Disaggregating Customer-Level Behind-the-Meter PV Generation Using Smart Meter Data and Solar Exemplars

Fankun Bu, *Graduate Student Member, IEEE*, Kaveh Dehghanpour, Yuxuan Yuan, *Graduate Student Member, IEEE*, Zhaoyu Wang, *Senior Member, IEEE*, and Yifei Guo, *Member, IEEE*

***Abstract*—Customer-level rooftop photovoltaic (PV) has been widely integrated into distribution systems. In most cases, PVs are installed behind-the-meter (BTM), and only the net demand is recorded. Therefore, the native demand and PV generation are unknown to utilities. Separating native demand and solar generation from net demand is critical for improving grid-edge observability. In this paper, a novel approach is proposed for disaggregating *customer-level* BTM PV generation using low-resolution but widely available hourly smart meter data. The proposed approach exploits the strong correlation between monthly nocturnal and diurnal native demands and the high similarity among PV generation profiles. First, a joint probability density function (PDF) of monthly nocturnal and diurnal *native* demands is constructed for customers *without* PVs, using Gaussian mixture modeling (GMM). Deviation from the constructed PDF is utilized to probabilistically assess the monthly solar generation of customers *with* PVs. Then, to identify *hourly* BTM solar generation for these customers, their estimated monthly solar generation is decomposed into an hourly timescale; to do this, we have proposed a maximum likelihood estimation (MLE)-based technique that utilizes hourly typical solar exemplars. Leveraging the strong monthly native demand correlation and high PV generation similarity enhances our approach's robustness against the volatility of customers' hourly load and enables highly-accurate disaggregation. The proposed approach has been verified using real native demand and PV generation data.**

***Index Terms*—Rooftop photovoltaic, distribution system, Gaussian mixture model, maximum likelihood estimation.**

I. Introduction

In PRACTICE, customer-level rooftop PVs are integrated into distribution systems at behind-the-meter (BTM), where only the net demand is recorded. The measured net demand equals native demand minus the PV generation, which are unknown to utilities separately. The native demand refers to the original demand consumed by home appliances. The invisibility of native demand and BTM solar generation poses challenges in distribution network design [1], [2], operation and expansion [3]–[5], load/PV generation forecasting [6], [7], and demand response [8], [9]. Thus, disaggregating PV generation from net demand is of significance to utilities.

Previous works regarding PV generation disaggregation can be classified into two categories based on the scale of solar power: *Class I - Customer-level approaches:* Customer-level BTM PV generation disaggregation can provide more fine-grained spatial granularity to utilities. Thus, the separated PV generation and native demand for individual customers can be aggregated to obtain the estimate at any higher levels, i.e., service transformer, feeder, or substation. In [10], customer PV generation is estimated by combining a PV performance model with a clear sky model, and using meteorological/geographical data. In [11], a non-intrusive load monitoring (NILM) approach is proposed to disaggregate customers' PV generation from their net demand using measurements with 1-second resolution. In [9], [12], a data-driven method is proposed for estimating the capacity and power output of residential rooftop PVs using customers' net load curve features. In [13], [14], a physical PV performance model is combined with a statistical load estimation model, along with weather data to identify key PV array parameters. The disadvantages of previous customer-level approaches are as follows: dependency on the availability of accurate native demand exemplars, unavailability of PV model parameters, requiring high-resolution sensors and weather data. These obstacles make the previous methods susceptible to the uncertainties of customer behavior and rooftop solar power generators, which result in a decline in disaggregation accuracy.

Class II - System-level approaches: Many previous works have proposed methods to disaggregate solar power from net demand at transformer, feeder, or regional levels. In [15], a data-driven approach is presented for separating the aggregate solar power of groups of customers using their service transformer measurements. In [16], an exemplar-based disaggregator is proposed to separate the output power of an unobservable solar farm from the feeder-level μPMU measurements, using power measurements of nearby observable PV plants and irradiance data. In [6], a regional-scale equivalent PV station model is

Manuscript received September 1, 2020; revised January 26, 2021 and April 14, 2021; accepted April 17, 2021. This work was supported in part by the National Science Foundation under Grant EPCN 2042314 and in part by the
Q1 Advanced Grid Modeling Program at the U.S. Department of Energy Office
Q2 of Electricity under Grant DE-OE0000875. Paper no. TPWRS-01498-2020.
(*Corresponding author: Zhaoyu Wang.*)

The authors 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; yifeig@iastate.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2021.3074614.

Digital Object Identifier 10.1109/TPWRS.2021.3074614

proposed to represent the total generation of small-scale PVs. The model parameters are optimized using known solar power data. In [17], a data-driven approach is proposed to estimate the total rooftop PV generation in a region by installing temporary sensors to measure representative solar arrays. Furthermore, previously in [18], we developed a game-theoretic data-driven approach for disaggregating the PV generation of sizeable groups of customers using solar and load exemplars. However, Class II approaches lack sufficient accuracy for performing customer-level PV disaggregation.

Considering the shortcomings of previous approaches, we propose a novel customer-level solar power disaggregation technique. Our basic idea is to first estimate each customer's *monthly* BTM PV generation and then decompose it into hourly solar power using solar exemplars. Note that in geographically bounded distribution systems, solar exemplars can be easily constructed from observable PVs due to the strong spatial correlation in weather data. Merely having solar exemplars is not sufficient to estimate the unknown PV generation; the relationship between the solar exemplar and unknown PV generation needs to be identified. One promising solution is to construct native demand exemplars. However, accurate customer-level native demand exemplar at the hourly timescale cannot be obtained due to high load uncertainties. To tackle this problem, we exploit an observation from our real smart meter data that the monthly nocturnal and diurnal native demands are highly correlated. Note that this high correlation applies to customers both with and without PVs. Then, identifying the relationship between the solar exemplar and unknown PV generation comes down to making the known monthly nocturnal native demand and the estimated monthly *diurnal* native demand optimally conform to the observed correlation. In other words, to avoid directly identifying the relationship at the hourly timescale, we first identify it at the monthly timescale and then extend the identified relationship to the hourly timescale.

More specifically, the first step is to construct the joint probability density function (PDF) of monthly nocturnal and diurnal native demands for *customers without PVs*. This will be done using a Gaussian Mixture Model (GMM) technique [19], which has demonstrated significant flexibility in forming smooth approximations to arbitrarily-shaped PDFs. The constructed joint PDF captures the monthly load characteristics of customers without PVs; hence, this joint PDF serves as a benchmark for evaluating the deviations caused by monthly BTM solar generation for **customers with unobservable PVs**. The second step is to project the obtained customer-level monthly solar estimations onto hourly values; to do this, the monthly BTM solar generations are represented as a linear weighted summation of solar exemplars with hourly resolution. The weights are optimized using a constrained maximum likelihood estimation (MLE) process, and will be leveraged to disaggregate the hourly net demand of customers with BTM PV generators. To enhance the robustness of MLE against anomalous data, a penalty term is integrated into the weight identification process. Throughout the paper, vectors are denoted using bold italic letters, and matrices are denoted as bold non-italic letters.

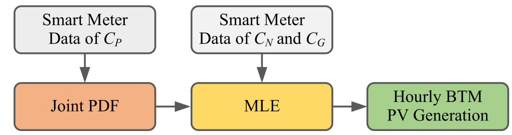


Fig. 1. Overall structure of the proposed customer-level BTM PV generation disaggregation method.

