A review on the integration of probabilistic solar forecasting in power systems

Binghui Li*, Jie Zhang*

*The University of Texas at Dallas, Richardson, TX 75080, USA

Abstract

As one of the fastest growing renewable energy sources, the integration of solar power poses great challenges to power systems due to its variable and uncertain nature. Therefore, there are imperative needs to increase flexibility, reliability, energy efficiency, and power quality for today’s power system with growing solar power penetration. A wide range of research has been conducted to promote the integration of solar power. The value of forecasts over multiple time horizons to numerous aspects of real-world power systems is being increasingly recognized in the recent decades. While the current use of probabilistic forecasts in power systems is limited, enormous amount of research has been conducted to promote the adoption of probabilistic forecasts and many methods have been proposed. This paper gives a comprehensive review on how probabilistic solar forecasts are utilized in power systems to address the challenges. Potential methods to deal with uncertainties in power systems are summarized, such as probabilistic load flow models, stochastic optimization, robust optimization, and chance constraints. In addition, specific areas where these methods can be applied in conjunction with probabilistic forecasts are discussed and state-of-the-art studies are summarized.

Keywords: Probabilistic forecasting, Solar power, Power markets, Solar integration

1. Introduction

The U.S. has witnessed a fast growth of solar and wind power in the recent decade. From 2005 to 2015, onshore wind capacity grows more than eightfold and grid-connected solar PV capacity increases by more than a factor of 109 [Beiter et al., 2017; Wiser and Bolinger, 2017], owing primarily to cost and performance improvements as well as policy incentives [Barbose et al., 2016; Wiser and Bolinger, 2017]. At the end of 2018, the total installed capacity of utility scale solar power (including both solar PV and solar thermal) is 32,239 MW and it accounts for less than 2% of U.S. electricity consumption [U.S. EIA, 2018]. Fig. 1 shows the installed capacities of grid-connected solar and wind power in the U.S. by state. In addition, U.S. EIA [2018] estimates that there is 19,547 MW of small scale solar PV installed. As one of the fastest growing renewable energy sources, the increasing penetration of solar power has profound impact on the grid reliability and security due to their variable and uncertain nature [Katiraei and Aguirre, 2013; Bird et al., 2013; Hossain and Mahmud, 2014]. A well-known example is the “duck curve” in California Independent System Operator (CAISO) [Denholm et al., 2015; CAISO, 2016]: CAISO has the largest installed solar power capacity across all wholesale power markets in the U.S., which results in drastic changes of net load during sunrise and sunset and sometimes the profiles of net load resemble a duck. Power system operators have to take measures to mitigate the impacts, such as increasing the procurement of operating reserves [ERCOT, 2010b], or introducing new market products [Navid and Rosenwald, 2013; CAISO, 2016a]. However, these measures usually come at the cost of a reduced market efficiency.
A promising way to address the challenges is to use forecasts. Although forecasts, such as electricity price forecasts (Weron, 2014; Nowotarski and Weron, 2018) and load forecasts (Hong and Fan, 2016; Feng et al., 2019), have been used in power systems for a long time (Hong, 2014), the use of forecasting models of renewable energy production is relatively new. In a survey by Widiss and Porter (2014), ten of the 13 surveyed operating entities in the Western Interconnection started to use forecasts of renewable generation in 2007 or later. The recent adoption of renewable energy forecasts is largely due to the growth of wind and solar power and it has made great contributions to the reliable and economic operation of power systems. For example, Xcel Energy, one of the largest utilities in the U.S., develops their wind forecasting system in collaboration with National Center for Atmospheric Research (NCAR) and uses it in both day-ahead trading and real-time operation (Mahoney et al., 2012). The use of forecasting results in an estimated $8 millions of savings over the course from 2009 to 2011 due to more efficient commitment and dispatching of fossil fuel resources. In addition, simulations of the operation of a variety of real-world power systems over a wide range of timescales also suggest many benefits from the improvement of forecasting techniques, such as reduction in risks of market imbalance, improvement of system reliability, decrease of reserve requirements, savings in electricity prices and system operation costs, and increased market revenues (Kraas et al., 2013; Luoma et al., 2014; Garrigle and Leahy, 2015; Hodge et al., 2015; Swinand and O'Mahoney, 2015; Kaur et al., 2016; Martinez-Ando et al., 2016; Cui et al., 2018).

The benefits in terms of power system cost-effectiveness and security have prompted the adoption of forecasts in many markets. In most markets, forecasts of load and renewable energy production are used to determine commitment and dispatch of their generating units. For example, in CAISO, day-ahead load and real-time renewable generation (solar and wind) are generated by machine-learning based methods, which are then used as inputs to reliability unit commitment (UC) and economic dispatch (ED) models for market clearing (CAISO, 2020). Orwig et al. (2015) gave a review on recent efforts of integrating forecasts...
of variable energy generation in the U.S. markets and it showed that although forecasts have been widely
used in many power markets in the U.S., most of them only focus on deterministic forecasts, i.e., point
forecasts. Surprisingly, even though probabilistic forecasts are also provided by forecasting vendors, power
system operators tend to neglect them (Witliss and Porter, 2014) due to a lack of proper ways to use them.
While the role of probabilistic forecasts becomes increasingly important (Haupt et al., 2019), there is still
no systematic way to integrate them in real-world power systems.

Numerous studies have been conducted to promote the integration of forecasting in the recent decade
and many papers have given in-depth and comprehensive reviews of the latest progress in this area. For
example, Bessa et al. (2017) and Ahmed and Khalid (2019) provided reviews on methods to handle vari-
ability and uncertainty in power system operations. In addition, most of the current reviews place emphasis
on the use of wind power forecasts. This is largely due to the greater penetration level of wind power than
solar power, and the higher maturity level of wind power forecasting. In fact, compared with wind power
forecasting (Feng et al., 2017; Sun et al., 2019a, b, 2020; Hong et al., 2016), Haupt et al. (2019) pointed out that both determin-
istic and probabilistic solar power forecasting are in a less mature stage due to solar’s less penetration level
worldwide. Although several studies (Bessa et al., 2014; Ahmed and Khalid, 2019) also included application
of solar power forecasting in power systems, most of them only focused on deterministic forecasts. The im-
portance of probabilistic forecasts is being increasingly recognized by utilities and power system operators,
and numerous methods exist in the literature where probabilistic forecasts are adopted for a variety of goals,
such as scheduling, operation, and energy trading. However, to the authors’ best knowledge, there is not a
comprehensive and systematic review on the use of probabilistic solar forecasts in power systems.

Therefore, this study is intended to fill the gap in the review literature by providing a comprehensive
review on the use of probabilistic solar forecasting in power systems, including the current market practice,
potential methods proposed in academic publications, and state-of-the-art studies in this area. Note that
this paper does not intend to provide a review on forecasting techniques per se, and interested readers are
referred to more comprehensive review work (Inman et al., 2013; Tuohy et al., 2015; Antonanzas et al.,
2016; van der Meer et al., 2018c; Sobri et al., 2018; Yang et al., 2018; Sweeney et al., 2019). This review
has three potential groups of audience. The first group consists of scholars from research institutes and
universities, to whom this review may be of interest since the research frontiers summarized can inspire new
research directions. The second group includes utilities and system operators or coordinators, who may use
the research and market implementation summarized in this paper as guidelines to better use probabilistic
solar forecasting and to design new market products for improved market efficiency. The last group consists
of policymakers such as public utility commissions or federal regulating bodies, and this review can help
them design policies to further promote the integration of renewable energy.

Figure 1: Total grid-connected solar PV and thermal capacities by state in the U.S. at the end of 2018. Data source: U.S. EIA
(2018). Note that grey indicates zero capacity installed.

The contributions of this paper are as follows:

1. First, we provide a broad overview on current market implementations of probabilistic solar forecasts
and on-going research directions.

2. Second, potential methods to integrate probabilistic forecasts into power system practice are discussed, including stochastic optimization, robust optimization, and chance-constrained optimization.

3. Third, studies are reviewed and grouped by their methods, types of forecasts, and objectives. Different applications, such as market scheduling and operation, ancillary services, are discussed in detail.

This paper is structured as follows. Section 2 gives a brief introduction to probabilistic solar forecasting. Section 3 reviews the current use of probabilistic forecasts, including both wind and solar power, in current power markets. Section 4 gives an in-depth discussion about commonly used methods to integrate probabilistic forecasts into power systems. Section 5 further discusses the areas where these methods can be applied. Section 6 concludes the paper.

2. A brief introduction to probabilistic solar forecasting

Forecasting is essentially the process of predicting the future based on past and current data. In the case of solar forecasting, the process becomes predicting future solar irradiance or power productions from historical and present meteorological observations, which are usually provided by local weather stations, remote sensing (satellite imaging), or local sensing (sky imaging, pyranometers). The first step is often to develop a model to characterize the mapping between input variables (meteorological observations) and output variables (predicted variables, such as irradiance or solar power). Depending on whether a physical model (or PV performance model in some literature) is used, solar forecasting can be divided into physical methods and non-physical methods (Antonanzas et al., 2016; van der Meer et al., 2018c). Physical methods predict future solar irradiance or solar power by modeling detailed physical processes between the predicted variables and meteorological observations. One example is the class of physical satellite models for solar irradiance forecasting, where radiative transfer models are enhanced by satellite observations of current atmospheric conditions (Inman et al., 2013). In addition, such methods also include numerical weather prediction (NWP), where measured meteorological parameters, such as temperature, pressure and humidity, are used as inputs and the atmospheric evolution process is explicitly simulated by solving a set of governing partial differential equations (Bauer et al., 2015). These equations are usually quite complex due to the highly non-linear and chaotic characteristics of the atmospheric process. Solving these equations require advanced numerical techniques and are typically quite computationally intensive. High performance computing platforms are often used. Examples of physical models include the North American Mesoscale Forecast System (NAM) developed by the National Centers for Environmental Prediction (NCEP), and the forecasts provided by the European Centre for Medium-Range Weather Forecasts (ECMWF).

