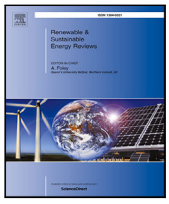
Renewable and Sustainable Energy Reviews 140 (2021) 110735

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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



Post-processing in solar forecasting: Ten overarching thinking tools

Dazhi Yang [a,*], Dennis van der Meer [b]

[a] Singapore Institute of Manufacturing Technology, Agency for Science, Technology and Research (A*STAR), Singapore
[b] Department of Civil and Industrial Engineering, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Keywords:
Solar forecasting
Post-processing
Review
Probabilistic forecasting

ABSTRACT

Forecasts are always wrong, otherwise, they are merely deterministic calculations. Besides leveraging advanced forecasting methods, post-processing has become a standard practice for solar forecasters to improve the initial forecasts. In this review, the post-processing task is divided into four categories: (1) deterministic-to-deterministic (D2D) post-processing, (2) probabilistic-to-deterministic (P2D) post-processing, (3) deterministic-to-probabilistic (D2P) post-processing, and (4) probabilistic-to-probabilistic (P2P) post-processing. Additionally, a total of ten overarching thinking tools, namely, (1) regression (D2D), (2) filtering (D2D), (3) resolution change (D2D), (4) summarizing predictive distribution (P2D), (5) combining deterministic forecasts (P2D), (6) analog ensemble (D2P), (7) method of dressing (D2P), (8) probabilistic regression (D2P), (9) calibrating ensemble forecasts (P2P), and (10) combining probabilistic forecasts (P2P), are proposed. These thinking tools can be thought of as the "style" or "mechanism" of post-processing. In that, the utilization of thinking tools circumvents the common pitfalls of classifying the literature by methods (e.g., statistics, machine-learning, or numerical weather prediction), which often leads to a "who used what method" type of roster review that is clearly ineffective, non-exhaustive, and dull. When myriads of post-processing methods are mapped to countable few thinking tools, it allows solar forecasters to enumerate the styles of adjustment that could be performed on a set of initial forecasts, which makes a post-processing task clearly goal-driven. Besides the thinking tools, this paper also emphasizes on the value of post-processing, and provides an outlook for future research. Although this paper is revolved around solar, the materials herein discussed can also be applied to wind and other forecasting areas.

1. Introduction

A good weather forecast is characterized by its consistency, quality, and value [1]. Consistency is defined by the correspondence between forecasts and judgment; quality refers to the correspondence between forecasts and observations; and value denotes the incremental benefits of forecasts to users, e.g., in monetary term. These three types of goodness of forecasts were first outlined by Allan Murphy [1] in his seminal essay on "What is a good forecast?" Since then, the weather forecasting community has made remarkable advancements towards these goals.

In the field of energy meteorology, the two topics, namely, solar irradiance forecasting and solar power forecasting, are jointly known as "solar forecasting" [2]. Modern solar forecasting is said to start from the early 2010s; it is surely new. However, most would agree that solar forecasting has come a long way, in a very brief period of time. Despite being relatively underdeveloped as compared to other energy forecasting domains, such as wind, load, or electricity price forecasting—so

Abbreviations: AnEn, analog ensemble; ARIMA, autoregressive integrated moving average; ANN, artificial neural network; AOD, aerosol optical depth; BLP, beta-transformed linear pool; BNI, beam normal irradiance; CCM, cross-correlation method; CDF, cumulative distribution function; CH-PeEn, complete-history persistence ensemble; CRPS, continuous ranked probability score; DTW, dynamic time warping; ECDF, empirical cumulative distribution function; ECMWF, European Centre for Medium-Range Weather Forecasts; EMOS, ensemble model output statistics; ETS, exponential smoothing; GAMLSS, generalized additive models for location, scale and shape; GHI, global horizontal irradiance; HRRR, High-Resolution Rapid Refresh; IGN, ignorance score; JMA, Japan Meteorological Agency; KMA, Korea Meteorological Agency; LSTM, long short-term memory; MOS, model output statistics; MSE, mean square error; NAM, North American Mesoscale; NCAR, National Center for Atmospheric Research; NWP, numerical weather prediction; NGR, nonhomogeneous Gaussian regression; PDF, probability density function; PeEn, persistence ensemble; PIT, probability integral transform; PV, photovoltaic; QR, quantile regression; QRF, quantile regression forest; QRNN, quantile regression neural network; RAMS, Regional Atmospheric Modeling System; RAP, Rapid Refresh; RMSE, root mean square error; SVR, support vector regression

* Corresponding author.
E-mail address: yangdazhi.nus@gmail.com (D. Yang).

https://doi.org/10.1016/j.rser.2021.110735
Received 29 May 2020; Received in revised form 8 January 2021; Accepted 12 January 2021
Available online 22 January 2021
1364-0321/

Nomenclature

Indexes

τ	$\tau \in [0,1]$, used to index quantiles
$\tilde{t}$	Half-window size, only appeared in AnEn formulation
h	Forecast horizon
i	Index of the fitting sample, $i = 1, \dots, n$
j	Index of the component models (when used in subscripts), or index of the weather variables (when used in superscripts), $j = 1, \dots, m$
m	Number of component models, or number of weather variables, depending on context
n	Size of the training sample
n'	Size of the entire sample, thus, $n' - n$ gives the size of the verification sample
N_v	Number of weather variables, only appeared in AnEn formulation
t	Index of the verification sample, $t = n+1, \dots, n'$

Symbols

β_0, β_j	Coefficients for regression
$\mathbb{E}$, $\mathbb{V}$	Expectation and variance operators
F, f	Generic variables denoting CDF and PDF
F_{ij}, f_{ij}	The ith initial predictive distribution from the jth component model in the fitting sample
F_{tj}, f_{tj}	The tth initial predictive distribution from the jth component model in the verification sample
g	The general notation for functions, e.g., the link function, or a nonlinear function
G_i	The combined predictive distribution corresponding to F_{ij}, $j = 1, \dots, m$
L, U	Lower and upper bounds of a prediction interval
q_τ	$q_\tau = F^{-1}(\tau)$, denotes the τth quantile
w_0, w_j	Weights for combination
X	Generic variable denoting "forecasts"
x_i	The ith initial forecast in the fitting sample, $i = 1, \dots, n$
$x_i^{(j)}$	The ith initial forecast of the jth weather variable in the fitting sample
x_t	The tth initial forecast in the verification sample
$x_t^{(j)}$	The tth initial forecast of the jth weather variable in the verification sample
x_{ij}	The ith initial forecast from the jth component model in the fitting sample
x_{tj}	The tth initial forecast from the jth component model in the verification sample
Y	generic variable denoting "observations"
y_i	The ith observation in the fitting sample
y_t	The tth post-processed forecast in the verification sample
y_t^*	The tth observation in the verification sample

claimed by general-purpose forecasting experts [3]—thousands of articles on solar forecasting have been published in the past decade, see [2,4] for general reviews on solar forecasting. It is, now, well accepted that solar forecasting has evolved into a stand-alone forecasting domain with a unique set of tools and salient features [5]. Most notably is the use of sky/shadow cameras, satellite imagery, numerical weather prediction (NWP), and other physically-based methods, which clearly differentiates solar forecasting from other energy forecasting domains. Moreover, as exemplified by a recent paper by Nouri et al. [6], the skills required to generate good solar forecasts are highly specialized, and are often associated with a steep learning curve. That said, the general goal of solar forecasting is no different from other weather forecasting, in that, one aims for consistency, quality, and value.

A major drawback of physically-based methods is our incomplete understanding about the dynamical laws governing weather events [7]. Even if a complete understanding for a particular subject matter could be achieved, it is often the data that limits the accuracy of a physically-based model—one simply cannot cover the Earth's surface with weather stations, radiosondes, and buoys [8]. Therefore, what solar forecasters wish to receive from physically-based models is a good initial guess. And the remaining task is to post-process that initial guess, with data-driven and machine-learning models.

In a forecasting context, *post-processing* refers to procedures performed, to further improve the goodness of forecasts, after some initial forecasts have been generated. As trivial as it sounds, if a forecaster knows that the forecasts generated by a forecasting system are always 10% higher than the actual value, he would certainly deduct whatever forecasts he receives from that system by 10%. Clearly, the actual situation is a lot more complex than that, but the general idea applies—a forecaster adjusts the forecasts based on past experience. Such past experience could be derived from data or be based on subjective judgment. The former is almost always used in forecasting in physical science, whereas the latter is frequently involved during forecasting in a social setting [9]. This review classifies, compiles, summarizes, and synthesizes various post-processing frameworks used by solar forecasters. These frameworks are referred to as *thinking tools*, that is synonymous with "styles" or "mechanisms".

There is a clear distinction between a post-processing thinking tool and a post-processing method. The former refers to a general strategy to achieve a goal, where the particular pathway taken to get there is not emphasized. On the other hand, a method is but one of the pathways. To give perspective, regression is a thinking tool, whenever the relationship between a predictand and several predictors is sought. In achieving that goal, one can opt for multiple linear regression, lasso regression, ridge regression, multilayer perceptron, support vector regression, radial basis function network, and the list goes on, as long as the data permits the choice. The endless list of available methods makes enumeration an impossible task. To that end, consolidating thinking tools is obviously the better strategy to review the existing post-processing methods in the literature. That said, a major component of any review paper is to define systematic typologies for the topic at hand, hence, a typology of post-processing is presented next.

1.1. Typology of post-processing

There ought to be more than one way to classify post-processing methods. Most obviously, one can classify post-processing methods based on the subject to which they belong, e.g., statistics or machine learning. Alternatively, one may consider dividing the methods into univariate and multivariate, endogenous and exogenous, batch and sequential, among other classifiers. However, in the context of forecasting, classifying post-processing methods based on the nature of forecast—i.e., deterministic or probabilistic[1]—seems particularly attractive.

[1] It is noted that there are different opinions on what can be considered as a deterministic or probabilistic forecast. In this paper, the word "deterministic" means "point" or "single-valued", whereas the word "probabilistic" refers to any forecast representation that offers a notion of uncertainty, such as a predictive distributions, a set of quantiles, a prediction interval, or an ensemble of point forecasts—such interpretation is inline with the latest recommendation from the energy forecasting literature [5].

The reason for opting this typology is due to the recent surge in recognition of the importance of probabilistic modeling in the solar community [10]. In contrast, deterministic solar forecasting was the norm at the beginning. This could be largely attributed to the fact that many physically-based methods, such as image-based forecasting or NWP, are seen as deterministic, in that, the forecast is simply a number. While this was anachronistic (e.g., ensemble NWP has been popular since at least the 1960s), solar forecast practitioners often do not have access to ensemble NWP forecasts, nor were they familiar with those statistical methods for uncertainty quantification. Consequently, virtually all of the most influential early works on physically-based solar forecasting report only deterministic forecast performance [e.g., 11–14]. Today, it is quite clear that deterministic forecasts can be converted to probabilistic forecasts seamlessly, and vise versa. Hence, as the name "post-processing" suggests, it is quite natural to perform the classification based on the direction of that conversion.

In this regard, post-processing methods are herein classified into four categories:

1. deterministic-to-deterministic post-processing;
2. probabilistic-to-deterministic post-processing;
3. deterministic-to-probabilistic post-processing; and
4. probabilistic-to-probabilistic post-processing.

In what follows, each of the four categories of methods will be reviewed, with respect to various thinking tools for each class. However, several fundamental concepts need to be made clear first, to facilitate the understanding of subsequent materials.

1.2. Basic concepts of probabilistic forecasts of continuous variables

This paper is restricted to probabilistic forecasts of continuous variables, such as solar irradiance, clear-sky index, or solar power. Probabilistic forecasts of a continuous variable may take the form of predictive distributions, quantile forecasts, or prediction intervals. More specifically, whether or not a forecast is probabilistic depends on whether it allows uncertainty quantification. Hence, ensemble forecasts are also probabilistic, which will be discussed in Section 1.3.

A statistical distribution can be parametric, semiparametric (i.e., mixture), or nonparametric. Since predictive distributions are statistical distributions, they can also be classified as such. To give perspective, when a normal predictive distribution is used, it is parametric, and is fully characterized by two parameters, namely, mean and variance. When the predictive distribution is a mixture of two normal distributions—e.g., to annotate the chance of having clear and cloudy sky conditions—it is semiparametric. Lastly, when the predictive distribution is composed of a mixture of many normal distributions, each corresponding to a different outcome of the forecast, it becomes a kernel density representation, and thus is nonparametric. Fig. 1 provides an illustrative example of these types of predictive distribution.

The cumulative distribution function (CDF) of a random variable X is $F(x) = \mathbb{P}(X \leq x)$, where $\mathbb{P}$ denotes probability. The inverse CDF, or quantile function, is defined by

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}, \quad \text{for } \tau \in [0, 1]. \tag{1}$$

The value of $F^{-1}(\tau)$ is known as the τth quantile. One often writes $q_\tau = F^{-1}(\tau)$. Hence, probabilistic forecasts are often represented in terms of quantiles, instead of predictive distributions. For example, in the Global Energy Forecasting Competition 2014 (GEFCom2014) [3], the contestants were asked to submit the 1st, 2nd, … , 99th percentiles, i.e., q_τ with $\tau = 0.01, 0.02, \ldots, 0.99$.

With the concept of quantiles, another type of probabilistic forecasts can be defined, i.e., prediction interval (PI). A PI contains a lower bound and an upper bound, denoted with L and U, respectively. With a nominal coverage probability of $(1 - \alpha)$ centered on the median, the bounds are simply:

$$[L, U] = [q_{\alpha/2}, q_{1-\alpha/2}]. \tag{2}$$

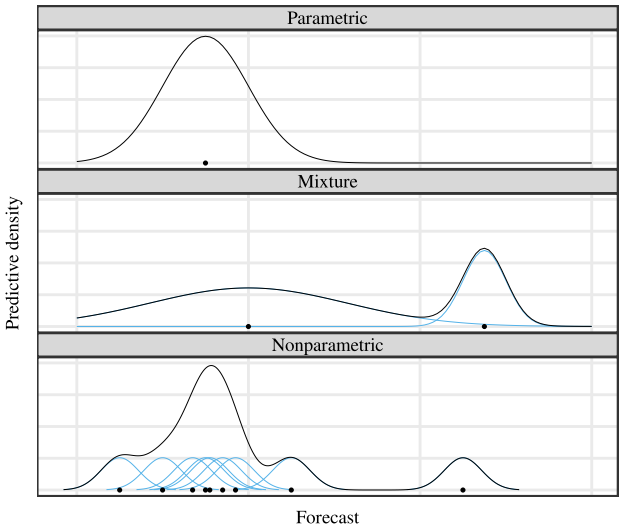


Fig. 1. Three types of predictive distribution (parametric, mixture, and nonparametric).

For instance, if $\alpha = 0.05$, which corresponds to a 95% nominal coverage probability, the lower and upper bounds of the 95% PI are the 0.025th and 0.975th quantiles. It should be noted that PIs need not be symmetric around the point forecast, but they often are.

1.3. Probabilistic representation using ensemble forecasts

A special form of probabilistic representation of a future event is an ensemble of forecasts. Ensemble forecasting produces different versions of the forecast of the same event using different *data*, *models*, and *parameters*. Roulston and Smith [15] presented the glossary of different ensemble forecast types, in which *dynamical ensemble* and *poor man's ensemble* are well known by solar forecasters, despite their perhaps unfamiliar names.

Dynamical ensemble is an ensemble obtained by evolving perturbed initial conditions forward in time, using the same dynamical model. It is a form of *data ensemble*. The 51-member ensemble forecasts produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) is a dynamical ensemble. On the other hand, poor man's ensemble refers to an ensemble that consists of forecasts generated by different models. For example, if one gathers the forecasts generated by different NWP models, such as Rapid Refresh (RAP), High-Resolution Rapid Refresh (HRRR), or North American Mesoscale (NAM) model, the forecaster obtains a poor man's ensemble. The poor man's ensemble is a form of *model ensemble*, e.g., see [16]. The third type of commonly used ensemble, which was not mentioned in [15], is a *parameter ensemble*. Considering a machine-learning model with some parameter, θ, to be tuned, instead of using just one θ value, one could use a set of θ's, namely, $\boldsymbol{\Theta} = \{\theta_1, \ldots, \theta_m\}$, each producing a forecast.

The forecasts produced by ensemble members are known as *component forecasts* [17]. Given m deterministic component forecasts, a predictive distribution can be constructed using several approaches. Firstly, one can assume that these component forecasts are samples from a known (semi)parametric statistical distribution with unknown parameter θ. Then, by estimating θ from the samples, the predictive distribution can be characterized. For example, if a normal predictive distribution, $\mathcal{N}(\mu, \sigma^2)$, is assumed, then μ takes the value of the ensemble mean, whereas σ^2 takes the value of the ensemble variance. Secondly, it is possible to use kernel density estimation, as shown in the bottom panel of Fig. 1.

Another group of approaches is to construct an empirical CDF (ECDF) from the deterministic component forecasts. Three such approaches have been summarized by Lauret et al. [18]. Since a CDF

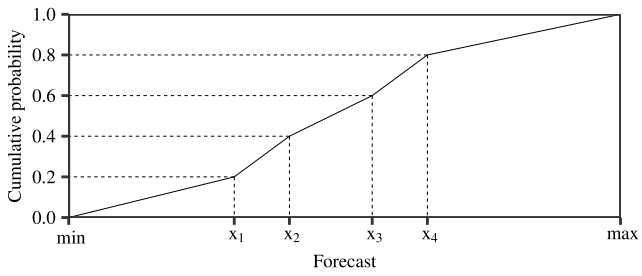


Fig. 2. One approach to construct ECDF from m deterministic component forecasts. A case with $m = 4$ is illustrated, and the forecasts are denoted as $x_1, \ldots, x_4$. The minimum and maximum forecasts can be assigned based on experience, e.g., assigning zero and the extraterrestrial irradiance to min and max, respectively.

provides a continuous mapping from a real line to the [0,1] interval, i.e., $F : \mathbb{R} \rightarrow [0, 1]$, the component forecasts act as "knots" in ECDF construction. As a result, the three approaches only differ from each other in terms of how those "knots" are joined or connected. Fig. 2 illustrates one of the three approaches—uniform spacing with linear interpolation. In this approach, the m component forecasts are simply the $1/(m + 1), 2/(m + 1), \ldots, m/(m + 1)$ quantiles. For more information on ECDF construction from component forecasts, the reader is referred to [18].

The individual component forecasts themselves could also be probabilistic, each carrying a notion of uncertainty. In this case, it is possible to further develop additional perturbed forecasts around each component forecast. This approach often relies on the forecast error statistics. Such an ensemble generated around an individual member is known as a *daughter ensemble* [15]. When all daughter ensembles from all component forecasts are used collectively, it is called a *hybrid ensemble* [15].

1.4. An overview on forecast conversion

A good deterministic forecast is simply a point value that reflects the single best guess of a forecaster, hence it is also known as a *point forecast*, a *single-valued forecast* or a *best-guess forecast*. Despite the name "best guess", it may be further adjusted. Such interventions aim at correcting the bias, and sometimes, the variance, of the initial forecasts, using frameworks such as regression or filtering. However, the latest research also reveals another type of post-processing on deterministic forecast, namely, forecast resolution change. In that, a temporally coarse forecast can be downscaled to forecasts with a higher temporal resolution, using pattern matching or synthetic generation. Henceforth, the deterministic-to-deterministic post-processing is abbreviated as D2D.

Probabilistic-to-deterministic post-processing, or P2D, may appear trivial, because one could simply summarize a predictive distribution through a statistical functional, namely, mean, median, or occasionally, other quantiles. However, the choice of the statistical functional requires some attention. This is related to the aforementioned concept of consistency. Besides summarizing predictive distribution, when the probabilistic representation is in the form of deterministic ensemble forecasts, various *forecast combination* methods exist. Forecast combination has been extensively studied in general-purpose forecasting. In solar forecasting and some sister domains, forecast combination is also known as *model blending*.

The most representative framework for deterministic-to-probabilistic post-processing, or D2P, is analog ensemble (AnEn). With a single realization of a future event, the forecaster seeks its analogs—weather patterns that resemble each other—from the past. Subsequently, past observations that correspond to these analogs are used as equally likely outcomes of that future event. AnEn relies on forecast error statistics. Instead, it is also possible to just sample, i.e., bootstrap, the past forecast errors, and "dress" them on the deterministic-style forecast. Although this may appear to be straightforward, when the samples are drawn from conditional error distributions, the procedure to determine an appropriate set of conditional variables could become tedious. Lastly, many regression methods, e.g., linear, quantile, or random forest regression, allow uncertainty quantification. A third class of methods is to conduct (probabilistic) regression between observations and forecasts.

Lastly, probabilistic-to-probabilistic post-processing, or P2P, calibrates ensemble forecasts, or combines different probabilistic forecasts, to form a final probabilistic forecast. Improving the calibration can be done by mapping ensemble forecasts to a parametric predictive distribution under some objective function, or by dressing additional members or kernels onto an ensemble forecast with some constraints. In terms of probabilistic forecast combination, one can either aggregate predictive distributions, predictive quantiles, or prediction intervals. These are fairly new subject areas, and not many references on solar forecasting are available. Therefore, the references for this part mostly come from other related domains, where early formalisms have been established.

1.5. The goal and structure of this review

Fig. 3 graphically shows the four categories of post-processing described above, as well as their respective thinking tools.

> These four categories and ten thinking tools are basically meant as a typology of post-processing. They help forecasters to crystallize what various post-processing methods are essentially doing. In turn, by making correspondence with these categories and thinking tools in future works, a unified understanding on post-processing could be achieved.

The remaining part of the review is organized as follows. The four categories of post-processing are separated into Sections 2–5, respectively. However, it is noted that the sizes of these sections are not balanced, due to the different amount of literature available. For instance, P2P post-processing is new, even in the field of general-purpose forecasting. Hence, it clearly calls for more attention from solar forecasters, and thus is bulkier than the rest.

In each of the four sections, respective thinking tools are motivated accordingly. Subsequently, each thinking tool is placed and discussed in a subsection. Relevant references are provided: whereas the important ones are discussed in more details, suggested readings are only introduced briefly. It should be clarified that, in the literature, post-processing that leverages more than one thinking tool is rare. This is because of the goal-driven nature of post-processing, e.g., removing bias or calibrating ensemble forecasts, which should be relaxed into a single-stage problem instead of a multi-stage one. In this regard, methods hybridizing two thinking tools are not discussed in this review.

Section 6 provides an outlook into the future of solar forecast post-processing. Several emerging topics, such as hierarchical forecasting, spatial post-processing, or irradiance-to-power conversion, are discussed. These topics are thought to bring major value-add to the large-scale integration of solar power. On this point, Section 7 revolves around the value of post-processing, and outlines how post-processed forecasts can attain value, particularly, in a grid-integration context. Concluding remarks follow at the end.

2. Deterministic-to-deterministic post-processing

As stated earlier, the bulk of the solar forecasting literature on physically-based methods, namely, camera-based, satellite-based, and NWP-based solar forecasting, only focuses on deterministic predictive performance. Naturally, solar forecasters have proposed many D2D

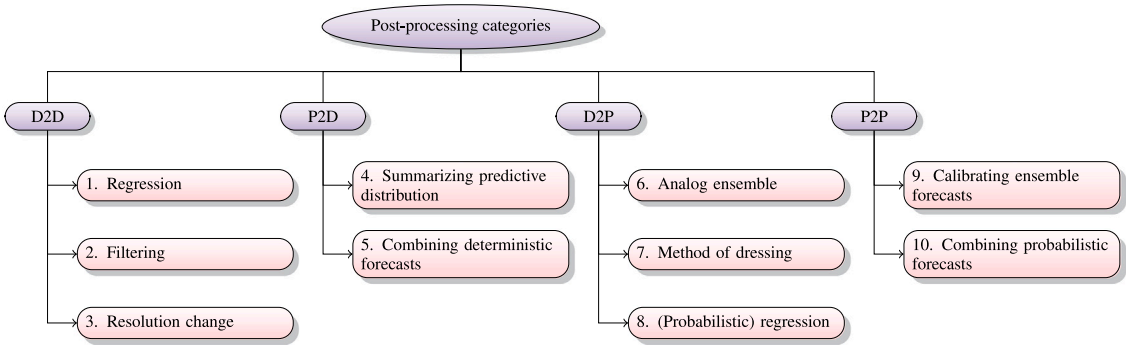


Fig. 3. Typology (blue) and thinking tools (red) of post-processing in solar forecasting. The letters "D" and "P" refer to deterministic forecasts and probabilistic forecasts, respectively, indicating the direction of conversion.

post-processing methods to improve these physically-based forecasts. For legacy reasons—particularly due to the typology introduced by Inman et al. [19] and Kleissl [20] in their early reviews—unlike the other sections where materials are arranged according to Fig. 3, the survey in this section is sub-divided based on post-processing used in (1) camera-based forecasting, (2) satellite-based forecasting, and (3) NWP-based forecasting. In this way, readers who are already familiar with the early literature can immediately appreciate how post-processing can be applied to existing deterministic solar forecasting methods. But before that, the general frameworks of D2D post-processing are first motivated and elaborated.

Thinking tool 1 (*Regression*)**.** *Regression establishes a relationship between initial forecasts and observations. The predictand of such a regression is the observation (or sometimes, bias of the initial forecast), whereas the predictors are (potentially multivariate) forecasts.*

D2D post-processing is quite versatile, in a sense that any sequential hybrid of two or more deterministic prediction models would belong to this class. The general idea of D2D post-processing is to establish a deterministic correspondence between forecasts and observations via historical samples. Subsequently, when a new forecast becomes available, one can use that established correspondence to perform a deterministic re-forecast. Arguably, the simplest D2D method is a linear regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{3}$$

where $i = 1, \dots, n$ indexes the fitting samples; x_i denotes ith historical forecast; y_i denotes ith historical observation; and ε_i is a zero-mean homoscedastic error. Then, with the estimated values of β_0 and β_1—often denoted using $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively—one can issue a post-processed forecast, y_t, after a new forecast, x_t, is issued by a forecasting system:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t. \tag{4}$$

It should be noted that, usually, when a prediction or a forecast is denoted, it often comes with a hat, i.e., $\hat{y}_t$; since the entire review is about issuing forecasts, the hat notation is saved throughout, see nomenclature for details. Additionally, in this paper, *verification samples* are indexed using t, with $t = n+1, n+2, \dots, n'$. By using the simple regression procedure in Eqs. (3) and (4), the overall bias in the initial forecasts can be effectively removed. A good collection of popular naïve bias-correction methods is available in [21]. Notwithstanding, a vital prerequisite for D2D post-processing is that the error characteristics of the forecasting system must be consistent. In an extreme case, if all forecasts used during model training over-predict, whereas all subsequently forecasts for testing under-predict, the trained model will not work.

Thinking tool 2 (*Filtering*)**.** *Filtering stabilizes forecasts over several forecast horizons. It also allows sequential update of forecasts as more observations become available.*

Besides bias removal, it is also possible to correct the variance, or dispersion, of the forecast. The variance in this case refers to the spread of a series of out-of-sample deterministic forecasts, and one should not confuse the wording with variance of a predictive distribution. To stabilize the variance, various filters are useful in making the series of forecasts less dispersed. The primary reason for filtering forecasts is to reduce the unnatural spikes in forecasts, which often contribute significantly towards the mean squared error. While it is possible to filter the forecasts after they are generated, another common approach is to directly embed the filter during forecasting, as in the case of classic time series forecasting methods. The reader is referred to two books by Box et al. [22] and Hyndman et al. [23] for a complete guide on time series forecasting.