The main contributions of our paper are summarized as follows: (1) Our approach takes full advantage of the strong similarity among small-scale rooftop PV generations. This similarity is due to the fact that the PVs installed within a spatially-bounded distribution system are subject to nearly identical meteorological inputs. (2) The proposed technique utilizes the significant correlation between monthly nocturnal and diurnal native demands. In this way, our approach avoids the direct use of hourly native demand, which is highly volatile at the customer level [20], [21]. (3) Our approach innovatively leverages a soft margin to mitigate the impact of anomalous data samples of solar exemplars. The introduction of this penalty term enhances the robustness of our approach against abnormal measurements.

The rest of the paper is organized as follows: Section II introduces the overall framework for customer-level BTM PV generation disaggregation and describes smart meter dataset. Section III presents the process for constructing joint PDF of monthly diurnal and nocturnal native demands. Section IV describes the procedure of formulating and solving MLE to perform disaggregation. In Section V, case studies are analyzed. Section VI discusses the relevant applications of the disaggregated estimates and Section VII concludes the paper.

II. Overall Disaggregation Framework and Dataset Description

A. Overall Framework

In distribution systems, residential customers can be typically categorized into three types: (I) C_P is the set of customers *without* PVs whose native demand is recorded by smart meters. (II) C_G denotes the small group of customers *with* PVs whose PV generation and native demand are both observable separately. (III) C_N represents the set of customers with PVs whose *net demand* is recorded by smart meter, while their native demand and PV generation are not separately visible. Our goal is to disaggregate PV generation and native demand from the net demand of individual customers in C_N.

The overall process is illustrated in Fig. 1: First, the known monthly nocturnal and diurnal native demands of customers in C_P are employed to construct a joint PDF using GMM modeling technique. This joint PDF is constructed based on a sizeable number of customers without PVs. Then, for each customer in C_N, the unknown PV generation is optimally estimated by performing MLE, and using the constructed joint PDF, known monthly net demand and solar exemplars.

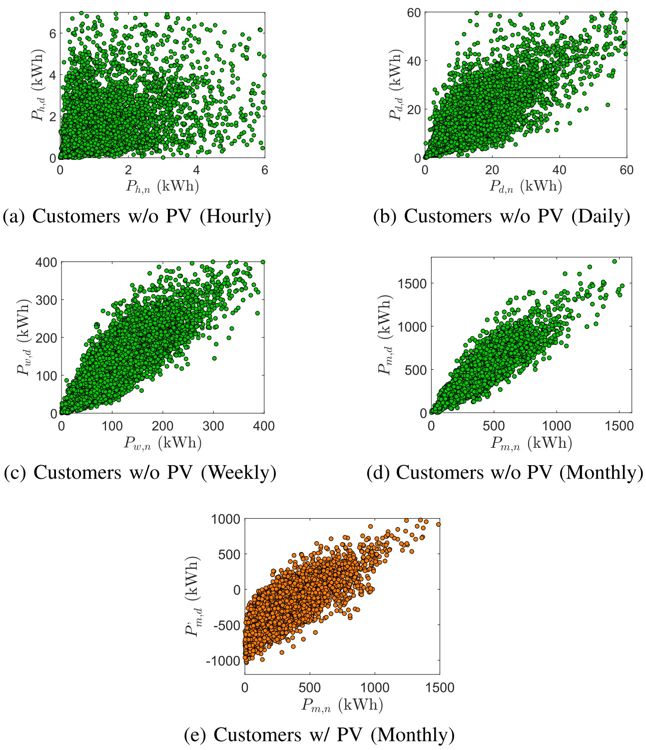


Fig. 2. Observations from real smart meter data.

B. Dataset Description

The hourly native demand data used in this paper are from a Midwest U.S. utility [22], and the hourly PV generation data are from a public dataset [23]. The time range of solar power is one year, and the time range of native demand of customers without PVs is three years. The test system consists of 1120 customers, of which 480 are residential customers without PVs and 237 are residential customers with PVs. Net demand data is obtained by aggregating customers' PV generation and native demand data.

III. Statistical Modeling of Monthly Native Demand

A. Findings From Real Smart Meter Data

One key finding which sets the foundation for the proposed disaggregation approach is that the correlation between nocturnal native demand and the diurnal native demand increases as the observation timescale increases. This finding is illustrated in Fig. 2, where, $P_{h,d}$, $P_{d,d}$, $P_{w,d}$, and $P_{m,d}$ denote the diurnal native demands measured on hourly, daily, weekly, and monthly basis, respectively. $P_{h,n}$, $P_{d,n}$, $P_{w,n}$, and $P_{m,n}$ denote the nocturnal native demands at the corresponding timescales, respectively. $P'_{m,d}$ denotes the monthly diurnal net demand of customers with PVs. Numerically, the correlation coefficients corresponding to Fig. 2(a)–2(d) are 0.56, 0.77, 0.89, and 0.91, respectively. In our paper, we employ the strong correlation of monthly native demand to perform disaggregation. The importance of this correlation is that it can be leveraged to reveal the monthly BTM generation of customers with PVs. For instance, consider two customers, one with PV and one without PV. These two customers can have statistically-similar monthly nocturnal net demand, however, their monthly diurnal net demand will be significantly different from a statistical perspective due to BTM PV generation at daytime. Specifically, Fig. 2(e) shows the nocturnal-diurnal net demand distribution for customers with PV which is significantly different from Fig. 2(d). Thus, the distribution shown in Fig. 2(d), which represents the behavior of customers *without* PV, can be used as a benchmark to determine whether a customer has BTM PV generation and estimate the monthly solar power. These findings have inspired us to construct a joint distribution of monthly nocturnal and diurnal *native* demands of customers *without* PVs to evaluate the deviation caused by the BTM PV generation of customers *with* PVs. These deviations correspond to monthly BTM solar generation.

B. Constructing the Nocturnal-Diurnal Native Demand PDF

We use a parametric PDF estimation technique known as GMM to construct the joint distribution of known monthly nocturnal and diurnal native demands of customers without PVs. A GMM is a linear combination of Gaussian components, and has demonstrated high flexibility and robustness in modeling arbitrary distributions [24]. Since utilities have access to a large amount of native demand data, the constructed GMM-based joint PDF is able to probabilistically describes the quantitative relationship between the monthly nocturnal native demand and monthly diurnal native demand for customers without PVs. The native demand of customers *with* PVs also follow this joint PDF, while their observed monthly net demand can deviate from the joint distribution. Compared with empirical histograms, the GMM-based PDF only has a limited number of parameters, therefore, it can be conveniently leveraged for estimating the BTM PV generation of the customers *with* PVs. In our problem, the GMM approximation model can be described as follows:

$$f(P_{m,n}, P_{m,d}|\boldsymbol{\Lambda}) = \sum_{k=1}^{S} \theta_k g_k(P_{m,n}, P_{m,d}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (1)$$

where, $f(\cdot,\cdot)$ denotes the approximated joint PDF, $P_{m,n}$ and $P_{m,d}$ denote the monthly nocturnal and diurnal native demands of customers without PVs (i.e., customers belonging to C_P), respectively. $\boldsymbol{\Lambda}$ denotes the parameter collection, $\{S, \theta_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$, which needs to be learned based on known native demand data. S denotes the total number of Gaussian components. θ_k's are the weights corresponding to the bi-variate Gaussian components $g_k(\boldsymbol{Z}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with $\boldsymbol{Z} = [P_{m,n}, P_{m,d}]$, which satisfy $\sum_{k=1}^{S} \theta_k = 1$ and $0 \leq \theta_k \leq 1$. The bi-variate Gaussian component is defined as

$$g_k(\boldsymbol{Z}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{Z}-\boldsymbol{\mu}_k)^{\top}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{Z}-\boldsymbol{\mu}_k)\right\}, \quad (2)$$

where, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the Gaussian component mean vector and covariance matrix, respectively.