By contrast, non-physical approaches belong to the family of data-driven approaches, which depend purely on historical data and do not model the physical process of meteorological evolution. Due to their heavy reliance on statistical tools, they are also called statistical methods. Commonly used statistical techniques include regressive models and artificial intelligence models (Wang et al., 2019a). Without explicit considering the atmospheric evolution process, non-physical models are sometimes called “black-box models”, whereas physical models are called “white-box models”. Despite the lack of clear representation of physical processes, the performance of non-physical models can sometimes be surprisingly good. Typically, physical models present better accuracy in day-ahead and longer forecast timeframes, while statistical models tend to work better over intra-hour time scales (Freiber et al., 2015; Tuohy et al., 2015). In real-world practice, multiple models can be used such that it can seamlessly shift from one method to another to reflect the optimal method that provides the best performance over a particular look-ahead period (Widiss and Porter, 2014). For example, in the Solar and Wind Integrated Forecast Tool (SWIFT), a service that provides short-term solar and wind forecasts (15 min to day-ahead) for Hawaii’s electric utilities, multiple forecasting techniques are integrated, which include NWP models, statistical models and models based on satellite images (Freedman et al., 2013).

In addition, hybrid approaches, which combine techniques from physical and statistical approaches, can also be found in the literature. These methods apply post-processing techniques after physical approaches are done to improve the performance and construct a density function (Bakker et al., 2019). For example,
Almeida et al. (2017) developed a post-processing technique based on quantile regression forests and applied it to a set of meteorological variables (solar radiation, cloud cover, temperature, wind speed, etc.) from an NWP model. By comparing it with the forecast from a pure physical approach, they concluded that the hybrid approach improved the results by presenting less bias. Sperati et al. (2016) took an ensemble forecast from ECMWF and applied a set of statistical methods to remove bias and create probability distribution functions (PDFs). Yang (2019) used Kalman filtering to post-process day-ahead NWP models. In the analog ensemble (AnEn) method, a deterministic NWP model is used and in order to account for forecasting errors, AnEn looks back into past NWP forecasts and selects a group of forecasts that are made under similar meteorology conditions to the current forecast. The past forecasts and observations are then used to construct the distributions of forecasting errors of the current forecasts (Cervone et al., 2017; Yang and Alessandrini, 2019). More comprehensive reviews on the hybrid approaches can be found in Voyant et al. (2017) and Akhter et al. (2019).

Most forecasts used in current power systems are deterministic. Conventionally, deterministic forecasts, or point forecasts, only provide one single value for each time point and often fail to account for the uncertainties associated with the forecasts. Probabilistic forecasts explicitly account for the uncertainty information, which usually take the form of probability distributions. While the above categorization is based on forecasting techniques, probabilistic forecasting can also be categorized based on the form of their output, which are more related to their applications in power systems. Depending on whether an explicit parametric form is used, probabilistic forecasting can be divided into parametric approach, where the forecasted random variable is assumed to follow a prior distribution, and non-parametric approach, where no such assumptions are made. A popular parametric approach is to express the forecast as the sum of a point forecast and a distribution of forecasting errors. Typically, the point forecast is generated by regressive or physical models and the forecasting errors are assumed to be normal, such as in David et al. (2016) and Lorenz et al. (2009). However, an increasing number of studies call this approach into question since the assumed normal distribution is not well supported due to a lack of symmetry in the observed data (Mathiesen et al., 2013; David et al. 2016, 2018). In addition, many studies pointed out the limitation of unimodal distributions due to the multi-modality of the observed distributions (Munkhammar et al., 2017; Smith et al., 2017). The asymmetric and multi-modal characteristics inspire the use of other types of distributions, such as Beta distribution (Lorenz et al., 2007), Gamma distribution (Bracale et al., 2013), and mixture models (Hollands and Suehrcke, 2013; Kakimoto et al., 2019). Such distributions are often used to directly model clear-sky index or irradiance.

By contrast, without explicitly making assumptions on distributions, non-parametric methods provide a better track of the observed asymmetry and thus better performance (Mathiesen et al., 2013). Almeida et al. (2017) produced probabilistic forecasts of solar radiation by applying two parametric approaches and five non-parametric approaches to post-process meteorological parameters obtained from NWP models. The comparison of a set of evaluation metrics indicated better performance from non-parametric methods. In a review study by van der Meer et al. (2018a), the family of non-parametric approaches dominate the methods of probabilistic solar forecasting by accounting for 69% of all reviewed studies, and the most commonly used non-parametric method is quantile regression (Lauret et al., 2017). In contrast to conventional linear regression, where the objective function is to minimize the sum of square errors, quantile regression minimizes the sum of mean absolute errors with asymmetric weights to obtain prediction quantiles. Quantile regression is usually better at quantifying asymmetric distributions and adopted in numerous studies (Almeida et al., 2017; Lauret et al., 2017; David et al., 2018; Panamtash et al., 2020). Other notable methods include Gaussian process (van der Meer et al., 2018a), Markov-chain mixture distribution model (Munkhammar et al., 2019), and lower upper bound estimate (LUBE) (Ni et al., 2017).

Another family of non-parametric forecasts is ensemble forecast. An ensemble consists of a set of different forecasts, which collectively produce better predictive performance than any single one. Depending on how the members are produced, ensemble forecasts can be further divided into ensemble NWP models, ensemble time series models, and ensemble machine learning models. Ensemble forecasts are widely used in NWP to quantify uncertainties caused by model imperfection and initial condition uncertainties (Leutbecher and Palmer, 2008; Bauer et al., 2015). They can be created by either adding perturbations to the initial state and boundary conditions of one model, or using different numerical models or physics schemes. For
example, the NAM ensemble is known as the Short Range Ensemble Forecast (SREF) (Du et al., 1997). However, the uncertainty spread from the ensemble forecast is known to be conservative, which means it tends to underestimate the prediction uncertainty. Therefore, post-processing based on statistical methods are usually employed to improve the forecasts, such as in Sperati et al. (2016). While ensemble NWP forecasts can provide valuable information regarding weather uncertainties, a disadvantage is their high computational intensity due to the repetitive use of NWP models. Other forms of ensemble forecasts, such as ensemble time series (Yang and Dong 2018; van der Meer et al., 2018a) and ensemble machine learning models (Raza et al. 2018), can also be found in the literature. van der Meer et al. (2018a) developed Gaussian process ensembles by employing a Gaussian mixture model and demonstrated that the GMM ensemble outperforms individual Gaussian process models. Ensemble learning methods take advantage of multiple learning algorithms to improve the predictive performance. Two popular methods to create ensembles are bootstrap aggregating (i.e., bagging) and boosting (Opitz and Maclin 1999). For example, by applying bagging to decision tree learning, quantile random forests (QRF) can be used to make probabilistic forecasts (Almeida et al. 2015). A review on ensemble forecasts can be found in Ren et al. (2015).

Different from parametric forecasts, non-parametric forecasts can take a variety of forms depending on the adopted methods, such as quantiles (Bacher et al. 2009; Golestaneh et al. 2016), uncertainty intervals (Almeida et al. 2015; Wang and Jia 2015), and kernel density estimations (Yamazaki et al. 2016). However, the basic quantity to be considered is the quantile forecast, since the other two may be expressed in terms of two or more quantiles (Pinson et al. 2009). For example, Almeida et al. (2015) developed a probabilistic solar power forecast based on quantile regression forests to output the 90% predictive interval by taking the difference between the 95th and 5th percentiles. Fig. 2 shows the quantiles and uncertainty intervals of a probabilistic solar power forecast. Note that during early morning and late afternoon, the distribution of the forecast appears to be asymmetric. In addition, the results of ensemble forecasts usually consist of a set of deterministic forecasts that spread a certain interval, from which the forecasting errors can be characterized by either prediction intervals or density functions (Sperati et al. 2016; Uno et al. 2018). To construct a CDF from ensemble forecasts, the ensemble members are first sorted in an ascending order and can be viewed as discrete estimates of a CDF. For continuous variables, such as GHI or solar power, different methods can be used to construct a continuous CDF by assigning probability masses between two neighboring discrete CDF values (Lauret et al. 2019). Fig. 3 summarizes the types of probabilistic forecasting techniques, their uncertainty representations, and corresponding usages in power system applications, which gives a high-level view of the workflow of using probabilistic forecasts.

![Figure 2: Probabilistic solar power forecasts displayed in the form of uncertainty intervals. The forecasts are produced by using the method from Sun et al. (2019b). The upper and lower edges of each uncertainty interval represent a percentile. For example, the upper and lower edges of the 10% uncertainty interval represent the 55th and 45th percentiles, respectively.](image)

### 3. Current market use of probabilistic forecasting

Probabilistic forecasting is not new to real-world power system operators. As early as 2012, a survey was conducted by Porter and Rogers (2012) to give an overview of the adoption of forecasts of variable generation in the Western Interconnection based on phone interviews with main balancing areas (BAs). They showed that probabilistic wind and solar forecasts had been prepared both internally by the BAs and externally.
by forecasting vendors. These forecasts are usually produced by ensemble forecasting and take the form of predictive intervals. However, these probabilistic forecasts rarely play any role in the decision-making process of real-world power systems and deterministic forecasts are still used predominantly (Bessa et al., 2017). What’s more, a follow-up of the 2012 survey conducted by Widiss and Porter (2014) showed that many system operators even reportedly ignored the provided predictive intervals and chose instead to use a single likeliest production value. The follow-up survey also showed that all surveyed entities used their wind forecasts in day-ahead unit commitment and most of them used forecasts for intra-day unit commitment and determination of the requirements of operating reserves. Other studies also showed that forecasts are primarily used for market scheduling and operation, determination of ancillary services, and improvement of situational awareness (Mahoney et al., 2012; Porter and Rogers, 2012; Widiss and Porter, 2014; Orwig et al., 2015; Tuohy et al., 2013).

