and periodic usage is determined by subtracting the current reading from the previous one. SMs often use a meter multiplier to calculate the actual kilowatt-hour consumption. The multiplier is preceded by an "X" and marked on the front of the SM. Thus, the monthly usage times the multiplier is used to calculate the monthly bill.

Compared to AMR meters, which only record monthly energy data, SMs for single-phase residential or small commercial customers can typically record real energy consumption (kWh) and the instantaneous voltage magnitude (V) at 15-, 30- or 60-min intervals. For three-phase large commercial and industrial customers, utilities typically use a 15-min meter-reading interval to collect the real energy consumption (kWh), reactive energy consumption (kVArh), and instantaneous voltage magnitude (V) for each phase. For some large-scale industrial customers who operate sensitive machinery, SMs can be activated for measuring current transients and harmonics. Figure 2 shows an example of SM data. Each customer has an account number, and its energy usage data are recorded at each time stamp. In addition to the usage data, SMs can monitor the energized/de-energized status of customers. When an SM realizes that it is going to lose power, it sends a "last-gasp signal" to utilities for outage notification. Furthermore, when a customer calls to report a suspicious outage, the SM provides a meter-pinging function to determine if the customer has actually lost power, which can eliminate time-consuming truck rolls to verify power outages.

SM Data Communication and Storage

To provide near-real-time information, data communication is a critical technical requirement. AMI communication networks need to deliver accurate, reliable, and massive data streams in a timely manner. In the United States, the ANSI C12.18 standard defines a table structure for passing data between an SM and a utility. As shown in Figure 3, a typical AMI communication network has two layers. The first layer connects data concentrators with a head-end system (HES). An HES is hardware and software that can receive and transmit data and store short-term consumption data to support customer billing. The second layer links multiple neighboring SMs with a data concentrator. The AMI communication network can consist of either wireless or fixed-wired connections. Fixed-wired connections include power line carriers, fiber-optic cables, telephone dial-up modems, and digital subscriber lines. Wireless communication options include cellular



(a) (b)

Figure 1. Examples of (a) an electromechanical meter and (b) an SM.

Account		Time	KWH or V	Time	KWH or V	Time
100000001	KWH	201704010100	0.392	201704010200	0.257	201704010300
100000001	VOLTS	201704010100	239	201704010200	239	201704010300
100000002	KWH	201704010100	0.245	201704010200	0.204	201704010300
100000002	VOLTS	201704010100	241	201704010200	240	201704010300
100000003	KWH	201704010100	1.479	201704010200	0.417	201704010300
100000003	VOLTS	201704010100	240	201704010200	239	201704010300
100000004	KWH	201704010100	1.009	201704010200	0.555	201704010300
100000004	VOLTS	201704010100	241	201704010200	237	201704010300
100000005	KWH	201704010100	0.798	201704010200	0.809	201704010300
100000005	VOLTS	201704010100	239	201704010200	238	201704010300
100000006	KWH	201704010100	0.109	201704010200	0.188	201704010300
100000006	VOLTS	201704010100	241	201704010200	240	201704010300
100000007	KWH	201704010100	1.199	201704010200	1.512	201704010300
100000007	VOLTS	201704010100	241	201704010200	240	201704010300
100000008	KWH	201704010100	0.422	201704010200	0.419	201704010300
100000008	VOLTS	201704010100	239	201704010200	239	201704010300
100000009	KWH	201704010100	2.288	201704010200	2.278	201704010300
100000009	VOLTS	201704010100	243	201704010200	242	201704010300
100000010	KWH	201704010100	0.223	201704010200	0.257	201704010300
100000010	VOLTS	201704010100	242	201704010200	241	201704010300

Figure 2. Exemplary SM data from utilities.

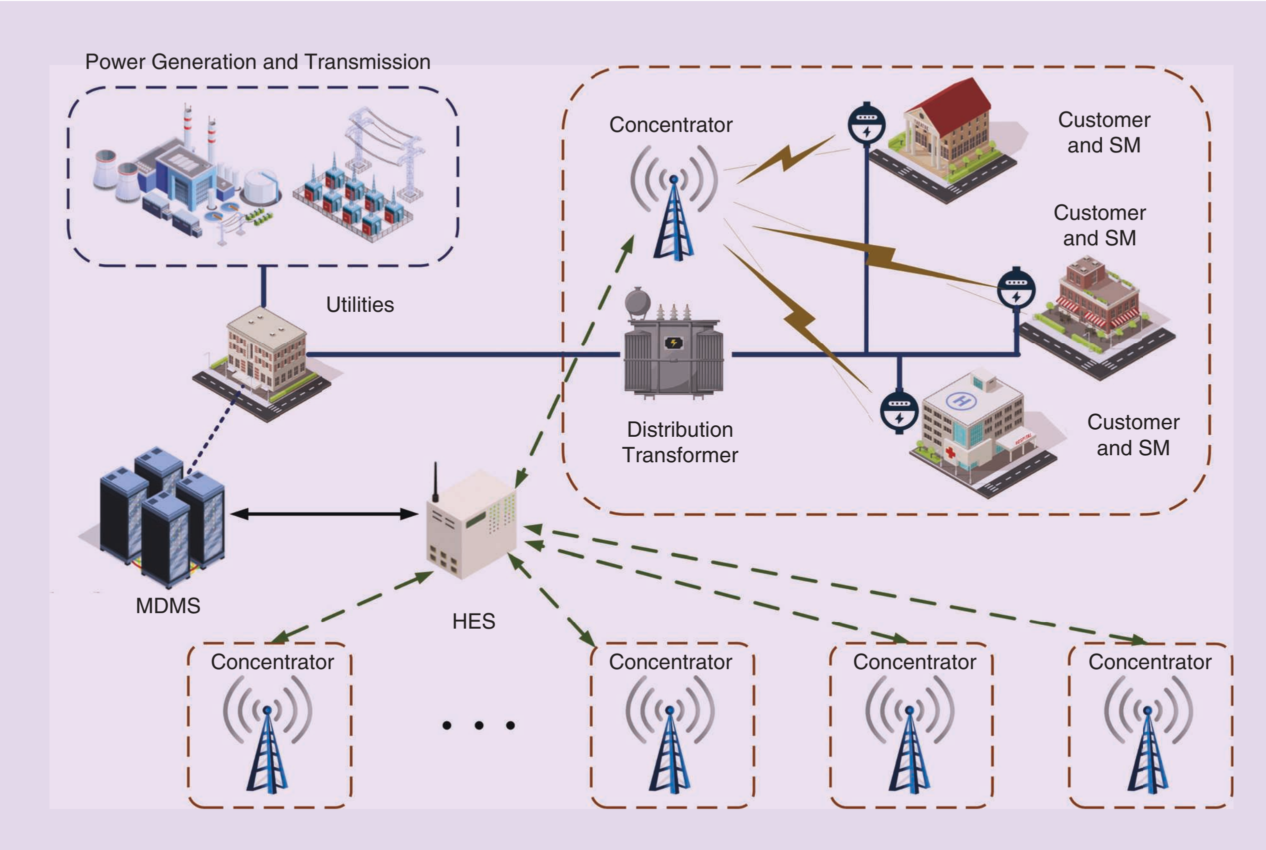


Figure 3. SM data collection and communication. HES: head-end system.

communications, Wi-Fi, low-power, long-range wireless networks, Zigbee, and Wi-SUN. Currently, there is no single communications solution that is optimal for all utilities. Utilities normally combine multiple technologies to build their communication networks. The selection of communication techniques is highly dependent on the geographic location and grid infrastructure. For example, urban utilities face very different communication challenges than rural utilities.

