IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 June 2024, accepted 5 July 2024, date of publication 9 July 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3425590

RESEARCH ARTICLE

A Deep Learning Framework for Net Load Forecasting With Unsupervised Behind-the-Meter Disaggregated Data

CHAICHAN THEPPROM, NATAWUT NUPAIROJ, (Member, IEEE), AND PEERAPON VATEEKUL, (Senior Member, IEEE)
Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
Corresponding author: Peerapon Vateekul (peerapon.v@chula.ac.th)

ABSTRACT Recently, distributed photovoltaic (PV) generation has increased significantly, leading to a high penetration of behind-the-meter (BTM) solar generation systems. In this work, we aim to improve net load forecasting by disaggregating BTM components to provide better representation. For the disaggregation process, we propose an unsupervised contrastive-based optimization method for estimating BTM PV generation from the net load at the aggregated level. Our proposed method uses a deep neural network to leverage the strong correlation between solar irradiance and PV generation. This means that our proposed method is independent of the availability of BTM data and the assumption of a physical model. Furthermore, to obtain the best forecasted trends on the disaggregated series (pure load and PV generation), various recent forecasting models have been compared i.e. DeepAR, Temporal Fusion Transformer (TFT), and Time-series Dense Encoder (TiDE). The experiment is conducted on two real-world electricity prosumption datasets collected from New York and Texas. Results show that the net load forecasting on the disaggregated series outperforms the net load series directly. Such an improvement is due to the accuracy of our unsupervised disaggregation of the BTM data, proving superior to the semi-supervised technique.

INDEX TERMS Behind-the-meter, net load, PV generation, net load disaggregation, unsupervised learning.

I. INTRODUCTION

Nowadays, distributed photovoltaic (PV) generation has a huge influence on electric power distribution. Such an impact is driven by the concern of the greenhouse effect that leads to global warming. Many countries are trying to reduce their dependence on fossil energy and are committed to a carbon neutrality goal [1]. Commercially, PV systems are seeking ways to reduce installation costs to achieve parity with fossil fuels. Therefore, a comprehensive cost analysis of feed-in tariffs and net metering is necessary [2], [3]. It is noted that distributed PV generation can effectively reduce greenhouse gas emissions, creating greater interest in PV generation [4]. Due to financial constraints, privacy concerns, and the net metering policy in certain countries, most PV systems are installed behind-the-meter (BTM). BTM makes PV generation invisible to utility operators.

Owing to the increase in PV penetration, households have installed smart meters. Smart meters are electricity consumption monitors and send real-time information about energy usage to energy suppliers. Power monitored on smart meters is aggregated load i.e. net load between the BTM pure load and PV generation. Hence, net load forecasting becomes a more challenging task because of the different nature of BTM components. Net load forecasting is important for power system operators to make decisions on power scheduling, demand response, and energy storage capacity planning. For grid stability and overgeneration as well as for efficient operation of the power system, the ability to accurately forecast net load is crucial. Power system operators must maintain grid stability and avoid overgeneration.

The associate editor coordinating the review of this manuscript and approving it for publication was Aysegul Ucar.

In previous works on net load forecasting, several techniques have been proposed to address the uncertainty of electricity demand and its components. These techniques include probabilistic forecasting [5], [6] to facilitate better decision-making, signal decomposition [7], [8] to extract hidden information from time series data, and net load disaggregation to estimate pure load and solar generation (also called PV generation) that lie behind-the-meter. In this work, we focus exclusively on enhancing net load forecasting through unsupervised disaggregation. For simplicity, most prior studies on net load forecasting have generally been categorized into two groups: the direct approach and the additive approach [9].

Firstly, the direct approach, also called "the integrated approach, is the traditional way of forecasting based on the net load series directly [10], [11], [12], [13]. Secondly, the additive approach is based on the addition of two forecasted series: BTM pure load and PV generation series [5], [14], [15]. The direct approach is a simple time series forecasting task but is unable to understand the impact of PV generation that penetrates the pure load due to the invisible BTM data. In contrast, for better representation, the additive approach requires more data and can capture the impact of BTM PV generation on the net load. Therefore, when PV penetration in the system is high, the additive approach is more suitable for net load forecasting.

For the additive approach, the forecasted result is based on a two-time series, including pure load and PV generation: both are components of the net load at BTM. Many previous works have been proposed to estimate BTM components by disaggregating them for net load series, such as supervised learning [5], semi-supervised learning [16], [17], [18], [19], [20], and unsupervised learning [21], [22]. However, in the real environment, obtaining BTM data to disaggregate the net load is quite impossible, making prior works impractical for real world applications. As such, the most practical solution is an unsupervised approach, which often relies on the assumption of a physical model of the PV system [23]. Thus, using a non-parametric model (deep learning network), the process of disaggregation can be further improved by reducing the requirement of technical variables in PV systems and meteorological data.

In this work, our primary objective is to forecast net load at a day-ahead level with a 15-minute interval using an additive approach. The study is divided into two main parts. First, we propose a novel unsupervised net load disaggregation method to identify pairs of timestamps with identical loads but different PV penetrations. This method, along with the contrastive loss function, enables us to estimate pure load and PV generation from the available net load data. Second, we compare modern time series forecasting models to evaluate the effectiveness of the proposed disaggregation method. Our results demonstrate that forecasting net load using the additive approach significantly improves accuracy compared to the direct approach due to net load disaggregation. We wish to expand upon previous works. The main contributions of this work are as follows:

- A contrastive-based BTM pure load proxy timestamp selection, based on the correlation between solar irradiance and PV generation as well as the correlation of electricity consumption across seasons, is proposed This method operates independently of the actual BTM data.
- An unsupervised long short-term memory (LSTM) network, utilizing meteorological data e.g. solar irradiance and temperature to estimate BTM PV generation from the net load at the aggregated level, is applied This method is independent of the physical model assumption.
- To evaluate the effectiveness of the proposed disaggregation method, various forecasting models, such as DeepAR [24], Temporal Fusion Transformer (TFT) [25], and Time-series Dense Encoder (TiDE) [26], are compared to evaluate the effectiveness of the proposed disaggregation method.
- To evaluate the performance of the proposed method, experiments on two real-world electricity datasets, conducted in New York and Texas, are referred to as the Pecan Street dataset [27]. Results show that net load disaggregation outperforms baseline methods, significantly improving net load forecasting accuracy.

To facilitate the understanding of the paper, we provide a list of nomenclature in Appendix VII. The rest of this paper is organized as follows: Section II discusses the related works. Section III declares the problem statement. Section IV explains the proposed method. Section V presents the experiment. Section VI shows the results. Finally, Section VII concludes the work.

II. RELATED WORK

In this section, we will give a brief overview of the relevant previous work in the fields of net load forecasting and disaggregation.

A. PREVIOUS NET LOAD FORECASTING WORKS

In Table 1, prior forecasting works comparing direct and additive approaches are shown. Kaur et al. [14] conducted a comparative study of net load forecasting methods for the operation and management of microgrids with high PV penetrations. It is seen that the direct model outperformed the additive model [14]. Garcia-Garrido et al. [28] compared the performance of direct and additive net load forecasting approaches using autoregressive exogenous (ARX) and multi-layer perceptron (MLP). The study found that MLP outperformed ARX for both net load forecasting approaches [28]. Falces et al. [29] examined the performance of direct and additive net load forecasting approaches, using random forest, extreme gradient boosting, and support vector regression models: the additive approach provided better forecasting performance compared to the direct approach [29]. Stratman et al. [15] proposed a two-stage net load forecasting framework and a compensator for

correcting the error of the net load forecast, revealing that the proposed additive method reduced the forecasting error compared to the direct method [15]. Jia et al. [30] introduced BTM component predictors for PV generation and pure load forecasting with the help of a net load disaggregation technique. Forecasting BTM components separately and combining them outperformed direct net load forecasting, even using net load and weather forecast data as input features [30].

TABLE 1. Previous net load forecasting works comparing the direct and additive approaches.