Despite the lack of wide recognition of the value of probabilistic forecasts, some utilities plan to transit to the use of probabilistic forecasts and many market operators plan to use probabilistic forecasts in their energy management systems, as shown in Porter and Rogers (2012) and Widiss and Porter (2014). Lately, Haupt et al. (2019) provided an overview on the current adoption of probabilistic forecasts of renewable energy in utilities and power markets worldwide. Most of them are focused on wind power due to its higher penetration. For example, the Southwest Power Pool (SPP) developed a process to improve situational awareness of the impacts of forecasting errors by conducting daily studies with different levels of errors and resource flexibility. By scrutinizing the uncertainty spread of a set of unique NWP models, system operators quantify the uncertainty levels associated with the forecasts and decide whether additional generating units need to be committed ahead of time to mitigate potential resource insufficiency. In addition, the Electric Reliability Council of Texas (ERCOT) has developed an online tool to incorporate probabilistic wind power forecasts into the real-time decision process by displaying percentiles of wind power forecasts to inform system operators of probable resource shortage. ERCOT also launched a first-of-its-kind wind forecasting tool designed to help system operators prepare for wind ramps (ERCOT, 2010a). The tool, known as ERCOT Large Ramp Alert System (ELRAS), is intended to improve situational awareness of ERCOT’s system operators in the control room by providing the likelihood and magnitude of wind ramp events for
the next 6 hours (Sharma et al., 2012; Maggio et al., 2010).

Due to the continuing growth of solar penetration, integration of probabilistic solar forecasts into power system practice can also be seen, especially in the areas where solar power plays an important role. Similar to the use of probabilistic wind forecasts, most of these applications are used for improvement of situational awareness and event forecasting. The Hawaiian Electric Company (HECO), which runs several isolated power grids with a high and growing penetration of solar and wind power, has supported the development of SWIFT in partnership with AWS Truepower and the Electric Power Research Institute (EPRI) (Orwig et al., 2015; Nakafuji et al., 2012). The forecasts are presented to system operators in tabular and graphical forms to display percentiles of solar and wind power forecasts, and probabilities of ramp events. The SWIFT forecasts are currently used in the decision-making process of hour-ahead dispatch and to inform the system operator of excessive ramp rates (Nakafuji et al., 2012; Nakafuji and Gouveia, 2016).

There are a lot of on-going research projects for better integration of probabilistic forecasts. Most of them focused on the determination of operating reserves, such as regulation (Matos and Bessa, 2011; Etingov et al., 2018). Other current e↵orts include the Department of Energy (DOE) Solar Forecasting 2 projects (Golnas, 2018), where utility companies and ISOs collaborate with research institutes and universities to promote the integration of probabilistic solar power forecasts in power systems by determining the requirements of ancillary services, such as non-spinning reserves and ramping reserves (Tuohy et al., 2019; Hobbs et al., 2019; Hodge et al., 2019).

In summary, the use of probabilistic solar forecasts is still in its early stage. Probabilistic forecasts are only used in a primitive way in most systems, and there is not a systematic way to integrate them into system operations and scheduling routines. However, operators start to realize the importance of probabilistic forecasts in shaping more cost-e↵ective, stable and reliable power systems and many research projects are on-going. Numerous studies exist in the literature, which can be used to facilitate the use of probabilistic forecasts. In the next sections, we will give a comprehensive review on these methods and how they can be used in conjunction with probabilistic forecasts.

4. Potential methods to use probabilistic forecasts

This section is intended to give an overview of the potentially useful approaches to using probabilistic solar forecasts in power systems. Generally speaking, these approaches are not new and historically they are widely used to deal with uncertainties in many other models. For example, stochastic optimization and robust optimization are widely employed in the decision making of many engineering and financial problems. The approaches used in probabilistic power flow models, such as the Monte Carlo (MC) method, are also regular methods in probability theory and statistics.

The underlying models of many power system problems can be further categorized into power flow (PF) models and optimal power flow (OPF) models. The PF model is the fundamental model used in steady-state analysis of an interconnected power system during normal operations. The system is assumed to be operating under a balanced condition and is represented by a single-phase network. It is the backbone of many analyses in power systems, such as transient stability and contingency analysis (Saadat, 2011; Wood et al., 2013). The basic form of the PF model is a set of non-linear equations. The OPF model is used to minimize the total generation cost of a power network subject to constraints such as resource availability, limits of line flow, and balance between supply and demand (Wood et al., 2013). Typically, PF equations are embedded as security constraints, either in the linearized (DC-OPF) or complete form (AC-OPF). The OPF model has several variants when additional operational constraints are enforced, such as UC and ED models, which are widely used in power market scheduling and operations. The core of these models can be reduced to optimization models. A general optimization problem can take the following form:

\[
\begin{align*}
\min_x & \quad f(x, \omega) \\
\text{s.t.} & \quad g(x, \omega) = 0 \\
& \quad h(x, \omega) \leq 0
\end{align*}
\]
where $x$ is a vector of decision variables, and $\omega$ represents a vector of uncertain parameters, which are represented with random variables. Function $f$ is the objective function and multivariate functions $g$ and $h$ constitute a set of equality and non-equality constraints. In this formulation, $x$ can represent the commitment decisions or dispatch setting points of thermal generators in a UC or ED model, or voltage amplitudes and phase angles in an OPF problem. The constraint can be PF equations that ensure power flow balance at each bus. The random variable $\omega$ can be forecasts of load, solar power, or wind power. Depending on the form of uncertainty representation, a variety of methods can be adopted to obtain the optimal solution under uncertainties. Table 1 simply compares three commonly used optimization models under uncertainties, which will be detailed later.

### Table 1: Comparison of uncertainty representations in optimization models.

<table>
<thead>
<tr>
<th>Models</th>
<th>Uncertainty representation</th>
<th>Objective</th>
<th>Technical challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic optimization</td>
<td>Scenarios</td>
<td>Expectation</td>
<td>Low scalability</td>
</tr>
<tr>
<td>Robust optimization</td>
<td>Uncertainty sets</td>
<td>Worst case solution</td>
<td>Conservative solutions</td>
</tr>
<tr>
<td>Chance-constrained optimization</td>
<td>Chance constraints</td>
<td>No change</td>
<td>Intractability</td>
</tr>
</tbody>
</table>

#### 4.1 Probabilistic PF and OPF models

As the fundamental model in many power system problems, a PF model gives voltage amplitudes and phase angles at each bus and line flow on each branch of a power network under a given system state, which consists of active and reactive power injections, voltage angles and amplitudes at selected buses. Mathematically, the PF problem is formulated as a set of non-linear equations and solved by iterative algorithms, such as Gauss-Seidel or Newton-Raphson (Saadat, 2011). The PF problem is conventionally solved under deterministic system state without considering uncertainties. By contrast, probabilistic PF (PPF, a.k.a. probabilistic load flow, or PLF in some literature) takes into account the uncertainties of system state and reflects the uncertainties explicitly in the output. Instead of giving point estimates for unknown variables, PLF provides probability distributions and/or statistical moments.

Basic techniques of PPF include analytical approaches, Monte Carlo simulations, and approximate methods (Prusty and Jena, 2017). Analytical approaches are traditionally applied to linearized PF formulations, where unknown variables can be expressed as linear combinations of input variables. Therefore, convolution (Allan et al., 1981) and cumulant (Zhang and Lee, 2004; Usaola, 2009; Wu et al., 2016) based methods can be used. However, due to the linearization of the PF model and assumed independence of variables, the applicability of analytical approaches is usually limited by their lack of accuracy and inability to provide voltage estimates. This issue is particularly prominent in distribution networks, where prerequisites for linearization are generally not satisfied. Therefore, the MC method is often used in the PPF model of distributed networks (Romero-Ruiz et al., 2016; Kabir et al., 2016; Giraldo et al., 2019; Constante-Flores and Illindala, 2019). In addition, many studies suggest that correlations across variables have profound impact on the assessment results (Morales et al., 2010; Usaola, 2010; Ciapessoni et al., 2014). Therefore, recent efforts also give explicit consideration of the correlations and non-linear PF formulations (Wang et al., 2017).

The Monte Carlo method estimates the distributions of output variables by running a deterministic PF model under a set of randomly generated scenarios, which correspond to different system states. The results from all the scenarios are then used to estimate the distributions of the output variables. Since the deterministic model can be a full PF formulation, the performance of the MC method depends largely on the number and quality of the generated scenarios. Clearly, the results of the MC method are more accurate if more scenarios are used. Basic sampling techniques usually use the inverse CDF of the sampled random variables. Various techniques exist to improve the accuracy of Monte Carlo samples, such as Latin Hypercube Sampling (LHC) (Mckay et al., 2000) and importance sampling (Rubinstein and Kroese, 2016). A comprehensive review of sampling techniques can be found in Homen-de-Mello and Bayraksan (2014). In many studies, the MC method is used as baseline against which more advanced methods are compared (Kabir et al., 2016). However, to obtain accurate results, a large number of scenarios are often needed, which can
be computationally expensive. Due to the high computational intensity of MC simulations, many efforts have been made to improve the efficiency (Hamon et al., 2016; Kabir et al., 2016; Huang et al., 2019; Su 2005; Morales and Perez-Ruiz, 2007), which represent a trade-off between accuracy and computational burden. While similar to the MC method in the sense that the PEM approach also evaluates deterministic cases at a selected set of system states, the difference lies in the generation of scenarios, where the MC set consists of randomly generated scenarios and the PEM scenarios are selected following a predefined rule. In addition, as opposed to the MC method where full parametric distribution functions of the random variables are often required, the PEM approach only needs the first few statistical moments of the random variables. The results are used to derive statistical moments of unknown random variables, from which their full probability distributions can be estimated. Due to the typically fewer number of evaluated scenarios, PEM is more efficient than the MC method. Additional techniques can be combined with PEM to further reduce the time consumption and to improve its applicability in large-scale systems (Ciapessoni et al., 2018). It has been utilized in a number of studies, such as security assessment, congestion management and estimation of transfer capability (Ciapessoni et al., 2014; Romero-Ruiz et al., 2016; Adinolfi et al., 2016). Another advantage of the two scenario based methods, i.e., the MC and PEM methods, is their capability to capture correlations across variables, since the dependency across variables can be explicitly represented in the scenario generation process (Liu and Kuireghian, 1986; Saunders, 2014).