When SM data are collected by the HES, an MDMS is used for long-term data storage and management. The MDMS will import SM data and then validate, clean, and process them before using them for utility applications, such as geographic information systems and outage management systems. In the United States, not all utilities have their own MDMS because of financial constraints; small utilities prefer partnering with third-party data companies to store and process their SM data.

SM Data Applications

In the United States, federally sponsored programs have been promoting the widespread adoption of AMI in distribution systems. But, to date, most utilities use SMs for remote billing and outage notifications only, without exploring insights or gaining actionable information, which significantly undervalues the SM data. The fundamental reason is that SM data are limited by several disadvantages. Specifically, unlike the microphasor measurement units (μPMUs) that provide high-resolution (e.g., 120 samples/s) and synchronized phasor data, SMs collect only low-resolution and unsynchronized energy and voltage magnitude measurements. Furthermore, SM data suffer from data-quality problems caused by communication failures, meter malfunctions, and human errors. Figure 4 presents several common SM data-quality issues, including duplicates, constant zero readings, missing data, and outliers.

Despite the data-quality difficulties, SM data are a good resource for enhancing distribution grid monitoring and control, thanks to extensive customer-side installations. Compared to μPMUs, which are installed only on critical nodes in distribution systems, an SM can be installed for each customer. According to the U.S. Energy Information Administration, SM installations have grown dramatically since 2011. By the end of 2020, an estimated 107 million SMs were developed, which covers nearly 75% of the total U.S. households. As the number of SMs increases, massive customer-side data are collected by utilities. For the utilities that have millions of observable customers, i.e., customers with SMs, the total SM data with 15-min resolution can reach 10 TB per year. Such a massive data set contains rich information on customer consumption behaviors and system operation insights, which provides an opportunity to perform data-driven local control. The next section will introduce several advanced applications of SM data enabled by machine learning techniques, which can greatly improve distribution grid modeling and operation, thus facilitating the BTM load control.

Advanced Applications of SM Data Enabled by Machine Learning Techniques

With the advent of machine learning technologies, researchers now have powerful tools to further exploit the undiscovered values of SM data. Figure 5 summarizes some existing and emerging applications of SM data enabled by machine learning. This section focuses on four of these topics, including load profiling, demand-side flexibility estimation, BTM solar disaggregation, and grid modeling.

Customer Load Profiling

Customers' typical load profiles are valuable for utilities in understanding customer consumption behaviors. With typical load profiles, each customer can be linked to one of the predefined classes. This information is utilized extensively in distribution system operations, such as rate design, distribution system state estimation, load forecasting, and network planning. The traditional approach to obtain typical load profiles is to simply average the consumptions of residential, commercial, and industrial customers, respectively. However, this method ignores different socioeconomic factors and weather conditions, thus reducing the profile representativeness. By using

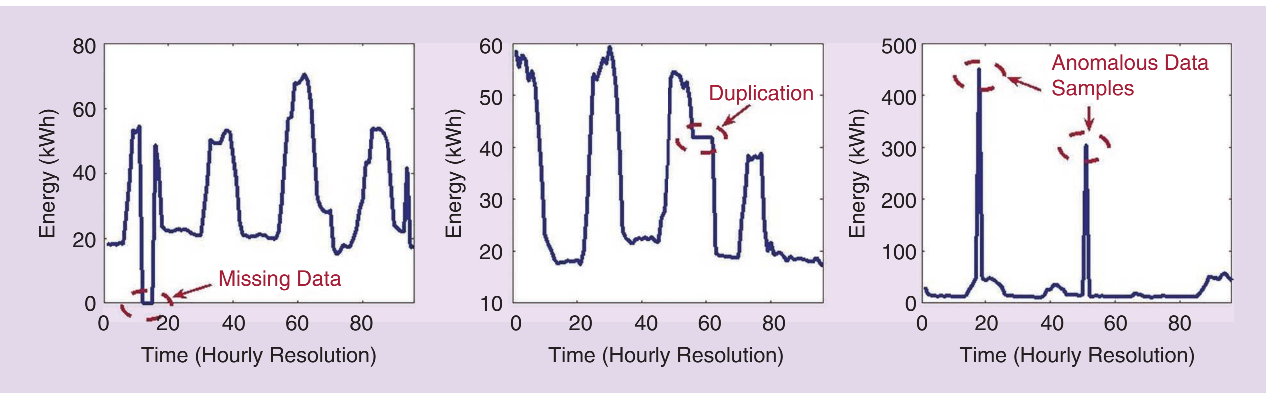


Figure 4. Exemplary SM data-quality problems.

machine learning techniques, load profiling can be cast as an unsupervised clustering problem and solved by multiple well-developed algorithms, such as k-means and hierarchical clustering. There are several challenges in performing load profiling by using these algorithms. First, most clustering algorithms suffer from the curse of dimensionality posed by time-series SM data because of the divergence of the Euclidean distance in a high-dimensional space. The second issue involves some algorithm-specific limitations. For example, hierarchical clustering is sensitive to outliers. The k-means algorithm can only handle clusters with spherical or ellipsoid symmetry, which is not necessarily true of SM data sets. The third challenge is to find the optimal values of hyperparameters, such as the number of typical load profiles and scaling factors in kernel functions.

