This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3606526

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

WLS-based behind-the-meter PV generation disaggregation using partial AMI measurements with first-order time scaling

DONGHO HYUN[1,2], (Student Member, IEEE), JAEBEOM IM[3], (Member, IEEE), JAEPIL BAN[1,2], (Member, IEEE), HYEONOK LIM[4], (Member, IEEE), JONGMIN JO[4], (Member, IEEE), SEONGSOO CHO[4], (Member, IEEE) and YOUNG-JIN KIM[3], (Senior Member, IEEE),

[1]School of Electronic Engineering, Kumoh National Institute of Technology, Gumi 39177, Republic of Korea
[2]Department of Aeronautic, Mechanical, and Electronic Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, Republic of Korea
[3]Department of Electrical Engineering, Pohang Unversity of Science and Technology, Pohang 37673, Republic of Korea
[4]Korea Electric Power Research Institute Power Distribution Research Institute, Daejeon 34056, Republic of Korea

Corresponding author: Jaepil Ban (jpban@kumoh.ac.kr) and Young-Jin Kim (powersys@postech.ac.kr)

This work was supported in part by the Korea Electric Power Research Institute(KEPRI) grant funded by the Korea Electric Power Corporation(KEPCO)(R19DG05) and in part by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Ministry of Trade, Industry&Energy(MOTIE) of the Republic of Korea(No.20193710100061)

ABSTRACT The increasing penetration of distributed photovoltaic (PV) generation systems due to the expansion of renewable energy has posed challenges to power grid operators, particularly in managing the uncertainty caused by behind-the-meter (BTM) PV systems. These systems are unmeasured, complicating gross load estimation and grid management. To address this issue, a novel disaggregation method utilizing net load measurements and partially aggregated advanced metering infrastructure (AMI) data is proposed. The proposed method uses a weighted least squares (WLS) method by assigning higher weights to nighttime errors to mitigate daytime errors caused by PV generation. Additionally, a first-order time-scaling method is introduced to correct temporal discrepancies caused by partially aggregated AMI data. The proposed method is validated using data provided by the Korea Electric Power Research Institute (KEPRI).

INDEX TERMS Advanced measurement infrastructure, gross load estimation, load disaggregation, non-intrusive load monitoring, photovoltaic generation.

I. INTRODUCTION

As carbon neutrality accelerates, the use of renewable energy has surged in recent years [1], [2], [3]. For example, the South Korean government has set a target to increase the share of renewable energy to over 20% by 2030 [4]. In many countries, including Denmark, Ireland, and Germany, the penetration rate of distributed renewable energy sources has surpassed 20% at the national level [5]. However, the high penetration of distributed resources introduces new challenges for power grid operators as it causes issues such as voltage stability and supply-demand imbalance [6]. In particular, most distributed photovoltaic (PV) generation systems are installed behind the meter (BTM) and system operators do not directly measure them [7]. According to [8], some companies share PV data with system operators in a non-real-time, retrospective manner. [9] indicates that National Grid is gradually developing infrastructure for real-time DER data sharing, but such capabilities have not yet been fully established. These practices remain limited and are not systematically implemented in many regions [10]. As a result, the net load measured at substations appears as a combination of the gross load and BTM PV generation, introducing significant uncertainty in forecasting gross load and, hence, challenging grid management recently.

Various disaggregation methods have been developed to separate the BTM PV generation from the net load profile measured at substations. For example, [11] proposed a model-based method that combines a clear sky generation model and weather-solar efficiency effects to estimate BTM PV generation. [12] proposed a maximal information coefficient-based correlation analysis for BTM PV estimation. In [13], an iterative algorithm was proposed to learn a physical model for estimating BTM PV generation and a mixed Hidden Markov model for estimating gross load. However, these studies often rely on detailed PV system specifications or can yield inaccurate results when the actual PV installation location deviates significantly from the assumed location.

In this regard, data-driven methods, which do not require physical models or PV system specifications, have been investigated [14], [15], [16]. In [14], a method using dimension reduction and mapping functions was proposed to estimate BTM PV generation, dealing with limited observable PV sites. In [15], a contextually supervised source separation method was proposed for real-time PV generation monitoring. [16] proposed a method for estimating community-level BTM PV generation using a federated learning-based Bayesian neural network. In [17], an adaptive machine learning framework was proposed for BTM PV estimation. However, these data-driven methods require the ground truth of the BTM PV generation, meaning that they cannot be used when there is no ground-truth data available.

To overcome the lack of BTM PV generation data, researchers have investigated methods using PV proxy such as solar irradiation [18], [19], [20]. In [18], a linear regression model was used for disaggregation using net load data and irradiation data. [19] introduced PV and gross load sensitivity models to estimate the BTM PV generation, using net load and meteorological data. However, since these methods rely on irradiation data to estimate PV generation, there is a nonlinear discrepancy between the irradiation data and the BTM PV generation. This issue has also been highlighted in previous studies [20], [21]. [22] proposed a game-theoretic data-driven method that uses only smart meter data to estimate BTM PV generation. However, it can demand a significant amount of library data for the accurate estimation.

Meanwhile, the adoption of advanced metering infrastructure (AMI) has seen remarkable growth. For example, in the United States, the number of operational AMI meters increased nearly twelve-fold between 2007 and 2017, resulting in AMI meters accounting for 51.9% of all meters [23]. Researchers have investigated the application of AMIs for load disaggregation [24], [25], [26]. However, these studies primarily focused on non-intrusive load monitoring of individual household appliances rather than estimating aggregated substation-level loads. In addition, while the AMI can offer opportunities for accurate disaggregation, its penetration rate varies across countries and regions, resulting in only partial measured data availability depending on areas.

