Renewable and Sustainable Energy Reviews 160 (2022) 112224

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

A review of behind-the-meter solar forecasting

Burcin Cakir Erdener [a,*], Cong Feng [b], Kate Doubleday [a,b], Anthony Florita [b], Bri-Mathias Hodge [a,b]

[a] *Electrical, Computer, and Energy Engineering, and Renewable and Sustainable Energy Institute, University of Colorado Boulder, Boulder, CO, USA*
[b] *Power Systems Engineering Center, National Renewable Energy Laboratory, Golden, CO, USA*

ARTICLE INFO

Keywords:
Behind-the-meter
Solar forecasting
Solar photovoltaic

ABSTRACT

Solar photovoltaic systems largely integrated within the distribution grid are operated 'behind-the-meter' and power generation cannot be directly monitored by most utilities. The increasing penetration of behind-the-meter solar photovoltaic systems can deter efficient network and market operations due to variability and uncertainty in net load, which is exacerbated by limited visibility and the difficulty in analyzing the hosting capacity. Risk introduced by behind-the-meter solar contributions may hinder reliable and secure grid operations due to biased system monitoring and forecasts. Accurate behind-the-meter estimations, together with capacity and specification forecasts, thus play a key role in balancing supply and demand and this article reviews the pertinent literature, identifying key characteristics and predictive methods for efficient behind-the-meter solar photovoltaic generation. Forecasting is central to methods herein. The fundamental characteristics of behind-the-meter solar forecasting, including which methods are applicable for scenario-driven use cases, are driven by the metrics most useful for system-wide performance evaluation. To this aim, the literature is reviewed with a focus on forecasting applications for aggregate, regional behind-the-meter generation useful to bulk system and utility operations. As distinguished from net load forecasting, subtleties in these coincident tasks are explored before concluding with recommendations for current practice and future implementations.

1. Introduction

The global electricity generation capacity of installed photovoltaic (PV) solar power has expanded rapidly over the past decade and exceeded 635 GW at the end of 2019 [1]. Current estimates indicate that the total installed capacity will increase six-fold over 2018 levels by 2030 and reach > 8000 GW by 2050 [2]. According to the International Energy Agency (IEA), half of the growth in solar energy capacity globally will be small-scale, distributed generations connected to the electric distribution system, as opposed to large, centralized plants feeding into the transmission system [3]. Most of these residential, commercial, and/or rooftop PV systems are installed behind-the-meter (BTM), mainly due to lower consumer costs, state-mandated renewable policies, trends towards a decentralized grid, and the need for innovative business models such as third-party ownership [4]. The increase in on-site energy generation due to the large number of installed BTM PV systems has caused changes in the operation and planning of power systems, and load forecasting in particular during critical times of the day. In the presence of BTM PV systems, observing their activity separate from loads can be difficult because most of the time only the accumulated net load reading is available to the utility.[1] However, in large quantities, BTM PV systems can significantly alter the shape of regional net load profiles and pose balancing and reliability challenges.

In geographies like Australia, Germany, Hawaii or California, fast growth rates of residential solar PV adoptions are accelerating the need for BTM visibility and forecastability. In Australia, BTM solar system had an 8.03 GW total capacity with 2.06 million PV installations in 2019 [5]. With such a high penetration, the quality of load forecasts can be degraded, impeding reliable and efficient network operations [6]. In Hawaii, 21% of single-family homes have rooftop solar and mid-day loads show significant down-ramping at peak PV generation and up-ramping with its diminished performance near the latter portions of the diurnal cycle; regional net load concerns necessitate further examination of the phenomena [7]. Fig. 1 depicts (a) the progression of the net load profile with additional PV generation in power system operations, the so-called "duck curve" with the belly of the curve growing with increasing levels of BTM solar penetration, and (b) the risk introduced with shoulder period ramping can be more accurately anticipated with probabilistic rather than point forecasts. With increasing amounts of BTM solar throughout the world, the capability to forecast, estimate, and control BTM PV systems is becoming important to smart grid operations [8].

* Corresponding author.
E-mail address: Burcin.CAKIRERDENER@colorado.edu (B.C. Erdener).

[1] Net load here is defined as the customer's energy consumption minus its solar generation.

https://doi.org/10.1016/j.rser.2022.112224
Received 26 March 2021; Received in revised form 30 November 2021; Accepted 1 February 2022
Available online 24 February 2022
1364-0321/Published by Elsevier Ltd.

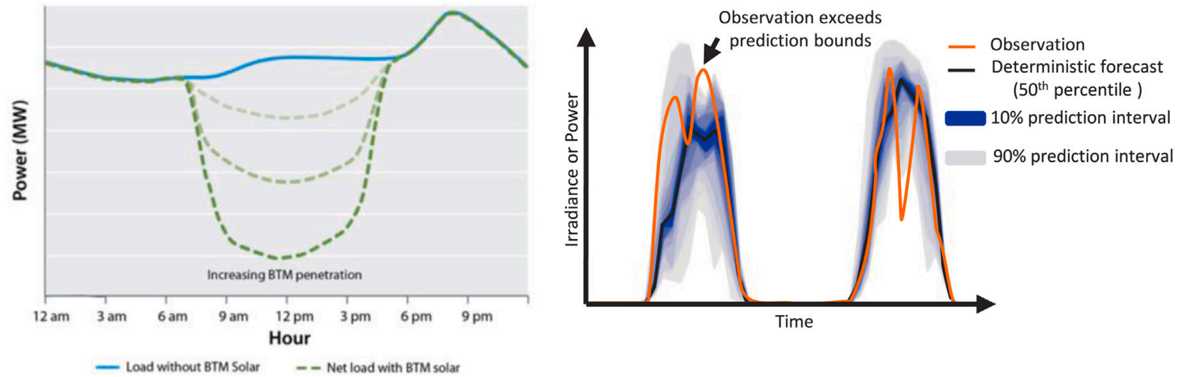

Fig. 1. (a) net load profiles with increasing penetration of BTM PV generation, adapted from [8]; (b) illustrative comparison of a point forecast and a probabilistic interval forecast of the 10%, 20%, …, 90% central intervals. The 50th percentile of a probabilistic forecast can also interpreted as a point forecast.

The system operators' visibility and forecastability of BTM PV system performance is currently latent to load expectations, owing to sparse PV penetration in many regions, and cannot be easily uncovered from existing measurements [9]. Explicit prediction of BTM PV system performance will enable reference signaling for both enhanced grid control and superior load-side dispatchability in conjunction with human-driven activities and demands. As with utility-scale PV systems, BTM PV is a variable and uncertain weather-dependent resource. However, the uncertainty in anticipating power generation from BTM systems is compounded by systemic missing data or information; e.g., lack of records or knowledge concerning installed BTM PV systems interferes with accurate time-series performance forecasting, making advanced energy management for load shaping and modulation (or optimal demand response) currently untenable at scale. Unlike utility-scale PV plants, specifications on residential PV systems are largely absent for commercial forecasters and system operators alike — or key data is often only partially known [10,11]. In an effort to minimize risks, utilities typically require permits and enforce interconnection requirements for installing BTM solar generation [12]. Nonetheless, the number of unregistered or unknown BTM PV systems is significant for a variety of reasons, including installations made prior to regulations being enacted, lack of awareness about the rules, owners avoiding permitting fees, and discrepancies between the reported and the installed configurations [13]. As the inadequate and biased information on the capacity and specification of BTM PV systems undermine the reliable operation of power systems, it is an essential step to include capacity estimation methods in the context of BTM PV generation forecasts.

While the literature on estimating the output of individual PV units is rich, the literature concerning aggregated output of a large fleet of unobserved BTM PV forecasting is limited. The fact that the actual production values are unknown to the entity conducting the forecasting is the main issue for forecasting the aggregated output of BTM PV installations. While individual units are metered by local power utilities, the collected data is often stored within the distribution company domain and used for settlement purposes. There is no infrastructure available to monitor and link the accumulated PV production values to transmission-level energy management systems in real time. [14].

Some researchers focus only on the prediction of solar irradiance, considering the fact that accurate irradiance forecasting is a prerequisite to accurate forecasting of the PV system's power output. Review of statistical approaches and techniques based on cloud images and numerical weather prediction models used for irradiance forecasting can be found in [15]. A detailed overview of irradiation forecasting using machine learning approaches is provided in [16] and a comparison and evaluation of several machine learning techniques for predicting the daily global solar radiation can be found in [17]. A more recent review on irradiance forecasting can be found in [18] where a large number of publications on the topic are reviewed with a text mining approach.

Forecasting output of PV power from individual systems are also considered in several reviews. For example, statistical and physical methods used to forecast solar irradiance and output PV power are reviewed in [19–21]. A comprehensive recent review on solar PV power forecasting techniques using time-series statistical, physical, and ensemble methods is provided in [22] including analysis of metrics assessment. Some review papers only focus on a certain forecasting methodology, such as ensemble models [23] and artificial intelligence (AI) [24], and probabilistic solar forecasting [25].

A 2014 review considered both generation and load forecasting describing net load forecasting at the scale of a single building, which included a PV component as part a commercial building's net load [26]. In a more recent review, literature on probabilistic solar power and load forecasting methods are reviewed separately, to map the opportunities for probabilistic net load forecasting [27]. However the study includes the papers published before 2017 and does not focus on the additional modeling efforts necessary for producing BTM solar generation forecasts.

The available literature revealed that particular attention to forecasting techniques for BTM PV output generation field is still limited and an updated literature review summarizing previous studies is required. Therefore this article reviews recent work addressing the added complexity of forecasting PV located behind the meter. The contribution of this paper can be summarized as follows.

Different from the existing review literature on solar power forecasting, this paper focuses on forecasting efforts required for invisible PV generation. While building- and neighborhood-scale BTM forecasting methods are also introduced, this review emphasizes methods for aggregate regional BTM forecasting that are most useful to power system operators. This includes methods to identify under-reported BTM PV systems and estimate their capacity and specifications — steps unnecessary for utility-scale PV forecasting but crucial to fill gaps in BTM registration data. Additionally, it differs significantly from previous reviews as the main objective is to provide a better understanding of the data flows for BTM forecasting, considering different users and use cases, and to discuss the methods used for each particular case, with a particular focus on regionally aggregated forecasts for power system operators. Finally, recent papers on net load forecasting which is an important application of BTM forecasting have also been reviewed. The rest of the manuscript is organized as follows: Section 2 presents fundamentals of BTM solar forecasting, including different forecast users and applications, spatiotemporal scales, and point versus probabilistic formats. Section 3 details capacity and specification estimation for BTM PV. Section 4 reviews current implementations of BTM PV forecasting methods in the literature. Section 5 considers existing applications in net load forecasting. Finally, conclusions and recommendations on research directions are discussed in Section 6.

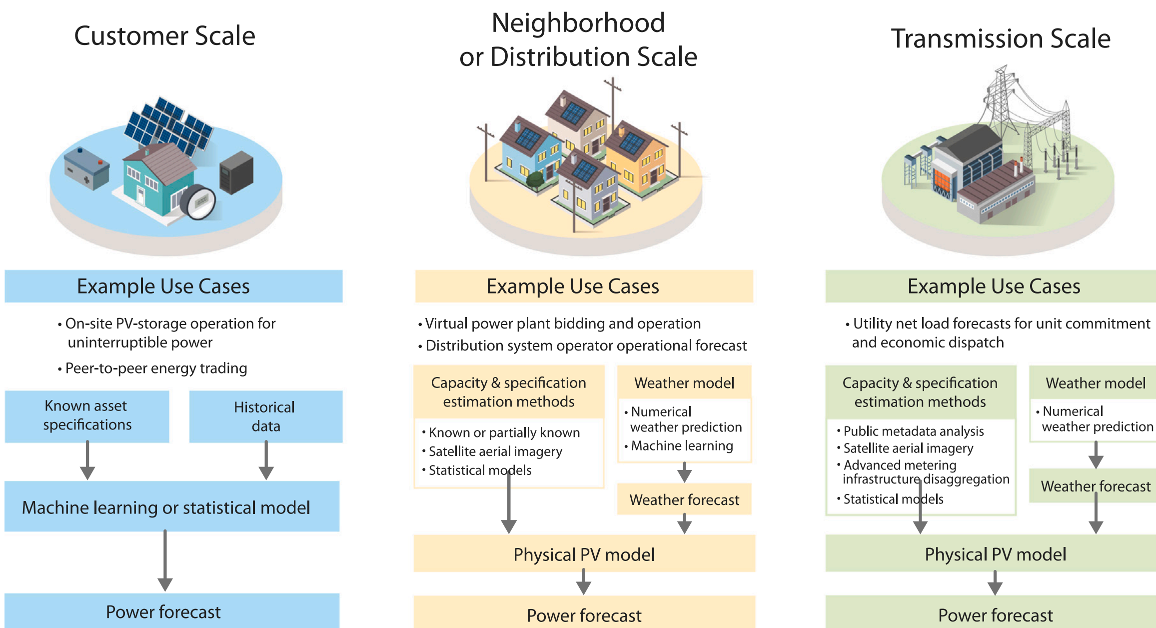

Fig. 2. Example approaches to BTM solar PV forecasting for different users and use cases.

2. Fundamentals of behind-the-meter solar forecasting

A variety of forecasting methods are suitable for PV power forecasting, in general, and BTM solar forecasting in particular. To select the best method for a given application, these methods can be grouped across multiple axes, such as their users and use cases, temporal and spatial resolutions, input data requirements, and approach to uncertainty quantification. This section introduces these fundamental characteristics that can be used to distinguish among methods, beginning with the needs of the users these forecasts are ultimately intended to serve.

2.1. Users and use cases

BTM solar forecasting has applications offering value to various entities in the power grid, including consumers, distributed energy resource (DER) aggregators, and distribution and transmission system operators. Fig. 2 illustrates example forecasting approaches based on these different use cases and the likely information available to the forecaster.

