Behind-the-Meter Solar Disaggregation: The Value of Information

S. Mahdi Noori R.A. Masoume Mahmoodi Ahmad Attarha José Iria Paul Scott Dan Gordon

The Australian National University, College of Engineering, Computing and Cybernetics

***Abstract*—In recent decades, the increasing adoption of rooftop solar photovoltaic (PV) systems has posed new challenges for distribution system operators (DSOs). DSO typically only have access to each customer's net generation/consumption. However, since these PV systems are commonly installed behind customers' meters, the DSOs are unable to monitor their generation, leading to a lack of observability that is critical for maintaining grid security. One potential solution to this issue is behind-the-meter solar disaggregation techniques, which can estimate behind-the-meter activity without the need for additional meter installations. These techniques utilise various sources of information, such as weather data and customer reactive power consumption, which can be obtained through third-party providers or direct contracts with customers. However, operators must determine which pieces of information are the most beneficial before investing in them. This paper studies seven different behind-the-meter solar disaggregation techniques that use varying sets of information and compare them based on their scalability, complexity, and required data. We also conduct numerical experiments using two publicly available datasets to evaluate the accuracy of these techniques.**

***Index Terms*—Renewable integration, distribution system, solar-demand disaggregation, non-intrusive load monitoring**

I. Introduction

The urgent need to address climate change has led to the rapid deployment of renewable energy sources, particularly rooftop solar photovoltaic (PV) generation. However, the volatile and distributed nature of PV generation poses challenges for the operation and planning of electrical distribution systems [1], [2]. Although the rollout of smart meters in many countries has increased the observability of the distribution network, these devices only measure the aggregate energy transactions of an entire building, leaving behind-the-meter information hidden. Consequently, distribution system operators and electric utilities struggle to maintain balance between demand and generation, plan reliably for grid reinforcements, and forecast demand/solar, despite having access to measurements from multiple locations on the grid [3].

Disaggregating energy consumption/generation to different sources using an aggregated measured value has been a topic of extensive research since 1985 [4]. This area of study, also known as non-intrusive load monitoring (NILM), involves developing algorithms that can estimate appliance-by-appliance energy consumption information from an aggregated value [5]. Our focus in this paper is on behind-the-meter solar disaggregation (BSD) algorithms, a special case of NILM. When no extra information is available, and only the aggregated measurements of solar and demand are given, the BSD problem is equivalent to an underdetermined system of linear equations with infinitely many solutions. Consequently, the BSD literature explores ways to introduce additional information to the problem of finding a unique solution that approximates the "true value" of solar and demand [6]. Examples of these approaches are those using: weather data [7], geographically close solar exemplars [3] and [8], customers load consumption without PV that are electrically close to the BSD target customers [9], and reactive power [10].

The existing literature assumed access to only certain information and developed their algorithm based on this assumption. However, electrical utilities can acquire most of the above information at a cost. For example, companies like Solcast [11] provide historical-geographical weather and solar irradiance at different time resolutions. Utilities may also gain access to information on whether customers have installed PV systems. Therefore, before investing in buying each type of information, operators need to understand how much each piece of information can assist them in more precisely disaggregating solar and demand. Also, it is important to clearly understand the impact of accessing different types of information on customers' privacy. For example, the approach in [9] requires access to many customers' electrical demand information, which may require additional investment in keeping customers' information safe. In contrast, the approach in [7] can work with the electrical demand of a single customer, thus avoiding such extra costs.

This paper studies seven techniques to disaggregate behind-the-meter solar generation and demand from their aggregated values, each using a different set of information. Inspired by the available literature, we develop these techniques using various optimisation models and machine learning algorithms. Our analysis compares the scalability, complexity, and information requirement of each approach. To analyse the accuracy of the BSD techniques, we use two publicly available large datasets, i.e., Ausgrid [12], and NextGen [13] datasets. Our results indicate that incorporating reactive power as an additional source of information yields superior accuracy and computational efficiency compared to other approaches.

The rest of this paper is organised as follows. Section II introduces the mathematical notation and the BSD problem formulation. Section III details the BSD techniques. Section IV reports the numerical experiments, and Section V concludes this paper.