To learn $\boldsymbol{\Lambda}$, first, a dataset is constructed based on smart meter measurements of customers in C_P. In practice, $P_{m,n}$ and $P_{m,d}$

of customers in C_P are known to utilities and can be obtained from hourly smart meter readings in each month:

$$P_{m,n} = \sum_{t \in I_n} P_h(t), \tag{3a}$$

$$P_{m,d} = \sum_{t \in I_d} P_h(t), \tag{3b}$$

where, $P_h(t)$ denotes the native demand reading at the t'th hour, I_n and I_d denote the sets of nighttime and daytime hours, respectively. Then, we can obtain the matrix of monthly demands by concatenating all customers' monthly native demand pairs:

$$\mathbf{Z} = [\mathbf{Z}(1), \dots, \mathbf{Z}(N_c)]^{\mathsf{T}} \tag{4}$$

where, N_c denotes the total number of customers, and $\mathbf{Z}(j)$ denotes a matrix of monthly nocturnal and diurnal native demand pairs of the j'th customer which is organized as follows:

$$\mathbf{Z}(j) = \begin{bmatrix} P_{m,n}(j,1) & P_{m,d}(j,1) \\ P_{m,n}(j,2) & P_{m,d}(j,2) \\ \vdots & \vdots \\ P_{m,n}(j,N_m) & P_{m,d}(j,N_m) \end{bmatrix}^{\mathsf{T}} \tag{5}$$

where, N_m is the total number of months. Then, we can obtain a dataset of observed monthly demand samples, $\{\boldsymbol{Z}(1), \dots, \boldsymbol{Z}(N')\}$, through partitioning $\mathbf{Z}$ by rows, where, $N' = N_c \times N_m$.

Thus, the problem of GMM approximation boils down to finding optimal parameter collection $\mathbf{\Lambda}^*$ that best fits the obtained dataset of monthly native demands, $\mathbf{Z}$, by assuming that the data samples are drawn independently from the underlying distribution. The most well-established idea for learning GMM parameters is to solve an optimization problem [19], [25], whereby the objective function can be formulated to maximize data likelihood, as follows:

$$\max_{\mathbf{\Lambda}} \prod_{i'=1}^{N'} f\left(\boldsymbol{Z}(i')|\mathbf{\Lambda}\right), \tag{6}$$

By taking the logarithm of objective function, (6) is rewritten as follows:

$$\max_{\mathbf{\Lambda}} \sum_{i'=1}^{N'} \ln\left\{f(\boldsymbol{Z}(i')|\mathbf{\Lambda})\right\}. \tag{7}$$

The optimization problem in (7) is solved using the expectation-maximization algorithm [19].

Based on the identified optimal GMM parameter collection from (7), $\mathbf{\Lambda}^*$, the joint PDF of monthly nocturnal and diurnal native demands can be specifically written as

$$f(P_{m,n}, P_{m,d}) = \sum_{k=1}^{S^*} \theta_k^* g_k^*(P_{m,n}, P_{m,d}), \tag{8}$$

where,

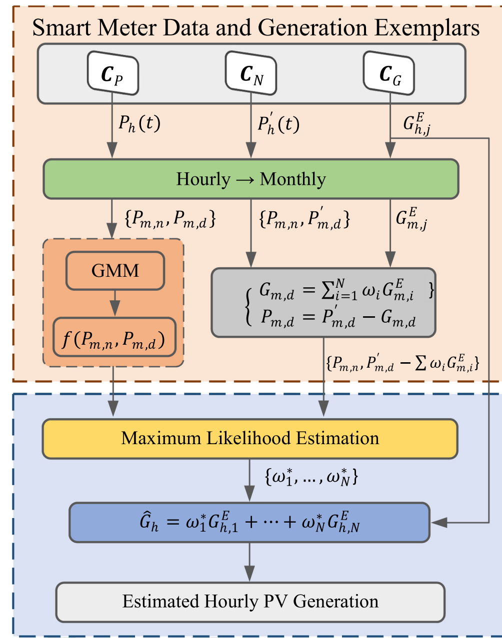


Fig. 3. Detailed structure of the proposed solar disaggregation approach for each customer.

$$g_k^*(P_{m,n}, P_{m,d}) = \frac{1}{2\pi\sigma_{P_{m,n},k}^* \sigma_{P_{m,d},k}^* \sqrt{1-\rho_k^{*2}}} \exp\left\{-\frac{1}{2(1-\rho_k^{*2})}\left[\frac{(P_{m,n}-\mu_{P_{m,n},k}^*)^2}{\sigma_{P_{m,n},k}^{*2}} + \frac{(P_{m,d}-\mu_{P_{m,d},k}^*)^2}{\sigma_{P_{m,d},k}^{*2}} - \frac{2\rho_k^*(P_{m,n}-\mu_{P_{m,n},k}^*)(P_{m,d}-\mu_{P_{m,d},k}^*)}{\sigma_{P_{m,n},k}^* \sigma_{P_{m,d},k}^*}\right]\right\}, \tag{9}$$

where, S^* and θ_k^* are the learned number of mixture Gaussian components and mixture weights, respectively. $\mu_{P_{m,n},k}^*$, $\mu_{P_{m,d},k}^*$, $\sigma_{P_{m,n},k}^*$, $\sigma_{P_{m,d},k}^*$, and ρ_k^* denote the learned mean, variance, and correlation of $P_{m,n}$ and $P_{m,d}$ for the k'th component, respectively.

Using GMM and the learned parameters, the joint distribution of monthly nocturnal and diurnal native demands is optimally represented. This joint distribution can serve as a benchmark for detecting and examining the discrepancy caused by BTM PV generation.

IV. Customer-Level Solar Disaggregation via MLE

In this section, we disaggregate solar generation from net demand for *each customer* with BTM PV using the constructed joint PDF, along with the measured net demand and solar exemplars. The detailed disaggregation process for each customer in C_N is illustrated in Fig. 3.

