A Data-Driven Game-Theoretic Approach for Behind-the-Meter PV Generation Disaggregation

Fankun Bu, *Student Member, IEEE*, Kaveh Dehghanpour, *Member, IEEE*, Yuxuan Yuan, *Student Member, IEEE*, Zhaoyu Wang, *Member, IEEE*, and Yingchen Zhang, *Senior Member, IEEE*

***Abstract*—Rooftop solar photovoltaic (PV) power generator is a widely used distributed energy resource (DER) in distribution systems. Currently, the majority of PVs are installed behind-the-meter (BTM), where only customers' *net demand* is recorded by smart meters. Disaggregating BTM PV generation from net demand is critical to utilities for enhancing grid-edge observability. In this paper, a data-driven approach is proposed for BTM PV generation disaggregation using solar and demand exemplars. First, a data clustering procedure is developed to construct a library of candidate load/solar exemplars. To handle the volatility of BTM resources, a novel game-theoretic learning process is proposed to adaptively generate optimal composite exemplars using the constructed library of candidate exemplars, through repeated evaluation of disaggregation residuals. Finally, the composite native demand and solar exemplars are employed to disaggregate solar generation from net demand using a semi-supervised source separator. The proposed methodology has been verified using real smart meter data and feeder models.**

***Index Terms*—Rooftop solar photovoltaic, distribution system, source disaggregation, game theory.**

I. Introduction

In PRACTICE, the majority of residential rooftop PVs are installed behind-the-meter (BTM), where only the net demand is recorded, which equals native demand minus the solar power generation. Therefore, PV generation is usually invisible to distribution system operators. This invisibility, along with the stochastic nature of solar power, can cause new problems for utilities, such as inaccurate load forecasting and estimation [1], [2], inefficient service restoration [3], [4], and sub-optimal network expansion decisions [5]–[7]. Thus, it is of significance to disaggregate PV generation from net demand to enhance grid-edge observability. One solution to this problem is to monitor each single rooftop PV generation by installing extra metering devices. However, due to the large number of distributed PVs, this option comes at a significant cost for utilities.

To avoid costly metering infrastructure expansion, two categories of approaches have been proposed in the literature to disaggregate PV generation from net demand:

Category I. Model-based methods: Parametric models along with weather information have been used to estimate solar generation [8]. In [1], a virtual equivalent PV station model is utilized to represent the total generation of BTM PVs in a region, where model parameters are obtained by solving an optimization problem. In [9], the clear sky generation model is combined with a physical PV panel model to estimate solar generation. This model-based framework requires meteorological data, precise geographic information, and accurate physical characteristics of PV arrays. The major shortcoming of model-based solutions for solar disaggregation is the unavailability and uncertainty of model parameter information [10], which is further complicated by limited access to unauthorized BTM installations [5]. Moreover, model-based solutions are subject to gross overestimation of solar generation in case of BTM PV failure [11].

Category II. Data-driven methods: As the advanced metering infrastructure (AMI) has been widely deployed in distribution systems in recent years, utilities have gained access to large amounts of smart meter data [12], [13]. To mine the hidden information contained within various data sources with both sufficient temporal and spatial granularity, data-driven approaches have been proposed by researchers for different applications, such as energy disaggregation [14], load forecasting [1], load management [15] and fault detection [11]. In particular, learning-based approaches have drawn significant attention among both researchers and industry practitioners. Measurement data from various sources, including smart meters, supervisory control and data infrastructure (SCADA), and micro-phasor measurement units (μPMU), have been utilized to perform solar disaggregation from net demand. In [16], a data-driven approach is proposed based on dimension reduction and mapping functions using PV generation measurement data from temporarily-installed sensors. In [17], a linear proxy-based estimator is developed to disaggregate a solar farm generation from feeder-level measurement using μPMU data, along with the measured power profile of nearby PV plants, and global horizontal irradiance (GHI) proxy data. In [18], a PV generation disaggregation approach is presented for groups of residential customers, under the assumption that their aggregate active power is measured at the point of common coupling (PCC)

Q1

Manuscript received July 15, 2019; revised November 23, 2019; accepted January 11, 2020. This work was supported by Advanced Grid Modeling Program at the U.S. Department of Energy Office of Electricity under Grant DE-OE0000875. Paper no. TPWRS-01023-2019. *(Corresponding author: Zhaoyu Wang.)*

F. Bu, K. Dehghanpour, Y. Yuan, and Z. Wang are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: fbu@iastate.edu; kavehdeh1@gmail.com; yuanyx@iastate.edu; wzy@iastate.edu).

Y. Zhang is with the National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: yingchen.zhang@nrel.gov).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2020.2966732

to the grid. In [19], a non-intrusive load monitoring (NILM) approach is proposed to disaggregate PV generation from net demand using measurements with 1-second resolution. In [10], the capacity of residential rooftop PVs are estimated using customers' net load curve features. In addition, several data-driven methods have been applied for disaggregating house-level demand into appliance-level energy [14], [20], [21]. However, these energy disaggregation approaches also require input data with high temporal resolution in the range of a few seconds. Datasets with this level of granularity are not generally available to utilities, which applying these approaches difficult for solving solar disaggregation problem. To sum up, the chief limitations of previous data-driven methods are: dependence on installation of costly μPMU devices and high-resolution sensors throughout the network, availability of massive PV generation data, vulnerability to customer behavior volatility, and the inability to estimate all relevant parameters of BTM PV generators, including panel orientations, which impact the time-series generation profile.

Considering the drawbacks of previous methods and the emerging of smart meter data source, in this paper, a novel game-theoretic data-driven approach is proposed for disaggregating PV generation using only smart meter data. The proposed approach exploits the observed correlations within real utility data. The basic idea is to use the native demand and PV generation of fully observable customers to disaggregate the native demand and PV generation of customers with only known net demand. Accordingly, a spectral clustering (SC) algorithm is employed to construct solar and native demand *candidate exemplars* using the data from fully observable customers, which are then stored in an exemplar library. Next, PV generation disaggregation is formulated as a nested bi-layer optimization problem: At the outer layer, a *semi-supervised signal separation* (SSS) algorithm receives the composite native demand and solar exemplars from the inner-layer to disaggregate the native demand and PV generation from customers' net demand. The outer layer of the solar disaggregation process is subject to the response of the inner layer, at which a learning mechanism is developed to find optimal weights that are assigned to candidate native demand and solar generation exemplars to construct *composite exemplars*. This mechanism is based on the concept of repeated games with vector payoff (RGVP) [22] in game theory. While game theory has been previously applied in power system studies [15], [23], [24], we have not found application of RGVP theory to address the solar disaggregation challenge as presented in this paper. The learned weights are continuously updated over time using the disaggregation residual as a feedback signal. The purpose of this novel adaptive composite exemplar construction strategy is to provide optimal response to the volatile and variable behavior of customers and solar generation profiles at the grid-edge. The proposed method is validated using advanced metering infrastructure (AMI) data from our utility partners. In this paper, vectors are denoted using bold italic lowercase letters and matrices are denoted as bold uppercase letters.

The main contributions are summarized as follows:

- A data-driven learning-based approach is proposed for disaggregating BTM PV generation using only smart meter data. This method has been numerically compared with a model-based benchmark and has shown to have considerable improvements under incomplete information of BTM PV parameters.
- To find the hidden native demand and solar power values corresponding to different patterns, a closed loop game-theoretic approach is designed to learn the weights assigned to candidate exemplars for composite exemplar construction.
- The time-varying weights that are used for exemplar construction significantly enhance the adaptability of the disaggregator to unknown abnormal BTM events, such as PV failure and unauthorized installation and expansion of solar arrays.

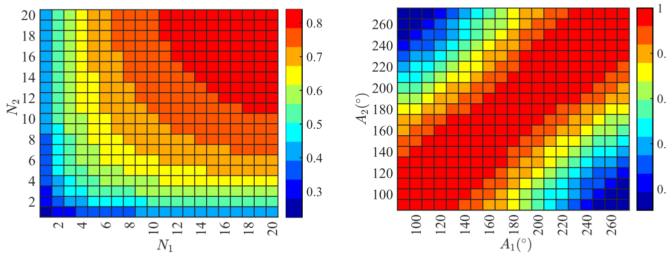
The rest of the paper is organized as follows: Section II introduces the overall framework of BTM PV generation disaggregation approach and describes smart meter dataset. Section III proposes the method for constructing candidate exemplars. Section IV describes the procedure of disaggregating BTM PV generation and native demand from net demand. In Section V, a game-theoretic learning process is presented to obtain optimal composite PV generation and native demand exemplars. In Section VI, case studies are analyzed and Section VII concludes the paper.

II. Proposed BTM PV Generation Disaggregation Framework and Dataset Description

A. Overall Framework of the Proposed Approach

In this paper, the customers are classified into three types: (I) S_P denotes the set of fully observable end-users without PVs, whose native demands are directly measured by smart meters. Note that the net demand of these customers equals their native demand. (II) S_G denotes the set of fully observable customers with PV generation resources. Both the native demand and the solar generation of these customers are measured separately. (III) S_N represents the group of customers with BTM PVs and net demand measurements. The native demand and solar generations of these customers are unknown to the utilities. The goal of this paper is to separate aggregate BTM PV generation of groups of customers in S_N.