Thinking tool 3 (*Resolution Change*)**.** *The temporal resolution of data and that of the required forecasts often differ. For upscaling, one simply aggregates the data before forecasting, whereas for downscaling, the initial low-resolution forecasts need to be converted to the required high-resolution forecasts.*

Recent research reveals another type of D2D post-processing, namely, forecast resolution change, in which the forecasts generated at one resolution are post-processed to forecasts at another resolution. The need for resolution change originates from two aspects: (1) when multiple sets of forecasts (with different resolutions) are generated for the same future event, some aggregation consistency is required, in that, several high-resolution forecasts should be summed up (or averaged out) exactly to a low-resolution forecast; and (2) when high-resolution forecasts are needed (e.g., a 5-h block of 15-min forecasts are required for the real-time market of the California Independent System Operator), but only the low-resolution forecasts are available (e.g., hourly forecasts generated by an NWP model). Whereas the first aspect, namely, aggregation consistency, has been fully resolved by Yang et al. [24]—using a method known as the *temporal hierarchical reconciliation*, first proposed in [25]—the second aspect is detailed in Section 2.3.3.

2.1. Post-processing camera-based forecasts

The basic workflow of camera-based forecasting is (1) capturing consecutive images of the sky, (2) processing the images, (3) retrieving cloud properties, such as cloud height, opacity, or motion, (4) projecting the cloud field forward in time, and (5) performing ray-tracing. The reader is referred to [26] for an overview on this topic. Obviously, the main difficulty for camera-based forecasting lies in the cloud property retrieval and ray-tracing. With a single camera, it is virtually impossible to determine cloud-base height—geometrical reconstruction of clouds requires two or more cameras [27,28]. Consequently, the single-camera forecasting methods often yields negative forecast skill, with reference to the clear-sky persistence (diurnal-cycle adjusted persistence). As for

ray-tracing, some authors circumvent the step, by directly feeding the processed image pixel values into machine-learning models as exogenous inputs, and thus make forecasts from a "black box" [e.g., 29,30], albeit there is an obvious decline in such works due to the lack of physical considerations.

To boost the accuracy of raw camera-based forecasts, Chu et al. [31] used an artificial neural network (ANN) as a post-processing tool. The ANN used both the recent measurements of power output from a 48 MW PV system and the raw camera-based forecasts as inputs, and output final forecasts with improved quality. It was shown that the skill score of 15-min-ahead forecasts increased from −18.9% to 26.2%, whereas the improvement for 5-min-ahead forecasts is from −71.3% to 15.1%.

The ANN approach used by Chu et al. [31] is a regression-based post-processing procedure. On the other hand, a filtering-based approach was shown in [32]. Using a two-camera system, and through a sequence of forecasting steps, including cloud detection and classification, cloud-base height estimation, cloud motion vector estimation, and clouds/shadows-to-irradiance conversion, 1-min-resolution beam normal irradiance (BNI) forecasts over a 15-min horizon were obtained. Nonetheless, Blanc et al. [32] noted that these forecasts often contain very-short-term spikes, that may be attributed to false cloud detection or cloud occlusions. In this regard, applying a median filter, with a 5-min window, to the forecast time series was found effective in filtering out these spikes, with a trade-off in terms of increasing misaligned ramps. The overall correlation between forecasts and observations, however, increased from 0.65 to 0.76, after applying the median filter. For a theoretical discussion on median filter, and its distinction from the traditional linear filters, the reader is referred to the book by Tukey [33], where the median filter was popularized.

Besides directly post-processing the forecasts, appropriate procedures may also be applied to results from different stages in a camera-based forecasting workflow. For instance, it is common to post-process the pixel-wise velocity field, obtained through optical flow or cross-correlation method (CCM), into a single average cloud velocity vector before advection [12,34]. When the initial optical flow or CCM leaves void areas, one needs to post-process and fill those voids [35]. Since it is not the focus of this review to discuss the permutation and combination of post-processing with various steps in camera-based forecasting, a few names are recommended instead—Jan Kleissl, Philippe Blanc, Carlos Coimbra, and Zhenzhou Peng—their works jointly cover most, if not all, possibilities.

2.2. Post-processing satellite-based forecasts

The workflow of satellite-based forecasting is similar to that of camera-based forecasting, where forecasters leverage cloud properties derived from visible and infrared images taken by instruments onboard geostationary weather satellites, and subsequently advect the cloud field forward in time that leads to forecasts. Due to its large spatial extent (usually between ±65° in latitude), satellite-based forecasting is deemed suitable for intra-day forecasting up to 6-h ahead. Historically, satellite images are only available at an hourly resolution, limiting the forecast resolution. The latest generation of weather satellites, however, is able to capture images at an sub-hourly scale, e.g., the 16-channel multispectral imager onboard Himawari-8 provides full disk observations every 10 min, and observations over Japan every 2.5 min. For a fairly up-to-date overview on the satellite-based forecasting workflow, the reader is pointed to the reviews by Miller et al. [36], Blanc et al. [37].

Needless to say, when cloud properties are used in forecasting, post-processing is imperative. The first and foremost challenge is a phenomenon known as the spatial mismatch [38,39]. Because satellite-derived irradiance represents the pixel-average value, it is necessary to downscale that value to a point location where the PV system resides, see [40–42] for example. Particularly important for high-resolution data (such as the 5 min and 2 km resolution of the National Solar Radiation Database) are the shadows cast by clouds, which need to be adjusted for sensor viewing parallax displacement and combined with solar geometry and terrain height to determine the actual location of cloud shadows [36]. Another contributor to the cloud-to-irradiance conversion is aerosol. Aerosols attenuate and scatter the incoming solar radiation, and thus must be considered during radiative transfer. The spatial mismatch between the pixel-average aerosol optical depth (AOD) and point-location AOD often benefits from a scale-height correction [43].

Another class of approaches to generate satellite-based forecasts is to directly construct statistical or machine-learning forecasting models on the satellite-derived irradiance [44]. These approaches are no different from forecasting through a ground-based sensor network, where auto- and cross-correction of irradiance time series are exploited [45,46]. Stated differently, the pixels of a satellite-derived irradiance database can be viewed as sensors, located regularly in space. This types of data is known as the lattice data, for which a large number of spatio-temporal inference and prediction methods are available, as best summarized in the book by Cressie and Wikle [47].

Satellite-derived irradiance is generally regarded as a suboptimal source of solar radiation data. Hence, it is necessary to perform site adaptation [21,48], which is a form of preprocessing. Nonetheless, *ex ante* bias correction via preprocessing does not eliminate the need to perform *ex post* processing. A method based on Kalman filtering was demonstrated graphically in [37].

2.3. Post-processing NWP forecasts

NWP models consider the atmospheric dynamics. Hence, saying NWP models are complex would be an understatement. With a few notable exceptions [e.g., 49–52], the utilization of NWP in solar forecasting is limited mostly to downloading and post-processing the forecasts issued by national weather centers. However, because solar irradiance is but one of hundreds of output variables, it is often subject to significant bias that varies with season, sky condition, climate, among other factors.

2.3.1. Model output statistics

One of the earliest attempts of correcting NWP-based solar forecasts was made by Lorenz et al. [53], who considered a very popular approach called model output statistics (MOS). The term "MOS" was first coined by Glahn and Lowry [54]. The basic concept of MOS is to express the bias of the NWP forecast as a regression. For example, a fourth-degree polynomial was used in [53],

$$\text{bias}_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_i^{(j)}, \tag{5}$$

where $x_i^{(j)} \in \{\cos^4 z_i, \dots, \cos z_i, \pi_i^4, \dots, \pi\}$; z is the zenith angle, which can be calculated through a solar positioning algorithm; and π is the clear-sky index forecast (the ratio between surface irradiance forecast and its clear-sky expectation) issued by the NWP model. Once the unknown coefficients, $\beta_0, \beta_1, \dots, \beta_8$, are fitted, one can predict the bias whenever a new forecast is issued:

$$\text{bias}_t = \hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_j x_t^{(j)}. \tag{6}$$

The scatter plots between bias and $\cos z$, π are visualized in Fig. 4, using forecasts issued by the NAM model at the Desert Rock station, Nevada, over the year 2016. The need for using polynomial regression—beyond linear and quadratic fit—is apparent.

That said, it is again noted that there are infinite possibilities to construct such regressions. Any data-driven or machine-learning method with regression capability would suffice. For example, neural-network-based models were demonstrated in [55,56], whereas stepwise

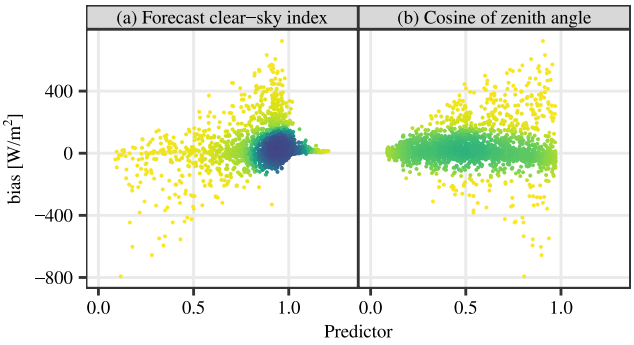


Fig. 4. Forecast bias of the North American Mesoscale model against (a) the forecast clear-sky index, and (b) the cosine of zenith angle. The data come from the Desert Rock station, Nevada, over the year 2016. Darker color indicates more points in the neighborhood.

regression was used in [57,58]. Moreover, one can also employ hybrid models that leverage several regression methods in cascade [59,60]. To add more facets to the problem, one may also include different input variables, as long as those input variables can be obtained at the same time when forecasts are issued. This basically means that any combination of NWP output variables (such as temperature, humidity, pressure, or wind speed) can be paired with any (hybrid) regression method (such as multiple linear regression, lasso, multilayer perceptron, or radial basis function) to make a "novel" MOS method. To correct such false sense of novelty, Yang [61] analyzed the properties of Lorenz's MOS, i.e., Eq. (5), that led to its success. Subsequently, a total of four criteria were proposed to align future MOS proposals, so that their incremental contribution to the literature can be quantifiable. More specifically, any proposed MOS method should be: (1) flexible and robust in handling different input-variable sets, (2) computationally efficient in terms of model diagnosis time, (3) easily implementable with high reproducibility, and (4) able to attain decent improvements over Lorenz's MOS.

Moving away from methodology, several other aspects of MOS require attention. Firstly, as mentioned earlier, the performance of NWP models is often season-, weather-, and climate-dependent. Hence, it would not be meaningful to use the same set of regression coefficients or model parameters to post-process NWP forecasts throughout a year. Instead, regressions with a rolling-window setup, or regime switching, are preferred, as shown in [62].

Secondly, MOS coefficients need to be fitted, in that, ground-based observations are required to compute the bias. Since ground-based measurement is not available everywhere, one has to rely on alternative observations—remote-sensed irradiance data. NWP post-processing via satellite-derived irradiance was proposed in [61]. Although post-processing NWP forecasts using satellite-derived irradiance would lead to marginal loss of accuracy as compared to that using ground-based measurements, gridded datasets are available for all locations on Earth, resolving the issue on training data availability—see [63] for a world-wide validation of hourly data from 8 gridded irradiance databases over 27 years. On this point, the more general idea of gauging forecast accuracy without *in situ* measurements, but using satellite-derived irradiance, has also been discussed [64,65].

2.3.2. Kalman filtering

Another particularly popular D2D post-processing method is Kalman filtering. Unlike all previously mentioned methods that post-process forecasts in batch, Kalman filtering is a sequential method. Being a remarkable engineering innovation, Kalman filtering has been extensively studied and used in various scientific domains. The reader is referred to [22] for a systematic introduction to its applications in time series forecasting. It should be noted that Kalman filtering is able to produce forecasts on its own, see [66] for a case study, however, in most solar forecasting works, it is used as a post-processing model.

Pelland et al. [67] gave an early attempt of using Kalman filtering to post-process NWP forecasts. A rather simple two-component observation model was adopted. This setup was subsequently extended to a three-component model by Diagne et al. [68]. However, as pointed out by Yang [69], Kalman filtering, as a sequential procedure, requires the most recent observation values to update its state equation in each iteration. Hence, when day-ahead NWP forecasts are passed through the filter, it becomes a sequence of hour-ahead forecasts. In this regard, most Kalman filtering implementations in the solar forecasting literature cannot be considered as a day-ahead forecast post-processing method, since they changes the forecast horizon from day-ahead to hour-ahead.

Yang [69] discussed two possible remedies for the forecast-horizon issue, but both were thought to have limited post-processing power. In the conclusion of [69], the author noted that Kalman filtering through an ordered set of analog forecasts [70] might be of interest, but has not been considered in a solar forecasting context.

2.3.3. Changing forecast resolution

Although dynamical models can evolve an analysis forward in time with an arbitrary step size, most of the NWP forecasts are issued in hourly or 3-hourly intervals. This is mostly due to storage concerns. Nonetheless, most power system operations, especially those for the real-time markets, require forecasts to be submitted at a higher resolution, e.g., 5-min or 15-min intervals [71,72]. To that end, if a forecaster wishes to leverage NWP output for intra-hour and intra-day forecasting in an operational context, the coarse hourly forecasts need to be downscaled to high-resolution ones. It is noted that the word "downscaling" and "upscaling" should not be confused. Since it is the timescale that is being changed, converting from low-resolution (big timescale) to high-resolution (small timescale) forecasts is downscaling.

Because most of the solar forecasting works are still purely academic exercises, studies that follow the exact operational requirements are extremely rare. To the best of our knowledge, there is only *one* paper that depicts such a case study for intra-day forecasting. Yang et al. [73] performed operational solar forecasting under the real-time market setting of the California Independent System Operator (CAISO). Starting from a time series of NWP forecasts issued by NAM, a pattern-matching algorithm was used to match that hourly NWP forecast time series to hourly-aggregated ground-based measurements observed in the past. Through this step, the best candidate, in terms of the smallest Euclidean distance, can be found. Subsequently, the high-resolution (15-min) version of this candidate was used as forecasts—recall the hourly measurement time series being matched was aggregated from high-resolution data. This process is illustrated in Fig. 5. A similar forecasting framework was presented for DNI forecasting in [74], where the authors referred to their method as "dynamic paths".

Whereas the similarity between a forecast time series and an observation time series can be defined through Euclidean distance [73], it is also possible to employ other similarity measures, in particular, those based on dynamic time warping (DTW). However, as the historical dataset gets large, the search for best candidate becomes slow, because of the necessity of computing all-pair Euclidean distances or DTWs. Although fast search methods are rarely studied in solar forecasting, the reader is expected to benefit from reading related publications in the computer science literature, e.g., [75–77], where fast similarity-search algorithms are available at bulk quantities.

Aside from pattern-matching-based methods, synthetic generation is another possible class of approaches that allows downscaling. Synthetic generation of high-resolution solar irradiance is typified by the works of Jamie Bright, Âzeddine Frimane, and Joakim Munkhammar [e.g., 78–81]. Markov chain is a popular approach to synthetically generate high-resolution data, where subsequent high-resolution samples are

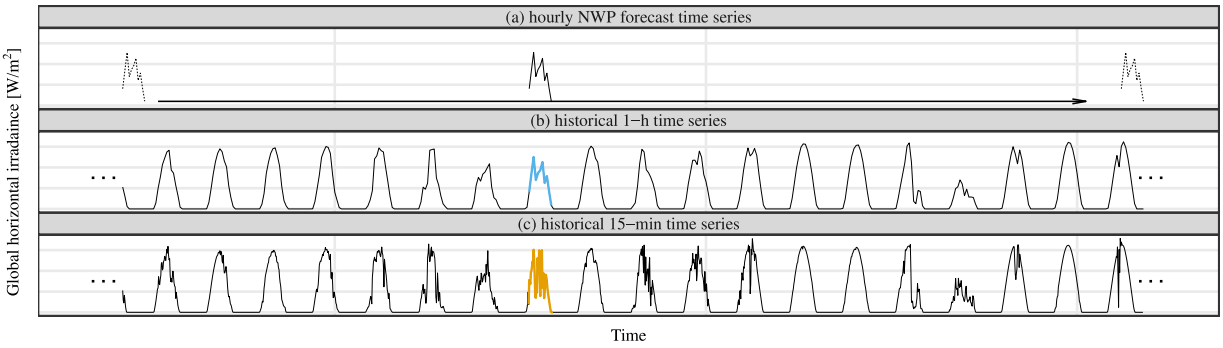


Fig. 5. An illustration of the method for changing forecast resolution used in [73]. For a given 1-h NWP forecast time series, see panel (a), the forecaster seeks a pattern from a large historical database that best resembles the forecast series, see panel (b). After the best match (shown in Malibu) is found, the corresponding high-frequency history sub-series (ground measurements aggregated to 15-min resolution, shown in gamboge), is used as the downscaled forecasts.

governed by a transition probability matrix derived from historical samples [82,83].

That said, regardless of using pattern matching or synthetic generation, the downscaled forecasts can only be close to the corresponding observations, statistically, i.e., in terms of distribution. In other words, one can ensure the variability of the downscaled forecasts to be similar to that of the observations, it is not possible to magically infer intra-hour cloud transient from hourly data alone. As a result, the ramps of the downscaled forecasts may miss the actual materialized ramps quite terribly. This is however not an issue for large-scale grid-integration applications, where fluctuations from many PV systems over a large area tend to cancel out—a phenomenon known as *geographical smoothing* [84,85].

3. Probabilistic-to-deterministic post-processing

Whereas probabilistic forecasts offer information related to uncertainty associated with forecasts, a one-number summary is needed if the user of the forecast is only interested in the best guess. In contrary to the common belief that P2D is straightforward—e.g., why not just calculate or retrieve the mean—the theory for summarizing probabilistic representations of forecasts ought to be carefully argued.

Thinking tool 4 (*Summarizing Predictive Distribution*)**.** *In P2D post-processing, predictive distribution should be summarized with respective to a forecast directive, while maintaining the consistency between forecasts and judgment.*

Thinking tool 5 (*Combining Deterministic Forecasts*)**.** *Given m deterministic component forecasts, a forecaster is tasked to optimally combine them, and issue a final forecast that (hopefully) has an improved predictive performance over the best component forecast.*

Sections 1.2 and 1.3 have outlined the various forms of probabilistic forecasts that a forecaster is able to receive. On one hand, if a probabilistic forecast is in the form of a predictive distribution, one seeks to summarize it through a statistical functional. On the other hand, if a probabilistic forecast is in the form of an ensemble, different ways to combine the component forecasts are sought. These two thinking tools of P2D post-processing are elaborated in Sections 3.1 and 3.2 , respectively.

3.1. Summarizing a predictive distribution

A predictive distribution can be summarized either explicitly or implicitly. Explicit summary may appear straightforward: after a predictive distribution is obtained, one uses the mean of that predictive distribution as the deterministic forecast. However, summarizing using the mean is not always optimal, due to the potential conflict between forecasts and judgment. Section 3.1.1 gives a brief account for such conflict, as well as ways to resolve that.

Implicit summary could be thought of as a competitive model selection process, before any forecast is issued. This is typified by the automated forecasting using the autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS) families of models. It is well known that every ARIMA model is associated with a process order triplet, namely, (p,d,q), whereas the ETS state space model contains 30 variations based on different combinations of error (E), trend (T), and seasonal (S) components. The choice of (p,d,q) in ARIMA or the (E,T,S) triplet in ETS is often selected through an information criterion, such as the Akaike information criterion, see [86,87] for a demonstration of such procedures in a solar forecasting context. Since forecasts made using different choices of (p,d,q) or (E,T,S) could be considered as a parameter ensemble (recall the three types of ensemble in Section 1.3), which can be considered as a predictive distribution, the automated model selection process summarizes that predictive distribution. To that end, although the final forecasts appear to be deterministic, the method still could be considered as a post-processing step, where the post-processing is performed implicitly, in terms of selecting the best model from the model space. Following the argument, another class of implicit model selection method resides in the machine-learning literature—*ensemble learning*. Section 3.1.2 discusses that in details.

3.1.1. Consistency as a type of goodness of forecasts

The statistical theory of making and evaluating deterministic forecasts has been formalized by Gneiting [88]. The main theory is revolved around the concept of *scoring function*, denoted using S. For instance, given one forecast (X) and one observation (Y), the squared error is $S(X,Y)=(X-Y)^2$; the absolute error is $S(X,Y)=|X-Y|$; the relative error is $S(X,Y)=|(X-Y)/X|$; and one can see how scoring functions work. Through a simulation example, Gneiting [88] showed that during forecast verification, one constantly observes conflicting results, in which one method performs well under one scoring function can perform poorly under another scoring function. This is directly related to the notion of *consistency*, which is defined as the correspondence between forecasts and judgment. Formally, let x^* denote an evaluation of the functional, and let x be any forecast, a scoring function S is consistent for the functional $T(\cdot)$ if:

$$\mathbb{E}_F\left[S(x^*,Y)\right] \le \mathbb{E}_F\left[S(x,Y)\right], \tag{7}$$

for all F (predictive distribution), all $x^* \in T(F)$, and all x within its valid range. And it is *strictly consistent* if the equality implies that $x \in T(F)$. In words, the above inequality states that the expected loss of using a consistent functional to summarize a predictive distribution is less than or equal to the expected loss of any other way of summarizing the predictive distribution.

Whereas the original article spent several pages discussing and proving various rigorous definitions, theorems, and principles related to the concept of consistency, which makes most forecast practitioners bewildered and perplexed, the end results can easily be appreciated. The most notable result is that the squared error scoring function is strictly consistent for the mean functional relative to F with finite second moment, and the *pinball loss*, namely,

$$S_\tau(X,Y) = \begin{cases} \tau|X-Y| & \text{for } X \le Y, \\ (1-\tau)|X-Y| & \text{for } X \ge Y, \end{cases} \tag{8}$$

is strictly consistent for the τth quantile relative to F with finite first moment. It is also noted that when $\tau = 0.5$, the pinball loss reduces to the absolute error. Hence, the absolute error is strictly consistent for the median functional relative to F with finite first moment.

The notion of consistency allows a forecaster to optimally choose a functional to summarize a predictive distribution. Stated differently, suppose the forecaster is interested in having a deterministic forecast with the smallest squared error with respective to an available F, he shall choose the mean of that F. Similarly, if the smallest absolute error is of interest, then summarizing F using the median is obvious. In terms of verification, the correspondence between forecasts and judgment is ensured by selecting a scoring rule that is consistent with the forecast directive. For example, given the directive of "forecast the mean of your probabilistic judgment", the forecaster ought to report his forecasts based on squared error.

Nonetheless, Jolliffe [89] pointed out that such strategy is essentially circular, in that, a forecaster can start by choosing and optimizing a scoring function, and the consistent directive naturally follows. This is quite the norm in solar forecasting, or any other forecasting domain for that matter. For example, when a set of forecasts based on the least squares fit is compared to another set produced by minimizing the absolute error, when everything else is fixed (i.e, same data, same verification length, etc.), the set that uses least squares fit naturally would outperform the other, under squared loss. To that end, claims of superiority based on such experimental design are most likely inconclusive.

3.1.2. Ensemble learning

A special class of methods known as ensemble learning summarizes a predictive distribution *implicitly*. The notion of uncertainty is represented through the parameter space. The fundamental idea of ensemble learning is to train multiple base learners and combine their predictions into a single output that often result in a better performance than any other any base learner. Frameworks of ensemble learning include bagging, boosting, randomization, option trees, stacking, among others. Voyant et al. [90] provided a brief introduction to some of these ensemble-learning frameworks, as well as their applications in solar forecasting.

Ferlito et al. [91] considered three ensemble-learning models, namely, random forest, cubist, and extreme gradient boosting, to forecast the PV power output from a 1 kWp system with 5 variables. It was found that the training mode (online versus offline) and training data length have a non-negligible effect on forecast accuracy. Although such dependence is generally well known, one can never overstate the importance of ensuring sufficient samples for training. For a good example of train–test split in day-ahead, intra-day, and intra-hour scenarios, the reader is referred to the publicly available dataset provided by Pedro et al. [92].

Deep learning is gaining unprecedented attention in almost all scientific domains, and solar forecasting is no exception. Although deep learning is prone to over-fitting, it can be protected via dropout regularization, which has become a standard practice in deep learning. A dropout layer blocks a random set of cell units in each iteration of training, and those cells do not receive and do not transmit information. As noted by Srivastava and Lessmann [93], this procedure is essentially a form of bagging. An excellent example on day-ahead spatio-temporal solar forecasting using the long short-term memory (LSTM) network can be found in Srivastava and Lessmann [93], using the 2014 AMS Kaggle Competition dataset and an additional satellite-derived irradiance dataset.

Ensemble learning frameworks can also be used together with classic time series forecasting framework. Bergmeir et al. [94] evaluated an bagged version of ETS in conjunction with the STL decomposition, and showed that it outperforms the original ETS models consistently on the M3 Competition dataset. It should be noted that the bagging was performed on the remainder series after the STL decomposition, since one prerequisite is that the series being bootstrapped has to be stationary. What this implies is that in the context of solar forecasting, one cannot directly perform bagging on irradiance or power time series. Rather, bagging should at least be performed on the clear-sky index. However, as shown by Yang [95], even the best clear-sky models, such as REST2 [96] or McClear [97], do not lead to stationary clear-sky index. Therefore, further data treatment is needed before ensemble learning can be applied. One such example is given in [98], where the author showed that ARIMA can be used as a preprocessing tool, in that, the residuals are not only stationary but also normally distributed.

3.2. Combining deterministic forecasts

As noted in the introduction, ensemble forecasts are probabilistic. The process of condensing a set of ensemble forecasts to a single forecast is known as *forecast combination*. Ever since the early formalism of Bates and Granger [99], combining forecasts has been studied extensively in most, if not all, forecasting domains. With no surprise, the literature on forecast combination is so bulky, that any single attempt to summarize it would not be exhaustive. Fortunately, through time, a series of milestone reviews have been written by researchers from various fields. In this paper, some good reviews from different eras are listed, and the readers can expand their understanding from these reviews and the references therein. The recommended reviews on forecast combination include Clemen [100], Diebold [101], Diebold and Lopez [102], de Menezes et al. [103], Armstrong [104], Timmermann [105], Wallis [106], Atiya [107], following a chronological order.

3.2.1. Training-free heuristics

Given m component forecasts, $x_{t1}, \ldots, x_{tm}$, the general form of forecast combination is:

$$y_t = f(x_{t1}, \ldots, x_{tm}), \tag{9}$$

where y_t is the combined forecast, and f is some function. Clearly, the function can be linear or nonlinear, among which some can be quite complex. However, the most obvious choice, namely, the simple average, or $y_t = m^{-1}\sum_{j=1}^{m} x_{tj}$ is often found hard to beat [108,109]. It is noted by Armstrong [104] that one should use equal weights unless there is strong evidence to support unequal weighting of forecasts. This is particularly true when the forecaster is unsure about which method is the best. In solar forecasting, the performance of all models is subject to weather and climate conditions, see [110] for a case study with 68 machine-learning models at 7 geographically distinct locations. Hence, simple averaging appears to be an attractive choice when combining solar forecasts made using model ensembles (recall the three types of ensemble in Section 1.3).

That said, any individual forecast may be subject to large errors, e.g., due to human error or data error. In this case, the wisdom of crowds is often found useful to eliminate some of the obvious outliers in the ensemble. Instead of simple averaging, one can perform a trimmed averaging by removing the highest and lowest forecasts from an ensemble—just like how the final score of a diver is calculated after removing the highest and lowest scores. To ensure a sufficient sample size, Armstrong [104] recommended using trimmed averaging

only if there are at least five component forecasts. This directly links to another important consideration in forecast combination, namely, forecast diversity.

Diversifying ensemble forecast is a theory that is so basic and so well tested, yet so often overlooked in solar forecasting. For instance, André et al. [111] used only two models in their satellite-based forecasting, whereas Huang et al. [112] blended only three models in their NWP-based forecasting. These practices violate the *forecasting principles* of Armstrong [104], and thus should be avoided in future solar forecasting research. On the other hand, an outstanding application of forecast combination has been demonstrated by Haupt and Kosović [113], in which seven NWP models including the aforementioned NAM, HRRR, and RAP, are blended through a system called DICast.