Motivated by the above discussion, this study proposes a BTM PV generation disaggregation method that uses only net load measurement and partially measured aggregated AMI data, enhancing its practical applicability. Specifically, we propose a weighted least squares (WLS)-based disaggregation method by assigning higher weights to nighttime estimation errors, thereby reducing the impact of daytime estimation errors caused by unknown BTM PV generation. We analyze the correlation between the gross load and aggregated AMI profiles to verify their similarity. Subsequently, the aggregated AMI profile is then used for estimating gross load. In addition, we introduce a temporal first-order time-scaling to compensate temporal discrepancies arising from unmeasured portions in the aggregated AMI data. The proposed method is verified and analyzed using the hourly collected data,

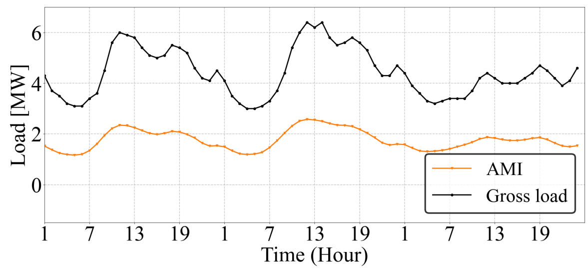

FIGURE 1. Aggregated AMI and gross load profiles

characterized by AMI penetration rates and weights between daytime and nighttime errors. The main contributions of this study are summarized as:

- The proposed BTM PV generation disaggregation method uses only net load measurement and partially aggregated AMI data, without requiring PV generation and irradiation profiles.
- A temporal first-order time-scaling method is proposed to mitigate temporal discrepancies arising from the partially aggregated AMI data.
- A WLS-based disaggregation method is proposed, which assigns higher weights to nighttime estimation errors to reduce the impact of daytime estimation errors caused by unknown BTM PV generation.

The remainder of this paper is organized as follows. Section II analyzes the aggregated AMI data and gross load and formulates the BTM PV generation disaggregation problem. Section III proposes a BTM PV disaggregation model introducing a first-order time-scaling method and presents a WLS-based estimation method that assigns higher weights to nighttime data. Section IV verifies the disaggregation performance of the proposed model and estimation method using hourly collected data. Finally, Section V concludes the paper and suggests directions for future research.

II. PROBLEM FORMULATION

The net load at the substation is represented as the sum of the gross load and BTM PV generation and expressed as follows

$$P_{\text{net}}(t) = P_{\text{grossload}}(t) + P_{\text{pv}}(t), \tag{1}$$

where $P_{\text{net}}(t)$ denotes the net load; $P_{\text{grossload}}(t)$ represents the gross load; $P_{\text{pv}}(t)$ is the PV generation for $t \in [t_1, t_2, ..., t_T]$ for $T > 0$. In this study, we consider the scenario that $P_{\text{net}}(t)$ can be measured at the substation, whereas $P_{\text{grossload}}(t)$ and $P_{\text{pv}}(t)$ are not measured directly by distribution system operators. In addition, in the scenario under consideration, AMI is installed for selected customers in the substation, providing their daily load profiles. As AMI data are gathered from a number of customer sites, the patterns of the aggregated AMI profiles can be similar to those of the gross load even if the magnitude of the aggregated AMI data is smaller than that of the gross load at each time.

We briefly analyze the similarity between the aggregated AMI and gross load profiles. Figure 1 shows the gross load

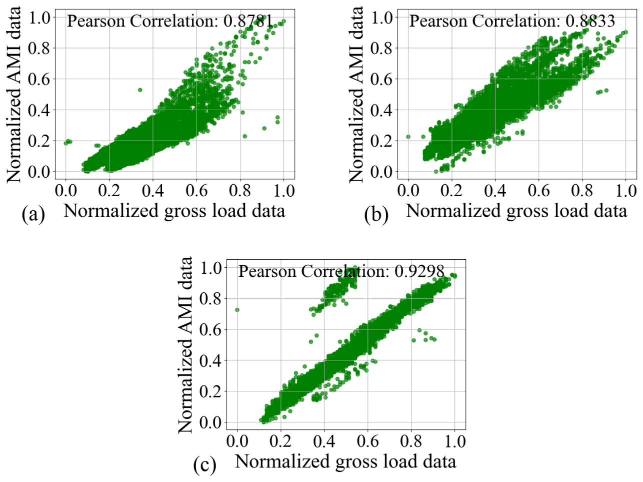

FIGURE 2. Correlation between the gross load and aggregated AMI data for each feeder with the AMI aggregation rate of (a) 41%, (b) 69%, and (c) 92%

and aggregated AMI data. The black line represents the gross load data within the feeder. The orange line represents the aggregated AMI data with the aggregation rate of 41%. A notable point in Figure 1 is that, despite only 41% of the gross load being captured by AMI, its profile shows a pattern very similar to that of the gross load profile. This similarity can be attributed to the smoothing effect on individual customer load variability when multiple customer load data are aggregated, resulting in more regular and predictable load patterns [27].

To quantitatively evaluate the profile similarity between the two datasets, we conducted a correlation analysis between the aggregated AMI data and the gross load based on different AMI aggregation rates. The results, presented in Figure 2, demonstrate consistently high correlation coefficients across all feeders, indicating that the gross load profile can be well represented by the aggregated AMI data, even at low aggregation rates. AMI and gross load data are based on customer consumption behavior and inherently reflect external factors such as weather, seasonal variations, and social activities. Therefore, a pattern based estimation method can be applied without relying on additional external variables.

Therefore, the goal of this study is to propose a method to accurately estimate the gross load $P_{\text{grossload}}(t)$ based on the aggregated AMI data without the information of the BTM PV generation, thereby demonstrating the potential to disaggregate BTM PV generation $P_{\text{pv}}(t)$ from the net load $P_{\text{net}}(t)$.