Customer-scale forecasts may be used for combined BTM PV and storage operations, whether for uninterruptible power or peer-to-peer trading applications [28,29]. At the distribution scale, forecasts of site-level BTM PV can be used for virtual power plant (VPP) bidding and operation and for operational forecasting of demand within a distribution service area [30,31]. For transmission system operators, longer-term predictions, used in transmission system operations compared to the forecasts used in energy storage operations and distribution system operations, may be used for utility net load forecasts utilized in the unit commitment and economic dispatch processes [9,32,33]. At both the distribution and transmission scales, BTM PV power forecasting requires additional considerations regarding installed capacity and specification estimation of the PV deployments [11]. The data related to each PV system, such as system location and orientation, manufacturer specifications, and capacity information is generally required to convert irradiance forecasts into power output by a PV system model. Once the desired forecasting application is identified, the necessary spatial and temporal resolutions, discussed in the next subsection, are more easily determined.

2.2. Spatial resolution

BTM forecasting approaches can be grouped into three spatial resolutions directly corresponding to the three user types (Fig. 2): the individual building level, the neighborhood level – representing a distribution feeder or campus, and the regional level – for balancing areas. On the one end of the spectrum, forecasting for an individual customer is relatively straightforward. The user knows system specification data and might have access to their own historical observations. Forecasts can be generated by any of a variety of machine learning or time-series methods proposed in the literature for forecasting PV output for point locations (e.g., utility-scale PV plants). For instance, the prediction of a 1 MW PV plant power production with support vector machines using several numerically predicted weather variables is presented in [34]. A similar approach is proposed for a grid-connected PV plant in [35], by using numerical weather prediction models combined with an artificial neural network-based model. A non-parametric PV model considering forecasts of meteorological variables and actual AC power measurements of PV plants is proposed in [36]. Uncertainty associated with the forecast of PV generation is analyzed in [37], for a solar house equipped with demand-side management techniques and a local storage system.

As a result, this category has been well covered in the recent review literature, and is thus not the primary focus herein. On the other end of the spectrum, forecasting methods for regionally aggregated BTM PV generation can be categorized into bottom-up and top-down approaches, selected based on the availability of the requisite PV system data [38].

Bottom-up forecasting is suitable when the data of all the individual installations are known, including their location, capacity, tilt, azimuth, etc. Well-defined physical models combined with irradiance and weather data can be used to directly predict the production at each PV installation and sum those over the region of interest [39]. For example, the California Independent System Operator collected all the data from all the PV systems in the state of California and then combined it with high-resolution irradiance values and weather forecasts to predict the total contribution of BTM solar systems in its grid [40]. Such a "bottom–up" approach is well suited for aggregating forecasts of utility-scale PV plants, and for understanding the localized impacts in regional analysis. This approach has the advantage of explicitly capturing each system's specifics, but it is more challenging for BTM forecasting where PV installation system data is often patchy.

When the goal is to estimate aggregated PV production or net load across a large region and/or when detailed data is not available, top-down BTM forecasting approaches may be more suitable. Top-down approach is based on combinations of the partially known data; e.g., if data from selected representative PV systems can be obtained, it can be used to estimate the total generation of the region [11]. This approach requires proper selection of the subset of systems and comparisons to metered data on a statistically significant number of sites for benchmarking purposes. In [41], upscaling from a representative set of PV systems is proposed to provide regional PV point forecasts by analyzing the orientation and module types of systems in Germany. The problem of poor representation of the region is considered in [42,43] by feeding a PV model with data from a larger sample of PV systems. Fuzzy confidence intervals are used in [14] for uncertainties in the input data with the advantage of collecting data from a few locations in the region rather than keep tracking of all of them.

If no data is available for any of the systems but aggregated regional production data is available, then regional output generation can be predicted directly using statistical models, such as auto-regressive moving average (ARMA) models [44,45], generalized auto-regressive integrated moving average (ARIMA) models [46] and Gaussian Conditional Random Fields (GCRF) models [47]. In the final case, when no data is available either for individual PV systems or for the region, irradiance forecasts can be transformed to power based on estimated installed capacity [48]".

Estimating the capacity and specifications of installed BTM PV systems is an active research area. When data is sufficiently known, regionally aggregated forecasts tend to benefit from lower forecasting errors due to spatial averaging and smoothing effects [49–51]. To estimate specifications, a likely first approach is to analyze public metadata to assess data quality and identify major trends, such the average capacity, tilt, and orientation of rooftop systems and the growth in installed capacity over time. Additional approaches include analysis of satellite aerial imagery and disaggregation of advanced measuring infrastructure (AMI) data, which are detailed in Section 3. At intermediate scales, VPP and distribution system operators might borrow from both the single-site and regionally aggregate approaches. VPP operators for instance might have access to private system data, allowing the use of bottom-up approaches less suitable for system operators. However, this review focuses primarily on the transmission scale in Fig. 2 and the methods particular to BTM forecasting at that resolution, including capacity and specification estimation and regional net load forecasting.

2.3. Temporal characteristics

PV system power output is highly dependent on weather and atmospheric conditions, especially the global horizontal irradiance (GHI), which has predictable daily and seasonal variations but also depends on changing local cloud cover. Different forecasting methods have varying strengths in anticipating these changes over the next few minutes, hours, or days. Two key parameters that describe the temporal specifications of a forecast are its resolution and horizon. Resolution describes the period between two forecast valid times (e.g., hourly or 5 minute resolution), and horizon describes the total duration of a sequence of forecasts (e.g., a sequence of 5 minute resolution forecasts over the next 1 h horizon). Forecasts can be broadly described by categorizing the horizon into very short-term (minutes to 6 h ahead, possibly with intra-hourly resolution), short-term (intra-day to day-ahead forecasts, typically with hourly resolution), and long-term (seasonal horizon with daily, monthly, or quarterly resolution). Typically increasing forecasting horizon will have negative impact on accuracy metrics, particularly for models solely depend on historical data [52]. The use case of the forecast will likely dictate its horizon and resolution requirements, and thus which methods are most suitable [53]. For instance, very short-term forecasts are highly beneficial for power smoothing and real-time power dispatching. At minutes-ahead timescales, climatic parameters are almost identical to the previous time steps, and BTM PV is most sensitive to rapid changes in cloud cover. Short-term forecasts are popular in load balancing, power plant management, and power system operations decisions [22]. Long-term forecasting is effective for maintenance scheduling of utility-scale plants and utility planning. Distribution and transmission system operators most often use short and medium forecasting but may also use long-term horizons for planning system operations and maintenance [54].

At the time scales suitable for power system operations, it is common to use solar irradiance forecasts based on numerical weather prediction (NWP) models that simulate upcoming atmospheric conditions as the foundation of the forecast. NWP models are computationally intensive simulations that require several hours to solve physics-based equations of atmospheric processes on a grid over the Earth's atmosphere. These models are typically provided by national and international weather agencies, such as the National Oceanic and Atmospheric Administration (NOAA) and the European Centre for Medium-Range Weather Forecasts (ECMWF) [55,56]. NOAA and ECMWF both provide a selection of NWP models at various spatial and temporal resolutions, with spatial grids of 3–28 km, temporal resolutions of 1–3 h, and forecasting horizons of a few hours to 16 days. NWP models typically perform well for time horizons more than 4 h [19], though recently developed models have shorter timescales, such as NOAA's High Resolution Rapid Refresh which has 15 minute resolution and is updated hourly [57]. From the coarse grid-scale forecast, statistical post-processing can be applied to correct for finer-scale topography and generate a forecast for a specific location [58]. For the purposes of BTM solar power forecasting, irradiance forecasts from a method such as NWP can be assumed to be available. While weather forecasting in general and irradiance forecasting in particular are not the focus of this review, additional information can be found in [15,17,16].

2.4. Input data used in BTM forecast entities

Power forecasting methods best suited to their users' needs will vary based on the input data available, such as system specifications, weather forecasts, and historical observations. A building owner/operator, for example, could generate a forecast for their own use will have most necessary details about the installed PV system. A power system operator generating a forecast of the aggregate BTM PV generation in their service territory might need to estimate system specifications and apply a time-series method appropriate for a wide-area forecasts, rather than point locations. The approach taken might be a single step that directly produces a power forecast, or it might require multiple steps with different sources of input data and pre-processing to impute missing data. Similar to utility-scale PV system forecasting, data-driven models (statistical models and machine learning algorithms), or physical models, can be used in BTM solar power forecasting. The statistical approach relies primarily on historical data to "train" models, with limited reliance on NWP and PV system models [59]. In statistical models, the starting point is a training data set that contains historical PV power, as well as various inputs or potential inputs involving meteorological data and PV asset specifications. This data set is used to train models – such as classical statistical models or artificial intelligence models – that output a forecast of PV power at a given time. In contrast, physical models do not use historical data but rather solar irradiance forecasts and PV system data to generate PV output forecasts. An example of a minimal set of PV system specification data is listed in Table 1 for the *PVWatts calculator*, which estimates the energy production and cost of energy of grid-connected PV systems [60]. A number of parameters are applicable to both utility-scale and BTM installations, though, for example, fixed installations are much more prevalent than tracking systems for residential and commercial rooftop installations.

Table 1
PV system specification parameters, [60].

Field	Units
System size	kW (DC)
Module type	Standard, premium, thin film
System losses	%
Array type	Fixed open rack, Fixed roof mount, 1-Axis, Backtracked 1-Axis, 2-Axis
Tilt angle	Degrees
DC/AC ratio	Ratio
Azimuth angle	Degrees
Inverter efficiency	%
Ground coverage ratio (1-axis only)	Fraction

BTM PV forecasts can rely on solar irradiance forecasts when the information and measurements related with the system are missing. Solar irradiance forecasts are generated by an NWP model that uses meteorological data combined with calculated solar zenith and azimuth angle data. Solar irradiance shows a strong positive correlation with PV power output [66]. Researchers have proposed several approaches to estimate solar irradiance including: statistical models [48,67], artificial neural networks [68–71], artificial intelligent techniques [72], cloud tracking [73,74], sky imagers [75,74,76], NWP models [77,78] and hybrid models [79–82]. These models use meteorological data such as GHI, air temperature, cloud cover, humidity, air pressure, wind pressure, etc., and physical parameters of solar position such as solar zenith, azimuth angle, as the input. Sources of meteorological data can include both empirical records and forecasts.

Although BTM solar data has low transparency and accessibility, there are still some data sets that can be used for BTM solar modeling and forecasting. These data sets usually contain BTM solar metadata, solar time series, or both. BTM solar meta data are in the form of BTM PV system specifications or remote sensing images that could be used to derive BTM PV capacity [83]. For example, an aerial image data set that contains 19,000 distributed PV systems with per-pixel labels [84]. The data set covers four cities in California, namely, Fresno, Stockton, Modesto, and Oxnard. This data set can be used to develop new non-intrusive PV detection methods for BTM PV modeling. A similar data set is provided by the DeepSolar platform [85]. The DeepSolar covers the contiguous U.S. Even though the data set has site-level data, it publishes solar deployments at the county-level to protect information privacy.

Solar time series data sets contain meteorological time series or PV power time series. PV power time series data sets usually rely on data donations from PV owners, therefore, only consist of a limited number of sites. For example, the Microgen Database collects generation data from over 7000 PV systems across the UK, which is conditionally open-source [61]. Ausgrid data set spans a 3 year period and consists of half-hour load and rooftop PV generation for 300 de-identified residential customers [86]. The Solcast hosts a public website where users are able to report their PV power. The Solcast publishes a data set with 1287 sites. To the best of our knowledge, there is no such data set that consists of a large number of BTM PV power time series in the US. Three popular data sets have meteorological time series either with high temporal resolution or with large spatial coverage, which are the NREL National Solar Radiation Database (NSRDB) [64], the NREL Solar Radiation Research Laboratory (SRRL) data set [65], and the data set from University of California, San Diego (UCSD) [63]. Table 2 lists some characteristics of these open-source time series data sets. Additionally, there are open-source tools, such as the OpenSolar [87], to ease the burden of solar data access and processing. Other commercial services are also available to provide BTM solar data, such as the BTM solar output from Genscape.

2.5. Point versus probabilistic forecasting

Conventional solar power forecasting produces a single value (series), or the conditional expectation of solar power output at a time point in the future, and commonly denoted as a "deterministic" or "point forecast". However, prediction involves uncertainty and probabilistic forecasting can directly address that which is related to time and space. PV variability is higher for a small-scale PV system as compared to a utility-scale one [88,89]. For residential PV generation, it is potentially more useful to take a probabilistic forecasting approach and quantify output according to prediction interval or a probability density function (PDF); i.e., the human-driven loads are also uncertain and convolving it with BTM PV determines the net load for tracking or control. Fig. 1(b) shows a comparison between point and probabilistic approaches with prediction interval levels, for estimating power or irradiance. As can clearly be seen from the figure, the probabilistic methods have better performance in capturing the full range of the observed data, and thus have advantages in particular applications where large errors can have strong consequences.