Similar to the PPF model, probabilistic OPF (POPF) is also used to address uncertainties in the OPF model. Since PF constraints are embedded in the OPF model implicitly, the methods mentioned above are also applicable to POPF, such as the MC (Xie et al., 2018) and PEM (Verbic and Canizares, 2006; Saunders, 2014) methods. In addition, techniques that are specifically used to handle uncertainties in optimization models can also be applied, which will be detailed in Sections 4.2 to 4.4.

### 4.2. Stochastic optimization

Stochastic optimization belongs to the class of methods for decision making under uncertainty (Birge and Louveaux, 2011). Compared with a deterministic optimization model, a stochastic optimization model optimizes stage-wise decision variables such that the probability-weighted objective function is minimized or maximized. In general, a stochastic optimization model includes multiple time periods separated by time stages at which uncertainties are partially or fully resolved. A stochastic optimization model can take the following form (Infanger, 1992):

\[
\begin{align*}
\min_{x, \omega} & \quad \mathbb{E}_{\omega \in \Omega} [f(x, \omega)] \\
\text{s.t.} & \quad g(x, \omega) = 0 \\
& \quad h(x, \omega) \leq 0 \\
& \quad \forall \omega \in \Omega
\end{align*}
\]

where \( \Omega \) represents the space of a random variable \( \omega \). In practice, the probability space \( \Omega \) can be represented by a set of discrete scenarios, i.e., \( \Omega = \{ \omega_k : k = 1, \cdots, K \} \) and the objective function becomes \( \sum_{k} \mathbb{P}(\omega_k) f(x, \omega_k) \).

Stochastic optimization has been widely used to assist decision-making in many fields, such as finance (Ziemba and Vickson, 2014), manufacturing (Buzacott and Shanthikumar, 1993), and long-term energy system planning (Hu and Hobbs, 2010; Labriet et al., 2012). In power systems, it is used to handle uncertainties in unit commitment, economic dispatch, optimal power flow, bidding strategy and capacity expansion (Zhou et al., 2016b; Alqurashi et al., 2016).

As the formulation in Eq. 2 suggests, a key step in stochastic optimization is the development of an event tree, which is populated with scenario outcomes and corresponding probabilities. The most commonly used method for scenario generation is to sample a given distribution at a node of the scenario tree, or by evolving a stochastic process according to an explicit formula (Kaut and Wallace, 2003). Since the inverse cumulative distribution function (CDF) is used in most scenario generation techniques, distributions of the random variables are often required to have an explicit parametric form. Therefore, parametric forecasts are a natural fit here. Typically used parametric distributions include normal distribution for
energy systems in household, buildings, or microgrids. Therefore, two-stage, or even multi-stage models are
with wholesale power markets, the decision-making process may also present multi-stage characteristics in
ferences are resolved, while the second-stage decisions are scenario specific. In addition, despite structural di-
second-stage decisions involve dispatch decisions of the committed units after uncertainties are resolved.
include commitment of thermal units based on forecasts of load and renewable power generation, while the
power markets can be formulated as a two-stage decision process, where the first-stage decisions usually
in the day-ahead (DA) market and adjust their dispatch strategies according to the latest load and renew-
sequential way, i.e., uncertainties are unfolded at more than one time stages and decisions in later stages
are made upon the realization of random variables. For example, to simulate the two-settlement market
process of real-world power markets, simulations usually include a day-ahead (DA) UC model and a real-
time (RT) ED model, where slow-starting thermal generating units determine their commitment statuses
in the day-ahead (DA) market and adjust their dispatch strategies according to the latest load and renew-
able power generation in the RT market. Following this path, the scheduling and operations in wholesale
power markets can be formulated as a two-stage decision process, where the first-stage decisions usually
include commitment of thermal units based on forecasts of load and renewable power generation, while the
second-stage decisions involve dispatch decisions of the committed units after uncertainties are resolved.
Note that the first-stage decisions remain constant across all scenarios as they are made before uncertainties
are resolved, while the second-stage decisions are scenario specific. In addition, despite structural differences
with wholesale power markets, the decision-making process may also present multi-stage characteristics in
energy systems in household, buildings, or microgrids. Therefore, two-stage, or even multi-stage models are
also employed to optimize operation of such energy systems. In such models, the first-stage variables usually include on/off status of co-generation units, thermal generators, or controllable appliances (Luo et al., 2019). For example, in determining the optimal operating schedule of a microgrid consisting of electricity, heating, and gas systems, the first-stage decision variables include scheduled operations of dispatchable resources, while corrective actions are applied in the second stage to correct deviations (Qiao et al., 2019).

4.3. Robust optimization

The limitations of stochastic optimization, as pointed out by Bertsimas et al. (2013), include the requirement of parametric distributions of random variables and the use of discrete scenarios that considerably increase the computational intensity. In contrast to stochastic optimization, robust optimization defines uncertainty sets to represent uncertainties. Following the form in Eq. (1), the robust counterpart of the optimization problem can be expressed as the following (Bertsimas et al., 2011):

\[
\min_{x} \mathbb{E}_{\omega} \left[ f(x, \omega) \right] \\
\text{s.t.} \\
g(x, \omega) = 0 \\
h(x, \omega) \leq 0 \\
\forall \omega \in \mathcal{U}
\]

where \( \mathcal{U} \) is an uncertainty set that represents all possible values of \( \omega \). The goal of robust optimization is to find the solution that is feasible to any realization of \( \omega \) in the set \( \mathcal{U} \). Therefore, if the set \( \mathcal{U} \) is continuous, there are an infinite number of constraints. Note that despite the similarity between the formulation of robust optimization in Eq. (3) and stochastic optimization in Eq. (2), the goal of robust optimization is to minimize the worst-case objective value over the feasible region, while stochastic optimization minimizes the probability weighted objective value. Previous studies suggest that while stochastic optimization is more computational demanding owing to the use of scenarios (Zheng et al., 2015; Cordova et al., 2018), the solutions from robust optimization tend to be more conservative (Guo and Zhao, 2018).

Robust optimization has been widely adopted in diverse fields (Bertsimas et al., 2011). In power systems, it has been used in transmission planning (Zhang and Conejo, 2018), capacity sizing (Wen et al., 2019), long-term capacity expansion (Henao et al., 2019), optimal scheduling and operation in bulk power systems (Zhu et al., 2019; Wei and Liu, 2019), and distributed energy management systems (Zhang et al., 2018; Ruiz Duarte and Pan, 2019; Brahimi and Amjadi, 2019; Wei et al., 2019b; Zhao et al., 2018; Soares et al., 2018). A central element in robust optimization is the uncertainty set, which is used to represent the uncertainties of random variables. An uncertainty set can be constructed in a variety of forms, such as discrete, box, ellipsoid (Calafiore and El Ghaoui, 2014), and polyhedral (Bertsimas et al., 2011). While most robust counterparts of an arbitrary optimization model are intractable, some models can be transformed and solved in a tractable way (Bertsimas et al., 2011). For example, robust linear optimization with an ellipsoidal uncertainty set can be transformed into a second order cone program (Calafiore and El Ghaoui, 2014), and when the uncertainty set is polyhedral, the model is equivalent to a linear optimization model (Ben-Tal and Nemirovski, 1999). Therefore, a critical step in the development of a robust optimization model is to construct a computationally tractable uncertainty set that can effectively represent uncertainties in real-world systems.

Historical data can be used to inform the construction of uncertainty set (Guan and Wang, 2014). For example, Veysi Raygani (2019) constructed an uncertainty set from intervals bounded by solar power productions during clear-sky and overcast conditions, which were further refined by including historical ramp rates of aggregated solar power generations and controlled by varying confidence intervals. In Soares et al. (2018), the uncertainties associated with solar PV and wind power in a distribution system were represented by a polyhedral uncertainty set constructed from the convex hull of a set of spatial-temporally correlated scenarios. In Wang et al. (2017), distributions of forecasting errors were constructed from historical data, and box uncertainty sets were then obtained by deterministic forecasts plus forecasting errors at a given predictive interval.

12
In addition, forecasts are widely used to construct uncertainty intervals, where the upper and lower bounds of a random variable are often given by its deterministic forecast plus a positive and a negative deviation term. Most frequently, box uncertainty sets are constructed such that the centers of uncertainty intervals are given by deterministic forecasts and a budget of uncertainty is used to control the conservativeness by limiting the sum of interval sizes of all random variables. Such implementations can be found in several studies (Guo and Zhao, 2018; Ebrahimi and Amjady, 2019; Wei et al., 2019b; Wei and Liu, 2019; Zhao et al., 2018; He et al., 2016). Given the CDF of a random variable, quantiles can also be used to construct uncertainty intervals. Huang et al. (2019a) used Gaussian distributions to model forecasting errors of load, wind, and solar power, and uncertainty intervals were given by predictive intervals. Another advantage of robust optimization models over stochastic optimization models is that non-parametric probabilistic forecasts, such as intervals and quantiles, can be directly used. For example, in Cordova et al. (2018), quantiles of probabilistic solar and wind power forecasts were used to construct box uncertainty sets and fed into a robust UC model. Apostolopoulou et al. (2018) used quantiles of solar power forecasts to dynamically adjust the risk levels in the determination of optimal dispatch schedules of a cascade hydroelectric power system coupled with solar photovoltaics.