III. BTM PV GENERATION DISAGGREGATION METHOD

A. OVERALL STRUCTURE OF THE PROPOSED METHOD

Figure 3 provides an overview of the proposed method for BTM PV generation disaggregation. The proposed method obtains two inputs to disaggregate BTM PV generation: the net load data measured at the substation level and the aggregated AMI data collected at the customer level. Then, the two input data are used for constructing the proposed first-order time-scaling-aided linear regression model, and the WLS-

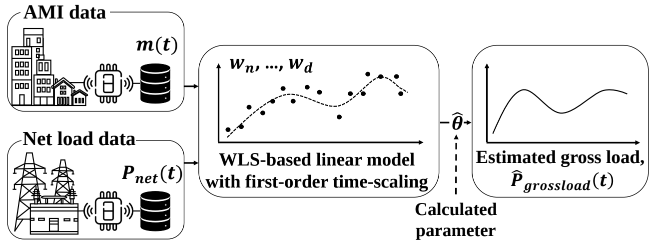

FIGURE 3. Overview of the proposed method for load/PV disaggregation

based gross load estimation method is then performed with the proposed weighting value selection strategy to obtain the unknown coefficients in the proposed linear model. Finally, using the obtained coefficients, the linear model can estimate the gross load accurately, disaggregating the BTM PV generation from the net load.

B. LINEAR MODEL WITH FIRST-ORDER TIME-SCALING COEFFICIENT

The proposed model is defined as follows:

$$\hat{P}_{\text{net}}(t) = \hat{P}_{\text{grossload}}(t) + \varepsilon(t), \tag{2a}$$

$$\hat{P}_{\text{grossload}}(t) = \theta_m m(t) + C_g + C_l l(t) + \xi(t), \tag{2b}$$

where $m(t)$ is the aggregated AMI data; θ_m represents the scaling factor of $m(t)$; C_g and C_l denote the coefficients for zero- and first-order regression terms, respectively; $\varepsilon(t)$ represents the error arising from the unknown PV generation; $\xi(t)$ denotes the gross load estimation error. It is worth noting that, to account for temporal trends in load estimation, we introduce a first-order time-scaling variable $l(t)$, which is defined as:

$$l(t) = \frac{t-1}{T-1}, \tag{3}$$

where T represents the number of measurements in a day. For example, $T = 24$ indicates we have one-hour period measurement data. Then, t takes values from 1 to 24, and $l(t)$ increases linearly from 0 to 1. Therefore, it effectively corrects offset mismatches between the estimated load and the net load during nighttime.

In this study, aggregated AMI data are used to estimate the gross load. However, since AMI systems are only partially deployed across distribution feeders, unmeasured portions of load may lead to temporal discrepancies. These discrepancies arise not only from the limited coverage of AMI systems but also from differences in load characteristics among customer types, especially when the data are biased toward specific customer groups. To clarify this, Figure 4 presents daily load profiles by customer type. In the example distribution feeder, General Use customers account for the majority, resulting in a gross load pattern with a distinct daytime peak. In contrast, other customer types, including Residential Power, have a smaller share and thus limited influence on the gross load. Although the aggregated AMI data may appear similar to the gross load, temporal discrepancies occur because the load

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3606526

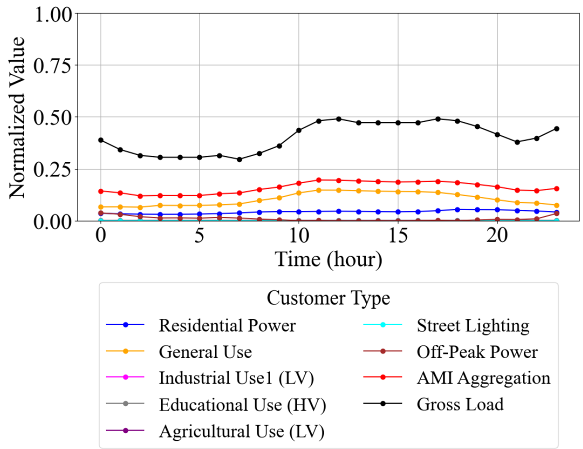

FIGURE 4. Normalized hourly load profiles for a specific day, including gross load, aggregated AMI, and customer-type-specific AMI data for feeders with AMI aggregation rates of 41%

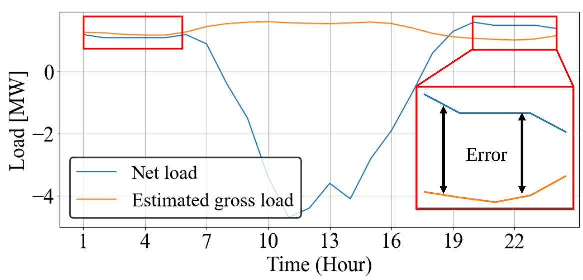

FIGURE 5. Net load and estimated gross load profile without applying the first-order time scaling

characteristics of unmeasured customer types are not fully reflected due to the low aggregation rate.

This issue is visually depicted in Figure 5, which shows the estimated gross load using aggregated AMI data without applying the first-order time-scaling adjustment. The figure compares the estimated gross load with the net load during nighttime. As a result, significant offset mismatches are observed in the early morning and afternoon/evening periods.