Probabilistic BTM solar generation forecasting has been considered in [90] to predict PV power generation for building energy management systems. The method first uses a clear sky model (CSM) to generate a point forecast, then calibrates the model to real-time PV measurements to estimate system losses and partial shading. Following this step, a residual function between the PV measurements and the calibrated CSM is determined for eight categories of cloudiness, training an ensemble of regression trees. Based on the distribution functions, next the 24 h forecast is generated with 1 h resolution. In the final step, an error analysis of point forecast is used to generate the probabilistic forecast values by cumulative distribution functions. Authors have proposed an alternative up-scaling approach in [42], to prevent the large errors obtained by using NWP variables as the input; the supposition being that when the set of reference plants is small, the real characteristics or weather conditions of unknown plants cannot be reflected. Therefore the proposed method use statistical information of PV plant parameters. To estimate the occurrence of these parameters, a set of reference PV systems with 35,000 PV plants is used, after which the plants are binned based on the location and the peak capacity. The final power values are weighted based on the frequency of occurrence of the considered configurations. Although the study does not provide probabilistic forecasts, it uses probabilistic information by estimating the most probable value by averaging the power values weighted by their frequency of occurrence. The uncertainty caused by missing information on the installed PV systems has also been considered in [43], in which the proposed model using a linear approach to obtain the unknown parameters by regression techniques. The deviations from initial state are restricted by imposing constraints. A Bayesian approach is used to estimate the parameters of the system. A comprehensive review of the current body of work on probabilistic forecasting of solar generation and load consumption can be found in [27].

2.6. Evaluation metrics

BTM solar forecasting involves both classification and regression problems, therefore, evaluation metrics for both problems are critical. Specifically, BTM PV detection is a classification task with discrete output, while BTM PV disaggregation and forecasting are a regression task with continuous output. This section will describe the two suites of metrics used in the assessment of model performance.

2.6.1. PV detection model performance evaluation

PV detection is typically a supervised learning problem; i.e., the model will be trained with, and tested by, a data set composed of training and validation subsets. Using aerial image-based methods as an example, the PV detection task is a binary classification task that classifies the PV or non-PV pixels. Evaluating the accuracy of these binary classification models starts by collating the results into a 2×2 confusion matrix, as shown in Fig. 3 [91]. The four groups of label-detection pairs are true positive (TP) indicating correct PV prediction, false positive (FP) indicating wrong PV prediction, true negative (TN)

Table 2
Open-source time series data sets for BTM modeling and forecasting.

Name/Reference	Parameters	Spatial coverage	Resolution/length
Microgen [61]	PV power, meta data	7000+ locations	30 min/1 year
Solcast [62]	PV power, meta data	1287 locations	10 min/7 months
UCSD [63]	Meteorological variables, NWP	1 location	1 min/3 years
NSRDB [64]	Meteorological variables	US	30 min/25+ years
NREL/SRRL [65]	Meteorological variables	1 location	1 min/40 years

Table 3
PV detection evaluation metrics.

Metric	Formula	Summary
Accuracy	$\frac{TP+TN}{TP+TN+FN+FP}$	The overall correctness of the PV detection
Recall	$\frac{TP}{TP+FN}$	The proportion of actual PV pixels correctly detected
Specificity	$\frac{TN}{TN+FP}$	The proportion of actual non-PV pixels correctly detected
Precision	$\frac{TP}{TP+FP}$	The success probability of detecting a correct PV pixel
F1 score	$\frac{2Precision \times Recall}{Precision+Recall}$	A weighted average of the precision and recall

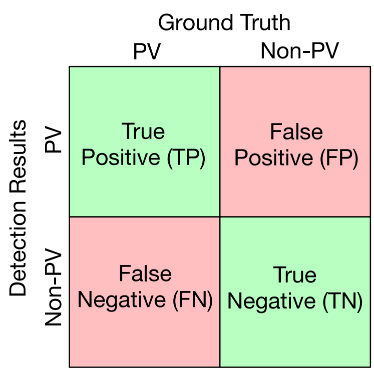

Fig. 3. A confusion matrix for PV detection.

indicating correct non-PV prediction, and false negative (FN) indicating wrong PV prediction. Based on the confusion matrix, a collection of metrics can be calculated and are usually used to quantify the PV detection model performance. Table 3 lists the definitions and summary of the confusion matrix-derived metrics [91]. The most straightforward metric is accuracy, which quantifies the overall correctness of the PV and non-PV pixel classification. However, PV pixels are sparse and rare in the aerial imagery in practice. That is to say, PV detection is a highly label-imbalanced classification problem. Imaging an area with much more non-PV pixels than PV pixels, as shown in Fig. 4, simply assigning all pixels to the non-PV class can achieve a 99.99% accuracy. This biased evaluation is common in most areas. Thus, there is a need to validate PV detection results from varying perspectives [92].

In addition to the crucial metrics of *accuracy*, there are other four widely-used confusion matrix-derived metrics. They are the *recall*, also known as the true positive rate or the sensitivity, the *specificity*, also known as the true negative rate, the *precision*, also known as the positive prediction value, and the *F1 score*, also known as the Dice coefficient. The value of these metrics ranges from 0 to 1, with 0 indicating the worst detection and 1 indicating the best detection. Among these five metrics, F1 is the only metric that measures the correctness from multiple perspectives. Other advanced classification metrics that are not widely-applied, but are valuable to comparable applications, include the Matthews correlation coefficient and the Jaccard index.

Most PV detection models determine the class through a detection probability and a threshold. For example, frequently the last layer in a neural network PV detection models use a sigmoid function or a softmax function. This output layer is a probability, between 0 and 1, and different thresholds lead to varying detection results. The receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) are performance measurements for the PV detection at various threshold settings [93]. The ROC curve is a probability curve that is plotted with the recall against the false positive rate, where false positive rate is the abscissa and recall the ordinate. AUC measures the degree of separability, indicating the models' capability in distinguishing between PV and non-PV classes. A larger deviation between the ROC curve and the diagonal line, and a larger AUC value (maximum AUC = 1) represents a better PV detection.

2.6.2. Forecast model performance evaluation

Once a method has been used to generate a solar power forecast, the quality of the method can be assessed using a validation data set of historical forecasts and observations. A framework for verifying point and probabilistic solar forecasts is then often employed. Generally, goodness of forecasts can be assessed by consistency, quality, and value, which were first proposed in the weather forecasting domain [94], then were introduced into the solar forecasting domain [95,96]. *Consistency* refers to the correspondence between forecasts and forecasters' judgement; *quality* refers to the correspondence between forecasts and the observations, which can be quantified by error metrics; and *value* refers to the benefit gained from the forecasts [94]. In power systems, value is mostly revealed by the economic or reliability benefits gained from the forecasts.

For point forecasts, the literature most commonly focuses on validating forecast quality and through error metrics, including mean bias error (MBE), mean absolute error (MAE), and root mean square error (RMSE) [97,98]. Definitions and formulations of point error metrics are extensively reviewed in [98]. One option – mean percentage absolute error (MAPE) – is controversial for solar applications. Engineers commonly rely on the MAPE to track dynamic percentage errors by considering the errors at every forecasting step separately, but percentage errors numerically explodes during periods with weak irradiance or power values, such as early mornings and late afternoons, so MAPE not recommended for solar applications [95]. To limit the impact of these outlier errors, it is common practice to use a standardized quality control filter, for example by removing all data point recorded with zenith angles above a threshold such as 85°. For a standard metric such as MAE or RMSE, a forecasting skill scores can show the improvement of a model over a baseline reference method [99]:

$$\text{Skill score} = \frac{\text{Proposed method score} - \text{Reference method score}}{\text{Ideal score} - \text{Reference method score}} \tag{1}$$

Of the three ideal characteristics a forecast, value is challenging and rarely assessed for any weather-dependent resource, including BTM PV. A value assessment likely requires down-stream power system simulations. For example, the 95th percentile of forecast errors was

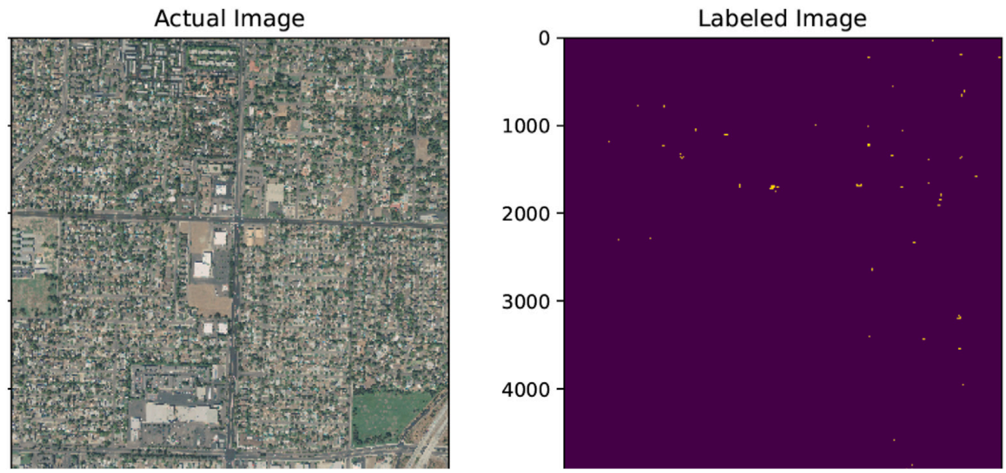

Fig. 4. A 2 km^2 urban area and its annotated PV systems. The yellow polygons are distributed PV panels. The non-PV/PV pixel ratio of this area is 10,000. This ratio will be much higher in rural areas.

used in Ref. [98] to quantify the operating reserves and characterize the solar forecast accuracy. Forecasters may consider over-forecasts and under-forecasts equally, but in practice solar over-forecasts and under-forecasts may have greatly different impacts. When considering utility-scale PV, a severe over-forecast can result in system reliability concerns, while an under-forecast can be simply resolved with increased curtailment. In a BTM context where PV systems are not curtailable, both over- and under-forecasts could cause reliability impacts if an under-forecast results in net load below the minimum output of must-run generators. Although it is hard to develop forecast value metrics that maximizes benefits of every power energy system individual, there is an emerging need and promising opportunity to establish methodologies for value-oriented metrics to guide solar forecasting. One example can be found in [100], in which the authors proposed a novel operational metric that assesses the lowest cost of operationally delivering perfect forecasts. This cost reflects the amount of solar production curtailment and backup storage required to rectify all over/under-prediction scenarios.

When considering a probabilistic forecast, verification commonly employs graphical diagnostic tools – natural for spread, comparative information formats – conveyed in quantiles or distributions. The three key concepts in probabilistic forecast verification are reliability, sharpness, and resolution [101]. The *reliability*, sometimes termed calibration, refers to the accuracy of forecast probability in correspondence to the true probability of a forecast occurring. The *sharpness* refers to the concentration or narrowness of the predictive distributions. The *resolution* indicates how well the probabilistic forecasts distinguish among different situations; resolution is most nebulous of the three and is typically assessed in concert with the other two. Most probabilistic verification tools are developed to quantify or visualize at least one of the three characteristics of probabilistic PV forecasts, and a framework of probabilistic solar verification tools examined [96] and reference methods employed [102].

Reliability of the probabilistic PV forecasts is the first priority to power system operators due to the importance of contingency preparedness for potential, critical tail events. To visualize probabilistic PV forecast reliability, a probability integral transform (PIT) histogram or a reliability diagram or quantile–quantile (Q–Q) plot are most widely used. A PIT histogram bins the frequency with which different bins of the cumulative distribution function (CDF) are actually observed, where a properly calibrated forecast has a uniform PIT histogram [96]. As illustrated in Fig. 5(a), a PIT histogram be used to visually diagnose reliability deficiencies, such as a ∪-shape indicating under-dispersion (forecast intervals too clustered) or a ∩-shape indicating over-dispersion (forecast intervals too broad). A reliability diagram is very similar, except it plots the observed frequency for each quantile instead of using bins; i.e., the distributions for perfectly reliable forecasts are plotted along the 45-degree diagonal.

Sharpness is also important to power systems, especially for their economic operations — it is more useful for making cost-effective decisions, but only if the forecast is reliable [101]. Sharpness is quantified by the mean size of the central prediction intervals for different nominal coverage rates [96]:

$$\overline{\delta}_\rho = \frac{1}{N}\sum_{n=1}^{N} F_n^{-1}\left(1-\frac{\rho}{2}\right) - F_n^{-1}\left(\frac{\rho}{2}\right), \tag{2}$$

where ρ indicates the central interval of interest (i.e., $\rho = 0.9$ for 90% prediction interval), N is the number of points in the validation set, and F_n^{-1} is the inverse CDF at time n. This is a function of the forecasts only, not the observations, and so has no ability to assess the forecast's calibration. Smaller values always indicate better sharpness. Fig. 5(b) [103] shows an example sharpness diagram (i.e., δ-diagram), which visualizes the average sharpness across multiple choices for ρ.

In practice, all three characteristics can be numerically assessed in a single proper metric, the continuous rank probability score (CRPS). For a single time point, the CRPS is formulated as:

$$\text{CRPS} = \int_{-\infty}^{+\infty} [\hat{F}_i(x) - F_i(x)]^2 dx, \tag{3}$$

where $\hat{F}$ is the forecast CDF and F is the CDF of the observation — a step function that jumps from 0 to 1 at the point of observation [99,96]. Fig. 6 visualizes CRPS, which integrates the squared difference between these two functions to captures both reliability and sharpness. Typically, CRPS is assessed on average over all time points in the validation set N. To look at certain areas of interest, average CRPS can be decomposed into the Brier score at individual power levels (less relevant for solar forecasting due to the diurnal trend in expected power) or the quantile score or pinball loss at individual quantiles [104,102,105]. The literature also contains relics of other metrics that have ambiguous interpretation and use has been subsequently discouraged [106,96]; e.g., prediction interval coverage probability (PICP), prediction interval normalized averaged width (PINAW), coverage width-based criterion (CWC).