The basic idea of the proposed BTM PV generation disaggregation approach is based on the observations that (1) the native demand of sufficiently-large groups of customers are highly correlated, (2) the PV generation of customers with similar orientation are highly correlated, and (3) the correlation between native demand and PV generation is very small. These three observations can be corroborated using real native demand and PV generation data. Fig. 1(a) shows the correlation between the native demands of two groups of fully observable customers in S_P and S_G, where, N_1 and N_2 denote size of each group. It can be seen that as the number of customers in each group increases, the correlation between the aggregate native demands of the customer rises as well. Fig. 1b illustrates the impact of PV panel azimuth on the pairwise PV generation correlation of customers, where A_1 and A_2 denote the azimuths of two PV panels. It can be seen that as the difference between the azimuths of PV panels decreases the correlation between the solar power increases significantly. Hence, the similarity in solar generation is mainly due to similar panel orientations as expected [9], [18].



(a) Native demand (b) BTM PV generation

Fig. 1. Observed correlations from real smart meter data.

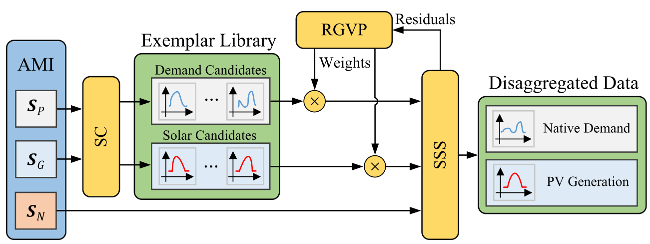


Fig. 2. Overall structure of the proposed BTM PV generation disaggregation method.

In contrast to the significant pairwise correlation between the native demands of groups of customers and that of BTM solar power outputs of PVs with similar orientations, the correlation between native demand and BTM PV generation is significantly small and less than 0.3. This small correlation can be further corroborated by the mismatch between the native demand and PV generation peak times. Specifically, PV units generally output their maximum power during noon, while the native demand usually peaks in the afternoon or early evening [25]. The small correlation can be explained by the PVs' zero-output during nighttime, which results in a decline in the correlation between PV generation and native demand. These three observations set the foundation for constructing native demand and PV generation exemplars using the data of customers in the sets S_P and S_G, to approximate the unobservable native demand and BTM PV generation of customers in the set S_N.

The overall disaggregation process is performed for groups of customers connected to the same lateral or secondary distribution transformer [6]. The components of this process are shown in Fig. 2:

Component I. Exemplar Library Construction: The library consists of typical candidate native demand and solar generation profiles. This library is constructed based on the data of customers in the sets S_P and S_G using a spectral clustering (SC) algorithm. The SC algorithm automatically identifies customers with similar native load patterns and solar generation profiles. The previously-discussed correlations are used as a measure of similarity within the clustering algorithm. The outcomes of the SC method are the native demand and solar power cluster centers that are added as candidate exemplars to the library.

Component II. Composite Exemplar Construction: A weighted averaging operation is performed over the candidate native demand and PV generation exemplars within the library to generate composite native demand and solar generation exemplars.

Component III. BTM PV Generation Disaggregation: An SSS method is developed to disaggregate the net demand of customers in S_N by finding the optimal coefficients that determine the share of composite native demand and solar power exemplars within the net demand. The objective of the coefficient optimization process is to minimize the disaggregation residuals.

Component IV. Game-theoretic Learning: A RGVP-based learning process is designed to assign and update the weights for each candidate exemplar in the library over time. These updated weights are then used to generate the composite native demand and solar generation exemplars for the next time point (Component II). This game-theoretic mechanism adaptively revises the behavior of the disaggregator in response to the time-varying solar power and native demand. The disaggregation process can be converted into a nested bi-layer optimization problem formulated as follows:

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

$$\text{s.t.} \quad \boldsymbol{p}_t+\boldsymbol{g}_t=\boldsymbol{p}_t^n \tag{1b}$$

$$\boldsymbol{p}_t^C=[\boldsymbol{p}_t^{c_1},\ldots,\boldsymbol{p}_t^{c_M}]\boldsymbol{\omega}_t^* \tag{1c}$$

$$\boldsymbol{g}_t^C=[\boldsymbol{g}_t^{c_1},\ldots,\boldsymbol{g}_t^{c_N}]\boldsymbol{\theta}_t^* \tag{1d}$$

$$\{\boldsymbol{\omega}_t^*,\boldsymbol{\theta}_t^*\}=\underset{\boldsymbol{\omega}_t,\boldsymbol{\theta}_t}{\text{argmin}}\ \Phi_\lambda(\boldsymbol{p}_{t-1}^n,\mathbf{P}_t^c,\mathbf{G}_t^c,\alpha_{t-1}^*,\beta_{t-1}^*) \tag{1e}$$

$$\text{s.t.} \quad \sum_{i=1}^{M}\omega_{i,t}=1,\omega_{i,t}\geq 0 \tag{1f}$$

$$\sum_{j=1}^{N}\theta_{j,t}=1,\theta_{j,t}\geq 0 \tag{1g}$$

where, $||\cdot||_2$ denotes l_2-norm. Note that all the demand and generation variables in this equation are defined over a time window of length T, where a vector $\boldsymbol{x}_t$ represents data samples of variable x in the time window $[t-T+1,t]$ as, $\boldsymbol{x}_t=[x(t-T+1),\ldots,x(t)]$. The objective of the *outer layer* is to minimize the summation of the overall disaggregation residuals, consisting of two components: 1) the difference between the actual native demand, $\boldsymbol{p}_t$, and its alternative epitome $\boldsymbol{p}_t^C\alpha_t$, and 2) the difference between the actual BTM PV generation, $\boldsymbol{g}_t$, and its alternative epitome $\boldsymbol{g}_t^C\beta_t$. Here, α_t and β_t determine the proportions of demand and solar powers within the net demand, for the given composite native demand and solar exemplars, denoted by $\boldsymbol{p}_t^C$ and $\boldsymbol{g}_t^C$, respectively. Constraint (1b) ensures that the summation of the native demand and PV generation equals the observed net demand, $\boldsymbol{p}_t^n$, which is measured by AMI. Constraints (1c) and (1d) represent the construction of composite native demand and BTM PV generation exemplars, where $\boldsymbol{p}_t^{c_i}$ and $\boldsymbol{g}_t^{c_j}$ are the candidate native demand and BTM PV generation exemplars, respectively. The composite exemplar construction process employs the weight vectors, $\boldsymbol{\omega}_t=[\omega_{1,t},\ldots,\omega_{M,t}]$ and $\boldsymbol{\theta}_t=[\theta_{1,t},\ldots,\theta_{N,t}]$, where $\omega_{i,t}$ and $\theta_{j,t}$

are the weights corresponding to candidate exemplars $\boldsymbol{p}_t^{c_i}$ and $\boldsymbol{g}_t^{c_j}$, respectively. Note that each candidate exemplar represents the typical native demand/solar generation profile in the time window $[t-T+1, t]$, which are stored in the exemplar library. The objective of the *inner layer*, (1e), is to minimize the parameterized *potential function*, Φ_λ, of the game-theoretic learning process, with parameter λ, to reduce the long-term estimation regret, where $\mathbf{P}_t^c = [\boldsymbol{p}_t^{c_1}, \ldots, \boldsymbol{p}_t^{c_M}]$ and $\mathbf{G}_t^c = [\boldsymbol{g}_t^{c_1}, \ldots, \boldsymbol{p}_t^{c_N}]$ are the native demand and solar generation candidate exemplar libraries, respectively. λ is a user-defined parameter that determines the speed of updating of the weights in the game-theoretic framework (i.e., higher λ implies faster updates). Note that the inner layer optimizes the weights at time t using the measured net demand and the outcome of the outer layer at time $t-1$. Constraints (1f) and (1g) ensure that the weights assigned to the candidate exemplars are non-negative and have l_1-norms equal to one. The game-theoretic process assigns higher weight values to candidate exemplars that have higher impact on reducing the overall disaggregation residuals.

B. Dataset Description

In this paper, net demand of individual customers is constructed by subtracting real BTM PV generation from real native demand. The hourly native demand and PV generation data are from Midwest U.S. utilities. The time rage of the dataset is one year and it contains 1120 customers and 337 PVs. The nominal capacity of PVs ranges from 3 kW to 8 kW. These data are available online [26].