3.2.2. Linear combination

An extension of the simple averaging method is to assign different weights for different component models. Some of the most popular methods are summarized in [102]. More specifically, these methods can be categorized into variance–covariance methods and regression-based methods. An example of the variance–covariance methods is combination through mean square error (MSE), i.e.,

$$y_t = \sum_{j=1}^{m} \frac{\frac{1}{\widehat{\mathrm{MSE}}_j}}{\sum_{j=1}^{m} \frac{1}{\widehat{\mathrm{MSE}}_j}} x_{tj}, \tag{10}$$

where $\widehat{\mathrm{MSE}}_j$ is the *estimated* MSE of the jth component model. Similarly, the linear regression,

$$y_t = \hat{w}_0 + \sum_{j=1}^{m} \hat{w}_j x_{tj}, \tag{11}$$

with the *estimates* of some unknown weights, $\hat{w}_0, \hat{w}_1, \ldots, \hat{w}_m$, typifies the regression-based methods. In one way or another, these combination methods require training. The popular statistics software R has several packages, e.g., the `ForecastComb` package [114], that offer standard implementations of these variance–covariance and regression-based methods.

One problem of forecast combination is that when m is large, e.g., $m > 100$, the contribution from each component model becomes opaque. Stated differently, it is generally unclear which predictor (component forecast) contributes more towards the final predictand (final combined forecast). This is known as the predictor selection problem, and existing solutions are available. In particular, the reader is referred to the books by Hastie et al. [115] and Miller [116] for overview and fundamentals of penalized regression. A ridge-regression-based solar forecast combination is demonstrated in [117], in which the authors considered an 158-member ensemble coming from 6 meteorological centers, including ECMWF, China Meteorological Administration, UK MetOffice, Korea Meteorological Administration, Centro de Previsão Tempo e Estudos Climáticos, and Météo-France. Another solar forecasting application with large m was presented in [118], where the lasso regression is used together with a pre-selection algorithm.

3.2.3. Nonlinear combination

The regression-based combination also allows the use of nonlinear functions, in particular, the machine-learning models. The overall goal is to find out how each member of the ensemble forecast is linked to the corresponding observation, as well as how different members interact. The two major tasks that can be accomplished by machine learning are classification and regression. From this viewpoint, one can certainly expect the model pool to be large. Moreover, machine learning has become an integral part of many modern data-oriented analyses, and standard implementations are available at bulk in most programming languages and software packages. For example, the `caret` package in R contains 135 machine-learning models that support regression, with most being nonlinear.

Among these available machine-learning models, many are ensemble-learning models. Nonetheless, the utilization of ensemble learning in forecast combination should not be confused with its earlier usage in Section 3.1.2. In Section 3.1.2, ensemble learning is a forecasting tool, which takes various predictors, such as the lagged versions of the forecast variable or other weather variables, and issues forecasts. Here, ensemble learning is a post-processing tool, and is used to summarize ensemble forecasts produced by whatever models into a single forecast.

Rodríguez-Benítez et al. [119] and Huertas-Tato et al. [120] presented a two-part work on intra-hour and intra-day solar forecasting in the Iberian Peninsula. In part 1 of the work, four models, namely, satellite-based forecasting using cloud motion vector, clear-sky persistence, WRF-Solar, and a hybrid satellite-NWP model called CIADCast, were used to forecast GHI and BNI for various horizons ranging from 15 min to 6 h with steps of 15 min, at four Spanish sites. Subsequently in part 2, these component forecasts were combined via support vector regression (SVR), with either linear or nonlinear kernels. The study considered two forecast combination strategies: the first used horizon-specific SVR models and the second used a single SVR for all horizons. While the differences between these two strategies were marginal, they both improved the root mean square error (RMSE) of the best component models further, up to 17% for GHI and 16% for BNI.

3.2.4. Combining solar forecasts with different prediction mechanisms

Unlike forecasting in a social setting, solar forecasting has more in terms of prediction mechanisms, e.g., physically-based, data-driven, or machine-learning methods. The importance of combining forecasts generated by methods with different prediction mechanisms cannot be overstated. Such a unique advantage must be exploited.

Yang and Dong [121] considered an application of day-ahead solar power forecasting. In an initial setup, only statistical and machine-learning models were combined, and some marginal improvements over the best component model were observed. Notwithstanding, an NWP model was later added to the mix, and subsequently, a significant boost in accuracy was achieved. In the analysis, Yang and Dong [121] noted that the NWP model by itself does not outperform the earlier time series models. However, the error characteristics of the NWP forecasts are distinct from that of the time series models, which leads to increased diversity in the model pool, as shown in Fig. 6. The same time series–NWP combination was later repeated in an irradiance forecasting scenario [73], and the same findings apply.

4. Deterministic-to-probabilistic post-processing

Forecasts are always wrong. The trick is to understand how wrong a forecast can be, and what the probabilities are about it coming out in different ways. Based on this viewpoint, relying on a deterministic best guess is always suboptimal in terms of understanding a forecast. Often times, all one has is a single time series of deterministic forecasts, e.g., NAM forecasts downloaded from the National Centers for Environmental Prediction (NCEP) server. In these situations, quantifying the uncertainty associated with that set of deterministic forecasts becomes a major issue. It is here that D2P post-processing provides some possibilities that address these types of problems.

Thinking tool 6 (*Analog Ensemble*)**.** *Weather patterns often repeat. Given a future weather pattern characterized by a set of forecasts, a forecaster can infer the uncertainty about the forecasts from the performance of similar weather patterns—i.e., analogs—in the past.*

The idea central to D2P post-processing is to leverage the past experience collected about a particular forecasting system, subject to a specific weather condition/pattern. Particularly important is how those past forecasts behave with respect to past observations. Because near things are more alike, it is reasonable to assume that if the current

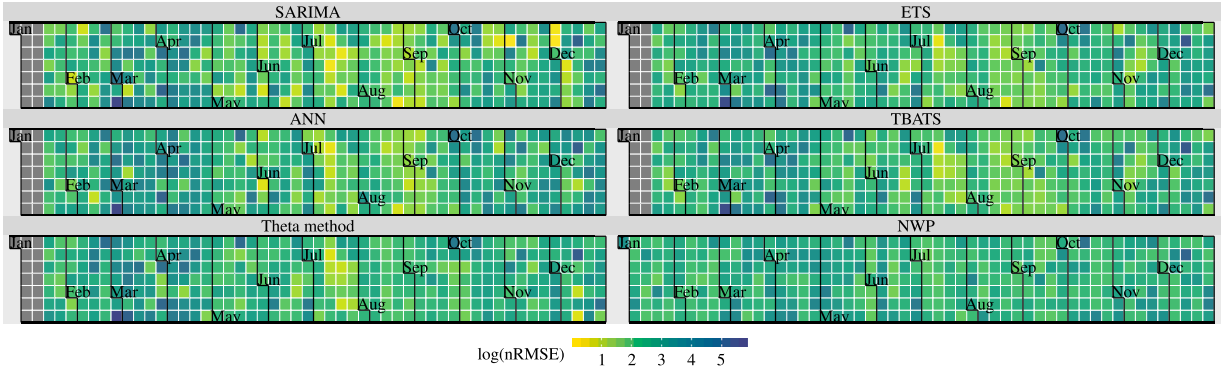


Fig. 6. Daily nRMSEs of hourly day-ahead forecasts using (1) seasonal ARIMA, (2) ETS, (3) ANN, (4) TBATS, (5) the theta method, and (6) NWP, over a period of one year, see [121] for more details on data. Log nRMSE is used in plotting to improve visual contrast. The five time series models show correlated errors—observe the yellow blocks in the summer months—whereas NWP forecast errors are less correlated, and thus introduce diversity.

weather pattern is similar to some past patterns, then whatever materialized in the past is likely to happen again. In weather forecasting, the notion of similarity is defined by weather analogs. Section 4.1 details the procedure for post-processing through analog ensemble (AnEn), as well as some popular extensions of it. The reader should note that the idea of AnEn can be extended to other matching strategies, e.g., matching most recent observations to historical observations, see [122] for a discussion. This review however only focuses on the classical AnEn.

Thinking tool 7 (*Method of Dressing*). *For a given forecasting model, its future error behavior can be inferred from the past performance of the same model. It is therefore reasonable to dress some of the past errors onto the current deterministic forecasts, as a quantification of uncertainty.*

The usual requirement for using AnEn is that the forecasts needs to be issued over a range of horizons following the same dynamical trajectory, e.g., 0–84-h-ahead forecasts issued by NAM. Moreover, dynamical models produce forecasts on a collection of atmospheric variables that are closely interlinked with each other. Hence, having the similarity computed based on multiple variables is no doubt preferred [15]. Nonetheless, if the deterministic-style forecasts are generated using a data-driven or machine-learning model, one usually does not have the luxury to define similarity through analogs. In this case, the method of dressing, which bootstraps the past errors and dresses them on the current deterministic forecast, could be used. This is discussed in Section 4.2. The major difference between AnEn and method of dressing is that the former uses m past observations as an ensemble forecast, whereas the latter dresses m past errors onto the same current forecast.

Thinking tool 8 (*Probabilistic Regression*). *Many regression methods allow uncertainty quantification for nonhomogeneous errors, in particular, generalized additive model (GAM) and quantile regression. By regressing a set of historical observations onto their corresponding historical (multivariate) deterministic forecasts, one can arrive at a predictive distribution or a set of predictive quantiles.*

As seen in the D2D post-processing, one can investigate the causal relationship between predictors and predictand through regression. In other words, the mean of the forecast is modeled as a function of the predictors. An immediate extension to that thinking tool is to model other parameters of a distribution, e.g., the variance or the skewness, using the same set of predictors. The generalized additive models for location, scale and shape (GAMLSS) provides such treatments, and generates parametric predictive distributions. On the other hand, for nonparametric predictive distributions, the well-established quantile regression (QR), and its many variations, such as the quantile regression forest or the quantile regression neural network, are available. However, one should note that not all regressions have well-accepted methods for generating predictive distribution. For instance, computing the standard error for the lasso regression remains challenging [123]. This thinking tool will be discussed in Section 4.3.

4.1. Analog ensemble

The word "analog" was first coined by Lorenz [124] to describe weather patterns that resemble each other. By definition, AnEn is a similarity-search method. In the load forecasting literature, AnEn is known as the similar-day method [125], in which the similarity is usually based on variables that affect the electricity consumption, such as the day of the week, season of the year, or weather patterns. In the machine-learning literature, AnEn is known as the k-nearest neighbor [126], where similarity is usually defined in the feature space. To that end, AnEn, as a post-processing method, has applications far beyond weather forecasting.

4.1.1. AnEn on NWP forecasts

The first major work on AnEn in solar forecasting is thought to be [127], in which the authors presented a case study on solar power forecasting at three PV farms in Italy, using weather forecasts obtained from the Regional Atmospheric Modeling System (RAMS). In fact, most works by this group of authors are related to AnEn; the literature is hence fairly concentrated on these authors. In [127], the advantages and disadvantages of AnEn in contrast to dynamical ensemble were put forth. That is, AnEn has advantages in terms of computation speed, as well as circumventing the need to master the sophisticated NWP models. On this point, AnEn is much favored by forecast practitioners, who have no resource nor knowledge to generate high-quality NWP forecasts on their own. On the other hand, AnEn requires a "frozen" dynamical model, i.e., the NWP model ought to behave consistently in the past and present. However, the frequent configuration updates on operational NWP models could prevent the generation of skillful analogs [127]. Additionally, the requirement on the amount of historical forecast–observation pairs is high. Using too small a training database dramatically reduces the accuracy of AnEn, due to the lack of appropriate analogs in the database.

The AnEn forecast-generating process starts with some N_v-variate forecasts, denoted with

$$\mathcal{X}_t = \begin{pmatrix} x_{t-\tilde{t}}^{(1)} & x_{t-\tilde{t}+1}^{(1)} & \cdots & x_t^{(1)} & \cdots & x_{t+\tilde{t}-1}^{(1)} & x_{t+\tilde{t}}^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ x_{t-\tilde{t}}^{(N_v)} & x_{t-\tilde{t}+1}^{(N_v)} & \cdots & x_t^{(N_v)} & \cdots & x_{t+\tilde{t}-1}^{(N_v)} & x_{t+\tilde{t}}^{(N_v)} \end{pmatrix}^{\top},$$

where t denotes the forecast timestamp, $\tilde{t}$ is the user-defined half-window size, and the superscripts index the weather variable (e.g., irradiance, temperature, or pressure). In the language of computer scientists, $\mathcal{X}_t$ is called the *query*, i.e., the pattern to be matched. Similarly, a

past forecast, $\mathcal{A}_i$, can be expressed using:

$$\mathcal{A}_i = \begin{pmatrix} x_{i-\tilde{t}}^{(1)} & x_{i-\tilde{t}+1}^{(1)} & \cdots & x_i^{(1)} & \cdots & x_{i+\tilde{t}-1}^{(1)} & x_{i+\tilde{t}}^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ x_{i-\tilde{t}}^{(N_v)} & x_{i-\tilde{t}+1}^{(N_v)} & \cdots & x_i^{(N_v)} & \cdots & x_{i+\tilde{t}-1}^{(N_v)} & x_{i+\tilde{t}}^{(N_v)} \end{pmatrix}^{\top}, \tag{12}$$

where i annotates a time in the past, i.e., $i \in \{1, 2, \ldots, t-h-\tilde{t}\}$. Here, h denotes the forecast horizon. By computing similarities between $\mathcal{X}_t$ and all $\mathcal{A}_i$'s, the m candidates with highest similarities are considered to be the analogs of $\mathcal{X}_t$. Subsequently, the m past observations that corresponds to these analogs jointly form an ensemble forecast, see Fig. 7.

The original formulation for analog similarity, d, was given in [127], in the form of a weighted Euclidean distance:

$$d\left(\mathcal{X}_t, \mathcal{A}_i\right) = \sum_{j=1}^{N_v} w_j \sqrt{\sum_{k=-\tilde{t}}^{\tilde{t}} \left(x_{t+k}^{(j)} - x_{i+k}^{(j)}\right)^2}, \tag{13}$$

where the weights w_j can be optimized from the data, e.g., by minimizing the continuous ranked probability score (CRPS). This type of minimization will be more discussed in Section 5.1. In NWP-based AnEn, setting $\tilde{t} = 1$ is usually sufficient [128,129], hence, for each of the N_v variables, three consecutive forecasts are used for matching.

Eq. (13) is used in several other publications of the authors [e.g., 129,130]. However, there are also some inconsistent definitions of similarity among the authors' works. For example, in [131], the two summations in Eq. (13) are swapped,

$$d\left(\mathcal{X}_t, \mathcal{A}_i\right) = \sum_{k=-\tilde{t}}^{\tilde{t}} \sqrt{\sum_{j=1}^{N_v} \omega_j \left(x_{t+k}^{(j)} - x_{i+k}^{(j)}\right)^2}. \tag{14}$$

Due to the square root, which will affect the optimization, weights ω_j in Eq. (14) are unlikely to be the same as w_j in Eq. (13). In one way or another, the inclusion of weights is to differentiate the strength of relationship between various predictors and the forecast variable [129]. As long as the optimization is done properly, the forecast performance is guaranteed to increase, due to increased calibration.

On this point, another alternative formulation was given in [122], where the lag and lead versions of the same forecast variable are simply regarded as distinct variables:

$$d(\mathcal{X}_t, \mathcal{A}_i) = \sqrt{\sum_{j=1}^{J} \omega_j \left(x_t^{(j)} - x_i^{(j)}\right)^2}, \tag{15}$$

where $J = N_v \cdot (2\tilde{t} + 1)$ is the total dimension defining the weather pattern. A distinct advantage of using Eq. (15) over using Eq. (13) or (14) is that it allows ultra-fast computation through *distance profile* and *kd-tree*. These two computation tricks will be discussed shortly after, in Section 4.1.3.

Regardless of which similarity measure is used, the forecaster still needs to decide on the number of analogs to be used. It is known *a priori* that too small an m will lead to an overconfident forecast, whereas too big an m will lead to an underconfident forecast. A table of CRPSs using different m can be found in [128]. But based on experience, setting $m = 20$ often leads to optimal results [127–130].

4.1.2. AnEn on satellite-based forecasts

Since AnEn operates based on similarity search, it can also be applied to satellite-based forecasting. Recall that whenever remote-sensing data are used in solar forecasting, the cloud properties are almost always the key to issuing good forecasts. Watanabe and Nohara [132] employed a random forest regression to map cloud properties (derived from satellite observations) to a set of time series features (e.g., mean, standard deviation, skewness, or kurtosis) that characterize the clearness index (the ratio between irradiance and the corresponding extraterrestrial irradiance). In this case, the weather analogs are defined through these time series features. Subsequently, the current features were matched to historical features, and the top m candidates were selected. Finally, m historical irradiance time series corresponding to the m analogs were used as predictions. However, it is noted that [132] is not a forecasting paper, but the concept therein discussed can be extended to a forecasting context with some straightforward modifications. For good references on time series features and their applications, the reader is referred to the works of Rob Hyndman and colleagues [e.g., 133–135]

The idea of using features as analogs introduces new facets to the AnEn problem. Most notably is the applications in spatio-temporal solar forecasting, where high-dimensional gridded data is involved. To address the curse of dimensionality, feature reduction techniques, such as principal component analysis (PCA) [136] or quadtree [137], can be applied to satellite imagery prior to AnEn [138]. This method is also suitable for gridded NWP forecasts [131]. Notwithstanding, one needs to be aware of the concept of *decorrelation distance* when applying dimension-reduction techniques to gridded data. It is well known that the spatial correlation between two locations in an irradiance random field decreases with separation distance [139,140]. Ignoring the decorrelation distance results in a false sense of similarity, which often deteriorates the performance. Hence, the features should only be sought within the correlation mask, see Fig. 3 in [138] and Fig. 8 in [139]. Another important point to take note is that features need to be computed based on the deseasonalized variables, particularly for irradiance, where the diurnal cycle prevents a forecaster from identifying true weather analogs. There are however some notable exceptions, where the absolute values of the weather variable matter. For instance, for relative humidity, it matters how close it is to 100%.

4.1.3. Computational issues

Despite being more computationally efficient than running NWP with multiple starts, the scalability of AnEn is still an issue when the procedure is performed and repeated on very large datasets, such as in a grid-integration scenario where thousands of PV systems are involved. In this regard, Cervone et al. [141] presented a simulation study that leverages the parallel computing power of the Yellowstone supercomputer at the National Center for Atmospheric Research (NCAR), which has 141,140 cores.

Leveraging powerful hardware is no doubt a viable solution when resource permits. A much preferred approach, however, is to improve the computational efficiency from an algorithmic perspective. AnEn as a similarity-search problem enjoys many properties that allow ultra-fast computation. In the field of time series data mining, motif discovery is a frequently encountered yet tedious task—motifs are sub-sequences of a longer time series that are very similar to each other. Hence, many fast algorithms have been proposed, among which the *distance profile* by Mueen et al. [75] is deemed to be the world's fastest similarity-search algorithm under Euclidean distance. The core concept of distance profile is using fast Fourier transform to convert a convolution problem to a multiplication problem. An application of distance profile on an AnEn problem has been put forward by Yang and Alessandrini [122]. The reader is referred to [122] for mathematical details and computer code.

A limitation of the distance-profile-based AnEn is that the query (current set of forecasts) and data (historical forecasts) have to be arranged as time series. Although possible, it is rather inconvenient for AnEn with multivariate forecasts. To address this issue, Yang [142] presented another option for ultra-fast AnEn, using k-dimensional tree (kd-tree). The kd-tree is a data structure for organizing points in a k-dimensional space by recursive partitioning (splitting). It has been studied extensively by computer scientists since at least the 1970s [143].

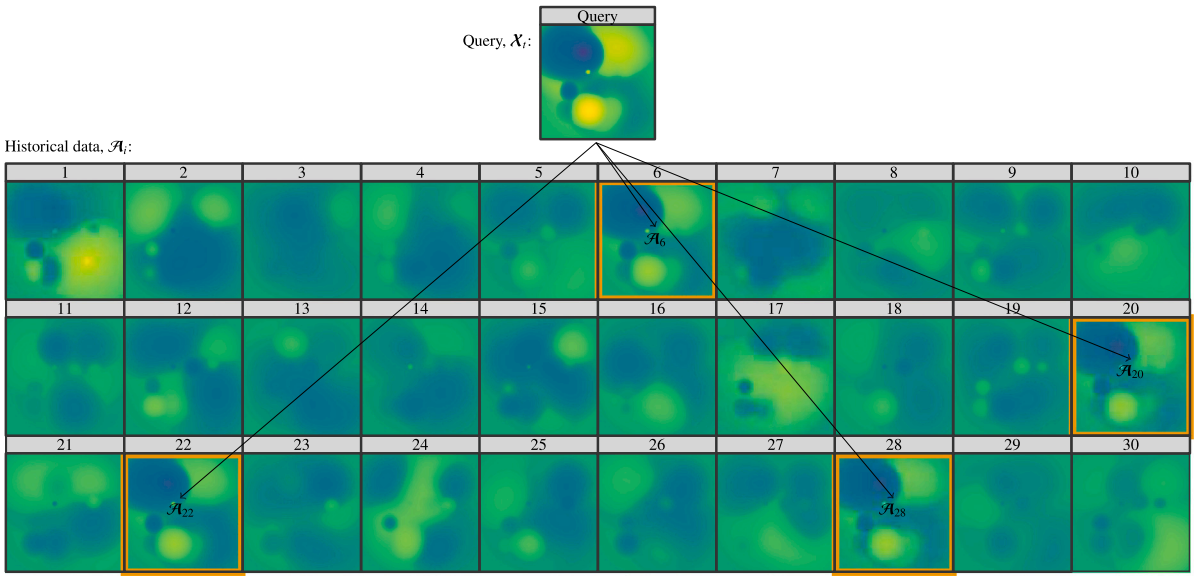


Fig. 7. The current weather pattern $\mathcal{X}_t$—represented by a matrix—is matched to all historical patterns in the database (in this case, $\mathcal{A}_i$, $i = 1, \ldots, 30$). Top m analogs (in this case, $m = 4$) are found. Subsequently, the corresponding historical observations (in this case, y_6, y_{20}, y_{22}, and y_{28}) are grouped and used as an ensemble forecast.

Being a very popular algorithm, there are many implementations of kd-tree. In [142], the Approximate Nearest Neighbor Library [144,145] written in C++ is used. The results of kd-tree-based AnEn is compared to a brute-force search method (as used by NCAR at the moment), and an improvement by three orders of magnitude on computation speed was observed—with the same hardware, a full day's computation can be done within a mere 100 s [142]. It is also worth mentioning that both the distance-profile-based Yang and Alessandrini [122] and kd-tree-based AnEn [142] algorithms have been vetted against a standard test script issued by Stefano Alessandrini, NCAR, and show identical results, i.e., the search is exact.

4.2. Method of dressing

The methods for generating probabilistic forecasts through dressing deterministic forecasts can be classified based on whether or not the predictive distributions are empirical. An empirical predictive distribution is formed by dressing m past errors onto the current forecast—these are known as *dressing errors* [15]. In the simplest case, consider a persistence model on clear-sky index, where the forecast for $t + 1$ is simply the observed value at t. To obtain an empirical predictive distribution, m past persistence forecast errors from $t - 1, \ldots, t - m$ are dressed onto the point forecast, in this case, the observation at t. This method was used by Gneiting et al. [146]. Here, it is important not to mix up this method with persistence ensemble (PeEn), which uses m most recent observations as ensemble members, even though their forecast performance is likely similar.

On the other hand, instead of using an empirical predictive distribution, one can assume a parametric or mixture distribution. The most familiar examples reside in the time series forecasting literature. Many statistical forecasting models, such as some models of ARIMA and ETS families, assume the forecast error has a normal distribution. The mean of the normal predictive distribution is simply the point forecast, whereas the variance depends on the standard error, estimated from in-sample fit. In some cases, the variance of a forecast can be expressed analytically, whereas in other cases, one has to use simulation to generate many possible future sample paths that jointly define the predictive variance.

4.2.1. Empirical approaches

The empirical error dressing is generic, since it applies to all deterministic forecasting models, e.g., David et al. [147] applied the method on an autoregressive model. However, the single most important drawback of this method originates from the fact that forecast errors are often, if not always, nonhomogeneous. Pinson and Kariniotakis [148] noted that it is unlikely that the prediction errors from the recent past would be representative of the current uncertainty. Therefore, the need for conditional sampling of dressing errors is apparent.

Considering the diurnal cycle of solar irradiance, the most straightforward strategy is to sample the dressing errors from the past instances that have the same time-of-day as the forecast timestamp. For example, if the forecast needs to be issued for 14:00, then only errors from previous 14:00 instances would be sampled. Furthermore, as weather evolves continuously, sampling errors from continuous timestamps—atmospheric conditions that follow the same trajectory—is thought advantageous. In other words, dressing errors are sampled in blocks. A solar forecasting case study using block bootstrapping has recently been put forward by Yang [149], using the algorithm proposed in [150].

A prerequisite for conditional sampling is that the conditional variables at the forecast timestamps need to be known in advance. Therefore, one ought to be careful not to use future measurements when defining the forecast conditions. One counter example is to use the actual forecast error as the conditional variable to select past errors, which leads to an incestuous situation. In solar forecasting, there is a list of variables that can be deterministically calculated for any future time, such as the solar zenith angle or clear-sky irradiance. Hence, using these calculatable variables as conditional variables is appropriate. For instance, Grantham et al. [151] binned the past errors according to elevation angle and hour angle—when a forecast becomes available, the dressing errors are only sampled from the corresponding bin. Alternatively, one can define the forecast situation based on the forecasts themselves, e.g., if the forecast clear-sky index is greater than 0.9, errors from the past forecast clear-sky situations are sampled and dressed; otherwise, errors from the past forecast non-clear-sky situations are sampled and dressed.

Worth noting is that there is not a clear-cut boundary separating different forecast conditions, e.g., if one were to separate clear-sky conditions from cloudy ones through the forecast clear-sky index value, a Boolean interpretation might not suffice, considering the chance of misclassification considering thin clouds and/or aerosols. In this regard,

the authors of [148,152] proposed a fuzzy inference model for prediction intervals, in which the wind power forecast condition is divided into three ranges (low, medium and high), to which three trapezoidal fuzzy sets are associated. This idea can readily be transferred to a solar forecasting scenario.

4.2.2. Dressing through an assumed distribution

It is often reasonable to assume that the forecast errors of an *adequate* forecasting model follow a normal distribution. In this case, after the standard deviation of the h-step-ahead forecast (denoted using σ_h) is estimated, the prediction interval is given as: $X \pm c\sigma_h$, where X is the point forecast and c depends on the coverage probability, e.g., $c = 1.96$ for 95% prediction interval. The analytic expression for σ_h using linear-regression-based forecasting models can be found in Chapter 13 of [153]. The analytic expressions of σ_h of ARIMA and ETS families of models can be found in Chapter 5 of [22] and Chapter 6 of [23], respectively. Ideally, the prediction distributions should consider all sources of uncertainty from the model, parameter, and the future innovation. However, due to practical difficulty, σ_h usually takes only the uncertainty in the future innovations and the uncertainty in the estimated parameters into account [23].

Munkhammar et al. [83] developed a model called Markov-chain mixture (MCM) distribution model, which dresses a mixture of uniform distributions onto the point forecast. MCM uses some past observations to construct a transition probability matrix $\boldsymbol{M}$. To achieve this, the range (maximum minus minimum) of training samples are first divided into m equal-size bins, and a number indicating the bin index is assigned to each observation, depending on which bin it falls into. The $[i,j]$th entry of $\boldsymbol{M}$ denotes the probability of transition from bin i to bin j. Subsequently, when a new observation (say inside the ith bin) becomes available, the corresponding forecast follows a piecewise uniform predictive distribution characterized by the probabilities in the ith row of $\boldsymbol{M}$.

