Maximal Information Coefficient Based Residential Photovoltaic Power Generation Disaggregation

Zhaohua Chen
Dept. of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
2111904198@mail2.gdut.edu.cn

Keda Pan
Dept. of Control Engineering
Guangdong University of Technology
Guangzhou, China
1111904017@mail2.gdut.edu.cn

Chun Sing Lai
Dept. of Electronic and Electrical Engineering
Brunel University London
London, UK
chunsing.lai@brunel.ac.uk

Zhenhao Li
Qingyuan Power Supply Bureau
Guangdong Power Grid Co., Ltd
Qingyuan, China
leechanhou@163.com

Zhuoli Zhao
Dept. of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
zhuoli.zhao@gdut.edu.cn

Loi Lei Lai
Dept. of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
l.l.lai@gdut.edu.cn

***Abstract*—Due to policy support, low cost and easy applicability, distribution photovoltaic systems (DPVSs) are increasingly popular among residential community. However, small-scale DPVSs of less than 10 kWp are always installed behind the meter (BTM), which results in the invisible of the photovoltaic (PV) power generation. Only access of composite power data can result in non-optimal distribution network control and optimization, leading to a series of energy management problems. In order to solve the aforementioned problems, this paper proposes a BTM composite power disaggregation method focusing on small-scale DPVSs, with only composite power data of residential users in a community, without relying on weather data and models assumption. Considering that community users' DPVSs usually exhibit approximate output characteristics, neighboring composite power is used to extract PV power generation information as mutual proxies. After obtaining approximate PV proxy data by subtracting composite power of inter-users, a grid search algorithm guided by Maximal Information Coefficient (MIC) is performed to obtain final PV power generation disaggregation results. The proposed method is evaluated using data gathered from residential customers located in Ithaca, New York and Austin, Texas in real-life scenarios. Testing results show that our proposed method achieve considerable disaggregation accuracy in the absence of solar radiation and temperature data as compared to other state-of-art methods.**

Keywords—Behind-the-meter, photovoltaic power generation disaggregation, correlation analysis, residential.

I. INTRODUCTION

Tightening of non-renewable resources and environmental degradation issues is the driving force for the transformation of traditional fossil fuel energy-dependent power system. Therefore, increasingly utilizing of renewable resources are emerging in different countries. Countries like Iceland, Norway, Costa Rica, Brazil and Canada have achieved 100%, 97%, 93%, 76% and 62% renewable grid, respectively [1]. Photovoltaic (PV), due to its massive and accessible energy that receive from the sun and price reduction on the installation of PV system, has a huge growth prospect with the low levelized cost of energy [2],[3]. Considering not occupying external areas and to reduce transmission losses, a large number of PV systems are installed rooftop [4]. According to the Solar Power Europe 2019, rooftop PV is estimated with installation of 44 GW with low scenario and 76.5 GW with high scenario [5]. However, most of the rooftop PV systems are installed behind the meter (BTM) without energy of demand and PV power generation separated metering, which will result in the troublesome of energy management, energy storage sizing and protection system setting. Moreover, the issue of power privacy is of concern, customers not sharing of details of residential energy information can also create problems of management of utilities operators. Therefore, designing methods of estimating BTM PV power generation is very meaningful.

The present research can be divided into model assumption based and data-driven approach according to the way of modelling.

The common feature of the model assumption based methods represented by [6], [7], and [8] is that the disaggregated PV is calculated from the assumed PV system geometry and generation output characteristics, which will lead to excessive transfer consistency between solar radiation and outgrowth, and the estimation error would persist once the assumed model is different from, the reality. The model assumption based method [9] disaggregates the BTM PV by building PV power generation model under clear sky condition and modifies it by the universal weather-solar effect to circumvent the errors brought about by specific kinds of PV output characteristics assumptions, but assumption about the geometric architecture is still unavoidable and the requirement for residential idleness limits the application scenarios of the model.