III. Candidate Exemplar Library Construction and Composite Exemplar Generation

The time-series data of fully observable customers in S_P and S_G are leveraged to construct the candidate native demand and solar generation exemplar library using an SC method [27], [28]. In performing SC, two graphs are developed independently, each corresponding to the typical native demand and the typical PV generation libraries. The general steps in constructing the candidate exemplar libraries and composite demand/solar exemplars are as follows:

Step I. Developing similarity graphs: The basic idea of SC technique is to solve the clustering problem using graph theory. To achieve this, a similarity graph, $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, is developed using AMI data. $\mathbf{V}$ denotes the set of vertices of the graph and $\mathbf{E}$ is the set of edges connecting vertices. In our work, for the native demand, the average daily load profiles of customers in set S_P are defined as the graph vertices. For PV generation, the normalized solar power profiles of the PVs in set S_G are defined as graph vertices. The basic idea is to connect the vertices in $\mathbf{V}$ that are similar to each other. We have used a Gaussian kernel function as a measure of similarity, as shown below:

$$\mathbf{W}_{i,j} = \exp\left(\frac{-||\mathbf{V}_i - \mathbf{V}_j||_2^2}{\rho_i \rho_j}\right) \tag{2}$$

where, $\mathbf{W}_{i,j}$ is the weight assigned to the edge connecting vertices $\mathbf{V}_i$ and $\mathbf{V}_j$, and ρ_i and ρ_j are tunable scaling parameters for vertices $\mathbf{V}_i$ and $\mathbf{V}_j$. The two vertices, $\mathbf{V}_i$ and $\mathbf{V}_j$, are connected when the weight of the corresponding edge, $\mathbf{W}_{i,j}$, is larger than 0 (i.e., they have non-trivial similarities).

Step II. Developing Graph Laplacian matrices: Based on similarity graphs obtained from demand/solar power data, the clustering process is transformed into a graph partitioning problem, which cuts a graph into multiple smaller sections by removing edges. The graph partitioning can be conducted in different ways according to different objective functions. In this paper, the objective function is to roughly maximize the dissimilarity between the different graph clusters while minimizing the similarity within each cluster [27]:

$$N_G = \min_{\mathbf{C}_1, \ldots, \mathbf{C}_\mu} \sum_{i=1}^{\mu} \frac{\varphi(\mathbf{C}_i, \mathbf{V} \setminus \mathbf{C}_i)}{d(\mathbf{C}_i)} \tag{3}$$

where, μ is the number of clusters, $\mathbf{C}_i$ is the i'th cluster in graph $\mathbf{G}$, $\mathbf{V} \setminus \mathbf{C}_i$ represents the vertices of $\mathbf{V}$ that are not in $\mathbf{C}_i$, $\varphi(\mathbf{C}_i, \mathbf{V} \setminus \mathbf{C}_i)$ represents the sum of the weights in $\mathbf{C}_i$ and $\mathbf{V} \setminus \mathbf{C}_i$, $d(\mathbf{C}_i)$ denotes the sum of the weights of the vertices in $\mathbf{C}_i$. It has been shown that the partitioning problem (equation (3)) can be solved using the eigenvectors of the normalized Graph Laplacian matrix [27], $\mathbf{L}$, as a reduced-order representation of the original data. The Laplacian is obtained as follows:

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} \tag{4}$$

where, $\mathbf{D}$ is a diagonal matrix whose diagonal elements equal the sum of elements in each row of $\mathbf{W}$. To obtain Laplacian eigenvalues, $\{\mu_1, \mu_2, \ldots, \mu_n\}$, and the corresponding eigenvectors, eigen-decomposition is performed on $\mathbf{L} \in \mathbb{R}^{n \times n}$. The first k eigenvectors corresponding to the first k largest eigenvalues are concatenated into a new matrix $\mathbf{E} \in \mathbb{R}^{n \times k}$.

Step III. Obtaining candidate exemplars: The matrix $\mathbf{E}$ can be considered as the new representation of dataset, which embeds vertices in a lower-dimension space. It has been shown that this new matrix improves the cluster-properties of the data [27]. Then, k-means algorithm is performed to cluster the rows of $\mathbf{E}$. To find the optimal number of clusters, the modified Hubert Γ statistic index is adopted for calibration [29]. After that, customers in the sets S_P and S_G are classified into M and N clusters, respectively. The corresponding candidate native demand and solar generation exemplars, $\boldsymbol{p}_t^{c_i}$ and $\boldsymbol{g}_t^{c_j}$, are obtained using the cluster centers, which equal the average demand/solar powers for the customers belonging to each cluster.

Step IV. Constructing composite exemplars: The input weights from the RGVP module are used to build composite native demand and PV generation exemplars through an averaging process over the candidate exemplars:

$$\boldsymbol{p}_t^C = \sum_{i=1}^{M} \boldsymbol{p}_t^{c_i} \omega_{i,t} \tag{5a}$$

$$\boldsymbol{g}_t^C = \sum_{j=1}^{N} \boldsymbol{g}_t^{c_j} \theta_{j,t} \tag{5b}$$

The weights, $\omega_{i,t}$ and $\theta_{j,t}$, are obtained from the RGVP-based learning process, which is elaborated in Section V.

IV. BTM PV Generation and Native Demand Disaggregation

Using the constructed composite native demand and solar generation exemplars for customers in the sets S_P and S_G, the task of SSS is to estimate the coefficients of the composite exemplars, α_t and β_t, which are unknown *a priori*. These coefficients determine the optimal disaggregation of the measured net demand for customers in S_N. This problem is formulated as a residual minimization problem:

$$\min_{\boldsymbol{p}_t,\boldsymbol{g}_t,\alpha_t,\beta_t} \quad \frac{1}{2}(||\boldsymbol{p}_t - \boldsymbol{p}_t^C\alpha_t||_2^2 + ||\boldsymbol{g}_t - \boldsymbol{g}_t^C\beta_t||_2^2) \tag{6a}$$

$$\text{s.t.} \quad \boldsymbol{p}_t + \boldsymbol{g}_t = \boldsymbol{p}_t^n \tag{6b}$$

where, the disaggregation residual is defined using the l_2-norm, which yields a convex and differentiable optimization problem that can be efficiently represented as a least-squares problem, assuming that the measurement noise follows a Gaussian distribution. Leveraging the negligible correlation between the native demand and PV generation, the optimization problem can be solved efficiently in real-time using *normal equations*, which are based on introducing Lagrange multipliers to the constraint (6b) and employing the gradient of the objective function (6a). This process yields the following optimal solutions:

$$\begin{bmatrix}\alpha_t^* \\ \beta_t^*\end{bmatrix} = (\mathbf{X}_t{}^{\mathsf{T}}\mathbf{X}_t)^{-1}\mathbf{X}_t{}^{\mathsf{T}}\boldsymbol{p}_t^n \tag{7}$$

where, $\mathbf{X}_t = [\boldsymbol{p}_t^C, \boldsymbol{g}_t^C]$. Note that although $\boldsymbol{p}_t$ and $\boldsymbol{g}_t$ are decision variables of interest in equation (6), they cannot be recovered explicitly, as described in [30]. Instead, the optimal coefficients α_t^* and β_t^* can be explicitly optimized and are used to approximate the disaggregated native demand, $\hat{\boldsymbol{p}}_t$, and BTM PV generation, $\hat{\boldsymbol{g}}_t$, for customers in the set S_N as follows:

$$\hat{\boldsymbol{p}}_t = \boldsymbol{p}_t^C\alpha_t^* \tag{8a}$$

$$\hat{\boldsymbol{g}}_t = \boldsymbol{g}_t^C\beta_t^* \tag{8b}$$

V. RGVP-Based Weight Learning

In practice, the weights assigned to the candidate exemplars, $\boldsymbol{\omega}_t$ and $\boldsymbol{\theta}_t$, are unknown *a priori* due to the unobservability of real native demand and PV generation of customers in the set S_N. In this section, a novel adaptive game-theoretic learning process is designed to learn these weights and then to generate and update optimal composite exemplars over time. The main idea is to handle the variations and volatility of unknown native demand and BTM solar generation by minimizing the long-term disaggregation residuals.

The weight updating process is cast as a repeated game model with vector payoff, in which two components are defined: a *player* and a set of *experts* [22], which in our problem correspond to the disaggregator and the candidate exemplars, respectively. The experts provide "advice" (i.e., typical load/solar patterns) to the player, who then combines them to obtain the composite demand/solar exemplars. To do this, the player assigns weights to each constructed candidate exemplar and performs weighted averaging to build time-series composite exemplars, $\boldsymbol{p}_t^C$ and $\boldsymbol{g}_t^C$. To optimize these weights, the player determines *regret* values for the advice of each expert. The regret value for each expert represents the difference between the player's loss when using the composite exemplars and the loss when using the candidate exemplar. Intuitively, a negative loss indicates how much the player is better off by performing disaggregation using composite exemplars instead of each candidate exemplar. A potential function is used to minimize the magnitude of accumulated regret vector over time by employing a gradient-based search process to optimize weight values. After that, the optimal weights are passed on to construct composite exemplars and perform the disaggregation process for customers in the set S_N, as described in Sections III and IV. The RGVP steps are shown in Fig. 3 and described as follows:

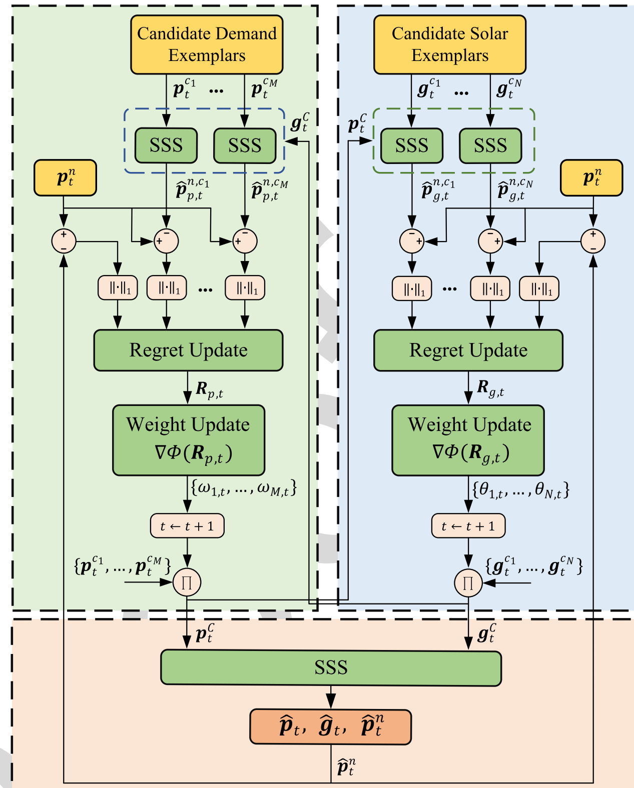


Fig. 3. Detailed structure of the RGVP module from Fig. 2.