979-8-3503-1809-8/23/$31.00 ©2023 IEEE

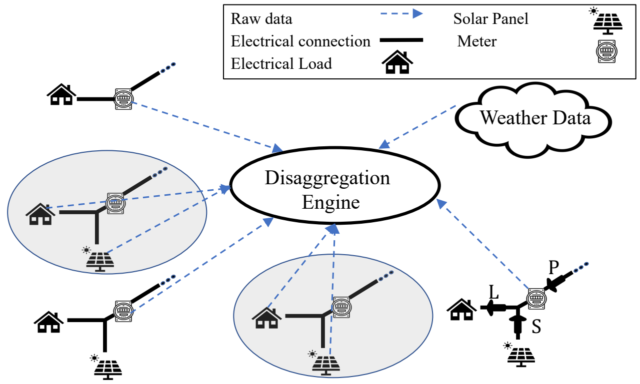

Fig. 1. Disaggregation Engine structure. The techniques in the engine use three types of information to behind-the-meter disaggregate solar and demand from the meter measured power that are: *i*) meter measured power information; *ii*) solar and demand information from nodes in close proximity (shown with green shades); *iii*) weather data.

II. Problem Formulation

This section introduces various disaggregation algorithms used in this article. These algorithms are classified based on how much information they need to perform disaggregation. We first begin by introducing the modelling notation.

A. Modelling Notation

Let set $\mathcal{N}$ with cardinality N denote all nodes with electrical injection/consumption to the grid. Also, let set $\mathcal{T}$ with cardinality T denote all the time steps in a dataset on which we aim to perform the disaggregation. Let $l_{i,t} \in \mathbb{R}_+$, $s_{i,t} \in \mathbb{R}_+$ and $p_{i,t} \in \mathbb{R}$ denote the behind-the-meter electricity consumption, behind-the-meter solar generation and net demand from the grid at node i and time step. Also, let vectors $L \in \mathbb{R}_+^{(N\times T)}$, $S \in \mathbb{R}_+^{(N\times T)}$ and $P \in \mathbb{R}^{(N\times T)}$ respectively collect $l_{i,t}$, $s_{i,t}$ and $p_{i,t}$ for all the nodes and time steps.

B. Solar-Demand Disaggregation Problem

The problem of solar-demand disaggregation can be mathematically defined as finding L and S given $\boldsymbol{P}$ (we use bold symbols to distinguish between the variables and parameters) so that:

$$L - S = \boldsymbol{P}. \tag{1}$$

This problem is a set of linear equations with $2 \times N \times T$ variables and $N \times T$ linear equality constraint. The problem is an underdetermined linear problem with infinitely many solutions, as the constraints (one for each node and time) are linearly independent.

The disaggregation algorithms, by introducing extra information, such as weather data or time-of-day impact on the solar and demand values, explore dependence between the variables. These algorithms also commonly convert problem (1) to an optimisation problem to take the most out of the introduced dependencies between the variables.

III. Disaggregation Algorithms

A. Minimum Solar Generation

This technique considers the minimum solar generation at every node and time step, and it does not require any additional information. It is formulated as a linear optimisation problem as follows:

$$\min_{S\geq 0, L\geq 0} \quad \mathbb{1}^\intercal S \tag{2a}$$
$$s.t.$$
$$L - S = \boldsymbol{P}, \tag{2b}$$

where $\mathbb{1}$ is a vector of all one, and operator $()^\intercal$ denotes vector transpose. Since all the constraints in (2b) are independent and assuming positive solar generation and demand, the optimisation problem is equivalent to assigning positive net demand to load and negative net demand to solar generation. Thus, the solution to (2b) can be obtained simply as follows:

$$L = \max(\boldsymbol{P}, 0), \quad S = -\min(\boldsymbol{P}, 0). \tag{3}$$

B. Same Irradiance

This technique assumes that all the PV panels in an area get the same solar irradiance as they are geographically close. Thus, it approximate the solar generation $s_{i,t}$ with $\beta_t \times \overline{\boldsymbol{s}}_i$, where β_t denotes the solar irradiance at time t , and $\overline{\boldsymbol{s}}_i$ denotes the PV system size at node i. Notice that the PV system size is considered to be known, as most system operators mandate the customers to report their PV system installation size. In cases where the operator does not have access to this information, the PV system size can be approximated as the minimum value of the net demand at a node in the dataset.