To address these challenges, we have utilized a graph-based clustering approach called *spectral clustering* to classify the typical load profiles. The proposed method is robust for high-dimensional SM data since it uses the distance on a graph rather than the Euclidean distance. Moreover, to avoid the manual selection of hyperparameters, a self-tuning strategy has been applied to calculate the k-nearest neighbors-based local distance for each customer. The Davies–Bouldin validation index has been leveraged to determine the optimal number of typical load profiles. Figure 6 shows 22 seasonal typical load profiles with their proportions. These results are based on three-year SM data from about 3,000 residential customers in the midwestern United States. As shown in the figure, in spring, about 38% of customers have a peak demand in the evening (8 p.m.). In summer, the customer load behavior is different; most customers show similar behavioral tendencies, and the peak demand occurs during the afternoon interval (from 4 p.m. to 6 p.m.). A possible reason is the higher energy consumption from air conditioning in time periods with higher temperatures. Fall has more typical load profiles than other seasons because of the greater variability in customer behaviors. Based on our observations, about 35% of customers have double peaks, one in the morning (from 7 a.m. to 9 a.m.) and the other in the evening periods (from 6 p.m. to 10 p.m.), such as C12, C13, and C14. In winter, customer consumption behaviors are similar to those in spring. This is probably because the two seasons have a similar meteorology in the midwestern United States.

Demand-Side Flexibility Quantification

System peak demand is one of the critical concerns for utilities. At peak times, the marginal cost of energy procurement can be 10 times higher than for off-peak times. Flexibility quantification refers to estimating the portion of system peak demand that can be reduced or shifted. In practical systems, utilities use the customer daily peak demand to approximate flexibility for peak-shaving programs. However, this approximation can cause a considerable error since individual customers' peak demands do

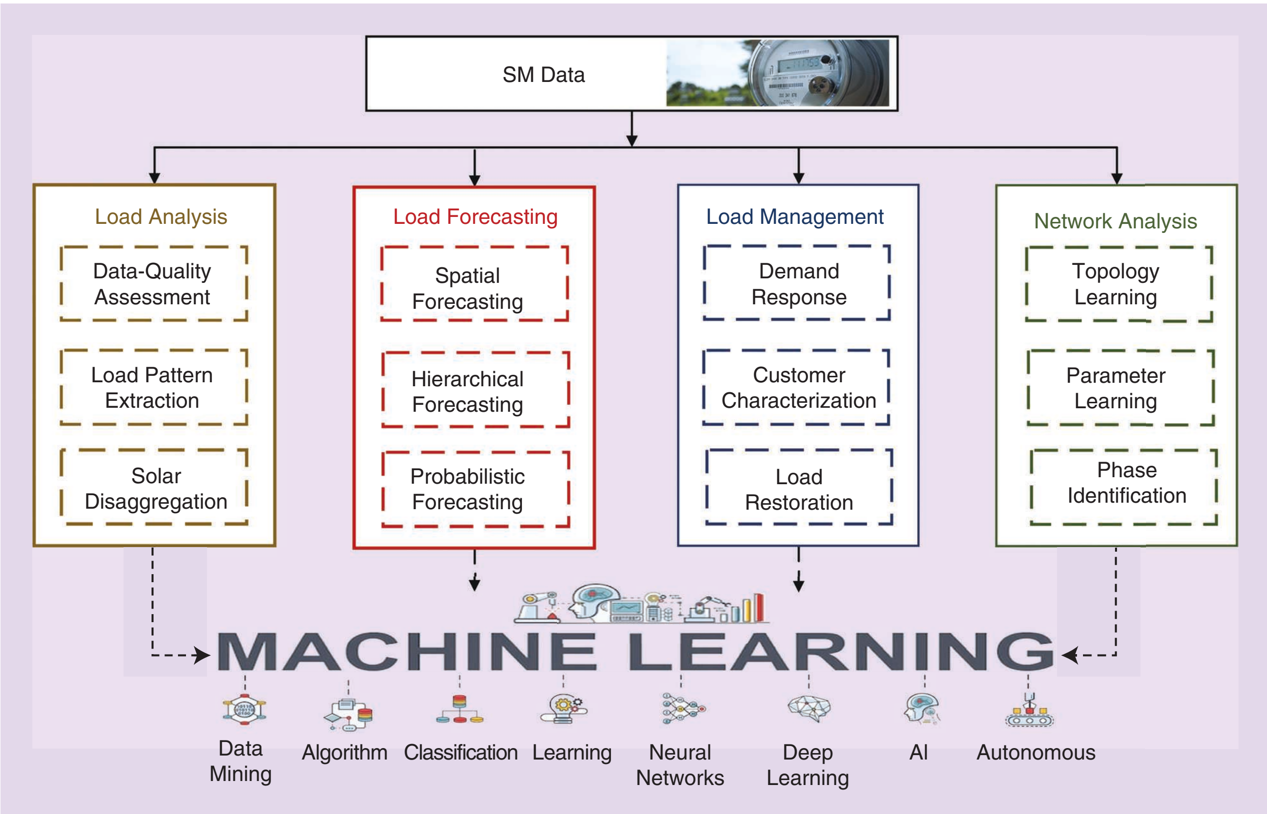


Figure 5. A summary of advanced SM data applications. AI: artificial intelligence.

not necessarily coincide with the system peak, which poses a challenge to BTM load control. A better solution is to calculate the ratios of individual customers' demands during the daily peak load times of the system to the daily system peak demand, which is called the *coincident monthly peak contribution* (CMPC). For observable customers with SMs, the CMPC can be directly computed based on real-time SM data and system demand data recorded by a supervisory control and data acquisition system. For unobservable customers without SMs, a weighted cluster-wise regression method can be used to estimate the CMPC using their monthly billing information. A flowchart of this method is described in Figure 7. The basic idea is to exploit the strong correlation between the CMPC and monthly energy consumption when the customers' load profiles are similar. Based on the validation of a real SM data set, the estimated values can accurately track a customer's real CMPC, as shown in Figure 8.

BTM Solar Disaggregation

To date, most residential rooftop PVs are installed BTM. Hence, as described in Figure 2, SMs can only record a

Figure 6. (a) Seasonal typical load profiles. (*Continued*)