Step I. Initialization: $t \leftarrow t_0$; uniform distribution is used to initialize the weights assigned to candidate exemplars, i.e., $\omega_{i,t} \leftarrow \frac{1}{M}$ and $\theta_{j,t} \leftarrow \frac{1}{N}$.

Step II. Construct the latest composite exemplars: Receive candidate exemplars from exemplar libraries constructed using SC. Then, assign $\boldsymbol{\omega}_t$ and $\boldsymbol{\theta}_t$ to these candidate exemplars to construct composite native demand and PV generation exemplars as shown in Section III.

Step III. Disaggregation using the latest composite exemplars: Pass the generated composite exemplars, $\boldsymbol{p}_t^C$ and $\boldsymbol{g}_t^C$, to SSS (Section IV) to perform disaggregation. The estimated net demand is calculated using the disaggregated native demand and solar generation as follows:

$$\hat{\boldsymbol{p}}_t{}^n = \hat{\boldsymbol{p}}_t + \hat{\boldsymbol{g}}_t \tag{9}$$

Step IV. Determine disaggregation residual: The disaggregation residual for the composite exemplars is obtained using the measured and estimated net demand as follows:

$$e_t^C = ||\hat{\boldsymbol{p}}_t{}^n - \boldsymbol{p}_t^n||_1 \tag{10}$$

where, $||\cdot||_1$ denotes l_1-norm.

Step V. Disaggregation using the latest candidate native demand exemplars: Instead of using composite native demand exemplar for disaggregation, candidate native demand exemplars are leveraged to perform SSS. To do this, the candidate native demand and composite PV generation exemplar pairs $\{\boldsymbol{p}_t^{c_i}, \boldsymbol{g}_t^C\}$ are passed to the SSS in parallel, $\forall i \in \{1,\ldots,M\}$. The outcomes are the disaggregated native demand and solar generation for each pair, denoted as $\hat{\boldsymbol{p}}_{p,t}^{c_i}$ and $\hat{\boldsymbol{g}}_{p,t}^{c_i}$, respectively. These obtained signals are used to reconstruct the net demand corresponding to each candidate native demand exemplar, $\hat{\boldsymbol{p}}_{p,t}^{n,c_i}$, $\forall i \in \{1,\ldots,M\}$, as follows:

$$\hat{\boldsymbol{p}}_{p,t}^{n,c_i} \leftarrow \hat{\boldsymbol{p}}_{p,t}^{c_i} + \hat{\boldsymbol{g}}_{p,t}^{c_i} \tag{11}$$

Finally, the disaggregation residual corresponding to each candidate native demand exemplar, $\boldsymbol{p}_t^{c_i}$, is obtained as shown below:

$$e_{p,t}^{c_i} = ||\hat{\boldsymbol{p}}_{p,t}^{n,c_i} - \boldsymbol{p}_t^n||_1 \tag{12}$$

Step VI. Disaggregation using the latest candidate solar generation exemplars: The process introduced in Step V is performed symmetrically over candidate solar generation exemplars. Accordingly, the composite native demand and candidate PV generation exemplar pairs $\{\boldsymbol{p}_t^C, \boldsymbol{g}_t^{c_j}\}$ are passed to the SSS, $\forall j \in \{1,\ldots,N\}$, where the disaggregated native demand and solar generation are obtained, denoted as $\hat{\boldsymbol{p}}_{g,t}^{c_j}$ and $\hat{\boldsymbol{g}}_{g,t}^{c_j}$, respectively. These disaggregated signals are then used to reconstruct the net demand, $\hat{\boldsymbol{p}}_{g,t}^{n,c_j}$, corresponding to each candidate solar exemplar and determine the disaggregation residuals $e_{g,t}^{c_j}$, similar to equation (12).

Step VII. Update candidate regrets: The player's instantaneous regrets for each native demand and solar generation candidate exemplars are calculated as follows:

$$r_{p,t}^{c_i} = e_t^C - e_{p,t}^{c_i} \quad i = 1,\ldots,M \tag{13a}$$

$$r_{g,t}^{c_j} = e_t^C - e_{g,t}^{c_j} \quad j = 1,\ldots,N \tag{13b}$$

here, $r_{p,t}^{c_i}$ and $r_{g,t}^{c_j}$ represent the regrets for candidate demand and solar exemplars, respectively, which are measured in terms of the payoffs in disaggregation residuals by following the advice of candidate exemplars instead of composite exemplars at time t. The cumulative regrets for the i'th candidate native demand exemplar and the j'th candidate PV generation exemplar are defined by summing up the instantaneous regret $r_{p,t}^{c_i}$ and $r_{g,t}^{c_j}$ for all previous time instants in $[t_0, t]$, as follows:

$$R_{p,t}^{c_i} = \sum_{t'=t_0}^{t} r_{p,t'}^{c_i} \quad i = 1,\ldots,M \tag{14a}$$

$$R_{g,t}^{c_j} = \sum_{t'=t_0}^{t} r_{g,t'}^{c_j} \quad j = 1,\ldots,N \tag{14b}$$

TABLE I
NUMBER OF CUSTOMERS AND PVS IN LATERALS

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
N_P	22	26	30	36	25	27	30	24	37	23
N_G	13	13	21	21	18	16	14	9	20	14
	L11	**L12**	**L13**	**L14**	**L15**	**L16**	**L17**	**L18**	**L19**	
N_P	16	26	29	37	67	13	43	24	33	
N_G	10	14	24	17	42	9	23	14	25	

By assigning accumulated regret values to each expert, the regret vectors are obtained as $\boldsymbol{R}_{p,t} = [R_{p,t}^{c_1},\ldots,R_{p,t}^{c_M}]^\mathsf{T}$ and $\boldsymbol{R}_{g,t} = [R_{g,t}^{c_1},\ldots,R_{g,t}^{c_N}]^\mathsf{T}$ for candidate native demand and solar generation exemplars, respectively.

Step VIII. Update weights: The goal of RGVP is to reduce the magnitude of the accumulated regret vectors $\boldsymbol{R}_{p,t}$ and $\boldsymbol{R}_{g,t}$. To do this, *potential functions* are assigned to these accumulated vector spaces. These scalar potential functions are increasing with respect to the advisors' accumulated regrets. Hence, reducing the accumulative regret is transformed into minimizing the values of these potential functions [22]. In this paper, we have adopted exponential potential functions, the gradients of which are used to update the weights as follows [22]:

$$\omega_{i,t+1} = \nabla\Phi_\lambda(\boldsymbol{R}_{p,t})_i = \frac{e^{\lambda R_{p,t}^{c_i}}}{\sum_{j=1}^{M} e^{\lambda R_{p,t}^{c_j}}} \quad i = 1,\ldots,M \tag{15a}$$

$$\theta_{j,t+1} = \nabla\Phi_\lambda(\boldsymbol{R}_{g,t})_j = \frac{e^{\lambda R_{g,t}^{c_j}}}{\sum_{i=1}^{N} e^{\lambda R_{g,t}^{c_i}}} \quad j = 1,\ldots,N \tag{15b}$$

where,

$$\Phi_\lambda(\boldsymbol{u}) = \frac{1}{\lambda}\ln\left(\sum_{i=1}^{L} e^{\lambda u_i}\right) \tag{16}$$

is an exponential potential operator with $\boldsymbol{u} = [u_1,\ldots,u_L]^\mathsf{T}$, λ is a positive tunable parameter indicating the updating speed of weights, which is adopted from literature as $\lambda = \sqrt{8\ln(L)/T}$ [22], with $L = M$ or $L = N$.

Step IX. Moving the disaggregation window: $t \leftarrow t+1$; go back to Step II.

An algorithmic overview of the aforementioned steps of BTM PV generation disaggregation is summarized in Algorithm 1.

VI. CASE STUDY

In this section, the proposed BTM PV generation disaggregation method is verified using real smart meter data described in Section II and the one-line diagram of a 240-node distribution grid presented in [26]. The proposed approach is applied to disaggregate single-phase lateral- and transformer-level PV generation over a one-year data period. The number of customers, N_P, and BTM PV generators, N_G, connected to the system laterals are shown in Table I.