To exemplify the procedure, consider a three-bin scenario: the clear-sky index ranging from 0 to 1.2 are divided into $[0, 0.4)$, $[0.4, 0.8)$, and $[0.8, 1.2)$ bins. Suppose the transition probability matrix trained from historical data is:

$$\boldsymbol{M} = \begin{pmatrix} 0.1 & 0.7 & 0.2 \\ 0.5 & 0.2 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}, \tag{16}$$

then for a current clear-sky index observation of 0.5—inside the second bin—the predictive PDF of the corresponding forecast is:

$$f = \begin{cases} 0.5/0.4, & \text{for } 0 \le X < 0.4; \\ 0.2/0.4, & \text{for } 0.4 \le X < 0.8; \\ 0.3/0.4, & \text{for } 0.8 \le X < 1.2, \end{cases} \tag{17}$$

where the denominator 0.4 is the bin width in this example. Clearly, MCM is calibrated by construct, but generally lacks sharpness.

4.3. Probabilistic regression

GAMLSS are univariate distributional regression models, where all the parameters of the assumed distribution for the response can be modeled as additive functions of the explanatory variables. QR and its variants are extensions of linear regression, and are used when the assumptions, such as linearity, homoscedasticity, or normality, are not met. It is worth noting that both GAMLSS and QR can generate forecasts on their own, but in this paper, it is their post-processing capability that is discussed. Stated differently, in the current context, the regressors are the raw forecasts, whereas the regressand is the post-processed forecast.

4.3.1. Generalized additive models for location, scale and shape

To understand GAMLSS, consider a multiple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i^{(1)} + \cdots + \beta_m x_i^{(m)} + \varepsilon_i, \tag{18}$$

where y_i is the ith observation, $x_i^{(j)}$ with $j = 1, \ldots, m$ is the jth regressor (e.g., forecasts of a weather variable), and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is a zero-mean homoscedastic error term. The specification of Eq. (18) is equivalent to:

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma^2\right), \tag{19}$$

$$\mu_i = \beta_0 + \beta_1 x_i^{(1)} + \cdots + \beta_m x_i^{(m)}, \tag{20}$$

or in matrix notation:

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{\mu}, \sigma^2\right), \tag{21}$$

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \tag{22}$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(m)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)} \end{pmatrix} \tag{23}$$

is an $n \times p$ design matrix with $p = m + 1$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, and $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$.

To extend the linear regression model to the generalized linear model (GLM), two modifications are in place. Firstly, the normal distribution of $\boldsymbol{y}$ is replaced by the exponential family of distributions, see [153] for details of exponential family. Examples of exponential family distributions are gamma and inverse Gaussian distributions. Secondly, the relationship between $\boldsymbol{\mu}$ and $\boldsymbol{X}\boldsymbol{\beta}$ is modeled through a *link function*, $g(\cdot)$. An example of $g(\cdot)$ is the logarithm function. Mathematically, denoting the parameters of an exponential family distribution by $\boldsymbol{\phi}$,

$$\boldsymbol{y} \sim \text{ExpFamily}\left(\boldsymbol{\mu}, \boldsymbol{\phi}\right), \tag{24}$$

$$g(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta}. \tag{25}$$

These two modifications allow more flexible modeling of the causal relationship between regressors and regressand.

The GLM can be further extended to GAM by allowing smoothing for the continuous variables among the regressors. Suppose there are J continuous variables in $\boldsymbol{X}$, then

$$\boldsymbol{y} \sim \text{ExpFamily}\left(\boldsymbol{\mu}, \boldsymbol{\phi}\right), \tag{26}$$

$$g(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + s_1\left(\boldsymbol{x}^{(1)}\right) + \cdots + s_J\left(\boldsymbol{x}^{(J)}\right), \tag{27}$$

where $s_1, \ldots, s_J$ are variable-specific smoothing functions, and $\boldsymbol{x}^{(j)}$ is the length-n vector of the jth continuous variable, i.e.,

$$\boldsymbol{x}^{(j)} = \begin{pmatrix} x_1^{(j)} & x_2^{(j)} & \cdots & x_n^{(j)} \end{pmatrix}^\top. \tag{28}$$

Last but not least, GAMLSS extends GAM by adding in analogous modeling of other distribution parameters. For example, for a two-parameter distribution $D(\boldsymbol{\mu}, \boldsymbol{\sigma})$, GAMLSS is in the form:

$$\boldsymbol{y} \sim D\left(\boldsymbol{\mu}, \boldsymbol{\sigma}\right), \tag{29}$$

$$g_1(\boldsymbol{\mu}) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + s_{11}\left(\boldsymbol{x}^{(11)}\right) + \cdots + s_{1J_1}\left(\boldsymbol{x}^{(1J_1)}\right), \tag{30}$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{X}_2\boldsymbol{\beta}_2 + s_{21}\left(\boldsymbol{x}^{(21)}\right) + \cdots + s_{2J_2}\left(\boldsymbol{x}^{(2J_2)}\right). \tag{31}$$

In this way, one can see how the GAMLSS framework is general, in terms of modeling parametric distribution. A large class of distributions can be characterized by location, scale and shape, and hence the name, GAMLSS. The book by Stasinopoulos et al. [154] provides a complete guide on GAMLSS, with its implementation in R.

At the moment, GAMLSS is not a popular method in the field of solar forecasting, as evidenced by its low usage rate. Nonetheless, it has

been used as a stand-alone forecasting model in [155], and as a post-processing method in Bakker et al. [156]. More specifically, Bakker et al. [156] used GAMLSS with gamma and truncated-normal distributions to post-process a large collection of output variables from the high-resolution non-hydrostatic NWP model HARMONIE-AROME and the atmospheric composition model CAMS. In terms of its deterministic performance, the two GAMLSS models performed similarly to the other nonparametric post-processing methods. However, in terms of probabilistic performance, it appeared that the nonparametric techniques (which will be discussed in Section 4.3.2) are more advantageous.

4.3.2. Quantile regression and variants

The motivation for developing QR aligns with the motivation for developing GAMLSS. Recalling that for linear regression, the coefficients $\boldsymbol{\beta}$ are obtained through least squares. For example, for ordinary least squares, the optimal choice of $\boldsymbol{\beta}$ minimizes the sum of squared errors, i.e.,

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2, \tag{32}$$

where

$$\boldsymbol{x}_i = \begin{pmatrix} 1 & x_i^{(1)} & \cdots & x_i^{(m)} \end{pmatrix}^\top, \tag{33}$$

not to be confused with $\boldsymbol{x}^{(j)}$ defined in the previous section. For quantile regression, instead of minimizing the squared loss, the pinball loss is minimized:

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i, \boldsymbol{x}_i^\top \boldsymbol{\beta}\right), \tag{34}$$

where

$$\rho_\tau(y_i, \boldsymbol{x}_i^\top \boldsymbol{\beta}) = \begin{cases} \tau |y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}| & \text{for } \boldsymbol{x}_i^\top \boldsymbol{\beta} \le y_i, \\ (1-\tau) |y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}| & \text{for } \boldsymbol{x}_i^\top \boldsymbol{\beta} \ge y_i, \end{cases} \tag{35}$$

is the pinball loss for the ith sample, *cf.* Eq. (8). Obviously, the parameter $\boldsymbol{\beta}$ differs for different τ values. Hence, if Q quantiles are needed, Q regressions need to be performed. The book by Koenker [157] provides a complete account for the basic theory of quantile regression.

Being a major powerhouse of nonparametric regression, the literature on QR is bulky. Naturally, the variations of QR are abundant. For example, analogous to the well-known lasso estimator:

$$\hat{\boldsymbol{\beta}}_\lambda = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1, \tag{36}$$

which shrinks the unconstrained least-squares estimator towards $\boldsymbol{\beta}_0$ (typically taken as zero), the QR-lasso estimator is given by:

$$\hat{\boldsymbol{\beta}}_{\tau,\lambda} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i, \boldsymbol{x}_i^\top \boldsymbol{\beta}\right) + \lambda \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1. \tag{37}$$

In the machine-learning domain, quantile regression neural network (QRNN) and quantile regression forest (QRF) are also available. Both QRNN and QRF are ingenious, yet very straightforward, extensions of QR. Anyone who is familiar with QR, neural network, and random forest could master both extensions in a few hours. These have thus been widely adopted as stand-alone probabilistic solar forecasting methods [147,158]. It should be noted that these methods can also be used to post-process ensemble forecasts, as discussed in Section 5.1.4.

That said, QR and its variants are less utilized as a post-processing tool for D2P conversion. The only known case in the solar forecasting literature is [156], who applied both QR and QRNN to a set of multivariate deterministic NWP forecasts. It is noted that the emphasis for D2P conversion using QR is that the design matrix $\boldsymbol{X}$ consists of different weather variables of the same set of forecasts. If the design matrix consists of different component forecasts, i.e., ensemble forecasts, QR then becomes a P2P post-processing tool. This will be more discussed in Section 5.

5. Probabilistic-to-probabilistic post-processing

As discussed in the introduction, a probabilistic forecast can be a distribution, a set of quantiles, a prediction interval, or simply an ensemble of (deterministic or probabilistic) component forecasts. This section discusses various post-processing procedures to improve the quality of probabilistic forecasts. In contrary to the thinking tools discussed in the P2D section, the post-processing thinking tools here are P2P. In other words, the final forecasts are still probabilistic, but need not have the same form as the initial probabilistic forecasts (e.g., ensemble forecasts can be post-processed to a parametric distribution).

The general strategy of producing good probabilistic forecasts is to maximize the sharpness of the predictive distributions, subject to calibration [146,159]. A forecast is said to be *probabilistically calibrated* when $Y \sim F$, which is to say that the observation is distributed according to the predictive CDF. This can be verified through the probability integral transform (PIT),

$$p_t = F_t(y_t^*), \tag{38}$$

which is the value of the tth predictive distribution (denoted with F_t) evaluated at the tth materialized observation (denoted with y_t^*) from the verification sample [17,160]. If the predictive distributions are calibrated, then the PIT, as a random variable, is uniformly distributed. However, one should note that a uniform PIT histogram (the histogram of p_t, for $t = n+1, \ldots, n'$, indexing the verification samples) is a necessary but not sufficient condition for calibration [146]. A more relaxed definition of calibration is simply that the claimed coverage probability of a PI is consistent with the nominal coverage in the long run [161]. On the other hand, sharpness refers to the concentration of a predictive distribution, or the width of a PI.

Thinking tool 9 (*Calibrating and Sharpening Ensemble Forecasts*)**.** *Ensemble forecasts are often not calibrated and/or not sharp. P2P post-processing through ensemble model output statistics and ensemble dressing could improve calibration and sharpness.*

When the initial probabilistic forecast is in the form of ensemble, one can convert it to an empirical predictive distribution using one of the three approaches described in [18]. Various quantiles and PIs can then be retrieved from the converted distribution. Nonetheless, a common pitfall of ensemble forecast (e.g. almost always for NWP) is a lack of diversity, which often leads to overconfident forecasts. In other cases, the ensemble forecasts may be too dispersed, which results in underconfident forecasts. In either case, it is quite necessary to *calibrate* and *sharpen* such incorrectly dispersed forecasts. One class of methods is known as ensemble model output statistics (EMOS), or more generally, GAMLSS, which converts ensemble forecasts to a calibrated parametric predictive distribution. Another class of methods seek to dress additional members or kernels onto ensemble forecasts. Lastly, it is possible to convert the initial ensemble in to an arbitrary set of predictive quantiles, using quantile regression and variants. All three classes of methods are discussed in Section 5.1.

Thinking tool 10 (*Combining Probabilistic Forecasts*)**.** *Similar to the case of combining deterministic forecasts, when multiple probabilistic forecasts are present, one can combine them to a final probabilistic forecast.*

EMOS and ensemble dressing both calibrate a single probabilistic forecast. Notwithstanding, calibration can also be done through forecast combination. Similar to the P2D case, where deterministic component forecasts are combined to a single deterministic forecast, P2P forecast combination aggregates different probabilistic forecasts to a final probabilistic forecast. Section 5.2 reviews methods for combining predictive distributions, quantiles, and prediction intervals.

5.1. Calibrating ensemble forecasts

Recall in the P2D section, combining forecasts is elaborated. Whereas an ensemble of deterministic forecasts naturally carries the notion of uncertainty, it is needs to be calibrated. Since calibration does not place any prerequisite on the form of the final calibrated forecasts, they can be (1) parametric predictive distribution (Sections 5.1.1 and 5.1.2), (2) ensemble forecasts with more members (Section 5.1.3), or (3) quantiles (Section 5.1.4). Subsequently, if a forecaster is interested in calibrated PIs, they can be extracted from any of the aforementioned representations of probabilistic forecasts.

5.1.1. Ensemble model output statistics

EMOS was proposed by Gneiting et al. [162] as a forecast post-processing tool. It takes ensemble forecasts as input, and outputs a parametric predictive distribution. In the original work [162], EMOS was demonstrated for 48-h forecasts of sea level pressure and surface temperature over the North American Pacific Northwest. Given a poor man's ensemble with m members at time t, denoted by $x_{t1}, \dots, x_{tm}$, EMOS assumes that these members are independent realizations from a normal predictive distribution,

$$y_t \sim \mathcal{N}\left(\hat{w}_0 + \hat{w}_1 x_{t1} + \cdots + \hat{w}_m x_{tm},\ \hat{\beta}_0 + \hat{\beta}_1 S_t^2\right), \tag{39}$$

i.e., its mean and variance are given by:

$$\mathbb{E}(y_t) = \hat{w}_0 + \hat{w}_1 x_{t1} + \cdots + \hat{w}_m x_{tm}, \tag{40}$$

$$\mathbb{V}(y_t) = \hat{\beta}_0 + \hat{\beta}_1 S_t^2, \tag{41}$$

where S_t^2 is the variance of the raw ensemble, and parameter $\hat{\theta} = \{\hat{w}_0, \hat{w}_1, \dots, \hat{w}_m, \hat{\beta}_0, \hat{\beta}_1\}$ is the *estimate* of unknown model parameter θ. Two approaches can be used to estimate θ, namely, maximum likelihood estimation and CRPS minimization. It is also noted that maximizing the log-likelihood is equivalent to minimizing the ignorance score (IGN), also known as the logarithm score.

Minimizing IGN or minimizing CRPS is an optimization problem, in that, an objective function is needed. Gneiting et al. [162] showed that the IGN and CRPS for n fitting samples, namely, $(y_i, x_{i1}, \dots, x_{im})$, $i = 1, \dots, n$, can be written as:

$$\text{IGN} = \frac{1}{2n} \sum_{i=1}^{n} \left[\ln(2\pi) + \ln\left(cS_i^2\right) + Z_i^2\right], \tag{42}$$

and

$$\text{CRPS} = \frac{1}{n} \sum_{i=1}^{n} \left(cS_i^2\right)^{\frac{1}{2}} \left\{ Z_i \left[2\Phi(Z_i) - 1\right] + 2\varphi(Z_i) - \frac{1}{\sqrt{\pi}} \right\}, \tag{43}$$

respectively, where Φ and φ are CDF and PDF of a standard normal distribution, and

$$Z_i = \frac{y_i - \left(w_0 + w_1 x_{i1} + \cdots + w_m x_{im}\right)}{\left(\beta_0 + \beta_1 S_i^2\right)^{\frac{1}{2}}} \tag{44}$$

is the ith standardized EMOS model error. Since both IGN and CRPS are strictly proper, a forecaster therefore needs to tailor the choice to the problem at hand.

Based on Eqs. (40) and (41), it is immediately clear that EMOS is a nonhomogeneous Gaussian regression (NGR), in which the variance of the predictand (i.e., the final post-processed forecast) is a function of the ensemble variance. When S_t^2 varies, for example, according to the clear and cloudy situations, the variance of the final predictive distribution follows such variation. However, it should be highlighted that the *mixing weights*, $w_1, \dots, w_m$, are fixed, suggesting that the contribution from each ensemble member is also fixed. Consequently, for sensible weight assignment, the error characteristics of the component models need to be consistent. Stated differently, if the error characteristics are inconsistent, it is not possible to assign meaningful weights. For instance, forecasts generated by the previously mentioned AnEn cannot be fed to Eq. (40), because the observations corresponding to the m best analogs generally do not possess any consistency. Nonetheless, if the spread of an AnEn forecast is underdispersed, the variance adjustment of EMOS could be beneficial.

In solar engineering, an EMOS application can be found in [163], in which the authors post-processed the ECMWF ensemble forecasts. However, as mentioned in Section 1.3, ECMWF dynamical ensemble forecasts are obtained by perturbing initial conditions. Since dynamical ensemble forecasts are equally likely, it is unreasonable to assume these forecasts to be *identifiable*. Here, identifiability means jth member forecasts in the ensembles consistently perform better or worse than its peers, which is a prerequisite for unequal weights assignment during EMOS. In this regard, the EMOS application in [163] appears to be ill-motivated. To correctly utilize EMOS, a poor man's ensemble is required, where the ensemble members come from different, identifiable sources, as demonstrated in [162,164–166]. It is noted that some references listed here are not forecasting applications, instead, [164,165] perform site adaptation of gridded irradiance and [166,167] perform solar modeling—these topics also belong to solar energy meteorology. Nonetheless, the generality of EMOS can be assumed at once; the data and code provided in [164,166] can be readily used for any solar forecasting exercise. Additionally, theoretical and empirical analyses on various statistical properties of EMOS are made available in [164], facilitating model diagnosis and verification for future applications.

5.1.2. Extensions of EMOS

Since EMOS is an NGR, it is immediately obvious that there are many variations and extensions to it. The first and foremost variation resides on the choice of predictive distribution. It has long been known to solar engineers that the distribution of clear-sky index is bimodal, in which the two peaks correspond to the clear and cloudy states of the atmosphere [136,168]. To that end, whether or not it is sufficient to use a unimodal predictive distribution is generally unknown. Moreover, since irradiance and clear-sky index are both non-negative quantities, it might be more appropriate to use a lower-tail-truncated predictive distribution to reflect the probability cut-off at zero. In other cases, the upper-tail-truncated predictive distribution may be appropriate to mark the extraterrestrial upper bound of solar irradiance or clear-sky index. On this point, the analytic CRPS expression of n training samples following a lower-tail truncated normal predictive distribution was first derived by Gneiting et al. [169], which can be used to replace Eq. (43) if needed. More general expressions for upper-tail-truncated normal density and two-sided truncated normal density can also be derived in a tedious but straightforward manner. In contrary, the current authors are not aware of any analytic expression of CRPS of samples that follow bimodal distributions.

Speaking of the choice of predictive distribution, the GAMLSS introduced in Section 4.3.1 is a more general form of EMOS. Instead of modeling the predictive variance using the ensemble variance, GAMLSS models the location, scale, and shape parameters of a predictive distribution using GAM. In contrast to its application in D2P post-processing, the inputs of GAMLSS in P2P post-processing are ensemble forecasts. Additionally, the parameter estimation of GAMLSS is more versatile, and does not rely on analytic expression of CRPS, which could facilitate implementation [154]. Yagli et al. [170] applied GAMLSS on ensemble solar forecasts generated using 20 data-driven models. Two distributions, namely, the truncated logistic distribution and the truncated skewed Student's t distribution, were considered; both were shown to produce more calibrated forecasts than that produced by the truncated normal distribution [170].

EMOS is also capable of performing variance correction on forecasts generated by AnEn. As mentioned earlier, when AnEn is used, Eq. (40) becomes inappropriate. In this case, there are two strategies to formulate the mean of the predictive distribution. First, one can use the observation that corresponds to the best analog as the mean of the predictive distribution. Second, one can use the mean of the m

observations corresponding to m best analogs as the mean of the predictive distribution. In either case, minimizing CRPS is mainly targeted to correct the dispersion of the AnEn forecasts. In this regard, the weights in analog similarity—recall Eqs. (13)–(15)—are no longer necessary, since the calibration is now performed through EMOS. Some authors refer to such methods for variance correction as *variance deficit* [171], which are primarily used to deal with overconfident ensemble forecasts.

5.1.3. Calibration via dressing

Dressing ensemble forecasts was first discussed by Roulston and Smith [15], who raised the concern regarding "double counting" the uncertainty using conventional dressing techniques. One example of double counting could be due to the fact that the uncertainty in the initial condition has been explicitly accounted for by the dynamical ensemble, and when each ensemble member is dressed with its own statistical errors, the final ensemble forecasts become too loose. To mitigate such undesirable effect, the best-member dressing method was proposed [15].

Suppose there are m ensemble members initially, denoted using x_{ij}, where $i = 1, \dots, n$, and $j = 1, \dots, m$. Given the corresponding observations y_i, a total of $n \times m$ errors are available:

$$\boldsymbol{E} = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix}, \tag{45}$$

where $e_{ij} = x_{ij} - y_i$. The traditional dressing through error statistics samples from the jth column of $\boldsymbol{E}$ and dresses them onto the jth ensemble member. On the other hand, the best-member dressing method first determines the row-wise smallest errors, e.g., for row i,

$$\epsilon_i = e_{ij^*}, \tag{46}$$

$$j^* = \underset{j}{\operatorname{argmin}} |e_{ij}|. \tag{47}$$

Subsequently, for each ensemble member, the dressing errors are all drawn from the set $\{\epsilon_1, \dots, \epsilon_n\}$. It is noted that when dealing with multivariate forecasts concurrently, i.e., forecasting for two or more variables, one needs to be aware of the pitfall of selecting the false best member, see [15] for more details.

Despite that the best-member dressing method could sharpen the post-processed forecasts, as compared to that using traditional statistical dressing, Wang and Bishop [172] noted that the method does not lead to reliable ensemble—a reliable ensemble should appear to be drawn from the same distribution as the verifying observations given the ensemble. Wang and Bishop [172] proposed a new dressing kernel that constrains the variance of the dressing perturbations, such that it is equal to the difference between the variance of the error of the underlying ensemble mean and the ensemble variance. An excellent toy example was provided in [172], graphically, making the technique immediately obvious to the readers.

The drawbacks of the improved best-member dressing method in [172] was further argued by Fortin et al. [173]. They noted through a synthetic example and showed that rescaling the error patterns can lead to ensemble forecasts with heavier tails than that of the observations. Secondly, the improved best-member dressing method in [172] only applies to underdispersive ensemble forecasts. Consequently, Fortin et al. [173] proposed a weighted-member dressing method, in which the ensemble members are dressed using samples drawn from different error distributions. More specifically, instead of drawing samples from an unconditional error pool, the new method draws sample from a conditional pool, e.g., for a member that ranked jth in the current ensemble, its dressing errors are drawn from the errors of jth-rank members observed in the database of past forecasts.

Another way to improve the calibration of dynamical ensemble forecasts via dressing is Bayesian model averaging (BMA). Introduced to the field of forecasting by Raftery et al. [161], BMA dresses each ensemble member with a PDF, and linearly combines those dressed PDFs with the weights representing the posterior model probability, which reflects the skill of each model on the training data. Fig. 8 provides an illustration of the BMA predictive distribution. Mathematically, for the ith training sample, BMA is simply:

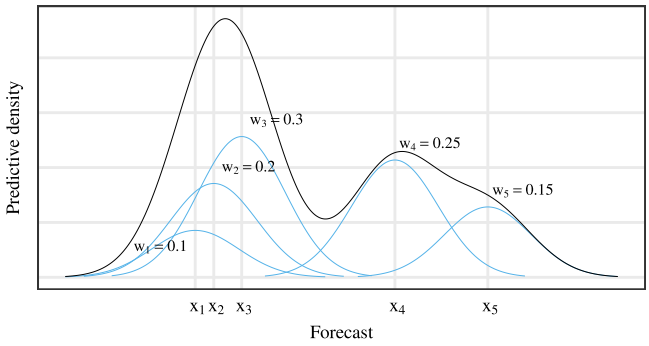


Fig. 8. An toy example of a BMA predictive distribution constructed from a 5-member bias-corrected ensemble. The weights for the five (dressed) Gaussian components are as indicated. The BMA predictive distribution is then obtained via the weighted sum. In this case, it shows bimodality.

$$g_i(y) = \sum_{j=1}^{m} w_j f_{ij}(y|x_{ij}), \tag{48}$$

where $f_{ij}(y|x_{ij})$ is the PDF dressed for the jth component forecast, $g_i(y)$ is the PDF of the combined forecast, and y is a generic variable notating the argument of the density functions. In [161], a normal PDF is used, namely,

$$Y|x_{ij} \sim \mathcal{N}(a_j + b_j x_{ij}, \sigma^2), \tag{49}$$

which is a common choice for surface temperature and sea level pressure. Parameters a_j and b_j are estimated by regressing y_i on x_{ij}, to correct the bias in raw ensemble forecasts. Subsequently, an expectation–maximization (EM) algorithm is used to estimate σ^2 and w_j. Raftery et al. [161] showed through numerous examples that the predictive densities of surface temperature were both better calibrated and sharper. As suggested by Eq. (48), BMA linearly combines PDFs, such linear combination of predictive distributions will be more discussed in Section 5.2.2.

In the field of solar forecasting, Aryaputera et al. [174] compared the choices of using the normal and skew-normal PDFs to dress the ensemble members, in forecasting the accumulated irradiance. Additionally BMA was compared to EMOS. Although intended for dynamical ensembles, Aryaputera et al. [174] applied BMA to a poor man's ensemble of dynamical ensembles (ECMWF, JMA and KMA). The authors showed that both EMOS and BMA improved calibration, sharpness, and CRPS. Furthermore, using the skew-normal PDF instead of the normal PDF improved the calibration further, because the skew-normal PDF fitted the data better. Doubleday et al. [175] later applied BMA in a solar power forecasting context. In their case-study, the raw NWP ensemble members were first preprocessed to forecast solar power at utility scale. To accommodate potential inverter clipping in their probabilistic forecasts, the authors proposed to use a mixture of a discrete and a continuous part, where the discrete probability of inverter clipping is estimated using logistic regression. For the continuous part—*cf.* Eq. (48)—the authors considered a beta distribution and compared it to a two-sided truncated normal distribution. Moreover, the authors compared their approach to EMOS and noted superior performance, while the beta distribution performed better than the truncated normal distribution. Overall, skill scores between 2%–36% were achieved.

5.1.4. Calibration via quantile regression

As mentioned in Section 4.3.2, QR is an approach to find conditional quantiles of the distribution of a random variable. In addition to being used as a forecasting technique, QR can be used as a P2P post-processing technique, with the aim of improving the calibration of NWP ensemble members that tend to be underdispersive [161]. To the best of the authors' knowledge, the first attempt to calibrate NWP ensemble members was performed by Bremnes [176], who used local QR to calibrate precipitation forecasts and showed significant improvement over the ECMWF Ensemble Prediction System (EPS). In the field of solar forecasting, Ben Bouallègue [177] used QR and QR-lasso to calibrate the underdispersive forecasts from the COSMO-DE-EPS ensemble system by developing a QR model for each quantile. For each QR model, the inputs include (1) an initial τth quantile retrieved from the sorted ensemble members (akin Fig. 2), (2) the square of that initial τth quantile, (3) the radiation at the top of the atmosphere, (4) a collection of 49 pre-selected weather variables, and (5) the product of the initial τth quantile with each of the 49 weather variables, making a total of 101 predictors. While both QR and QR-lasso produced calibrated forecasts, the QR-lasso outperformed the QR model in terms of the pinball loss as it benefited from the regularization of the high number of input variables.

It is important to note that the definition in Eq. (34) for linear QR can be extended to a nonlinear case:

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i, g(\boldsymbol{x}_i)\right), \tag{50}$$

where g represents some nonlinear function, $\boldsymbol{x}_i$ contains the forecasts of ensemble members of an NWP model, and ρ_τ represents the pinball loss function. As mentioned above, common options for g include ANN (i.e., the approach of QRNN) and gradient boosted regression trees (GBRTs). Whereas GBRTs support using pinball loss as the cost function, algorithmically, see `scikit-learn` documentation, QRNN requires modification of the pinball loss as it is not differentiable everywhere, which hinders the use of gradient-based training algorithms. To address this issue, Cannon [178] suggested to approximate the pinball loss function with the Huber function [179,180], which is a hybrid L_1/L_2-norm. The Huber function, $h(u)$, applies the L_2-norm to absolute values below a threshold value ϵ and the L_1-norm to absolute values above ϵ:

$$h(u) = \begin{cases} u^2/(2\epsilon), & \text{if } 0 \le |u| \le \epsilon, \\ |u| - \epsilon/2, & \text{if } |u| > \epsilon. \end{cases} \tag{51}$$

Here, ϵ is a small positive real number. In the default QRNN training routine, ϵ is set to 2^{-8} initially, and gradually reduces to 2^{-32} over subsequent iterations [178]. Massidda and Marrocu [181] used GBRTs with the pinball loss function to calibrate the ECMWF EPS forecasts, and compared the method to variance deficit calibration of the ensemble. The authors proposed to generate 51 clear-sky index forecasts for PV power by applying a deterministic model that depends on the azimuth and zenith, as well as the 51 EPS members, see Eq. (15) of [181]. Then, the 51 clear-sky index forecasts for PV power were divided into 9 quantiles, and fed to 51 GBRT models to produce a probabilistic forecast with the same number of quantile levels as the EPS. The rank histograms showed that calibration through variance deficit improved the underdispersed forecasts of the EPS but did not calibrate them entirely. Instead, linear QR and QR with GBRT produced calibrated probabilistic forecasts while also improving CRPS.