This technique formulates the solar-demand disaggregation problem as follows:

$$\min_{l\geq 0, \beta\geq 0} \quad \sum_{t=0}^{T} \beta_t \tag{4a}$$
$$s.t.$$
$$l_{i,t} - \beta_t \times \overline{\boldsymbol{s}}_i = \boldsymbol{p}_{i,t} \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}. \tag{4b}$$

After solving the optimisation (4) and determining β_t, it then calculate the solar generation as:

$$s_{i,t} = \boldsymbol{\beta}_t \times \overline{\boldsymbol{s}}_i \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}. \tag{5}$$

Finally, knowing the solar generation, the demand values can then be simply obtained as:

$$L = \boldsymbol{S} + \boldsymbol{P}. \tag{6}$$

C. Same Irradiance and Houses Without PV Installation

This approach considers the measurement from the houses in the network that have no PV installation. It then assumes that the average net consumption of the houses with and without PV installation should be close to each other at any time step. Also, it assumes that the solar irradiance of the houses that are in close proximity should be close to each other. Considering these assumptions, we then develop the following optimisation model to disaggregate behind-the-meter solar generation and demand.

Let non-empty sets $\mathcal{N}^g \subset \mathcal{N}$ and $\mathcal{N}^l \subset \mathcal{N}$ with cardinality N^g and N^l denote the house with and without PV generation, respectively. Also, let vector B_t collect $\beta_{i,t}$ for all the nodes with PV generation at time step t. We then write:

$$\forall t \in \mathcal{T}:$$

$$\min_{l\geq 0, \beta\geq 0} \quad |\sum_{i\in\mathcal{N}^l} \frac{\boldsymbol{l}_{i,t}}{\boldsymbol{N}^l} - \sum_{i\in\mathcal{N}^g} \frac{l_{i,t}}{\boldsymbol{N}^g}| + \frac{||B_t||_2^2}{\boldsymbol{N}^g} \tag{7a}$$

$$s.t.$$

$$l_{i,t} - \beta_{i,t} \times \overline{\boldsymbol{s}}_i = \boldsymbol{p}_{i,t} \quad \forall i \in \mathcal{N}. \tag{7b}$$

After solving the optimisation for every time-step, behind-the-meter solar generation for each node and time-step is obtained by multiplying the solar irradiance ($\boldsymbol{\beta}_{i,t}$) and the PV system size ($\overline{\boldsymbol{s}}_i$).

D. Constant Power Factor Demand

This approach requires the net reactive power consumption at each node and time step. Since it is measured at the connection point, it does not need any behind-the-meter information. Also, since at the moment most residential inverters work at the unity power factor[1], the measured reactive power at the connection point is only from household consumption. Thus, assuming that the power factor in each node does not change much during a day and the solar generation is equal to zero in the early hours of the day, we obtain the following constant for each day. Let set $\mathcal{D}$ denotes the days in the dataset. We then write:

$$K_{i,d} = \frac{\sum_{tt=0}^{\boldsymbol{M}} \dfrac{\boldsymbol{q}_{i,tt}}{\boldsymbol{p}_{i,tt}}}{\boldsymbol{M}} \quad \forall d \in \mathcal{D}, \tag{8}$$

where $\boldsymbol{M}$ is the number of time steps with zero solar generation in the early hours of the day. For example, if the dataset has a granularity of 60 or 15 minutes, $\boldsymbol{M}$ can be considered to be 4 or 16, respectively. Let set $\mathcal{T}_d$ denote the remaining time step of the day d. Then, the real power behind-the-meter electricity consumption can be obtained as:

$$l_{i,t} = K_{i,d} \times \boldsymbol{q}_{i,t} \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}_d, \forall d \in \mathcal{D}. \tag{9}$$

Next, knowing the demand values, the solar generation can then be simply obtained as:

$$S = \boldsymbol{L} - \boldsymbol{P}. \tag{10}$$

E. Measurements from Neighbouring Sites

This approach assumes to have access to solar generation data from a few sites located in the same area as the target nodes. It builds a linear solar mixture model to estimate each node's solar generation. The coefficient of the linear mixture model is then optimised to minimise the $l-2$ norm between the model output and the negative net measured at each node. It is formalised as follows.