The data-driven methods can avoid the pitfalls of models assumption altogether, but has higher exogenous data requirements in comparison. Most data-driven methods disaggregate composite power relying on proxy settings. In [10], by setting a PV proxy of unit capacity, estimated PV power generation was calculated by multiplying the unit PV proxy to the estimated installation capacity which inferred by a designed support vector machine (SVR) model. Similarly, but starting from energy demand, [11] formulated the target customer demand as a mixed behavioral composition of neighbors without PV installation (similar to a proxy of "consumer", but no additional installation was required) and disaggregated the composite power using multivariate linear programming in combination with solar radiation data. In [12], PV of individuals were disaggregated with the help of separated measured demand and PV power generation of aggregated customers on the feeder side using linear model with scale proxy data as an additional input. The study carried out in [13] with the disaggregation method does not employ any proxies while using a data-driven approach, but rather by minimizing the estimated error of composite power to search relevant parameters to disaggregate BTM PV power generation.

In summary, whether the method is model assumption based or data-driven, the disaggregation results of the target BTM composite power always require at least one exogenous data to derive the estimated demand or PV power generation. For the model assumption based method, the exogenous variable is mostly meteorological data, while for the data-driven method, it is mostly the demand or PV power generation data of proxies. A summary of the exogenous variables required for different studies is given in Table I below.

Bringing in exogenous variables, for example, the transposition error of meteorological data due to geographically distance between the collection device and the target users, and will increase the implementation cycle of the disaggregation methods such as the time required to collect data from newly installation proxy devices, which is detrimental to the practical implementation of composite power disaggregation. Therefore, we propose a data-driven method to disaggregate BTM PV of community individuals using only composite power data of residential users but not requiring meteorological data, PV power generation and demand proxies data for better application to real-world scenarios.

The rest of the paper is organized as follows: Section II presents the disaggregation methodology. Section III presents case study on two datasets to verify the effectiveness and superiority of the proposed method. Finally, conclusions and future work are presented in Section IV.

II. Proposed Methodology

A. Framework

PV power generation data are invisible in the BTM system. It is difficult to obtain PV power generation data only through the composite power data of a single residential user. In general, composite power data of residential users in a community have similar shape of PV power generation curves but with different capacity.

By subtracting composite power data between users with similar demand, the difference of PV power generation can be roughly obtained. Due to the strong linear relationship between PV power generation and capacity, the curve of the difference of PV power generation has a similar shape to the curves of PV power generation. The most suitable PV power generation curve can be obtained by Maximal Information Coefficient (MIC) between integrated composite power of all users and the PV power generation difference obtained before.

TABLE I. A Summary of the Exogenous Variables required for Different Studies

Study	Model type	Exogenous variable
[5], [6], [7], and [8]	Model assumption based	Solar radiation and temperature data
[9]	Data-driven	PV power generation of the PV proxy
[10]	Data-driven	Demand of the neighbor
[11]	Data-driven	PV power generation and demand of feeder level
[12]	Data-driven	Solar radiation and temperature data

Denoising technology can be used to obtain smoother PV data from the most suitable PV power generation curve by reducing the impact of residential demand. The most similar PV power generation curve after denoising can be seen as PV proxy of the whole community. A grid search algorithm is performed to obtain the residential disaggregated PV power generation data guided by MIC metric finally.

B. Estimation of PV proxy by composite power data

The weather conditions in the same community are similar, such as solar radiation and ambient temperature. Hence, the generation output characteristics of solar PV panels of residential users in a community are similar. Therefore, each user can be seen as an implied PV proxy for others. Subtracting the composite power of users with similar demand but different PV capacities can offset the demand between users, and obtain a PV power generation difference as PV proxy for the community

Composite power C (kW) is equal to demand D (kW) minus PV power generation P (kW) [14]. Suppose there are n users in the community, for user i , it could be present as:

$$C_{i,t} = D_{i,t} - P_{i,t}, i \in n \tag{1}$$

Since the basic demand of each user is unknown and different, demand of each user is required to be adjusted to the same magnitude. Thus, feature scaling of D is needed. Considering that C is equal to D at night (P equals to zero at night), feature scaling of D can be performed by dividing by the mean of composite power in nighttime as below:

$$C'_{i,t} = \frac{C_{i,t}}{\frac{1}{m}\sum_{t_1} C_{i,t_1}}, i \in n, t_1 \in night \tag{2}$$

Denominator of (2) is the mean of composite power in nighttime, m represents the number of total time points at night.

