An Unsupervised Similarity-based Method for Estimating Behind-the-Meter Solar Generation

Kang Pu* and Yue Zhao†

*Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA

†Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794, USA

Emails: {kang.pu, yue.zhao.2}@stonybrook.edu

Abstract—**Accurate knowledge of solar generation in power distribution systems provides great values to utilities for efficient and reliable distribution system operation. However, many solar PV resources are installed behind-the-meter (BTM), and as a result only the net load measurements are available to the utilities. In this paper, a high-performance method for disaggregating BTM solar generation traces from net load traces is developed. The algorithm takes the net load data measured by smart meters and other widely available environmental measurements (e.g., publicly monitored solar irradiance and temperature) as inputs, and disaggregates the net load traces into BTM solar generation and load traces. Notably, the proposed method does not rely on any separately metered data of BTM solar generation. Rather, in a fully unsupervised fashion, the proposed method effectively exploits the self-similarity and cross-customer similarity of customer loads to achieve accurate BTM solar disaggregation. The developed unsupervised method is evaluated on two real-world smart meter data sets collected from New York and Texas, and exhibits very high performance that closely approaches the ideal performance bound from supervised learning.**

I. Introduction

There has been a rapid and continuing growth of distributed generation, in particular rooftop solar, in the power distribution systems around the world. These distributed solar are typically connected to the grid "behind the meters" (BTM) installed by the electric utilities. Indeed, it is economically infeasible for the utilities to deploy separate meters and communication systems for all the distributed solar. As such, the utilities do not have access to separate readings of BTM solar, and only the "net loads" of customers are metered. It is however immensely valuable for the utilities to estimate the BTM solar generation traces [1]. Such estimation will enable utilities to greatly improve its efficiency and reliability of distribution system operation and planning.

There has been an extensive recent body of work addressing the BTM solar generation estimation problem [2]. A majority of them exploit the increasing deployment of advanced metering infrastructure (AMI) in the distribution systems. On the high level, some of the existing methods are fully data-driven, while the others in addition seek to exploit the knowledge from the physical models of solar generation. **Data-Driven Methods:** In [3], a linear proxy-based estimator that disaggregates solar generation at the substation level is built. Solar generation data from nearby PV systems are utilized to build the contextually supervised source separation model in [4] to disaggregate solar generation at an individual level. Dimension reduction and mapping functions are deployed to estimate total power generation of solar power sites in [5]. Information from customers without PV is utilized to build the consumer mixture model in [6]. In [7], solar exemplars are utilized for building a maximum likelihood based solar estimator. [8] introduces a game theory based method which casts the solar disaggregation problem as a bi-layer optimization problem. **Model-Based/Assisted Methods:** Although fully data-driven methods can achieve some success, however, many require the ground truth solar generation of at least a subset of the customers as supervision signals [4][7][8]. They also tend to require a significant amount of data to learn from especially if there is no ground-truth solar generation data available. Model-based/assisted methods generally require less data as the physical model can serve as strong regularization in learning. A probabilistic model is introduced in [9] for estimating BTM PV solar generation at an aggregate level. In [10] a disaggregation algorithm is developed based on (among other techniques) finding a sufficient amount of "clear sky" periods of time in the data. In [11], an algorithm is developed that iteratively learns a physical model and a mixed Hidden Markov model to estimate solar PV generation and electric load respectively. Other related works include [12] which studies disaggregating an aggregate load into individual loads by utilizing partially labeled data, and [13] which further utilizes faster sensors in the distribution system such as distribution Phasor Measurement Units.

In this paper, we developed a novel high-performance BTM solar generation estimation method that disaggregates BTM solar traces from net load traces measured by smart meters. Our focus here is to estimate BTM solar generation at certain *aggregate* levels (e.g., transformer and feeder levels). On the one hand, this addresses important needs for utilities. On the other hand, this also implies that the method can be adapted to use cases even in the absence of smart meters, as long as aggregate-level net load data are available. Importantly, our algorithm does not utilize any ground-truth BTM solar generation data at all. Such "unsupervised" nature of our method allows it to be widely applicable in practice. Specifically, leveraging a general physical model of solar generation, our method introduces a fundamental *similarity-based* principle for estimating the unknown physical model parameters of BTM solar. As such, the proposed approach is able to achieve great performance in an unsupervised fashion even with a limited amount of net load data. As demonstrated in our performance evaluation on two real-world data sets collected from NY and TX, USA, the latter of which containing only four weeks of data, the performance of our unsupervised learning method is very close to the performance bound

This work was supported by NYSERDA with Agreement Number 159613.

provided by the ideal supervised learning method.