Reference	Year	Resolution	Horizon	Winner approach	
				Direct	Additive
[14]	2016	Hourly	Short-term	X	
[28]	2020	Hourly	Day-ahead		X
[29]	2023	Hourly	Day-ahead		X
[15]	2023	Hourly, 15 min	Hour-ahead		X
[30]	2023	15 min	Day-ahead		X

B. PREVIOUS BTM ESTIMATION WORKS

Herein, we discuss previous works using the model-based approach. Chen and Irwin [21] introduced a method, combining a clear sky generation model and percent reduction from the maximum PV generation to create an equivalent PV system of target building. However, dependency on the geographical location of the PV system made the method inaccurate, since the geographical location was different from the actual location of the PV system [21]. Wang et al. [5] set up an equivalent PV system variable using maximal information coefficient-based correlation analysis, under the assumption of a PV physical model and known BTM data. Due to the lack of data from every PV site during a specific time frame, the method proved impractical in the BTM situation [5]. Pu and Zhao implemented a method that exploited self-similarity and cross-customer similarity of electricity consumption behavior to estimate aggregated-level BTM PV generation. Although the method was independent of the availability of BTM PV generation data, it required the assumption of a physical model of the PV system and the availability of consumption data from a group of customers who had not installed PV systems [16].

Currently, the new trend in research is a ''data-driven approach''. Shaker et al. [17] introduced a data-driven method to estimate BTM PV generation using dimension reduction and mapping functions, where a small number of fully observable sites and their BTM PV and capacity are known [17]. Cheung et al. [18] utilized neighboring households, which do not have a PV system installed, as a consumption proxy, to estimate the BTM PV generation of PV customers. This method assumes that the consumption behavior of a consumer is a mixture of latent behaviors and utilizes the night-time load of mixture proxy customers for optimizing a linear function to estimate BTM PV generation of PV customers [18]. Li et al. [19] employed a two-stage net load disaggregation approach to estimate BTM PV generation, using a small number of observable BTM PV generation and pure load by solving the optimization problem via the Lagrange multiplier [19]. Bu et al. [20] addressed a maximum likelihood estimation-based BTM PV generation estimator, using solar exemplars and representative profiles, which applied the correlation between monthly nocturnal and diurnal pure load [20].

These reviewed papers show that data-driven approaches needed to leverage the customer without a PV load profile or a small number of observable BTM data points to create a representative profile for estimating BTM PV generation. Pan et al. [22] introduced PV sensitivity and pure load sensitivity estimation models based on the assumption of the nearest neighbor consumption behavior at night-time and the similarity of solar irradiation, respectively. The method applying deep learning techniques and clustering of consumption behavior could disaggregate BTM PV tasks, using only net-load data and mainstream meteorological variables as exogenous variables [22].

In this work, we aim to conduct net load forecasting using an additive approach by disaggregating pure load and PV generation based on an unsupervised approach with non-parametric models (deep learning networks). To reduce dependence on BTM data of customers without PV, we adopt the assumption of self-similarity, as stated in [16].

III. PROBLEM STATEMENT

Due to the fact that net load measured by a smart meter is the electricity demand after BTM PV generation has been subtracted, the relationship between net load and BTM data can be written, as follows:

$$n_{c,t} = l_{c,t} - p_{c,t} \tag{1}$$

where $n_{c,t}$ is net load, $l_{c,t}$ is BTM pure load, and $p_{c,t}$ is BTM PV generation of customer c at time t, respectively. In this work, we aim to estimate BTM PV generation $p_{c,t}$ from net load n_t and BTM pure load l_t at aggregate level. Aggregate level refers to the group of customers who have installed PV systems, denoted as S in the same area at time t and can be written as:

$$n_t = \sum_{c \in S} n_{c,t} \tag{2}$$

where both BTM pure load (l_t) and BTM PV generation (p_t) have identical relationships. If we know BTM PV generation, we can find BTM pure load using the following equation:

$$l_t = n_t + p_t \tag{3}$$

After both components of the net load have been estimated, the net load forecasting task is conducted using the disaggregated series. The net load forecasting task is to predict the net load for the next day ahead with a 15 min interval (96-time steps). To evaluate the effectiveness

TABLE 2. Previous BTM disaggregation works.

Reference	Year	Dependence on BTM labels (Unsupervised)	Dependence on physical model	Disaggregation technique
[21]	2017	Unsupervised	Model-based	Maximum clear sky model and multi-step optimization
[5]	2017	Supervised	Model-based	Maximal information coefficient-based correlation analysis and multi-step optimization
[16]	2023	Semi-supervised	Model-based	Multi-step optimization
[17]	2015	Semi-supervised	Model-free	Principal component analysis and multiple of mapping techniques, i.e. linear regression, Kalman filter, multi-layer perceptron (MLP), and wavelet neural network
[18]	2018	Semi-supervised	Model-free	Consumer mixture model optimization
[19]	2021	Semi-supervised	Model-free	Multi-step optimization
[20]	2021	Semi-supervised	Model-free	Probability density function (PDF) fitting
[22]	2022	Unsupervised	Model-free	Nearest behavior and multi-step optimization

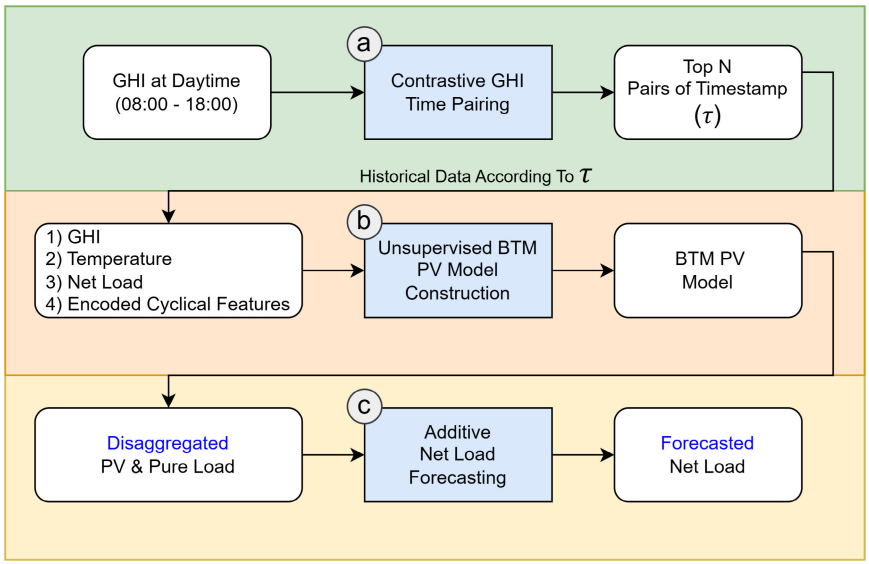

FIGURE 1. The overall framework of net load forecasting with information disaggregation. The proposed framework consists of three main steps: a) Time pairings selection based on contrastive GHI, b) Constructing the BTM PV generation estimation model in an unsupervised manner for net load disaggregation, and c) Net load forecasting using disaggregated BTM components.

of the proposed disaggregation method, the forecasting performance between the direct and additive approaches is compared.

IV. METHODOLOGY

This section describes the proposed net load disaggregation method for improving net load forecasting with invisible real BTM. In Fig. 1, the overall framework of the proposed method is shown. Our proposed method has three main steps. Firstly, contrastive-based timestamp pairings use a timestamp at daytime with low global horizontal irradiance (GHI) as a proxy of BTM pure load, using a score-based approach that considers the correlation between PV generation and GHI; selected timestamp pairs τ are used to optimize the mapping function. Secondly, unsupervised optimization uses the LSTM network to minimize the difference between the BTM pure load proxy and the estimated pure load (the difference between the net load and estimated BTM PV). This task is carried out to estimate BTM PV generation without the need for complex meteorological data, geographic information, or accurate physical characteristics of PV arrays. Finally, additive net load forecasting uses the disaggregated series to forecast the BTM pure load and PV generation separately. Then, both are combined to obtain the net load forecast. Details of each step are described in the following subsections.

Due to the absence of a BTM pure load and BTM PV generation data from fully observable customers or partially observable customers, timestamp pairs are selected for exploiting the BTM pure load from the net load data, as a proxy. Since PV penetration influences net load, the greater the PV generation, the greater the discrepancy between the net load and the pure load, and vice versa. In Fig. 2, the relationship between PV generation and pure load is shown. On sunny days, PV generation is high (according to GHI). Hence, the net load decreases and diverges from the pure load.

On cloudy days, PV generation is low such that the net load is closer to the pure load.