The tunable parameters in the proposed BTM PV generation disaggregation approach include the number of candidate native

Algorithm 1: BTM PV Generation and Native Demand Disaggregation From Net Demand.

1: **procedure** Initialization
2: $\quad t \leftarrow t_0, \omega_{i,t} \leftarrow \frac{1}{M}, i \in \{1, \ldots, M\}, \theta_{j,t} \leftarrow \frac{1}{N}, j \in \{1, \ldots, N\}$
3: **end procedure**
4: Receive $\{\boldsymbol{p}_t^{c_1}, \boldsymbol{p}_t^{c_2}, \ldots, \boldsymbol{p}_t^{c_M}\}$ and $\{\boldsymbol{g}_t^{c_1}, \boldsymbol{g}_t^{c_2}, \ldots, \boldsymbol{g}_t^{c_N}\}$ from SC
5: **procedure** Perform SSS using $\boldsymbol{p}_t^C$ and $\boldsymbol{g}_t^C$
6: $\quad \boldsymbol{p}_t^C \leftarrow \sum_{i=1}^{M} \boldsymbol{p}_t^{c_i} \omega_{i,t}, \boldsymbol{g}_t^C \leftarrow \sum_{j=1}^{N} \boldsymbol{g}_t^{c_j} \theta_{j,t}$
7: $\quad \mathbf{X}_t \leftarrow [\boldsymbol{p}_t^C, \boldsymbol{g}_t^C]$
8: $\quad \{\alpha_t^*, \beta_t^*\} \leftarrow (\mathbf{X}_t{}^\mathsf{T} \mathbf{X}_t)^{-1} \mathbf{X}_t{}^\mathsf{T} \boldsymbol{p}_t^n$
9: $\quad \hat{\boldsymbol{p}}_t \leftarrow \boldsymbol{p}_t^C \alpha_t^*, \hat{\boldsymbol{g}}_t \leftarrow \boldsymbol{g}_t^C \beta_t^*$
10: $\quad \hat{\boldsymbol{p}}_t{}^n \leftarrow \hat{\boldsymbol{p}}_t + \hat{\boldsymbol{g}}_t$
11: **end procedure**
12: **procedure** Perform SSS using $\boldsymbol{p}_t^{c_i}$ and $\boldsymbol{g}_t^C$
13: $\quad \mathbf{X}_{p,t}^{c_i} \leftarrow [\boldsymbol{p}_t^{c_i}, \boldsymbol{g}_t^C] \quad i = 1, \ldots, M$
14: $\quad \{\alpha_{p,t}^{c_i}{}^*, \beta_{p,t}^{c_i}{}^*\} \leftarrow (\mathbf{X}_{p,t}^{c_i}{}^\mathsf{T} \mathbf{X}_{p,t}^{c_i})^{-1} \mathbf{X}_{p,t}^{c_i}{}^\mathsf{T} \boldsymbol{p}_t^n$
15: $\quad \hat{\boldsymbol{p}}_{p,t}^{c_i} \leftarrow \boldsymbol{p}_t^{c_i} \alpha_{p,t}^{c_i}{}^*, \hat{\boldsymbol{g}}_{p,t}^{c_i} \leftarrow \boldsymbol{g}_t^C \beta_{p,t}^{c_i}{}^*$
16: $\quad \hat{\boldsymbol{p}}_{p,t}^{n,c_i} \leftarrow \hat{\boldsymbol{p}}_{p,t}^{c_i} + \hat{\boldsymbol{g}}_{p,t}^{c_i}$
17: **end procedure**
18: **procedure** Perform SSS using $\boldsymbol{p}_t^C$ and $\boldsymbol{g}_t^{c_j}$
19: $\quad \mathbf{X}_{g,t}^{c_j} \leftarrow [\boldsymbol{p}_t^C, \boldsymbol{g}_t^{c_j}] \quad j = 1, \ldots, N$
20: $\quad \{\alpha_{g,t}^{c_j}{}^*, \beta_{g,t}^{c_j}{}^*\} \leftarrow (\mathbf{X}_{g,t}^{c_j}{}^\mathsf{T} \mathbf{X}_{g,t}^{c_j})^{-1} \mathbf{X}_{g,t}^{c_j}{}^\mathsf{T} \boldsymbol{p}_t^n$
21: $\quad \hat{\boldsymbol{p}}_{g,t}^{c_j} \leftarrow \boldsymbol{p}_t^C \alpha_{g,t}^{c_j}{}^*, \hat{\boldsymbol{g}}_{g,t}^{c_j} \leftarrow \boldsymbol{g}_t^{c_j} \beta_{g,t}^{c_j}{}^*$
22: $\quad \hat{\boldsymbol{p}}_{g,t}^{n,c_j} \leftarrow \hat{\boldsymbol{p}}_{g,t}^{c_j} + \hat{\boldsymbol{g}}_{g,t}^{c_j}$
23: **end procedure**
24: **procedure** Update Regret and Weights (Demand)
25: $\quad r_{p,t}^{c_i} = ||\hat{\boldsymbol{p}}_t{}^n - \boldsymbol{p}_t^n||_1 - ||\hat{\boldsymbol{p}}_{p,t}^{n,c_i} - \boldsymbol{p}_t^n||_1$
26: $\quad R_{p,t}^{c_i} = \sum_{t'=t_0}^{t} r_{p,t'}^{c_i}$
27: $\quad \omega_{i,t+1} \leftarrow e^{\lambda R_{p,t}^{c_i}} / \sum_{j=1}^{M} e^{\lambda R_{p,t}^{c_j}} \quad i = 1, \ldots, M$
28: **end procedure**
29: **procedure** Update Regret and Weights (PV)
30: $\quad r_{g,t}^{c_j} = ||\hat{\boldsymbol{p}}_t{}^n - \boldsymbol{p}_t^n||_1 - ||\hat{\boldsymbol{p}}_{g,t}^{n,c_j} - \boldsymbol{p}_t^n||_1$
31: $\quad R_{g,t}^{c_j} = \sum_{t'=t_0}^{t} r_{g,t'}^{c_j}$
32: $\quad \theta_{j,t+1} \leftarrow e^{\lambda R_{g,t}^{c_j}} / \sum_{i=1}^{N} e^{\lambda R_{g,t}^{c_i}} \quad j = 1, \ldots, N$
33: **end procedure**
34: $t \leftarrow t + 1$
35: Go to Step 4

demand exemplars, M, the number of candidate PV generation exemplars, N, and the disaggregation time-window length, T. To optimize the number of clusters, M and N, the modified Hubert Γ statistic index is calculated for different number of clusters by running SC [29]. Then, the optimal values of M and N are determined by finding the knee point of Γ curve as presented in [31]. In our case, M and N are optimized at 4 and 3, respectively. To tune the length of the moving time window, a grid search method was employed to find the minimum net demand estimation residual in terms of mean absolute percentage error ($MAPE$), calculated as follows:

$$MAPE = \frac{100\%}{K} \cdot \sum_{t=1}^{K} \left| \frac{\hat{p}^n(t) - p^n(t)}{\frac{1}{K} \sum_{t=1}^{K} |p^n(t)|} \right| \tag{17}$$

where, K is the total number of net demand samples. In our case, the identified optimal value of T is 96 hours.

These calibrated parameters are then used to perform BTM PV generation disaggregation. Fig. 5 shows the disaggregated PV generation and native demand for a lateral within a one-week period. It can be seen in Fig. 5(a) that the disaggregated PV generation closely fits the actual unobservable solar power, which indicates a high disaggregation accuracy. Also, as demonstrated by the disaggregated and actual PV generation curves during the last day, despite high variations in PV generation profile, the proposed approach can still provide satisfying accuracy. The disaggregated native demand and the actual native demand are shown in Fig. 5b, where the disaggregated native demand can accurately capture the load variations as well.

The proposed approach has been applied to 19 laterals in a real distribution system to test its generalizability. The lateral divisions are illustrated in Fig. 4, where the nodes that have observable PVs are shown with brown color. The smart meter data from individual customers are aggregated to obtain lateral-level demand profiles, in which the utilities have shown significant interest. In our case study 5% of all customers are fully observable with known PV generation and native demand data (the set S_G). The remaining 95% belong to the sets S_N and S_P, where only either the net demand or the native demand is observable, respectively. The idea is to use the data of S_G and S_P to disaggregate PV generation of S_N. Using the actual PV generation and native demand as ground truths, the fitness of the disaggregated PV generation and native demand are evaluated in terms of $MAPE$, as shown in Table II, where the solar and demand disaggregation accuracy are denoted as g_m and p_m, respectively. As can be seen, g_m ranges from 4% to 8%, and p_m ranges from 6% to 10%. Note that in the case of solar disaggregation using feeder-level demand profile measurements, e.g., SCADA power flow data, the network losses should be removed before disaggregation. To do this, previous works have proposed several power flow-based techniques to estimate and eliminate the network losses [32], [33]. After this, PV generation disaggregation can be performed using our proposed method.

The case study is conducted on a standard PC with an Intel(R) Xeon(R) CPU running at 3.70 GHz and with 32.0 GB of RAM. PV generation disaggregation is performed for each lateral illustrated in Fig. 4 over a year, and the computational time ranges from 4.20 seconds to 4.64 seconds.