II. Problem Formulation

Consider a set of smart meter data. A customer can possibly have solar PV behind the meter. The smart meters measure the *net loads* of the customers. Customer k's smart meter reading at time t is denoted by

$$n_k(t) = l_k(t) - g_k(t), \tag{1}$$

where $n_k(t), l_k(t)$ and $g_k(t)$ are the net load, load and solar generation (all of which are real power), respectively. The smart meter data are collected at discrete time instances such as every 15 minutes. We denote such time indices by $t = 1, \ldots, T$. Furthermore, for a set of customers denoted by S, we denote the *total* net load of the customers in this set by

$$n_S(t) = \sum_{k \in S} n_k(t), \tag{2}$$

and similarly define $l_S(t), g_S(t)$. Based on the smart meter readings, i.e., $\{n_k(t), \forall k, t\}$, our objective is the following: for an "aggregate" customer by summing over a set of individual customers denoted by A, disaggregate the sum-net-load $n_A(t)$ into $l_A(t)$ and $g_A(t)$, so that $n_A(t) = l_A(t) - g_A(t)$.

Clearly, to perform the above disaggregation, additional information beyond the metered net load data themselves is needed. We will discuss in detail in the following sections what and how such information can be exploited to achieve our objective with high performance.

III. Proposed Methodology

In this section, we discuss the methods for disaggregating net load $n_A(t)$ into load $l_A(t)$ and solar generation $g_A(t)$. We first highlight the following remark in detecting whether a customer has solar PV installed or not.

Remark (Detection of Presence of BTM Solar PV)**.** *By observing the smart meter data of a customer, whether this customer has solar PV or not can be detected with perfect accuracy using the following simple rule: a) If the net load $n_k(t)$ has ever reached a negative value, meaning that $g_k(t) > l_k(t)$, the customer must have solar PV; And b) If no such negative net load is observed within a sustained period (empirically, one month is sufficient,) the customer must not have PV.*

The above rule of detecting the presence of BTM PV achieves perfect accuracy in all the real-world data sets we examined. We now continue to address the main problem in this section — BTM solar generation estimation for an aggregation of customers.

A. Physical Model of Solar Generation

For an installed solar panel, its solar generation follows an accurate physical model that depends on a number of physical quantities including technical parameters of the solar panel and weather related variables. We start with the following equation [14]:

$$P_t \approx C \frac{I_{PV,t}}{I_{ref}} \left[1 - \mu \left(T_{PV,t} - T_{ref}\right)\right], \tag{3}$$

where P_t is the solar generation; C denotes the capacity of the panel; $I_{PV,t}$ denotes the solar irradiance that strikes on the solar panel (W/m^2); $T_{PV,t}$ denotes the temperature ($^\circ$C) of PV cells; I_{ref} and T_{ref} are reference irradiance and cell temperature with 1000 (W/m^2) and $25\,^\circ$C as their typical values respectively; μ denotes the temperature coefficient with $-0.5\%/^\circ$C as a typical value [10]. The cell temperature is related to many factors such as weather metrics and solar irradiance, and can be approximated by the following [14]:

$$T_{PV,t} = T_{A,t} + \frac{I_{PV,t}}{800} \times \left(N_{oct} - 20\right), \tag{4}$$

where $T_{A,t}$ denotes the ambient air temperature ($^\circ$C) and N_{oct} is the Nominal Operating Cell Temperature with $48\,^\circ$C as a typical value [10]. The solar irradiance received on solar panels can be calculated as follows:

$$\begin{aligned} I_{PV,t} &= I_{0,t}\tau_{b,t}(\sin\alpha\cos\beta + \cos\alpha\sin\beta\cos(\gamma - A)) \\ &+ I_{d,t}\left(\frac{1+\cos\beta}{2}\right) + \left(I_{b,t} + I_{d,t}\right)\rho_t\left(\frac{1-\cos\beta}{2}\right). \end{aligned} \tag{5}$$