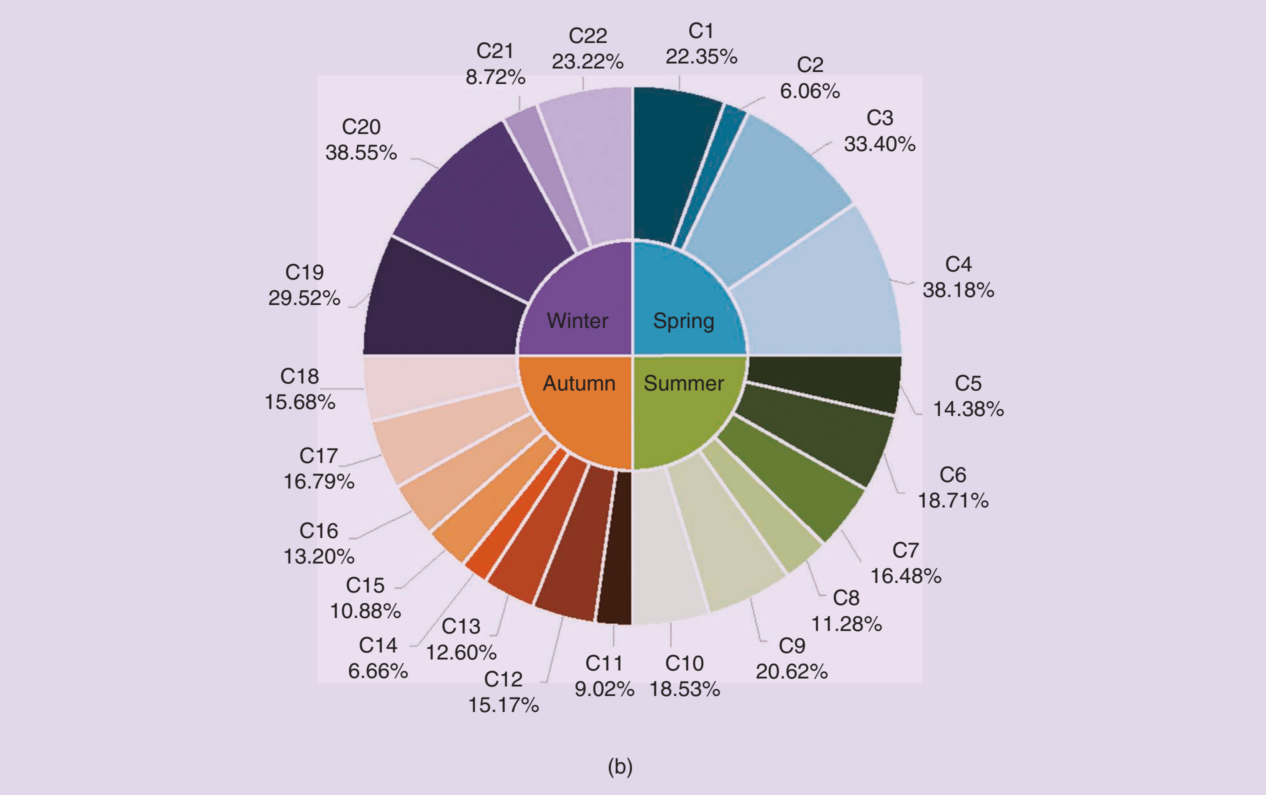


Figure 6. (*Continued*) (b) Proportions of typical load profiles.

customer net demand that equals the native load consumption minus the PV generation. While customers have SMs, PV generation and customer native demand remain invisible to utilities. This invisibility greatly hinders the implementation of distributed load control. One simple solution is to install additional PV meters for monitoring each individual PV, which is cost-prohibitive. As an alternative low-cost solution, researchers have investigated solar disaggregation methods to infer customer native load consumption and solar generation, which can be classified into two groups: model-based methods and data-driven methods. Model-based methods use multiple PV parameters, such as location, size, inverter efficiency, tilt, and orientation, along with weather information to estimate solar generation. The customer native demand then can be obtained by subtracting the estimated solar generation from SM data. Nevertheless, PV parameters can be incomplete or unavailable in practice, which makes the actual implementation of these methods costly. In contrast, data-driven methods do not require physical parameters and rely only on historical solar generation and customer consumption data to build the mapping functions. The main issue with the existing data-driven models is the availability of historical solar data, especially for BTM systems.

To address these shortcomings, a probabilistic learning-based disaggregation model has been proposed to estimate the customer demand and PV generation without using a historical solar data set. The flowchart of this method is described in Figure 9. The basic idea is to exploit the temporal correlation between nocturnal ($P_{m,n}$) and diurnal native demands ($P_{m,d}$) and the spatial correlation between unknown BTM PVs and solar examples in the same distribution system. The temporal correlation is modeled as a joint data distribution using a Gaussian mixture model (GMM). Three typical solar exemplars, including PVs facing east, south, and west, are utilized, demonstrating distinguishable features, as shown in Figure 10. Figures 11 and 12 present the disaggregated native demand and PV generation curves, respectively, for one customer over two weeks using the proposed method. It can be observed that the disaggregated curve closely tracks the actual profile, regardless of the consumption volatility on some days.

Topology and Parameter Identification

A complete and accurate system model is essential for modern distribution system operations and BTM load control. However, there are still many utilities that do not have any digital system models and rely on paper-based maps to do their work. These maps are usually outdated and incomplete because of frequent system expansions and DER installations. This lack of system knowledge inhibits effective system monitoring and control. One approach for solving this problem is to widely perform field inspections to verify system connections and line parameters. This manual solution is labor intensive and