3. Capacity and specification estimation for BTM PV

The integration of massive quantities of solar PV systems on distribution networks can result in technical challenges in both planning and operations. For instance, most interconnection standards do not require remote monitoring and control of inverters, which can lead

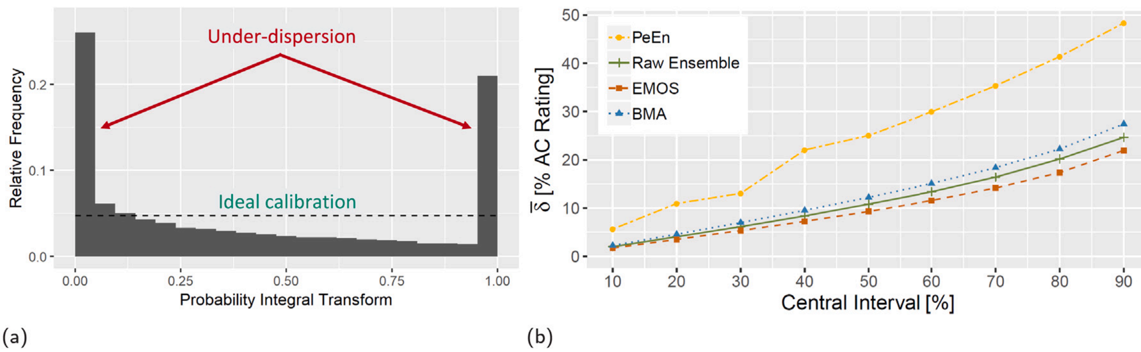

Fig. 5. Examples of the 5(a) PIT histogram and 5(b) sharpness diagram. The PIT histogram can illustrate regular bias and under- or over-dispersion when the shape deviates from the ideal uniform distribution. A sharpness diagram here demonstrates that one of four example forecasting methods (persistence ensemble, PeEn) is on average much less sharp than the other three. 5(b) is reprinted, with permission, from [102].

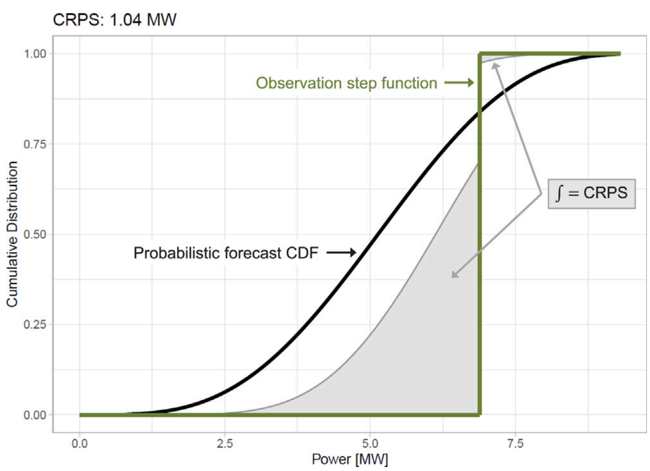

Fig. 6. For a single time point, the CRPS is derived from the mismatch between the forecast CDF (black) and the observation CDF (green), which is simply a step function at the observation — here, 6.8 MW. The squared difference between the two functions ($[\hat{F}_i(x) - F_i(x)]^2$) defines the area (gray) that constitutes the CRPS.

to additional operational challenges [107]. Additionally, the effect of the each PV system on the distribution system is highly dependent on its installation configuration, condition, and location. Therefore, it is clearly important for the operator to know the capacity of installed BTM solar PV [108]. However, by definition, BTM solar PV panels cannot be monitored by the utilities, and often even their basic capacity information is unknown. In some cases although this information may be publicly available, it may not be reliable (or available to the correct utility employees) as some clients do not follow the utilities' regulations for registrations or, once registered, expand the installed capacity without notice.

BTM PV can have significant impacts on load profiles at the building, distribution, and transmission system levels. The distribution system operator may only notice slight, continual changes in the electricity demand profiles, or electricity revenue streams, but cannot immediately determine the capacity of BTM solar PV systems integrated in the network due to the existing uncertainties in load patterns. In terms of demand response capacity estimation, demand response aggregators must determine how much demand response capacity they have in order to develop a realistic bidding strategy in the power market. The BTM PV deployments have a considerable impact on load profiles and, as a result, on the demand capacity available. Significant errors will appear in demand response capacity estimation if the actual BTM PV capacity information is unknown. Similarly, lack of knowledge about the installed capacity of BTM PV will also result in biased load prediction results and higher errors in net load forecasting [32]. Therefore an important phase in BTM PV generation forecasting is to include a capacity and specification estimation procedure for determining the condition of the network at any particular time. Some researchers focus on predicting the aggregate output power or capacity of all BTM solar PV panels in a specific region [109,110,9,111]. Fewer works consider estimating the output power or capacity of individual BTM solar PV panels [112,13,113,32,108]. Despite the higher accuracy of the latter approach, it is difficult to detect the capacity of every PV installation without detailed meter or interconnection data. The capacity and specification estimation models are generally used for distribution and transmission scale aggregate estimates. In some cases, primarily at the transmission scale, publicly available metadata sets containing PV system power measurements and time steps can be found for research purposes [62]. In the case of a lack of any data, the models that can be used for estimation of capacity and specification can be classified as satellite aerial image-based methods and time series disaggregation methods. These two groups are reviewed below. Summaries of the publications reviewed in detail pertaining to satellite aerial image-based capacity and specification estimation, disaggregation-based methods are available in Tables 4 and 5, respectively.

3.1. Satellite aerial image-based methods

One non-intrusive approach to derive the distributed PV capacity is detecting PV arrays from satellite aerial images, an example of which is illustrated in Fig. 7. In the computer vision field, it is a image classification problem. Satellite images with PV labels are first used to train machine or deep learning models, which are used to detect rooftop PV systems in images without annotations. PV capacity can be derived from the size and orientation of the PV arrays detected within the images. The recent development of machine or deep learning and remote sensing techniques makes it possible to accurately detect PV panels over the large geographic areas necessary for transmission-level estimations. Like most machine learning applications, shallow learning algorithms were first used in conjunction with feature engineering in the satellite image processing. For example, building a support vector machine (SVM) model to learn from the color, shape, and texture of 100 aerial images to detect PV objects [114]; the SVM classifier was able to identify the PV object with a 94% accuracy. However, the case study was conducted on a small data set and the false positive rate was high. The same authors optimized the features used in shallow learning models, including raw pixels, local color statistics, and textons (fundamental micro-structures in natural images) to detect the PV pixels [115]. Random forests (RF), a more advanced classifier, detected PV arrays with a better ROC curve. However, other accuracy metrics were not reported. Two other features, the pixel means and variances of different sizes of windows, were used as input to RF models for PV

Fig. 7. A satellite aerial image-based PV detection procedure. PV panels are detected by well-trained machine/deep learning models in a small image patch that is extracted from a large image patch around the National Renewable Energy Laboratory campus. PV capacity can be derived from the detected images based on the pixel area and the roof tilt angle.

Table 4
Satellite image-based PV detection articles.

Reference	Method	Input	Output	Evaluation
[114]	SVM	Aerial image	PV object	Accuracy=0.94
[115]	RF	Aerial image features	PV pixel	ROC curve
[112]	RF	Aerial image features	PV pixel	Accuracy=0.90
[118]	CNN	Aerial image	PV object	Accuracy=0.90
[116]	CNN	Aerial image	PV object	Accuracy=0.87
[117]	CNN	Aerial image	PV object	Precision=0.93
[119]	CNN	Aerial image	PV object	Precision=0.95
[120]	CNN	Aerial image	PV pixel	Precision=0.94
[121]	CNN	Aerial image	PV pixel	ROC curve
[85]	CNN	Aerial image	PV pixel	Precision=0.93
[122]	CNN	Aerial image	PV object	Precision=0.91

SVR: support vector machine; RF: random forest; CNN: convolutional neural network. The difference between image features and images is that the former one needs additional feature engineering, but the latter one can be directly input into the detection models. The difference between PV object and PV pixel in the output is that the former one can only provide the existence of the PV panels in a image, but the latter one can provide the location and the shape.

pixel detection with a 90% detection accuracy. It is not clear if the model could be applied to a larger spatial-scale with the same accuracy.

Satellite aerial image-based PV detection accuracy has been significantly improved by deep learning techniques. Deep learning models learn patterns and features from images automatically without heavy feature engineering. Convolutional neural networks (CNNs) have been used [116] to detect PV panels in 670 Google Maps images with 87% accuracy. However, the case study was based on an unrealistic assumption that PV arrays had a 50% chance to appear in the images. In practice, PV detection is an extremely unbalanced classification, where PV objects and pixels are sparse and rare. This is the biggest challenge in satellite image-based PV detection [117]. On the other hand, some have tried to reflect the label imbalance nature of PV detection but resulted in low detection accuracy: CNN model to detect PV panel objects over a 135 km^2 area with 80% true positive rate and 72% precision [118]. In another example, the recall was sacrificed to provide PV detection with high precision [119].

Efforts have sought to improve the adaptability of the detection models. For example, transfer learning has been used to leverage the weights trained with the other image data set for fine-tuning [119]. However, the pre-trained network did not improve the PV detection accuracy. The PV detection accuracy also varied significantly between different geographic regions. For example, a CNN-based network yielded PV detection with an average 76% precision, 77% recall, and 76% F1 score in three cities in California [123]. However, the same architecture's detection in two municipalities in Connecticut had an average 88% precision, 83% recall, and 85% F1 score. Compared to utility-scale PV detection (i.e., 93.76% precision, 92.18% recall, and 92.55% F1 score), the accuracy of the distributed PV detection was noticeably lower [120]. One recent work developed a deep learning framework, called the SolarForecast, to automatically localize solar PV panels in the U.S [85]. The CNN model was trained with 366,467 images sampled from over 50 cities/towns across the U.S., which achieved 93.1% precision in residential areas and 93.7% precision in non-residential areas.

Only a limited number of studies have taken the next step to derive distributed PV installed capacities after detecting PV arrays from satellite images [85]. To simulate distributed PV power time series, more parameters, including azimuth, shading, and irradiance time series should be provided. While there are some open-source meteorological data sets available, other parameters are always unknown, which makes distributed PV forecasting more challenging. With the prevalence of open-source research and data, more efforts and attention will likely be focused on this field.

Another related but distinctively different topic is solar technical potential estimation [124]. Most research in this topic relies on light detection and ranging (LIDAR), geographic information system data that contains building footprints or land-use data to derive capacity potentials. Since there is no PV information involved, the derived PV capacity potentials cannot represent the actual capacity of installed BTM PV systems, but rather an upper bound on that capacity. Therefore, this research topic is beyond the scope of our research.

3.2. Disaggregation methods

A family of methods to obtain the capacity and specification data for BTM PV is disaggregation of the net load readings from advanced measuring infrastructure (AMI) meters. A summary of studies using disaggregation for capacity and specification estimation is provided in Table 5.

AMI meters are starting to be widely used and it is anticipated that the number of such meters increase exponentially [128]. The readings obtained from an AMI meter can be used to estimate the generation output of the BTM PV solar system and data related to its capacity and specifications. Some studies require additional data together with net load readings, for instance weather related data; e.g., GHI, air temperature and weather condition classification [125,113,32,108]. In [125], a model based on historical and meteorological data is

Table 5
Publications considering disaggregation approaches.

Reference, Year	Disaggregation method	Forecast horizon	Input data	Location	Performance measures
[125], 2015	PV plant model	Short-term	PV plant coordinates Module tilt and azimuth angles Global horizontal irradiation Air temperature Data from two PV plants 5 and 15 min intervals	Switzerland	Simulation error
[13], 2016	Change-point detection algorithm Permutation test with Spearman's rank coefficient	Short-term	Electricity consumption 40 customers 15 min interval	A city in U.S.A.	Correlation strength analysis
[113], 2017	Clear sky generation model Universal weather-solar effect using machine learning	Short-term	Building's location (latitude and longitude) Minimal amount of historical net meter data 100 solar power buildings 1 h interval	N/A	MAPE
[9], 2017	Maximal information coefficient based correlation analysis Correlation analysis based on copula theory	Weekly	Actual load data PV output Ambient air temperature Estimated PV panel temperature 1 h interval	ISO-NE zone Maine, New England, U.S.A	MAPE, RMSE
[32], 2018	Support vector classification based detection model Support vector regression based capacity estimation	Short-term	Electricity consumption PV output power data Net load data 183 customers 1 h interval	Austin/Texas, U.S.A.	MAPE, R^2
[108], 2019	Machine learning PV load decoupling	1 week	Electricity consumption 300 customers 1 h interval	Sydney, Australia	MAPE
[126], 2019	Physical PV system performance model Hidden Markov model regression	1 month	Pecan Street data set Net load Customer load Solar PV generation data The solar irradiance and weather data 197 customers 15 minute interval	Austin/Texas, U.S.A.	MSE, MASE, CV
[127], 2020	Deep neural network	1 year	Pecan Street data set 1300 customers 1, 5, 15, 30, 60 min intervals	Pecan Street data set	MAPE, MAE

MAPE: Mean absolute percentage error; MSE: Mean squared error; MASE: Mean Absolute Scaled Error; MAE:Mean Absolute Error; CV: Coefficient of variation

proposed to estimate PV system parameters such as: module azimuth and tilt angle, power curve, etc. The method first simulates the PV system power output and then tries to minimize the differences in the parameters between the observed and the simulated PV output. Performance of the model is tested by using measurements from two PV plants. The average tilt and azimuth estimation errors were found to be 0.75 degrees and 4 degrees, respectively. A black box method, called SunDance is proposed in [113] for disaggregating solar generation from net load data. In addition to the historical meter data, the model uses the location of a building and a CSM to estimate the irradiance. The proposed model has two key modules, a clear sky solar generation module which use historical net energy meter data and a module to map multiple weather metrics to the expected percentage reduction in clear sky solar irradiance potential. SunDance is evaluated by using metered net load data for 100 buildings and the results show that its accuracy is acceptable without access to any solar training data from a deployment. According to the findings, the errors for the estimated tilt and azimuth for one of the buildings in the study is found to be 1 degree and 1–5 degrees, respectively. One-class support vector classification (SVC) based model is proposed in [32] to detect BTM solar sites. The proposed model then estimates the capacity of the detected solar panels by using a bootstrap-support vector regression (SVR) model. The algorithm is tested on a realistic data set from 183 residential customers and according to the results, the value of MAPE is found to be between 5%–7% and the value of R^2, between 0.86–0.92. Machine learning has been used to disaggregate the output power of each individual BTM solar PV system from net load data [108]. The algorithm starts with the capacity estimation using a multiple support vector regression-based ensemble model. In the second stage, the calculated capacity is multiplied by the output of a standard distributed solar system to obtain the forecasts of each individual system.