A. MLE for Optimizing Solar Exemplar Weights

In a geographically bounded distribution system, it can be assumed that different PV arrays are subject to nearly identical meteorological inputs. Under this condition, the signature of an individual PV's generation profile is primarily determined by PV array's maximum power output and azimuth. The maximum power output determines the magnitude of generation curve [9], and the azimuth determines the "skewness" of generation profile [15]. Using the solar power curve of a south-facing PV array as a benchmark, the solar power curve of an east-facing PV array is left-skewed. A west-facing PV array has a right-skewed solar power curve. Therefore, the unknown BTM PV generation can be reliably represented using known generation profiles of BTM PVs (belonging to C_G) with typical orientations that serve as exemplars:

$$G_{m,d} = \sum_{i=1}^{N} \omega_i G^E_{m,i} = \boldsymbol{\omega}^\mathsf{T} \boldsymbol{G}^E_m, \tag{10}$$

where, N is the total number of solar exemplars, $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_N]^\mathsf{T}$ denotes an *unknown* weight vector to be optimized, and $\boldsymbol{G}^E_m = [G^E_{m,1}, \ldots, G^E_{m,N}]^\mathsf{T}$ denotes the PV generation vector of solar exemplars, where, $G^E_{m,i}$ is obtained by converting the known hourly diurnal PV generation into monthly diurnal solar power exemplars:

$$G^E_{m,i} = \sum_{t \in I_d} G^E_{h,i}(t), \tag{11}$$

where, $G^E_{h,i}(t)$ is the PV generation of the i'th exemplar at the t'th hour. Therefore, disaggregating BTM PV generation of each customer in C_N comes down to finding optimal coefficients assigned to known solar exemplars. To do this, first, we represent the unknown monthly diurnal native demand using the known monthly net demand and monthly PV generation of solar exemplars:

$$P_{m,d} = P'_{m,d} - \boldsymbol{\omega}^\mathsf{T} \boldsymbol{G}^E_m. \tag{12}$$

where, $P'_{m,d}$ is the known monthly net demand which can be obtained as follows:

$$P'_{m,d} = \sum_{t \in I_d} P'_h(t), \tag{13}$$

where, $P'_h(t)$ denotes the recorded net demand at the t'th hour.

Since the monthly nocturnal and diurnal native demands of customers *with* PVs probabilistically follow the constructed GMM-based joint PDF, by substituting (12) into (8), we can represent the distribution function for customers with BTM PVs as follows:

$$f\left(P_{m,n}, P'_{m,d} - \boldsymbol{\omega}^\mathsf{T} \boldsymbol{G}^E_m\right). \tag{14}$$

Note that (10)–(14) apply to each month, and we do not add the dimension of month into these equations for the sake of conciseness. Then, the exemplar weight optimization is formulated as an MLE problem over all months, as described as follows:

$$\boldsymbol{\omega}^* = \max_{\boldsymbol{\omega}} \left\{ \prod_{t'=1}^{M} f(P_{m,n}(t'), P'_{m,d}(t'), \boldsymbol{G}^E_m(t') | \boldsymbol{\omega}) \right\}, \tag{15}$$

where, M is the total number of months.

Algorithm 1: Disaggregating BTM PV Generation and Native Demand from Net Demand for *Each Customer*.

```
Classify residential customers into three types: C_P, C_G, and C_N
procedure Data Conversion
  For customers in C_P:
  P_{m,n} ← Σ_{t∈I_n} P_h(t), P_{m,d} ← Σ_{t∈I_d} P_h(t)
  For customers in C_G:
  G^E_{m,i} ← Σ_{t∈I_d} G^E_{h,i}(t)   i = 1, ..., N
  For customers in C_N:
  P_{m,n} ← Σ_{t∈I_n} P'_h(t), P'_{m,d} ← Σ_{t∈I_d} P'_h(t)
end procedure
procedure Construct Nocturnal-Diurnal Native Demand PDF
    For customers in C_P:
    Λ ← {θ_k, μ_k, Σ_k}   k = 1, ..., S
    Λ* ← max_Λ Σ_{i'=1}^{N'} ln{f(P_{m,n}, P_{m,d}|Λ)}
end procedure
procedure Perform MLE for Optimizing Weights
    For customers in C_N:
    P_{m,d} ← P'_{m,d} − ω^T(G^E_m)
    Solve optimization in (16) to obtain ω*
end procedure
procedure Estimate Hourly BTM PV Generation and Native Demand
    For customers in C_N:
    Ĝ_h ← (ω*)^T G^E_h, P̂_h ← P'_h − Ĝ_h
end procedure
```

Further, the optimization solution should be subject to multiple constraints to enforce the identified PV generation to be non-positive and the estimated native demand to be non-negative. Finally, by taking logarithm of (15) and introducing the constraints, the complete optimization problem is elaborated as follows:

$$\max_{\boldsymbol{\omega}} \left\{ \sum_{t'=1}^{M} \ln \left[f(P_{m,n}(t'), P'_{m,d}(t'), \boldsymbol{G}^E_m(t') | \boldsymbol{\omega}) \right] \right\} - \frac{1}{2} \lambda ||\boldsymbol{\beta}||^2_2, \tag{16a}$$

$$\text{s.t.} \quad (\boldsymbol{\omega}^\mathsf{T} \mathbf{G}^E_h)^\mathsf{T} \leq \mathbf{0}, \tag{16b}$$

$$\boldsymbol{P}'_h - (\boldsymbol{\omega}^\mathsf{T} \mathbf{G}^E_h)^\mathsf{T} \geq \boldsymbol{\beta}, \tag{16c}$$

$$\boldsymbol{\beta} \leq \mathbf{0}, \tag{16d}$$

where, $\mathbf{G}^E_h = [\boldsymbol{G}^E_h(1), \ldots, \boldsymbol{G}^E_h(N_h)]$ denotes a matrix of hourly PV generation solar exemplars' time series, $\boldsymbol{G}^E_h(\tau) = [G^E_{h,1}(\tau), \ldots, G^E_{h,N}(\tau)]^\mathsf{T}, \tau = 1, \ldots, N_h$ denotes the vector of solar exemplars' generation readings at the τ'th hour, N_h denotes the total number of hourly demand readings, $\boldsymbol{P}'_h$ denotes the time-series hourly net demand readings and $\mathbf{0}$ represents a zero vector. In addition to maximizing the likelihood function shown in (15), a l_2-norm penalty term, $-\frac{1}{2}\lambda||\boldsymbol{\beta}||^2_2$, is added into the objective function, where, $\lambda \geq 0$ is a tuning parameter and $\boldsymbol{\beta}$ is a vector with non-positive elements. Constraint (16b) ensures that the estimated hourly PV generation is non-positive. Constraints (16c) and (16d) ensure that the estimated time-series native demand is larger than a non-positive vector whose l_2-norm is penalized in the objective function. This penalty term is based on the following consideration: In practice, it is common for the solar generation to have data quality problems. For example, PV arrays can stop running due to solar panel failures, and the recorded anomalous samples are usually smaller than the unrecorded expected values. For the customers whose PV generation is supposed to be disaggregated from the known net demand, the unwanted PV failure does not cause significant disaggregation error. This is because the relatively smaller anomalous PV generation samples cause an unwanted rise in the net demand readings only for a limited number of samples. These larger net demand readings can still give us positive estimated native demand values, since the native demand is estimated by subtracting the disaggregated BTM PV generation from net demand. In comparison, the anomalous readings of *solar exemplars* can cause a negative estimated native demand, which brings significant estimation errors. This is because removing a zero or near-zero PV generation from a negative net demand measurement gives us a negative estimated native demand value. Thus strictly constraining the estimated native demand to be non-negative can cause unwanted errors. Therefore, we have leveraged a soft margin to penalize the effect of anomalous data. Since the purpose of introducing the penalty term is to allow for a small number of negative native demand estimates, the value of tuning parameter, λ, should be chosen in a way to ensure that the number of negative native demand estimates is close to the number of solar exemplars' anomalous data samples. The MLE problem in (16) is solved via numerical optimization using interior-point methods.