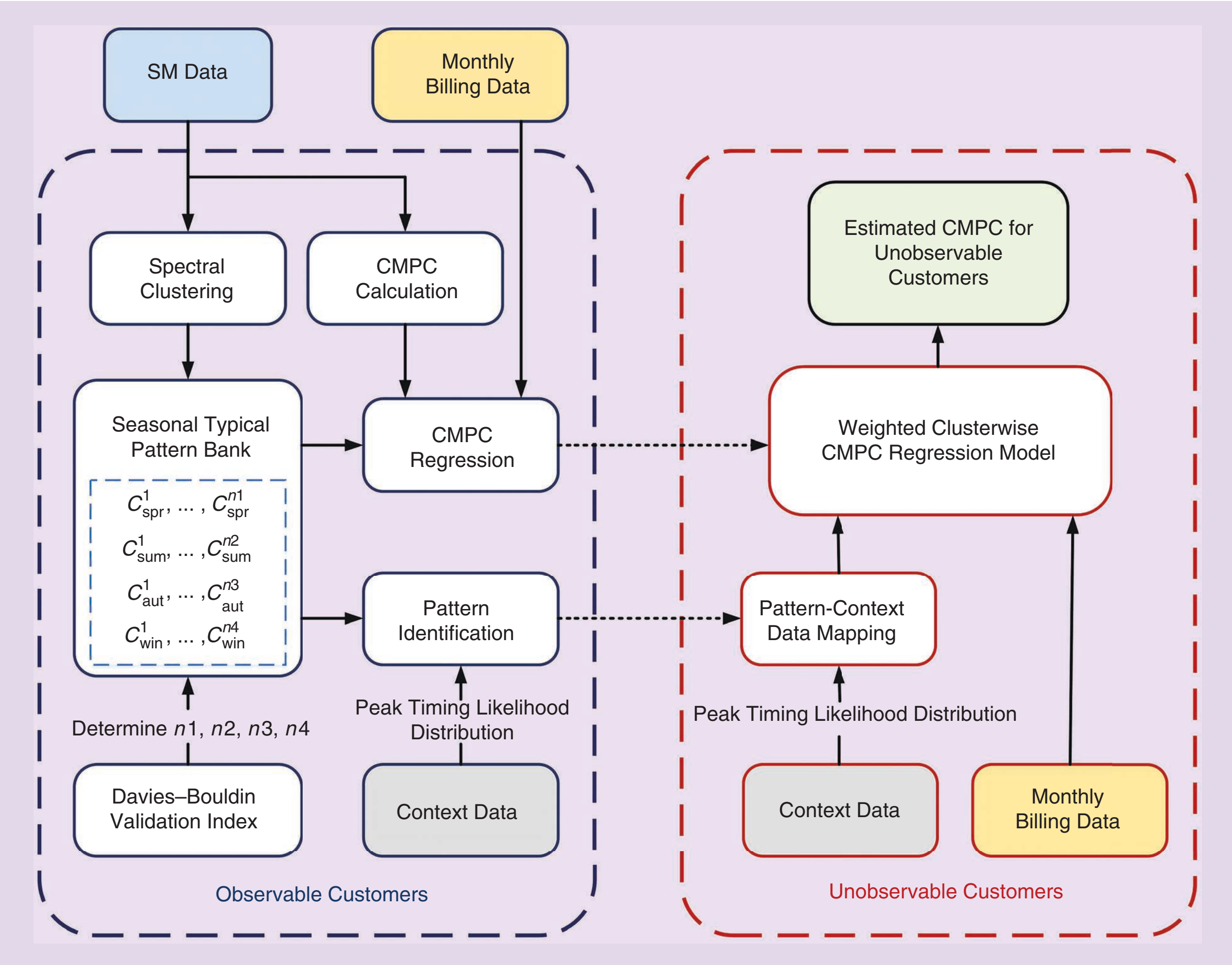


Figure 7. The flowchart of a weighted clusterwise regression model.

time consuming, especially for large-scale systems with thousands of overhead lines and underground cables. Hence, some studies have made efforts to explore data-driven approaches to capture the inherent dependencies among field measurements for topology and parameter identification. Specifically, the term *topology identification* refers to finding the connectivity of different nodes in entire networks. The term *parameter identification* means calculating the resistance (R) and reactance (X) of each branch. In the literature, existing data-driven methods generally rely on full coverage of μPMUs and the availability of prior knowledge of the network, such as the R/X ratios of all branches. The reason for these assumptions is that a μPMU can provide high-granular phase angle information, while the prior knowledge of the network can significantly reduce the search space of the optimization process. Nevertheless, these assumptions are not necessarily applicable to practical distribution systems, especially for old systems that need to perform topology and parameter identification. To tackle these shortcomings, a novel two-stage data-driven framework has been proposed to identify network topology and parameters separately, using only SM measurements and a low-cost conductor library.

In the first stage, the procedure of identifying topology from SM data is cast as a graph theory problem: how to approximately estimate the weighted Laplacian matrix of the network. The basic idea is that the

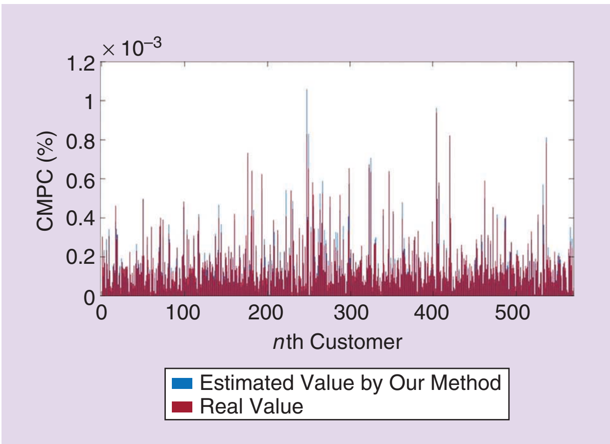


Figure 8. A comparison of the estimated CMPC and the actual value.

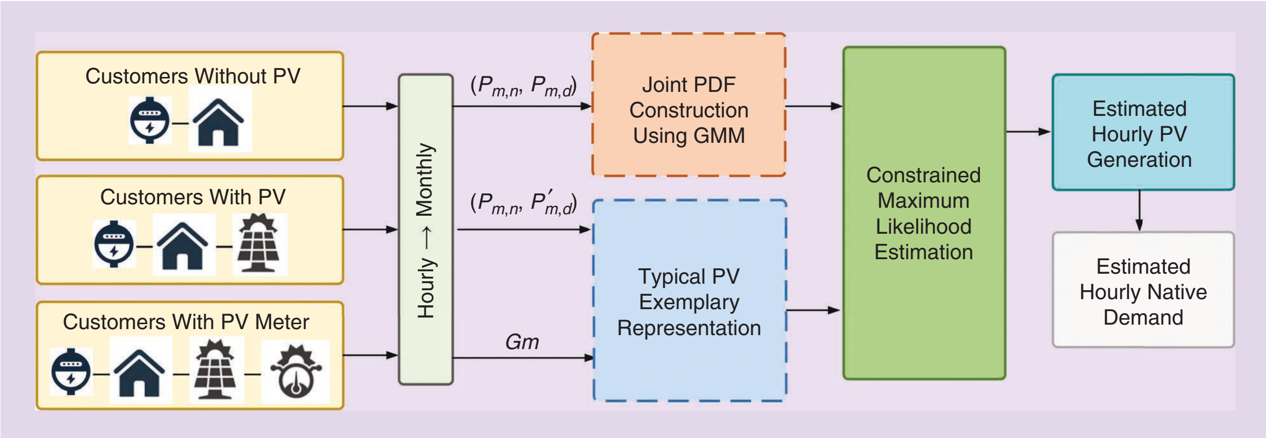


Figure 9. The flowchart of a BTM solar disaggregation method. GMM: Gaussian mixture model.