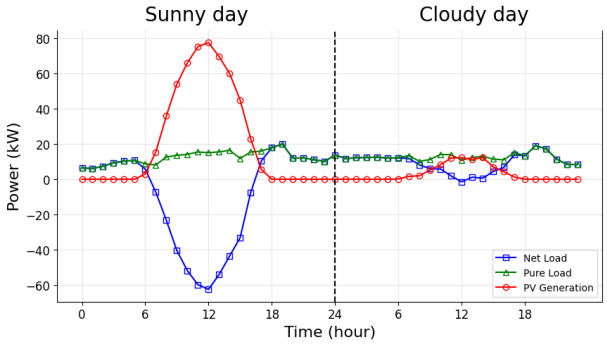

FIGURE 2. Comparison of net load and its component between sunny day and cloudy day.

A. CONTRASTIVE GHI TIME PAIRING

Further, selection criteria in terms of a score-based approach that considers GHI during the daytime (from 8:00 to 16:00) are proposed. When GHI is low during a cloudy period, contrastive pairing is used to select the timepoint, as a proxy of BTM pure load. When GHI is high during a sunny period, the timestamp to estimate the BTM PV generation is compared. Contrastive pairing criteria can be formulated in accordance with Eqs. (4)-(8). The total score in Eq. (8) is a summation of scores that focuses on two main aspects:

1) The similarity of consumption behavior across time, including Eqs. (4) and (5) considers the difference of the min of the day, and the difference of the day of the year, respectively.
2) The discrepancy of GHI between the two timepoints, including Eqs. (6) and (7) considers the difference between maximum GHI and GHI of t_2; the other one is the difference between GHI between t_1 and t_2, respectively.

The intention of these scores is to overcome the challenge of the invisible BTM data by assuming that each pair of timepoints has different BTM PV generation (GHI is high at t_1 and low at t_2) but similar BTM pure load.

$$S1 = \frac{1440 - |MOD_{t_1} - MOD_{t_2}|}{1440} \tag{4}$$

$$S2 = \frac{365 - |DOY_{t_1} - DOY_{t_2}|}{365} \tag{5}$$

$$S3 = \frac{|GHI_{max} - GHI_{t_2}|}{GHI_{max}} \tag{6}$$

$$S4 = \frac{|GHI_{t_1} - GHI_{t_2}|}{GHI_{t_1}} \tag{7}$$

$$Score(t_1, t_2) = S1 + S2 + S3 + S4 \tag{8}$$

where MOD_{t_i} is min of day, DOY_{t_i} is day of year at time t_i, respectively. GHI_{max} is maximum GHI.

To avoid the dissimilarity of consumption behavior when selecting timepoints that are too far apart, we set additional constraints: the difference of min of day and day of year between t_1 and t_2 should be less than 120 min (2h) and 14 days, respectively. In Fig. 3, an example of contrastive GHI pairings, where the search origin (t_1) at 12/05/2018 12:00 noon and the possible pair members (t_2) on the time window of 2 h and 14 days, is shown.

As for the pair of timepoints, the first timepoint (t_1) is represented by the daytime timepoint when GHI is high. The second timepoint (t_2) is represented by the daytime timepoint when GHI is low. Such timepoints are assumed to be proxies of BTM pure load due to the negative correlation between GHI and net load. Finally, the set of timepoints with the top N highest score is denoted as $\tau = \left[\left(t_{1,1}, t_{1,2}\right), \left(t_{2,1}, t_{2,2}\right), \left(t_{3,1}, t_{3,2}\right), \cdots, \left(t_{N,1}, t_{N,2}\right)\right]$ being selected for the unsupervised optimization step.

B. UNSUPERVISED BTM PV GENERATION MODEL CONSTRUCTION

In this step, we aim to construct an unsupervised LSTM network to estimate BTM PV generation using τ, which is the set of time pairs selected from the previous step. Due to the absence of BTM PV generation data, it is not practical to estimate BTM PV generation through supervised or semi-supervised learning. However, the unsupervised optimization method is more suitable for estimating BTM PV generation without the need for real BTM data. To achieve this, considering the strong correlation between GHI and PV generation, a deep neural network is used to convert GHI and covariates into the estimated BTM PV generation. Hence, to estimate the BTM PV generation, considering the strong correlation between GHI and PV generation. we use a deep neural network.

As discussed in Section III and according to Eq. (3), BTM pure load can be estimated when BTM PV generation and net load are known. Moreover, the assumption of time pair is that BTM PV generation is different: GHI is high at t_1 and low at t_2 but very close to the BTM pure load. Therefore, to minimize the difference between the BTM pure load estimated from each member in the time pair (t_1, t_2), the problem as an unsupervised optimization problem can be formulated, as follows:

$$\min_{f(x)} \sum_{(t_1,t_2)\in\tau} \left[\left(n_{t_1} + f(x_{t_1})\right) - \left(n_{t_2} + f(x_{t_2})\right) \right]^2 \tag{9}$$

$$\text{s.t.} f(x) \geq 0 \tag{10}$$

where x_t represents a group of input features at time t, including GHI, temperature, net load, and encoded cyclical features (i.e., minute of the day and month of the year). The output of the mapping function, $f(x)$, is formed by the deep neural network (LSTM). LSTM is used to estimate BTM PV generation. All pairs of timestamps in τ undergo optimization, which iteratively updates the parameters of the mapping function to minimize the custom loss function.

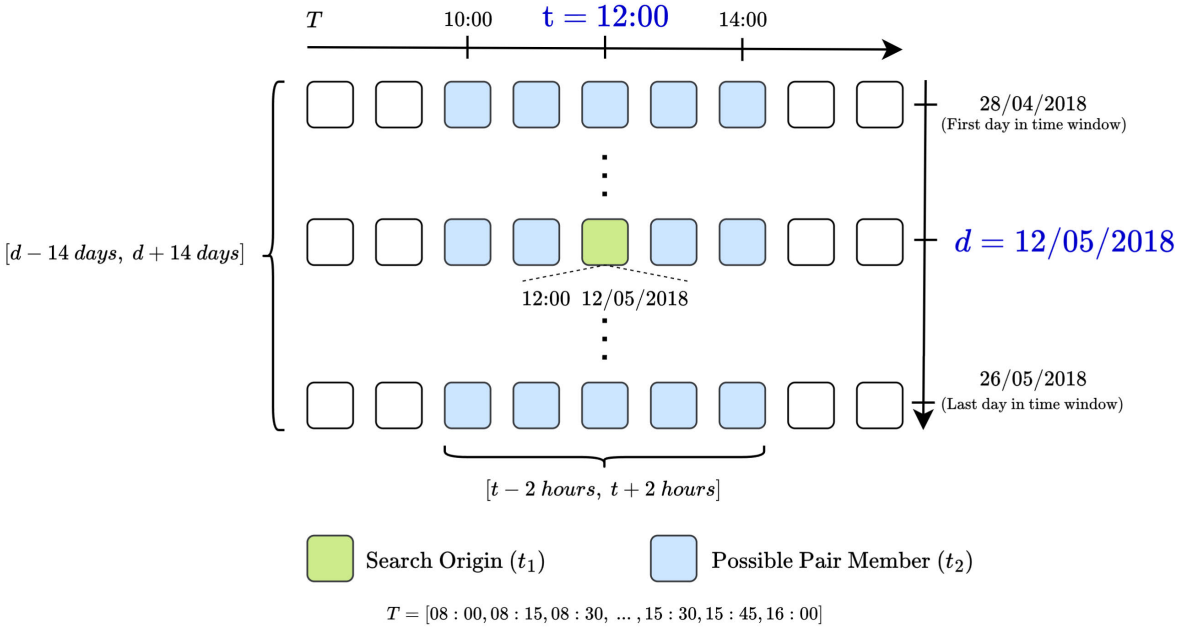

FIGURE 3. Schema of time pairings at 12:00 noon on 12/05/2018 based on a time window of 2 h and 14 days.

LSTM is a variant of recurrent neural network (RNN) that can capture long-term dependencies in time series data [31]. Due to the vanishing gradient problem and the exploding gradient problem in RNN, LSTM is designed to solve these problems, using a gating mechanism i.e. an input gate, a forget gate, and an output gate. To control the flow of information in the network, a memory cell is also utilized. Further, the LSTM network is used in this work to disaggregate BTM PV generation from the net load data and the selected BTM pure load proxy.