The notations are explained as follows. $\tau_{b,t}$ and ρ_t denotes the atmospheric transparent coefficient and surface albedo with 0.74 and 0.2 as typical values respectively [15]. $I_{0,t}, I_{d,t}$, and $I_{b,t}$ are the direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and direct horizontal irradiance, respectively. DNI and DHI are directly measured and can be obtained from the NSRDB dataset [16]. The direct horizontal irradiance can be calculated as follows:

$$I_{b,t} = I_{0,t}\tau_{b,t}\sin\alpha, \tag{6}$$

$$\alpha = 90^\circ - \zeta, \tag{7}$$

where α and ζ denote the elevation of the sun and zenith of the sun receptively, and the latter can also be obtained from the NSRDB dataset [16]. β and γ in (5) are the tilt angle and azimuth angle of the solar PV panels which are typically unknown to the utilities and need to be estimated. A in (5) denotes the sun's azimuth and can be calcuated as follows:

$$\cos A = \frac{\sin\delta\cos\phi - \cos\delta\sin\phi\cos\omega}{\cos\alpha}, \tag{8}$$

where δ and ϕ denote the declination angle and geographical latitude which can be readily calculated based on time.

In sum, we denote the overall physical model of solar generation by $f(\boldsymbol{x}(t); \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ contains the three panel-dependent model parameters — panel capacity C, tilt angle β, and azimuth angle γ — that are unknown and need to be estimated, and $\boldsymbol{x}(t)$ contains all the relevant input measurement data (i.e. DHI, DNI, solar zenith angle, temperature, longitude, latitude, time of day, and day of year.) Notably, all these input data $\boldsymbol{x}$ (or at least reasonable estimates of them) are indeed available to the utilities. As such, the problem of BTM solar generation estimation is then cast as the problem of estimating the unknown physical model parameters $\boldsymbol{\theta}$. While the above physical model is derived for a single solar panel, we will use it as an approximate model for the collection of a set of solar panels of an aggregation of customers.

B. Fundamental Idea: Similarity Based Unsupervised Learning of Physical Model Parameters

In principle, as $\boldsymbol{\theta}$ consists of only *three* parameters to learn — panel capacity, tilt angle, and azimuth angle, if we can obtain just *three equations* of these parameters, $\boldsymbol{\theta}$ can then be solved. In an ideal "supervised learning" situation, if we can observe the solar generation $g(t)$ at three different times t_1, t_2 and t_3, by solving the three equations $\{f(\boldsymbol{x}(t_i);\boldsymbol{\theta}) = g(t_i),\ i = 1,2,3\}$, we can obtain an estimate of the physical model parameters $\boldsymbol{\theta}$.

However, BTM solar generation $g(t)$ is not known and is precisely what we want to estimate based on the other information that is indeed known — net load data $n(t)$, and other external input information $\boldsymbol{x}(t)$. To estimate the physical model parameters $\boldsymbol{\theta}$, we again seek to find *a sufficient number of equations that constrain* $\boldsymbol{\theta}$ *but without relying on any information of the solar generation* g. To achieve this, let us consider the following conceptual example.

If we know for a fact that, at two different time instances t and t', the loads are the same, we would have the following series of equations:

$$\begin{aligned} l(t) = l(t') &\Leftrightarrow n(t) + g(t) = n(t') + g(t') \\ &\Leftrightarrow n(t) + f(x(t);\boldsymbol{\theta}) = n(t') + f(x(t');\boldsymbol{\theta}). \end{aligned} \quad (9)$$

We note that, as long as the external input variables at these two time instances, $\boldsymbol{x}(t)$ and $\boldsymbol{x}(t')$, are *not identical*, (9) is a valid/non-trivial equation of $\boldsymbol{\theta}$ because, other than $\boldsymbol{\theta}$, everything in (9) is *known*. In other words, even if we do not know the BTM load $l(t)$ (or equivalently the BTM solar generation $g(t)$), knowing relations such as $l(t) = l(t')$ would still provide a valid equation of $\boldsymbol{\theta}$.

In practice, knowing only the net loads $n(t)$, we cannot know for sure if the load $l(t)$ at two different times are exactly the same. Nonetheless, we can make predictions of when the loads are likely to be *sufficiently similar*. As such, if we can collect a sufficiently large number of *pairs of time instances at which the loads are sufficiently similar*, an estimate of the physical model parameters $\boldsymbol{\theta}$ can be computed by solving an over-determined set of equations as a learning problem. In the next section, we formalize and present the details of the above similarity based approach.