Similar to stochastic optimization models, robust optimization models can also be constructed in a way that uncertainties are unfolded over stages. Such models are also called adaptive models as the later stage decision variables are adapted after uncertainties are resolved (Bertsimas et al., 2011; Guo and Zhao, 2018). For example, a two-stage robust optimization model can be used to simulate the market scheduling process, where the first-stage decisions include commitment of slow-starting units and the second-stage decisions involve dispatch decisions of the committed units under the worst case scenario (Veysi Raygani, 2019; Guo and Zhao, 2018; Ebrahimi and Amjady, 2019; Wei and Liu, 2019; Zhao et al., 2018). In Veysi Raygani (2019), a two-stage robust optimization model was developed, where the first-stage decisions involved purely commitment statuses of thermal units. In modeling the energy management of microgrids, first-stage decisions may include commitment of co-generation units (Guo and Zhao, 2018), diesel turbines and charging/discharging statuses of storage devices (Zhao et al., 2018), or even the day-ahead flexibility contracted in a distribution system (Soares et al., 2018). Multistage models can also be found in the literature. For example, Lorca and Sun (2017) argued that while most two-stage UC models assume uncertainties are resolved as soon as all unit commitment statuses are fixed, in real-world the uncertainties are not fully resolved until the end of the model horizon. Therefore, they developed a multistage robust UC model to simulate the real-world process of uncertainty realization, where uncertainties were resolved by stage.

### 4.4. Chance-constrained optimization

Given the optimization problem with uncertainty in Eq. (1), chance constraints usually take the following form:

$$\Pr(h(x, \omega) \leq 0) \geq p$$

where, $p$ is an assigned reliability level. The chance constraint indicates that the inequality constraint in Eq. (1) is satisfied up to a given level of $p$.

Chance constraints are frequently used to enforce system reliability. Conventional system reliability is often enforced using strict equality or inequality constraints in the form of load demand balance or reserve availability. By contrast, chance constraints typically enforce security constraints by limiting system risk indices, such as loss-of-load probability (LOLP) and expected energy not served (EENS) (North American Electric Reliability Corporation, 2018). By limiting the LOLP, Wu et al. (2014) used a chance constraint to guarantee the availability of sufficient reserves in the real-time dispatch. Instead of placing upper or lower bounds for system reserve margins, Zhao et al. (2018) limited the probability of reserve shortage in a hybrid concentrating solar power (CSP) and wind power system. In designing a risk-limiting strategy to optimize the restoration of a distribution network after contingency, Wang et al. (2019b) formulated a chance constraint by setting an upper bound for the probability of restored demand not being met by available resources. In van der Heijden et al. (2017), a chance constraint was developed by limiting the probability of net load being greater than the power purchased from long-term power purchase agreement (PPA). In addition, chance constraints can also be used to limit line flows in power flow constraints. In Wu...
et al. (2014) and Labin et al. (2016), chance constraints were enforced to probabilistically bound the line flow within limits under uncertainties of load or renewable generation. To construct chance constraints, parametric distributions of the uncertain variables are often required, however, statistical moments are also used in some studies (Zhao et al. 2019).

Due to the non-linearity of chance constraints, it is often necessary to transform them into tractable forms before it can be solved efficiently. Most chance-constrained optimization models are hard to solve because it is not guaranteed to be convex in general (Calafiore and El Ghaoui, 2014). However, under certain circumstances, chance constraints can be transformed into linear constraints. One prominent example is chance-constrained linear optimization, which is widely utilized in the literature. A general chance-constrained linear optimization model can be expressed in the following form:

\[
\min_x \ c^T x \\
\text{s.t.} \quad \Pr(a_i^T x \leq b_i) \geq p_i, i = 1, \ldots, m
\]

where both the coefficient vector \( a_i \) and constant term \( b_i \) can be random.

When \( a_i \) is constant and \( b_i \) is random, the chance constraint can be directly transformed into a linear inequality using the inverse CDF function, if \( b_i \) can be represented by parametric distribution functions. This type of formulation is widely used in most power system models with chance-constraints. A commonly used distribution is normal distribution (Gao et al., 2019; Ciftci et al., 2019; Wu et al., 2014). For example, Wu et al. (2014) used a censored normal distribution to represent the forecasting errors of load and renewable power productions and the chance constraints can be directly transformed into linear constraints by the probit function, i.e., the quantile function of the standard normal distribution. Wang et al. (2019b) used Gaussian mixture model (GMM) to account for spatio-temporal correlations of renewable power output from different sites, and the distribution of aggregated output was given directly based on the linear invariance property of GMM. The probabilistic constraint was then transformed into a linear constraint using the inverse CDF function. Non-parametric methods can also be used, such as in van der Heijden et al. (2017), the distributions of solar irradiance were obtained by fitting non-parametric distributions to historical hourly solar irradiance data using kernal density estimation (KDE), which was then used to transform chance-constraints into linear inequalities.

In addition, when \( a_i \) follows a normal distribution and \( p_i \geq 0.5 \), the chance constraint can be transformed into a second-order cone constraint (Calafiore and El Ghaoui, 2014; Boyd and Vandenberghe, 2004). For example, Bienstock et al. (2014) transformed a chance-constrained DC-OPF model into a second-order cone program and solved it with commercial convex solvers.

4.5. Hybrid models

Uncertainties in power systems can also be represented by hybrid models. Such approaches are usually used to model multiple uncertain variables, where different constraints may be applied to uncertain variables of different characteristics. Due to the combination of multiple techniques, hybrid methods can be more flexible in dealing with different forms of uncertainty representation. In a typical way, scenarios are used to represent discrete uncertain variables and the uncertainties of variables with continuous distributions can be controlled by robust constraints or chance constraints. For example, in Huang et al. (2019a), the uncertainties associated with continuous variables were modeled with robust constraints, while discrete variables were modeled with scenarios, which were then used as input to stochastic optimization. In Dominguez et al. (2012) and He et al. (2016), the uncertainties associated with solar power output and load were simultaneous considered, where the solar power uncertainty was formulated as robust constraints while scenarios were used to represent electricity price uncertainty. By contrast, in Abdel-Karim et al. (2018), scenarios were used to represent uncertainties in solar and wind power productions, market prices and temperature, whereas box uncertainty sets were constructed to represent end-use demand uncertainty.

Chance constraints can also be incorporated in robust or stochastic optimization models (Wang et al., 2012). Such models usually include constraints with respect to system reliability. In Gao et al. (2019), a stochastic UC model was developed, where the uncertainties in renewable energy productions were represented by scenarios and the balance between demand and supply was enforced with a chance constraint. More
frequently, chance constraints are used to restrict system reliability metrics, such as LOLP and EENS (Roald et al., 2016; Giraldo et al., 2019).

5. Integration of probabilistic solar forecasting in power systems

5.1. Market scheduling and operations

The U.S. power industry has been restructured dramatically by the wave of deregulation in the past, in which wholesale markets have replaced conventional vertically integrated monopolies in many places (Hobbs and Oren, 2019). Although differences may exist in the design and structures of different markets, most real-world wholesale electricity markets implement a two-settlement system that consists of a day-ahead market and a real-time market (Stoft, 2002). The day-ahead market is a financially-binding forward energy market that usually runs one day ahead of the operating day based on forecasts of demand and renewable energy productions. The day-ahead market is a financially-binding forward energy market that usually runs one day ahead of the operating day based on forecasts of demand and renewable energy productions. The differences between day-ahead commitments and actual realizations are balanced in the real-time market, where market participants buy and sell wholesale electricity during the course of the operating day. Both markets are settled independently. Following the two-settlement market design, UC and ED models are usually used to schedule the generating units by minimizing total costs. The UC models, which are run day ahead or in real time, are used to determine the commitment statuses of slow-starting units. The scheduled units will go online according to their commitments and the ED models, which run in real time, are used to produce a least-cost dispatch of online resources and calculate locational marginal prices (LMPs). As shown in Fig. 4, the UC model is run in both the day-ahead and the real-time markets, while the ED model is running every 5 minutes in the real-time market. Typically, both models are formulated as optimization models with security constraints to observe resource and transmission limits (Wood et al., 2013; Hobbs et al., 2006). Currently, market operators only run deterministic UC and ED models, which depend on deterministic forecasts of load and renewable energy productions. The uncertainties of load and renewable energy are conventionally covered by ancillary services, such as regulating reserves, spinning reserves and ramping reserves. However, a variety of probabilistic methods are proposed to address the challenges posed by the growth of renewable energy. Probabilistic forecasting naturally fits in the scope as the uncertainties can be represented in a variety of forms and used in corresponding methods.

![Diagram of market scheduling and operations](image)

**Figure 4:** Market process in CAISO. Note that the RTUC models run every 15 minutes and their horizons may vary, the RTED model runs every 5 minutes and typically covers 65 minutes (CAISO, 2020). The numbers in the parentheses indicate temporal horizons that are covered by different models. A trading hour is any hour during which trades are conducted in a CAISO market.
Table 2: Studies of market scheduling and operations. CC: Chance constraints. SO: Stochastic optimization. RO: Robust optimization.