QRF is a technique based on random forests (RF) that produces conditional quantiles instead of the conditional mean [182]. Owing to the prediction mechanism of RF, QRF does not directly minimize any loss function. Recall that an RF comprises many trees, and the forecast of the RF, given a new data point $\boldsymbol{x}_t$, is obtained by weighted averaging over all observations. Denoting the weight for the ith observation given $\boldsymbol{x}_t$ by $w_i(\boldsymbol{x}_t)$—the notation suggests that the weight changes with different $\boldsymbol{x}_t$ values—the forecast is simply:

$$y_t = \sum_{i=1}^{n} w_i(\boldsymbol{x}_t) \cdot y_i. \tag{52}$$

In the above equation, i.e., the forecast from an RF model, y_t is a deterministic forecast. QRF modifies Eq. (52),

$$y_t \sim F(y|\boldsymbol{x}_t) = \sum_{i=1}^{n} w_i(\boldsymbol{x}_t) \cdot \mathbf{1}_{y_i<y}, \tag{53}$$

where $\mathbf{1}_{y_i<y}$ is an indicator function, and here, y_t is a predictive distribution. To compute y_t, one needs to evaluate Eq. (53) for all $y \in \mathbb{R}$. In other words, the prediction of QRF is simply the weighted distribution of the observations, where the weights are identical to that of the RF [182]. Taillardat et al. [183] used QRF to calibrate the 35-member ensemble forecasts of surface temperature and wind speed of Météo-France and compared their results to the raw ensemble and EMOS calibrated ensemble. The authors showed that both EMOS and QRF successfully calibrated the ensemble forecasts but that QRF outperformed EMOS in terms of CRPS, since the former included additional output variables from the ensemble forecasts.

In two relevant studies, Bremnes [184,185] proposed to improve the calibration of wind speed forecasts of the ECMWF ensemble members using constrained QR splines (CQRS), and quantile function regression based on neural networks and polynomials (QPR). In the case of CQRS, the core idea is to use the B-spline of order k as $g(\boldsymbol{x}_i)$ and minimize Eq. (50) to find the B-spline coefficients [184]. Using splines has several advantages, such as that the spline function is bounded by its coefficients or that the spline function is monotonically increasing when the coefficients are monotonically increasing [184]. In order to prevent quantile crossing, the author suggested to let each quantile model depend on one covariate and to define the B-splines identically across the quantiles. Then, only the spline coefficients $\alpha_j, j = 1, \ldots, J$ are different among the quantiles and it suffices to add constraints that ensure ordered coefficients when estimating these or by simply reordering the coefficients afterwards. Bremnes [184] showed that the proposed method significantly outperformed the raw ensemble while modestly improving over the lognormal EMOS in terms of deviation from reliability and the pinball loss.

One issue with linear QR and CQRS is that of crossing quantiles as mentioned above. While Bremnes [184] suggested additional constraints to prevent crossing quantiles, the computational complexity introduced numerical issues. In order to prevent quantile crossing altogether, Bremnes [185] introduced QPR where all quantiles are modeled simultaneously by replacing $g(\boldsymbol{x}_i)$ in Eq. (50) with Bernstein basis polynomials of degree d and estimating the coefficients simultaneously via a neural network. The reason for using Bernstein basis polynomials is that if the coefficients are in non-decreasing order the resulting function will also be non-decreasing [185]. The coefficients are generated using a neural network that takes as input the ensemble forecasts and site specific information and outputs all coefficients simultaneously. The result is a non-decreasing set of forecasts that were compared against probabilistic forecasts from CQRS and monotonic composite quantile regression neural networks (MCQRNN). The author noted that while the performance improvement over CQRS in particular is relatively small (approximately 1% in terms of the pinball score), the main advantage of the proposed technique is that quantile crossing is prevented and that in contrast to CQRS, different ensemble variables can be included.

5.2. Combining probabilistic forecasts

Probabilistic forecasts can be issued in various forms, namely as predictive distributions, quantile forecasts, or prediction intervals. This section presents an overview of the heuristic, linear, and nonlinear combination methods that can be applied to the aforementioned forms of probabilistic forecasts.

5.2.1. Model-free heuristics

Heuristics for predictive distributions. Similar to the case of combining deterministic forecasts, heuristics can be used to combine probabilistic forecasts without training. Stone [186] formally introduced the "opinion pool" for predictive distributions $F_1, \dots, F_m$, which can be linearly combined as $G = \sum_{j=1}^{m} w_j F_j$ where $w_j \geq 0$ and $\sum_{j=1}^{m} w_j = 1$. Stone [186] suggested to set $w_1, \dots, w_m = 1/m$ to allow for a "democratic" prediction system where all experts are equally valued.

At forecast timestamp t, denoting the component predictive distribution with $F_{tj}(y)$, then the linear opinion pool with equal weights is:

$$G_t(y) = \frac{1}{m} \sum_{j=1}^{m} F_{tj}(y). \tag{54}$$

Furthermore, denoting the mean and variance of each $F_{tj}(y)$ as μ_{tj} and σ_{tj}^2, one can show that the mean and variance of $G_t(y)$ are:

$$\mu_{G_t} = \frac{1}{m} \sum_{j=1}^{m} \mu_{tj}, \tag{55}$$

$$\sigma_{G_t}^2 = \frac{1}{m} \sum_{j=1}^{m} \sigma_{tj}^2 + \frac{1}{m} \sum_{j=1}^{m} \left(\mu_{tj} - \mu_{G_t} \right)^2, \tag{56}$$

respectively, see Appendix for derivation. As location diversity, namely, the second term in Eq. (56), increases, the linear opinion pool's variance, namely, $\sigma_{G_t}^2$, also increases, resulting in an underconfident forecast. In other cases, Winkler et al. [187] argued that it is desirable to solicit forecasts from experts, who are likely to have similar training, see similar data, and use similar forecasting methods. This leads to overconfident forecasts. To address both potential issues, Jose et al. [188] introduced the exterior- and interior-trimmed opinion pool to improve the calibration. In a subsequent paper, the idea of trimmed opinion pool was applied in conjuncture with random forests [189].

To trim an opinion pool, the forecasts need to be ordered. Two approaches are available for that: (1) ordering based on μ_j (or the mean approach), and (2) ordering based on F_j (or the CDF approach) [188]. In the mean approach, suppose k forecasts will be trimmed from either side, then the exterior-trimmed average forecast is simply:

$$G_t(y) = \frac{1}{m - 2k} \sum_{j=k+1}^{m-k} F_{tj}(y). \tag{57}$$

Similarly, the interior-trimmed average forecast is:

$$G_t(y) = \frac{1}{m - 2k} \left(\sum_{j=1}^{m/2-k} F_{tj}(y) + \sum_{j=m/2+k+1}^{m} F_{tj}(y) \right). \tag{58}$$

This notation applies when m is even, whereas if m is odd, one can modify the notation using the floor operator. An illustration of different opinion pools with $m = 6$ and $k = 1$, using the mean approach, is shown in Fig. 9. In the CDF approach, instead of ranking the mean, one trims according to the CDF values over the range of the forecast variable; see [188] for more details and illustrative plots.

Heuristics for predictive quantiles. Probabilistic forecasts can be issued as quantiles. Suppose Q quantiles are issued for time t, by each of the m forecasters, the τth ensemble quantile can be denoted using $\{q_{\tau,t1}, \dots, q_{\tau,tm}\}$, with $\tau \in \{\tau_1, \dots, \tau_Q\}$. Quite obviously, one can use simple averages of the τth ensemble quantile:

$$q_{\tau,t} = \frac{1}{m} \sum_{j=1}^{m} q_{\tau,tj}. \tag{59}$$

Similarly, exterior- and interior-trimmed averaging could be used when more than five forecasters are available [104]. Instead of using mean, averaging through median is also another option.

Naïve sorting [190], on the other hand, treats the $Q \times m$ quantiles as point forecasts, and reorders them into a new sequence, denoted using $\boldsymbol{S} = \{s_1, \dots, s_{Q \times m}\}$. Subsequently, the forecast draws Q quantiles from $\boldsymbol{S}$. Since the forecasts are now sorted, it is also possible to average blocks of m forecasts, that is,

$$q_{\tau_l,t} = \frac{1}{m} \sum_{i=(l-1)\times m+1}^{l \times m} s_i, \tag{60}$$

where $l = 1, \dots, Q$.

Heuristics for prediction intervals. As for PIs, Gaba et al. [191] outlined six of such heuristics, namely, (1) the simple average, (2) the median, (3) and envelope approach, (4) an interior trimming method, (5) an exterior trimming method, and (6) probability averaging of endpoints. These heuristics were subsequently promoted by Grushka-Cockayne and Jose [192] to post-process the PIs submitted during the M4 forecasting competition [193–195]. With little surprise, interval aggregation was found to improve the calibration and accuracy.

Given m sets of component PIs for time t, denoted using $\boldsymbol{L}_t = \{L_{t1}, \dots, L_{tm}\}$ and $\boldsymbol{U}_t = \{U_{t1}, \dots, U_{tm}\}$, the simple average method generate the final PI through

$$[L_t, U_t] = \left[\frac{1}{m} \sum_{j=1}^{m} L_{tj}, \frac{1}{m} \sum_{j=1}^{m} U_{tj} \right], \tag{61}$$

the median method uses

$$[L_t, U_t] = \left[\text{median}(\boldsymbol{L}_t), \text{median}(\boldsymbol{U}_t) \right], \tag{62}$$

and the envelope approach takes

$$[L_t, U_t] = \left[\min(\boldsymbol{L}_t), \max(\boldsymbol{U}_t) \right], \tag{63}$$

as the final PI. On the other hand, the interior and exterior trimming methods remove a portion of ranked PI forecasts from inside and outside, respectively. More specifically, if $\{L_{t1}, \dots, L_{tm}\}$ and $\{U_{t1}, \dots, U_{tm}\}$ are ranked, from lowest to highest, interior trimming uses

$$[L_t, U_t] = \left[\frac{1}{m-k} \sum_{j=1}^{m-k} L_{tj}, \frac{1}{m-k} \sum_{j=k+1}^{m} U_{tj} \right], \tag{64}$$

whereas exterior trimming uses

$$[L_t, U_t] = \left[\frac{1}{m-k} \sum_{j=k+1}^{m} L_{tj}, \frac{1}{m-k} \sum_{j=1}^{m-k} U_{tj} \right], \tag{65}$$

where k is the number of forecasts to be trimming from each set of ranked bounds. Lastly, probability averaging of endpoints assumes the lower and upper bounds of the PI come from a normal distribution. In other words, given m normal distributions, with the jth one having a CDF F_{tj}, whose $(\alpha/2)$th and $(1-\alpha/2)$th quantiles are L_{tj} and U_{tj}, the method has the final PI, such that $\frac{1}{m} \sum_{j=1}^{m} F_{tj}\left(L_{tj}\right) = \alpha/2$ and $\frac{1}{m} \sum_{j=1}^{m} F_{tj}\left(U_{tj}\right) = 1 - \alpha/2$. The reader is referred to [191] for the pros and cons of each method.

A solar forecasting case study on PI averaging can be found in [73], in which the authors combined the underconfident PeEn forecasts and overconfident AnEn-based forecasts through simple averaging. Despite the simplicity of the approach, significant improvement in terms of probabilistic predictive performance was observed at all test sites used in that work.

5.2.2. Linear combination methods

Winkler [196] outlined three additional methods to the opinion pool, to decide on weights w_j: (1) based on some kind of ranking, (2) based on a self-rating, and (3) based on scoring rules where past predictive distributions are compared to observations. It is the latter method—the use of a scoring rule to determine how to optimally combine predictive distributions—on which we focus. In this case, scoring rules calculated using past forecasts are optimization cost functions. In the simplest case, one can optimize the combining weights by minimizing the IGN or CRPS of n historical forecasts. To give perspective of such optimization procedures, an illustration is provided here. The traditional linear pool (TLP) of the ith training sample is in the form:

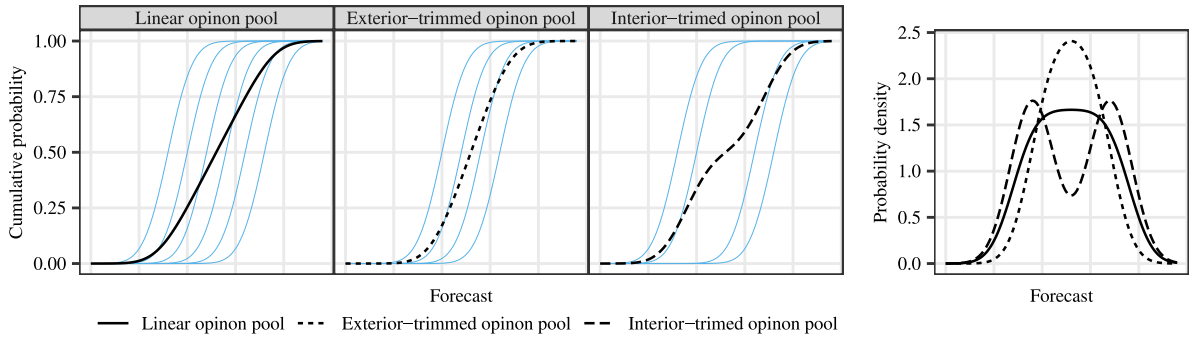


Fig. 9. CDFs of $m = 6$ component forecasts (blue), and that of linear, exterior-trimmed, and interior-trimmed opinion pools using the mean approach, with $k = 1$. The PDFs of these opinion pools are shown on the right.

$g_i(y) = \sum_{j=1}^{m} w_j f_{ij}(y)$. In this case, the likelihood of n samples is defined by:

$$\mathcal{L}_n(w_1, \dots, w_m) = \prod_{i=1}^{n} g_i(y_i), \tag{66}$$

where $i = 1, \dots, n$, whereas the log-likelihood is:

$$\ell_n(w_1, \dots, w_m) = \sum_{i=1}^{n} \log g_i(y_i) = \sum_{i=1}^{n} \log \left(\sum_{j=1}^{m} w_j f_{ij}(y_i) \right), \tag{67}$$

where $f_{ij}(y_i)$ is simply the value of f_{ij} evaluated at the ith observation, y_i. On the other hand, the average logarithmic score, i.e., IGN, over n samples is defined as:

$$\text{IGN} = \frac{1}{n} \sum_{i=1}^{n} \text{ign}_i = \frac{1}{n} \sum_{i=1}^{n} -\log g_i(y_i) = -\frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{m} w_j f_{ij}(y_i) \right). \tag{68}$$

It is immediately clear that $\text{IGN} = -n^{-1} \cdot \ell_n(w_1, \dots, w_m)$. Since n is fixed, which does not contribute to the optimization, maximizing the log-likelihood is equivalent of minimizing the averaged IGN. In most statistical software packages, the evaluation of PDFs and nonlinear optimization can be done with built-in functions. Hence, the remaining task is to simply write a cost function.

Predictive distributions. TLP has been proven to be highly capable of improving the calibration of underdispersed predictive distributions. Evidence exists in a variety of disciplines, e.g., [197–199]. In solar forecasting, there are but a few examples that applied linear pooling to improve the accuracy of probabilistic forecasts. In order to find the weights to combine the predictive distributions of three forecasting models, Bracale et al. [200] developed a single-objective optimization problem that minimizes the CRPS, as well as a multi-objective optimization problem that minimizes a combination of CRPS and the deviation from reliability. The selected forecasting models were a Bayesian model, a Markov-chain model, and a quantile regression model. The authors compared the PIT histograms of the individual models against those of the single- and multi-objective optimized combined predictive distributions. For the next-hour forecasts, the PIT histograms of the component forecasts were quite close to uniformity; consequently, the single-objective predictive distributions were overdispersed, whereas the multi-objective approach improved the calibration of the PIT histogram compared to the single-objective approach. Nonetheless, such improvement was not obvious when compared to the calibration of the component models. The main improvement was a reduction of approximately 3% in CRPS. The next-day forecast results revealed a similar pattern. An interesting study by Thorey et al. [201] aimed at linearly combining ensemble members and the deterministic forecasts of ECMWF and Météo France, and quantile forecasts derived from the residuals of the deterministic forecasts using a parameter-free online learning algorithm. The aim of the combination algorithm is to minimize the CRPS and comes with a theoretical guarantee on the performance. The results show that the method improves the calibration over the raw ensemble, which was underdispersive. Furthermore, the CRPS skill was improved for nearly all forecast horizons and location.

The main reason that linear pooling improves the calibration of underdispersed predictive distributions is that linear pooling increases dispersion [17]. Hora [202] and Ranjan and Gneiting [203] presented evidence that linear combinations of individual calibrated predictive distributions are uncalibrated, which is caused by the linear pool not being *flexibly dispersive*. An aggregation method is said to be flexibly dispersive when it comprises *combination formulas* $\mathcal{C}_\theta$—formulas that map m predictive distributions to a single distribution using parameter θ, e.g., $w_1, \dots, w_m$ in the case of TLP—such that the combined predictive distribution $\mathcal{C}_\theta(F_1, \dots, F_m)$ is neutrally dispersed, i.e., the variance of PIT is $1/12$, which is the variance of a uniform random variable over the range 0 to 1 [17]. Furthermore, Gneiting and Ranjan [17] defined the aggregation method to be *exchangeably* flexibly dispersive when $\mathcal{C}_\theta$ is *anonymous*. An example of an anonymous combination formula $\mathcal{C}_\theta$ is the opinion pool proposed by Stone [186] where the predictive distributions are equally weighted, while an example of an aggregation method is the linear pool. Gneiting and Ranjan [17] proved that the linear pool is not a flexibly dispersive aggregation method—*cf.* Theorem 3.1 (a) in [17]: "linearly combined forecast F is at least as dispersed as the least dispersed of the components $F_1, \dots, F_m$"—by giving the example where $Y \sim \mathcal{N}(0, 1)$ and the overdispersed predictive distributions $F_i \sim \mathcal{N}(0, j+1)$ for $j = 1, \dots, m$. In other words, it is straightforward to find an example where it is impossible to find a combination formula $\mathcal{C}_\theta\left(F_1, \dots, F_m\right)$ that results in a calibrated combined predictive distribution.

In addition to the linear pool, Gneiting and Ranjan [17] included *generalized* linear combination formulas for predictive cumulative distribution functions that were first introduced by Dawid et al. [204] for probability forecasts. Generalized linear combination formulas consist of a strictly monotone link function (recall the ones used in GAMLSS) [17],

$$g\left(G_i(y)\right) = \sum_{j=1}^{m} w_j g\left(F_{ij}(y)\right). \tag{69}$$

When $g(x) = x$, namely, the identity function, Eq. (69) is the linear pool. Gneiting and Ranjan [17] provided a list of link functions and the restrictions on the weights. To the authors' knowledge, the generalized linear pool has not been applied in the field of solar forecasting.

Besides evidence that the linear pool is not flexibly dispersive, Ranjan and Gneiting [203] also presented evidence that nonlinear combination methods can outperform linear pooling, which is the topic of Section 5.2.3.

Quantile forecasts. The methodology of linear combination of quantile forecasts is similar to that of predictive distributions. However, the major difference is the number of weights to be learned from the training data, which scales with the number of nominal probability levels τ that are considered. The linearly combined quantile forecast

thus becomes $q_\tau = \sum_{j=1}^{m} w_{\tau,j} q_{\tau,j}$, where $w_{\tau,j} \geq 0$ and $\sum_{j=1}^{m} w_{\tau,j} = 1$. The pinball loss function, defined in Eq. (8), is a proper score that can be used to find weights $w_{\tau,j}$ [205].

Bracale et al. [206], inspired by Wang et al. [190], took a more elaborate approach than Bracale et al. [200] and included periodicity and regularization when determining the quantile forecast combination weights, found by minimizing the pinball score. Three different models were used to produce quantile forecasts, namely: quantile k-nearest neighbor, quantile regression forests and quantile regression, based on the GEFCom2014 dataset. Unlike their previously discussed paper [200], Bracale et al. [206] did not include an extensive analysis via reliability diagrams or PIT histograms, which makes it challenging to determine whether the improvements of the combined predictive distributions over the individual models in terms of pinball score also reflect improved calibration.

Prediction intervals. Linear combination of prediction intervals is similar to linear combination of predictive distributions and quantiles, as we outlined above. Mathematically, it is defined as

$$\left[L_i, U_i\right] = \left[q_{\alpha/2,i}, q_{1-\alpha/2,i}\right] = \left[\sum_{j=1}^{m} w_j L_{ij}, \sum_{j=1}^{m} w_j U_{ij}\right]. \tag{70}$$

The main difference is the objective function, which should allow for the definition of a prediction interval. As mentioned in Section 1.2, a prediction interval is defined by a nominal coverage probability $1-\alpha$ and two quantiles that are centered around the median, i.e., a lower and upper bound. In order to ensure calibrated prediction intervals, the interval score is a proper score that can be applied to find weights w_j [146]

$$S_\alpha(L, U, Y) = (U - L) + \frac{2}{\alpha}(L - Y) \cdot \mathbf{1}_{Y<L} + \frac{2}{\alpha}(Y - U) \cdot \mathbf{1}_{Y>U}, \tag{71}$$

where $\mathbf{1}_{\text{cond}}$ is the indicator function, which takes the value 1 if "cond" is true and 0 otherwise. The interval score is closely related to the Winkler score [207]. Since the interval score is a negatively oriented score, it can be seen from Eq. (71) that it rewards sharp prediction intervals and penalizes poor reliability.

Ni et al. [208] applied the ensemble lower–upper bound estimation (ELUBE) method to estimate prediction intervals. The authors used the extreme learning machine (ELM), which is a single-layer feed-forward neural network, to produce prediction intervals. In order to create an ensemble, Ni et al. [208] used three activation functions, namely the sigmoid function, radial basis function and sine function. The objective function used to find the respective weights was the so-called coverage width criterion (CWC), which is a combination of the Winkler score [207] and another term that is based on the difference between the prediction interval coverage probability and the sharpness. It is important to note that CWC can lead to misinterpretations [18,209] and is preferably replaced by a proper score such as the interval score. Ni et al. [208] noted that the ensemble significantly outperformed the individual models in terms of the Winkler score. Unfortunately, no visual assessment of calibration was performed.

5.2.3. Nonlinear combination methods

In order to overcome the flexibility issues of TLP, Gneiting and Ranjan [17] proposed the spread-adjusted linear pool (SLP). SLP aims to adjust the spread of the individual predictive distributions through a strictly positive parameter c. SLP is defined as:

$$G_i(y) = \sum_{j=1}^{m} w_j F_{ij}^0\left(\frac{y - \mu_{ij}^*}{c}\right), \tag{72}$$

where μ_{ij}^* is the median of F_{ij} and $F_{ij} = F_{ij}^0\left(y - \mu_{ij}^*\right)$. For F_{ij} with $\mu_{ij} = \mu_{ij}^*$, the mean and variance of $G_i(y)$ are:

$$\mu_{G_i} = \sum_{j=1}^{m} w_j \mu_{ij}, \tag{73}$$

$$\sigma_{G_i}^2 = \sum_{j=1}^{m} w_j \sigma_{ij}^2 c^2 + \sum_{j=1}^{m} w_j \left(\mu_{ij} - \mu_{G_i}\right)^2. \tag{74}$$

These results are immediate, because $F_{ij}^0\left((y - \mu_{ij})/c\right)$ in Eq. (72) is in fact the CDF of the spread-adjusted predictand with mean μ_{ij} and variance $c^2\sigma_{ij}^2$, and thus the mean and variance equations of TLP shown in Appendix can be applied at once.

Underdispersed predictive distributions benefit from $c \geq 1$ while overdispersed or neutrally dispersed predictive distributions benefit from $c < 1$ [17]. As can be seen from Eq. (72), SLP reduces to TLP when $c = 1$. The flexibility of SLP is limited when the individual predictive distributions are underdispersed or neutrally dispersed and introducing a spread-adjustment parameter c for each component does not improve that [17]. In an attempt to calibrate the ECMWF ensemble prediction system by correcting the ensemble members using an autoregressive process based on recent errors (coined AR-EMOS), Möller and Groß [210] used SLP to combine the resulting predictive distributions with the predictive distributions acquired from the same ensemble through EMOS. Since EMOS showed to produce underdispersed predictive distributions on average while AR-EMOS produced overdispersed predictive distributions on average, combining EMOS and AR-EMOS with weights $w_1 = w_2 = 0.5$ and $c = 0.9$ substantially improved the calibration and sharpness of the predictive distributions. Möller and Groß [211] extended the AR-EMOS by including a heteroscedastic model of the variance that depended on a linear combination of the variance of the autoregressive errors and the current empirical variance of the corrected ensemble members. The authors showed substantial improvements over EMOS, particularly as the forecast horizon increased.

Another nonlinear combination method is the beta-transformed linear pool (BLP), which was introduced by Ranjan and Gneiting [203] and generalized by Gneiting and Ranjan [17]. It is formulated as [17]

$$G_i(y) = B_{\alpha,\beta}\left(\sum_{j=1}^{m} w_j F_{ij}(y)\right), \tag{75}$$

where $B_{\alpha,\beta}$ represents the beta CDF with parameters $\alpha > 0$ and $\beta > 0$. The weights in Eq. (75) are nonnegative and sum to one, same as the case of TLP. Moreover, BLP reduces to TLP when $\alpha = \beta = 1$. Gneiting and Ranjan [17] proved that when the weights $w_1, \ldots, w_m$ are fixed, the variance of the PIT can attain any value in the open interval $\left(0, \frac{1}{4}\right)$. Furthermore, Gneiting and Ranjan [17] proved that BLP is exchangeably flexibly dispersive with weights $w_1, \ldots, w_m = m^{-1}$.

Fatemi et al. [212] adopted BLP to combine two parametric predictive densities. The mean and variance of a future irradiance value were first forecast through two multiple regression models, separately, based solely on lagged irradiance variables. Subsequently, these forecast means and variances are converted into beta predictive distributions and two-sided power predictive distributions, with parameters optimized via maximum likelihood. Despite the BLP-combined forecasts showed marginal improvements over both component forecasts, the component models were in fact very poorly designed, because (1) no exogenous weather information was utilized whatsoever, and (2) the authors failed to notice the wide-accepted usage of clear-sky model in solar forecasting. To that end, the study in [212] was unable to reflect the state-of-the-art in solar forecasting, and the results were inconclusive.

5.2.4. Simulation study

In this section, we present a simulation study—adapted from Gneiting and Ranjan [17]—to show the efficacy of TLP, SLP and BLP, and compare it to the individual forecasts. The data is generated through a combination of standard normal random variables $X_0, X_1, X_2, X_3, \epsilon$:

$$Y = X_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \epsilon, \tag{76}$$

where ϵ represents an error term. In the simulation study, it is assumed that X_0 represents public information, e.g., from a weather service,

Table 1
Optimized parameters for TLP, SLP, and BLP.

	α	β	c	w_1	w_2	w_3
TLP	–	–	–	0.067	0.211	0.722
SLP	–	–	0.807	0.147	0.274	0.579
BLP	1.438	1.469	–	0.165	0.277	0.557

Table 2
The results of the individual and combined forecasts in terms of the variance of the PIT, the standard deviation of the forecasts and the averaged IGN over the training and test sets. Note that the PIT variance of neutrally dispersed forecasts is $\frac{1}{12}$ or approximately 0.083.