In contrast to the aforementioned methods, some studies have considered only AMI meter data as an input without additional data requirements. In [126], the authors proposed an integrated approach to disaggregate net load data and to estimate the solar PV system technical parameters. The proposed method combines a physical PV system model and a statistical load estimation model based on Hidden Markov regression. Results show that the proposed approach outperforms the disaggregation algorithm proposed in [113] and reduces the mean square error by 44%. In [127], only AMI data is used within a deep neural network approach for estimating PV size, tilt, and azimuth angle. The authors reported a mean absolute percentage error of 10.1% and 2.8% for the estimates of PV tilt and azimuth, respectively.

Time series approaches have been also used for disaggregation of net load. In [13] a three-step data-driven approach is proposed to first detect and then estimate BTM solar installations. Unauthorized PV installations are identified by focusing on abrupt changes in time series data using change-point detection. This step is followed by a statistical test to verify their existence. The last step combines the AMI meter data with local cloud coverage data in order to estimate BTM

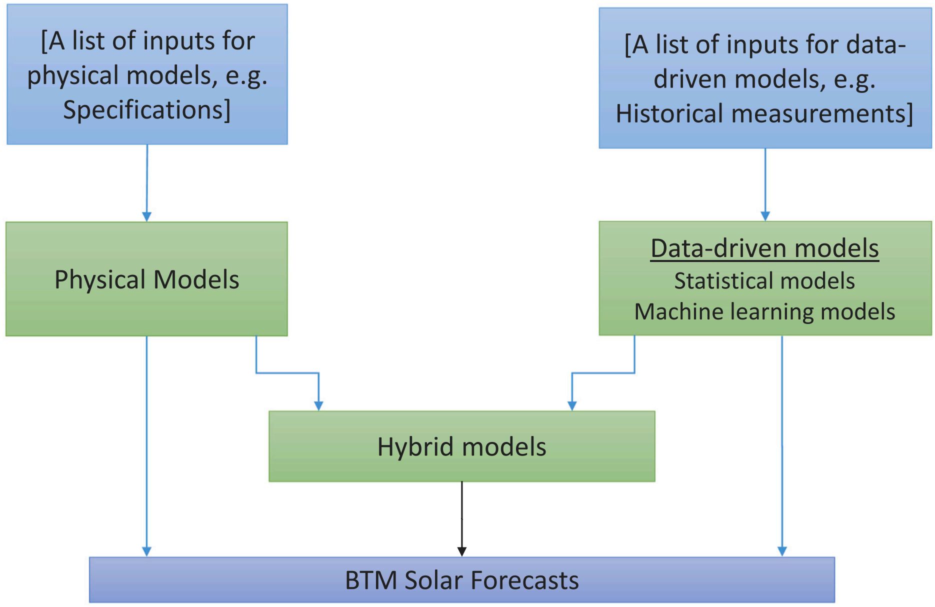

Fig. 8. BTM PV power estimation models.

PV parameters. The proposed method is evaluated by using actual AMI data under different scenarios. Based on the results, when local cloud coverage information is included, the error in the PV size estimation is found to be approximately 4% for one 5 kW PV installation. However when no cloud data is included, the error for the same 5 kW PV system is found to be 44%. Further studies have proposed methods based on the disaggregation of net load for forecasting BTM solar PV system generation, for systems with known capacity and specifications [111,129,6,130–133].

4. Forecasting methods for BTM solar generation

The methods commonly used in BTM solar PV forecasting are discussed in this section, including: physical, data-driven and hybrid models (Fig. 8). Both data-driven and physical models can be used for irradiance and BTM solar generation forecasting. When it comes to data-driven models, the approaches used for irradiance and power generation are similar as there is a strong positive correlation. However for physical forecasting, a PV system model is still required to convert the irradiance forecasts obtained from NWP. Extensive review articles regarding physical and data-driven models for solar irradiance forecasting and utility-scale PV power forecasting have been published in recent years [21,20,134]. Additionally the performance of 12 different forecasting models that forecast the day-ahead power production, such as regression, support vector regression, ensemble learning, deep learning and physical based techniques are compared in [135], as well as the effect of aggregating PV systems on the forecast model performance. As the main focus of this review is not on those methods this section aims to summarize the main features and key publications.

4.1. Physical models for BTM solar PV power forecasting

Physical models can be defined as theoretical simulation models which describe the physical relationships of a PV system. A physical model can combine the introduced meteorological data together with a PV system model in order to estimate PV power output. These models are mainly based on the main design parameters of the PV system where historical data is not required. If the system data is available, they can be used for converting irradiance forecasts to power output forecasts effectively [136]. However, the accuracy of the output degrades with sharp changes in meteorological variables [137]. Physical models can be very simple or highly complex, based on what input data is introduced to the model. These models are highly dependent on the input weather model performance, particularly if the weather is cloudy. In such a case, insufficient spatial and temporal resolution can result in high errors [20]. On contrary, the physical model itself is often the main source of error when combined with clear sky models [136]. In order to overcome this drawback, researchers have proposed various methods. In [50], an approach focusing on improving PV power estimation during periods of snow cover is proposed for 11 sites in Germany. The authors further proposed a method for regional PV point forecasts of up to 2 days-ahead using physical models and upscaling techniques [41]. Another approach is proposed in [52] to analyze forecasted variability in PV power generation due to clouds. The input data is obtained from measurements of 80 residential rooftop PV systems. A framework to model station-pair correlations of irradiance variability is proposed and optimal locations of irradiance sensors are recommended. Other studies using physical performance models for BTM PV generation are: [138,52,139,140]. As the focus of this article is not on physical performance models, we refer the reader to [136] for a more detailed review.

4.2. Data-driven models for BTM PV power forecasting

Data-driven methods are based on historical PV power output data and can be broadly categorized to include statistical and machine learning methods. A comprehensive review of data-driven models for solar forecasting can be found in [22].

4.2.1. Statistical models

Statistical models have been widely used for forecasting PV power output. Some of the well known statistical models can be defined as autoregressive moving average (ARMA), autoregressive moving average with exogeneous inputs (ARMAX), regression and exponential smoothing. ARMA works with stationary time series and can be used in non-linear systems to a certain degree. ARMAX is an improved form of ARMA that incorporates meteorological variables as exogenous inputs [46]. In recent papers, regression trees also have been widely employed for boosting and bagging purposes [141].

In statistical models, system design parameters are not required and historical irradiance and production data-sets can be enough to generate forecasts. These models are most commonly used for forecasting residential solar generation. A series of related studies considered representative sites and estimate the regional generation based on selected sites in California [109,110,14]. Instead of being purely data-driven, a method combining historical data with weather forecast data can be found in [142]. A data-driven inference model, built on a Bayesian network, is developed for a very short-term PV generation forecast (less than 30 min) with minute-level resolution.

4.2.2. Machine learning models

Recently many researchers have proposed different methods to forecast time series values based on machine learning methods. Machine learning techniques investigates and develops algorithms for learning from data, making estimations, and improving forecasts. Integrating machine learning approaches in grid analysis tools can be helpful to address data and operational problems that arise as with the increased penetration of BTM PV systems. The reviewed literature shows that ANN have been successfully applied for forecasting BTM PV energy supply due to its robustness and strong inference capabilities. These methods learn to recognize patterns in data using training data sets. Generally speaking, these models require historical data about weather forecasts, real power production, and environmental quantities, which could be seen as the main drawback. High dependence on prior knowledge, requirement of large amounts of data during training together with multi-layered structure may result in increased complexity.

Support vector machine (SVM) is a modern and reliable approach for non-linear BTM solar forecasting. Different from ANN, SVM has no local minima problem and it does not rely heavily on prior knowledge. However, parameters have a big impact on SVM, and accurate parameter selection is required [143]. In ELM, linear regression can be used to choose input weights and hidden node biases at random. A review on machine learning techniques can be found in [144].

Despite the challenges defined related with machine learning approaches, in [145] two machine-learning methods are proposed to estimate rooftop solar power from publicly available weather forecasts where only the PV system's geographic location and a small amount of past output data are required. A machine learning based approach for day-ahead forecasting of aggregated PV power generation in Australia is proposed in [146], with 30 min. time interval. Five machine learning and deep learning algorithms are analyzed together with two alternative feature sets as inputs to the forecasting algorithms. In [147], machine-learning approach is used for regional PV output forecasting with hourly resolution for up to two days ahead. The physical PV power model is summed for a whole region after being aggregated by geographical clusters. Parameters such as, irradiance, temperature, barometric pressure, and wind speed, which are utilized as inputs to calculate PV output power and irradiance, are forecasted with NWP model. Following the machine-learning approach, linear regression method is used to account for bias in computed power. The available historical power generation data along with NWP inputs are used within a machine learning approach in [14]. Fuzzy Arithmetic Wavelet Neural Networks (FAWNN) are used to develop the forecasting engine, providing fuzzy prediction intervals for any desired level. Another advanced hybrid deep learning-based technique for estimating daily PV power generation in 30 min. interval is proposed in [148]. In order to learn spatio-temporal patterns in complicated time series data, the approach integrates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models in an encoder–decoder architecture, and only prior solar power data is used as input in the prediction process.

It should be noted that the accuracy of the statistical forecasts highly depends on the length and quality of the historical data set, even in the most advanced deep learning methods. Training data sets longer than 1–3 years tend to increase forecasting accuracy [136].

4.3. Hybrid models for BTM solar PV power forecasting

In order to avoid the disadvantages of physical and statistical models, the best option is often to use hybrid models [149]. Any combination of two or more of the previously described methods leads to a hybrid model. The idea is to combine different models with unique features to overcome the negative performance characteristics of a single model and to improve the final forecast. In [150], a weather-based hybrid method including classification, training, and forecasting stages is proposed for 1 day-ahead hourly forecasting of PV power output. The forecasting phase is made by using a fuzzy inference method to select an appropriate trained model. In other research [151], a hybrid method comprised of a genetic algorithm with a neural network approach is used in which a genetic algorithm is used to adjust the weights and thresholds of the neural network. The main role of the genetic algorithm is to increase the accuracy of the main model.

Machine learning techniques combined with various physical approaches, such as NWP, clear sky, satellite imaging were proven to generate improved forecasting outcomes for long-term time horizons [152,153]. The hybrid models can be computationally complex as the number of strategies it incorporates grows. A balance between accuracy, computing complexity, and cost has to be ensured in order to get better performance in forecasting accuracy [144]. In comparison to any other individual technique, hybrid models are can be defined as efficient models when all of these limitations are kept within limits. More information on hybrid models can be found in [19]. Additionally, a complete review on the recent applications of artificial intelligence techniques; focusing particularly on machine learning, deep learning, and hybrid methods can be found in [154].

5. Application for BTM solar PV generation forecasting: net load forecasting

An important application area of BTM solar generation at transmission scale is net load forecasting. Net load forecasting is an important application area for BTM solar power at the transmission scale. They are frequently used in unit commitment and economic dispatch models. This section reviews the available literature on net load foresting and the summary of the papers is given in Table 6. Net load refers to the total system load, less the amount of load provided by the intermittent variable renewable sources. In a case where the intermittent energy source considered is solar PV, the net load refers to the total system load, less the demand met by solar PV. As discussed in Section 1, the penetration of BTM PV generation has a significant impact on net load. The relevance of addressing BTM PVs in conjunction with load has gained significant attention in recent years. Therefore, detailed modeling of BTM solar generation is essential to achieve high net load forecasting accuracy.

Forecasting net load described in the literature primarily utilizes two approaches, integrated and additive. In the integrated approach, historical data for the two sub-components of net load (PV generation and load data) are formed and combined and the resulting time series is considered as the forecasting input for the corresponding model [159,33,158,161]. In the additive approach, PV generation and load data are individually estimated and the results are used to perform net load forecasting [155,9,157]. In [159], a very short-term hybrid forecasting model is proposed by integrating phase space reconstruction and deep neural network (DNN) methods. The net-load time series are firstly reconstructed using the phase space reconstruction method. This method helps to better reflect the dynamic characteristics of the bus load, then the reconstructed data is fitted by the DNN to obtain the predicted value of the net load. A similar approach is also used in [33] for predicting solar-integrated, utility-scale feeder net load in the San Diego Gas & Electric operating region. The proposed hybrid model, consisting of ANN and SVR models, forecasts net load over a 10 to

Table 6
Publications considering net load forecasting.