In Algorithm 1, we summarize the proposed net load disaggregation method. Algorithm 1 consists of two main steps: contrastive-based timestamp pairings, and unsupervised optimization. During daytime from 8:00 to 16:00, between sunny and cloudy timepoints, timepoint pairs attaining top N criteria scores are searched as the proxy of BTM pure load. Then, the selected pair of timepoints are represented by τ. Subsequently, unsupervised optimization is applied. Next, the LSTM network is used to estimate BTM PV generation from the net load data and the selected BTM pure load proxy. Finally, the estimated BTM PV generation is used to estimate the BTM pure load, in accordance with Eq. (3).

Algorithm 1 Contrastive-Based BTM Solar Disaggregation

Input : Net load and GHI
procedure Net Load disaggregation module
  Select a set of timestamp pairs (τ) using contrastive-based scoring
  Initialize the LSTM network
  for epoch = 1 to max epoch **do**
    Initiaize $Total\_Loss = 0$
    for $(t_1, t_2) \in \tau$ **do**
      $Loss$ = Eqs. (9) and (10)
      $Total\_Loss$ += $Loss$
    end for
    Update the LSTM network parameters via backpropagation
  end for
end procedure
procedure BTM PV generation estimation module
  for $t \in$ Dataset **do**
    Estimate the BTM PV generation by inference the optimized LSTM network
  end for
end procedure
procedure BTM pure load extraction module
  Obtain the BTM pure load as in (3)
end procedure
Output : Estimated BTM PV generation and BTM pure load

C. ADDITIVE NET LOAD FORECASTING

After both BTM components have been estimated, we will conduct the net load forecasting task using the disaggregated series, according to the additive approach. BTM pure load and PV generation are forecasted separately. Then, the components are combined to obtain the net load forecast as the final result. Since we assume that direct measurements of BTM PV generation and pure load are unavailable, the disaggregated BTM PV generation and pure load are used as the ground truth values for training the additive forecasting models. In Fig. 4, the overall process of the direct and additive

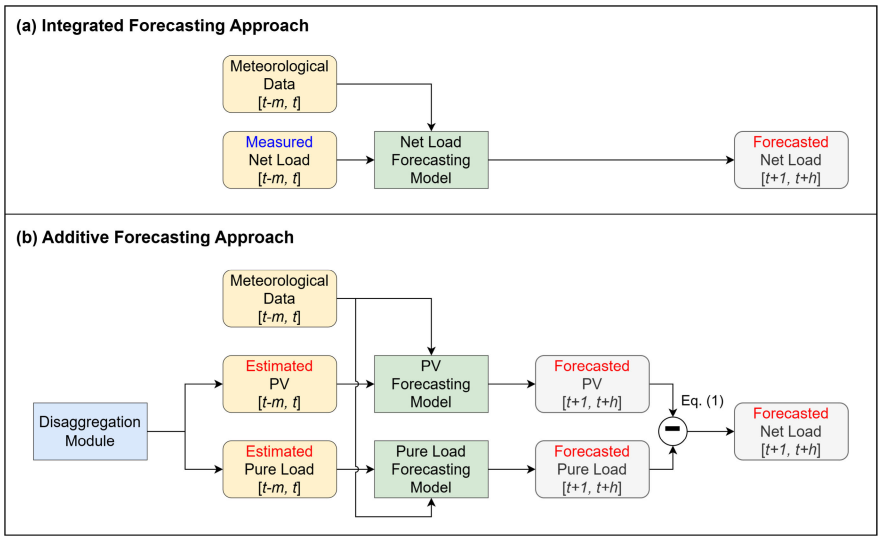

FIGURE 4. Schematic illustration of net load forecasting experiment scenarios between direct and additive approaches. The direct approach uses the net load as input as well as ground truth for the forecasting model. In contrast, the additive approach uses the estimated BTM components separately as the input and ground truth for the forecasting model. The range $[t-m, t]$ indicates the lookback period of the forecasting model. The range $[t+1, t+h]$ indicates the forecast horizon of the forecasting model, where t is the current time point, m is the number of time points in the lookback period, and h is the number of time points in the forecast horizon. In this work, we aim to forecast the net load for the next 96 time points i.e. day-ahead forecast.

net load forecasting approaches is illustrated. To compare the forecasting performance between the direct and additive approaches, we use various recent forecasting models such as DeepAR,TFT, and TiDE. The details of the forecasting models are described in the following subsections.

1) DEEPAR

DeepAR is a probabilistic forecasting model based on RNN that works like a likelihood-based model [24]. DeepAR learns the distribution of future values given past observations. Due to the probabilistic nature of DeepAR, it can predict the interval for capturing the uncertainty of the forecasted values. Furthermore, to improve the accuracy of forecasts, DeepAR can utilize a sequence of past values and additional features, such as categorical variables and other time dependent variables.

2) TEMPORAL FUSION TRANSFORMER (TFT)

TFT is a transformer-based deep learning model that is designed to handle multiple time series with different frequencies and irregular time intervals [25]. TFT uses a combination of attention mechanism and gating mechanism to capture long-term dependencies in time series data. Due to the capability of TFT to handle multivariate inputs, it can capture static (time-invariant) and dynamic (time-variant) features in time series data, providing interpretable features that are important for forecasted values. These findings are useful for understanding the impact of each feature on the forecasted values, making TFT suitable for forecasting time series data with multiple features.

3) TIME-SERIES DENSE ENCODER (TIDE)

TiDE is a time-series model that has been shown to outperform or be on par with state-of-the-art transformer models in long sequence time-series forecasting [26]. TiDE is able to handle covariates both from the past as well as known or unknown future. The model is constructed using a dense encoder and decoder architecture, with a temporal decoder to generate predictions. Despite its simplicity and absence of attention mechanisms, the model can capture long-range dependencies and outperforms other models in terms of forecasting accuracy and computational efficiency.

V. EXPERIMENTAL SETUP

In this section, we describe the setup of our experiment. Our experimental setup is divided into four parts: 1) experimental data, 2) experiments for BTM PV estimation, 3) experiments for net load forecasting, and 4) evaluation metrics.

A. EXPERIMENTAL DATA

To verify the effectiveness of the proposed method, we conducted experiments on two real-world datasets from Pecan Street Inc. The datasets contain the electricity consumption and BTM PV generation collected from 1) Ithaca, NY, and 2) Austin, TX. Both datasets were collected at 15 min intervals. The first dataset (Ithaca, NY) aggregated the electricity consumption and BTM PV generation of 14 homes with PV. The dataset spanned 6 months from 2019/5/1 to 2019/10/31. The second dataset (Austin, TX) comprised the electricity consumption and BTM PV generation of 19 homes

with PV systems. The dataset spanned 12 months from 2018/1/1 to 2018/12/31 [27].

Meteorological data were obtained from the National Solar Radiation Database (NSRDB) [32]. For our case study locations, NSRDB provided both solar radiation and comprehensive meteorological data. In our experiment, we used 4×4 km grid resolution at 30 min intervals, interpolated to 15 min intervals, applying linear interpolation to match the Pecan Street dataset. Both datasets were used to train and evaluate the proposed net load disaggregation method as well as to forecast net loads for both locations. Please note that we focused on the aggregated level. Moreover, we considered a time series for all households altogether.

B. EXPERIMENTS FOR BTM PV ESTIMATION

Herein, the proposed net load disaggregation method employed 1) the net load and 2) the meteorological data from NSRDB. Since we assume that BTM PV generation is invisible (unsupervised), there is no use of a measured pure load and PV generation for training and testing in this step. We only use BTM PV generation data (ground truth) to verify the accuracy of the estimated BTM PV generation. To evaluate the proposed method, we compare the results of the proposed method with the physical model and real BTM pure load depending on the optimization [16]. More specifically, the hyperparameters of the proposed LSTM model, determined by grid searching and tuning, are displayed in Appendix VII. Note that the number of timestamp pairs selected for estimating the BTM PV generation in this work are searched based on grid search. In Ithaca, NY, we used 5000 timestamp pairs, and in Austin, TX, we used 7500 timestamp pairs.

C. EXPERIMENTS FOR NET LOAD FORECASTING

In this section, we aim to evaluate the accuracy of our additive net load forecasting that is a summation of two forecasted BTM series: namely, 1) BTM pure load and 2) BTM PV generation. The dataset was split into three parts: training, validation, and testing set.