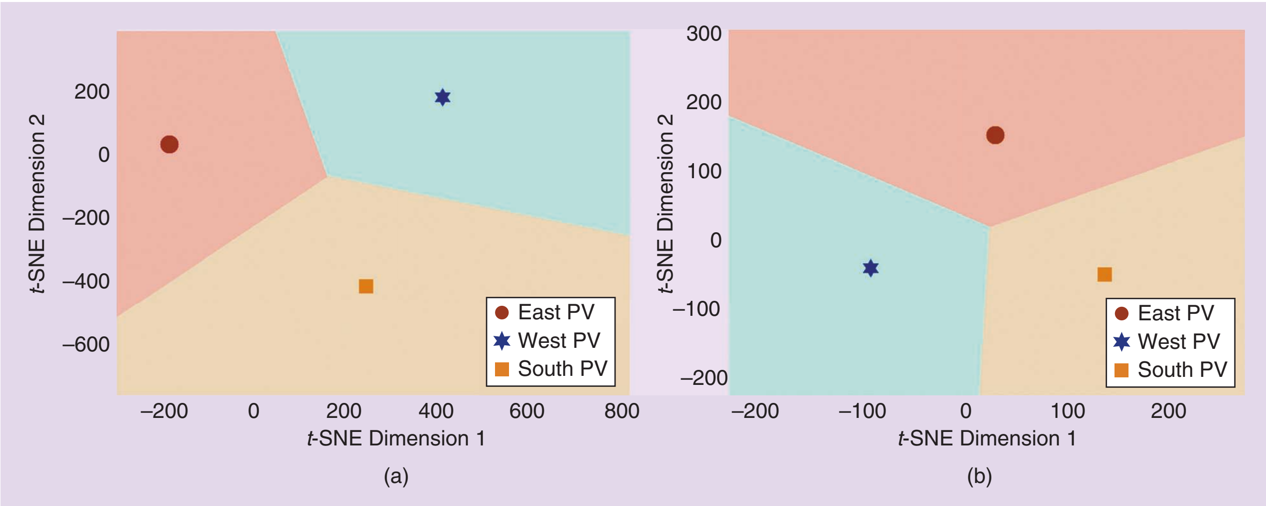


Figure 10. Visualizing the distinguishability of time-series PV generation curves of solar exemplars (a) hourly and (b) monthly. t-SNE: t-distributed stochastic neighbor embedding.

network connection information can be easily extracted by exploiting the unique property of the weighted Laplacian matrix. Furthermore, to relax the phase angle information, the branch-flow model is utilized to formulate our model. The proposed topology identification method has been tested on the IEEE 13- and 37-bus test feeders shown in Figure 13. The time-series nodal load demand is obtained from our real-world SM data sets. The voltage measurements are calculated using a power-flow analysis. The estimated weighted Laplacian matrix of each feeder is described in Figure 14. In the figure, the absolute value of each entry in the matrix is represented by the edge width. As can be observed, the connectivity of different nodes is obtained by distinguishing entries close to zero and large nondiagonal entries. Based on 500 Monte Carlo simulations, the proposed method can achieve

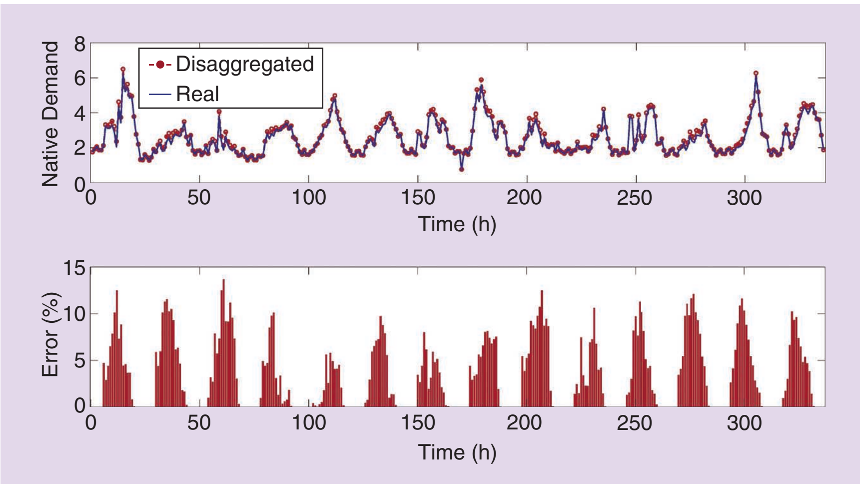


Figure 11. Two-week disaggregated native demand curves, along with the corresponding actual curves.

100% accuracy on 13- and 37-bus test feeders.

In the second stage, a novel bottom-up sweep parameter identification method has been developed to estimate R and X of each branch. The unique feature of our solution is that it does not infer parameters in a system-wide fashion but performs branch-level parameter estimation in an alternating manner. This alternating strategy solves the dimensionality issue and enables parallel computation of all line sections, thus providing good scalability for large distribution grids. To achieve this, a radial distribution network is first decomposed into multiple layers, as shown in Figure 15. Then, the proposed bottom-up sweep algorithm begins with the last branches at the bottom layer and estimates line parameters and power flows layer by layer in an alternate manner. For each branch, the line impedance estimation establishes the voltage drop using the nonlinear branch-flow model. The flowchart is demonstrated in Figure 16.

To narrow down the search space and handle the ill-conditioning issue, a library of R/X ratios is added as a constraint in the single branch parameter estimation. This library only requires knowledge of the types of conductors in a system without knowing the exact R/X ratio of each branch. Thus, information can easily be found in utility inventory records or guidance for distribution systems at a particular voltage level. Moreover, since the original optimization is nonconvex, we have applied the Big M technique and semidefinite programming relaxation to tackle the bilinear term and nonconvex quadratic equalities. After the relaxation, the single branch parameter estimation can be modeled as two optimization models: a nonlinear least absolute deviations (LAD) model and a nonlinear least squares (LS)

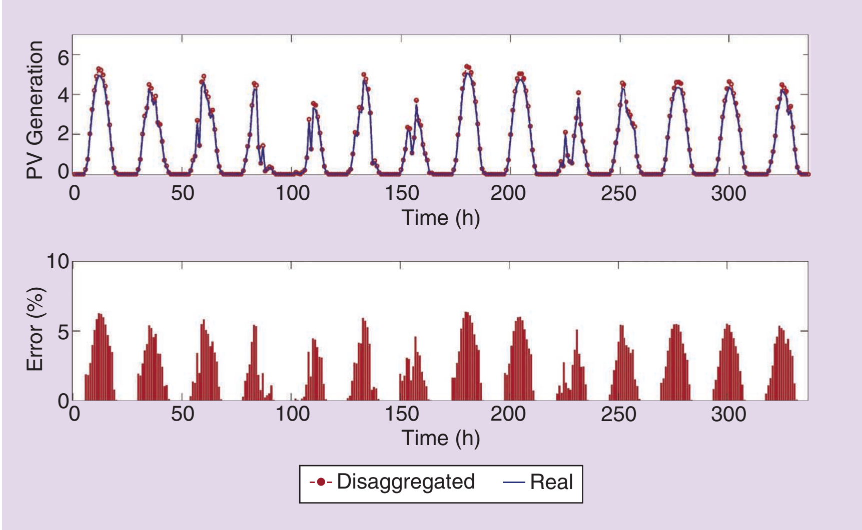


Figure 12. Two-week disaggregated PV generation curves, along with the corresponding actual curves.