In order to reduce the influence of demand of each user, a subtract operation of C' between users i and j after feature scaling of same numerical magnitude is performed as below:

$$C'_{i,t} - C'_{j,t} = (D'_{i,t} - D'_{j,t}) - (P'_{i,t} - P'_{j,t}); i \in n, j \in n, i < j \tag{3}$$

For simplicity, (3) is represented by (4).

$$\Delta C'_{ij,t} = \Delta D'_{ij,t} - \Delta P'_{ij,t} \tag{4}$$

The feature scaling makes $\Delta D'_{ij,t}$ in (4) a small value that can be seen as noise. Beside, due to the geographically close and strong linear relationship between PV power generation and capacity, the PV power generation of users i and j with difference capacity has similar output curves:

$$\alpha_j P'_{i,t} = \alpha_i P'_{j,t} \tag{5}$$

Where α_i and α_j in (5) represent the capacity of users i and j, respectively.

Obviously, $\Delta P'_{ij,t}$ also have strong linear relationship with $P'_{i,t}$ and $P'_{j,t}$:

$$\begin{cases} P'_{i,t} = \beta_{ij} \Delta P'_{ij,t} \\ \beta_{ij} = \alpha_i / (\alpha_i - \alpha_j) \end{cases}, \alpha_i \neq \alpha_j \tag{6}$$

Simultaneously, each user in a community has similar shape of PV power generation curves while with difference capacity. Thus, $\Delta P'_{ij,t}$ can be a PV proxy for each user of the whole community. However, the demand diversity of different users exists, it is necessary to filter $\Delta C'_{ij,t}$ in (4). Specifically, the aim of this step is to find a pair of users with $\Delta C'_{ij,t}$ mainly composed by $\Delta P'_{ij,t}$.

For n users in a community, $(n^2 - n)/2$ combinations of $\Delta C'_{ij,t}$ can be obtained from (4). The most suitable users i and j in which $\Delta C'_{ij,t}$ mainly is composed by $\Delta P'_{ij,t}$ can be selected through correlation analysis. MIC is used to correlation analysis between $\Delta C'_{ij,t}$ and integrated composite power in (7). MIC can quantify both linear and non-linear correlation relationships of pairs of variables. More details of MIC are given in [15]. If the value of correlation relationship between the integrated composite power and $\Delta C'_{ij,t}$ achieves the maximum, the most suitable users i and j are determined, and the corresponding $\Delta C'_{ij,t}$ is considered to be the highest percentage of $\Delta P'_{ij,t}$ among all combinations of $\Delta C'_{ij,t}$.

$$R_1 = \arg\max_{i,j} MIC(C'_{n,t}, \Delta C'_{ij,t}) \tag{7}$$

where, $C'_{n,t}$ is the integrated composite power of n users. Thus, the user matching process is converted into an optimization problem of (7). This is a non-analytical optimization problem, which cannot be solved by traditional mathematical analysis methods. A grid search algorithm guided by (7) is employed to find the most relevant i and j. For simplify, after finding the most relevant i and j, the corresponding $\Delta C'_{ij,t}$ is represent as ΔC_t.

To further reduce the impact of $\Delta D'_{ij,t}$, Variational Mode Decomposition (VMD) is applied to denoise ΔC_t. The constrained variational problem is represented as:

$$\begin{aligned} &\min_{v_k, w_k} \left\{ \sum_k^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ &s.t. \sum_k^K v_k = f(t). \end{aligned} \tag{8}$$

Where $f(t)$ is the signal of ΔC_t with frequency w and mode v, k is the index of mode, δ is the Dirac distribution, $*$ denoted the convolution operation, K is the total number of modes and the decomposition level. More details are given in [16].

The composite power data for residential users is a combination of relatively low-frequency $\Delta P'_{ij,t}$ and high-frequency $\Delta D'_{ij,t}$, thus, it is easy to decompose ΔC_t by VMD for the low-frequency part as $\Delta P'_{ij,t}$. The filtered $\Delta P'_{ij,t}$ can be seen as a finer PV proxy for each user of the whole community.