- In Ithaca, NY, the first 4 months (2019/5/1 to 2019/8/31) were used for training. The next month (2019/9/1 to 2019/9/30) was used for validation. The last month (2019/10/1 to 2019/10/31) was used for testing.
- In Austin, TX, the first 10 months (2018/1/1 to 2018/10/31) were used for training. The next month (2018/11/1 to 2018/11/30) was used for validation. The last month (2018/12/1 to 2018/12/31) was used for testing.

To forecast the net load for the day ahead with a 15 min interval, we used DeepAR, TFT, and TiDE models. To evaluate the effectiveness of the proposed disaggregation method, the results of the net load forecasting between the direct and additive approaches are compared.

In Table 3, the input variables used for forecasting are shown. The weather variables, such as GHI and temperature, were selected based on their correlation with the target. Cyclical features, including hour, day of the week, and month, were used to help the models learn daily and weekly patterns. These cyclical features were encoded, using cyclical index transformation via sine and cosine functions. All input variables were normalized, using min-max normalization.

TABLE 3. Input variables used for forecasting.

Model	Input variables
Net load forecast	Net load, temperature, GHI, hour, day of weak, and month
PV generation forecast	Estimated PV generation, temperature, GHI, hour, and month
Pure load forecast	Estimated pure load, temperature, hour, day of weak, and month

All forecasting models were configured to predict the net load for the day ahead with a 15 min lead time (next 96 steps) and a three-day lookback period (288 steps). In Appendix VII, the hyperparameters for DeepAR, TFT, and TiDE models are given. These forecasting models are implemented, using the Darts library [33] in Python 3.8.18.

D. EVALUATION METRICS

For evaluating the performance of net load disaggregation and net load forecasting, we use the root mean square error (RMSE), mean absolute error (MAE), and symmetric mean absolute percentage error (sMAPE). To show the percentage error between 0% and 100% for interpretability, sMAPE is adjusted. The reason for using sMAPE instead of mean absolute percentage error (MAPE) is to avoid the problem of generating infinite or undefined values when the actual value of BTM PV generation is zero or close to zero. The equations for RMSE, MAE, and sMAPE are shown as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{11}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{12}$$

$$\text{sMAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)} \times 100\% \tag{13}$$

where N is the total number of samples, i is the index of each value, y_i is the ground truth value, and $\hat{y}$ is the predicted value.

It is noted that previous studies on net load disaggregation show that percentage evaluation metrics are highly sensitive when low BTM PV generation occurs at dusk and dawn [21], [34]. To avoid this problem, by referring to [22] and [34], we will only evaluate the performance of BTM PV generation estimation results in sMAPE during daytime from 8:00 to 16:00, when solar irradiance trends are high. In contrast, the net load forecasting performance will be evaluated for the whole day.

TABLE 4. Performance of BTM PV generation estimation in Ithaca, NY dataset. Boldface refers to the winner.

Method	RMSE (kW) ↓	MAE (kW) ↓	sMAPE (%) ↓
Baseline [16]	4.06	2.22	8.64
Proposed GHI time pair with physical model	5.92	3.33	12.87
Proposed GHI time pair with LSTM	**2.72**	**1.56**	**7.49**

TABLE 5. Performance of BTM PV generation estimation in Austin, TX dataset. Boldface refers to the winner.

Method	RMSE (kW) ↓	MAE (kW) ↓	sMAPE (%) ↓
Baseline [16]	9.78	5.52	21.77
Proposed GHI time pair with physical model	9.95	5.64	23.16
Proposed GHI time pair with LSTM	**7.06**	**4.03**	**15.90**

VI. EXPERIMENTAL RESULTS

In this section, we present the results of the net load disaggregation and net load forecasting experiments on two datasets in Ithaca, NY and Austin, TX. There are two main parts: BTM PV estimation results and net load forecasting results.

A. BTM PV ESTIMATION RESULTS

In this subsection, we compare the PV estimation results of our unsupervised model and the baseline model [16], which is a semi-supervised approach and based on a physical model [23]. Our method is further improved by replacing the physical model with a deep learning model (LSTM).

In Tables 4 and 5, the disaggregation results for Ithaca, NY, and Austin, TX are shown. Results indicate that the proposed method with LSTM is the winner. Hence, the estimated BTM PV generation is closer to the real BTM PV value, with RMSE and MAE of 2.72 kW and 1.56 kW for Ithaca, NY, and 7.06 kW and 4.03 kW for Austin, TX, respectively. Moreover, the sMAPE of the estimated BTM PV generation at daytime (from 8:00 to 16:00) proved to be 7.49% for Ithaca, NY, and 15.90% for Austin, TX, being lower than the baseline method. Although the proposed method is an unsupervised approach, it is still superior to the baseline. In addition, performance can really be improved by replacing the physical model with LSTM since it is a non-parametric approach and requires fewer assumptions (parameters), including meteorological data, geographic information, and accurate physical characteristics of PV arrays.

In Fig. 5, sample results of BTM PV generation estimation in one week, 2019/10/21 - 2019/10/28 for Ithaca, NY, and 2018/10/05 - 2018/10/12 for Austin, TX, are shown. It is seen that the proposed method (green line) has a shape and amplitude that is closer to the real data (blue line) compared to the baseline in both datasets.

In Fig. 6, the hourly absolute error of the BTM PV generation disaggregation is shown for both datasets. Results of the proposed method (green box) are compared with the baseline method (grey box). In Ithaca, NY, the proposed method has a lower median and interquartile range of the absolute error compared to the baseline method, indicating

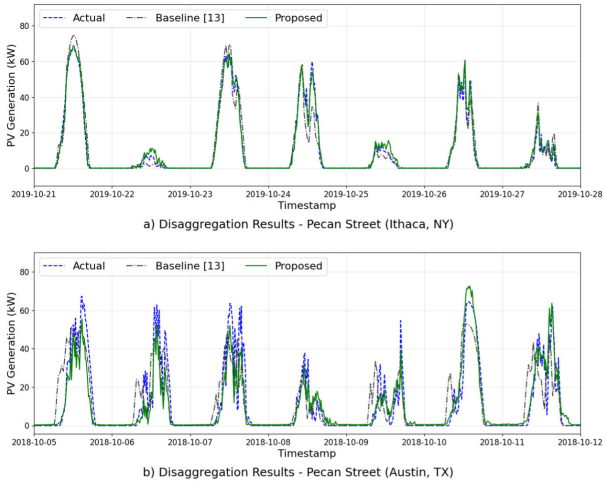

FIGURE 5. Disaggregation results showing the comparison of different methods between our method (green line) and the baseline [16] (grey line), where the actual line (ground truth) refers to the blue one.

that the proposed method is more stable and robust as well as accurate in estimating BTM PV generation across different time periods of the day. In Austin, TX, the proposed method tends to have a higher median and interquartile range of absolute error compared to the baseline method in the afternoon period. Such an outcome is due to the dissimilarity in consumption behavior.

For more details of our disaggregation results, we focus on the selected time pairs based on Eqs. (4)-(8) wherein each pair should have a similar pure load and different PV generation. Tables 6-8 illustrate the selected time pairs in Ithaca, NY during morning, noon, and afternoon, respectively. Results demonstrate that each time pair really follows the assumption with similar pure load ($|\Delta l_t(kW)|$) and different PV generation ($|\Delta p_t(kW)|$). For example, the first pair in Table 6 is at 2019-07-25 08:45:00 and 2019-07-22 08:00:00. The difference in their pure load is very low (only 0.22 kW) while the difference in PV generation is high (44.41 kW).