Let set $\mathcal{G}$ with cardinality G denote the sites with known solar generation. Also, let variable $w_{i,k}$ denote the coefficient of the k-th site with known solar generation for node i. The linear mixture model then states the relation between the generation at node i and the generation of the known sites as:

$$s_{i,t} = \sum_{g\in\mathcal{G}} (w_{i,g} \frac{\boldsymbol{p}_{g,t}}{\overline{\boldsymbol{s}}_g}) \overline{\boldsymbol{s}}_i \tag{11}$$

Defining vectors

$$W_i = [w_{i,0}, ..., w_{i,G}] \in \mathbb{R}^G$$

$$\boldsymbol{P^G} = \Big[\frac{[\boldsymbol{p}_{g,0}, ..., \boldsymbol{p}_{g,T}]^\intercal}{\overline{\boldsymbol{s}}_g} \;\; \forall g \in \mathcal{G} \Big] \in \mathbb{R}^{G\times T}$$

$$\boldsymbol{B}_i = \frac{[-min(\boldsymbol{p}_{i,0}, 0), ..., -min(\boldsymbol{p}_{i,T}, 0)]}{\overline{\boldsymbol{s}}_i} \in \mathbb{R}^T$$

We then formulate the following optimisation model to disaggregate solar and demand at node i for all $t \in \mathcal{T}$:

$$\min_{0\leq W_i \leq 1} \quad ||W_i \times \boldsymbol{P^G} - \boldsymbol{B}_i||_2^2 \tag{12a}$$

$$s.t.$$

$$\mathbb{1}^\intercal \, W_i = 1. \tag{12b}$$

After solving the optimisation (12) and finding the optimum coefficient matrix W, we first use (11) to retrieve behind-the-meter solar generation. We then obtained behind-the-meter demand using (6).

F. Weather Data

This technique approximates a customer's home load assuming access to weather data and other known data, such as the hour of the day. It trains a non-linear function that maps these data to the customer's load. Here, we use the approach suggested in [14] that considers the ambient temperature, exponentially weighted moving average of temperature over the last 24 hours, time of the day, and a binary variable that indicates if it is a weekday or weekend as explanatory variables. We use a similar iterative algorithm as suggested in [8] to train the non-linear function iteratively, using the random forest regression model [15] to minimise the error of mapping the explanatory variables to the customer's load overall time steps. We then update behind-the-meter demand and solar generation in each iteration. The algorithm stops when the solar or demand changes in two consecutive iterations, is smaller than a threshold, or it reaches the maximum number of iterations. This algorithm is formalised as follows:

Let variable $\boldsymbol{\Theta} \in \mathbb{R}^{4\times T}$ collect the explanatory variables. Also, let vectors $L_i \in \mathbb{R}_+^T$, $S_i \in \mathbb{R}_+^T$ and $P_i \in \mathbb{R}^T$ respectively collect $l_{i,t}$, $s_{i,t}$ and $p_{i,t}$ for all the time steps. We then define function $f_i : \mathbb{R}^{4\times T} \to \mathbb{R}_+^T$ that maps variable $\boldsymbol{\Theta}$ to L_i. We suggest the following algorithm:

[1]Note that many countries, such as Australia, the grid code mandates that residential inverters should have the capability to provide reactive power support by working at constant power factor or following a standard piece-wise linear volt-VAR function. However, in most cases, these functionalities are not being used, and the inverters are working at the unity power factor.

Algorithm 1 Solar demand disaggregation using weather data

Initialise: $L_i^{[0]} = \max(P_i, 0),\ S_i^{[0]} = -\min(P_i, 0),\ iter = 0$
while $\max(S_i^{[iter]} - S_i^{[iter-1]}) > \epsilon$ and $iter \leq iterMax$ **do**:
  Train $f_i^{[iter]}$ using $L_i^{[iter]}$ and Θ
  $L^{[iter+1]} = f_i^{[iter]}(\Theta)$
  $S_i^{[iter+1]} = L_i^{[iter+1]} - P_i$
  $iter \ \ =+ \ 1$