C. WLS-BASED BTM PV DISAGGREGATION

This section proposes the WLS method to obtain the coefficients of the proposed model, thereby estimating the gross load using available data. The proposed method assigns different weights to data points during the optimization process to reflect the varying importance of data over time [28]. The WLS method is formulated as follows:

$$\mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{X}\boldsymbol{\theta}, \tag{4}$$

where $\mathbf{W}$ is a diagonal matrix containing weights. The coefficient vector $\boldsymbol{\theta}$ is estimated by minimizing the weighted sum of squared residuals, formulated as follows:

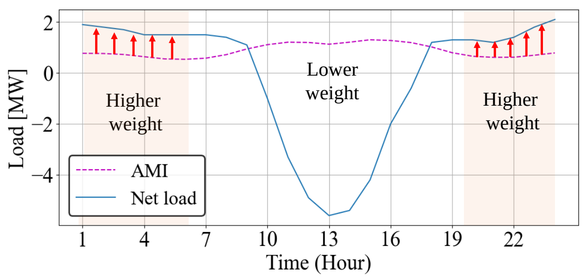

FIGURE 6. Net load and aggregated AMI profiles

$$\begin{aligned} \min_{\theta_m, C_g, C_l} \quad & \|\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\|_2^2 \\ \text{subject to} \quad & \theta_m \geq 0, \\ & C_g \geq 0, \\ & C_l \in (-\infty, \infty), \end{aligned}$$

where

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_T \end{bmatrix}, \mathbf{X} = \begin{bmatrix} m(t_1) & 1 & l(t_1) \\ m(t_2) & 1 & l(t_2) \\ \vdots & \vdots & \vdots \\ m(T) & 1 & l(T) \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} P_{\text{net}}(t_1) \\ P_{\text{net}}(t_2) \\ \vdots \\ P_{\text{net}}(T) \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \theta_m \\ C_g \\ C_l \end{bmatrix}.$$

The proposed WLS method can effectively estimate the gross load using the proposed model (2). Nighttime data are not affected by solar power generation and exhibit high reliability between net load and total load. Therefore, WLS assigns higher weights to nighttime data to maximize optimization accuracy. On the other hand, daytime data has high uncertainty due to unknown BTM PV generation. Thus, lower weights are assigned to mitigate distortions that may occur due to errors arising from unknown BTM PV generation (Fig. 6).

D. METHOD OF ASSIGNING WEIGHTS

Appropriate selection of the weighting matrix $\mathbf{W}$ is crucial, as it can significantly affect the model's performance. The weighting matrix is defined as follows:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{24} \end{bmatrix}, \tag{5}$$

$$w_t = \begin{cases} w_n, & \text{if } t \in t_{\text{night}} \\ w_d, & \text{otherwise} \end{cases}, \tag{6}$$

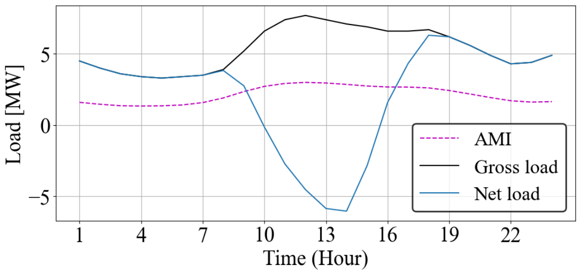

FIGURE 7. The example of the three types of data

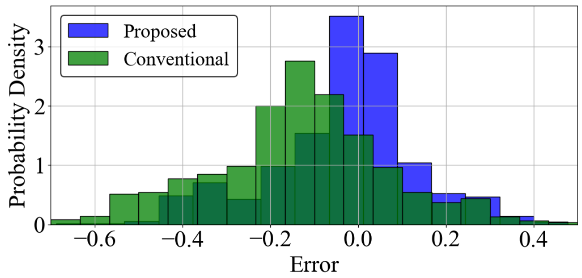

FIGURE 8. Probability-based comparison of the proposed and conventional models

where w_n represents the weight for the nighttime, and w_d represents the weight for the daytime; t_{night} denotes the prescribed time interval for nighttime. To reflect the importance of nighttime data, we can set w_n larger than w_d. A key consideration when setting the weights is that the duration of the nighttime period should be flexibly adjusted according to regional or seasonal variability. In addition, the magnitude of the weights is determined at the user's discretion. To provide better understanding, we analyze the effect of different weight values in Section IV-D.

IV. EXPERIMENTAL RESULTS

A. DATASET

The datasets used in this study were provided by the Korea Electric Power Research Institute (KEPRI). The datasets consist of hourly data collected from January 1st to December 31st in 2023. They were collected from three anonymous regions with different aggregation rates. The datasets for each region consist of the following three types: AMI, gross load, and net load data.

1) AMI data: this data aggregates the measured demand of customers within each region without including PV generation. The AMI aggregation rates of the three regions are 41%, 69%, and 92%.
2) Gross load data: gross load data were measured at the substations of feeders that do not contain BTM PV generation. These data serve as the ground truth data for evaluating the performance of the estimated gross load results.
3) Net load data: To construct net load data that includes BTM PV generation and gross load data, BTM PV generation data were collected from locations near the aforementioned three regions for the same time period. Subsequently, the net load data were constructed by adding the PV generation data to gross load data for each region.

The example of the three types of data for a day is shown in Figure 7. Considering regional and seasonal conditions in the test environment, nighttime weight hours were assigned as $t_{\text{night}} = \{22, 23, 24, 1, 2, 3, 4, 5\}$.

B. EVALUATION METHOD

In this study, we evaluate the accuracy of the proposed method using data collected hourly over the course of one year. The proposed method is evaluated using the mean absolute error (MAE) and root mean square error (RMSE) metrics, which are defined as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |p_i - \hat{p}_i| \,,$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \hat{p}_i)^2},$$

where p_i represents the actual value at time i, $\hat{p}_i$ represents the predicted value at time i, and N is the total number of data points.

C. COMPARISON ANALYSIS OF ESTIMATION MODELS

In this subsection, we present a comparative analysis between the proposed and conventional models, where we define the conventional model as one without first-order time-scaling. Table 1 presents the prediction errors of the models based on the evaluation method described in Section IV-B, showing the prediction errors of the proposed model and conventional model for a feeder with an AMI aggregation rate of 41%. In this process, arbitrary weights were applied, with the daytime weight set to 1 and the nighttime weight set to 10. The analysis results, presented in Table 1,indicate that the proposed model achieved approximately a 9.12% reduction in MAE and a 17.59% reduction in RMSE compared to the conventional model.