In conclusion, we show that our disaggregation method is the winner. Hence, net load time series is disaggregated into

TABLE 6. Sample of time pair (t_1, t_2) in morning period.

t_1	t_2	$l_{t_1}(kW)$	$l_{t_2}(kW)$	$\lvert\Delta l_t(kW)\rvert \downarrow$	$p_{t_1}(kW)$	$p_{t_2}(kW)$	$\lvert\Delta p_t(kW)\rvert \uparrow$
2019-07-25 08:45:00	2019-07-22 08:00:00	11.84	12.06	0.22	46.768	2.354	44.41
2019-07-13 09:15:00	2019-07-22 08:15:00	11.85	11.86	0.01	46.220	2.751	43.47
2019-08-25 09:30:00	2019-09-04 10:00:00	8.14	11.84	3.70	45.601	9.262	36.34
2019-08-05 10:00:00	2019-07-22 08:00:00	13.10	12.06	1.04	66.431	2.354	64.08
2019-10-05 10:45:00	2019-10-04 09:00:00	15.20	15.20	0.00	70.685	2.279	68.41

TABLE 7. Sample of time pair (t_1, t_2) in noon period.

t_1	t_2	$l_{t_1}(kW)$	$l_{t_2}(kW)$	$\lvert\Delta l_t(kW)\rvert \downarrow$	$p_{t_1}(kW)$	$p_{t_2}(kW)$	$\lvert\Delta p_t(kW)\rvert \uparrow$
2019-10-21 12:00:00	2019-10-31 11:30:00	14.60	14.89	0.29	67.44	5.61	61.83
2019-10-13 12:00:00	2019-10-06 10:00:00	17.54	21.03	3.49	74.40	5.54	68.86
2019-09-17 12:45:00	2019-09-26 11:00:00	15.91	14.63	1.28	75.52	3.48	72.04
2019-08-05 12:45:00	2019-08-08 14:00:00	11.76	14.17	2.41	79.15	0.25	78.90
2019-10-01 12:30:00	2019-10-06 10:00:00	24.48	21.03	3.45	67.50	5.54	61.96

TABLE 8. Sample of time pair (t_1, t_2) in afternoon period.

t_1	t_2	$l_{t_1}(kW)$	$l_{t_2}(kW)$	$\lvert\Delta l_t(kW)\rvert \downarrow$	$p_{t_1}(kW)$	$p_{t_2}(kW)$	$\lvert\Delta p_t(kW)\rvert \uparrow$
2019-08-05 13:30:00	2019-08-18 15:45:00	13.47	13.55	0.08	75.73	3.98	71.75
2019-08-19 14:45:00	2019-08-18 15:00:00	14.33	15.51	1.18	50.06	1.54	48.52
2019-09-19 13:15:00	2019-10-03 15:15:00	15.38	13.75	1.63	73.67	1.77	71.90
2019-10-11 13:15:00	2019-10-17 14:15:00	20.40	20.97	0.57	65.75	3.53	62.22
2019-09-20 13:45:00	2019-10-03 15:30:00	12.17	13.75	1.58	65.83	1.90	63.93

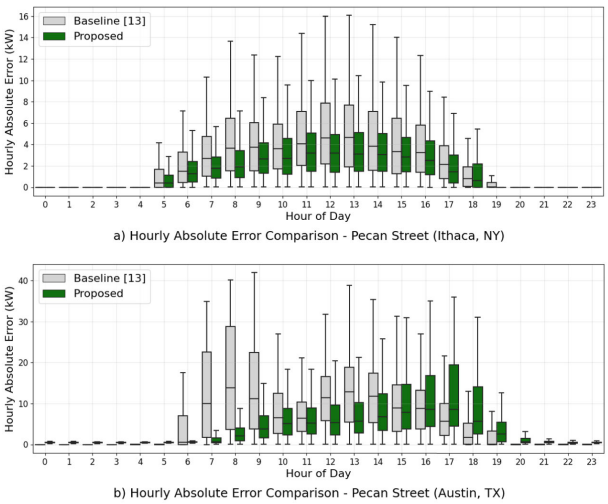

FIGURE 6. Hourly absolute error for each hour comparison of PV disaggregation between baseline and proposed method.

two time series: 1) BTM PV generation and 2) BTM pure load. In the next section, these extracted time series will be utilized for net load forecasting.

B. NET LOAD FORECASTING RESULTS

In this section, we use the results of the day-ahead net load forecasting on net load and estimated BTM data from both Ithaca, NY, and Austin, TX, datasets using three modern time series forecasting models. Such models can handle long sequence time-series forecasting tasks, including DeepAR, TFT, and TiDE. The forecast results of direct and additive approaches are compared. In Tables 9 and 10, the evaluation results show that the additive approach can outperform the direct approach in terms of RMSE, MAE, and sMAPE for both datasets and all models. The additive approach can lower RMSE, MAE, and sMAPE on average for all models by 8%, 15%, and 6% in Ithaca, NY, and 13%, 16%, and 3% in Austin, TX, respectively. Of the three forecasting models, the TiDE model shows the best performance overall. Results demonstrate that the proposed net load disaggregation method can improve net load forecasting performance.

TABLE 9. Forecast evaluation results for Ithaca, NY. There are three forecasting techniques comparing between direct and additive approaches. The number in parentheses refer to a difference between additive and direct approaches as an improvement. Boldface refers to the winner.

Method	RMSE (kW) ↓	MAE (kW) ↓	sMAPE (%) ↓
DeepAR - Direct	16.39	11.65	42.86
DeepAR - Additive	**14.91 (8%)**	10.43 (10%)	38.51 (4%)
TFT - Direct	20.76	14.50	47.55
TFT - Additive	17.55 (15%)	11.64 (19%)	38.47 (9%)
TiDE - Direct	15.75	11.75	41.11
TiDE – Additive	15.35 (2%)	**9.82 (16%)**	**34.78 (6%)**
%Average improvement	**8%**	**15%**	**6%**

TABLE 10. Forecast evaluation results for Austin, TX. There are three forecasting techniques comparing between direct and additive approaches. The number in parentheses refer to a difference between additive and direct approaches as an improvement. Boldface refers to the winner.

Method	RMSE (kW) ↓	MAE (kW) ↓	sMAPE (%) ↓
DeepAR - Direct	18.48	12.75	36.81
DeepAR - Additive	14.00 (24%)	9.43 (26%)	31.57 (5%)
TFT - Direct	17.53	11.19	33.41
TFT - Additive	15.56 (11%)	9.89 (11%)	32.23 (1%)
TiDE - Direct	13.24	9.43	32.85
TiDE – Additive	**12.60 (4%)**	**8.35 (11%)**	**29.02 (4%)**
%Average improvement	**13%**	**16%**	**3%**

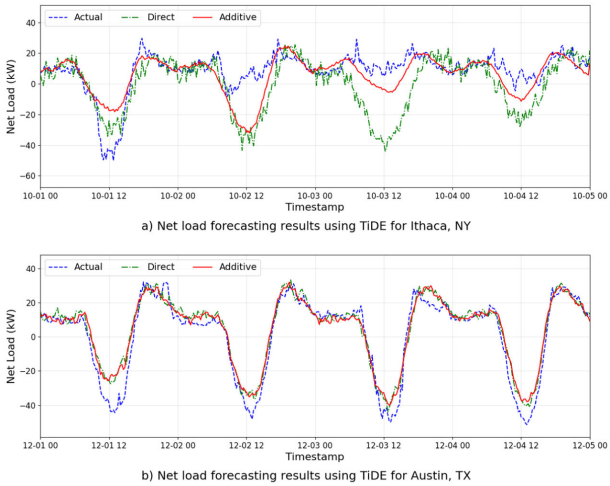

FIGURE 7. Net load forecast results of TiDE model.

In Fig. 7, the predicted net load for both the direct and additive methods using the TiDE model is displayed. Thus, the additive method can better capture the pattern of the BTM pure load at night than the direct method, as the BTM forecasting model is separated.

Additionally, in Fig. 8, forecast results for the additive approach using the TiDE model show the results of a) BTM PV generation, b) BTM pure load separately, and c) net load forecast as the sum of both components for the Austin, TX dataset. Results reveal that BTM PV generation and BTM pure load trends have different behaviors, which is consistent with the additive forecasting approach's assumption that forecasting BTM PV generation and pure load separately will provide a better trend capture. Moreover, the forecast results of the BTM pure load show superiority in capturing the trend of the BTM pure load, contrary to BTM PV generation, which is more volatile and difficult to forecast due to external factors, such as weather conditions and the condition of PV panels. The most crucial factor is the duration of the collected data (4 months of training data in Ithaca, NY, and 8 months in Austin, TX). Such a time period may not be enough to capture the pattern of BTM PV generation long sequence time-series forecasting tasks.