B. *Estimating Hourly PV Generation and Native Demand*

By solving the optimization (16), we can obtain the optimized weight vector, $\boldsymbol{\omega}^*$, which is utilized to estimate the unknown hourly BTM PV generation of customers with PVs:

$$\hat{\boldsymbol{G}}_h = (\boldsymbol{\omega}^*)^{\mathsf{T}} \mathbf{G}_h^E. \tag{17}$$

Further, the hourly native demand can be estimated by subtracting the disaggregated BTM PV generation from known net demand readings:

$$\hat{\boldsymbol{P}}_h = \boldsymbol{P}'_h - \hat{\boldsymbol{G}}_h. \tag{18}$$

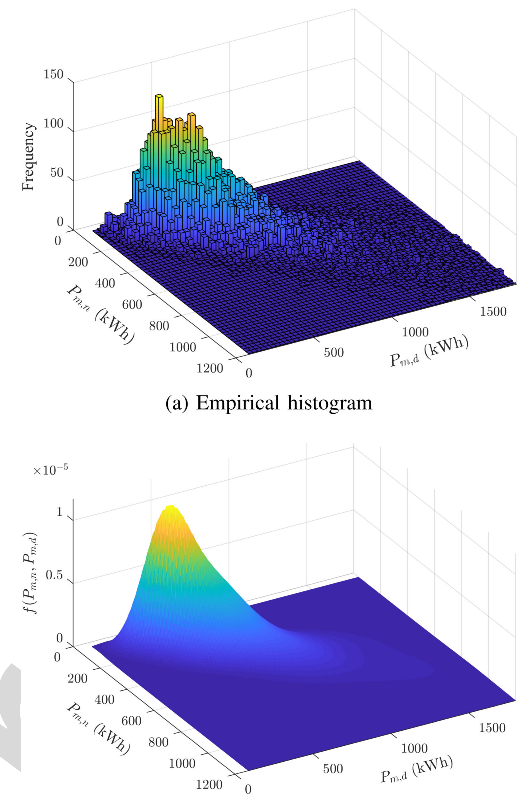
An algorithmic overview of the aforementioned steps of BTM PV generation disaggregation is summarized in Algorithm 1.

V. Case Study

In this section, the proposed customer-level rooftop BTM solar power separation approach is verified using real smart meter and PV generation data described in Section II.

A. *Assessing Statistical Behavior of Customers*

The empirical histogram and the GMM-based estimation of $f(P_{m,n}, P_{m,d})$ are shown in Fig. 4(a) and Fig. 4(b), respectively.



(a) Empirical histogram

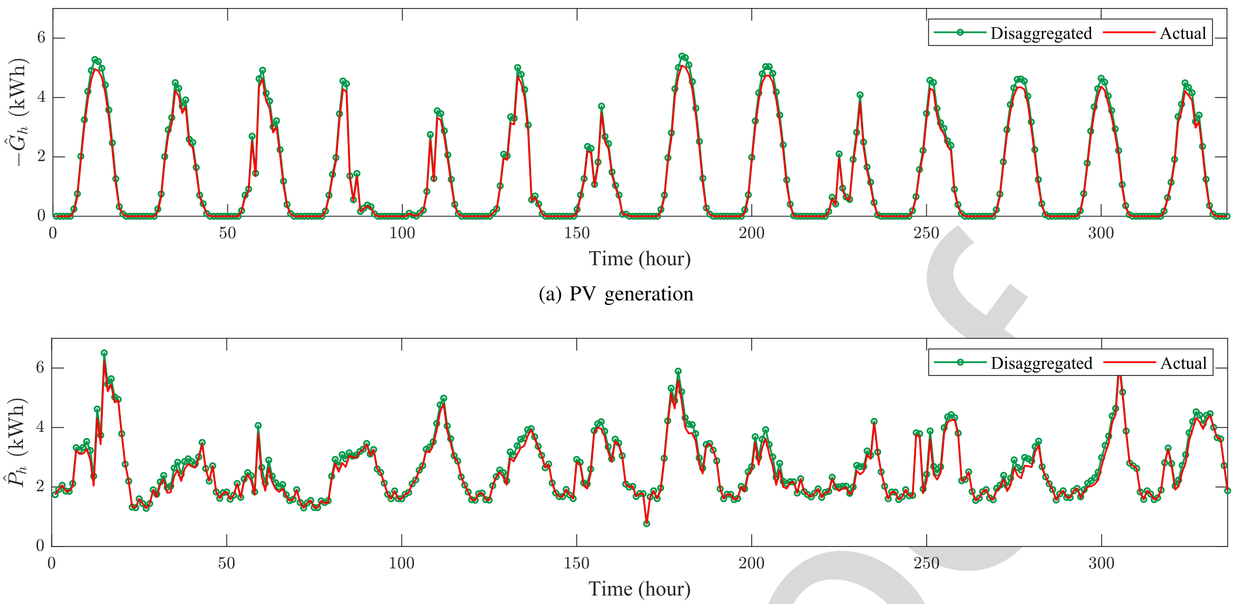
(b) GMM-based estimation

Fig. 4. Joint PDF estimation of monthly nocturnal and diurnal native demands.

Comparing these two figures, it can be seen that GMM is able to accurately model the joint distribution of monthly nocturnal and diurnal native demands using smooth parametric Gaussian density functions. Also note that the joint PDF surface is quite narrow, i.e., the data is highly concentrated around the linear representative of nocturnal and diurnal demands. This corroborates the high correlation between monthly nocturnal and diurnal native demands observed in Fig. 2(d).

B. *BTM PV Generation Disaggregation Validation*

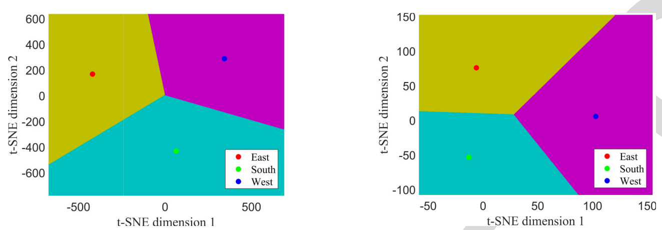
Using the constructed GMM-based joint PDF, along with the known monthly net demand of customers with PVs and PV generation of solar exemplars, we can solve the MLE problem described in (16). When selecting solar exemplars, it is demonstrated that on average, three exemplars can sufficiently represent the PV generation profiles, and introducing additional solar exemplars does not bring further disaggregation accuracy improvement [18]. Thus, we have selected three typical solar power curves from C_G corresponding to PVs facing east, south and west, respectively. Fig. 5 shows disaggregated PV generation and native demand curves of one customer over two weeks, along with corresponding actual profiles. In Fig. 5(a), it can be seen that the disaggregated curve closely fits the actual profile, regardless of the solar volatility on some days. This shows the accurate diaggregation capability of our proposed method and also corroborates our observation that PV generation profiles



(a) PV generation

(b) Native demand

Fig. 5. Two-week disaggregated PV generation and native demand curves, along with corresponding actual curves.



(a) Hourly

(b) Monthly

Fig. 6. Visualizing the distinguishability of time-series PV generation curves of solar exemplars.

with similar PV array orientations are highly correlated. Fig. 5(b) shows the disaggregated and actual native demand profiles. It can be observed that despite the uncertain and volatile pattern of native demand, the disaggregated curve can still fit the real profile.