In addition, Figure 8 shows the error distribution during 22:00–24:00. The proposed model exhibited a bias of -0.0357 and variance of 0.0312, while the conventional model showed a bias of -0.1380 and variance of 0.0429. These results quantitatively demonstrate that the proposed model achieved a significantly lower bias, with a reduction of approximately 74.11%, and a smaller variance, reduced by 27.32%, compared to the conventional model.

D. EFFECT OF WEIGHT VARIATIONS IN THE WLS METHOD

In this subsection, the effects of varying nighttime weights on the disaggregation performance were analyzed. The proposed model was evaluated with nighttime weights set to 1,

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3606526

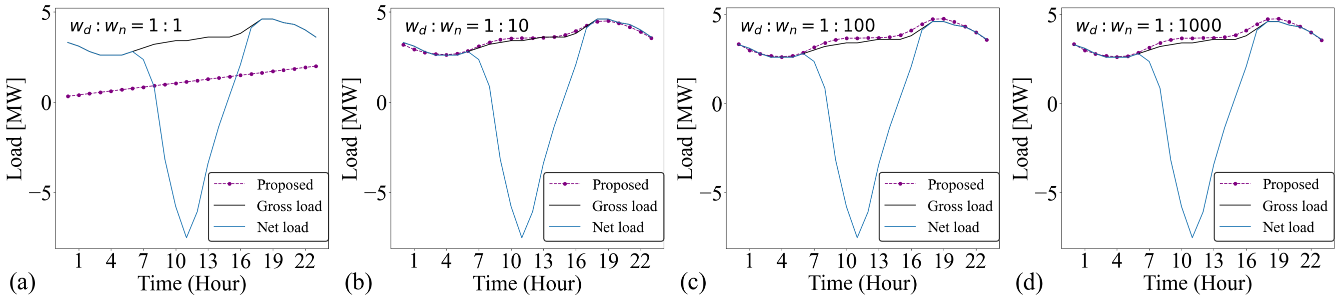

FIGURE 9. Comparison of WLS-based model performance with various weights applied to the feeder with an AMI aggregation rate of 69% on November 5, 2023

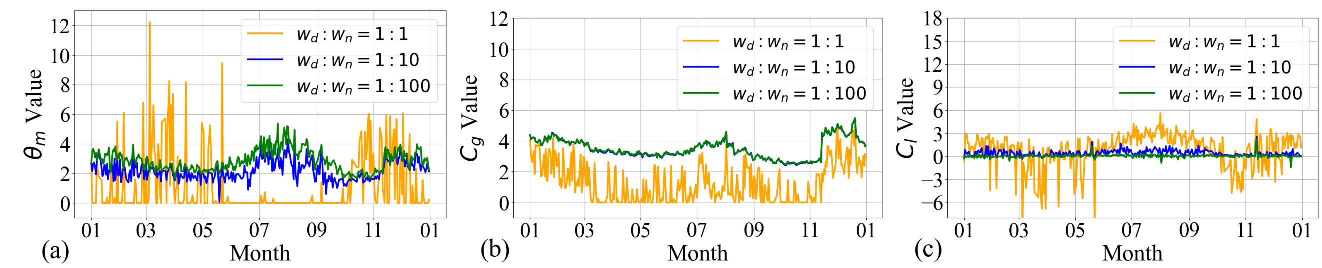

FIGURE 10. The results of analyzing the feeder with an AMI aggregation rate of 69% using the WLS-based model with various weighting ratios of the optimization coefficients (a) θ_m, (b) C_g, and (c) C_l

Model	MAE	RMSE
Proposed	0.3349	0.4731
Conventional	0.3685	0.5741

TABLE 1. Comparison between models with and without first-order time-scaling

10, 100, and 1000, and the resulting performance changes were compared. Figure 9 illustrates the prediction results for a specific day of a feeder with an AMI aggregation rate of 69% under various weighting configurations. Figure 9a shows the results when a 1:1 weighting ratio was applied, meaning that the standard least square method was used. The result shows that it failed to achieve accurate predictions, particularly during the daytime. When the weighting ratio was set to 1:10 or higher, prediction accuracy improved (Fig. 9). Table 2 shows that among the tested weights, a nighttime weight of 10 achieved the best performance. Increasing the nighttime weight to 100 or higher did not result in further improvements.

Figure 10 shows the obtained coefficients of the proposed model for the one-year data under various weight ratios in the WLS method. When the weight ratio of $1:1$ was applied, the variance of the obtained coefficients was large. This is because the coefficients were adjusted to minimize the large errors of the daytime period, which includes the errors caused by the unknown BTM PV generation. In the case of a weight ratio of $1:10$, the variance of the coefficients was relatively small. Since larger weights were assigned for nighttime errors than those for daytime errors, the coefficients were obtained to mainly reduce nighttime errors, which reduces the influence of the large daytime errors caused by unknown BTM PV generation, thereby decreasing the variance of the coefficients.

$w_d : w_n$	MAE	RMSE
1:1	2.5747	2.8904
1:10	0.2810	0.3772
1:100	0.3104	0.4462
1:1000	0.3114	0.4480

TABLE 2. Performance of the proposed method with various weights

E. EFFECTS OF AMI AGGREGATION RATES

In this subsection, the performance of the proposed method is evaluated under various AMI aggregation rates. Specifically, a weighting ratio of 1:10 was applied for AMI aggregation rates of 41% and 69%, and a ratio of 1:100 was applied for 92%, as these settings yielded the lowest estimation errors in each case. Figure 11a shows that even when the AMI aggregation rate is as low as 41%, the proposed method yields relatively accurate estimates by successfully capturing the

AMI aggregation rate	MAE	RMSE
41%	0.3349	0.4731
69%	0.2810	0.3772
92%	0.1200	0.1900