C. Algorithm Design for BTM Solar Generation Estimation

Consider a set of customers in an area with BTM solar PV, denoted by S_P, and the aggregate customer formed by summing them up. In real-world situations, there are typically also a significant number of customers who do *not* have PV. Our first step is to *form an aggregate "non-PV" customer* by summing up a set of customers *in the same area but without PV*, denoted by S_N. Given the regularities of human behaviors especially on the aggregate level, we will exploit the inherent *similarity between the load pattern* of the aggregate PV customer and that of the aggregate non-PV customer:

$$\frac{1}{|S_P|} l_{S_P}(t) \approx \kappa \frac{1}{|S_N|} l_{S_N}(t). \quad (10)$$

For short, we term the aggregate PV customer "customer S_P" and the aggregate non-PV customer "customer S_N".

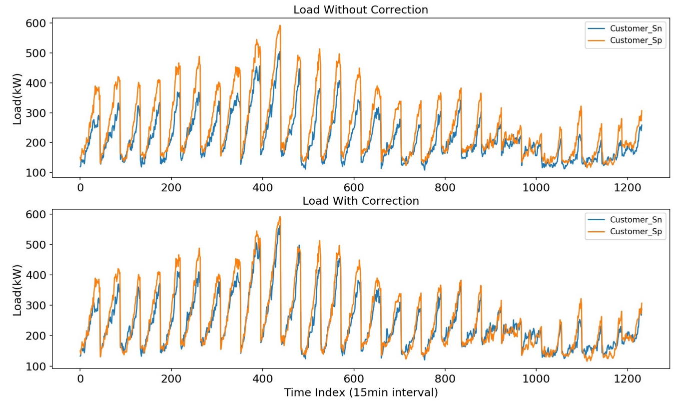

Figure 1: Load scale correction for an aggregation of customers in Austin, TX.

$|S_P|$, $|S_N|$ are the total number of PV customers and non-PV customers. κ is a scale factor used to correct the load scale difference between customer S_P and customer S_N. To estimate κ, we first select a set of time slots, denoted by s_{low}, from 8am-6pm with very low solar irradiance (i.e., $GHI < 20$). κ is then calculated as $\kappa = \frac{\sum_{t \in s_{low}} n_{S_P}(t)}{\sum_{t \in s_{low}} n_{S_N}(t)}$. An example demonstrating the similarity between the loads of customers S_P and S_N, before and after the scale correction, is plotted in Figure 1.

Self-Similarity Over Time: The first key step in our approach is to select pairs of time slots in which customer S_P likely has similar loads. We term such a pair of time slots a "similar time pair". Since the loads of customer S_P are not measured, we leverage the following facts to find such pairs:

- *Temporal correlations*: The loads in time slots close (or similar) to each other are likely to be similar due to the smoothness/regularity of the aggregate human behaviors.
- *Cross-Customer correlations*: If the loads of customer S_N, which are equal to the metered net loads due to the absence of PV, are similar in a pair of time slots, it is likely that the loads of customer S_P are also similar in the same pair of time slots due to the inherent behavioral similarity between the two aggregate customers.

Accordingly, we design a scoring system to measure the similarity between any two time slots. The similarity score s is composed of day (of year) similarity s_1, minute (of day) similarity s_2 and load similarity s_3 which are calculated as follows:

$$s_1 = \frac{365^{\frac{1}{n_1}} - (|DOY(t_i) - DOY(t_i')|)^{\frac{1}{n_1}}}{365^{\frac{1}{n_1}}}, \quad (11)$$

$$s_2 = \frac{1440^{\frac{1}{n_2}} - (|MOD(t_i) - MOD(t_i')|)^{\frac{1}{n_2}}}{1440^{\frac{1}{n_2}}}, \quad (12)$$

$$s_3 = \frac{MLD^{\frac{1}{n_3}} - (|l_{S_N}(t_i)) - l_{S_N}(t_i')|)^{\frac{1}{n_3}}}{MLD^{\frac{1}{n_3}}}, \quad (13)$$