<table>
<thead>
<tr>
<th>Study</th>
<th>Uncertainties</th>
<th>Form of uncertainties</th>
<th>Uncertainty representation in the model</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. (2015)</td>
<td>Solar, wind, load</td>
<td>Point + error distribution (uniform)</td>
<td>Scenarios</td>
<td>SO</td>
</tr>
<tr>
<td>Córdova et al. (2018)</td>
<td>Solar, wind</td>
<td>Quantiles</td>
<td>Uncertainty set</td>
<td>RO</td>
</tr>
<tr>
<td>Gao et al. (2019)</td>
<td>Solar, wind</td>
<td>Point + forecasting error</td>
<td>Scenarios, chance constraints</td>
<td>CC, SO</td>
</tr>
<tr>
<td>Hanhuawei et al. (2017)</td>
<td>Solar</td>
<td>Day-ahead, parametric, Gaussian</td>
<td>Scenarios</td>
<td>SO</td>
</tr>
<tr>
<td>Huang et al. (2019a)</td>
<td>Solar, wind, load, generator outage</td>
<td>Point + forecasting error</td>
<td>Uncertainty set</td>
<td>RO</td>
</tr>
<tr>
<td>Lorca and Sun (2017)</td>
<td>Solar, wind</td>
<td>Point + forecasting error</td>
<td>Uncertainty set</td>
<td>RO</td>
</tr>
<tr>
<td>Ming et al. (2018)</td>
<td>Solar, hydro</td>
<td>Point + forecasting error</td>
<td>Scenarios</td>
<td>SO</td>
</tr>
<tr>
<td>Moazzami et al. (2018)</td>
<td>Wind, load</td>
<td>Point + forecasting error</td>
<td>Scenarios</td>
<td>SO</td>
</tr>
<tr>
<td>Osório et al. (2015)</td>
<td>Solar, wind</td>
<td>Historical distribution</td>
<td>Discretized states</td>
<td>MC</td>
</tr>
<tr>
<td>Quan et al. (2015, 2016)</td>
<td>Solar, wind, load, generator outage</td>
<td>Prediction intervals</td>
<td>Scenarios</td>
<td>SO</td>
</tr>
<tr>
<td>Roald et al. (2016)</td>
<td>Solar, wind, load</td>
<td>N/A</td>
<td>PDF</td>
<td>CC</td>
</tr>
<tr>
<td>Veysi Raygani (2019)</td>
<td>Solar</td>
<td>N/A</td>
<td>Uncertainty set</td>
<td>RO</td>
</tr>
</tbody>
</table>

Given the probabilistic description of solar power uncertainty, distributions of key system indicators, such as costs, energy not supplied and power production from generators, can be built by solving a set of Monte Carlo scenarios (Osório et al., 2015). In addition, scenarios can also be used as input to stochastic optimization models (Zheng et al., 2015; Moazzami et al., 2018; Gao et al., 2019; Hanhuawei et al., 2017). A typical approach formulates the UC model as a two-stage stochastic optimization model, where the commitment decisions are formulated as first-stage decision variables (Zheng et al., 2015; Wu et al., 2007; Wang et al., 2005). In these studies, probabilistic forecasts are commonly represented by the sum of point forecasts and forecasting errors, which are assumed to follow certain parametric distributions and can be used to generate scenarios. However, due to the intrinsic computational complexity of integer programming, the use of scenarios and the inclusion of security constraints, such models are usually difficult to scale up to real-world sized systems as a result of the curse of dimensionality. By comparison, robust optimization is less computationally demanding and does not require parametric distributions of random variables (Zheng et al., 2015; Pandžić et al., 2016), therefore is more likely to be applied to large-scale systems. For example, Bertsimas et al. (2013) developed a two-stage adaptive robust optimization UC model and applied it in a large scale power system operated by the ISO New England, which has 312 generating units, 174 loads, and 2816 nodes. Although parametric distributions are more common in the literature, non-parametric forecasts, such as prediction intervals and quantiles, can be directly used to construct uncertainty intervals (Veysi Raygani, 2019). Córdova et al. (2018) used quantiles from probabilistic forecasts directly as uncertainty sets to run a robust UC model. By comparing the results with a stochastic optimization model, it was concluded that the robust optimization model presented similar results with less time consumption.

While a day-ahead UC model typically runs once per day, intra-day UC and ED models are repeated periodically to take advantage of the latest forecasts with shorter look-ahead horizons. Market scheduling
may benefit from frequent intra-day UC and ED runs as latest forecasts become available (Warrington et al., 2016). In addition, the horizon of real-world UC and ED models typically span multiple time periods and the forecast horizons of these periods may differ, where forecasts with longer look-ahead time tend to be more uncertain. To simulate the varying degree of uncertainties, multi-stage models with rolling horizons can be used (Xie et al., 2014; Lorca and Sun, 2015). Zhu et al. (2019) developed a multi-timescale robust dispatch model to vary the level of robustness to reflect more accurate forecast as time evolves forward. Lorca and Sun (2017) developed a multistage robust UC model and applied their model to a Polish 2736-bus system with 60 wind farms and 30 solar farms under high-dimensional uncertainties.

5.2. Behaviors of market participants

From an economic perspective, the scheduling and operation of a power market is intended to achieve market equilibrium at a point where the social welfare is maximized. Such models are usually adopted by system operators to clear the market given the supply and demand bids from market participants. However, from the perspective of a market participant, such models no longer apply since the purpose of a market participant is to maximize profit or minimize cost. Market participants can be an end-use customer who purchases electricity from the wholesale power market, a supplier who provides electricity, or a microgrid that exchanges electricity with the main grid. For a consumer, the purpose is often to minimize its cost through optimizing its behavior, while the purpose of a supplier is to maximize its profit of selling electricity to the market. Therefore such problems can be formulated as optimization models.

Traditional market participants usually only include dispatchable resources, such as fossil fuel fired power plants. However, operations of solar power plants are highly weather dependent. Besides, the growing penetration of distributed solar resources, such as rooftop PV panels, brings additional layers of complexity to the optimization of market participant behaviors. Therefore, forecasts can be used to make more reliable and cost-effective strategies for market participants. In van der Heijden et al. (2017), historical distributions of solar irradiance conditioned on time-of-day and season were used to determines the optimal power purchase agreement (PPA) of a small PV-rich community by limiting the chance of net demand exceeding the PPA. Angizeh and Parvania (2018) characterized the output uncertainty from a solar power plant using parametric probabilistic forecasting, which was utilized in a two-stage stochastic optimization model to determine the optimal operating schedule of a large customer equipped with a solar power plant by exchanging electricity with the main power grid. Similarly, in Abedinia et al. (2019), a stochastic optimization model with robust constraints was developed to determine the optimal bidding and offering strategy of a large consumer, who has its own fleet of generators and meanwhile can purchase electricity from bilateral contracts, wholesale markets, or retail markets.
Table 3: Studies of behaviors of market participants. CC: Chance constraints. SO: Stochastic optimization. RO: Robust optimization.

<table>
<thead>
<tr>
<th>Study</th>
<th>Uncertainties</th>
<th>Form of uncertainties</th>
<th>Uncertainty representation in the model</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abedinia et al. (2019)</td>
<td>Solar, wind, load, market price</td>
<td>Point + forecasting error</td>
<td>Uncertainty interval, scenarios</td>
<td>RO + SO</td>
</tr>
<tr>
<td>Apostolopoulou et al. (2018)</td>
<td>Solar</td>
<td>Point + forecasting error</td>
<td>Uncertainty set</td>
<td>RO</td>
</tr>
<tr>
<td>Attarha et al. (2019)</td>
<td>Solar, market price</td>
<td>N/A</td>
<td>Uncertainty interval</td>
<td>RO</td>
</tr>
<tr>
<td>Dominguez et al. (2012)</td>
<td>Solar, market price</td>
<td>History</td>
<td>Uncertainty interval, scenarios</td>
<td>RO + SO</td>
</tr>
<tr>
<td>He et al. (2016)</td>
<td>Solar, market price</td>
<td>History</td>
<td>Uncertainty interval, scenarios</td>
<td>RO + SO</td>
</tr>
<tr>
<td>Kakimoto et al. (2019)</td>
<td>Solar</td>
<td>KDE</td>
<td>KDE</td>
<td>Analytical</td>
</tr>
<tr>
<td>van der Heijden et al. (2017)</td>
<td>Solar</td>
<td>History</td>
<td>Chance constraints</td>
<td>CC</td>
</tr>
<tr>
<td>Wang et al. (2017)</td>
<td>Solar, wind, market price</td>
<td>Point + forecasting error</td>
<td>Uncertainty set</td>
<td>RO</td>
</tr>
</tbody>
</table>

The use of probabilistic methods can help suppliers optimize their operating strategies for a variety of purposes. For example, Kakimoto et al. (2019) used the forecasted distributions of solar power production to analytically design a strategy by maximizing the total profits of bidding into the day-ahead spot market in Japan. Dominguez et al. (2012) and He et al. (2016) used thermal energy storage to improve the flexibility of solar thermal power plants. Both studies determined the optimal bidding strategy of selling electricity to the market under uncertain solar irradiance and volatile market prices. In addition to the energy market, He et al. (2016) also bid into the ancillary service market. In several studies, solar resources are coupled with other types of resources to improve the stability and flexibility of the combined power output. Attarha et al. (2019) optimized the bidding strategy of a solar PV power plant coupled with a battery storage system by maximizing its revenue in ERCOT, where the uncertainties of solar power productions and electricity prices were represented by polyhedral uncertainty sets. Apostolopoulou et al. (2018) demonstrated that the efficiency of a cascade hydroelectric power system can benefit from coupling with solar PV resources. They developed a robust optimization model to determine the optimal dispatch strategy under uncertainties of solar power production and inflow of the cascade.

Unlike in market scheduling models, where market prices are endogenously given by marginal costs under market equilibrium, market participants are typically assumed to be price takers and have no influence on the prices. Therefore, some models also take price volatility into consideration. For example, in Dominguez et al. (2012) and He et al. (2016), market price uncertainty was modeled using scenarios while solar power uncertainty was represented by uncertainty intervals. Wang et al. (2017) used point forecasts plus forecasting errors of wind and solar power to construct uncertainty sets, which were used in a robust optimization model to determine the optimal bidding strategy of a microgrid by maximizing its revenue from day-ahead markets of energy and ancillary services.