C. Estimation of Residential PV generation by MIC

To simplify the explanation, the PV proxy $\Delta P'_{ij,t}$ obtained after VMD represents as ΔP_t. PV power generation of each residential users is equal to ΔP_t multiplied by a coefficient β, for user i:

$$P'_{i,t} = \beta_i \Delta P_t; i \in n \tag{9}$$

Where β_i in (9) is solved by a grid search algorithm with metric of (10):

$$R_2 = \arg\min_{\beta_i} MIC(C'_{i,t} + \beta_i \Delta P_t, \Delta P_t) \tag{10}$$

By manually setup the lower bound $\beta_{i,\min}$, upper bound $\beta_{i,\max}$ and discretization step $\Delta\beta_i$, an exhaustive search is performed to the manually specified subset of β_i. If the correlation between $C'_{i,t} + \beta_i \Delta P_t$ and ΔP_t achieves its minimum value, the first component would be seen as $D'_{i,t}$ which has very low correlation with PV power generation, and β_i is determined from (10). Then the residential PV power generation $P'_{i,t}$ can be obtained according to (9). Finally, the residential PV power generation $\hat{P}_{i,t}$ for user i can be obtained after inverse transformation of $P'_{i,t}$ by multiplying the denominator in (2).

The algorithm to disaggregate residential PV power generation from composite power is summarized in Algorithm 1 below.

Algorithm 1: Algorithm for residential PV power generation disaggregation from composite power

Input: Composite power data C_t of n users in a community.

Output: Estimated PV power generation $\hat{P}_{k,t}$ of user k.

1: Feature scaling of $C_{i,t}$ for all residential users using (2), obtain $C'_{i,t}$.
2: Obtain the integrated composite power $C'_{n,t}$.
3: **for** $i = 1$ to n **do**
4: **for** $j = i + 1$ to n **do**
5: Obtain $\Delta C'_{ij,t}$ from (3) and (4).
6: Calculate $MIC(C'_{n,t}, \Delta C'_{ij,t})$.
7: **end for**

end for

Obtain the most relevant ΔC_t according to (7).

Denoise ΔC_t by VMD, obtain ΔP_t as PV proxy.

for $\beta_k = \beta_{k,\min} : \Delta\beta_k : \beta_{k,\max}$ **do**

  Calculate $MIC(C'_{k,t} + \beta_k \Delta P_t, \Delta P_t)$.

end for

Obtain β_k according to (9).

Calculate $P'_{k,t} = \beta_k \Delta P_t$.

Inverse transformation of $P'_{k,t}$ in (2), obtain $\hat{P}_{k,t}$.

Return $\hat{P}_{k,t}$.

III. CASE STUDY

A. Dataset and evaluation metrics

To evaluate the proposed residential BTM PV power generation disaggregation method for real scenarios, open-source datasets which located in Austin, Texas and Ithaca, New York are applied. After data pre-processing of complementing, dataset located in Austin, Texas have 24 residential users with metering time from 01/01/2018 to 30/12/2018 with 15-minute interval, while Ithaca, New York have 18 with metering time from 01/05/2019 to 31/10/2019 with 15-minute interval. Both datasets provide real-world composite power data and PV power generation data, so this experiment is completely based on actual scenarios. Daytime of the experiment is set from 6:30 to 17:30.

The root mean square error (RMSE) and the coefficient of variation (CV) are used to evaluate the disaggregation accuracy. Considering the large number of zero points in PV data, which is unavailable for commonly used MAPE, CV is used instead. RMSE and CV are calculated as follows: $P_{i,d,t}$ in CV is another expression of $P_{i,t}$, where d and t represent the day and the moment of the day, respectively.

$$RMSE=\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{P}_{i,t}-P_{i,t})^2} \tag{11}$$

$$CV=\frac{1}{N}\sum_{d=1}^{N}\frac{\sqrt{\sum_{t=1}^{T}(\hat{P}_{i,d,t}-P_{i,d,t})^2}}{\sum_{t=1}^{T}P_{i,d,t}} \tag{12}$$

B. Correlation analysis of PV generation inter-users

The key point of proposed method is the assumption of (5) that all residential users in a community have similar shape of PV power generation curves. To illustrate the rationality of this assumption, Pearson Correlation Coefficient (PPC) [17] is used to evaluate the linear correlation of PV power generation between each residential users of the datasets. PPC is a measure of linear correlation between two sets of data. Figs. 1 and 2 show the heatmap of PPC matrix of Austin and Ithaca, respectively. Both the horizontal axis and the vertical axis of the figures are the user ID of the community. The result of PPC have a value between -1 and 1, when a value close to 1 indicates that the linear correlation is stronger.