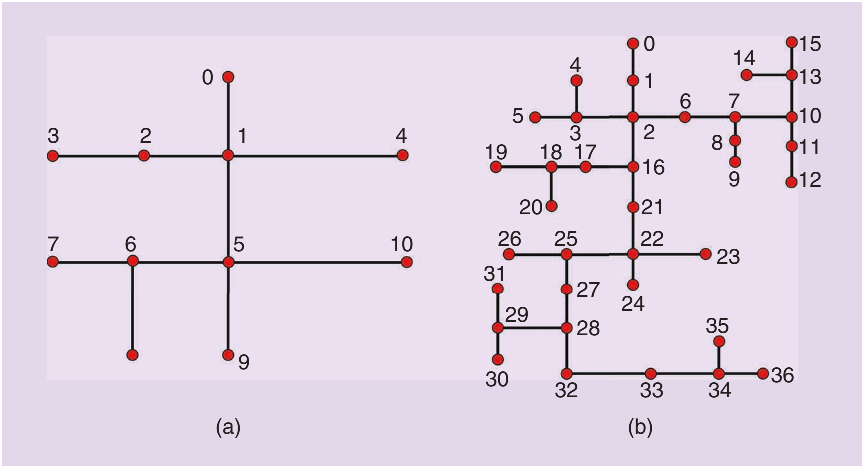


Figure 13. One-line diagrams of two test feeders: (a) IEEE modified 13-bus feeder and (b) IEEE 37-bus feeder.

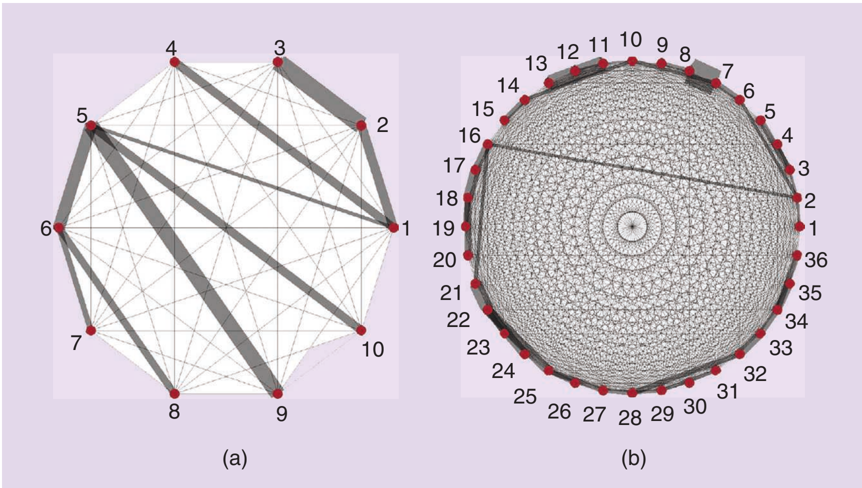


Figure 14. The estimated weighted Laplacian matrices of two test feeders using the proposed topology identification model: (a) IEEE modified 13-bus feeder and (b) IEEE 37-bus feeder.

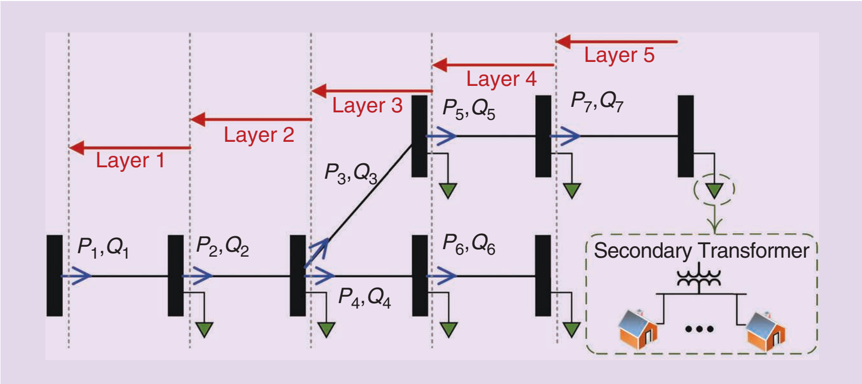


Figure 15. An example of a five-layer distribution network. P and Q denote the real and reactive power flowing from the upstream bus.

model. The performance of our solution has been validated on IEEE 13-bus and 37-bus test feeders. The results are depicted in Figures 17 and 18. It can be observed that the proposed model precisely recovers the line impedance of each branch under all test cases. In terms of R, the largest relative errors are less than 0.001%; as for X, the largest relative errors are less than 0.0005%.

Open Source SM Data Set

Because of data privacy and concerns, most utilities are hesitant to share their system models and SM data with the public, which poses a great challenge to many researchers. To bridge this gap, with permission from our utility partner, we have shared a real distribution grid model with one-year SM data for researchers and engineers to perform validation and demonstration. The system consists of three feeders, 240 primary buses, and multiple grid components, such as substation transformers with load tap changers, secondary distribution transformers, and capacitor banks. There are 1,120 customers, and all of them are equipped with SMs to record hourly real energy consumption (kWh). One can download the data set at http://wzy.ece.iastate.edu/Testsystem.html, including the system description, SM data, OpenDSS model, and MATLAB code for quasi-static time-series simulation.

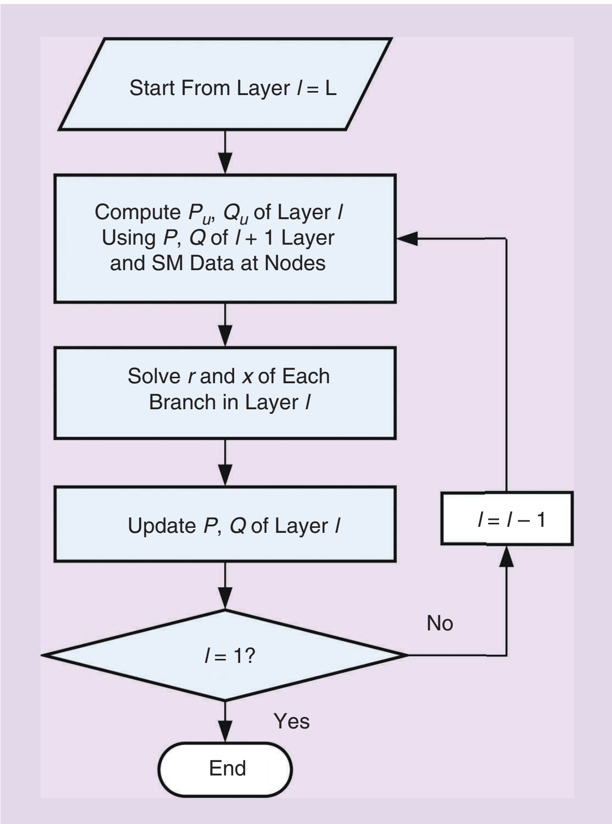


Figure 16. The flowchart of a bottom-up alternating method.