	$\mathbb{V}$(PIT)	σ	IGN$^{\text{train}}$	IGN$^{\text{test}}$
f_1	0.083	1.99	2.125	2.094
f_2	0.083	1.99	2.101	2.113
f_3	0.084	1.73	1.980	1.980
TLP	0.068	1.97	1.949	1.949
SLP	0.078	1.72	1.923	1.927
BLP	0.082	1.65	1.913	1.915

while X_1, X_2, X_3 represent information, e.g., from measurements, measured by forecasters 1, 2 and 3, respectively. Each forecaster therefore issues the following probabilistic forecasts:

$$f_1 = \mathcal{N}\left(X_0 + a_1 X_1, 1 + a_2^2 + a_3^2\right), \tag{77}$$

$$f_2 = \mathcal{N}\left(X_0 + a_2 X_2, 1 + a_1^2 + a_3^2\right), \tag{78}$$

$$f_3 = \mathcal{N}\left(X_0 + a_3 X_3, 1 + a_1^2 + a_2^2\right), \tag{79}$$

where $a_1 = a_2 = 1$ and $a_3 = 1.4$. Both the training and test data are of length $T = 1500$ and the parameters $w_1, w_2, w_3, c, \alpha, \beta$ are found by minimizing the IGN, which is a strictly proper score, defined in Eq. (68). The log-likelihood (negative of the IGN) of TLP can be calculated using Eq. (67), while the log-likelihood of SLP can be calculated as follows:

$$\ell_n(w_1, \dots, w_m, c) = \sum_{i=1}^{n} \log\left(\frac{1}{c}\sum_{j=1}^{m} w_j f_{ij}\left(\frac{y_i - \mu_{ij}}{c}\right)\right), \tag{80}$$

where μ_{ij} is the median of the jth component forecast. To calculate the log-likelihood of the beta-transformed linear pool, we use the derivation by Gneiting and Ranjan [17]:

$$\begin{aligned}\ell_n(w_1, \dots, w_m, \alpha, \beta) = \sum_{i=1}^{n} \Bigg\{ &(\alpha - 1)\log\left(\sum_{j=1}^{m} w_j f_{ij}(y_i)\right) \\ &+ (\beta - 1)\left(1 - \sum_{j=1}^{m} w_j f_{ij}(y_i)\right)\Bigg\} \\ &+ \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} w_i f_{ij}(y_i)\right) - n \cdot \log \mathrm{B}(\alpha, \beta),\end{aligned} \tag{81}$$

where B represents the beta function, not to be confused with the beta distribution.

Since all parameters should be greater or equal to zero and the weights should sum to one, we use `Rsolnp` [213] to optimize the logarithmic score using the aforementioned constraints. Since the optimization algorithm is sensitive to initial starting parameters, we run the algorithm 500 times and select the best outcome.

Table 1 presents the optimal parameters for the three combination techniques. As can be seen, the third forecaster is valued more due to the fact that the variance of his or her forecast is lower and the forecast is therefore sharper. Table 2 presents the quantitative results of the individual and the combined probabilistic forecasts in terms of the variance of the PIT, the standard deviation of the predictive distributions and the logarithmic score over the train and test sets. In addition, Fig. 10(a) presents the resulting PIT histograms. Table 2 indicates that the predictive distributions f_1, f_2 and f_3 are close to neutral dispersion, which results in overdispersed forecasts when f_1, f_2 and f_3 are combined linearly, as shown by the PIT histogram of the TLP. The PIT histogram of the SLP is closer to neutral dispersion than the PIT histogram of the TLP, as also indicated by its variance, but the combination method worsens the calibration of the individual forecasts as was expected since the flexibility is limited when the individual forecasts are neutrally dispersed [17]. The variance of the BLP PIT is close to neutrally dispersed, as can also be observed from the PIT histogram in Fig. 10(a). Moreover, the sharpness of the BLP is superior to the other combination techniques and the individual forecasts. Consequently, the logarithmic score on both the train and test sets is lowest for the BLP. Finally, Fig. 10(b) presents the predictive densities f_1, f_2 and f_3 in light blue, TLP, SLP and BLP in black and the observation in orange. As Fig. 10(b) shows, the location variation of the individual forecasts increases the variance of the TLP forecast substantially. In contrast, the SLP and BLP forecasts are sharper than the TLP forecast and the BLP technique produces the sharpest predictive density.

6. Outlook

From the review in the previous four sections, it should be clear that D2D, P2D, D2P, and P2P post-processing methods in solar forecasting have different levels of utilization and maturity. Nonetheless, they all revolve around the quantity for which the initial forecasts are generated, i.e., the predictand. Stated differently, if an initial forecast is made at location $\boldsymbol{s}$, the information used in post-processing also come from location $\boldsymbol{s}$, and if an initial forecast is made for irradiance, the post-processed forecast would also be for irradiance. What if forecasts from several locations over an area need to be post-processed simultaneously? What is the implication on post-processing when such initial forecasts have spatio-temporal interactions? What if the final quantity of interest (e.g., PV power) is different from the predictand (e.g., irradiance)?

Clearly, moving forward, it is necessary to place our research emphasis on those less developed and relatively immature aspects of solar forecast post-processing. In what follows, an outlook is presented, and various emerging technologies and trending research topics on post-processing solar forecasts are presented. The selection of these topics is based on importance and rarity. In other words, most of the topics discussed below are thought critical to advance the field further, yet, solar forecasters have put very little, if not no, attention thus far.

6.1. Hierarchical forecasting

A forecasting task is often associated with time series that have aggregation constraints due to geographical or temporal groupings. For example, the tourist arrivals for each state of United States should sum up exactly to the total tourist arrival for the country; the loads at distributions feeders should equal to the load at the corresponding transmission node minus losses. In solar forecasting, this is also the case, for example, the PV power produced by each string ought to be equal to the total generation of the PV plant, see Fig. 11 for an expanded illustration. Under such scenarios, hierarchical forecasting is needed. Hierarchical forecasting, also known as forecast reconciliation [214] or high-dimensional forecasting [215], is a trending topic, not only in solar forecasting, but forecasting in general. In fact, GEFCom2017 was dedicated to the hierarchical load forecasting problem [216], despite that only a small percentage of submissions fully leveraged the hierarchical information.

Due to the different data and information sets available at different levels and nodes of a hierarchy, the base forecasts—that is, forecasts generated for each series in a hierarchy, i.e., each node in Fig. 11—usually do not sum up, unless very naïve forecasting models, such as persistence, are used [214]. This is known as *aggregation inconsistency* or *incoherence*. To that end, hierarchical forecasting has two distinct advantages over conventional forecasting: (1) the final forecasts are

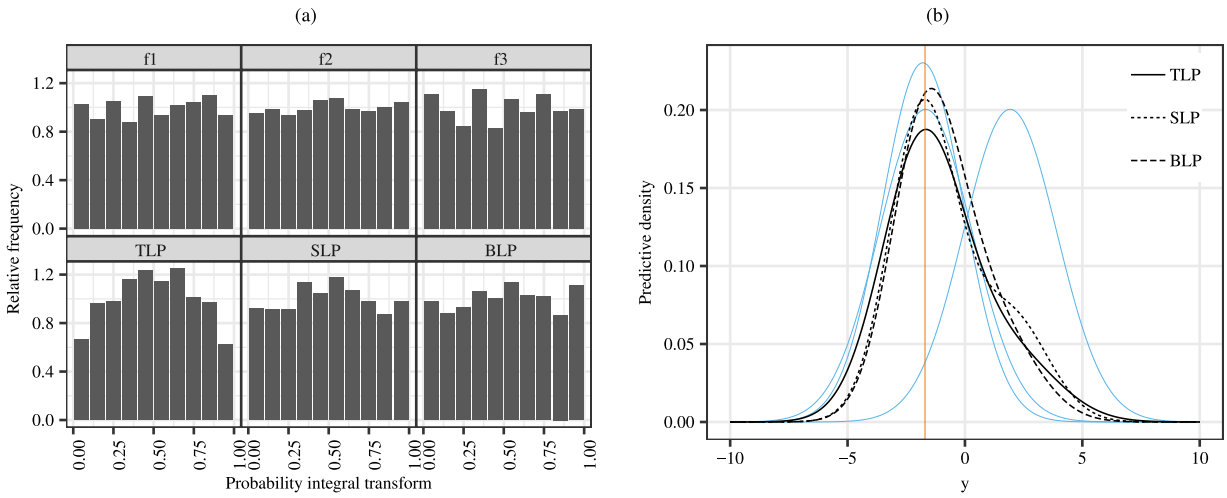


Fig. 10. The results of the simulation study. In (a), the PIT histograms of the individual (top row) and combined (bottom row) probabilistic forecasts. As the PIT histograms of the individual forecasts are close to uniform, the resulting PIT histogram of the TLP is overdispersed. In (b), the individual predictive densities in light blue, the combined predictive densities in black and the observation in orange.

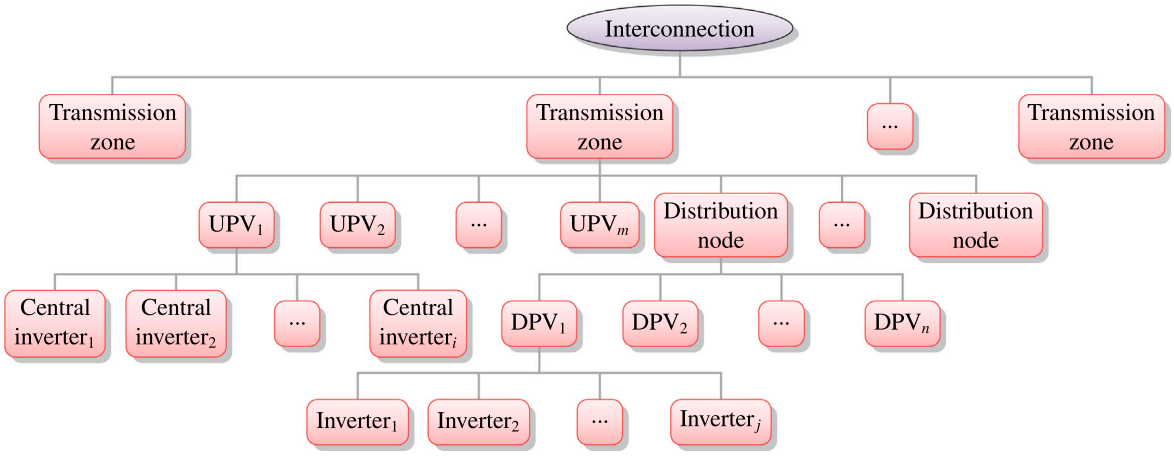


Fig. 11. An illustration of the hierarchical nature of grid-tied PV plants in a power system [217]. Abbreviations: UPV (utility-scale PV); DPV (distributed PV).

aggregate consistent and coherent, and (2) the final forecasts are often more accurate than base forecasts.

Traditional hierarchical forecasting can be categorized into three types, namely, bottom-up, middle-out, and top-down approaches. Each of these approaches only uses partial information from a hierarchy, e.g., the bottom-up approach only forecasts the bottom-level series, and adds those forecasts to yield middle- and top-level forecasts, so that the forecasts are naturally coherent. However, since part of the information is being ignored (e.g., bottom-up approach does not consider top- and middle-level information, or top-down approach does not consider bottom- and middle-level information), the performance of traditional methods is often suboptimal. Therefore more advanced methods for hierarchical forecasting are desired.

Recent advances of hierarchical forecasting can be attributed mostly to Rob Hyndman's research group, see [218] for an early formalism. The core idea in optimal hierarchical forecasting is to post-process the base forecasts under a regression framework. Over a series of publications, computation issues that hinder the large-scale applications of hierarchical forecasting have been addressed [214,219,220]. Subsequently, these advanced methods of hierarchical forecasting have been transferred to the solar forecasting community.

Yang et al. [217] presented the first work in hierarchical solar forecasting, which depicts a case study using a geographical hierarchy defined over a hypothetical transmission grid over California with 318 simulated PV systems. In a follow-up work, a temporal hierarchy was put forth using the same data [24]; the reason for having temporal hierarchies has been briefly mentioned in Section 2. Subsequently, Yagli et al. [221] continued from both works, and proposed a procedure called sequential reconciliation, which takes both geographical and temporal hierarchies into consideration, in a sequential manner. Very recently, Yang [149] further expanded the envelop of hierarchical solar forecast post-processing into probability space, and elaborated nonparametric probabilistic forecast reconciliation via a block bootstrapping method. In parallel, a parametric approach was demonstrated in [222]. In general, the quality of the final reconciled solar forecasts depends on two elements: (1) the accuracy of base forecasts [223], and (2) the choice of reconciliation technique [217].

6.2. *Spatial post-processing*

In the weather forecasting community, it is customary to verify forecasts spatially [224–227]. It is known *a priori* that spatial forecast verification uses different "truth" from forecast verification at point locations. For point-location forecast verification, the truth is almost always ground-based observations made at weather stations. However, since spatial forecasts are gridded data, it is necessary to used gridded data to verify those. More specifically, when verifying NWP forecasts, spatially, one must leverage remote-sensed data, because ground-based

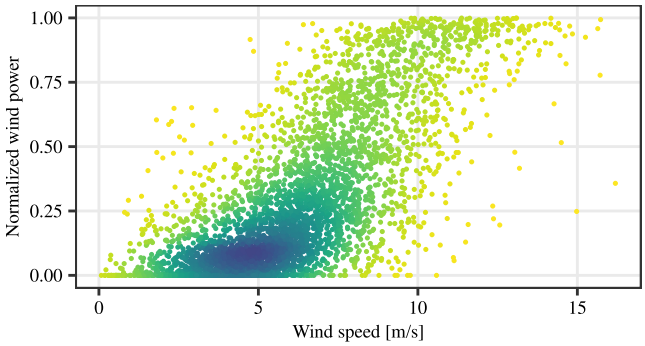


Fig. 12. A visualization of wind power curve using the GEFCom2014 data [3]. Wind speed is measured at a height of 100 m. Darker color indicates more points in the neighborhood.

monitoring stations are sparse, and thus do not provide sufficient coverage that can support spatial verification.

To the best of our knowledge, there is not any work on spatial forecast verification from the solar forecasting community. One major reason could be that high-quality remote-sensed radiation datasets were lacking, historically. Notwithstanding, with the advent of modern algorithms for deriving solar radiation from remote-sensing data, the latest generation of satellite-derived irradiance might offer such an opportunity [111]. Latest validation studies [63,228] suggested that the hourly satellite-derived irradiance data often has a RMSE of less than 15%, with negligible biases, and the numbers drop below 10% after site adaptation [164]. Some researchers also noted that state-of-the-art satellite-derived irradiance data is able to help quality control and detect calibration drift of ground measurements [229,230].

It is believed that gridded irradiance from remote-sensing data is going to further improve in time. Hence, the need for developing post-processing methods with satellite-derived irradiance is apparent. As mentioned earlier, such works already exist in the literature [61], despite that the application was restricted to point locations. Moving forward, one may consider matching the spatio-temporal structures of NWP forecasts and the satellite-derived irradiance, to achieve better outcomes. Spatio-temporal data-fusion methods for large-scale remote-sensing data, such as fixed-rank kriging [231,232] or Gaussian process [233,234], can be readily applied.

6.3. Irradiance-to-power conversion

In the wind forecasting literature, wind speed and wind power are related through a *power curve*. An example power curve is shown in Fig. 12, using a subset of data provided during GEFCom2014 [3]. It is apparent that the mapping from speed to power is non-injective. To that end, estimating wind power curve has been a major focus of the wind forecasting community [235–238]. Although the mapping from irradiance to PV power is also non-injective by nature, enthusiasm of developing solar power curve is not found in the solar forecasting community, in terms of converting irradiance forecasts to PV power forecasts.

Generally, the approaches for converting irradiance to PV power can be classified into direct and indirect ones. Direct approaches exploit the irradiance–power relationship through data-driven model; this is the setup of GEFCom2014 [3]. On the other hand, indirect approaches take two or three steps, each associated with a physical model. Given a PV array with a certain tilt and azimuth (jointly known as orientation), the global tilted irradiance (TGI) can be estimated from GHI and BNI via a so-called "transposition model" [239]. If BNI forecasts are unavailable, they can be derived from GHI forecasts through a so-called "separation model" [240]. Once the TGI is obtained, it is used together with other variables, such as ambient temperature, nominal operating cell temperature, or wind speed, to jointly estimate the PV power through a PV model [241].

As of now, the most accurate separation model is the one proposed by Yang and Boland [242], fairly recently. In contrast, the most accurate transposition model has been the Perez model [243], ever since it was proposed in 1990. The reader is referred to Yang [239], Yang and Gueymard [167], Gueymard and Ruiz-Arias [240] for the latest worldwide performance comparisons of different separation and transposition models in the literature. The PV model, however, largely depends on the granularity of modeling. The most significant contributor of errors is the overall system efficiency, η. There are countless inter-related factors that affect η, e.g., soiling, shading, potential-induced degradation, light-induced degradation, micro-cracks on cells, inverter efficiency, wiring loss, just to name a few. Hence, how much of this information can a modeler access, directly affects the accuracy of the PV model.

Despite solar engineers have been working on these topics for decades, as exemplified by the large collection of these solar models in the Python `pvlib` package, works that leverage such knowledge for irradiance-to-power conversion during forecasting are extremely rare. To the best of the authors' knowledge, there is no work that compares the direct and indirect conversion approaches. Hence, moving forward, this aspect ought to be addressed in depth, in order for solar forecasting to progress, in a way that wind forecasting did.

6.4. Stochastic simulations based on state-of-the-art forecasts

Stochastic simulations are used to quantify various uncertainties involved in a modeling process. Depending on the input uncertainty, it is of interest to observe how different outputs can turn out probabilistically. Typically, stochastic simulations require probabilistic forecasts that extend over various horizons, e.g., 1–24-h ahead, in order to simulate under uncertainty. Since probability distributions are defined on an interval along the real line there are infinitely many possible strategies, which renders a stochastic simulation computationally infeasible. Instead, one can sample a scenario from the consecutive probability distributions through the *inverse* probability integral transform $y_t = F_t^{-1}(u_t)$, where u_t is a sample from the standard uniform distribution $U[0,1]$ and F_t^{-1} is the *inverse* CDF or quantile function. In order for such a scenario to be meaningful, the random numbers for the entire forecast horizon $u_1, \ldots, u_{24}$ should be autocorrelated so that meaningful strategies can be found [215]. One way of generating correlated random numbers is through a copula, which is a multivariate CDF whose marginals are uniformly distributed on the interval $[0,1]$. This is particularly relevant because the definition of the copula allows for sampling correlated random uniform numbers from the multivariate distribution. Using the aforementioned inverse probability integral transform, it is subsequently possible to generate scenarios, also referred to as time trajectories or space–time trajectories in case spatial information is included as well [215].

Golestaneh et al. [244] employed a Gaussian copula to generate PV power space–time trajectories and noted that further research into suitable copulas is necessary. van der Meer et al. [245] followed up on the study by Golestaneh et al. [244] and tested four different copulas to generate space–time trajectories for probabilistic clear-sky index forecasts of the NREL pyranometer network on Oahu, Hawaii, and additionally introduced multivariate verification techniques developed in the atmospheric sciences. While the empirical copula was most flexible, each application requires careful verification similar to selecting the appropriate forecast model. Despite their relevance to the solar forecasting research field, the studies by Golestaneh et al. [244], van der Meer et al. [245] did not consider the application of the generated scenarios to power systems. To that end, Toubeau et al. [246] employed a copula to model temporal and cross-variable dependencies for short-term scheduling in power markets and showed that their approach increased the profits by 10%. Other approaches exist in the

literature, although it should be noted that stochastic simulation in power systems is a bulky domain, perhaps even bulkier than solar forecasting. Hence, only a handful of works are mentioned in this paper.

Ammar and Sharaf [247] discussed voltage regulation in the distribution network with the presence of centralized PV plants. A probabilistic profile of PV power generation was obtained from historical hourly ambient temperature and in-plane irradiance. In a distribution system probabilistic load flow study [248], Latin hypercube sampling was used to model the interaction of PV generation and load. Moving down the power system hierarchy from distribution level to user level, Zhou et al. [249] presented a sizing problem for home energy management system with a hybrid of PV and energy storage. To model the weather- and season-dependent variability of PV generation, the annual PV output data were divided into 12 scenarios according to weather conditions and seasons. In a similar study dealing with the resilience and operational cost in a commercial building, the uncertainty in PV power generation is realized through the remainder of a seasonal autoregressive integrated moving average model [250].

In all of the above-mentioned examples of stochastic simulation, the probabilistic representation of PV power generation was obtained from a single time series realization. This introduces some significant issues, among which the insufficient and under-represented PV power variability is most obvious. From a solar engineer's viewpoint, the overly smooth irradiance profiles appeared in the plots of those studies are far from the reality. Analogously, claiming the results from such stochastic simulations are acceptable is similar to the situation where one claims merit of a forecasting model solely based on its performance under clear skies, which is highly inappropriate.

A common issue with multidisciplinary studies is knowledge imbalance. For instance, power system engineers are familiar with the stochastic simulation methods, however, they are generally not aware of the latest developments in solar energy meteorology, particularly, probabilistic solar forecasting. The situation reverses when it comes to solar forecasting, that is, once the forecasts are produced, solar forecasters rarely present any case study on how to use them. This issue is known for decades, but very little has been done to address it, despite some scattered efforts from different individuals, e.g., through organizing special issues or establishing collaborations among different laboratories [251]. Unfortunately, we do not have much to offer in terms of strategies that can improve such pitfalls, but hope that the readers of this paper can become aware of this situation and initiate remedies on their own. On this point, there are however a handful of studies that leverage the state-of-the-art solar forecasting in power system simulation under high PV penetration, e.g., [252,253], which can typify the future works that are desirable. The reader is also referred to Li and Zhang [254] for a latest review on integration of probabilistic solar forecasting in power systems.

7. On the value of post-processing

The aim of post-processing is to improve the initial forecasts. No doubt, this usually refers to the improvement in forecast quality. After the discussions in the earlier sections, it should also be clear that the improvement can also mean a more appropriate representation of forecasts, depending on the preference of the user of the forecasts. For instance, the P2D post-processing, which summarizes a prediction distribution or combines several deterministic forecasts, improves the interpretability by trimming the information set. Similarly, the D2P post-processing, which dresses more forecasts or elicits quantiles from deterministic forecasts, improves the quantification of uncertainty.

Indeed, in the opening paragraph of the paper, three types of goodness of forecasts outlined by Murphy [1] are introduced. The third type of goodness—value—is the incremental benefits of forecasts to users. Here, the word "incremental" has to be emphasized. In a general sense, the incremental value of a forecast can only be acquired when it influences the decision made by users. In other words, the value denotes the difference gained (or occasionally, lost), e.g., in a monetary sense, after the forecast is utilized, with reference to a no-forecast situation. Such a gauge is often *ex post*. That is, the cost–loss ratio is computed after a sample of forecast–observation pairs has been collected.

Surely, there are pathways connecting the three types of goodness [1], and the one between value and quality is particularly apparent. It is generally the case that higher forecast quality suggests higher value, despite their relationship is almost surely nonlinear. On this point, it is important to measure quality, in an incremental sense, in order to ensure such improved quality is quantifiable. Stated differently, whenever a post-processing technique is adopted during a forecasting process, forecasts before and after post-processing must be contrasted, not only to each other, but with respect to a reference method. Section 7.1 further elaborates on this point, and it reviews the standard of reference used in both deterministic and probabilistic solar forecasting.

A forecast by itself has no value. Its value-add must be quantified by its users. In an extreme case, even when the forecast is perfect, it does not possess value if it cannot be integrated by forecast users into their decision-making process. Section 7.2 provides an example of such situations. It then follows to discuss a remedy proposed recently, by three groups of researchers, independently. The emphasis in Section 7.2 is on how to correctly describe a forecasting exercise in an operational context, and unfortunately, a vast majority of the present solar forecasting works fail to comply to such standard.

The users of forecasts, from a grid-integration aspect, can be divided into three main groups, namely, the independent system operators (ISOs), market participants, and solar power plant owners. Naturally, their needs differ. From an ISO's perspective, (regional) solar power generation forecasts are used during system planning and operations, with the goal of reducing reserves. In this case, regardless of forecast horizon, the value of forecasts is realized in terms of the reduced amount of reserves. For market participants and solar energy system owners, forecasts are used for optimal bidding in order to maximizing their profit. Hence, the value of forecasts is in terms of monetary rewards received from the power grid, subject to the cost for power production. Section 7.3 reviews, very briefly, some of the lessons learned on converting forecasts to value.

7.1. Reference solar forecasts

Most generally, the accuracy of weather forecasts is location- and time-period-dependent. In solar forecasting, the performance of a same algorithm at a desert location and at a tropical rain forest location can be dramatically different. Similarly, the forecasts made over a time period with predominant clear-sky conditions can be a lot more accurate and those made over a time period with mostly cloudy skies. In this regard, skill scores, which allow comparisons of forecasts made at different locations and over different time periods, have be ubiquitously used in meteorology [255,256].

In this section, following the notation used in the previous sections, (irradiance or solar power) forecasts are denoted with X, whereas the observations are denoted with Y. It is absolutely important to note that reference forecasts in solar forecasting are never generated using Y, directly. Rather, they are always generated using clear-sky index, and back-transformed to irradiance or power for verification [95].

The *skill score*, s, is computed using one minus the ratio between a negatively oriented accuracy measure (i.e., the perfect accuracy is 0) of some forecasts of interest and that of reference forecasts. Mathematically, denoting the accuracy measure using A, the skill score is expressed as:

$$s = 1 - \frac{A(X,Y)}{A(R,Y)}, \tag{82}$$

where R denotes the reference irradiance or power forecasts. For example, if A is RMSE, then

$$s = 1 - \frac{\mathrm{RMSE}(X,Y)}{\mathrm{RMSE}(R,Y)}, \tag{83}$$

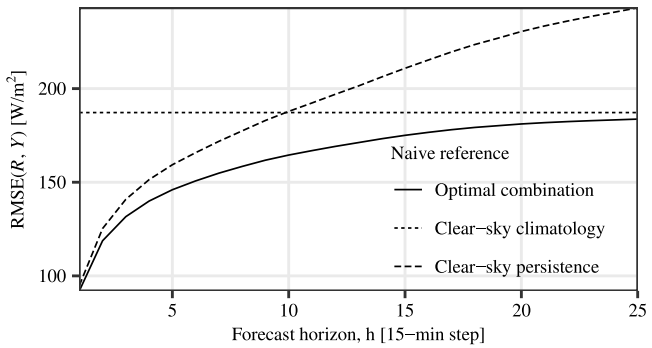


Fig. 13. RMSEs of h-step-ahead forecasts on 15-min GHI, at Table Mountain, Colorado (40.12498° N, 105.23680° W, 1689 m), over 2015–2018. Three naïve reference methods, namely, clear-sky climatology, clear-sky persistence, and their optimal convex combination, are applied on the clear-sky index retrieved using the REST2 model [96]. Subsequently, the forecast clear-sky index is converted back to GHI for verification.

and if A is CRPS, the expression for s is obtained by replacing "RMSE" in Eq. (83) with "CRPS". One can see that in order to have a positive score, the accuracy measure of the forecasts of interest must be lower than the accuracy measure of the reference forecasts, i.e., $A(X,Y) < A(R,Y)$.