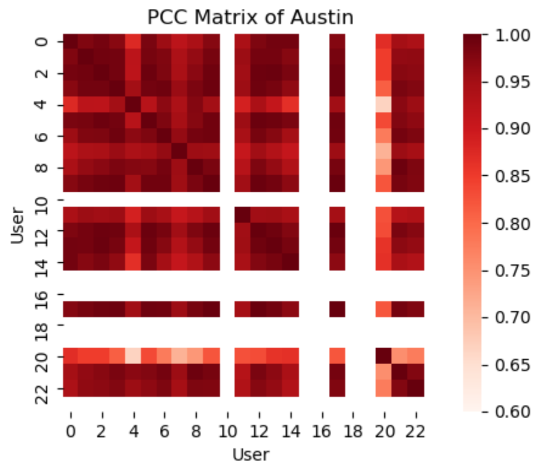

Fig. 2. Pearson correlation coefficient matrix of Austin

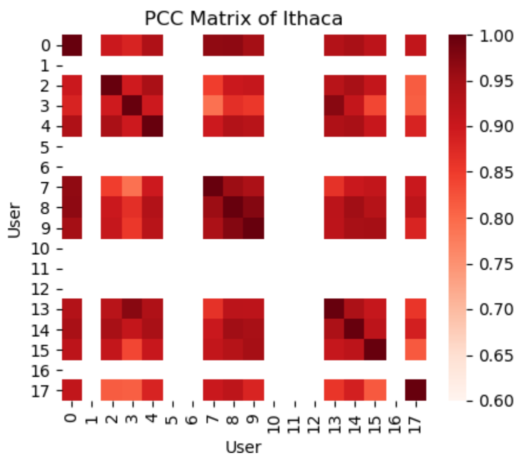

Fig. 1. Pearson correlation coefficient matrix of Ithaca

It can be seen from the figure that PV power generation data between each residential user has a high degree of linear correlation, mostly above 0.9. Users in white part of the figure have no PV equipment installed with considering the capacity α in (5) as zeros. Therefore, the assumption is rationality and feasible.

It is worth noting that, due to the data access permission restrictions, only open source dataset, in which users are not geographically concentrated enough, is used in this experiment. It will affect the disaggregation accuracy to a certain extent. When collected users data are more geographically concentrated, such as under the same feeder, the correlation coefficient will be higher, which is more in line with the ideal situation.

C. Method performance and comparation

In this section, the performance of proposed method is compared with two other state-of-art composite power disaggregation methods proposed in [5] and [12]. Noteworthy, the model assumption based on method [5] and the data-driven based method [12] both require exogenous variable of solar radiation and temperature data as input, while our method only require the same type of residential composite power data in the community as input, which is more accessible. The required solar radiation and temperature data in method [5] and method [12] for corresponding period are obtained from National Solar Radiation Database (NSRDB). Due to the NSRDB only supply 30-minute interval meteorological data, linear interpolation is applied to obtain 15-minute interval meteorological data.

The experiment is to verify composite power disaggregation performance of residential users. User #2 in both datasets equipped with PV system is randomly selected as experimental object. Figs. 3 and 4 show the disaggregation results of user #2 of difference method in a week. Tables II and III present the disaggregation results of user #2 of difference method in RMSE and CV.

It can be seen from Tables II and III that disaggregation results of the three methods are similar, and proposed method obtains the best disaggregation results of RMSE in Austin dataset, but the overall gap is very small. However, the proposed algorithm can obtain results comparable to other two state-of-the-art methods without relying on exogenous solar radiation and temperature data, which is a great advantage in practical implementation. This is a relatively ideal result and illustrates the feasibility of the proposed method. Disaggregation results of New York dataset are a little worse than other two methods. The reason may be that New York dataset has fewer residential users than Austin dataset, thus it is difficult to match users with similar electricity demand, but it is still considerable under this data requirement.

TABLE II. RMSE AND CV OF VARIOUS DISAGGREGATION METHODS OF USER #2 IN AUSTIN, TEXAS

Evaluation method	Method [5]	Method [12]	Proposed Method
RMSE(kW)	0.550	0.543	0.540
CV(%)	8.537	7.870	8.235

TABLE III. RMSE AND CV OF VARIOUS DISAGGREGATION METHODS OF USER #2 IN ITHACA, NEW YORK

Evaluation method	Method [5]	Method [12]	Proposed Method
RMSE(kW)	1.006	0.926	1.070
CV(%)	7.434	6.939	9.556

It can be seen from Figs. 3 and 4 that the data is a small-scale PV power generation system for a single residential user, thus, PV power generation curve has more glitched and fluctuations. Disaggregation results between method [5] and method [12] are relatively similar, and the overall shape of both method are relatively smooth, but they are difficult to learn the real PV fluctuations. The proposed method can better adapt the fluctuation of real PV curve due to the directly calculation of composite power.