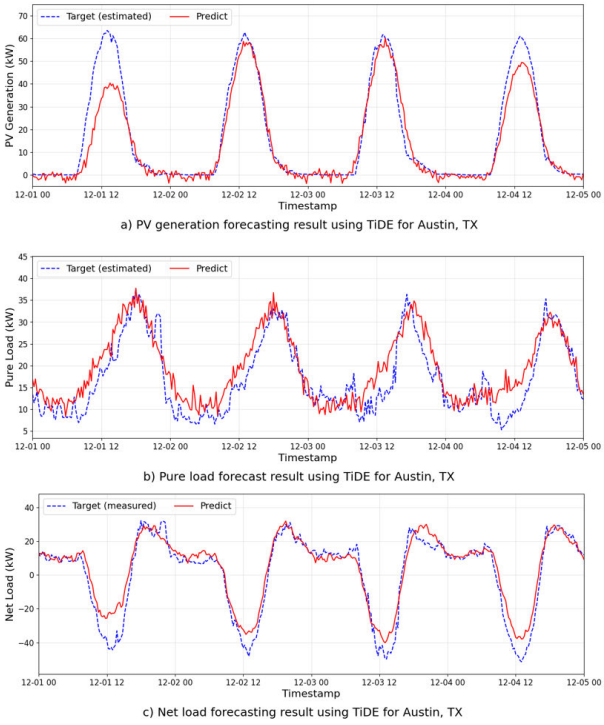

FIGURE 8. Forecast result of PV generation, pure load, and net load of additive approach for Austin, TX.

Further, to analyze how unexpected weather changes affect forecasting results, we examined two periods based on GHI data during the testing phase for the Ithaca, NY, and Austin, TX datasets, as shown in Fig. 9. Specifically, we selected a sunny period characterized by high GHI and a cloudy period characterized by low GHI, as follows:

- For the Ithaca, NY dataset: sunny period spans from 2019-10-08 to 2019-10-11, cloudy period spans from 2019-10-16 to 2019-10-18, and entire selected period spans from 2019-10-08 to 2019-10-18.
- For the Austin, TX dataset: sunny period spans from 2018-12-01 to 2018-12-04, cloudy period spans from 2018-12-05 to 2018-12-08, and entire selected period spans from 2018-12-01 to 2018-12-08.

The net load forecast results of the winning model, TiDE with an additive approach, were evaluated across sunny, cloudy, and entire selected periods. Tables 11 and 12 present partial evaluation results for the Ithaca, NY and Austin, TX datasets. Both tables indicate that the sunny period had lower RMSE, MAE, and sMAPE values compared to the other periods, particularly the entire selected period. This demonstrates that unexpected weather changes significantly impact forecasting accuracy. It underscores a major challenge in time series forecasting: the need to capture unexpected changes in external factors, such as weather conditions in this case.

In conclusion, we demonstrate the effectiveness of the proposed net load disaggregation method in improving the net load forecasting performance using the additive approach. Results show that the proposed method can enhance the

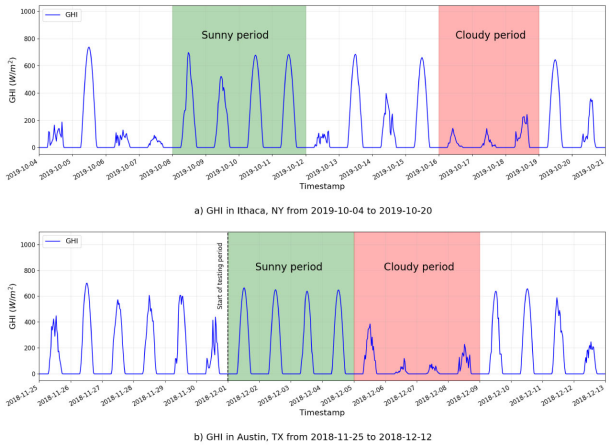

FIGURE 9. GHI during the testing set period from Ithaca, NY and Austin, TX datasets.

TABLE 11. Evaluation results using the Ithaca, NY dataset for the entire selected period (shown above), separated into sunny and cloudy periods, compared to the overall results of the testing data (shown below).

Period	RMSE (kW) ↓	MAE (kW) ↓	sMAPE (%) ↓
Entire selected	15.26	9.89	32.90
• Sunny	13.11	8.30	26.28
• Cloudy	15.22	10.23	39.49
Overall	15.35	9.82	34.78

TABLE 12. Evaluation results using the Austin, TX dataset for the entire selected period (shown above), separated into sunny and cloudy periods, compared to the overall results of the testing data (shown below).

Period	RMSE (kW) ↓	MAE (kW) ↓	sMAPE (%) ↓
Entire selected	10.67	7.51	27.33
• Sunny	9.94	6.97	22.08
• Cloudy	11.34	8.05	32.58
Overall	12.60	8.35	29.02

forecasting performance by separating the BTM components and forecasting them separately. Results also show that the TiDE model can provide the best forecasting performance among the three models. Furthermore, the additive forecasting results signify that the TiDE model forecasts the BTM pure load more accurately than the BTM PV generation, owing to the distinct behavior of the two components and external factors influencing the BTM PV generation.

VII. CONCLUSION

In this work, we aim to improve day-ahead net load forecasting based on an additive approach to add up two forecasted time series that represent behind-the-meter (BTM) components, including pure load and PV generation. Because BTM components are invisible, we developed an unsupervised net load disaggregation method to estimate the BTM PV generation from the net load. The model was based on LSTM with an unsupervised contrastive loss from a set of selected time pairs. After disaggregating BTM time series, we have compared many modern time series forecasting techniques, including DeepAR, TFT, and TiDE.

TABLE 13. Model hyperparameters for the proposed BTM PV generation estimation model.

Hyperparameter	Value
Layer type	LSTM
Number of neurons	256^a-256^a-512^b-256^b
Time step of LSTM	1
Batch size	64
Number of epochs	300
Optimizer	Adam
Learning rate	0.001

[a] denotes the number of neurons in the LSTM layer, and the others are dense layers.
[b] denotes that the ReLU activation function is applied to the layer.

The experiments were conducted on two real-world datasets from Pecan Street Inc. (Ithaca, NY, and Austin, TX). Results show that the proposed unsupervised disaggregation method based on LSTM, compared to the semi-supervised method, was the winner in terms of RMSE, MAE, and sMAPE, corresponding to 2.72 kW, 1.56 kW, and 7.49% in Ithaca, NY, and 7.06 kW, 4.03 kW, and 15.90% in Austin, TX. These results verify that the proposed method is superior even in the absence of BTM data to estimate BTM PV generation. For net load forecasting, the additive approach is seen to outperform the direct approach in terms of average percentage improvement in RMSE, MAE, and sMAPE for both datasets and all models, corresponding to 8%, 15%, and 6% in Ithaca, NY, and 13%, 16%, and 3% in Austin, TX.

Since the duration of the collected data may not be sufficient to capture the pattern of BTM PV generation in long sequence time-series forecasting tasks, we will explore electricity prosumption using BTM data over a longer period in future work.

APPENDIX A NOMENCLATURE

Abbreviations

ARX	Autoregressive Exogenous.
BTM	Behind-the-Meter.
GHI	Global Horizontal Irradiance.
LSTM	Long Short-Term Memory.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MLP	Multilayer Perceptron.
PV	Photovoltaic.
RMSE	Root Mean Squared Error.
RNN	Recurrent Neural Network.
sMAPE	Symmetric Mean Absolute Percentage Error.
TFT	Temporal Fusion Transformer.
TiDE	Time-series Dense Encode.

Symbols

Δl_t	Different of pure load from timepair in time interval t.
Δp_t	Different of PV generation from timepair in time interval t.
τ	Set of timepair samples.

TABLE 14. Specific hyperparameters of DeepAR.

Hyperparameter	Value
Hidden size	256
LSTM layers	2
Likelihood model	Gaussian
Dropout rate	0.3
Batch size	512
Epochs	100
Optimizer	Adam
Initial learning rate	0.001
Learning rate scheduler	ReduceLROnPlateau
Loss function	Beta negative log-likelihood

TABLE 15. Specific hyperparameters of TFT.

Hyperparameter	Value
Hidden size	256
LSTM layers	2
Attention heads	4
Hidden size for processing continuous variables	8
Likelihood model	Quantile regression with set of quantile {0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.99}
Dropout rate	0.5
Batch size	192
Epochs	100
Optimizer	Adam
Initial learning rate	0.0001
Learning rate scheduler	ReduceLROnPlateau
Loss function	Quantile loss

DOY_t	Day of year in time interval t.
GHI_t	GHI in time interval t.
GHI_{max}	Maximum GHI from dataset.
l_t	Aggregated pure load in time interval t.
$l_{c,t}$	Pure load for the customer c in time interval t.
MOD_t	Minute of day in time interval t.
n_t	Aggregated net load in time interval t.
$n_{c,t}$	Net load for the customer c in time interval t.
p_t	Aggregated PV generation in time interval t.
$p_{c,t}$	PV generation for the customer c in time interval t.
S	Set of customer.
$S1, S2, S3, S4$	Criterion score.
$Score(t_1, t_2)$	Sum of criterion score in time interval t_1 and t_2.