5.3. Energy management of microgrids

The electric power industry is undergoing a paradigm shift, where large centralized power plants are being replaced by small distributed energy resources (Vadar 2021). Scheduling and operation models can also be found in studies of distributed systems, such as home energy management (Luo et al. 2019), community level energy systems (Zhao et al. 2018), etc. A common property is that they can be operated either as a part of
the main power grid when synchronously connected, or in an island mode when disconnected (Huang et al., 2019). These power systems can be categorized as microgrids. Microgrids emerge as a result of technology advancement in two-way smart metering, new sensor and controlling devices, energy storage systems, and distributed energy resources. A microgrid is defined as an interconnected system that includes a group of suppliers and consumers of different energy commodities. Similar to bulk power systems, the growing penetration of renewable resources also poses great threats to the cost-effectiveness and reliability of microgrids. Therefore, modeling uncertainties becomes increasingly crucial in these studies. For simple system with small numbers of appliances, heuristic methods, such as exhaustive enumeration can be used. For example, El-Baz et al. (2018a) used a machine-learning method to produce day-ahead probabilistic solar power forecasting for a household rooftop PV system. The forecasts are used in a home energy management system to determine the optimal demand-side management plan for a set of household appliances (El-Baz et al., 2018a). By comparing the results with a deterministic forecast, they concluded that the probabilistic forecasting resulted in better performance in terms of self-sufficiency and self-consumption. Approaches to dealing with uncertainties in bulk power systems can also be adopted in microgrid scheduling and operations, such as chance constraints (Vergara-Dietrich et al., 2019), Ciftci et al. (2019), stochastic optimization (Luo et al., 2019; Lin et al., 2019), and robust optimization (Ebrahimi and Amjady, 2019; Guo and Zhao, 2018; Zhao et al., 2018). Note that microgrids can act as a whole to provide energy and ancillary services to the main grid (Wang et al., 2017). In these cases, the behavior of the microgrid resembles an ordinary market participant and such studies are discussed in Section 5.2.

In contrast to the operation of bulk power systems, where electricity is usually the only energy carrier, regional microgrids may include other forms of energy carriers, energy conversion technologies and end-use demands. Many studies include combined heat and power units (Guo and Zhao, 2018; Eladl and ElDesouky, 2019; Ahn et al., 2019), natural gas flow (Ciftci et al., 2019; Qiao et al., 2019; Rakipour and Barati, 2019), heating and cooling demands (Li et al., 2018; Guo and Zhao, 2018), and even electric vehicle behaviors (Derakhshandeh et al., 2017). The inclusion of additional forms of energy carriers and end-use demand types add additional layers of complexity in these studies. For example, correlations across different forms of end-use demands can be considered and usually represented by scenarios. Ahn et al. (2019) used multivariate normal distributions to represent the correlations across electricity load, heat load, cooling load, and solar irradiance, which were used to generate samples for their Monte Carlo simulation. In Qiao et al. (2019), the correlation between electricity demand and heat demand was modeled with joint distributions to generate scenarios for a two-stage stochastic optimization model. Meanwhile, scenarios of solar PV power production were generated independently. The uncertainties of loads and solar power production were represented collectively by the Cartesian product of the two sets of scenarios. Furthermore, in the model formulation of microgrids, the main power grid is typically modeled as an energy source or sink that can exchange electricity with the microgrid at market prices (Guo and Zhao, 2018). Therefore, price uncertainties are often considered in such models as well.

Several studies simulate the two-settlement system in the bulk electricity market by modeling the operation of microgrids in a hierarchical way. Such models usually determine the on/off status of controllable appliances in the day-ahead schedule, and optimize the power generation/consumption/procurement in the real-time schedule (Fawum et al., 2015; Shams et al., 2018; Luo et al., 2019). The models are constructed such that operating schedules can be updated when the latest forecasts become available. Ebrahimi and Amjady (2019) used day-ahead forecasts of renewable energy production, load and market prices as inputs to an adaptive robust optimization model to optimally schedule the commitment of microturbines, which were assumed to be slow-starting units. In addition, in many studies, a rolling horizon optimization, where the prediction horizon is continuously rolling forward as the latest forecasts become available, is used (Silvente et al., 2015; Luo et al., 2019; Palma-Behnke et al., 2013). For example, Luo et al. (2019) used a rolling horizon optimization scheme to integrate the latest forecasts of solar power productions into a real-time dispatch strategy. Scenarios and point forecasts were used in a mixed manner to relieve the computation burden and enable a longer optimization horizon.
5.4. Ancillary service markets

Although energy and power are the primary service provided by power markets, a variety of ancillary services are also introduced (Stoft, 2002). While depending on specific market, the names, types and functions of the ancillary services may vary (Zhou et al., 2016a), most of them ensure that the supply of delivered power is reliable and of high quality. One of the most important ancillary services is operating reserves (Zhou et al., 2016a; Denholm et al., 2019), which are used to maintain stability of system frequency, an indicator of the balance of real power. Operating reserves are provided by a set of qualified resources with different technical characteristics that are deployed over a wide range of response times. For example, in ERCOT, regulating reserves are typically deployed within seconds to 5 minutes, spinning reserves are required to response within 10 minutes, and non-spinning reserves are deployed within 30 minutes (ERCOT, 2020). Market operators usually use predetermined reserve margins to guide the procurement of operating reserves and a critical step is to estimate the reserve requirement, where an increased reserve margin results in a high reliability level, but at the cost of a greater budget. The uncertainties of renewable energy also bring greater challenges to the determination process and the role of accurate renewable power forecasting becomes increasingly important in the process.
Table 5: Determination of operating reserves using probabilistic solar forecasting. CC: Chance constraints. SO: Stochastic optimization. RO: Robust optimization.

<table>
<thead>
<tr>
<th>Study</th>
<th>Uncertainties</th>
<th>Form of uncertainties</th>
<th>Uncertainty representation in the model</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etingov et al. (2018)</td>
<td>Solar, wind, load</td>
<td>Histograms</td>
<td>Non-parametric distribution, histograms</td>
<td>Analytical</td>
</tr>
<tr>
<td>Gao et al. (2019)</td>
<td>Solar, wind, load</td>
<td>Point + forecasting error</td>
<td>Chance constraints + scenarios SO + CC</td>
<td></td>
</tr>
<tr>
<td>Huang et al. (2019a)</td>
<td>Solar, wind, generator outage</td>
<td>Point + forecasting error</td>
<td>Uncertainty sets + scenarios RO + SO</td>
<td></td>
</tr>
<tr>
<td>Yan et al. (2017)</td>
<td>Solar, load</td>
<td>Statistical, ANN</td>
<td>Parametric distribution</td>
<td>Analytical</td>
</tr>
</tbody>
</table>

Methodologies to determine an appropriate operating reserve level have been constantly evolving (Holttinen et al., 2008; Ela et al., 2010). The requirements of operating reserves are determined based on the uncertainty levels of net load, i.e., the total electric demand minus wind and solar generation. Net load represents the part of load that must be met by dispatchable resources, such as fossil fuel fired units and nuclear power plants. To determine the requirements of operating reserves, market operators usually depend on distributions of forecasting errors of net load, and the operating reserves are required to cover certain predictive intervals. The requirements of reserves are given by percentiles that correspond to pre-specified reliability levels to inform the procurement of regulating reserves (Bruninx and Delarue, 2014, 2017). A typical method is based on the standard deviation of net load and a specific reliability interval. Since net load is a linear combination of multiple components, the uncertainty of net load can be decomposed into uncertainties of all constituent components. One method discussed in Holttinen et al. (2012) used arithmetic or geometric sums of all individual uncertainties to estimate the operating reserves. This method often overestimates the needs and tends to be conservative. The other method in Holttinen et al. (2012) assumed all components followed independent Gaussian distributions and used square sums to estimate the uncertainty levels of net load. Similar study can be found in Yan et al. (2017), where square sums of the prediction errors of load and PV power are used to provide spinning reserves. However, the distributions of solar power and wind power usually follow non-Gaussian distributions. For example, Bruninx and Delarue (2017) used three distributions to model wind power forecasting errors: Gaussian, β-distribution, and Lévy distribution. Therefore, another method to estimate the distribution of net load is convolution (Zhang et al., 2016; Etingov et al., 2018), which can be applied to estimate the distribution of the linear combination of multiple non-Gaussian random variables, even if they are correlated (Li et al., 2020).

In many markets, system operators use historical data to construct the distributions due to its simplicity. However, prior distributions are usually static and fail to reflect future system states. Therefore, probabilistic forecasts can be used to inform system operators of the requirements (Matos and Bessa, 2011). Huang et al. (2019a) determined the requirements of operating reserves by satisfying specific system reliability requirements. Similarly, Gao et al. (2019) found the optimal operating reserve requirements using chance constrains to limit both LOLP and the probability of spillage of wind and solar power. Hypothetical parametric distributions were used to construct the probability distributions. Etingov et al. (2018) developed a method to aggregate day-ahead probabilistic forecasts of load, wind and solar in CAISO and the distribution of net load was used to inform the procurement of regulating reserves in the day-ahead market. Studies show that probabilistic forecasts can improve the cost-effectiveness of market operation without compromising system reliability (Vos and Driesen, 2014; Etingov et al., 2018).