G. Proxy Measurements and Weather Data

This approach, inspired by the approach suggested in [8], builds on approaches III-F and III-E. More specifically, it uses an iterative approach that in each iteration, it first uses the following optimisation model to update the solar generation:

$$W_i^{[iter]} := \underset{0 \leq W_i^{[iter]} \leq 1}{\arg\min} \ ||W_i^{[iter]} \times \boldsymbol{P^G} - \frac{\boldsymbol{S}_i^{[iter]}}{\overline{\boldsymbol{s}}_i}||_2^2 \quad (13a)$$

$$s.t.$$

$$\mathbb{1}^\intercal \ W_i^{[iter]} = 1. \quad (13b)$$

Then, similarly to algorithm 1, it uses a random forest regression model to estimate customer's load using weather data. This approach can be expressed by the following algorithm:

Algorithm 2 Solar demand disaggregation using the measurement from neighbouring nodes and weather data

Initialise: $S_i^{[0]} = -\min(P_i, 0)$, Solve (13) to obtain $W_i^{[0]}$, $iter = 0$
while $\max(S_i^{[iter]} - S_i^{[iter-1]}) > \epsilon$ and $iter \leq iterMax$ **do**:
  Solve optimisation (13) to obtain $W_i^{[iter]}$
  Use (11) to update $S_i^{[iter]}$
  $L_i^{[iter]} = S_i^{[iter]} + P_i$
  Train $f_i^{[iter]}$ using $L_i^{[iter]}$ and Θ
  $L^{iter+1} = f_i^{[iter]}(\Theta)$
  $S_i^{[iter+1]} = L_i^{[iter+1]} - P_i$
  $iter \ \ =+ \ 1$

IV. Numerical Experiments

A. Data

We use the Ausgrid [12] and NextGen [13] datasets to evaluate the estimation accuracy of different solar-demand disaggregation techniques described in the Section III. We also obtain weather data for the same period and location of the above electricity data from [11]. These real-world and publicly available datasets are briefly introduced in the following.

1) NextGen dataset: This dataset is made public from the Next Generation (NextGen) Residential Battery Storage trial in the Australian Capital Territory (ACT) in 2016. It includes a 5-minute resolution data of real and reactive power of demand and solar generation for 100 customers over the period of 1 January 2018 to 31 December 2018.

2) Ausgrid dataset: This dataset provides a half-hour real power electricity data for demand and solar generation of 300 homes with rooftop solar systems from the Ausgrid's electricity network in New South Wales, Australia. This relatively large dataset spans from July 1, 2010, to June 30, 2013. Note that since this dataset does not include reactive power, we only test the performance of the *Constant Power Factor Demand* technique on the NextGen dataset.

B. Simulation Platform

All the simulations are carried out using Python scripts. The optimisations are implemented using the Pyomo package and CPLEX is used as the solver. We made all the codes available online in an open-source package in [16]. The experiments are carried out on a server with a 128-core CPU and 128GB RAM.

C. Evaluation Metrics

To evaluate the accuracy of each technique, we calculate the distance between the inferred solar generation and the true value of the solar generation in each house. To do so, the mean square error (MSE) and the mean absolute error (MAE) are used, which are formulated as follows:

$$MSE_i = \frac{||S_i - \hat{S}_i||_2^2}{T}, \quad MSE = \frac{\sum_i^N MSE_i}{N} \quad (14a)$$

$$MAE_i = \frac{\sum_{t=1}^T \frac{|s_{i,t} - \hat{s}_{i,t}|}{s_{i,t}}}{T}, \quad MAE = \frac{\sum_i^N MPE_i}{N}, \quad (14b)$$

where symbol ($\hat{}$) is used to indicate the inferred solar generation values.

We are also interested in investigating the scalability of each approach. To do so, we analyse each approach based on their computational time and memory by exploring their type, e.g., categorising them based on being a linear problem (LP), quadratic problem (QP) and non-convex, and their number of variables (NoV) and constraints (NoC). We then check if these approaches can be done separately for each time and each customer, which allows parallel computation. The actual solve time for different techniques over the entire two datasets is also reported in Section IV-E.

D. Detailed Results for a Particular Day and Customer

We first report the simulation results for a randomly selected day and customer. Fig. 2 shows the real and inferred solar generation for the seven techniques described in Section III. For brevity, we refer to the techniques by their section number, e.g. *Tech. A* refers to the Minimum Solar Generation technique. We can see that the *Constant Power Factor Demand* approach can more accurately separate demand and solar generation from the net measurement than the alternative approaches. However, we must point out that this might not be generalisable to all other networks. For example, similarly to our assumption in this technique, the PV system works at the unity power factor, which may not be the case in some networks.