Reference, Year	Forecast method	Forecast horizon	Input data	Location	Performance measures
[155], 2016	Clear sky model Autoregressive model Autoregressive model with exogenous input Support vector regression	Short-term	Heating Ventilation and Air Conditioning (HVAC) load PV power generation Load demand from grid 1 h interval	University of California, U.S.A	MAPE, MBE, MAE, RMSE
[9], 2017	Maximal information coefficient based correlation analysis Correlation analysis based on copula theory	Weekly	Actual load PV power generation Ambient air temperature Estimated PV panel temperature 1 h interval	ISO-NE zone Maine, New England, U.S.A	MAPE, RMSE
[33], 2017	Artificial Neural Network Support Vector Machine	10, 20, and 30 min	Actual load Solar penetration Feeder length 10 min interval	San Diego Gas & Electric utility company, U.S.A	MBE, MAPE, rRMSE
[156], 2017	Dynamic Gaussian process Quantile regression	Seasonal	Residential electricity consumption Rooftop PV power generation 30 min interval	Sydney, Australia	PINAW, PICP, CRPS
[157], 2018	Gumbel copula based joint probability distribution Grey index model	Seasonal	Actual load Wind and Solar generation 5 min interval	Bonneville Power Admin., U.S.A.	MAE, PE, MAPE
[158], 2019	Bayesian deep learning	Day-ahead	Actual load PV power generation 30 min interval	Sydney, Australia	RMSE, MAE, NRMSD, MAPE, Pinball, Winkler
[159], 2019	Phase space reconstruction Deep neural network	Ultrashort-term	Net load 5-min interval	N/A	RMSE, MAPE
[160], 2020	Artificial neural network Multilayer Perception	Day ahead	Net load Irradiation 1 h interval	N/A	MAE
[161], 2020	Gradient boosting machine Support vector machine	Day ahead	Net load Irradiation 1 h interval	Texas, U.S.A	nMAE, nRMSE, MAPE
[162], 2020	Deep neural network Wavelet transform	12, 24, 48 and 168 h	Meteorological data Wind and solar generation Actual load Energy price 1 h interval	Germany	MAE, MAPE, RMSE

MBE: Mean bias error; RMSE: Root mean square error; PINAW: Prediction interval normalized average width; PICP: Prediction interval coverage probability; CRPS: Continuous ranked probability score; PE: Persistence ensemble; NRMSD: Normalized root mean square deviation; Pinball: Pinball loss function; Winkler: Winkler score

30 min horizon for four solar-integrated, utility-scale feeders in Southern California. combining ANN and SVR models, forecasts net load for four solar-integrated utility-scale feeders in Southern California over a 10 to 30 minute horizon. The proposed hybrid method performance is improved by removing low-frequency load variation due to human activities and incorporating sky image features into daytime forecasts after separating training between daytime and nighttime. In [161], day-ahead, data-driven feeder-level net load forecasts in northern Texas are generated by considering weather forecasts and BTM solar PV generation. The generation of BTM PV is estimated by using the PV penetration and forecasted solar irradiance within a SVR model. The estimated BTM PV output of the feeder is then used together with forecasted weather to train extreme hot/cold load forecasting models by using a gradient boosting machine (GBM) machine-learning method. Finally, in [158], a probabilistic model based on Bayesian probability theory and deep learning is applied to time-series aggregation after clustering consumers into two groups with "visible" BTM systems (i.e., specifications are known) and "invisible" ones. The main reason for clustering is to enhance the performance of the aggregated net load forecasting, by providing a deep learning model for each cluster. Several visibility levels are employed to analyze the effectiveness of the clustering approach. An additive approach is used in [9], for addressing the question of net load forecasting for high penetration scenarios of BTM solar PV systems. First, the capacity of BTM PV is estimated by a maximal information coefficient based correlation analysis and a grid search. The measured net load profile is then broken down into three parts which include solar PV output, the residual (net load minus estimated actual load, plus estimated PV output), and the actual load. Afterwards, in order to obtain a probabilistic net load forecast, correlation analysis based on copula theory is conducted on the distributions and dependencies of the forecasting. Another application of additive approach can be found in [157], in which the wind and PV outputs are considered within a net load forecast model. The proposed model uses a Gumbel copula to model the joint probability distribution of load, wind and solar generation forecasting errors. First, the correlation of meteorological parameters, including humidity, wind speed, and temperature with renewable generation and load is calculated. Then, both generation and load are predicted to obtain final net load forecast. Similar to [157], authors in [162] also consider solar and wind generation uncertainties within the net load forecast, but also include the effects of energy prices on the load forecast. The authors proposed a deep neural network approach combined with the wavelet transform to increase model sensitivity to the variations of the input data.

A comparison of additive and integrated model is performed in [155], focusing on the significance of net load forecasting for operation and management for microgrids with high PV penetrations. Solar power is forecasted using SVR based model and an adaptive clear sky model. Load demand is forecasted by different approaches, including auto-regressive models with or without exogenous input, as well as SVR. For the additive forecasting approach, solar power and load demand are forecasted separately and then combined to obtain net load. In the integrated approach, solar power is used as an input to net load forecasting model. The results showed that integrated model outperforms the additive model in terms of all error metrics. The effects of aggregation of customers and increased penetration of PV power on prediction intervals are analyzed in [156]. Authors proposed a dynamic Gaussian process and quantile regression to produce probabilistic forecasts on data obtained from the area of Sydney, Australia. Additionally, seasonal effects are examined on the results. Finally, in [160], an ANN-based model is used to predict short-term high-resolution residential net load profiles. A case study of a real neighborhood including 75 single-family houses is used to evaluate the performance of the model.

6. Conclusions and future research directions

BTM PV monitoring, forecasting and control has the potential to significantly impact power system operations. Specifically, high solar PV penetrations lead to supply and demand imbalance, resulting in erroneous net load forecasts. Accurate and robust forecasting of BTM solar generation is essential to mitigate the consequences of these imbalances in a cost effective manner and to maintain system reliability. Forecasts of BTM PV are needed by multiple stakeholders in the power system and can be classified – from the customer scale to aggregate, regional BTM generation – with top-down approaches focused on bulk system and utility operations, while bottom-up approaches seek to scale detailed, hierarchical models. However, precise estimations of installed BTM PV capacity are challenging in many locales, as the generation data from BTM systems is only partially recorded or completely missing. Additional estimation techniques such as capacity and specification estimation are often needed to produce a reasonable BTM PV forecast; i.e., innovation is required for this critical step in the BTM forecasting algorithms. Ignoring capacity estimates, particularly in the case of high BTM PV penetration could have impacts on the following technical challenges: matching, safety, uncertainty, and system adequacy. A difference in PV capacity information causes a prediction error of solar power generation amount by area (reduces prediction accuracy), making it harder to compute the correct power system hosting capacity. Lack of capacity information might also result in a variety of safety issues, such as over-voltage and back-feeding, which can harm system facilities.

The growth in BTM PV installations is likely to continue as PV panel prices continue to decrease and distributed storage installations are following similar trajectories. To ensure their continued interconnection and the maintenance of power system for reliability standards, impacts on the overall system should not be ignored and explicitly considered in operational analysis. Moreover, BTM storage technologies, make it possible to use stored energy another degree-of-freedom. Directly controlling the customer's net load profile, especially when a rooftop PV system is coupled with a battery, requires the capacity of the battery to be jointly estimated. The literature on the inclusion of distributed storage technologies in bulk power system operations is nascent, with BTM battery capacity estimation only considered in [10], in which local solar irradiance information is used to estimate both the hidden rooftop solar PV generation and hidden battery capacity in buildings.

The most critical and challenging factor of the general BTM problem(s) is the prediction accuracy necessary; i.e., for enhanced BTM solar PV system integration, which has direct and tangent applications alike dependent on whether the reference signal, for example, is a PV inverter status or instantaneous neighborhood-level (target) load. An increase in estimation accuracy can be achieved by combining strengths of various methods through hybrid approaches; e.g., stacking and blending machine learning models. Hybrid approaches can then perform better and result in reduced errors. Because of the stochastic nature of solar generation and uncertainties in data for BTM solar PV system, the accuracy level of power estimation is generally poor. The literature explicitly focused on uncertainty quantification within BTM systems is immature. Probabilistic approaches could be an effective means to provide a better representation of the underlying uncertainties. The combination of probabilistic approaches with hybrid methods could thus help optimize the stability and safety of grid operations in both practice and counterfactual hedging.

Another critical factor inherent to the BTM problem, only touched upon with this review, and could be a potential future research direction is the implications for next-generation demand response. Whether load shaping and modulation is beneficial to the smart grid in question, likely coupled to grid-tied batteries, is constrained by (1) the computational effort of expanding degrees-of-freedom in the controllability of loads, and (2) the implied control over humans' demand (or desire) for electricity during critical events or volatile periods. Coordinating all the methods, techniques, and considerations of this article with human-in-the-loop activity is not trivial. Next generation demand response programs, perhaps relying on real-time marginal emissions rates, seek to optimize load-side controls but this entails evermore degrees-of-freedom for hierarchical controls or other cascading logic. Clearly, a balance between the top-down (e.g. weather) and bottom-up data (e.g. residential BTM PV panel specification database) is needed to maximize value to all stakeholders.

In summary, this article reviewed BTM solar power requirements, techniques, and implications. The models used for BTM solar power generation forecast models were classified according to the users, i.e., customers, distribution and transmission system operators. Each type of user was then introduced based on their use cases, respectively. Several application challenges exist in respective modes, as discussed according to currently adopted practices. This article's comparative discussions seek to help professionals decide the appropriate combination of top-down and bottom-up modeling and data technique to improve existing tools for users and use case combinations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under Solar Energy Technologies Office Agreement Number 33505. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

[1] Jäger-Waldau Arnulf. Snapshot of photovoltaics—February 2020. Energies 2020;13(4):930.

[2] IRENA. Future of solar photovoltaic: Deployment, investment, technology, grid integration, and socio-economic aspects. International Renewable Energy Agency; 2019.

[3] World Health Organization, et al. Tracking SDG 7. Washington, DC: World Bank; 2019.

[4] Balestrieri Michael, James Anthony, Kedis Matthew, Gonzales Frank M. Mitigation of grid susceptibility caused by behind-the-meter solar generation. In: 2020 IEEE conference on technologies for sustainability. IEEE; 2020, p. 1–8.

[5] Tidemann Carl E, Bright Jamie M, Engerer Nicholas A. Solar forecasting as an enablement tool for the distribution system operator (DSO). In: 2019 IEEE 46th photovoltaic specialists conference. IEEE; 2019, p. 1637–44.

[6] Bu Fankun, Dehghanpour Kaveh, Yuan Yuxuan, Wang Zhaoyu, Zhang Yingchen. A data-driven game-theoretic approach for behind-the-meter generation disaggregation. IEEE Trans Power Syst 2020.

[7] Anukoolthamchote Pam Chasuta, Assané Djeto, Konan Denise Eby. Net electricity load profiles: Shape and variability considering customer-mix at transformers on the island of Oahu, Hawai'i. Energy Policy 2020;147:111732.

[8] Denholm Paul, O'Connell Matthew, Brinkman Gregory, Jorgenson Jennie. Overgeneration from solar energy in California: A field guide to the duck chart. Technical report, Golden, CO, United States: National Renewable Energy Lab. (NREL); 2015.

[9] Wang Yi, Zhang Ning, Chen Qixin, Kirschen Daniel S, Li Pan, Xia Qing. Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV. IEEE Trans Power Syst 2017;33(3):3255–64.

[10] Padullaparthi Venkata Ramakrishna, Sarangan Venkatesh, Sivasubramaniam Anand. Suncover: Estimating the hidden behind-the-meter solar rooftop and battery capacities in grids. In: 2019 IEEE power & energy society innovative smart grid technologies conference. IEEE; 2019, p. 1–5.

[11] Killinger Sven, Lingfors David, Saint-Drenan Yves-Marie, Moraitis Panagiotis, van Sark Wilfried, Taylor Jamie, et al. On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth, tilt, capacity, yield and shading. Sol Energy 2018;173:1087–106.

[12] Wirth Harry, Schneider Karin. Recent facts about photovoltaics in Germany, volume 92. Fraunhofer ISE; 2015.

[13] Zhang Xiaochen, Grijalva Santiago. A data-driven approach for detection and estimation of residential PV installations. IEEE Trans Smart Grid 2016;7(5):2477–85.

[14] Shaker Hamid, Manfre Daniel, Zareipour Hamidreza. Forecasting the aggregated output of a large fleet of small behind-the-meter solar photovoltaic sites. Renew Energy 2020;147:1861–9.

[15] Diagne Maimouna, David Mathieu, Lauret Philippe, Boland John, Schmutz Nicolas. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew Sustain Energy Rev 2013;27:65–76.

[16] Voyant Cyril, Notton Gilles, Kalogirou Soteris, Nivet Marie-Laure, Paoli Christophe, Motte Fabrice, et al. Machine learning methods for solar radiation forecasting: A review. Renew Energy 2017;105:569–82.

[17] Wang Lunche, Kisi Ozgur, Zounemat-Kermani Mohammad, Salazar Germán Ariel, Zhu Zhongmin, Gong Wei. Solar radiation prediction using different techniques: Model evaluation and comparison. Renew Sustain Energy Rev 2016;61:384–97.

[18] Yang Dazhi, Kleissl Jan, Gueymard Christian A, Pedro Hugo TC, Coimbra Carlos FM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101.

[19] Ahmed Razin, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew Sustain Energy Rev 2020;124:109792.

[20] Antonanzas Javier, Osorio Natalia, Escobar Rodrigo, Urraca Ruben, Martinez-de Pison Francisco J, Antonanzas-Torres Fernando. Review of photovoltaic power forecasting. Sol Energy 2016;136:78–111.

[21] Inman Rich H, Pedro Hugo TC, Coimbra Carlos FM. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39(6):535–76.

[22] Sobri Sobrina, Koohi-Kamali Sam, Rahim Nasrudin Abd. Solar photovoltaic generation forecasting methods: A review. Energy Convers Manage 2018;156:459–97.

[23] Ren Ye, Suganthan PN, Srikanth N. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renew Sustain Energy Rev 2015;50:82–91.

[24] Raza Muhammad Qamar, Nadarajah Mithulananthan, Ekanayake Chandima. On recent advances in PV output power forecast. Sol Energy 2016;136:125–44.

[25] Li Binghui, Zhang Jie. A review on the integration of probabilistic solar forecasting in power systems. Sol Energy 2020;207:777–95.