Furthermore, it is of interest to examine the variations of the obtained game-theoretic weights corresponding to different candidate exemplars. In Fig. 6a, it can be seen that the weights assigned to the three candidate PV generation exemplars for one of the laterals converge to approximately 0.8, 0.1, and 0.1, through the learning process, after which the weight values remain nearly stable. Similar characteristics can be observed in Fig. 6b, which shows the variations of weights assigned to the candidate native demand exemplars for the same lateral.

To validate RGVP, Fig. 8 employs error histograms to demonstrate the performance of the disaggregation process corresponding to RGVP-based composite exemplars and individual candidate exemplars, where $e_{g_t} = g(t) - \hat{g}_(t)$ and $e_{p_t} = p(t) - \hat{p}(t)$ represent solar and demand disaggregation errors, respectively. As can be seen, the error distribution under composite exemplars

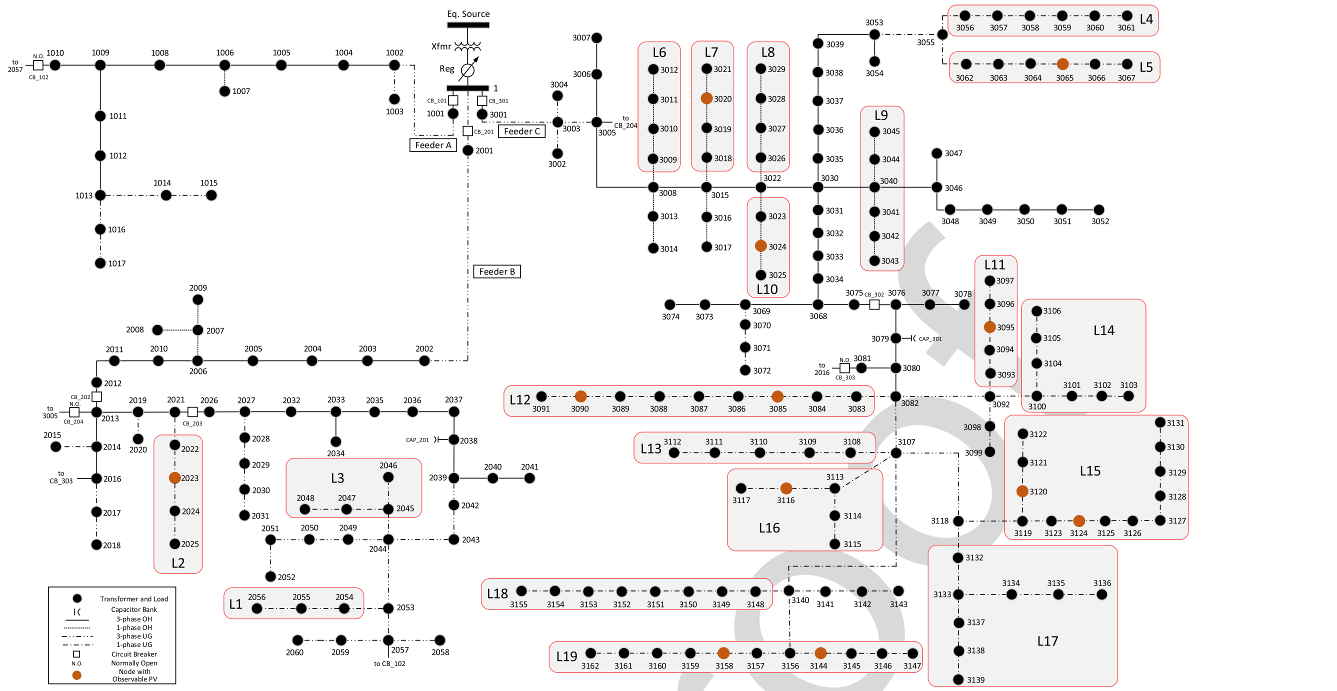
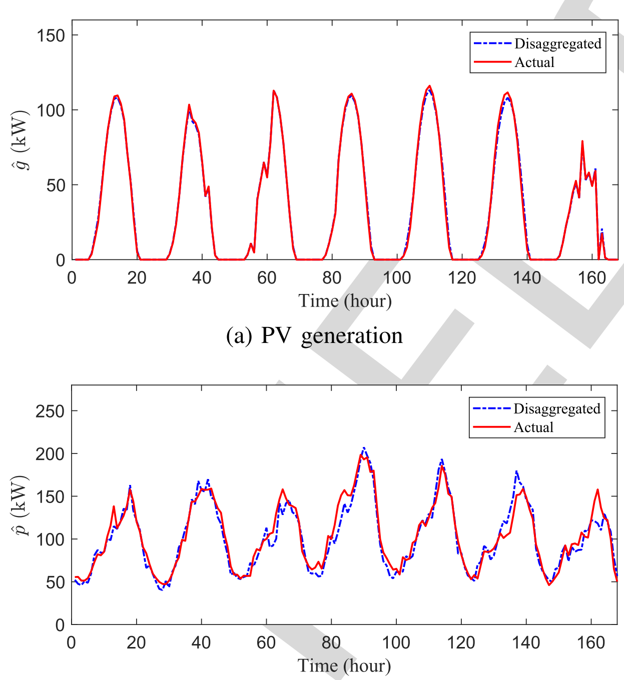


Fig. 4. One-line diagram of a real distribution system.



(a) PV generation

(b) Native demand

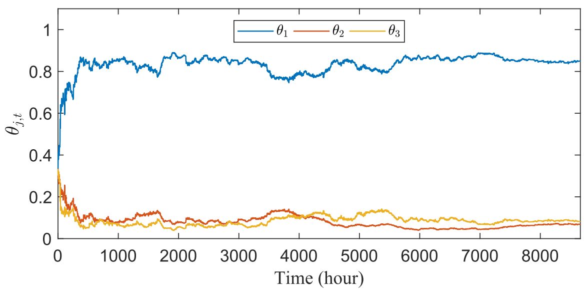
Fig. 5. PV generation and native demand profiles of a lateral within a one-week period.

for both solar and native demand show considerably lower variance (i.e., higher precision), compared to those of candidate exemplars. This implies that using the game-theoretic learning process to construct composite exemplars leads to improvements in the disaggregation accuracy on average.

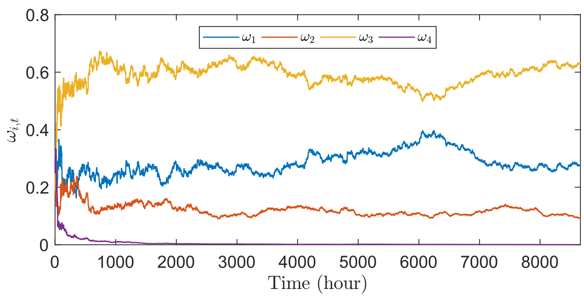
TABLE II
PV Generation and Native Demand Disaggregation MAPE

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
g_m (%)	6.10	8.09	4.43	6.68	7.38	5.34	6.81	8.08	4.86	4.96
p_m (%)	8.44	8.47	7.52	7.84	8.03	7.99	7.72	8.45	7.52	8.96
	L11	**L12**	**L13**	**L14**	**L15**	**L16**	**L17**	**L18**	**L19**	
g_m (%)	6.33	6.12	5.79	6.97	3.66	5.58	4.32	4.58	3.44	
p_m (%)	9.47	7.23	9.04	7.52	6.59	9.66	6.62	8.38	7.51	

Since the proposed disaggregator depends on exemplary profiles, it is critical to conduct additional numerical analysis to capture the sensitivity of disaggregation accuracy with respect to the number of observable customers in S_G and S_P. This has been done by reducing the percentage of observable PVs, n_g, from 1.5% to 0.1% in 0.1% steps (1 PV per step) and performing disaggregation at each step. The average $MAPE$s of disaggregated PV generation and native demand are plotted against the percentage of observable PVs, as shown in Fig. 7. As can be seen in Fig. 7a, once n_g drops below 0.3%, the average $MAPE$ significantly increases. In the worst case, when only 1 customer is observable ($n_g \approx 0.1\%$), the average MAPE is about 28%. This is consistent with expectations: (1) Only a single observable PV cannot represent all other PVs, since the PV panel orientation has significant impact on PV generation profile; (2) As n_g increase from 0.1% to 0.3%, the average $MAPE$ significantly decreases, because the unobservable PVs can be better represented using more diverse PV generation profiles; (3) When n_g is larger than 0.3%, further increase in

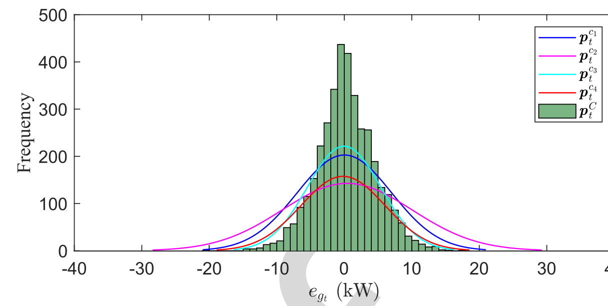


(a) Candidate solar generation weights

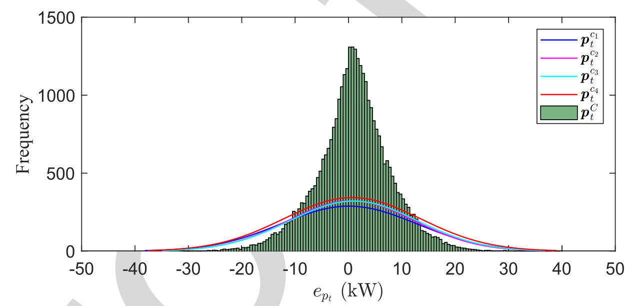


(b) Candidate native demand weights

Fig. 6. The time-series weights assigned to the candidate PV generation and native demand exemplars.