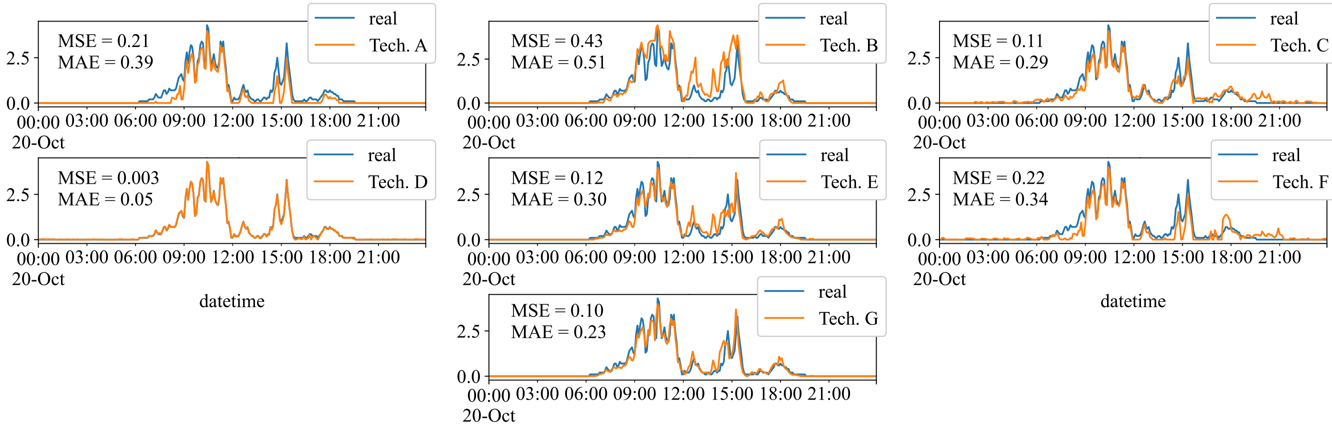

Fig. 2. Detailed simulation result for all the seven techniques described in Section III using the NextGen data for the 20th of October at a randomly selected customer. The legends in the figures refer to each technique's section number. For example, *Tech1* refers to the Minimum Solar Generation technique.

We now investigate the scalability of each approach. Model *Minimum Solar Generation* III-A is a linear problem that can be separated for each time and each customer. Model *Same Irradiance* III-B is also a linear problem that can be separated for each time step. Model *Same Irradiance and Houses Without PV Installation* III-C has a convex quadratic function as the objective function and a set of linear constraints, and thus, it is a convex quadratic problem (CQP). This problem can be separated for each time step, given that in each problem, we have access to the measurement from the other houses without PV installation. Model *Constant Power Factor Demand* III-D is a linear problem that can be separated for each customer and each day, e.g., given 30-minute resolution data, the problem includes 48 time-steps. Model *Measurements from Neighbouring Sites* III-E, similar to model III-C, is a quadratic problem that can be separated for each customer, given it has access to the PV measurement from the known neighbouring sites. Models *Weather Data* III-F and *Proxy Measurements and Weather Data* III-G include using a random forest regression model, which is a supervised learning algorithm and is a non-convex problem. On top of that, in these techniques, we iteratively solve the non-convex problem in the hopes of convergence! Although we have not seen a divergence in our simulation, extra care should be in place when using these techniques. We will provide more discussions on this in Section IV-E.

Table I summarises our discussion above. In this table, NoC, NoV, Sep T and Sep. C refer to the number of variables, the number of constraints, separable for each time step and separable for each customer, respectively. Also, we use the symbol $(.) \times (.)$ to separate the number of variables and constraints in each subproblem. For example, NoC $= 1 \times NT$ means there is one constraint in each of the N multiplied by T problems, while $N \times T$ means there are N constraints in each of the T problems. Also, note that we left the number of variables and constraints empty for the last two techniques, as they depend on the number of iterations it takes for the algorithms to converge.

TABLE I
SCALABILITY OF SOLAR DEMAND DISAGGREGATION TECHNIQUES.