APPENDIX B
HYPERPARAMETERS OF PV ESTIMATION MODEL

In Table 13, the hyperparameters of the proposed LSTM model are shown.

TABLE 16. Specific hyperparameters of TiDE.

Hyperparameter	Value
Hidden size (encoder and decoder)	1024
Number of encoder layers	2
Number of decoder layers	2
Hidden size (temporal decoder)	128
Layer normalization	True
Reversible instance normalization	True
Dropout rate	0.5
Batch size	64
Epochs	300
Optimizer	Adam
Initial learning rate	0.0001
Learning rate scheduler	ReduceLROnPlateau
Loss function	Mean squared error

APPENDIX C
HYPERPARAMETERS OF FORECASTING MODELS

In Tables 14, 15, and 16, the hyperparameters of the DeepAR, TFT, and TiDE models are shown, respectively.

REFERENCES

[1] C. S. Lai, G. Locatelli, A. Pimm, X. Wu, and L. L. Lai, "A review on long-term electrical power system modeling with energy storage," *J. Cleaner Prod.*, vol. 280, Jan. 2021, Art. no. 124298.
[2] C. S. Lai, Y. Jia, L. L. Lai, Z. Xu, M. D. McCulloch, and K. P. Wong, "A comprehensive review on large-scale photovoltaic system with applications of electrical energy storage," *Renew. Sustain. Energy Rev.*, vol. 78, pp. 439–451, Oct. 2017.
[3] C. S. Lai and M. D. McCulloch, "Levelized cost of electricity for solar photovoltaic and electrical energy storage," *Appl. Energy*, vol. 190, pp. 191–203, Mar. 2017.
[4] X. Xu, J. Li, Y. Xu, Z. Xu, and C. S. Lai, "A two-stage game-theoretic method for residential PV panels planning considering energy sharing mechanism," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3562–3573, Sep. 2020.
[5] 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.
[6] F. Lin, Y. Zhang, K. Wang, J. Wang, and M. Zhu, "Parametric probabilistic forecasting of solar power with fat-tailed distributions and deep neural networks," *IEEE Trans. Sustain. Energy*, vol. 13, no. 4, pp. 2133–2147, Oct. 2022.
[7] A. Stratigakos, A. Bachoumis, V. Vita, and E. Zafiropoulos, "Short-term net load forecasting with singular spectrum analysis and LSTM neural networks," *Energies*, vol. 14, no. 14, p. 4107, Jul. 2021.
[8] S. H. Kim, G. Lee, G.-Y. Kwon, D.-I. Kim, and Y.-J. Shin, "Deep learning based on multi-decomposition for short-term load forecasting," *Energies*, vol. 11, no. 12, p. 3433, Dec. 2018.
[9] B. C. Erdener, C. Feng, K. Doubleday, A. Florita, and B.-M. Hodge, "A review of behind-the-meter solar forecasting," *Renew. Sustain. Energy Rev.*, vol. 160, May 2022, Art. no. 112224.
[10] Y. Chu, H. T. C. Pedro, A. Kaur, J. Kleissl, and C. F. M. Coimbra, "Net load forecasts for solar-integrated operational grid feeders," *Sol. Energy*, vol. 158, pp. 236–246, Dec. 2017.
[11] F. Mei, Q. Wu, T. Shi, J. Lu, Y. Pan, and J. Zheng, "An ultrashort-term net load forecasting model based on phase space reconstruction and deep neural network," *Appl. Sci.*, vol. 9, no. 7, p. 1487, Apr. 2019.

[12] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2020.

[13] M. Sun, C. Feng, and J. Zhang, "Factoring behind-the-meter solar into load forecasting: Case studies under extreme weather," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2020, pp. 1–5.

[14] A. Kaur, L. Nonnenmacher, and C. F. M. Coimbra, "Net load forecasting for high renewable energy penetration grids," *Energy*, vol. 114, pp. 1073–1084, Nov. 2016.

[15] A. Stratman, T. Hong, M. Yi, and D. Zhao, "Net load forecasting with disaggregated behind-the-meter PV generation," *IEEE Trans. Ind. Appl.*, vol. 59, no. 5, pp. 5341–5351, Oct. 2023.

[16] 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)*, Jan. 2023, pp. 1–5.

[17] 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.

[18] C. M. Cheung, W. Zhong, C. Xiong, A. Srivastava, R. Kannan, and V. K. Prasanna, "Behind-the-meter solar generation disaggregation using consumer mixture models," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2018, pp. 1–6.

[19] 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.

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

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

[22] 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.

[23] R. Ross Jr., "Flat-plate photovoltaic array design optimization," in *Proc. 14th Photovolt. Spec. Conf.*, 1980, pp. 1126–1132.

[24] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *Int. J. Forecasting*, vol. 36, no. 3, pp. 1181–1191, Jul. 2020.

[25] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021.

[26] A. Das, W. Kong, A. Leach, S. Mathur, R. Sen, and R. Yu, "Long-term forecasting with TiDE: Time-series dense encoder," 2023, *arXiv:2304.08424*.

[27] C. Holcomb, "Pecan Street Inc.: A test-bed for NILM," in *Proc. Int. Workshop Non-Intrusive Load Monitor*. New York, NY, USA: ACM, 2012, pp. 3:1–3:8. [Online]. Available: https://www.pecanstreet.org/

[28] E. Garcia-Garrido, M. Mendoza-Villena, P. M. Lara-Santillan, E. Zorzano-Alba, and A. Falces, "Net demand short-term forecasting in a distribution substation with PV power generation," in *Proc. E3S Web Conf.*, vol. 152. Les Ulis, France: EDP Sciences, 2020, p. 01001.

[29] A. Falces, C. Capellan-Villacian, M. Mendoza-Villena, P. J. Zorzano-Santamaria, P. M. Lara-Santillan, E. Garcia-Garrido, L. A. Fernandez-Jimenez, and E. Zorzano-Alba, "Short-term net load forecast in distribution networks with PV penetration behind the meter," *Energy Rep.*, vol. 9, pp. 115–122, May 2023.

[30] M. Jia, K. Pu, and Y. Zhao, "Net load forecast based on behind-the-meter disaggregation of smart meter data," in *Proc. 57th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2023, pp. 1–6.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[32] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (NSRDB)," *Renew. Sustain. Energy Rev.*, vol. 89, pp. 51–60, Jun. 2018.

[33] J. Herzen et al., "Darts: User-friendly modern machine learning for time series," *J. Mach. Learn. Res.*, vol. 23, no. 124, pp. 1–6, 2022.

[34] 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, Nov. 2019, Art. no. 113595.

CHAICHAN THEPPROM received the B.Sc. degree in computer science from the Department of Computer Science, Prince of Songkla University, Thailand, in 2022. He is currently pursuing the master's degree in computer science with the Department of Computer Engineering, Chulalongkorn University, Thailand. His research interests include deep learning, pattern recognition, signal processing, and time series forecasting.

NATAWUT NUPAIROJ (Member, IEEE) received the B.Eng. degree from Chulalongkorn University and the M.S. and Ph.D. degrees in computer science from Michigan State University. He is currently an Assistant Professor with the Department of Computer Engineering, Chulalonkorn University, and the Director of the Center of Excellence in Digital and AI Innovation for Mental Health (AIMET), Faculty of Engineering, Chulalongkorn University. His research focuses on applying big data and AI technology in multidisciplinary areas, especially in energy and healthcare. His current research interests include MLOps, big data architecture, high-performance ML model inferencing, healthcare process mining, and AI applications in healthcare.

PEERAPON VATEEKUL (Senior Member, IEEE) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Miami (UM), Coral Gables, FL, USA, in 2012. Currently, he is an Associate Professor with the Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. His research interests include machine learning, data mining, deep learning, text mining, and big data analytics. To be more specific, his works include variants of classification (hierarchical multi-label classification), natural language processing, data quality management, and applied deep learning techniques in various domains, such as medical images and videos, satellite images, meteorological data, and text.