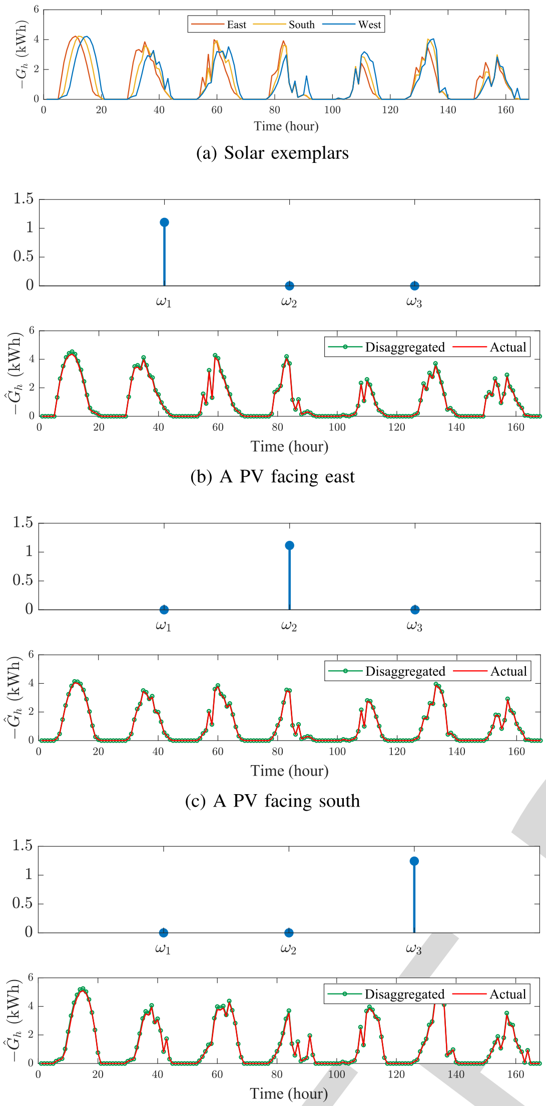
It is of importance to examine the representative feature of typical solar exemplars. In (10), the unknown BTM PV generation is represented using known generation profiles of solar exemplars. Therefore, these PV generation profiles which serve as exemplars should be distinguishable, otherwise, multiple solutions of weights with the same losses can be derived in the MLE optimization process. We have employed a dimensionality reduction technique known as t-SNE to visualize the dissimilarities among PV generation profiles of solar exemplars [26]. Note that each time point is treated as one dimension in our problem. The dimensions of hourly and monthly PV generation time series are reduced for convenient visualization, as shown in Fig. 6. Fig. 6(a) shows the reduced two-dimensional solar power exemplars based on the hourly PV generation of PVs facing east, south and west. As can be seen, the solar exemplars are demonstrated to be distinct. Similarly, the monthly PV generation of solar exemplars also demonstrate distinguishable features, as shown in Fig. 6(b). This is consistent with our observation that solar generation profiles are primarily determined by PV panel orientations in geographically bounded distribution systems.

It is of significance to test whether the proposed approach can track the appropriate exemplars (east, south or west) in the disaggregation process. Fig. 7(a) shows PV generation curves of the three exemplars facing east, south and west. We can see that PVs with different orientations show distinct profile skewness. Fig. 7(b) shows the disaggregated and real PV generation curves of a PV facing east, along with the optimized weights assigned to the three solar exemplars. It can be seen that the weight corresponding to the first exemplar (i.e., PV facing east) is much larger compared to the other two weights, which validates the tracking ability of our proposed approach. This verification can also be observed in Fig. 7(c) and 7(d), which show the weights, disaggregated and actual PV generation curves of PVs facing south and west, respectively. In all cases, our method has accurately detected the correct underlying BTM PV panel orientations.

The proposed customer-level BTM solar separation approach is applied to all 237 customers with PVs, and the disaggregation accuracy for each customer is evaluated in terms of mean absolute percentage error ($MAPE$), which is calculated as follows:

$$MAPE = \frac{100\%}{N_h'} \cdot \sum_{t=1}^{N_h'} \left| \frac{\hat{O}_h(t) - O_h(t)}{\frac{1}{N_h'} \sum_{t=1}^{N_h'} |O_h(t)|} \right| \tag{19}$$

where, N_h' denotes the total number of non-zero PV generation observations for an individual customer, O_h can be P_h or G_h. Fig. 8 shows the distribution of disaggregation error for all



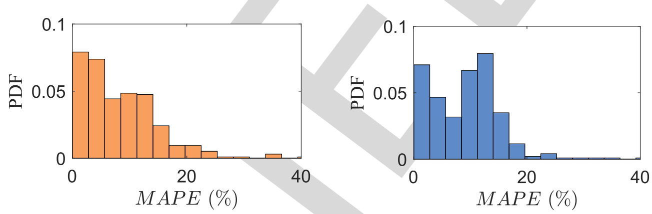
(a) Solar exemplars

(b) A PV facing east

(c) A PV facing south

(d) A PV facing west

Fig. 7. The proposed approach can correctly track proper solar exemplars to perform disaggregation.



(a) PV generation (b) Native demand

Fig. 8. Empirical distribution of $MAPE$ of disaggregated estimates.

customers in terms of $MAPE$. As can be seen, majority of the $MAPE$s are less than 20%. This effectively demonstrates the generalization ability of our proposed method. Table I summarises the empirical cumulative distribution function (CDF) of disaggregation $MAPE$. As can be seen, for the disaggregated hourly PV generation, 80% of the $MAPE$s are less than 13.5%.

TABLE I
EMPIRICAL CDF OF DISAGGREGATION $MAPE$

Empirical CDF	0.2	0.4	0.6	0.8	1.0
$MAPE$ of $\hat{G}_h$ (%)	2.5	4.8	9.7	13.5	33.4
$MAPE$ of $\hat{P}_h$ (%)	3.1	8.3	12.3	14.9	29.1

(a) PV generation (b) Native demand

Fig. 9. Empirical distribution of $RMSE$ of disaggregated estimates.

TABLE II
EMPIRICAL CDF OF DISAGGREGATION $RMSE$

Empirical CDF	0.2	0.4	0.6	0.8	1.0
$RMSE$ of $\hat{G}_h$	0.06	0.12	0.21	0.31	4.51
$RMSE$ of $\hat{P}_h$	0.10	0.28	0.73	1.08	3.85

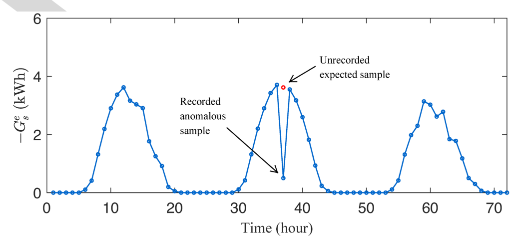


Fig. 10. A solar exemplar with an anomalous sample due to PV failure.

Regarding the disaggregated hourly native demand, 80% of the $MAPE$s are less than 14.9%. This effectively verifies the disaggregation accuracy of our proposed approach.