Technique	NoC	NoV	Sep. T	Sep. C	Complexity
Tech. A	$1 \times$ N T	$2 \times$ N T	Y	Y	LP
Tech. B	(N + 1) $\times$ T	N $\times$ T	Y	N	LP
Tech. C	3N $\times$ T	3N $\times$ T	Y	N	CQP
Tech. D	$1 \times$ N T	$1 \times$ N T	Y	Y	LP
Tech. E	GT $\times$ N	1 $\times$ T	N	Y	CQP
Tech. F	-	-	N	Y	NP-hard
Tech. G	-	-	N	Y	NP-hard

E. Average Results Over the Entire Datasets

Figures 3 and 4 show the boxplot of MSE and MAE of the described techniques over the entire NextGen and Ausgrid datasets. From Fig. 3, we can see that the *Constant Power Factor Demand* technique clearly outperforms other approaches. Also, we can see that the *Weather Data* and *Proxy Measurements and Weather Data* techniques (Tech. E and Tech. G) overall produce less error compared to alternative approaches. This is also shown in Table II, where the average values of MSE and MAE over all the customers are reported. As expected, since techniques *Minimum Solar Generation* and *Constant Power Factor Demand* are linear and separable both in time and for each customer, it takes less time to solve the disaggregation problem using them. We can also see that using the *Weather Data* (Tech. E) technique to solve the problem takes more time than *Proxy Measurements and Weather Data* (Tech. G) technique. This is interesting since in Tech. G, in addition to training the non-linear function f, as is done in Tech. F, we need to solve the optimisation problem (13). The reason is that, overall, Tech. G takes fewer iterations to converge than Tech. F. An example of this is shown in Fig. 5.

V. Conclusion

We study seven techniques to disaggregate behind-the-meter solar generation and demand from their aggregated values. These techniques use optimisation models and machine learning algorithms to incorporate a different set of information

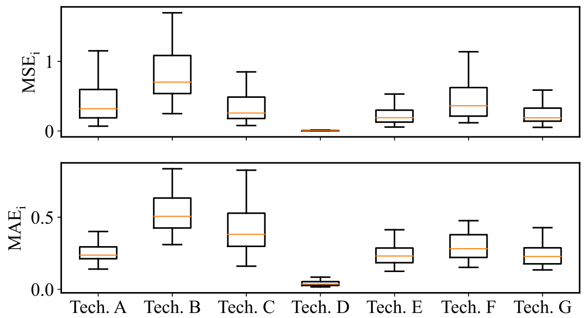

Fig. 3. Boxplot of MSE_i and MAE_i the seven techniques described in Section III for the customers in the NextGen data.

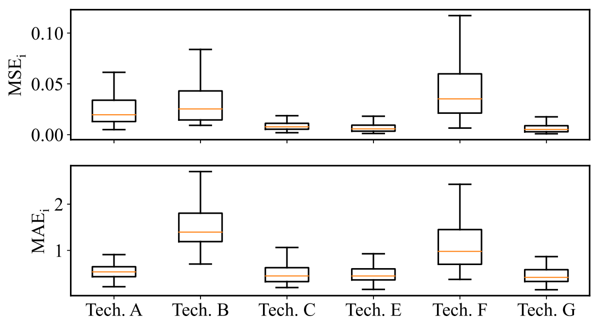

Fig. 4. Boxplot of MSE_i and MAE_i the seven techniques described in Section III for the customers in the Ausgrid data.

TABLE II
SIMULATION RESULTS OVER THE ENTIRE DATASETS.

Data/Techniques		T.A	T.B	T.C	T.D	T.E	T.F	T.G
Nex.	MSE(%)	42.2	84.9	45.4	6.1	23.4	49.6	24.7
	MAE	0.25	0.57	0.43	0.04	0.25	0.30	0.26
	Time(s)	3	1456	3640	8	12	1984	753
Aus.	MSE(%)	2.89	4.28	9.92	-	0.92	4.36	0.88
	MAE	0.55	1.55	0.63	-	0.50	1.29	0.47
	Time(s)	3.8	595	1840	-	38	680	192