$$s(t_i, t_i') = s_1 + s_2 + s_3, \quad (14)$$

where $DOY(t_i)$, $MOD(t_i)$ and MLD denotes the day of year, minute of day of time t_i and the maximum load difference of customer S_N. n_1, n_2, n_3 are three parameters that indicate our tolerance of dissimilarity (the higher the value the lower the tolerance). In our experiment, n_1, n_2, n_3 are set to be 1.5, 2.0, and 3.0 respectively. Finally, a number (in our study 250) of top time pairs with the highest similarity scores are selected to form a training data set denoted by $\mathcal{T}$. In a spirit similar to (9), a loss term is then defined as

$$L^{(1)} = \sum_{(t_i, t_i') \in \mathcal{T}} \left((n_{S_P}(t_i) + f(\boldsymbol{x}_{S_P}(t_i); \boldsymbol{\theta}_{S_P}) \right) - \left(n_{S_P}(t_i') + f(\boldsymbol{x}_{S_P}(t_i'); \boldsymbol{\theta}_{S_P}) \right)^2. \quad (15)$$

Cross-Customer Similarity: We further define another loss term motivated by the similarity between customer S_P and S_N (cf. (10)):

$$\begin{aligned} L^{(2)} &= \sum_{t=1}^{T} \left(n_{S_P}(t) + f(\boldsymbol{x}_{S_P}(t); \boldsymbol{\theta}_{S_P}) - \kappa \frac{|S_P|}{|S_N|} l_{S_N}(t) \right)^2 \\ &= \sum_{t=1}^{T} \left(n_{S_P}(t) + f(\boldsymbol{x}_{S_P}(t); \boldsymbol{\theta}_{S_P}) - \kappa \frac{|S_P|}{|S_N|} n_{S_N}(t) \right)^2. \end{aligned} \quad (16)$$

We then combine the two loss terms to form the final loss function to minimize, and solve for the physical model parameters $\boldsymbol{\theta}_{S_P}$:

$$\min_{\boldsymbol{\theta}_{S_P}} L^{(1)} + \beta L^{(2)}, \quad (17)$$

where β is a weight that balances the two losses. Notably, other than the model parameters $\boldsymbol{\theta}_{S_P}$ to learn, all the values in $L^{(1)}$ and $L^{(2)}$ (cf. (15) and (16)) are either known or have reasonable estimates based on measurements.

Algorithm 1 Similarity-based BTM Solar Disaggregation

```
Input: x(t) and initialization of model parameters θ
Select a set of time slot pairs T for customer S_P
for epochs in 1 to maxiter do
    Initialize Loss = 0
    for (t_i, t'_i) in T do
        Calculate self-similarity loss L^(1) and cross-customer similarity loss L^(2) as in (15) and (16)
        Calculate total loss for this training pair Loss_i = L_i^(1) + βL_i^(2)
        Loss = Loss + Loss_i
    end for
    Update model parameters θ by backpropagation
end for
Output: Estimated physical model parameters θ
```

IV. Numerical Evaluation

A. Experimental Setup

In this section, we present numerical evaluation of the algorithms developed above for BTM solar disaggregation. The evaluation is performed on two real-world data sets from Pecan Street Inc. [17], one collected near Ithaca, New York and another in Austin, Texas. Crucially, these data sets are collected with BTM solar generation measured, so that evaluations by comparing the estimated solar generation with the ground truths can be performed. The solar irradiance data, solar zenith angle data, and temperature data are collected from NSRDB with a 4 km $\times$ 4 km grid resolution and a 30-min granularity. As we will perform BTM solar energy disaggregation at a 15-min granularity, the NSRDB data is converted into 15-min interval data using linear interpolation.

Table I: Performance of BTM Solar Generation Estimation

	RMSE	MASE	CV
Aggregate, Ithaca, NY (Supervised)	3.39	1.32	0.29
Aggregate, Ithaca, NY (Unsupervised)	**3.63**	**1.43**	**0.31**
Aggregate, Austin, TX (Supervised)	37.19	1.41	0.27
Aggregate, Austin, TX (Unsupervised)	**37.94**	**1.49**	**0.28**

In what follows, three metrics will be used to measure the performance of the method: a) Root-Mean-Square Error (RMSE), b) Mean Absolute Scaled Error (MASE), and c) Coefficient of Variation (CV):