TABLE 3. Gross load estimation errors for various AMI aggregation rates

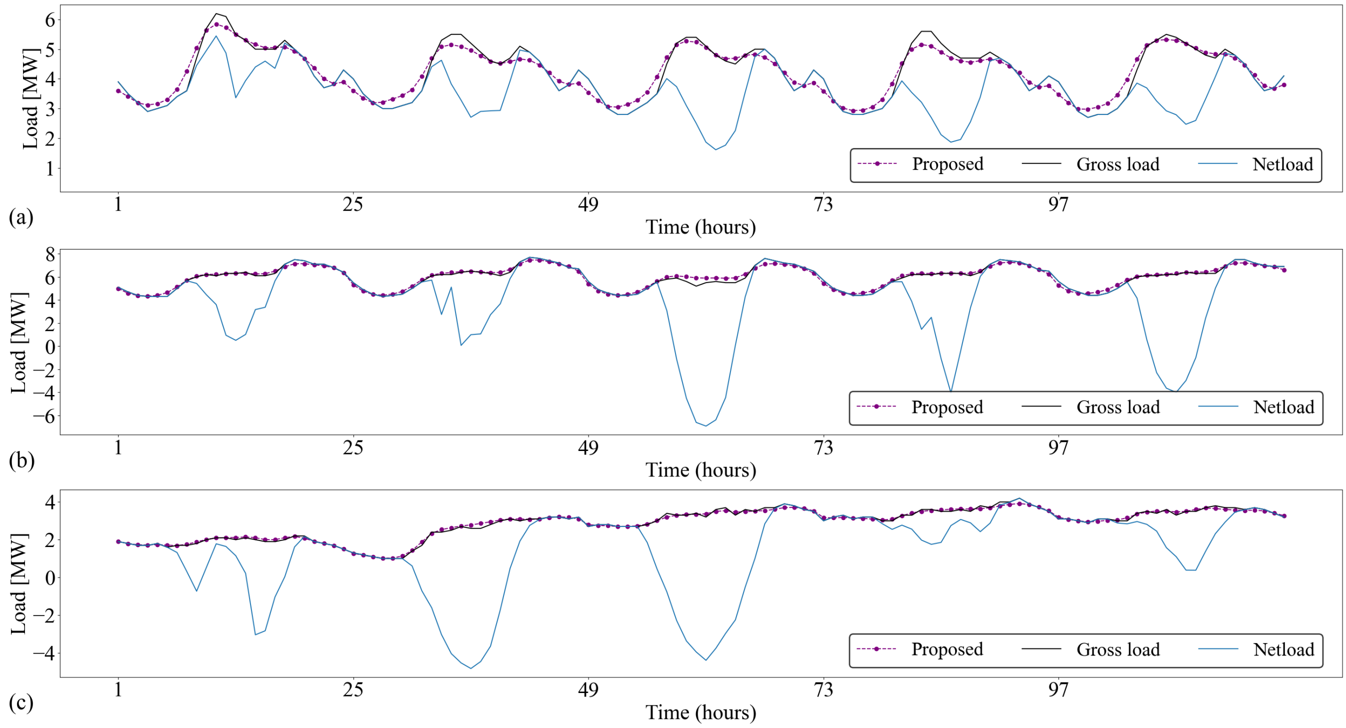

FIGURE 11. BTM PV disaggregation results for each feeder with an AMI aggregation rate of (a) 41%, (b) 69%, and (c) 92%

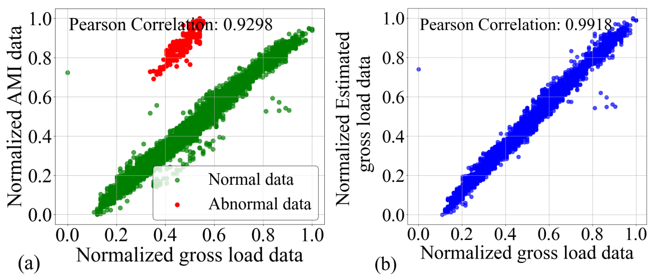

FIGURE 12. (a) Correlation between the gross load and the aggregated AMI data for a feeder with 92% AMI aggregation rate (b) Correlation between the gross load and the estimated gross load obtained by the proposed method for the same feeder

overall trend of the gross load, despite slight deviations in short-term variability. This demonstrates the method's reliable performance under constrained data conditions.

Figures 11b and 11c demonstrate that as the AMI aggregation rate increases to 69% and 92%, the proposed method exhibits significantly improved agreement with the gross load. The time-resolution accuracy also improves, and the proposed method more precisely captures the load trends. This trend is also reflected in Table 3, where higher AMI aggregation rates correspond to lower estimation errors. These findings reinforce the expected benefit of increasing AMI aggregation for load estimation. More importantly, however, they highlight the robustness of the proposed method in scenarios where AMI data is only partially available.

F. ROBUSTNESS TO AMI MEASUREMENT ANOMALIES

This subsection analyzes the robustness of the proposed method against anomalies present in AMI data. Figure 12a illustrates the correlation between the aggregated AMI data and the gross load for a feeder with an AMI aggregation rate of 92%, where each data point is color-coded based on its correlation characteristics. The green points represent data within the normal category, where the AMI data exhibit a relatively linear relationship with the gross load. In contrast, the red points represent abnormal data, where the AMI values are excessively high compared to the gross load, indicating local nonlinearity. This observed nonlinearity is likely caused by anomalies due to data collection or processing errors. Since a considerable amount of time has passed since the data were collected, it is difficult to determine the exact causes of these anomalies. Therefore, this study does not interpret them as definitive errors but considers them likely measurement irregularities due to unidentified causes.