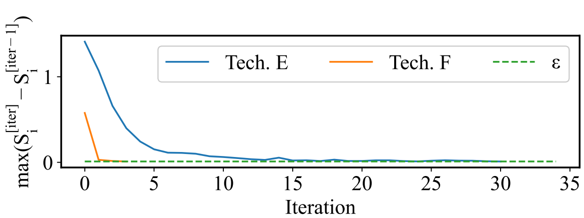

Fig. 5. An example of the convergence of Tech. F and Tech. G using the NextGen data for a randomly selected customer. In our experiments, we consider $\epsilon = 0.01$ and the maximum number of iterations equals 50.

to find the solution. Our analyses and experiments show that: *i)* using reactive power to solve the BSD problem provides better accuracy and computational efficiency compared to the other information; *ii)* assuming that all the customers receive a similar irradiance creates a considerable error in the solar and demand estimates (more than the base case of assuming that solar and demand are equal to the power injection and consumption, respectively; *iii)* iteratively using machine learning techniques is prone to convergence issues, and extra care should be in place when using such techniques.

A further extension to this work would be to conduct a cost analysis study on the impact of various BSD techniques. For example, the study can consider the cost of obtaining the additional information, extra computation and required expenses in each BSD technique and analyse the financial return of having a more accurate BSD technique for system operators and customers.

ACKNOWLEDGEMENT

This work was done in the framework of the project “Converge”. This project received funding from the Australian Renewable Energy Agency (ARENA) as part of ARENA’s Advancing Renewables Program.

REFERENCES

[1] S Mahdi Noori RA, Mark Burgess, Masoume Mahmoodi, Ahmad Attarha, and Paul Scott. An adjustable scenario optimisation approach in operating PV-rich distribution systems. *IEEE Trans. Power Syst.*, 2022.

[2] Masoume Mahmoodi, Ahmad Attarha, S Mahdi Noori RA, Paul Scott, and Lachlan Blackhall. Adjustable robust approach to increase DG hosting capacity in active distribution systems. *Electric Power Systems Research*, 211:108347, 2022.

[3] Fankun Bu, Kaveh Dehghanpour, Yuxuan Yuan, Zhaoyu Wang, and Yifei Guo. Disaggregating customer-level behind-the-meter PV generation using smart meter data and solar exemplars. *IEEE Trans. Power Syst.*, 36(6):5417–5427, 2021.

[4] George W Hart. *Prototype nonintrusive appliance load monitor: Progress report 2*. MIT Energy Laboratory, 1985.

[5] Pascal A Schirmer and Iosif Mporas. Non-intrusive load monitoring: A review. *IEEE Trans Smart Grid*, 2022.

[6] Xinlei Chen and Omid Ardakanian. Solar disaggregation: State of the art and open challenges. In *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, pages 6–10, 2020.

[7] Dong Chen and David Irwin. Sundance: Black-box behind-the-meter solar disaggregation. In *Proceedings of the eighth international conference on future energy systems*, pages 45–55, 2017.

[8] Xinlei Chen and Omid Ardakanian. Data efficient energy disaggregation with behind-the-meter energy resources. *Sustainable Energy, Grids and Networks*, 32:100813, 2022.

[9] Joe Brown, Alessandro Abate, and Alex Rogers. Disaggregation of household solar energy generation using censored smart meter data. *Energy and Buildings*, 231:110617, 2021.

[10] Evangelos Vrettos, EC Kara, EM Stewart, and Ciaran Roberts. Estimating PV power from aggregate power measurements within the distribution grid. *J. Renew. Sustain. Energy*, 11(2):023707, 2019.

[11] Nicholas Engerer. Global solar irradiance data via API (2018), the Australian National University data commons.dataset. *Online: https://solcast.com*, 2018.

[12] Ausgrid, Solar home electricity data. *[Online]. Available: https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data.*

[13] Marnie Shaw, Bjorn Sturmberg, Lin Guo, Xinyu Gao, Elizabeth Ratnam, and Lachlan Blackhall. The nextgen energy storage trial in the act, australia. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, pages 439–442, 2019.

[14] Xinlei Chen, Moosa Moghimi Haji, and Omid Ardakanian. Disaggregating solar generation using smart meter data and proxy measurements from neighbouring sites. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, pages 225–230, 2021.

[15] Cory Lesmeister. *Mastering machine learning with R*. Packt Publishing Ltd, 2015.

[16] S. Mahdi Noori R. A. and D. Gordon. Converge load forecasting package. *https://pypi.org/project/converge-load-forecasting/*, Nov. 2022.