D. Ablation experiments of community user scales

The number of residential users in the dataset is an important parameter of proposed method. In theory, the larger the group of community, the easier it is to match users with similar demand. Experiments are conducted to verify the feasibility of the method under different community user scales.

The experiment uses Austin dataset, and experiment object is still user #2, randomly select other users in the dataset to form 3 new datasets of 8, 16, and 24 users. Table IV shows the disaggregation results of user #2 of difference user scales.

It can be seen from Table IV that as the number of users in dataset decreases, disaggregation accuracy continues to decrease. The main reason is that when the number of users in dataset are too small, the probability of matching a suitable user will decrease, and the sub-suitable user will be matched, lead to a decrease in disaggregation accuracy.

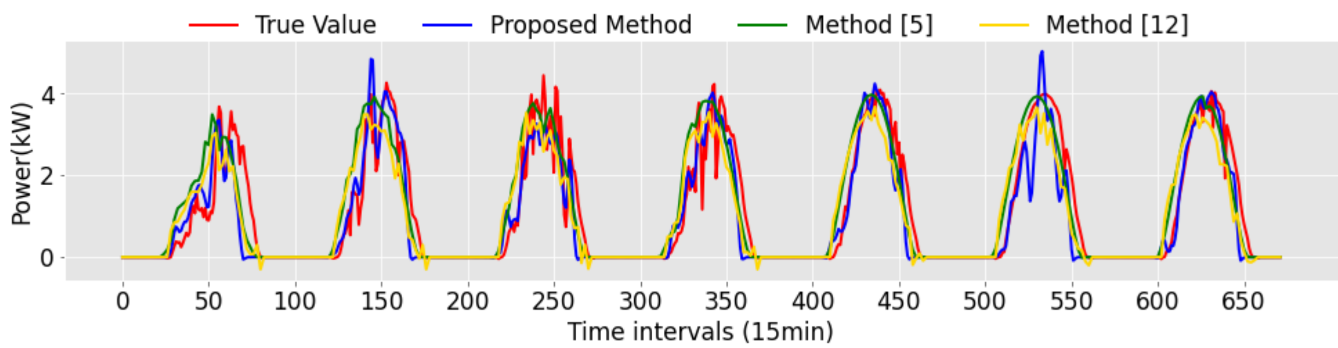

Fig. 3. Disaggregation results of user #2 in Austin, Texas

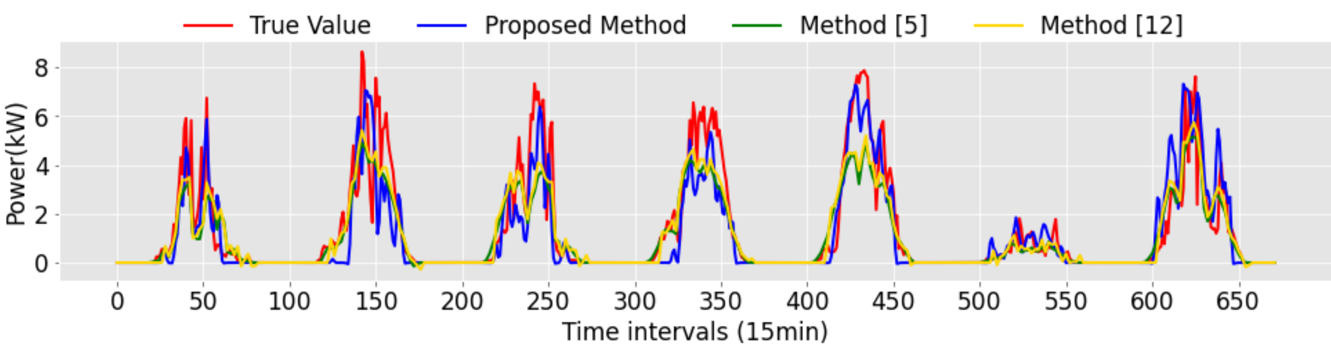

Fig. 4. Disaggregation results of user #2 in Ithaca, New York

TABLE IV. RMSE AND CV OF VARIOUS USER SCALES OF AUSTIN, TEXAS

Evaluation Metric	Number of Users in Dataset		
	8	16	24
RMSE(kW)	0.766	0.658	0.540
CV(%)	12.334	8.936	8.235

Therefore, for the system operator, if the composite power data of a large number of users under the same feeder is available, the proposed method may be able to obtain a higher disaggregation accuracy.