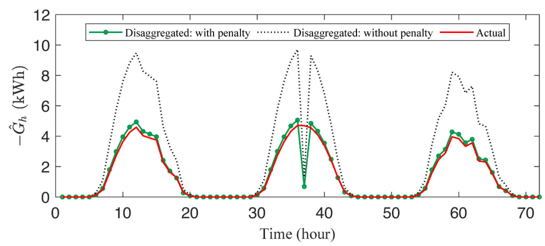
The disaggregation accuracy for each customer is also evaluated using $RMSE$, which is computed as follows:

$$RMSE = \sqrt{\frac{\Sigma_{t=1}^{N'_h}(\hat{O}_h(t) - O_h(t))^2}{N'_h}}. \quad (20)$$

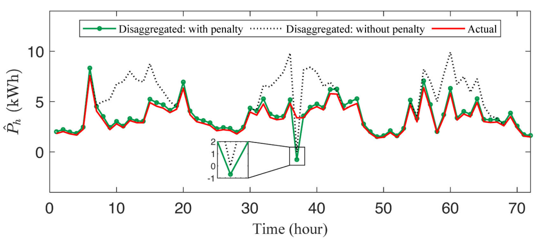
Fig. 9 shows the empirical distributions of the $RMSE$ of disaggregated estimates based on all customers' computed $RMSE$s. It can be seen that most PV generation and native demand $RMSE$s are less than 0.5 and 1.5, respectively. Also, the empirical CDF of disaggregation $RMSE$ is calculated for a comprehensive examination, as shown in Table II.

C. Testing the Robustness of the Proposed Approach

It is common for a practical metering system to have a small number of anomalous measurements in solar exemplars, as shown in Fig. 10, where the unrecorded expected generation is denoted as a red circle. The typical reason for anomalous solar power data samples is PV failure, which causes the recorded data

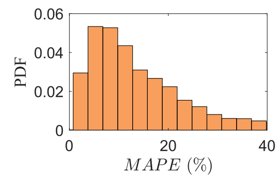


(a) PV generation

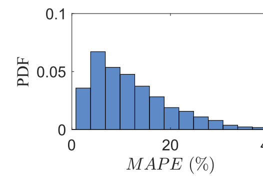


(b) Native demand

Fig. 11. The introduction of penalty term significantly improves disaggregation accuracy and robustness.

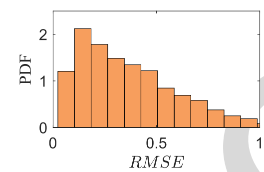


(a) PV generation

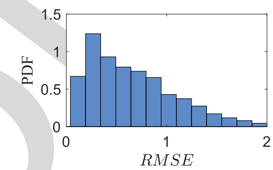


(b) Native demand

Fig. 12. Empirical distributions of $MAPE$ of disaggregated estimates obtained using the Bi-Modeling method.



(a) PV generation



(b) Native demand

Fig. 13. Empirical distributions of $RMSE$ of disaggregated estimates obtained using the Bi-Modeling method.

samples to be smaller than the unrecorded expected values. As previously elaborated in Section IV, a penalty term is included in (16) to mitigate the effect of solar exemplar's anomalous samples. Therefore, it is crucial to test the usefulness of the penalization mechanism. Note that the results in Section V-B are obtained using (16) with a penalty term. Thus, to conduct a performance comparison, we alter (16) to obtain a new optimization formulation with the penalty term omitted, as expressed as follows:

$$\max_{\boldsymbol{\omega}} \quad \sum_{t'=1}^{M} \ln\left[f(P_{m,n}(t'), P'_{m,d}(t'), \boldsymbol{G}_m^E(t')|\boldsymbol{\omega})\right], \tag{21a}$$

$$\text{s.t.} \quad (\boldsymbol{\omega}^\mathsf{T}\mathbf{G}_h^E)^\mathsf{T} \leq \mathbf{0}, \tag{21b}$$

$$\boldsymbol{P}'_h - (\boldsymbol{\omega}^\mathsf{T}\mathbf{G}_h^E)^\mathsf{T} \geq \mathbf{0}. \tag{21c}$$

Then, using the solar exemplar with an anomalous sample in Fig. 10, we utilize (16) and (21) to perform disaggregation, respectively. Fig. 11 compares three-day disaggregated PV generation and native demand curves based on (16) and (21), respectively. The actual solar power and native demand curves are also plotted as benchmarks. In Fig. 11(a), it can be seen that the disaggregated PV generation curve using (16) can closely fit the actual curve except for at the hour that the solar exemplar's anomalous sample appears. In comparison, the disaggregated PV generation curve using (21) significantly deviates from the actual benchmark. Regarding the disaggregated native demand, we can draw the same conclusion by observing Fig. 11(b). The overestimation of PV generation and native demand using (21) is due to the constraint that forces the estimated native demand to be strictly non-negative, as shown in Fig. 11(b). In contrast, our approach presented in (16) allows a negative native demand estimate to mitigate the anomalous samples' impact. To sum up, the introduction of penalty into the MLE optimization significantly enhances the robustness of our proposed approach against anomalous data.

D. Performance Comparison

It is vital to compare the performance of our proposed approach with other methods. Since the proposed approach in [14] has been demonstrated to have a relatively better performance than previous methods, we first apply the proposed approach in [14] to conduct PV generation disaggregation using our dataset and then compare its performance with our approach. The approach to be compared is denoted as Bi-Modeling, which employs a statistical model and a physical model to represent the native load and the PV generation, respectively. The Bi-Modeling method utilizes the observable net load series and weather data to optimize model parameters iteratively. A threshold is set to evaluate whether the two models reach a consensus. The results obtained by applying the Bi-Modeling method to our dataset are shown in Figs. 12 and 13. It can be seen that our approach has a better performance than the Bi-Modeling method in terms of the $MAPE$ and $RMSE$ of PV generation by comparing Fig. 12(a) and 13(a) with Fig. 8(a) and 9(a), respectively. In terms of *native demand* disaggregation error comparisons (obtained from Fig. 8(b), Fig. 9(b), Fig. 12(b), and Fig. 13(b)), the results are inconclusive. Further results in terms of *average* $MAPE$ and $RMSE$ are examined as shown in Table III, and it can be seen that our approach demonstrates smaller disaggregation errors. Note that no single method alone is best in all situations.

VI. Application Discussion

It is essential to discuss how the disaggregated PV and native demand can be used in practice. These estimates target static applications since the sampling rates of widely available smart meter data are 1-hour, 30-min, or 15-min. To further explain the

TABLE III
AVERAGE $MAPE$ AND $RMSE$ OF ESTIMATES

Metrics	Our Approach	Bi-Modeling
Average $MAPE$ of $\hat{G}_h$ (%)	10.2	16.1
Average $MAPE$ of $\hat{P}_h$ (%)	9.64	12.4
Average $RMSE$ of $\hat{G}_h$	0.23	0.38
Average $RMSE$ of $\hat{P}_h$	0.61	0.69

usefulness of our approach, we primarily focus on three specific applications:

A. Native Load Monitoring and Forecasting

Since small-scale rooftop PVs can be disconnected or otherwise absent without prior knowledge, utilities usually adopt a conservative approach in distribution system studies and do not treat small PVs as reliable sources [3]. As a result, utilities use the *native* load for conducting conservative scenario analysis instead of the *net* load. Therefore, it is crucial for utilities to monitor the actual native load. In most cases, small-scale rooftop PVs are installed BTM, and only the net load is recorded. Thus, it is necessary to disaggregate the unknown native load and PV generation from the known net load. Our proposed approach can directly provide utilities the estimated native load, which can be further utilized for system operation and design.

The disaggregated estimates can also be used for native load forecasting. As the PV penetration level increases, the native load can be seriously masked by PV generation. Under this condition, it is necessary to separate the native load from the net load first and then perform native load forecasting. For this application, our proposed approach can provide native load estimates to train native load forecasting models.

B. Demand Response

Due to the existence of BTM PVs, the native demand is masked by PV generation. However, the majority of demand response schemes are designed for native load controlling [9]. Under this condition, the unknown native demand hinders utilities from applying demand response schemes efficiently because of the invisibility of the real power consumption. Therefore, the native demand of individual customers needs to be separated from the net demand, as our proposed approach fulfills.