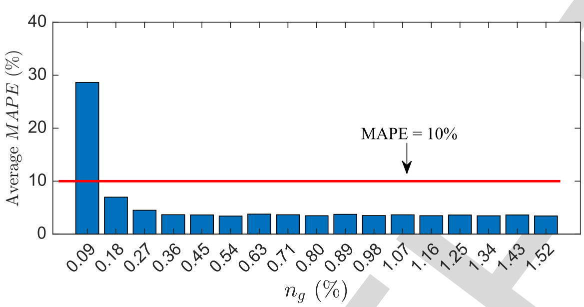


(a) Solar power estimation error

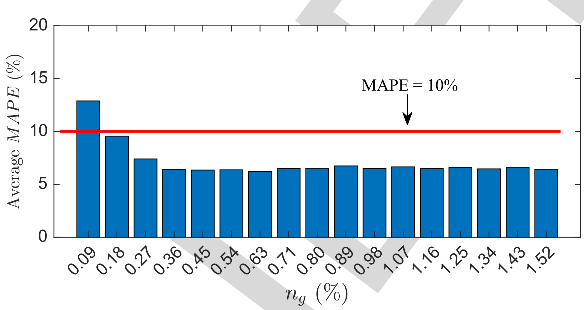


(b) Native demand estimation error

Fig. 8. Error distribution under composite exemplar and candidate exemplars.

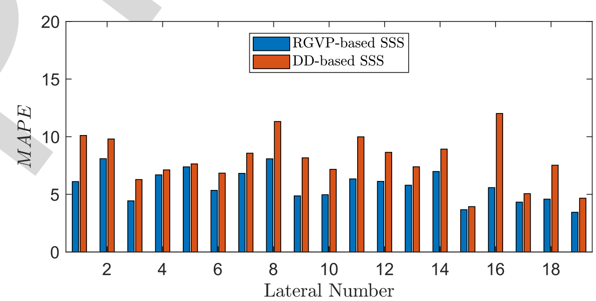


(a) PV generation

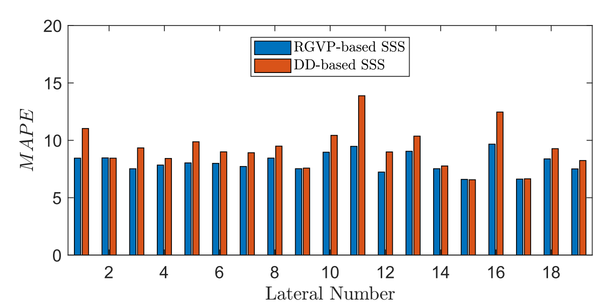


(b) Native demand

Fig. 7. Sensitivity analysis for the number of observable PVs.



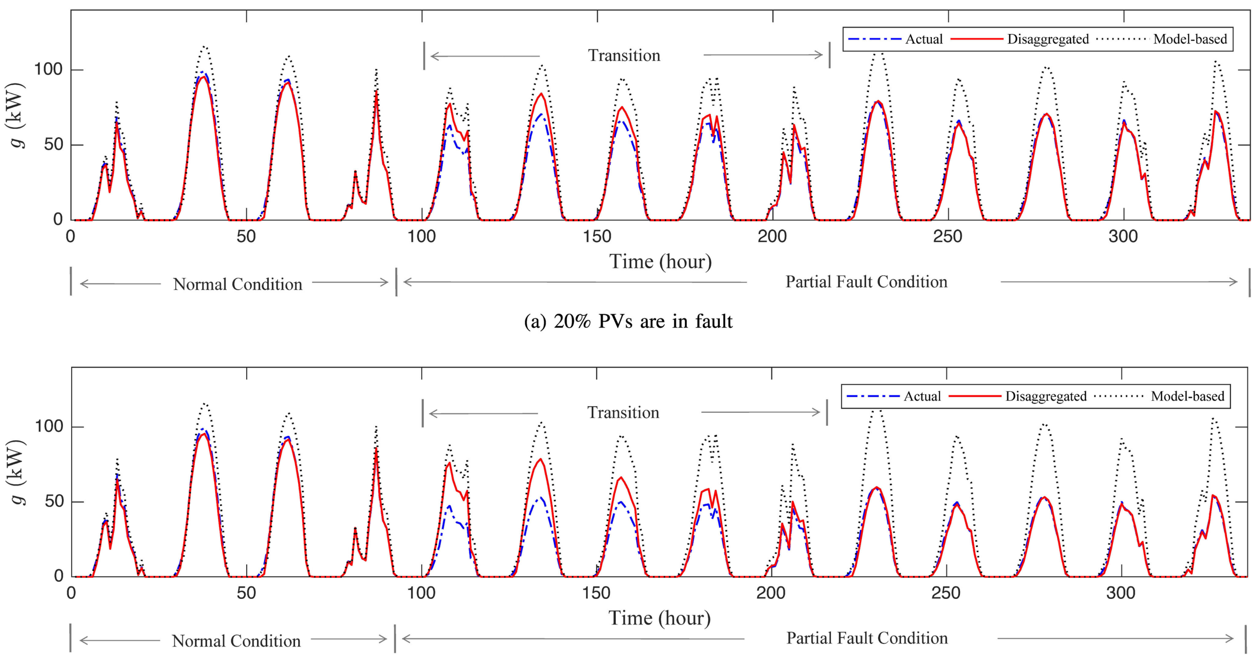
(a) PV generation estimation MAPE



(b) Native demand estimation MAPE

Fig. 9. MAPE comparison using RGVP-based SSS and DD-based SSS.

n_g does not lead to any noticeable accuracy improvements since the redundant exemplary PV generation profiles do not contain much additional information. In Fig. 7b, a similar decreasing trend in average $MAPE$ for the disaggregated native demand can also be observed, which is consistent with the constraint in Equation (1b). Also, we have performed numerical sensitivity analysis to capture the impact of number of observable customers in S_P on disaggregation accuracy. Similar to the case of S_G, as the number of customers with observable native demand increases the disaggregation accuracy improves. In our case studies, the minimum required customers with observable demand is 18%. Note that these numbers are case-dependent, and vary for different customer behaviors in different regions.



(a) 20% PVs are in fault

(b) 40% PVs are in fault

Fig. 10. Adaptability of the proposed data-driven approach.

To further demonstrate the advantage of our proposed approach, we have conducted simulations in two scenarios where a certain percentage of BTM PVs (20% and 40%) have stopped running without the utility's knowledge. A model-based method [34] has been used as a benchmark for comparison with our data-driven technique. In Fig. 10, the real PV generation curves, the disaggregated curves from the data-driven method and the estimated curves from PV model are plotted with different PV failure percentages for comparison. As can be seen, regardless of the percentage of faulty PVs, the model-based method cannot detect PV failure and cannot adjust to PV generation estimation. This inflexibility is due to the model's inability to adapt to changes in system conditions, which are BTM and unknown. In contrast, our data-driven approach displays high adaptability to PVs' failure conditions. Specifically, although the real PV generation decreases to a lower level due to PV failure, our data-driven disaggregator can track this change after a short transition phase. The small difference between the real and the estimated curves, both before the PVs' failure and after the transition, demonstrates satisfying disaggregation accuracy. The adaptability of the proposed disaggregator can also be extended to unauthorized PV installation and expansion, since essentially both partial failure and installation can be translated into changes in BTM capacity of generation.

An alternative approach to the proposed RGVP is to directly perform PV generation disaggregation using candidate exemplars without developing composite exemplars. For abbreviation, we denote this approach as "direct disaggregation-based SSS" (DD-based SSS). The solar-demand disaggregation accuracy of the DD-based SSS and the proposed RGVP-based SSS are shown in Fig. 9. It can be seen that the proposed RGVP-based SSS outperforms DD-based SSS in terms of $MAPE$. The reason for this better performance is that the RGVP-based method can identify the candidate exemplars that are highly correlated with the BTM real load/solar powers.

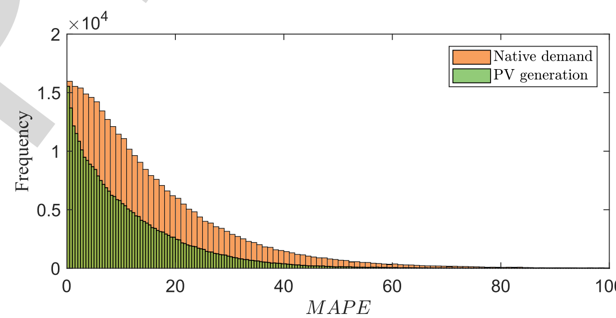


Fig. 11. Estimation MAPE distribution for PV generation and native demand at secondary transformer level.