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \left(g_k(t) - \hat{g}_{k(t)} \right)^2}, \quad (18)$$

$$MASE = \frac{1}{N} \sum_{k=1}^{N} \frac{T-1}{T} \frac{\sum_{t=1}^{T} \left(g_k(t) - \hat{g}_{k(t)} \right|}{\sum_{t=2}^{T} |g_k(t) - g_k(t-1)|}, \quad (19)$$

$$CV = \frac{1}{N} \sum_{k=1}^{N} \left(\sqrt{\frac{\sum_{t=1}^{T} \left(g_k(t) - \hat{g}_k(t) \right)^2}{T}} \Big/ \frac{1}{T} \sum_{k=1}^{T} g_k(t) \right). \quad (20)$$

B. BTM Solar Generation Estimation

The first data set on which we evaluate the performance contains smart meter data collected from 25 customers near Ithaca, NY, 14 of whom have BTM solar PVs spanning a 6 months period from 5/1/2019 to 10/31/2019. Importantly, not only do we evaluate the performance of our unsupervised learning method, we also evaluate the performance of *supervised learning* by utilizing the ground truths of solar generation in the data set to train the physical model parameters. Given the physical model, the supervised learning approach provides the best possible performance and hence a *performance bound* for all unsupervised learning methods. The performance gap between the unsupervised and the supervised learning methods can then be used to evaluate how well the unsupervised learning method works.

For this data set, the achieved performance is summarized in Table I. We observe that our algorithm (unsupervised) achieves performance that is very close to the ideal performance bound offered by the supervised approach. To visualize the performance comparison, we plot in Figure 2 the ground truths, estimates from supervised learning, and our unsupervised estimates of the solar generation in four typical weeks from May, July, August, and October respectively in 2019. We can indeed observe that our unsupervised learning algorithm estimates the BTM solar generation very well, and the estimates between the ideal supervised and our unsupervised approaches are nearly identical. The fact that even the supervised approach still does not recover the ground truth perfectly can be explained by factors including a) our physical model of solar generation, while accurate, is still imperfect,

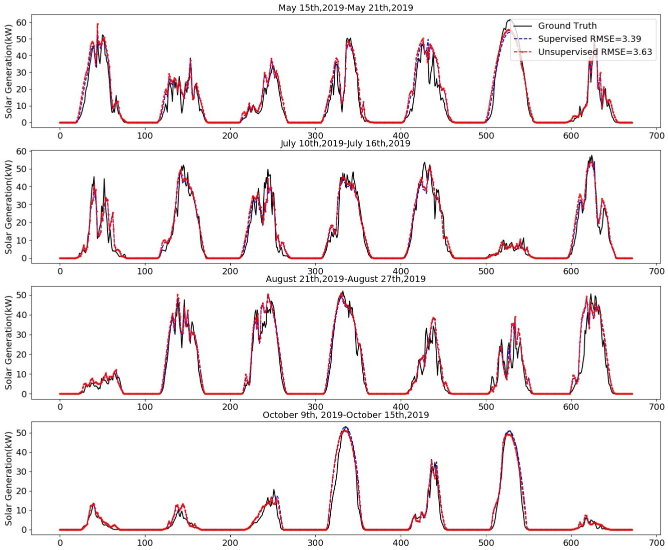

Figure 2: Performance of BTM solar generation estimation for an aggregation of customers, Ithaca, NY.

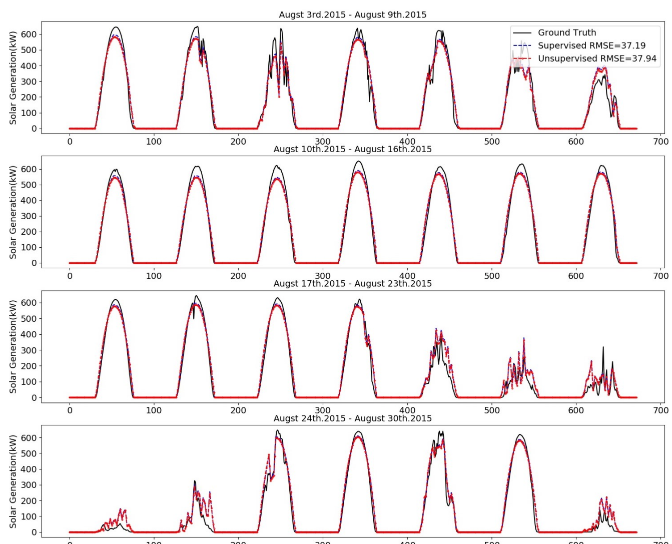

Figure 3: Performance of BTM solar generation estimation for an aggregation of customers, Austin, TX.

and b) the solar irradiance data from NSRDB are collected at locations not exactly where the customers are.