C. Service Restoration

Another application is relevant to service restoration. When restoring cold loads, more power will be drawn by air-conditioning appliances than in normal operation. This power increase is caused by the simultaneous restarting of a large number of appliances and can be several times larger than the normal load. Thus, this abnormal load should be estimated for developing optimal service restoration tactics. One typical way of estimating the abnormal load is to multiply the normal native load before outage by a ratio from a look-up table [3], [27]. To do this, we need to separate the normal native load from the net load. Leveraging the disaggregated native load estimate obtained from our approach can be used in optimizing restoration strategies.

VII. CONCLUSION

This paper presents a novel robust approach to disaggregate invisible customer-level BTM PV generation and native demand from net demand using smart meter data and solar exemplars. The proposed method employs a limited number of observable solar power exemplars to represent the invisible BTM PV generation. Also, the proposed approach innovatively leverages the significant correlation between nocturnal and diurnal native demands at the timescale of monthly to alleviate the hourly native demand's volatility. In addition, a penalty term is innovatively integrated into the estimation problem to tackle anomalous samples of solar exemplars due to PV failures. The numerical experiments verify that the approach is able to perform disaggregation with excellent accuracy and robustness, which further improves utilities' situational awareness of grid-edge resources.

The key findings of the paper are summarized as follows: (1) Using real BTM PV generation and native demand data, we have observed that the hourly generation series of a PV can be sufficiently represented using solar power outputs of PVs with similar orientations. In comparison, the hourly customer-level native demand shows higher volatility. (2) Despite the uncertainty of *hourly* native demand, the monthly nocturnal and diurnal native demands are highly correlated. This has inspired us to first estimate the monthly PV generation, then decompose it into hourly solar power. (3) The anomalous data of PV generation is common in practice, and can cause significant disaggregation error. This has motivated us to introduce a penalty term into MLE to reduce the impact of solar exemplars' anomalous samples.

REFERENCES

[1] 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.

[2] Y. Zhang, J. Wang, and Z. Li, "Uncertainty modeling of distributed energy resources: Techniques and challenges," *Curr. Sustain. Energy Rep.*, vol. 6, no. 2, pp. 42–51, Jun. 2019.

[3] R. Seguin, J. Woyak, D. Costyk, J. Hambrick, and B. Mather, "High penetration PV integration handbook for distribution engineers," *Nat. Renewable Energy Lab.*, Golden, CO, USA, 2016. Q3

[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] K. Sun, Y. Hou, W. Sun, and J. Qi, *Renewable and Energy Storage in System Restoration*. Hoboken, NJ, USA: Wiley-IEEE Press, 2019.

[6] 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.

[7] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Convers. Manage.*, vol. 212, no. 112766, pp. 1–14, May 2020.

[8] Z. Xuan, X. Gao, K. Li, F. Wang, X. Ge, and Y. Hou, "PV-load decoupling based demand response baseline load estimation approach for residential customer with distributed PV system," *IEEE Trans. Ind Appl.*, vol. 56, no. 6, pp. 6128–6137, Nov./Dec. 2020.

[9] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, N. Duić, and T. Wang, "Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation," *Appl. Energy*, vol. 253, 2019, Art no. 113595.

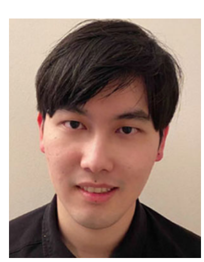
[10] D. Chen and D. Irwin, "Sundance: Black-box behind-the-meter solar disaggregation," in *e-Energy*, May 2017, pp. 16–19.
[11] C. Dinesh, S. Welikala, Y. Liyanage, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Non-intrusive load monitoring under residential solar power influx," *Appl. Energy*, vol. 205, pp. 1068–1080, Aug. 2017.
[12] 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.
[13] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Estimation of behind-the-meter solar generation by integrating physical with statistical models," in *Proc. IEEE SmartGridComm*, Oct. 2019, pp. 1–6.
[14] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Joint estimation of behind-the-meter solar generation in a community," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 682–694, Jan. 2021.
[15] 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.
[16] 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, pp. 112–121, 2018.
[17] 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.
[18] 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, Jul. 2020.
[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2009.
[20] S. Wang, Y. Dong, L. Wu, and B. Yan, "Interval overvoltage risk based PV hosting capacity evaluation considering PV and load uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2709–2721, May 2020.
[21] F. Bu, K. Dehghanpour, Y. Yuan, and Z. Wang, "Quantifying load uncertainty using real smart meter data," in *Proc. IEEE Smart Grid Commun.*, 2020, pp. 1–6.
[22] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A time-series distribution test system based on real utility data," in *Proc. North Amer. Power Symp.*, Oct. 2019, pp. 1–6.
[23] K. Nagasawa, C. R. Upshaw, J. D. Rhodes, C. L. Holcomb, D. A. Walling, and M. E. Webber, "Data management for a large-scale smart grid demonstration Project in Austin, Texas," in *ASME 2012 6th Int. Conf. Energy Sustainability, Parts A. B.*, Jul. 2012, pp. 1027–1031.
[24] D. A. Reynolds, "Gaussian mixture models," in *Encyclopedia Biometrics* 2nd ed. 2015, pp. 827–832.
Q4 [25] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. Springer Series in Statistics, 2001.
Q5 [26] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *JMLR*, vol. 9, pp. 2579–2605, 2008.
[27] K. P. Schneider, E. Sortomme, S. S. Venkata, M. T. Miller, and L. Ponder, "Evaluating the magnitude and duration of cold load pick-up on residential distribution using multi-state load models," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3765–3774, Sep. 2016.

Fankun Bu (Graduate 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 for NARI Technology Company Ltd., Nanjing, China. From 2013 to 2017, he was an Electrical Engineer for State Grid Corporation of China, Jiangsu, Nanjing, China. His research interests include distribution system modeling, smart meter data analytics, renewable energy integration, and power system relaying.

Kaveh Dehghanpour 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. His research interests include application of machine learning and data-driven techniques in power system monitoring and control.

Yuxuan Yuan (Graduate Student Member, IEEE) received the B.S. degree in 2017 in electrical and computer engineering 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 (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiaotong University, Shanghai, China, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA. He is currently a Harpole-Pentair Assistant Professor with Iowa State University, Ames, IA, USA. His research interests include optimization and data analytics in power distribution systems and microgrids. 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 Economic Development Authority. He is the Chair of IEEE Power and Energy Society (PES) PSOPE Award Subcommittee, Co-Vice Chair of PES Distribution System Operation and Planning Subcommittee, and Vice Chair of PES Task Force on Advances in Natural Disaster Mitigation Methods. He is the Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, IEEE TRANSACTIONS ON SMART GRID, IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY, IEEE POWER ENGINEERING LETTERS, and *IET Smart Grid*. He was the recipient of the National Science Foundation CAREER Award, the IEEE PES Outstanding Young Engineer Award, and the Harpole-Pentair Young Faculty Award Endowment.



Yifei Guo (Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Shandong University, Jinan, China, in 2014 and 2019, respectively. He is currently a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. From 2017 to 2018, he was a Visiting Student with the Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark. His research interests include voltage or var control, renewable energy integration, wind farm control, distribution system optimization and control, and power system protection.