Figure 12b presents the correlation between the estimated gross load generated by the proposed method and the gross load data. Despite the presence of nonlinear anomalies in the AMI data observed in Figure 12a, the proposed method mitigates the impact of such abnormal segments and shows an overall high linear correlation. Figure 13 compares the proposed method with the gross load. The red dashed line corresponds to the nonlinear data segments identified in Figure 12a, while the green dashed line represents a portion of the relatively linear segments. Overall, the proposed method closely follows the trend of the gross load. However, between

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3606526

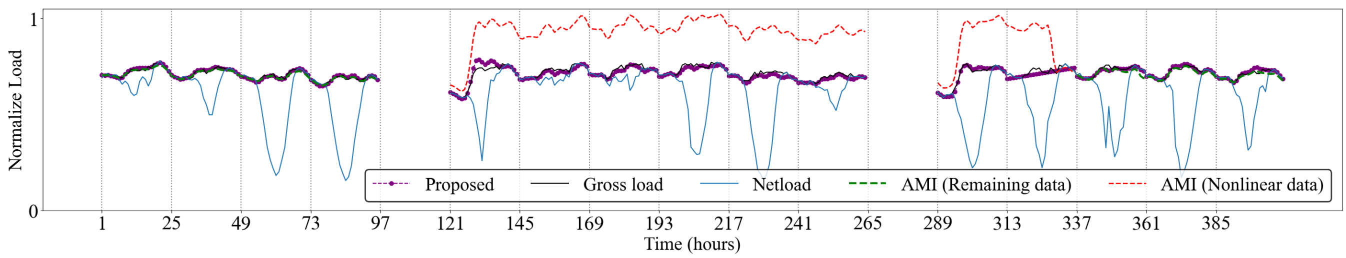

FIGURE 13. Time-series comparison between the BTM PV disaggregation results and the gross load, highlighting nonlinear (red dashed line) and linear (green dashed line) segments identified in Figure 12a

hours 313 and 337, some AMI data exhibited abnormal patterns that differed from the gross load, and due to the nature of the estimation method, which operates on a daily basis, the estimated values showed some discrepancies from the gross load. Nevertheless, overall, the proposed method suggests that it can reliably estimate the gross load even when local nonlinear anomalies exist in the AMI data.

V. CONCLUSION

This study proposed a method to disaggregate BTM PV generation in distribution grids with high BTM PV penetration by utilizing substation-level net load and partially aggregated AMI data. The proposed method introduced a first-order time-scaling coefficient to adjust temporal discrepancies caused by partial AMI aggregation. A WLS-based method was utilized to assign higher weights to nighttime data in order to reduce the impact of daytime PV generation variability. The proposed method included a comparison of estimation models, an analysis of the impact of weight variations in the WLS method, and an evaluation of AMI aggregation rates. The introduction of the first-order time-scaling coefficient was proven to effectively mitigate the impact of partial AMI aggregation, and the WLS-based method with higher weights assigned to nighttime data was confirmed to be effective in reducing daytime errors. The analysis of various AMI aggregation rates indicated that higher aggregation rates result in lower error rates. In addition, the proposed method demonstrated robustness to local anomalies in the AMI data. The proposed method is suitable when full AMI data collection is not feasible.

REFERENCES

[1] M. M. Haque and P. Wolfs, "A review of high PV penetrations in lv distribution networks: present status, impacts and mitigation measures," *Renew.Sustain.Energy Rev*, vol. 62, pp. 1195–1208, May. 2016.

[2] U.S. Energy Information Administration, *International Energy Outlook*. Washington, DC, USA: Government Printing Office, 2010.

[3] "What the duck curve tells us about managing a green grid," CAISO, Folsom, CA, USA, Rep., 2012.

[4] D. Li, J.-H. Bae, and M. Rishi, "A preference analysis for a peer-to-peer (P2P) electricity trading platform in South Korea," *Energies*, vol. 15, no. 21, p. 7973, Oct. 2022.

[5] B. Kroposki, B. Johnson, Y. Zhang, V. Gevorgian, P. Denholm, B.-M. Hodge, and B. Hannegan, "Achieving a 100% renewable grid: operating electric power systems with extremely high levels of variable renewable energy," *IEEE Power Energy Mag*, vol. 15, no. 2, pp. 61–73, Mar. 2017.

[6] R. Bayindir, S. Demirbas, E. Irmak, U. Cetinkaya, A. Ova, and M. Yesil, "Effects of renewable energy sources on the power system," in *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)*, pp. 388–393, IEEE, Nov. 2016.

[7] A. Dey, B. Chakraborty, S. Dalai, and K. Bhattacharya, "Insights and new practices for advanced metering infrastructure and smart energy metering framework in smart grid—a case study," in *Proc. IEEE Calcutta Conf. (CALCON)*, pp. 323–326, IEEE, Dec. 2022.

[8] OSIsoft, "New Services and Products Share Real-Time Distributed Data Across the Grid," 2017. [Online]. Available: https://www.ourenergypolicy.org/wp-content/uploads/2017/03/New_Services_and_Products_Share_Real_Time_Distributed_Data_Across_the_Grid_FINAL_4_.pdf

[9] National Grid, "2023 Distributed System Implementation Plan (DSIP) Update of Niagara Mohawk Power Corporation d/b/a National Grid," Jun. 2023. [Online]. Available: https://www.nationalgridus.com → DSIP Update 2023.

[10] N. Balakumar, L. Kristov, M. McDonnell, and M. Paterson, "Distribution System Operator (DSO) Initial Study," Nov. 2024. [Online]. Available: https://www.maine.gov/energy/sites/maine.gov.energy/files/meetings/DSO%20draft%20study%20results%20Nov%202024.pdf

[11] D. Chen and D. Irwin, "Sundance: black-box behind-the-meter solar disaggregation," in *Proc. 8th Int. Conf. Future Energy Syst*, pp. 45–55, May. 2017.

[12] 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.

[13] K. Pu and Y. Zhao, "An unsupervised similarity-based method for estimating behind-the-meter solar generation," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, pp. 1–5, IEEE, Jan. 2023.

[14] H. Shaker, H. Zareipour, 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.