[26] Lazos Dimitris, Sproul Alistair B, Kay Merlinde. Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. Renew Sustain Energy Rev 2014;39:587–603.

[27] Van der Meer Dennis W, Widén Joakim, Munkhammar Joakim. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512.

[28] Elsinga Boudewijn, van Sark Wilfried GJHM. Inter-system time lag due to clouds in an urban PV ensemble. In: 2014 IEEE 40th photovoltaic specialist conference. IEEE; 2014, p. 0754–8.

[29] Elsinga Boudewijn, van Sark Wilfried GJHM. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. Appl Energy 2017;206:1464–83.

[30] Huuki Hannu, Karhinen Santtu, Böök Herman, Lindfors Anders V, Kopsakangas-Savolainen Maria, Svento Rauli. Virtual power plant operation with solar power forecast errors and demand response. 2018, Available at SSRN 3277354.

[31] Huuki Hannu, Karhinen Santtu, Böök Herman, Lindfors Anders V, Kopsakangas-Savolainen Maria, Svento Rauli. Utilizing the flexibility of distributed thermal storage in solar power forecast error cost minimization. J Energy Storage 2020;28:101202.

[32] Wang Fei, Li Kangping, Wang Xinkang, Jiang Lihui, Ren Jianguo, Mi Zengqiang, et al. A distributed PV system capacity estimation approach based on support vector machine with customer net load curve features. Energies 2018;11(7):1750.

[33] Chu Yinghao, Pedro Hugo TC, Kaur Amanpreet, Kleissl Jan, Coimbra Carlos FM. Net load forecasts for solar-integrated operational grid feeders. Sol Energy 2017;158:236–46.

[34] Fonseca Jr Joao Gari da Silva, Oozeki Takashi, Takashima Takumi, Koshimizu Gentarou, Uchida Yoshihisa, Ogimoto Kazuhiko. Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan. Prog Photovolt, Res Appl 2012;20(7):874–82.

[35] Fernandez-Jimenez L Alfredo, Muñoz Jimenez Andrés, Falces Alberto, Mendoza-Villena Montserrat, Garcia-Garrido Eduardo, Lara-Santillan Pedro M, et al. Short-term power forecasting system for photovoltaic plants. Renew Energy 2012;44:311–7.

[36] Almeida Marcelo Pinho, Perpinan Oscar, Narvarte Luis. PV power forecast using a nonparametric PV model. Sol Energy 2015;115:354–68.

[37] Masa-Bote Daniel, Castillo-Cagigal Manuel, Matallanas E, Caamaño Martín Estefanía, Gutiérrez A, Monasterio-Huelín F, et al. Improving photovoltaics grid integration through short time forecasting and self-consumption. Appl Energy 2014;125:103–13.

[38] Fonseca Jr Joao Gari da Silva, Oozeki Takashi, Ohtake Hideaki, Takashima Takumi, Ogimoto Kazuhiko. Regional forecasts of photovoltaic power generation according to different data availability scenarios: A study of four methods. Prog Photovolt, Res Appl 2015;23(10):1203–18.

[39] Hoff T. Behind-the-meter (BTM) PV forecast integration with load forecasting. In: Utility variable generation integration group forecasting workshop. 2016.

[40] Braun Gerald W, Kleissl Jan. 0 Final report California renewable energy forecasting, resource data and mapping appendix A current state of the art in solar forecasting. 2011.

[41] Lorenz Elke, Scheidsteger Thomas, Hurka Johannes, Heinemann Detlev, Kurz Christian. Regional PV power prediction for improved grid integration. Prog Photovolt, Res Appl 2011;19(7):757–71.

[42] Saint-Drenan Yves-Marie, Good Garrett H, Braun Martin. A probabilistic approach to the estimation of regional photovoltaic power production. Sol Energy 2017;147:257–76.

[43] Saint-Drenan Yves-Marie, Vogt Stephan, Killinger Sven, Bright Jamie M, Fritz Rafael, Potthast Roland. Bayesian parameterisation of a regional photovoltaic model–Application to forecasting. Sol Energy 2019;188:760–74.

[44] Huang Rui, Huang Tiana, Gadh Rajit, Li Na. Solar generation prediction using the ARMA model in a laboratory-level micro-grid. In: 2012 IEEE third international conference on smart grid communications. 2012.

[45] Benmouiza Khalil, Cheknane Ali. Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. Theor Appl Climatol 2016;124(3–4):945–58.

[46] Li Yanting, Su Yan, Shu Lianjie. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. Renew Energy 2014;66:78–89.

[47] Zhang Bei, Dehghanian Payman, Kezunovic Mladen. Spatial-temporal solar power forecast through use of Gaussian conditional random fields. In: 2016 IEEE power and energy society general meeting. IEEE; 2016, p. 1–5.

[48] Reikard Gordon. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Sol Energy 2009;83(3):342–9.

[49] Lorenz Elke, Hurka Johannes, Karampela Giota, Heinemann Detlev, Beyer Hans Georg, Schneider Martin. Qualified forecast of ensemble power production by spatially dispersed grid-connected PV systems. Measurement 2007;1–7.

[50] Lorenz Elke, Heinemann Detlev, Kurz Christian. Local and regional photovoltaic power prediction for large scale grid integration: Assessment of a new algorithm for snow detection. Prog Photovolt, Res Appl 2012;20(6):760–9.

[51] Lorenz E, Kühnert J, Wolff B, Hammer A, Kramer O, Heinemann D. PV power predictions on different spatial and temporal scales integrating PV measurements, satellite data and numerical weather predictions. In: Proceedings of the 29th European photovoltaic solar energy conference and exhibition. 2014. p. 22–26.

[52] Lonij Vincent PA, Brooks Adria E, Cronin Alexander D, Leuthold Michael, Koch Kevin. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. Sol Energy 2013;97:58–66.

[53] Amral N, Ozveren CS, King D. Short term load forecasting using multiple linear regression. In: 2007 42nd international universities power engineering conference. IEEE; 2007, p. 1192–8.

[54] Nespoli Alfredo, Ogliari Emanuele, Leva Sonia, Massi Pavan Alessandro, Mellit Adel, Lughi Vanni, et al. Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. Energies 2019;12(9):1621.

[55] NOAA National Weather Service. Environmental Modeling Center. https://www.emc.ncep.noaa.gov/.

[56] European Centre for Medium-Range Weather Forecasts. Set I- Atmospheric model high resolution 10-day forecast (HRES). https://www.ecmwf.int/en/forecasts/datasets/set-i.

[57] NOAA Earth System Research Laboratory. High resolution rapid refresh. https://rapidrefresh.noaa.gov/hrrr/.

[58] Ruiz-Arias José A, Pozo-Vázquez David, Lara-Fanego Vicente, Santos-Alamillos Francisco J, Tovar-Pescador J. A high-resolution topographic correction method for clear-sky solar irradiance derived with a numerical weather prediction model. J Appl Meteorol Climatol 2011;50(12):2460–72.

[59] Pelland Sophie, Remund Jan, Kleissl Jan, Oozeki Takashi, De Brabandere Karel. Photovoltaic and solar forecasting: State of the art, volume 14. IEA PVPS Task; 2013, p. 1–36.

[60] Dobos Aron P. PVwatts version 5 manual. Technical report, Golden, CO, United States: National Renewable Energy Lab.(NREL); 2014.

[61] Solar Sheffield. Microgen database. Sheffield Solar-University of Sheffield; 2016, [Online]. Available: http://www.microgen-database.org.uk.

[62] Bright Jamie M, Killinger Sven, Engerer Nicholas A. Data article: Distributed PV power data for three cities in Australia. J Renew Sustain Energy 2019;11(3):035504.

[63] Pedro Hugo TC, Larson David P, Coimbra Carlos FM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. J Renew Sustain Energy 2019;11(3):036102.

[64] Sengupta Manajit, Xie Yu, Lopez Anthony, Habte Aron, Maclaurin Galen, Shelby James. The national solar radiation data base (NSRDB). Renew Sustain Energy Rev 2018;89:51–60.

[65] Stoffel T, Andreas A. Nrel solar radiation research laboratory (SRRL): Baseline measurement system (bms); golden, colorado (data). Technical report, Golden, CO, United States: National Renewable Energy Lab.(NREL); 1981.

[66] Liu Luyao, Liu Diran, Sun Qie, Li Hailong, Wennersten Ronald. Forecasting power output of photovoltaic system using a BP network method. Energy Procedia 2017;142:780–6.

[67] Bacher Peder, Madsen Henrik, Nielsen Henrik Aalborg. Online short-term solar power forecasting. Sol Energy 2009;83(10):1772–83.

[68] Guarnieri R, Martins F, Pereira E, Chou Sin Chan. Solar radiation forecast using artificial neural networks. National Institute for Space Research; 2008, p. 1–34.

[69] Mellit Adel, Pavan Alessandro Massi. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Sol Energy 2010;84(5):807–21.

[70] Mihalakakou G, Santamouris M, Asimakopoulos DN. The total solar radiation time series simulation in Athens, using neural networks. Theor Appl Climatol 2000;66(3–4):185–97.

[71] Hocaoğlu Fatih O, Gerek Ömer N, Kurban Mehmet. Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed forward neural networks. Sol Energy 2008;82(8):714–26.

[72] Sfetsos A, Coonick AH. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. Sol Energy 2000;68(2):169–78.

[73] Hammer Annette, Heinemann Detlev, Lorenz Elke, Lückehe Bertram. Short-term forecasting of solar radiation: A statistical approach using satellite data. Sol Energy 1999;67(1–3):139–50.

[74] Chow Chi Wai, Urquhart Bryan, Lave Matthew, Dominguez Anthony, Kleissl Jan, Shields Janet, et al. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Sol Energy 2011;85(11):2881–93.

[75] Lorenz Elke, Hurka Johannes, Heinemann Detlev, Beyer Hans Georg. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J Sel Top Appl Earth Obs remote sens 2009;2(1):2–10.

[76] Feng Cong, Zhang Jie. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. Sol Energy 2020;204:71–8.

[77] Grell Georg A. Adescription of the fifth-generation penn state/NCAR mesoscale model (MM5). Technical report NCAR/TN-398+STR, Boulder, CO: National Center For Atmospheric Research (NCAR); 1995.

[78] Skamarock William C, Klemp Joseph B, Dudhia Jimy, Gill David O, Barker Dale M, Wang Wei, et al. Adescription of the advanced research WRF version 2. Technical Report, Boulder, CO: National Center For Atmospheric Research (NCAR); 2005.

[79] Cao Shuanghua, Cao Jiacong. Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. Appl Therm Eng 2005;25(2–3):161–72.

[80] Cao JCSH Cao, Cao SH. Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. Energy 2006;31(15):3435–45.

[81] Cao Jiacong, Lin Xingchun. Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. Eng Appl Artif Intell 2008;21(8):1255–63.

[82] Ji Wu, Chee Keong Chan. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. Sol Energy 2011;85(5):808–17.

[83] Feng Cong, Liu Yuanzhi, Zhang Jie. A taxonomical review on recent artificial intelligence applications to PV integration into power grids. Int J Electr Power Energy Syst 2021;132:107176.

[84] Bradbury Kyle, Saboo Raghav, Johnson Timothy L, Malof Jordan M, Devarajan Arjun, Zhang Wuming, et al. Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. Sci Data 2016;3(1):1–9.

[85] Yu Jiafan, Wang Zhecheng, Majumdar Arun, Rajagopal Ram. Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the United States. Joule 2018;2(12):2605–17.

[86] Solar home electricity data. 2013, https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data. [Accessed 3 November 2021].

[87] Feng Cong, Yang Dazhi, Hodge Bri-Mathias, Zhang Jie. Opensolar: Promoting the openness and accessibility of diverse public solar datasets. Sol Energy 2019;188:1369–79.

[88] Marcos Javier, Marroyo Luis, Lorenzo Eduardo, Alvira David, Izco Eloisa. From irradiance to output power fluctuations: The PV plant as a low pass filter. Prog Photovolt, Res Appl 2011;19(5):505–10.

[89] El-Baz Wessam, Seufzger Michael, Lutzenberger Sandra, Tzscheutschler Peter, Wagner Ulrich. Impact of probabilistic small-scale photovoltaic generation forecast on energy management systems. Sol Energy 2018;165:136–46.

[90] El-Baz Wessam, Tzscheutschler Peter, Wagner Ulrich. Day-ahead probabilistic PV generation forecast for buildings energy management systems. Sol Energy 2018;171:478–90.

[91] Tharwat Alaa. Classification assessment methods. Appl Comput Inform 2020.

[92] Thabtah Fadi, Hammoud Suhel, Kamalov Firuz, Gonsalves Amanda. Data imbalance in classification: Experimental evaluation. Inform Sci 2020;513:429–41.

[93] Feng Cong, Mehmani Ali, Zhang Jie. Deep learning-based real-time building occupancy detection using ami data. IEEE Trans Smart Grid 2020;11(5):4490–501.

[94] Murphy Allan H. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather forecast 1993;8(2):281–93.

[95] Yang Dazhi, Alessandrini Stefano, Antonanzas Javier, Antonanzas-Torres Fernando, Badescu Viorel, Beyer Hans Georg, et al. Verification of deterministic solar forecasts. Sol Energy 2020.

[96] Lauret Philippe, David Mathieu, Pinson Pierre. Verification of solar irradiance probabilistic forecasts. Sol Energy 2019;194:254–71.

[97] Pedro Hugo TC, Coimbra Carlos FM. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol Energy 2012;86(7):2017–28.

[98] Zhang Jie, Florita Anthony, Hodge Bri-Mathias, Lu Siyuan, Hamann Hendrik F, Banunarayanan Venkat, et al. A suite of metrics for assessing the performance of solar power forecasting. Sol Energy 2015;111:157–75.