7.1.1. Climatology–persistence combination

For deterministic solar forecasting, the clear-sky persistence, or persistence calculated using clear-sky index, was the standard of reference for a long time. Denoting the clear-sky index as κ, the h-step-ahead clear-sky persistence forecast is given by:

$$R_t^{(p)} = \frac{Y_{t-h}}{C_{t-h}} \cdot C_t = \kappa_{t-h} \cdot C_t, \tag{84}$$

where C is the clear-sky expectation and the superscript "(p)" denotes persistence. In words, clear-sky persistence uses the most recent clear-sky index observation as the forecast, and adjusts it with the clear-sky expectation at the forecast timestamp. Besides persistence, climatology is also a popular Ref. [56]. The clear-sky climatology first issues the mean of κ as forecasts, and then back-transforms that to irradiance:

$$R_t^{(c)} = \mathbb{E}(\kappa) \cdot C_t, \tag{85}$$

where the superscript "(c)" denotes climatology. Because climatology does not distinguish near forecasts from the more distant ones, it has the same RMSE for all horizons.

However, a pair of references published recently argued that the accuracy of clear-sky persistence drops as the horizon gets longer, regardless of whether it is a day-ahead, intra-day, or intra-hour scenario [257,258]. Moreover, the accuracy of clear-sky persistence will become worse than clear-sky climatology beyond a threshold horizon. To that end, using a combination of climatology and persistence could ensure that the reference forecasts stay optimal. This is illustrated in Fig. 13.

The climatology–persistence combination has several attractive statistical properties, see Yang [95] for a discussion. Most notable is that the mixing weights to the climatology and persistence parts can be easily optimized. More specifically, the *optimal climatology–persistence combination* forecast is given by:

$$R_t^{(cp)} = \left[\gamma_h \kappa_{t-h} + (1-\gamma_h)\mathbb{E}(\kappa)\right] \cdot C_t, \tag{86}$$

where the superscript "(cp)" denotes climatology–persistence combination, γ_h is the lag-h autocorrelation of the κ time series, and $\mathbb{E}(\kappa)$ is the mean of κ. Computationally, this reference method can be easily calculated with two or three lines of code, hence poses no difficulty to anyone with basic computer literacy. Whereas the reader is referred to [257,258] for worldwide case studies made at various day-ahead, intra-day, and intra-hour horizons, the optimal climatology–persistence combination is herein recommended as the standard of reference in deterministic solar forecasting. In fact, the exact recommendation has been made by a group of 33 solar forecasting experts in a recent review article on verification of deterministic solar forecasting [259].

7.1.2. Complete-history persistence ensemble

Persistence and climatology can also be used as reference methods during probabilistic forecast verification. PeEn uses m most recent observations as ensemble members. Climatology, in a probabilistic sense, is simply the unconditional distribution of the forecast variable. Similarly to the deterministic forecasting case, both PeEn and climatology forecasts need to be generated using κ. More specifically:

$$R_t^{(p)} = \left\{\kappa_{t-h} \cdot C_t, \kappa_{t-h-1} \cdot C_t, \ldots, \kappa_{t-h-m+1} \cdot C_t\right\}, \tag{87}$$

$$R_t^{(c)} = F^{-1}(\tau) \cdot C_t, \quad \forall \tau \in [0,1], \tag{88}$$

where F is the CDF of κ. To make climatology comparable to PeEn, the number of τ's can choose to follow m.

As in the case of AnEn, an obvious issue with PeEn is that the arbitrarity in the choice of m. PeEn forecasts are likely to suffer from overconfidence, for small m, and underconfidence, for big m. When m is equal to the total number of training samples, PeEn becomes climatology. Another problem with using a big m in PeEn, or using climatology, is that such forecasts lack discrimination, in that, the same predictive density is assigned to all forecast situations. Notwithstanding, climatology is calibrated by construct, which is desirable on that note.

Recall that the general goal for probabilistic forecasting is to minimize sharpness subject to calibration. Hence, a natural strategy to improve upon climatology is to use a conditional climatology, which is also calibrated by definition. Yang [260] offered one such possibility, called the complete-history persistence ensemble (CH-PeEn). CH-PeEn uses time-of-day as the conditional variable, and a predictive distribution is constructed using all training samples that have the same time-of-day. The choice is motivated by the fact that the clear-sky models are imperfect, especially in early mornings and late afternoons. Therefore, the conditional distributions of clear-sky index for different time-of-day are also different, see [95] for visualizations and hypothesis tests.

That said, it is certainly possible to use some other conditional variables, e.g., zenith angle or clear-sky expectations. However, they are thought to be less desirable than using time-of-day, because discretizing these continuous variables requires choosing a bin size, which again introduces arbitrarity. In contrast, CH-PeEn is easy to implement, and does not contain any ambiguity. To that end, CH-PeEn has been recommended by various research groups [261–263], independently, to be used as the standard of reference in probabilistic solar forecasting.

7.2. Operational requirements

As discussed, very briefly, in Section 2.3.3, forecasting performed in an actual operational context is extremely rare in the current solar forecasting literature. The reason is thought two-fold: (1) different energy markets around the world have different operational time requirements, hence, producing forecasts at standardized horizons may facilitate benchmarking; and (2) mimicking an operational forecasting context significantly complicates the experiment design, see [73] for full details. The cumbersome implementation causes most academic solar forecasters to choose a simplified forecasting procedure instead.

Notwithstanding, in a wind forecast integration study [264], a group of employees from three ISOs of North America noted that "The ISO's hour ahead forecast (real time market) is an average energy forecast for the following operating hour. The ISO does not receive a professionally prepared forecast for these types of wind events. To meet this need, the ISO has developed a forecast that is based on a

15-min persistence model". This direct quote describes a potential non-compliance situation between the ISO's operational time requirements and the forecasts submitted by the forecast service providers. To that end, no matter how accurate those forecasts were, they possessed no value, because the forecasts could not benefit the users.

Recently, at least three groups of researchers (independently) noted the need to emphasize operational solar forecasting [73,261,265]. As a result, four time parameters are proposed to fully describe any submission requirements of operational forecasting. The forecast span ($\mathcal{S}$) refers to the time span between the first forecast and the last forecast for a single submission; the forecast resolution ($\mathcal{R}_f$) is the temporal resolution of the required forecasts; the forecast lead time ($\mathcal{L}$) denotes the difference of a forecast submission time and the first forecast timestamp; and the forecast update rate ($\mathcal{U}$) specifies how often the forecasts are updated. For example, the CAISO's real-time market requires 15-min forecasts out to 5 h with a lead time of 75 min to be submitted every hour, see Fig. 14. In this case, the quadruplet $\left\{\mathcal{S}^{5\text{ h}}, \mathcal{R}_f^{15\text{ min}}, \mathcal{L}^{75\text{ min}}, \mathcal{U}^{60\text{ min}}\right\}$ fully characterizes the time requirement. From Fig. 14, it can be seen that for every timestamp, multiple forecasts will be received at different times through out a day. This complicates the forecast verification. The reader is referred to [266] for more details on verifying operational forecasts.

7.3. Value of forecasts for grid integration

Given the examples in Section 7.2, it should now be well understood that in order to quantify the value of forecasts, the directives have to come from the users. Questions such as "how should forecasts be submitted", "how can forecasts be used", or "how can improvements in forecasts benefit users" ought to be answered by the users before any incremental benefits of forecasts can be quantified [267]. Otherwise, the perceived value of forecast is just mere speculation or hypothesis of a forecaster. It is no surprise that works that align with this principle are rare. The majority of works that attempt to quantify the value of forecasts are based on forecasters' own setups. The reader is referred to [263] a short bibliography on this topic, to [268] for a full-length review, and to [269] for a full-scale simulation study on solar power integration.

For instance, Almeida et al. [270] spent a section in their paper discussing the impact of forecasts on trading in an electric power market. It assumed two scenarios: (1) the market penalizes the daily energy forecast based on mean bias error, and (2) the market penalizes the hourly energy forecast based on mean absolute error. Nonetheless, these assumptions were not supported by any reference nor evidence, which then led to a conclusion that is not meaningful.

In contrast, an excellent example of estimating forecast value can be found in [265]. Several day-ahead and intra-hour forecasting scenarios were carefully designed according to the market operations in the Western Interconnection Energy Imbalance Market (WI EIM), United States. The impacts on reserve scheduling were studied based on three sets of forecasts generated using two reference methods and a state-of-the-art method (post-processed NWP for day-ahead forecasts, and genetic-algorithm optimized SVR for intra-hour forecasts), respectively. In short, this paper by Kaur et al. [265] echoed most of our previous recommendations on quantifying value of forecasts, including (1) making good reference forecasts, (2) following operational requirements, and (3) considering the problem from the users' perspective. Given the fact that the paper was published in 2016, Kaur et al. [265] showed impressive foresight, and hence sets the standard for forthcoming works on quantifying value of forecasts.

From a market participant's perspective, Antonanzas et al. [271] presented a case study on economic benefits of forecasting under the Iberian electricity market, using data from a 1.86 MW PV system over a period of six months. The economic model proposed in the paper considered five situations according to market needs and the energy production offsets. A total of nine forecasting models (four SVR models, four random forest models, and persistence) were contrasted in terms of their materialized profit under the economic model. It was concluded that better forecasts generate more value, in a linear fashion. The linearity is a direct result of the assumed linear economic model, which might be over-simplified as compared to the actual situation.

8. Concluding remarks

A weather forecast is essentially five-dimensional, spanning space, time, and probability [272], and thus the best judgment of a forecaster is necessarily probabilistic. It follows that solar forecasts ought to be generated probabilistically, that is, in terms of predictive distribution, quantiles, prediction interval, or ensemble of forecasts. However, deterministic (point, single-valued, or best-guess) forecasts are easier to interpret, and are often preferred by the non-technical audience. Hence, flexible conversion between deterministic and probabilistic forecasts, post forecast generation, is needed. Moreover, by post-processing forecasts based on accumulated knowledge, the consistency, quality, and value of the forecasts can be enhanced. In this review, a typology that divides the post-processing task into four categories is proposed. More specifically, based on the type of conversions, a post-processing model could be (1) deterministic-to-deterministic (D2D), (2) probabilistic-to-deterministic (P2D), (3) deterministic-to-probabilistic (D2P), or (4) probabilistic-to-probabilistic (P2P). Under this typology, ten overarching thinking tools (i.e., styles or mechanisms) for post-processing solar forecasts are summarized.

These four categories of conversion and ten thinking tools could unify the way solar forecasters view post-processing. They are exceptionally useful in a sense that they consolidate myriads of post-processing methods into a manageable set of classes, facilitating review, summarization, making new proposals relevant to post-processing.

In D2D post-processing, a forecaster is tasked to correct the bias and variance, as well as to change the temporal resolution of the initial forecasts issued by a forecasting system. **Regression**, **filtering**, and various techniques for **resolution change** are commonly used frameworks to achieve such goals. In P2D post-processing, **summarizing a predictive distribution** or **combining deterministic forecasts** are the two major ways to optimally deduce the best-guess forecasts from probabilistic representations of future events. Albeit seemingly simple, the underlying statistical theories of optimally summarizing probabilistic forecasts still require attention. **Analog ensemble**, which exploits similar weather patterns in the past and present, and **method of dressing**, which bootstraps the observed errors and dresses them onto the deterministic-style forecasts, lie at the core of D2P post-processing. Additionally, one can also use **probabilistic regressions**, such as generalized additive model or quantile regression, for D2P. Lastly, in P2P post-processing, **calibrating ensemble forecasts** and **combining probabilistic forecasts** are new to solar forecasting, and thus deserve most attention moving into the future.

Since one of the main aims of producing solar forecasts is to facilitate grid integration of variable solar energy, hierarchical forecasting, which models the power injection at different parts of a power system, is thought to be a trend. Then, it immediately follows that other spatial post-processing methods are of interest. Analogous to fact that spatial forecasting is more advantageous to point-location forecasting, in terms of capturing the dynamical irradiance process, it is believed that spatial post-processing offers more flexibility as well. To use the post-processed forecasts during grid integration, adequate skills on stochastic simulations are needed. At the moment, there is a clear disconnection between the power system community and the solar energy community. Given the multidisciplinary nature of grid integration, it is necessary to bridge the two domains.

Value—the incremental benefits of forecasts to users—is a very much overlooked aspect in solar forecasting. Therefore, future research should place its priority on developing new post-processing frameworks, rather than developing method-specific hybrids that are easily

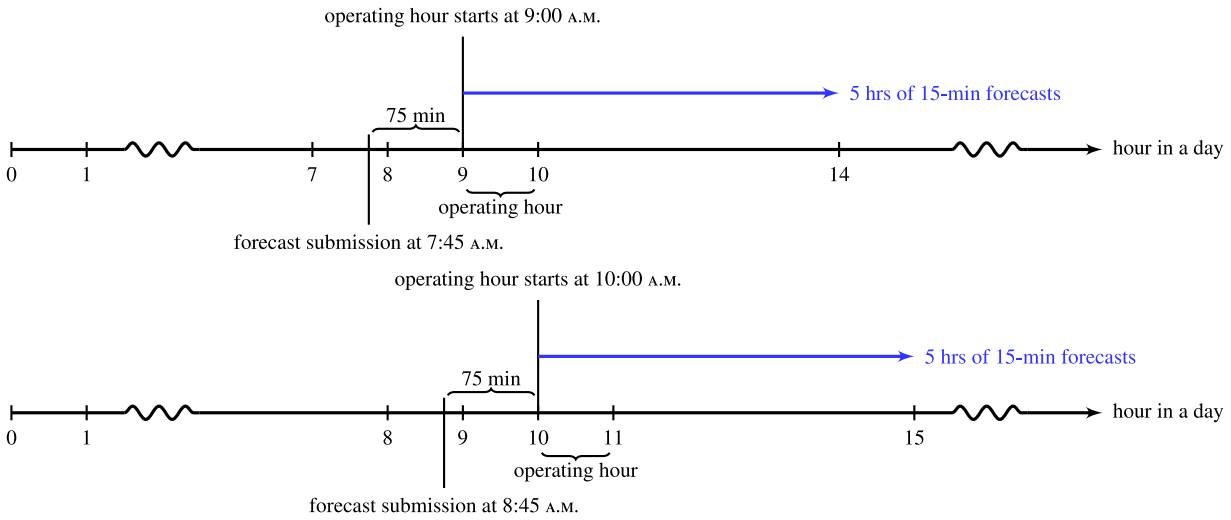


Fig. 14. Timeline of real-time market operation in CAISO. (top) For an operating hour starting at 9:00 A.M., 5 h of 15-min forecasts need to be submitted at 7:45 A.M., i.e., 75 min prior to the operating hour. (bottom) The submission requirements for the next operating cycle, which starts at 10:00 A.M..

replaceable. In order to demonstrate the value of post-processing, skill scores calculated based on the standard of reference allow comparison (with appropriate caveats) between forecasts made at different locations and over different time periods, as well as before and after post-processing. Submitting forecasts in exact required format is also critical to whether or not the forecasts can be appreciated by its users.

We end with a quote by Frank Vignola, a pioneer of modern radiometry and the owner of the longest-history high-resolution radiation measurement station: "*Too many people try to repeat what was done and do not add to the knowledge base. I try to show new ways to attack problems and do not always give a final answer. I try to leave enough for someone to follow what I did and improve on it. One of my biggest problems has been the uncertainty and systematic errors in the base data. That has prevented me from deriving concise models that incorporate systematic biases in the analysis. But as data gets better, people will see the old ideas and refine them and get better results as finally there are better instruments around. Forecasting is a growing field and has economic benefits. But there is a need to show people how to use the information and develop ways to facilitate its use. Also one should clearly identify the limits and problems.*" We fully agree with that.

Declaration of competing interest

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

Acknowledgments

D. van der Meer is funded by the projects "Probabilistic Forecasting for Battery Management" and "Development and evaluation of forecasting models for solar power and electricity use over space and time", both funded by the Swedish Energy Agency. This work also forms part of the Swedish strategic research program StandUp for Energy. The authors would like to thank Jan Kleissl from University of California San Diego (UCSD) for proofreading.

Appendix. Derivation of mean and variance expressions of a traditional linear pool

Given m component CDFs at time i, denoted using $F_{ij}(y)$, the traditional linear pool (TLP) suggests:

$$G_i(y) = \sum_{j=1}^{m} w_j F_{ij}(y). \tag{A.1}$$

The expectation of the combined distribution is:

$$\begin{aligned}\mathbb{E}_{G_i}(Y) &= \int_{-\infty}^{\infty} y \left[\sum_{j=1}^{m} w_j f_{ij}(y)\right] dy \\ &= w_1 \int_{-\infty}^{\infty} y f_{i1}(y) dy + \cdots + w_m \int_{-\infty}^{\infty} y f_{im}(y) dy \\ &= \sum_{j=1}^{m} w_j \mathbb{E}_{F_{ij}}(Y) \equiv \mu_{G_i},\end{aligned} \tag{A.2}$$

where

$$\mathbb{E}_{F_{ij}}(Y) \equiv \mu_{ij} = \int_{-\infty}^{\infty} y f_{ij}(y) dy. \tag{A.3}$$

On the other hand, the variance of the TLP distribution is:

$$\begin{aligned}\mathbb{V}_{G_i}(Y) &= \int_{-\infty}^{\infty} (y-\mu_{G_i})^2 \left[\sum_{j=1}^{m} w_j f_{ij}(y)\right] dy \\ &= \sum_{j=1}^{m} w_j \left\{\int_{-\infty}^{\infty} (y-\mu_{G_i})^2 f_{ij}(y) dy\right\} \\ &= \sum_{j=1}^{m} w_j \left\{\int_{-\infty}^{\infty} (y-\mu_{G_i}-\mu_{ij}+\mu_{ij})^2 f_{ij}(y) dy\right\} \\ &= \sum_{j=1}^{m} w_j \left\{\int_{-\infty}^{\infty} \left[(y-\mu_{ij})^2 + (\mu_{ij}-\mu_{G_i})^2 + 2(y-\mu_{ij})(\mu_{ij}-\mu_{G_i})\right] f_{ij}(y) dy\right\} \\ &= \sum_{j=1}^{m} w_j \mathbb{V}_{F_{ij}}(Y) + \sum_{j=1}^{m} w_j (\mu_{ij}-\mu_{G_i})^2 \equiv \sigma_{G_i}^2,\end{aligned} \tag{A.4}$$

where

$$\mathbb{V}_{F_{ij}}(Y) \equiv \sigma_{ij}^2 = \int_{-\infty}^{\infty} (y-\mu_{ij})^2 f_{ij}(y) dy. \tag{A.5}$$

Note that the last term in the second last row of Eq. (A.4), namely,

$$\begin{aligned}&\int_{-\infty}^{\infty} (y\mu_{ij} - \mu_{ij}^2 + \mu_{ij}\mu_{G_i} - y\mu_{G_i}) f_{ij}(y) \\ =&\mu_{ij} \int_{-\infty}^{\infty} y f_{ij}(y) dy - \mu_{ij}^2 \int_{-\infty}^{\infty} f_{ij}(y) dy + \mu_{ij}\mu_{G_i} \int_{-\infty}^{\infty} f_{ij}(y) dy - \mu_{G_i} \int_{-\infty}^{\infty} y f_{ij}(y) dy \\ =&\mu_{ij}^2 - \mu_{ij}^2 + \mu_{ij}\mu_{G_i} - \mu_{ij}\mu_{G_i} = 0.\end{aligned} \tag{A.6}$$

Therefore, one has:

$$\mu_{G_i} = \sum_{j=1}^{m} w_j \mu_{ij}, \tag{A.7}$$

$$\sigma_{G_i}^2 = \sum_{j=1}^{m} w_j \sigma_{ij}^2 + \sum_{j=1}^{m} w_j \left(\mu_{ij} - \mu_{G_i}\right)^2. \tag{A.8}$$

References

[1] Murphy AH. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecast. 1993;8(2):281–93. http://dx.doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

[2] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. In: Advances in Solar Resource Assessment and Forecasting. Solar Energy 2018;168:60–101. http://dx.doi.org/10.1016/j.solener.2017.11.023, URL http://www.sciencedirect.com/science/article/pii/S0038092X17310022.

[3] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. Int J Forecast 2016;32(3):896–913. http://dx.doi.org/10.1016/j.ijforecast.2016.02.001, URL http://www.sciencedirect.com/science/article/pii/S0169207016000133.

[4] Yang D. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). J. Renew. Sustain. Energy 2019;11(2):022701. http://dx.doi.org/10.1063/1.5087462.

[5] Hong T, Pinson P, Wang Y, Weron R, Yang D, Zareipour H. Energy forecasting: A review and outlook. IEEE Open Access J Power Energy 2020;7:376–88. http://dx.doi.org/10.1109/OAJPE.2020.3029979.

[6] Nouri B, Kuhn P, Wilbert S, Hanrieder N, Prahl C, Zarzalejo L, et al. Cloud height and tracking accuracy of three all sky imager systems for individual clouds. Solar Energy 2019;177:213–28. http://dx.doi.org/10.1016/j.solener.2018.10.079, URL http://www.sciencedirect.com/science/article/pii/S0038092X18310570.

[7] Larson VE. Forecasting solar irradiance with numerical weather prediction models. In: Solar energy forecasting and resource assessment. Elsevier; 2013, p. 299–318. http://dx.doi.org/10.1016/B978-0-12-397177-7.00012-7.

[8] Jones AS, Fletcher SJ. Data assimilation in numerical weather prediction and sample applications. In: Solar energy forecasting and resource assessment. Elsevier; 2013, p. 319–55. http://dx.doi.org/10.1016/B978-0-12-397177-7.00013-9.

[9] Makridakis S, Hyndman RJ, Petropoulos F. Forecasting in social settings: The state of the art. Int J Forecast 2020;36(1):15–28. http://dx.doi.org/10.1016/j.ijforecast.2019.05.011, M4 Competition. URL http://www.sciencedirect.com/science/article/pii/S0169207019301876.

[10] van der Meer D, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512. http://dx.doi.org/10.1016/j.rser.2017.05.212, URL http://www.sciencedirect.com/science/article/pii/S1364032117308523.

[11] Hammer A, Heinemann D, Lorenz E, Lückehe B. Short-term forecasting of solar radiation: A statistical approach using satellite data. Solar Energy 1999;67(1):139–50. http://dx.doi.org/10.1016/S0038-092X(00)00038-4, URL http://www.sciencedirect.com/science/article/pii/S0038092X00000384.

[12] Chow CW, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, Washom B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy 2011;85(11):2881–93. http://dx.doi.org/10.1016/j.solener.2011.08.025, URL http://www.sciencedirect.com/science/article/pii/S0038092X11002982.

[13] Mathiesen P, Kleissl J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states. Solar Energy 2011;85(5):967–77. http://dx.doi.org/10.1016/j.solener.2011.02.013, URL http://www.sciencedirect.com/science/article/pii/S0038092X11000570.

[14] Perez R, Lorenz E, Pelland S, Beauharnois M, van Knowe G, Hemker K, et al. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Solar Energy 2013;94:305–26. http://dx.doi.org/10.1016/j.solener.2013.05.005, URL http://www.sciencedirect.com/science/article/pii/S0038092X13001886.

[15] Roulston M, Smith L. Combining dynamical and statistical ensembles. Tellus A 2003;55(1):16–30. http://dx.doi.org/10.3402/tellusa.v55i1.12082.

[16] Bougeault P, Toth Z, Bishop C, Brown B, Burridge D, Chen DH, et al. The THORPEX interactive grand global ensemble. Bull Am Meteorol Soc 2010;91(8):1059–72. http://dx.doi.org/10.1175/2010BAMS2853.1.

[17] Gneiting T, Ranjan R. Combining predictive distributions. Electron J Stat 2013;7:1747–82. http://dx.doi.org/10.1214/13-EJS823.

[18] Lauret P, David M, Pinson P. Verification of solar irradiance probabilistic forecasts. Solar Energy 2019;194:254–71. http://dx.doi.org/10.1016/j.solener.2019.10.041, URL http://www.sciencedirect.com/science/article/pii/S0038092X19310382.

[19] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39(6):535–76. http://dx.doi.org/10.1016/j.pecs.2013.06.002, URL http://www.sciencedirect.com/science/article/pii/S0360128513000294.

[20] Kleissl J. Solar energy forecasting and resource assessment. Academic Press; 2013.

[21] Polo J, Fernández-Peruchena C, Salamalikis V, Mazorra-Aguiar L, Turpin M, Martín-Pomares L, et al. Benchmarking on improvement and site-adaptation techniques for modeled solar radiation datasets. Solar Energy 2020;201:469–79. http://dx.doi.org/10.1016/j.solener.2020.03.040, URL http://www.sciencedirect.com/science/article/pii/S0038092X20302784.

[22] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: Forecasting and control. John Wiley & Sons; 2015.

[23] Hyndman R, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing: The state space approach. Springer Science & Business Media; 2008.

[24] Yang D, Quan H, Disfani VR, Rodríguez-Gallegos CD. Reconciling solar forecasts: Temporal hierarchy. Solar Energy 2017;158:332–46. http://dx.doi.org/10.1016/j.solener.2017.09.055, URL http://www.sciencedirect.com/science/article/pii/S0038092X17308423.

[25] Athanasopoulos G, Hyndman RJ, Kourentzes N, Petropoulos F. Forecasting with temporal hierarchies. Eur J Oper Res 2017;262(1):60–74. http://dx.doi.org/10.1016/j.ejor.2017.02.046, URL http://www.sciencedirect.com/science/article/pii/S0377221717301911.

[26] Kazantzidis A, Tzoumanikas P, Blanc P, Massip P, Wilbert S, Ramirez-Santigosa L. Short-term forecasting based on all-sky cameras. In: Kariniotakis G, editor. Renewable Energy Forecasting. Woodhead Publishing Series in Energy, Woodhead Publishing; 2017, p. 153–78. http://dx.doi.org/10.1016/B978-0-08-100504-0.00005-6, URL http://www.sciencedirect.com/science/article/pii/B9780081005040000056.

[27] Peng Z, Yu D, Huang D, Heiser J, Yoo S, Kalb P. 3d cloud detection and tracking system for solar forecast using multiple sky imagers. Solar Energy 2015;118:496–519. http://dx.doi.org/10.1016/j.solener.2015.05.037, URL http://www.sciencedirect.com/science/article/pii/S0038092X15002972.

[28] Nguyen DA, Kleissl J. Stereographic methods for cloud base height determination using two sky imagers. Solar Energy 2014;107:495–509. http://dx.doi.org/10.1016/j.solener.2014.05.005, URL http://www.sciencedirect.com/science/article/pii/S0038092X14002333.

[29] Chu Y, Pedro HTC, Li M, Coimbra CFM. Real-time forecasting of solar irradiance ramps with smart image processing. Solar Energy 2015;114:91–104. http://dx.doi.org/10.1016/j.solener.2015.01.024, URL http://www.sciencedirect.com/science/article/pii/S0038092X15000389.

[30] Chu Y, Pedro HTC, Kaur A, Kleissl J, Coimbra CFM. Net load forecasts for solar-integrated operational grid feeders. Solar Energy 2017;158:236–46. http://dx.doi.org/10.1016/j.solener.2017.09.052, URL http://www.sciencedirect.com/science/article/pii/S0038092X17308393.

[31] Chu Y, Urquhart B, Gohari SMI, Pedro HTC, Kleissl J, Coimbra CFM. Short-term reforecasting of power output from a 48 MWe solar PV plant. Solar Energy 2015;112:68–77. http://dx.doi.org/10.1016/j.solener.2014.11.017, URL http://www.sciencedirect.com/science/article/pii/S0038092X14005611.

[32] Blanc P, Massip P, Kazantzidis A, Tzoumanikas P, Kuhn P, Wilbert S, et al. Short-term forecasting of high resolution local DNI maps with multiple fish-eye cameras in stereoscopic mode. AIP Conf. Proc. 2017;1850(1):140004. http://dx.doi.org/10.1063/1.4984512.

[33] Tukey JW. Exploratory data analysis, vol. 2. Addison-Wesley Publishing Company; 1977.

[34] Yang H, Kurtz B, Nguyen D, Urquhart B, Chow CW, Ghonima M, Kleissl J. Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy 2014;103:502–24. http://dx.doi.org/10.1016/j.solener.2014.02.044, URL http://www.sciencedirect.com/science/article/pii/S0038092X14001327.