IV. CONCLUTION AND FUTURE WORK

We propose a BTM residential PV power generation disaggregation method only using composite power data of users in a community, without relying on solar radiation and temperature data which are strong exogenous variables related to PV power generation. The ease access to composite power data contributed to the high universality of the proposed method. In the comparison of other two state-of-art methods that rely on solar radiation and temperature, proposed method has achieved considerable disaggregation accuracy.

This method is still relatively primitive with room for improvement. For example, it is difficult to match with the suitable users in composite power difference phase when the number of users is small and it is still difficult to eliminate for some unexplained fluctuations in VMD decomposition phase. Therefore, in future research, various integrated methods will be considered to optimize the composite power difference phase to eliminate the influence of user demand, and research for better filtering in VMD decomposition phase.

REFERENCES

[1] B. Johnson, P. Denholm, B. Kroposki, and B. Hodge, “Achieving a 100% renewable grid,” IEEE Power Energy Mag., no 2, pp. 61–73, Apr, 2017.

[2] O. Gandhi, D. S. Kumar, C. D. Rodríguez-Gallegos, and D. Srinivasan, “Review of power system impacts at high PV penetration Part I: Factors limiting PV penetration,” *Sol. Energy*, vol. 210, pp. 181–201, Feb, 2020.

[3] C. S. Lai *et al.*, “Levelized cost of electricity for photovoltaic/biogas power plant hybrid system with electrical energy storage degradation costs,” *Energy Convers. Manag.*, vol. 153, pp. 34–47, 2017.

[4] B. Uzum, A. Onen, H. M. Hasanien, and S. M. Muyeen, “Rooftop solar pv penetration impacts on distribution network and further growth factors—a comprehensive review,” *Electron.*, vol. 10, no. 1, pp. 1–31, Dec. 2020.

[5] Michael Schmela et al, “Global market outlook for solar Power/2018–2022,” Sept, 2018.

[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, Oct, 2017.

[7] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, “Estimation of behind-the-meter solar generation by integrating physical with statistical models,” 2019 IEEE Int. Conf. Commun. Control. Comput. Technol. Smart Grids, SmartGridComm 2019, 2019.

[8] 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.

[9] D. Chen and D. Irwin, “SunDance: Black-box behind-the-meter solar disaggregation,” e-Energy 2017 - Proc. 8th Int. Conf. Futur. Energy Syst., pp. 45–55, 2017.

[10] 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, pp. 113595, Nov, 2019

[11] C. M. Cheung, S. R. Kuppannagari, R. Kannan, and V. K. Prasanna, “Disaggregation of behind-the-meter solar generation in presence of energy storage resources,” 2020 IEEE Conf. Technol. Sustain. SusTech 2020, no. 1911229, 2020.

[12] M. Tabone, S. Kiliccote, and E. C. Kara, “Disaggregating solar generation behind individual meters in real time,” BuildSys 2018 - Proc. 5th Conf. Syst. Built Environ., pp. 43–52, 2018.

[13] 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. Informatics, vol. 14, no. 9, pp. 3904–3913, Sept. 2018.

[14] K. Pan, C. Xie, C. S. Lai, D. Wang, and L. L. Lai, “Photovoltaic output power estimation and baseline prediction approach for a residential distribution network with behind-the-meter systems,” *Forecasting*, vol. 2, no. 4, pp. 470–487, Oct. 2020.

[15] D. Reshef *et al.*, “Detecting novel associations in large data Sets,” *Sci. Transl. Med.*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.

[16] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Nov. 2013.

[17] J. Lee Rodgers and W. Alan Nice Wander, “Thirteen ways to look at the correlation coefficient,” *Am. Stat.*, vol. 42, no. 1, pp. 59–66, Jun. 1987.