[99] Gneiting Tilmann, Raftery Adrian E. Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 2007;102(477):359–78.

[100] Perez Richard, Perez Marc, Pierro Marco, Schlemmer Jim, Kivalov Sergery, Dise John, et al. Operationally perfect solar power forecasts: A scalable strategy to lowest-cost firm solar power generation. In: 2019 IEEE 46th photovoltaic specialists conference. vol. 2, IEEE; 2019, p. 1–6.

[101] Gneiting Tilmann, Balabdaoui Fadoua, Raftery Adrian E. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Ser B Stat Methodol 2007;69(2):243–68.

[102] Doubleday Kate, Jascourt Stephen, Kleiber William, Hodge Bri-Mathias. Probabilistic solar power forecasting using Bayesian model averaging. IEEE Trans Sustain Energy 2020.

[103] Doubleday Kate, Van Scyoc Hernandez Vanessa, Hodge Bri-Mathias. Benchmark probabilistic solar forecasts: Characteristics and recommendations. Sol Energy 2020;206:52–67.

[104] Gneiting Tilmann, Ranjan Roopesh. Comparing density forecasts using threshold-and quantile-weighted scoring rules. J Bus Econom Statist 2011;29(3):411–22.

[105] Steinwart Ingo, Christmann Andreas, et al. Estimating conditional quantiles with the help of the pinball loss. Bernoulli 2011;17(1):211–25.

[106] Pinson P, Tastu J. Discussion of "prediction intervals for short-term wind farm generation forecasts" and "combined nonparametric prediction intervals for wind power generation". IEEE Trans Sustain Energy 2014;5(3):1019–20.

[107] Seguin Rich, Woyak Jeremy, Costyk David, Hambrick Josh, Mather Barry. High-penetration PV integration handbook for distribution engineers. Technical report, Golden, CO, United States: National Renewable Energy Lab.(NREL); 2016.

[108] Li Kangping, Wang Fei, Mi Zengqiang, Fotuhi-Firuzabad Mahmoud, Duić Neven, Wang Tieqiang. Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation. Appl Energy 2019;253:113595.

[109] Shaker Hamid, Zareipour Hamidreza, Wood David. A data-driven approach for estimating the power generation of invisible solar sites. IEEE Trans Smart Grid 2015;7(5):2466–76.

[110] Shaker Hamid, Zareipour Hamidreza, Wood David. Estimating power generation of invisible solar sites using publicly available data. IEEE Trans Smart Grid 2016;7(5):2456–65.

[111] Kara Emre C, Roberts Ciaran M, Tabone Michaelangelo, Alvarez Lilliana, Callaway Duncan S, Stewart Emma M. Disaggregating solar generation from feeder-level measurements. Sustain. Energy Grids Netw. 2018;13:112–21.

[112] Malof Jordan M, Bradbury Kyle, Collins Leslie M, Newell Richard G. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. Appl Energy 2016;183:229–40.

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

[114] Malof Jordan M, Hou Rui, Collins Leslie M, Bradbury Kyle, Newell Richard. Automatic solar photovoltaic panel detection in satellite imagery. In: 2015 international conference on renewable energy research and applications. IEEE; 2015, p. 1428–31.

[115] Malof Jordan M, Bradbury Kyle, Collins Leslie M, Newell Richard G, Serrano Alexander, Wu Hetian, et al. Image features for pixel-wise detection of solar photovoltaic arrays in aerial imagery using a random forest classifier. In: 2016 IEEE International conference on renewable energy research and applications. IEEE; 2016, p. 799–803.

[116] Golovko Vladimir, Bezobrazov Sergei, Kroshchanka Alexander, Sachenko Anatoliy, Komar Myroslav, Karachka Andriy. Convolutional neural network based solar photovoltaic panel detection in satellite photos. In: 2017 9th IEEE International conference on intelligent data acquisition and advanced computing systems: Technology and applications. vol. 1, IEEE; 2017, p. 14–9.

[117] Golovko Vladimir, Kroshchanka Alexander, Bezobrazov Sergei, Sachenko Anatoliy, Komar Myroslav, Novosad Oleksandr. Development of solar panels detector. In: 2018 international scientific-practical conference problems of infocommunications. Science and technology. IEEE; 2018, p. 761–4.

[118] Malof Jordan M, Collins Leslie M, Bradbury Kyle, Newell Richard G. A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery. In: 2016 IEEE international conference on renewable energy research and applications. IEEE; 2016, p. 650–4.

[119] Malof Jordan M, Collins Leslie M, Bradbury Kyle. A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery. In: 2017 IEEE International geoscience and remote sensing symposium. IEEE; 2017, p. 874–7.

[120] Jie Yongshi, Yue Anzhi, Liu Shunxi, Huang Qingqing, Chen Jingbo, Meng Yu, et al. Photovoltaic power station identification using refined encoder–decoder network with channel attention and chained residual dilated convolutions. J Appl Remote Sens 2020;14(1):016506.

[121] Camilo Joseph, Wang Rui, Collins Leslie M, Bradbury Kyle, Malof Jordan M. Application of a semantic segmentation convolutional neural network for accurate automatic detection and mapping of solar photovoltaic arrays in aerial imagery. 2018.

[122] Mainzer Kai, Killinger Sven, McKenna Russell, Fichtner Wolf. Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques. Sol Energy 2017;155:561–73.

[123] Malof Jordan M, Li Boning, Huang Bohao, Bradbury Kyle, Stretslov Artem. Mapping solar array location, size, and capacity using deep learning and overhead imagery. 2019, arXiv preprint arXiv:1902.10895.

[124] Gagnon Pieter, Margolis Robert, Melius Jennifer, Phillips Caleb, Elmore Ryan. Estimating rooftop solar technical potential across the US using a combination of GIS-based methods, lidar data, and statistical modeling. Environ Res Lett 2018;13(2):024027.

[125] Saint-Drenan Yves-Marie, Bofinger Stefan, Fritz Rafael, Vogt Stefan, Good GH, Dobschinski Jan. An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation. Sol Energy 2015;120:479–93.

[126] Kabir Farzana, Yu Nanpeng, Yao Weixin, Yang Rui, Zhang Yingchen. Estimation of behind-the-meter solar generation by integrating physical with statistical models. In: 2019 IEEE international conference on communications, control, and computing technologies for smart grids. IEEE; 2019, p. 1–6.

[127] Mason Karl, Reno Matthew J, Blakely Logan, Vejdan Sadegh, Grijalva Santiago. A deep neural network approach for behind-the-meter residential PV size, tilt and azimuth estimation. Sol Energy 2020;196:260–9.

[128] Mohassel Ramyar Rashed, Fung Alan, Mohammadi Farah, Raahemifar Kaamran. A survey on advanced metering infrastructure. Int J Electr Power Energy Syst 2014;63:473–84.

[129] Sossan Fabrizio, Nespoli Lorenzo, Medici Vasco, Paolone Mario. Unsupervised disaggregation of photovoltaic production from composite power flow measurements of heterogeneous prosumers. IEEE Trans Ind Inf 2018;14(9):3904–13.

[130] Kloibhofer Simon, Stifter Matthias, Leimgruber Fabian, Rao Bharath-Varsh. Comparing and improving residential demand forecast by disaggregation of load and PV generation. CIRED-Open Access Proc J 2017;2017(1):1638–41.

[131] Stainsby Wendel, Zimmerle Daniel, Duggan Gerald P. A method to estimate residential PV generation from net-metered load data and system install date. Appl Energy 2020;267:114895.

[132] Shaffery Peter, Yang Rui, Zhang Yingchen. Bayesian structural time series for behind-the-meter photovoltaic disaggregation. In: 2020 IEEE power & energy society innovative smart grid technologies conference. IEEE; 2020, p. 1–5.

[133] Landelius Tomas, Andersson Sandra, Abrahamsson Roger. Modelling and forecasting PV production in the absence of behind-the-meter measurements. Prog Photovolt, Res Appl 2019;27(11):990–8.

[134] Ahmed Adil, Khalid Muhammad. A review on the selected applications of forecasting models in renewable power systems. Renew Sustain Energy Rev 2019;100:9–21.

[135] Visser Lennard, AlSkaif Tarek, van Sark Wilfried. Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. Renew Energy 2021.

[136] Mayer Martin János, Gróf Gyula. Extensive comparison of physical models for photovoltaic power forecasting. Appl Energy 2020;116239.

[137] Soman Saurabh S, Zareipour Hamidreza, Malik Om, Mandal Paras. A review of wind power and wind speed forecasting methods with different time horizons. In: North American power symposium 2010. IEEE; 2010, p. 1–8.

[138] Pelland Sophie, Galanis George, Kallos George. Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. Prog Photovolt, Res Appl 2013;21(3):284–96.

[139] Larson David P, Nonnenmacher Lukas, Coimbra Carlos FM. Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. Renew Energy 2016;91:11–20.

[140] Saint-Drenan Yves-Marie, Good Garrett H, Braun Martin, Freisinger Thomas. Analysis of the uncertainty in the estimates of regional PV power generation evaluated with the upscaling method. Sol Energy 2016;135:536–50.

[141] Persson Caroline, Bacher Peder, Shiga Takahiro, Madsen Henrik. Multi-site solar power forecasting using gradient boosted regression trees. Sol Energy 2017;150:423–36.

[142] Zhang Ruiyuan, Ma Hui, Hua Wen, Saha Tapan Kumar, Zhou Xiaofang. Data-driven photovoltaic generation forecasting based on a Bayesian network with spatial–temporal correlation analysis. IEEE Trans Ind Inf 2019;16(3):1635–44.

[143] Jang Han Seung, Bae Kuk Yeol, Park Hong-Shik, Sung Dan Keun. Solar power prediction based on satellite images and support vector machine. IEEE Trans Sustain Energy 2016;7(3):1255–63.

[144] Akhter Muhammad Naveed, Mekhilef Saad, Mokhlis Hazlie, Shah Noraisyah Mohamed. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. IET Renew Power Gener 2019;13(7):1009–23.

[145] Iyengar Srinivasan, Sharma Navin, Irwin David, Shenoy Prashant, Ramamritham Krithi. A cloud-based black-box solar predictor for smart homes. ACM Trans Cyber-Phys Syst 2017;1(4):1–24.

[146] Rana Mashud, Rahman Ashfaqur, Jin Jiong. A data-driven approach for forecasting state level aggregated solar photovoltaic power production. In: 2020 International joint conference on neural networks. IEEE; 2020, p. 1–8.

[147] Li Yuan, Sun Qian, Lehman Brad, Lu Siyuan, Hamann Hendrik F, Simmons Joseph, et al. A machine-learning approach for regional photovoltaic power forecasting. In: 2016 IEEE power and energy society general meeting. IEEE; 2016, p. 1–5.

[148] Almaghrabi Sarah, Rana Mashud, Hamilton Margaret, Rahaman Mohammad Saiedur. Forecasting regional level solar power generation using advanced deep learning approach. In: 2021 International joint conference on neural networks. IEEE; 2021, p. 1–7.

[149] Ogliari Emanuele, Dolara Alberto, Manzolini Giampaolo, Leva Sonia. Physical and hybrid methods comparison for the day ahead PV output power forecast. Renew Energy 2017;113:11–21.

[150] Yang Hong-Tzer, Huang Chao-Ming, Huang Yann-Chang, Pai Yi-Shiang. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. IEEE Trans Sustain Energy 2014;5(3):917–26.

[151] Tao Yuqi, Chen Yuguo. Distributed PV power forecasting using genetic algorithm based neural network approach. In: Proceedings of the 2014 international conference on advanced mechatronic systems. IEEE; 2014, p. 557–60.

[152] Dolara Alberto, Grimaccia Francesco, Leva Sonia, Mussetta Marco, Ogliari Emanuele. A physical hybrid artificial neural network for short term forecasting of PV plant power output. Energies 2015;8(2):1138–53.

[153] Eseye Abinet Tesfaye, Zhang Jianhua, Zheng Dehua. Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information. Renew Energy 2018;118:357–67.

[154] Mellit Adel, Massi Pavan Alessandro, Ogliari Emanuele, Leva Sonia, Lughi Vanni. Advanced methods for photovoltaic output power forecasting: A review. Appl Sci 2020;10(2):487.

[155] Kaur Amanpreet, Nonnenmacher Lukas, Coimbra Carlos FM. Net load forecasting for high renewable energy penetration grids. Energy 2016;114:1073–84.

[156] Van der Meer DW, Munkhammar Joakim, Widén Joakim. Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. Sol Energy 2018;171:397–413.

[157] Sreekumar Sreenu, Sharma Kailash Chand, Bhakar Rohit. Gumbel copula based aggregated net load forecasting for modern power systems. IET Gener Transm Distrib 2018;12(19):4348–58.

[158] Sun Mingyang, Zhang Tingqi, Wang Yi, Strbac Goran, Kang Chongqing. Using Bayesian deep learning to capture uncertainty for residential net load forecasting. IEEE Trans Power Syst 2019;35(1):188–201.

[159] Mei Fei, Wu Qingliang, Shi Tian, Lu Jixiang, Pan Yi, Zheng Jianyong. An ultrashort-term net load forecasting model based on phase space reconstruction and deep neural network. Appl Sci 2019;9(7):1487.

[160] Kobylinski P, Wierzbowski M, Piotrowski K. High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources. Int J Electr Power Energy Syst 2020;117:105635.

[161] Sun Mucun, Feng Cong, Zhang Jie. Factoring behind-the-meter solar into load forecasting: Case studies under extreme weather. In: 2020 IEEE power & energy society innovative smart grid technologies conference. IEEE; 2020, p. 1–5.

[162] Alipour Mohammadali, Aghaei Jamshid, Norouzi Mohammadali, Niknam Taher, Hashemi Sattar, Lehtonen Matti. A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration. Energy 2020;118106.