[15] E. C. Kara, C. M. Roberts, M. 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, Mar. 2018.

[16] J. Lin, J. Ma, and J. Zhu, "A privacy-preserving federated learning method for probabilistic community-level behind-the-meter solar generation disaggregation," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 268–279, Jan. 2022.

[17] R. Saeedi, S. K. Sadanandan, A. K. Srivastava, K. L. Davies, and A. H. Gebremedhin, "An adaptive machine learning framework for behind-the-meter load/PV disaggregation," *IEEE Trans. Ind. Inform.*, vol. 17, no. 10, pp. 7060–7069, Oct. 2021.

[18] Z. Zhao, D. Moscovitz, S. Wang, X. Fan, and L. Du, "Semi-supervised disaggregation of load profiles at transmission buses with significant behind-the-meter solar generations," in *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 1–5, IEEE, Oct. 2022.

[19] K. Pan, Z. Chen, C. S. Lai, C. Xie, D. Wang, X. Li, Z. Zhao, N. Tong, and L. L. Lai, "An unsupervised data-driven approach for behind-the-meter photovoltaic power generation disaggregation," *Appl. Energy*, vol. 309, Mar. 2022, Art. no. 118450

[20] D. Moscovitz, Z. Zhao, L. Du, and X. Fan, "Semi-supervised, non-intrusive disaggregation of nodal load profiles with significant behind-the-meter solar generation," *IEEE Trans. Power Syst.*, vol. 39, no. 3, pp. 4852–4864, May. 2024.

[21] A. Mokaribolhassan, G. Nourbakhsh, G. Ledwich, A. Arefi, and M. Shafiei, "Distribution system state estimation using PV separation strategy in LV feeders with high levels of unmonitored PV generation," *IEEE Syst. J.*, vol. 17, no. 1, pp. 684–695, Mar. 2023.

[22] 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.
[23] N. Chatterjee, R. Glick, and B. McNamee, "Assessment of demand response and advanced metering," Technical report, Federal Energy Regulatory Commission, Dec. 2019. [Online]. Available: https://www.ferc.gov/sites/default/files/2020-04/DR-AM-Report2019_2.pdf
[24] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, Jul. 2018.
[25] B. Zhao, M. Ye, L. Stankovic, and V. Stankovic, "Non-intrusive load disaggregation solutions for very low-rate smart meter data," *Appl. Energy*, vol. 268, Jun. 2020, Art. no. 114949.
[26] S. Dash and N. C. Sahoo, "Electric energy disaggregation via non-intrusive load monitoring: a state-of-the-art systematic review," *Elect. Power Syst. Res.*, vol. 213, Aug. 2022, Art. no. 108673.
[27] K. Li, J. Yan, L. Hu, F. Wang, and N. Zhang, "Two-stage decoupled estimation approach of aggregated baseline load under high penetration of behind-the-meter PV system," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 4876–4885, Nov. 2021.
[28] T. Asparouhov and B. Muthén, "Weighted least squares estimation with missing data," *Mplus Tech. Append.*, vol. 2010, no. 1–10, p. 5, 2010.

DONGHO HYUN (Student Member, IEEE) received the B.S. degree in electronic engineering from the Kumoh National Institute of Technology, Gumi, South Korea, in 2024, where he is currently pursuing the M.S. degree in electronic engineering. His current research interests include power system analysis, data analysis, and learning methods.

JAEBEOM IM (Member, IEEE) received his Ph.D. degree in Electrical Engineering from Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 2024. Since 2024, he has been a postdoctoral researcher at POSTECH. His research interests include power system state estimation, distributed optimization, wide-area monitoring systems, and big data processing for future smart grids.

JAEPIL BAN (Member, IEEE) received the Ph.D. degree in electrical engineering from Pohang University of Science and Technology, Pohang, South Korea, in 2020. From August 2020 to August 2021, he was a postdoctoral researcher at Pohang University of Science and Technology. Since 2021, he has been an Assistant Professor with the Kumoh National Institute of Technology. His research interests include robust and optimal control, hybrid systems, power system control, applications of reinforcement learning, and power system state and line parameter estimation.

HYEONOK LIM (Member, IEEE) received the B.S. and M.S degree in electrical engineering from Jeonbuk National University, Jeonju, South Korea, in 2018. Since 2018, she has been a senior researcher at KEPCO Research Institute. her research interest is DER integration in distribution system.

JONGMIN JO (Member, IEEE) received the B.S., the M.S and Ph.D degree in electrical engineering from Chungnam National University, Daejeon, South Korea, in 2013, 2015 and 2020, respectively. Since 2020, he has been a senior researcher at KEPCO Research Institute. His research interests include DER control with GFL/GFM inverter, voltage regulation in distribution network.

SEONGSOO CHO (Member, IEEE) received the B.S. Degree in electrical engineering from Konkuk University, Seoul, South Korea, in 1994, and the M.S. and Ph.D degree from Chungnam National University, Daejoen, South Korea, in 2009. Since 1993, he has been as a Chief researcher at the KEPCO Research Institute. His research interests are interconnection of DER, active voltage control, and evaluation on control setting options of smart inverter in distribution system.

YOUNG-JIN KIM (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2007 and 2010, respectively, and the Ph.D. degree in electrical engineering from Massachusetts Institute of Technology, Cambridge, MA, USA, in 2015. From 2007 to 2011, he was with Korea Electric Power Corporation as a Power Transmission and a Distribution System Engineer. He was also a Visiting Scholar with Catalonia Institute for Energy Research, in 2014, and a Postdoctoral Researcher with the Center for Energy, Environmental, and Economic Systems Analysis, Energy Systems Division, Argonne National Laboratory, from 2015 to 2016. He joined the Faculty with Pohang University of Science and Technology, where he is currently an Associate Professor with the Department of Electrical Engineering. His research interests include distributed generators, renewable energy resources, and smart buildings.