Another way to estimate the reserve size is through cost-benefit analysis. As the reserve size increases, the cost of operating the system becomes more expensive while the system reliability and adequacy improve, which can be valued by the product of value of lost load (VOLL) and expected energy not supplied as a function of the committed capacity. By evaluating the socio-economic benefits due to the improved system reliability and the costs associated with the increased reserve values, optimal reserve requirements can be determined. These studies usually co-optimize the energy provision and ancillary services by sequentially.
running UC and ED models, where operating reserves are formulated with additional constraints. In addition, by varying the reserve margins dynamically, the trade-off between total system costs and socio-economic benefits of operating reserves are explored, and additional auxiliary models can be employed to determine the optimal requirements (Ortega-Vazquez and Kirschen, 2007; Cui and Zhang, 2018; Zhang and McCalley, 2018). The benefits of operating reserves are constantly compared with other probabilistic methods, such as stochastic optimization and robust optimization. For instance, Papavassiliou et al. (2011) presented a two-stage scenario-based stochastic optimization model, which explicitly considered operating reserve provisions. They showed that their method outperformed approaches that determine reserve requirements based on peak-load or hourly forecasted load and wind power. Wang and Hobbs (2016) evaluated flexible ramp products by comparing results from a deterministic real-time UC model and a stochastic model. Bruninx and Delarue (2014) developed and compared three UC strategies: a deterministic UC model, a scenario-based stochastic UC model, and a deterministic UC model with probabilistic reserve constraints.

5.5. Power system security assessment

Up until now, the main emphasis of the discussed power system applications are placed on power system economics only. The other overriding factor in the operation of power systems is security. A power system needs to operate securely so that any possible single contingency does not cause the system to deviate from the normal state, which usually means there is no voltage violation or line overload (Wood et al., 2013). Since the state of a power system is constantly evolving, critical system state parameters such as currents, voltages and frequencies must be monitored and telemetered to the system operator periodically. The gathered information is used to inform the system operator of possible violations such that necessary remedial actions can be taken (Morison et al., 2004). To successfully meet this goal, it relies heavily on the application of good engineering tools, such as power flow models.

Simple analytical methods can be used directly to estimate system risk levels. For example, in Alamri et al. (2018), the impact on system reliability of a solar farm was estimated by both analytical and MC methods. In the analytical estimation of the PDF of solar power output, the forced outage rate is also explicitly included in addition to the solar radiation uncertainty. The PDF is then convolved with the load PDF to assess the overall system reliability in terms of LOLP and EENS. However, to give detailed risk information on specific lines or buses in a network, more sophisticated methods must be employed. Given probabilistic description of load and renewable power generation, a PPF model can be used in the assessment of power system security under uncertainties (Li et al., 2015; Hamon et al., 2016; Huang et al., 2019). Typically, probabilistic distributions of key state variables are given by the model and used to evaluate system risk indices, such as the probabilities of line overload and voltage violation. For example, Romero-Ruiz et al. (2016) evaluated the probabilities of line overload and voltage violation by scrutinizing the CDFs of all output variables in a distribution network. Kabir et al. (2010) used a PPF model to identify over-voltage issues in a residential distribution network with high penetration of PV generation. Chiapessoni et al. (2014) assessed the the operational risk of a Sicilian power system in terms of LOLP using PPF. They also demonstrated the influence on the risk levels from the forecast lead time and correlations across forecasting errors of renewable power. Similar framework was combined with principal component analysis (PCA) and used in a follow-up study of a pan-European grid with 9241 buses (Chiapessoni et al., 2018).

Results of the security assessments can then be used to inform corrective actions to restore system stability. For example, congestion management was used to relieve line overload by generation rescheduling and load shedding, which were usually formulated as optimization problems (Pillay et al., 2015; Romero-Ruiz et al., 2016; Reddy, 2017; Yusoff et al., 2017; Prajapati and Mahajan, 2019).

In some studies, the network security requirements are embedded in scheduling and operation models as additional constraints. Such constraints are used to ensure that the system remains secure despite the forecast uncertainties. This type of model formulation usually takes the form of hybrid model, where the security constraints can be formulated as chance constraints or robust constraints. For example, Roald et al. (2016) used a chance-constrained OPF model to limit the probabilities of line overload and reserve shortage at the same time. More discussions on the use of chance constraints are already given in Section 4.3.

Typically, the security indices, such as LOLP, EENS, probability of reserve shortage or line overload, can be restricted in this way. In Giraldo et al. (2019), an unbalanced three-phase distribution system was modeled
as a probabilistic OPF model under uncertainties of wind velocity, solar irradiance and demand. Lower and upper bounds of nodal voltages and current magnitudes were limited by robust constraints.

Table 6: Studies of power system security. CC: Chance constraints. PPF: Probabilistic power flow. MC: Monte Carlo. PEM: Point estimate methods.

<table>
<thead>
<tr>
<th>Study</th>
<th>Uncertainties</th>
<th>Form of uncertainties</th>
<th>Uncertainty representation in the model</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adinolfi et al. (2016)</td>
<td>Solar, wind, load</td>
<td>Point + forecasting error</td>
<td>PDF</td>
<td>PPF</td>
</tr>
<tr>
<td>Ciapessoni et al. (2014)</td>
<td>Solar, wind, load</td>
<td>Point + forecasting error</td>
<td>PDF</td>
<td>PPF</td>
</tr>
<tr>
<td>Ciapessoni et al. (2018)</td>
<td>Solar, wind, load</td>
<td>Point + forecasting error</td>
<td>PDF</td>
<td>PPF</td>
</tr>
<tr>
<td>Girakdo et al. (2019)</td>
<td>Solar, wind, load</td>
<td>Normal distribution</td>
<td>Scenarios + Uncertainty sets</td>
<td>CC PPF (PEM)</td>
</tr>
<tr>
<td>Kabir et al. (2016)</td>
<td>Solar, load</td>
<td>Normal distribution</td>
<td>Scenarios</td>
<td>PPF</td>
</tr>
<tr>
<td>Prajapati and Mahajan (2019)</td>
<td>Solar, wind, vehicle</td>
<td>Point + forecasting error</td>
<td>Scenarios</td>
<td>MC</td>
</tr>
<tr>
<td>Roald et al. (2016)</td>
<td>Solar, wind</td>
<td>Normal distribution</td>
<td>Chance constraints</td>
<td>CC OPF</td>
</tr>
<tr>
<td>Romero-Ruiz et al. (2016)</td>
<td>Solar, wind, load</td>
<td>Normal distribution</td>
<td>PDF</td>
<td>PPF</td>
</tr>
</tbody>
</table>

6. Conclusion

Forecasts are widely used for reliable and cost-efficient operations of power systems. However, most forecasts are deterministic forecasts. Although probabilistic wind forecasting is adopted in some markets, most of them use it to improve situational awareness and it rarely plays any important role in the decision-making process. In addition, compared with wind forecasting, probabilistic solar forecasting is still in its early stage. This study represents a first review on the use of probabilistic solar forecasting in power systems, which include a brief summary of probabilistic solar forecasting techniques, an overview of current use of probabilistic forecasting in power systems, and potential methods to adopt probabilistic forecasting in power systems. By giving a thorough review on the potential methods and latest trends in the literature to promote the integration of probabilistic solar forecasting, we have the following findings.

First, many probabilistic methods have been proposed in the literature to handle various forms of uncertainties, such as wind and load. Naturally, they can also be used to integrate probabilistic solar forecasting into the decision-making process of real-world power markets. Each method has its own advantages and disadvantages and a correct tool should be selected accordingly. For example, the spatial and temporal correlations across loads and renewable energy productions are crucial in the modeling process. It may have adverse impacts on power system efficiency and reliability to neglect such correlations. Scenario based methods, such as the MC method and stochastic optimization, are better at capturing such correlations, while they tend to be more computationally demanding and often require scenario reduction. In contrast, robust optimization requires less computation time and can be applied to large-scale power systems. However, the solution tends to be conservative.

In addition, while probabilistic forecasts can take a variety of forms, such as PDFs, uncertainty intervals or quantiles, they must be converted into a correct form to be used as inputs of a specific method. For example, scenarios are used in the MC method and stochastic optimization, while uncertainty sets must be constructed for robust optimization. Therefore, most of the reviewed studies have a dedicated section to discuss the representation of the modeled uncertainties. Our review indicates that parametric distributions
are more frequently used in most studies since they are the most fundamental form. However, as van der Meer et al. (2018c) point out, non-parametric forecasts, particularly quantiles, dominate the studies of probabilistic solar forecasting. Therefore, there is a clear gap between probabilistic solar forecasting and the use of probabilistic solar forecasting. Particularly, future studies should bridge the gap by placing more emphasis on the use of non-parametric probabilistic solar forecasting, as it is the predominant form and typically has better performance.

Despite the fact that most of the review studies explicitly take into account solar power uncertainties, only a few studies use de facto probabilistic forecasting techniques [El-Baz et al., 2018a,b; Córdova et al., 2018; Kakimoto et al., 2019] to produce uncertainty distributions or quantiles of solar power, while most studies use historical or even hypothetical data (i.e., a priori distributions) to construct parametric distributions of solar power uncertainty. Although it may be a direct result of the immature status of the forecasting technique, it implies that more studies should be conducted in this area. In particular, as suggested by some studies (Yang et al., 2019), the forecasts should be made specifically based on the timeline of real-world systems for better integration.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number DE-EE0008215.

Disclaimer

This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References


Hollands, K.G.T., Suehrcke, H., 2013. A three-state model for the probability distribution of instantaneous solar radiation, with


Hodge, B.M., Doubleday, K., Lara, J.D., Feng, C., Li, B., Zhang, J., Jascourt, S., Cassidy, C., Du, P., Chang, S., Sharma, S.,


Hollands, K.G.T., Suehrcke, H., 2013. A three-state model for the probability distribution of instantaneous solar radiation, with
doi:10.1016/j.solener.2013.07.007
doi:10.1016/j.ijforecast.2015.11.011
doi:10.1016/j.enconman.2015.02.083
doi:10.1016/j.apenergy.2015.11.003
doi:https://doi.org/10.1016/j.renene.2015.09.011
doi:10.1016/j.solener.2013.04.012
doi:10.3390/en10101591


Li, W., Thirumalana, K., Tushar, W., Yue, C., Chew, K.T., Tai, S., 2018. Improving the operation of solar water heating systems in green buildings via optimized control strategies. IEEE Transactions on Industrial Informatics 14, 1646–1655. doi:10.1109/TII.2018.2797018