Conclusion and Future Work

With the increasing penetration of DERs, distribution systems are gradually transforming from traditional grids to smart grids. Utilities need to improve systematic situational awareness to execute BTM load-control strategies. Although SM data may be of relatively low resolution and suffer from various data-quality problems, they can still provide valuable insights for utilities. In this article, we have discussed AMI data structure, communication, and several advanced applications enabled by machine learning, including typical load profiling, demand-side flexibility quantification, BTM solar disaggregation, and topology and line parameter identification. Previous works and our proposed solutions on these topics have been discussed and presented.

In 2020, investments in distribution systems were estimated to exceed US$41.8 billion. It is foreseeable that the increasing development of distribution systems will inevitably bring new challenges to BTM load controls and system operations. Hence, to ensure a reliable and resilient electricity supply, we have envisaged several research directions.

1) New sensor technologies, such as µPMUs, have recently been introduced to enhance real-time distribution system monitoring. These sensors can provide high-resolution voltage and current phasors not available from SMs. It will be interesting to investigate how the data from such sensors can be used to facilitate SM data analytics.
2) As the number of EVs continues to rise, the impact of EV charging on distribution grids is increasing. By combining EV charging profiles and SM data, the impact of the uncoordinated charging of EVs on distribution grids can be studied in detail. Moreover, thanks to real-time measurement and two-way communication capabilities, SMs have the potential to assist utilities in effectively managing various EV charging facilities.

3) Microgrids provide a promising solution to manage the power of different distributed generation. However, traditional mathematical model-based control schemes are not necessarily applicable to practical microgrids. Accurate and complete model information is difficult to obtain because of the widespread uncertain dynamics and disturbances in microgrids. Consequently, for the future smart grid, it is of great significance to explore data-driven adaptive control by taking advantage of a large amount of SM data.
4) A recent survey of 1,000 utilities in 10 countries showed that about 80% of utilities realize big data problems as crucial for smart grids. Current data-driven models are mainly conducted on megabit or gigabyte data sets, which may not be suitable for big data. In the near future, utilities will collect, store, and process terabyte SM data sets, which can cause a heavy burden in data analysis. Thus, high-performance algorithms, such as federated learning and parallel computing, should be further investigated to help with real-time SM applications.

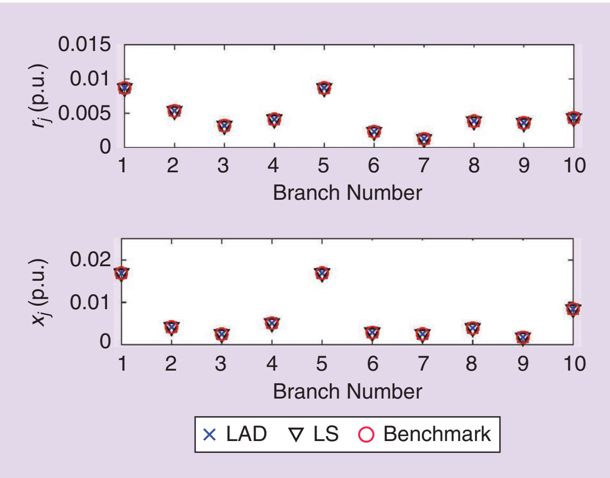


Figure 17. The results of line parameter estimation of a 13-bus test feeder. p.u.: per unit.

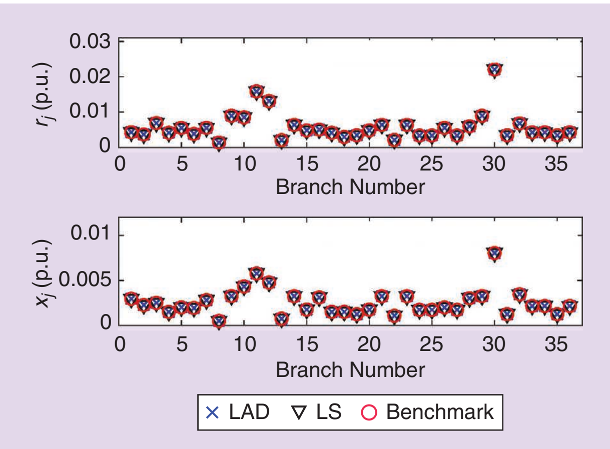


Figure 18. The results of line parameter estimation of a 37-bus test feeder.

5) Adopting the latest machine learning techniques on SM data analytics will receive increasing attention. These algorithms will provide good opportunities in further understanding customer behaviors. The critical issue is to develop data-driven models with high interpretability. This will help utility engineers to acknowledge machine learning techniques and apply them in real-world distribution systems.

Acknowledgments

This work was supported, in part, by both the National Science Foundation under EPCN 2042314 and the Advanced Grid Modeling Program at the U.S. Department of Energy Office of Electricity under Grant DE-OE0000875.

For Further Reading

"Distributed energy poised for explosive growth on the U.S. grid," Greentech Media, Boston, 2018. [Online]. Available: https://www.greentechmedia.com/articles/read/distributed-energy-poised-for-explosive-growth-on-the-us-grid#gs.08g2n2./

"Advanced metering infrastructure and customer systems," Office of Electricity Delivery and Energy Reliability, Washington, D.C., 2016. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/12/f34/AMI%20Summary%20Report_09-26-16.pdf

"Reports on demand response and advanced metering," Federal Energy Regulatory Commission, Washington, D.C., 2020. [Online]. Available: https://www.ferc.gov/industries-data/electric/power-sales-and-markets/demand-response/reports-demand-response-and#:~:text=Definition%20of%20Advanced%20Metering%3A%20Advanced,to%20a%20central%20collection%20point

Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A data-driven customer segmentation strategy based on contribution to system peak demand," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4026–4035, Sept. 2020. doi: 10.1109/TPWRS.2020.2979943.

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 Trans. Power Syst.*, vol. 35, no. 4, pp. 3133–3144, July 2020. doi: 10.1109/TPWRS.2020.2966732.

Y. Guo, Y. Yuan, and Z. Wang, "Distribution grid modeling using smart meter data," 2021, arXiv:2103.00660.

F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A time-series distribution test system based on real utility data," in *Proc. 2019 North Amer. Power Symp. (NAPS)*, Wichita, KS, pp. 1–6. doi: 10.1109/NAPS46351.2019.8999982.

P. Paudyal et al., "The impact of behind-the-meter heterogeneous distributed energy resources on distribution grids," in *Proc. 2020 47th IEEE Photovoltaic Specialists Conf. (PVSC)*, Calgary, AB, Canada, pp. 0857–0862.

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. doi: 10.1109/TSG.2018.2818167.

Biographies

Yuxuan Yuan (yuanyx@iastate.edu) is with Iowa State University, Ames, Iowa, 50010, USA.

Zhaoyu Wang (wzy@iastate.edu) is with Iowa State University, Ames, Iowa, 50010, USA.