The next data set on which we evaluate our performance contains smart meter data from 322 customers in Austin, TX, spanning a 4 weeks period from 8/3/2015 to 8/30/2015. 182 of which have BTM solar PVs, and we evaluate the performance for the aggregation of these customers with PVs. The performance is also summarized in Table I. The traces of the ground truths, estimates from supervised learning, and our estimates of the solar generation in these four weeks are plotted in Figure 3. We observe a similarly high performance of our algorithm which approaches the ideal supervised learning performance very closely.

V. Conclusion

We developed a novel unsupervised-learning-based method that estimates BTM solar generation based on net load data measured by smart meters. The method successfully exploits the self-similarity and cross-customer similarity of customer loads to achieve high BTM solar generation estimation accuracy. Evaluation of BTM solar generation estimation is performed for aggregations of customers based on two real-world data sets collected in New York and Texas. It demonstrates the very high performance of the proposed unsupervised method which closely approaches the ideal performance achieved by supervised learning.

References

[1] Erdener, B. C., Feng, C., Doubleday, K., Florita, A., and Hodge, B.-M., "A review of behind-the-meter solar forecasting," *Renewable and Sustainable Energy Reviews*, vol. 160, 2022.

[2] Chen, X. and Ardakanian, O., "Solar disaggregation: State of the art and open challenges," in *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, New York, NY, USA, 2020, p. 6–10.

[3] Kara, E. C., Roberts, C. M., Tabone, M., Alvarez, L., Callaway, D. S., and Stewart, E. M., "Disaggregating solar generation from feeder-level measurements," *Sustainable Energy, Grids and Networks*, vol. 13, pp. 112–121, 2018.

[4] Tabone, M., Kiliccote, S., and Kara, E. C., "Disaggregating solar generation behind individual meters in real time," in *Proceedings of the 5th Conference on Systems for Built Environments*. New York, NY, USA: Association for Computing Machinery, 2018, p. 43–52.

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

[6] Cheung, C. M., Zhong, W., Xiong, C., Srivastava, A., Kannan, R., and Prasanna, V. K., "Behind-the-meter solar generation disaggregation using consumer mixture models," in *2018 IEEE SmartGridComm*, 2018, pp. 1–6.

[7] Bu, F., Dehghanpour, K., Yuan, Y., Wang, Z., and Guo, Y., "Disaggregating customer-level behind-the-meter pv generation using smart meter data and solar exemplars," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5417–5427, 2021.

[8] Bu, F., Dehghanpour, K., Yuan, Y., Wang, Z., and Zhang, Y., "A data-driven game-theoretic approach for behind-the-meter pv generation disaggregation," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3133–3144, 2020.

[9] Wang, Y., Zhang, N., Chen, Q., Kirschen, D. S., Li, P., and Xia, Q., "Data-driven probabilistic net load forecasting with high penetration of behind-the-meter pv," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3255—-3264, 2018.

[10] Chen, D. and Irwin, D., "Sundance: Black-box behind-the-meter solar disaggregation," in *Proceedings of the Eighth International Conference on Future Energy Systems*, ser. e-Energy '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 45–55.

[11] Kabir, F., Yu, N., Yao, W., Yang, R., and Zhang, Y., "Joint estimation of behind-the-meter solar generation in a community," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 682–694, 2021.

[12] Li, W., Yi, M., Wang, M., Wang, Y., Shi, D., and Wang, Z., "Real-time energy disaggregation at substations with behind-the-meter solar generation," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2023–2034, 2020.

[13] Lin, S. and Zhu, H., "Enhancing the spatio-temporal observability of grid-edge resources in distribution grids," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5434–5443, 2021.

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

[15] Dobos, A. P., "PVWatts version 5 manual," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2014.

[16] Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., and Shelby, J., "The national solar radiation data base (nsrdb)," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 51–60, 2018.

[17] Holcomb, C., "Pecan Street Inc.: A Test-bed for NILM," in *International Workshop on Non-Intrusive Load Monitoring*, 2012.