The proposed BTM PV generation and native demand disaggregation approach is also applied to secondary transformers. Fig. 11 shows the transformer-level disaggregation $MAPE$ distribution. It can be seen that the proposed method is able to achieve an average solar disaggregation $MAPE$ of 10.0%, with an average native demand disaggregation $MAPE$ of 15.0%. As can be seen, disaggregation at transformer-level results in higher residuals compared to lateral-level due to increased grid-edge demand volatility.

VII. Conclusion

This paper presents a non-intrusive novel RGVP-based approach to disaggregate BTM solar generation from the net demand. The proposed method employs the data of fully observable customers to identify typical demand/generation patterns, and optimally combines these patterns to improve disaggregation performance over time. We have used real smart meter data and practical distribution system models from our utility partners to show that this technique is able to enhance solar disaggregation accuracy by adaptively updating the estimator's response to volatile BTM resources. This can enhance utilities' situational awareness of grid-edge resources without incurring extra metering investment costs. The key findings of the paper are summarized as follows:

- Using real smart meter data, we have observed that: the native demand of any two sizable groups of customers are highly correlated; any two PV generation profiles with similar orientations are significantly correlated; the correlation between PV generation and native demand is insignificant. Based on these three observations, we have proposed a novel data-driven PV generation disaggregator which only relies on utilities' existing smart meter data to separate native demand and BTM solar power.
- Numerical experiments have demonstrated that our approach can accurately perform solar generation disaggregation without knowing the specific parameters of BTM PV array and inverters, or weather information. This gives our method a considerable edge over parameter-dependent model-based techniques.
- The numerical experiments have also verified that the proposed disaggregator shows strong robustness and adaptability to unobservable BTM abnormal events, such as PV failure and unauthorized PV array installation/expansion.

The proposed approach shows satisfactory performance on feeder/lateral-level PV generation disaggregation in which our utility partners have shown great interest; however, when applied to individual customers' data, the disaggregation accuracy declines. In the future, we intend to address this challenge.

References

[1] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3255–3264, May 2018.

[2] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalo, "Synchronous pattern matching principle-based residential demand response baseline estimation: Mechanism analysis and approach description," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6972–6985, Nov. 2018.

[3] R. Seguin, J. Woyak, D. Costyk, J. Hambrick, and B. Mather, *High Penetration PV Integration Handbook for Distribution Engineers*. Golden, CO, USA: Nat. Renewable Energy Lab., 2016.

[4] B. Chen, C. Chen, J. Wang, and K. L. Butler-Purry, "Sequential service restoration for unbalanced distribution systems and microgrids," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1507–1520, Mar. 2018.

[5] X. Zhang and S. Grijalva, "A data-driven approach for detection and estimation of residential PV installations," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2477–2485, Sep. 2016.

[6] W. Kersting, *Distribution System Modeling and Analysis*. Boca Raton, FL, USA: CRC Press, 2011.

[7] F. Ding and B. Mather, "On distributed PV hosting capacity estimation, sensitivity study and improvement," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1010–1020, Jul. 2017.

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

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

[10] F. Wang *et al.*, "A distributed PV system capacity estimation approach based on support vector machine with customer net load curve features," *Energies*, vol. 11, no. 7, Jul. 2018, Art. no. 1750.

[11] Y. Zhao, R. Ball, J. Mosesian, J.-F. de Palma, and B. Lehman, "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2848–2858, May 2015.

[12] W. Yi, C. Qixin, H. Tao, and K. Chongqing, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, Mar. 2018.

[13] 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, Sep. 2018.

[14] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE Trans. Neural Netw. Learn. Syst*, to be published. Q2

[15] W. Liu, W. Gu, J. Wang, W. Yu, and X. Xi, "Game theoretic non-cooperative distributed coordination control for multi-microgrids," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6986–6997, Nov. 2018.

[16] H. Shaker, H. Zareipour, E. Muljadi, and D. Wood, "A data-driven approach for estimating the power generation of invisible solar sites," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2466–2476, Sep. 2016.

[17] E. C. Kara, C. M. Roberts, M. D. Tabone, L. Alvarez, D. S. Callaway, and E. M. Stewart, "Disaggregating solar generation from feeder-level measurements," *Sustain. Energy, Grids Netw.*, vol. 13, 2018. Q3

[18] F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised disaggregation of photovoltaic production from composite power flow measurements of heterogeneous prosumers," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3904–3913, Sep. 2018.

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

[20] G. S. Ledva, L. Balzano, and J. L. Mathieu, "Real-time energy disaggregation of a distribution feeder's demand using online learning," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4730–4740, Sep. 2018.

[21] W. Kong, Z. Y. Dong, J. Ma, D. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, Jul. 2018.

[22] M. Maschler, E. Solan, and S. Zamir, *Game Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[23] L. Han, T. Morstyn, and M. McCulloch, "Incentivizing prosumer coalitions with energy management using cooperative game theory," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 303–313, Jan. 2019.

[24] T. Dai and W. Qiao, "Finding equilibria in the pool-based electricity market with strategic wind power producers and network constraints," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 389–399, Dec. 2016.

[25] T. A. Short, *Electric Power Distribution Handbook*. Boca Raton, FL, USA: CRC Press, 2014.

[26] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A time-series distribution test system based on real utility data," 2019. [Online]. Available: https://arxiv.org/abs/1906.04078

[27] U. von Luxburg, "A tutorial on spectral clustering," *Statist Comput.*, vol. 17, pp. 395–416, 2007.

[28] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: American Mathematical Society, 1997.

[29] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, pp. 107–145, 2001.

[30] M. Wytock and J. Z. Kolter, "Contextually supervised source separation with application to energy disaggregation," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2014, pp. 486–492.

[31] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proc. 16th IEEE Int. Conf. Tools AI*, 2004, pp. 576–584.

[32] L. Queiroz, M. A. Roseli, C. Cavellucci, and C. Lyra, "Energy losses estimation in power distribution systems," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1879–1887, Apr. 2012.

[33] R. S. Rao, K. Ravindra, K. Satish, and S. V. L. Narasimham, "Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 317–325, Feb. 2013.

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

Fankun Bu (Student Member, IEEE) received the B.S. and M.S. degrees from North China Electric Power University, Baoding, China, in 2008 and 2013, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. From 2008 to 2010, he was a Commissioning Engineer with NARI Technology Co., Ltd., Nanjing, China. From 2013 to 2017, he was an Electrical Engineer with State Grid Corporation of China at Jiangsu, Nanjing, China. His research interests include distribution system modeling, smart meter data analytics, renewable energy integration, and power system relaying.

Kaveh Dehghanpour (Member, IEEE) received the B.Sc. and M.S. degrees in electrical and computer engineering from the University of Tehran, Tehran, Iran, in 2011 and 2013, respectively, and the Ph.D. degree in electrical engineering from Montana State University, Bozeman, MT, USA, in 2017. He is currently a Postdoctoral Research Associate with Iowa State University, Ames, IA, USA. His research interests include application of machine learning and data-driven techniques in power system monitoring and control.

Yuxuan Yuan (Student Member, IEEE) received the B.S. degree in electrical and computer engineering in 2017 from Iowa State University, Ames, IA, USA, where he is currently working toward the Ph.D. degree. His research interests include distribution system state estimation, synthetic networks, data analytics, and machine learning.

Zhaoyu Wang (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 2009 and 2012, respectively, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2012 and 2015, respectively. He is currently the Harpole-Pentair Assistant Professor with Iowa State University, Ames, IA, USA. He was a Research Aid with Argonne National Laboratory in 2013 and an Electrical Engineer Intern with Corning, Inc., in 2014. His research interests include power distribution systems and microgrids, particularly on their data analytics and optimization. He is the Principal Investigator for a multitude of projects focused on these topics and funded by the National Science Foundation, the Department of Energy, National Laboratories, PSERC, and Iowa Energy Center. He is the Secretary of IEEE Power and Energy Society (PES) Award Subcommittee, a Co-Vice Chair of PES Distribution System Operation and Planning Subcommittee, and the Vice Chair of PES Task Force on Advances in Natural Disaster Mitigation Methods. He is an Editor for the IEEE TRANSACTIONS ON POWER SYSTEMS, IEEE TRANSACTIONS ON SMART GRID, IEEE PES LETTERS, and IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY, and an Associate Editor of *IET Smart Grid*.

Yingchen Zhang (Senior Member, IEEE) received the B.S. degree from Tianjin University, Tianjin, China, in 2003, and the Ph.D. degree from Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2010. He is a Chief Scientist and a Manager of the Sensing and Predictive Analytics Group, National Renewable Energy Laboratory. He is also a Visiting Research Assistant Professor with the University of Denver and an Adjunct Faculty with Colorado State University. He has authored or coauthored more than 100 peer-reviewed publications and holds one U.S. patent. His key areas of expertise lie in sensing and measurement for power systems, artificial intelligence, predictive analytics, advanced energy management system for future grids, the impact of large-scale integration of renewable energies on power system operations.

He previously worked for the California ISO developing and implementing State Estimator and Network Applications. He serves as an Editor for the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY and IEEE POWER ENGINEERING LETTERS.