[35] Chow CW, Belongie S, Kleissl J. Cloud motion and stability estimation for intra-hour solar forecasting. Solar Energy 2015;115:645–55. http://dx.doi.org/10.1016/j.solener.2015.03.030, URL http://www.sciencedirect.com/science/article/pii/S0038092X15001565.

[36] Miller SD, Rogers MA, Haynes JM, Sengupta M, Heidinger AK. Short-term solar irradiance forecasting via satellite/model coupling. Solar Energy 2018;168:102–17. http://dx.doi.org/10.1016/j.solener.2017.11.049, Advances in Solar Resource Assessment and Forecasting. URL http://www.sciencedirect.com/science/article/pii/S0038092X17310435.

[37] Blanc P, Remund J, Vallance L. Short-term solar power forecasting based on satellite images. In: Kariniotakis G, editor. Renewable Energy Forecasting. Woodhead Publishing Series in Energy, Woodhead Publishing; 2017, p. 179–98. http://dx.doi.org/10.1016/B978-0-08-100504-0.00006-8, URL http://www.sciencedirect.com/science/article/pii/B9780081005040000068.

[38] Wu X, Xiao Q, Wen J, You D, Hueni A. Advances in quantitative remote sensing product validation: Overview and current status. Earth-Sci Rev 2019;196:102875. http://dx.doi.org/10.1016/j.earscirev.2019.102875, URL http://www.sciencedirect.com/science/article/pii/S001282521830151X.

[39] Molero B, Leroux DJ, Richaume P, Kerr YH, Merlin O, Cosh MH, et al. Multi-timescale analysis of the spatial representativeness of in situ soil moisture data within satellite footprints. J. Geophys. Res.: Atmos. 2018;123(1):3–21. http://dx.doi.org/10.1002/2017JD027478, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD027478.

[40] Yang D. Quantifying the spatial scale mismatch between satellite-derived solar irradiance and in situ measurements: A case study using CERES synoptic surface shortwave flux and the Oklahoma Mesonet. J. Renew. Sustain. Energy 2020;12(5):056104. http://dx.doi.org/10.1063/5.0025771.

[41] Perez R, Kivalov S, Schlemmer J, Hemker K, Hoff TE. Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. Solar Energy 2012;86(8):2170–6. http://dx.doi.org/10.1016/j.solener.2012.02.027, Progress in Solar Energy 3. URL http://www.sciencedirect.com/science/article/pii/S0038092X12000928.

[42] Antonanzas-Torres F, Martínez-de Pisón FJ, Antonanzas J, Perpinan O. Downscaling of global solar irradiation in complex areas in R. J. Renew. Sustain. Energy 2014;6(6):063105. http://dx.doi.org/10.1063/1.4901539.

[43] Gueymard CA, Yang D. Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of AERONET observations. Atmos Environ 2020;225:117216. http://dx.doi.org/10.1016/j.atmosenv.2019.117216, URL http://www.sciencedirect.com/science/article/pii/S1352231019308556.

[44] Dambreville R, Blanc P, Chanussot J, Boldo D. Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model. Renew Energy 2014;72:291–300. http://dx.doi.org/10.1016/j.renene.2014.07.012, URL http://www.sciencedirect.com/science/article/pii/S096014811400398X.

[45] Yang D, Gu C, Dong Z, Jirutitijaroen P, Chen N, Walsh WM. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. Renew Energy 2013;60:235–45. http://dx.doi.org/10.1016/j.renene.2013.05.030, URL http://www.sciencedirect.com/science/article/pii/S0960148113002759.

[46] Yang D, Ye Z, Lim LHI, Dong Z. Very short term irradiance forecasting using the lasso. Solar Energy 2015;114:314–26. http://dx.doi.org/10.1016/j.solener.2015.01.016, URL http://www.sciencedirect.com/science/article/pii/S0038092X15000304.

[47] Cressie N, Wikle CK. Statistics for spatio-temporal data. John Wiley & Sons; 2015.

[48] Polo J, Wilbert S, Ruiz-Arias J, Meyer R, Gueymard C, Súri M, et al. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. Solar Energy 2016;132:25–37. http://dx.doi.org/10.1016/j.solener.2016.03.001, URL http://www.sciencedirect.com/science/article/pii/S0038092X16001754.

[49] Wu E, Yang H, Kleissl J, Suselj K, Kurowski MJ, Teixeira J. On the parameterization of convective downdrafts for marine stratocumulus clouds. Mon. Weather Rev. 2020;148(5):1931–50. http://dx.doi.org/10.1175/MWR-D-19-0292.1.

[50] Wu E, Clemesha RES, Kleissl J. Coastal stratocumulus cloud edge forecasts. Solar Energy 2018;164:355–69. http://dx.doi.org/10.1016/j.solener.2018.02.072, URL http://www.sciencedirect.com/science/article/pii/S0038092X18302056.

[51] Sahu DK, Yang H, Kleissl J. Assimilating observations to simulate marine layer stratocumulus for solar forecasting. Solar Energy 2018;162:454–71. http://dx.doi.org/10.1016/j.solener.2018.01.006, URL http://www.sciencedirect.com/science/article/pii/S0038092X18300185.

[52] Jimenez PA, Hacker JP, Dudhia J, Haupt SE, Ruiz-Arias JA, Gueymard CA, et al. WRF-solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction. Bull Am Meteorol Soc 2016;97(7):1249–64. http://dx.doi.org/10.1175/BAMS-D-14-00279.1.

[53] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2009;2(1):2–10. http://dx.doi.org/10.1109/JSTARS.2009.2020300.

[54] Glahn HR, Lowry DA. The use of model output statistics (MOS) in objective weather forecasting. J Appl Meteorol 1972;11(8):1203–11. http://dx.doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

[55] Pereira S, Canhoto P, Salgado R, Costa MJ. Development of an ANN based corrective algorithm of the operational ECMWF global horizontal irradiation forecasts. Solar Energy 2019;185:387–405. http://dx.doi.org/10.1016/j.solener.2019.04.070, URL http://www.sciencedirect.com/science/article/pii/S0038092X19304177.

[56] Lauret P, Lorenz E, David M. Solar forecasting in a challenging insular context. Atmosphere 2016;7(2):18. http://dx.doi.org/10.3390/atmos7020018, URL https://www.mdpi.com/2073-4433/7/2/18.

[57] Verzijlbergh RA, Heijnen PW, de Roode SR, Los A, Jonker HJJ. Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications. Solar Energy 2015;118:634–45. http://dx.doi.org/10.1016/j.solener.2015.06.005, URL http://www.sciencedirect.com/science/article/pii/S0038092X15003138.

[58] Verbois H, Huva R, Rusydi A, Walsh W. Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. Solar Energy 2018;162:265–77. http://dx.doi.org/10.1016/j.solener.2018.01.007, URL http://www.sciencedirect.com/science/article/pii/S0038092X18300197.

[59] Pierro M, Bucci F, Cornaro C, Maggioni E, Perotto A, Pravettoni M, Spada F. Model output statistics cascade to improve day ahead solar irradiance forecast. Solar Energy 2015;117:99–113. http://dx.doi.org/10.1016/j.solener.2015.04.033, URL http://www.sciencedirect.com/science/article/pii/S0038092X15002212.

[60] Rincón A, Jorba O, Frutos M, Alvarez L, Barrios FP, González JA. Bias correction of global irradiance modelled with weather and research forecasting model over paraguay. Solar Energy 2018;170:201–11. http://dx.doi.org/10.1016/j.solener.2018.05.061, URL http://www.sciencedirect.com/science/article/pii/S0038092X18304948.

[61] Yang D. Post-processing of NWP forecasts using ground or satellite-derived data through kernel conditional density estimation. J. Renew. Sustain. Energy 2019;11(2):026101. http://dx.doi.org/10.1063/1.5088721.

[62] Mejia JF, Giordano M, Wilcox E. Conditional summertime day-ahead solar irradiance forecast. Solar Energy 2018;163:610–22. http://dx.doi.org/10.1016/j.solener.2018.01.094, URL http://www.sciencedirect.com/science/article/pii/S0038092X18301154.

[63] Yang D, Bright JM. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27years. Solar Energy 2020;210:3–19. http://dx.doi.org/10.1016/j.solener.2020.04.016, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20303893.

[64] Yang D, Perez R. Can we gauge forecasts using satellite-derived solar irradiance? J. Renew. Sustain. Energy 2019;11(2):023704. http://dx.doi.org/10.1063/1.5087588.

[65] Yagli GM, Yang D, Gandhi O, Srinivasan D. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl Energy 2020;259:114122. http://dx.doi.org/10.1016/j.apenergy.2019.114122, URL http://www.sciencedirect.com/science/article/pii/S0306261919318094.

[66] Soubdhan T, Ndong J, Ould-Baba H, Do M-T. A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: Application to solar and photovoltaic prediction. Solar Energy 2016;131:246–59. http://dx.doi.org/10.1016/j.solener.2016.02.036, URL http://www.sciencedirect.com/science/article/pii/S0038092X16001444.

[67] Pelland S, Galanis G, Kallos G. Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model. Prog. Photovolt., Res. Appl. 2013;21(3):284–96. http://dx.doi.org/10.1002/pip.1180, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.1180.

[68] Diagne M, David M, Boland J, Schmutz N, Lauret P. Post-processing of solar irradiance forecasts from WRF model at Reunion Island. Solar Energy 2014;105:99–108. http://dx.doi.org/10.1016/j.solener.2014.03.016, URL http://www.sciencedirect.com/science/article/pii/S0038092X14001509.

[69] Yang D. On post-processing day-ahead NWP forecasts using Kalman filtering. Solar Energy 2019;182:179–81. http://dx.doi.org/10.1016/j.solener.2019.02.044, URL http://www.sciencedirect.com/science/article/pii/S0038092X19301707.

[70] Delle Monache L, Nipen T, Liu Y, Roux G, Stull R. Kalman filter and analog schemes to postprocess numerical weather predictions. Mon. Weather Rev. 2011;139(11):3554–70. http://dx.doi.org/10.1175/2011MWR3653.1.

[71] Makarov YV, Etingov PV, Ma J, Huang Z, Subbarao K. Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures. IEEE Trans. Sustain. Energy 2011;2(4):433–42. http://dx.doi.org/10.1109/TSTE.2011.2159254.

[72] Makarov YV, Guttromson RT, Huang Z, Subbarao K, Etingov PV, Chakrabarti BB, et al. Incorporating wind generation and load forecast uncertainties into power grid operations. Technical report, (PNNL-19189). Richland, Washington: Pacific Northwest National Laboratory; 2010.

[73] Yang D, Wu E, Kleissl J. Operational solar forecasting for the real-time market. Int J Forecast 2019;35(4):1499–519. http://dx.doi.org/10.1016/j.ijforecast.2019.03.009, URL http://www.sciencedirect.com/science/article/pii/S0169207019300755.

[74] Fernández Peruchena CM, Gastón M, Schroedter-Homscheidt M, Kosmale M, Martínez Marco I, García-Moya JA, et al. Dynamic Paths: Towards high frequency direct normal irradiance forecasts. Energy 2017;132:315–23. http://dx.doi.org/10.1016/j.energy.2017.05.101, URL http://www.sciencedirect.com/science/article/pii/S0360544217308575.

[75] Mueen A, Zhu Y, Yeh M, Kamgar K, Viswanathan K, Gupta C, et al. The fastest similarity search algorithm for time series subsequences under euclidean distance. 2017, http://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html.

[76] Yeh C, Zhu Y, Ulanova L, Begum N, Ding Y, Dau HA, Silva DF, Mueen A, Keogh E. Matrix Profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). 2016, p. 1317–22. http://dx.doi.org/10.1109/ICDM.2016.0179.

[77] Rakthanmanon T, Campana B, Mueen A, Batista G, Westover B, Zhu Q, Zakaria J, Keogh E. Searching and mining trillions of time series subsequences under dynamic time warping. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. 2012, p. 262–70, 9.

[78] Bright JM, Smith CJ, Taylor PG, Crook R. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. Solar Energy 2015;115:229–42. http://dx.doi.org/10.1016/j.solener.2015.02.032, URL http://www.sciencedirect.com/science/article/pii/S0038092X15001024.

[79] Frimane Â, Soubdhan T, Bright JM, Aggour M. Nonparametric Bayesian-based recognition of solar irradiance conditions: Application to the generation of high temporal resolution synthetic solar irradiance data. Solar Energy 2019;182:462–79. http://dx.doi.org/10.1016/j.solener.2019.02.052, URL http://www.sciencedirect.com/science/article/pii/S0038092X19301781.

[80] Bright JM. The impact of globally diverse GHI training data: Evaluation through application of a simple Markov chain downscaling methodology. J. Renew. Sustain. Energy 2019;11(2):023703. http://dx.doi.org/10.1063/1.5085236.

[81] Frimane Â, Bright JM, Yang D, Ouhammou B, Aggour M. Dirichlet downscaling model for synthetic solar irradiance time series. J. Renew. Sustain. Energy 2020;12(6):063702. http://dx.doi.org/10.1063/5.0028267.

[82] Munkhammar J, Widén J. An N-state Markov-chain mixture distribution model of the clear-sky index. Solar Energy 2018;173:487–95. http://dx.doi.org/10.1016/j.solener.2018.07.056, URL http://www.sciencedirect.com/science/article/pii/S0038092X18307205.

[83] Munkhammar J, van der Meer D, Widén J. Probabilistic forecasting of high-resolution clear-sky index time-series using a Markov-chain mixture distribution model. Solar Energy 2019;184:688–95. http://dx.doi.org/10.1016/j.solener.2019.04.014, URL http://www.sciencedirect.com/science/article/pii/S0038092X19303469.

[84] Perez R, Kivalov S, Schlemmer J, Hemker K, Hoff T. Parameterization of site-specific short-term irradiance variability. Solar Energy 2011;85(7):1343–53. http://dx.doi.org/10.1016/j.solener.2011.03.016, URL http://www.sciencedirect.com/science/article/pii/S0038092X11000995.

[85] Lave M, Kleissl J, Arias-Castro E. High-frequency irradiance fluctuations and geographic smoothing. Solar Energy 2012;86(8):2190–9. http://dx.doi.org/10.1016/j.solener.2011.06.031, Progress in Solar Energy 3. URL http://www.sciencedirect.com/science/article/pii/S0038092X11002611.

[86] Yang D, Jirutitijaroen P, Walsh WM. Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy 2012;86(12):3531–43. http://dx.doi.org/10.1016/j.solener.2012.07.029, Solar Resources. URL http://www.sciencedirect.com/science/article/pii/S0038092X12003039.

[87] Yang D, Sharma V, Ye Z, Lim LI, Zhao L, Aryaputera AW. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. Energy 2015;81:111–9. http://dx.doi.org/10.1016/j.energy.2014.11.082, URL http://www.sciencedirect.com/science/article/pii/S0360544214013528.

[88] Gneiting T. Making and evaluating point forecasts. J. Amer. Statist. Assoc. 2011;106(494):746–62. http://dx.doi.org/10.2307/41416407, URL http://www.jstor.org/stable/41416407.

[89] Jolliffe IT. The impenetrable hedge: A note on propriety, equitability and consistency. Meteorological Applications 2008;15(1):25–9. http://dx.doi.org/10.1002/met.60, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.60.

[90] Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, Fouilloy A. Machine learning methods for solar radiation forecasting: A review. Renew Energy 2017;105:569–82. http://dx.doi.org/10.1016/j.renene.2016.12.095, URL http://www.sciencedirect.com/science/article/pii/S0960148116311648.

[91] Ferlito S, Adinolfi G, Graditi G. Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. Appl Energy 2017;205:116–29. http://dx.doi.org/10.1016/j.apenergy.2017.07.124, URL http://www.sciencedirect.com/science/article/pii/S0306261917309972.

[92] Pedro HTC, Larson DP, Coimbra CFM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. J. Renew. Sustain. Energy 2019;11(3):036102. http://dx.doi.org/10.1063/1.5094494.

[93] Srivastava S, Lessmann S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Solar Energy 2018;162:232–47. http://dx.doi.org/10.1016/j.solener.2018.01.005, URL http://www.sciencedirect.com/science/article/pii/S0038092X18300173.

[94] Bergmeir C, Hyndman RJ, Benítez JM. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. Int J Forecast 2016;32(2):303–12. http://dx.doi.org/10.1016/j.ijforecast.2015.07.002, URL http://www.sciencedirect.com/science/article/pii/S0169207015001120.

[95] Yang D. Choice of clear-sky model in solar forecasting. J. Renew. Sustain. Energy 2020;12(2):026101. http://dx.doi.org/10.1063/5.0003495.

[96] Gueymard CA. REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – Validation with a benchmark dataset. Solar Energy 2008;82(3):272–85. http://dx.doi.org/10.1016/j.solener.2007.04.008, URL http://www.sciencedirect.com/science/article/pii/S0038092X07000990.

[97] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos Meas Tech 2013;6(9):2403–18. http://dx.doi.org/10.5194/amt-6-2403-2013, URL https://www.atmos-meas-tech.net/6/2403/2013/.

[98] Yang D. On adding and removing sensors in a solar irradiance monitoring network for areal forecasting and PV system performance evaluation. Solar Energy 2017;155:1417–30. http://dx.doi.org/10.1016/j.solener.2017.07.061, URL http://www.sciencedirect.com/science/article/pii/S0038092X17306461.

[99] Bates JM, Granger CWJ. The combination of forecasts. J Oper Res Soc 1969;20(4):451–68. http://dx.doi.org/10.1057/jors.1969.103.

[100] Clemen RT. Combining forecasts: A review and annotated bibliography. Int J Forecast 1989;5(4):559–83. http://dx.doi.org/10.1016/0169-2070(89)90012-5, URL http://www.sciencedirect.com/science/article/pii/0169207089900125.

[101] Diebold FX. Forecast combination and encompassing: Reconciling two divergent literatures. Int J Forecast 1989;5(4):589–92. http://dx.doi.org/10.1016/0169-2070(89)90014-9, URL http://www.sciencedirect.com/science/article/pii/0169207089900149.

[102] Diebold FX, Lopez JA. Forecast evaluation and combination. In: Handbook of statistics, vol. 14. 1996, p. 241–68. http://dx.doi.org/10.1016/S0169-7161(96)14010-4, URL http://www.sciencedirect.com/science/article/pii/S0169716196140104.

[103] de Menezes LM, Bunn DW, Taylor JW. Review of guidelines for the use of combined forecasts. Eur J Oper Res 2000;120(1):190–204. http://dx.doi.org/10.1016/S0377-2217(98)00380-4, URL http://www.sciencedirect.com/science/article/pii/S0377221798003804.

[104] Armstrong JS. Combining forecasts. In: Principles of forecasting. Springer; 2001, p. 417–39.

[105] Timmermann A. Forecast combinations. In: Elliott G, Granger CWJ, Timmermann A, editors. Handbook of economic forecasting, vol. 1. 2006, p. 135–96. http://dx.doi.org/10.1016/S1574-0706(05)01004-9, URL http://www.sciencedirect.com/science/article/pii/S1574070605010049.

[106] Wallis KF. Combining forecasts – Forty years later. Appl Financ Econ 2011;21(1–2):33–41. http://dx.doi.org/10.1080/09603107.2011.523179.

[107] Atiya AF. Why does forecast combination work so well? Int J Forecast 2020;36(1):197–200. http://dx.doi.org/10.1016/j.ijforecast.2019.03.010, M4 Competition. URL http://www.sciencedirect.com/science/article/pii/S0169207019300779.

[108] Blanc SM, Setzer T. When to choose the simple average in forecast combination. J Bus Res 2016;69(10):3951–62. http://dx.doi.org/10.1016/j.jbusres.2016.05.013, URL http://www.sciencedirect.com/science/article/pii/S0148296316303952.

[109] Genre V, Kenny G, Meyler A, Timmermann A. Combining expert forecasts: Can anything beat the simple average? Int J Forecast 2013;29(1):108–21. http://dx.doi.org/10.1016/j.ijforecast.2012.06.004, URL http://www.sciencedirect.com/science/article/pii/S016920701200088X.

[110] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98. http://dx.doi.org/10.1016/j.rser.2019.02.006, URL http://www.sciencedirect.com/science/article/pii/S1364032119300905.

[111] André M, Perez R, Soubdhan T, Schlemmer J, Calif R, Monjoly S. Preliminary assessment of two spatio-temporal forecasting technics for hourly satellite-derived irradiance in a complex meteorological context. Solar Energy 2019;177:703–12. http://dx.doi.org/10.1016/j.solener.2018.11.010, URL http://www.sciencedirect.com/science/article/pii/S0038092X18311101.

[112] Huang J, Rikus LJ, Qin Y, Katzfey J. Assessing model performance of daily solar irradiance forecasts over Australia. Solar Energy 2018;176:615–26. http://dx.doi.org/10.1016/j.solener.2018.10.080, URL http://www.sciencedirect.com/science/article/pii/S0038092X18310594.

[113] Haupt SE, Kosović B. Variable generation power forecasting as a big data problem. IEEE Trans. Sustain. Energy 2017;8(2):725–32. http://dx.doi.org/10.1109/TSTE.2016.2604679.

[114] Weiss CE, Raviv E, Roetzer G. Forecast combinations in R using the ForecastComb package. R J. 2018;10(2):262–81. http://dx.doi.org/10.32614/RJ-2018-052.

[115] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media; 2009.

[116] Miller A. Subset selection in regression. CRC Press; 2002.

[117] Thorey J, Mallet V, Chaussin C, Descamps L, Blanc P. Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database. Solar Energy 2015;120:232–43. http://dx.doi.org/10.1016/j.solener.2015.06.049, URL http://www.sciencedirect.com/science/article/pii/S0038092X15003576.

[118] Yang D. Ultra-fast preselection in lasso-type spatio-temporal solar forecasting problems. Solar Energy 2018;176:788–96. http://dx.doi.org/10.1016/j.solener.2018.08.041, URL http://www.sciencedirect.com/science/article/pii/S0038092X18308120.

[119] Rodríguez-Benítez FJ, Arbizu-Barrena C, Huertas-Tato J, Aler-Mur R, Galván-León I, Pozo-Vázquez D. A short-term solar radiation forecasting system for the Iberian Peninsula. Part 1: Models description and performance assessment. Solar Energy 2020;195:396–412. http://dx.doi.org/10.1016/j.solener.2019.11.028, URL http://www.sciencedirect.com/science/article/pii/S0038092X19311296.

[120] Huertas-Tato J, Aler R, Galván IM, Rodríguez-Benítez FJ, Arbizu-Barrena C, Pozo-Vázquez D. A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning. Solar Energy 2020;195:685–96. http://dx.doi.org/10.1016/j.solener.2019.11.091, URL http://www.sciencedirect.com/science/article/pii/S0038092X19311910.

[121] Yang D, Dong Z. Operational photovoltaics power forecasting using seasonal time series ensemble. Solar Energy 2018;166:529–41. http://dx.doi.org/10.1016/j.solener.2018.02.011, URL http://www.sciencedirect.com/science/article/pii/S0038092X18301282.

[122] Yang D, Alessandrini S. An ultra-fast way of searching weather analogs for renewable energy forecasting. Solar Energy 2019;185:255–61. http://dx.doi.org/10.1016/j.solener.2019.03.068, URL http://www.sciencedirect.com/science/article/pii/S0038092X19302944.

[123] Kyung M, Gill J, Ghosh M, Casella G. Penalized regression, standard errors, and bayesian lassos. Bayesian Anal 2010;5(2):369–411. http://dx.doi.org/10.1214/10-BA607.

[124] Lorenz EN. Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci. 1969;26(4):636–46. http://dx.doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

[125] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. Int J Forecast 2016;32(3):914–38. http://dx.doi.org/10.1016/j.ijforecast.2015.11.011, URL http://www.sciencedirect.com/science/article/pii/S0169207015001508.

[126] Pedro HTC, Coimbra CFM. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. Renew Energy 2015;80:770–82. http://dx.doi.org/10.1016/j.renene.2015.02.061, URL http://www.sciencedirect.com/science/article/pii/S0960148115001792.

[127] Alessandrini S, Delle Monache L, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. Appl Energy 2015;157:95–110. http://dx.doi.org/10.1016/j.apenergy.2015.08.011, URL http://www.sciencedirect.com/science/article/pii/S0306261915009368.

[128] Alessandrini S, Delle Monache L, Sperati S, Nissen J. A novel application of an analog ensemble for short-term wind power forecasting. Renew Energy 2015;76:768–81. http://dx.doi.org/10.1016/j.renene.2014.11.061, URL http://www.sciencedirect.com/science/article/pii/S0960148114007915.

[129] Junk C, Delle Monache L, Alessandrini S, Cervone G, von Bremen L. Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. Meteorol Z 2015;24(4):361–79. http://dx.doi.org/10.1127/metz/2015/0659.

[130] Junk C, Delle Monache L, Alessandrini S. Analog-based ensemble model output statistics. Mon. Weather Rev. 2015;143(7):2909–17. http://dx.doi.org/10.1175/MWR-D-15-0095.1.

[131] Davò F, Alessandrini S, Sperati S, Delle Monache L, Airoldi D, Vespucci MT. Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. Solar Energy 2016;134:327–38. http://dx.doi.org/10.1016/j.solener.2016.04.049, URL http://www.sciencedirect.com/science/article/pii/S0038092X16300962.

[132] Watanabe T, Nohara D. Prediction of time series for several hours of surface solar irradiance using one-granule cloud property data from satellite observations. Solar Energy 2019;186:113–25. http://dx.doi.org/10.1016/j.solener.2019.05.004, URL http://www.sciencedirect.com/science/article/pii/S0038092X19304499.

[133] Hyndman RJ, Wang E, Laptev N. Large-scale unusual time series detection. In: 2015 IEEE international conference on data mining workshop. 2015, p. 1616–9. http://dx.doi.org/10.1109/ICDMW.2015.104.

[134] Wang X, Smith K, Hyndman R. Characteristic-based clustering for time series data. Data Min Knowl Discov 2006;13(3):335–64. http://dx.doi.org/10.1007/s10618-005-0039-x.

[135] Kang Y, Hyndman RJ, Smith-Miles K. Visualising forecasting algorithm performance using time series instance spaces. Int J Forecast 2017;33(2):345–58. http://dx.doi.org/10.1016/j.ijforecast.2016.09.004, URL http://www.sciencedirect.com/science/article/pii/S0169207016301030.

[136] Yang D, Dong Z, Lim LHI, Liu L. Analyzing big time series data in solar engineering using features and PCA. Solar Energy 2017;153:317–28. http://dx.doi.org/10.1016/j.solener.2017.05.072, URL http://www.sciencedirect.com/science/article/pii/S0038092X17304796.

[137] Yang D, Goh GSW, Jiang S, Zhang AN. Spatial data dimension reduction using quadtree: A case study on satellite-derived solar radiation. In: 2016 IEEE international conference on big data. 2016, p. 3807–12. http://dx.doi.org/10.1109/BigData.2016.7841052.

[138] Ayet A, Tandeo P. Nowcasting solar irradiance using an analog method and geostationary satellite images. Solar Energy 2018;164:301–15. http://dx.doi.org/10.1016/j.solener.2018.02.068, URL http://www.sciencedirect.com/science/article/pii/S0038092X18301993.

[139] Yang D, Dong Z, Reindl T, Jirutitijaroen P, Walsh WM. Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. Solar Energy 2014;103:550–62. http://dx.doi.org/10.1016/j.solener.2014.01.024, URL http://www.sciencedirect.com/science/article/pii/S0038092X14000425.

[140] Yang D. Kriging for NSRDB PSM version 3 satellite-derived solar irradiance. Solar Energy 2018;171:876–83. http://dx.doi.org/10.1016/j.solener.2018.06.055, URL http://www.sciencedirect.com/science/article/pii/S0038092X18306066.

[141] Cervone G, Clemente-Harding L, Alessandrini S, Delle Monache L. Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble. Renew Energy 2017;108:274–86. http://dx.doi.org/10.1016/j.renene.2017.02.052, URL http://www.sciencedirect.com/science/article/pii/S0960148117301386.

[142] Yang D. Ultra-fast analog ensemble using kd-tree. J. Renew. Sustain. Energy 2019;11(5):053703. http://dx.doi.org/10.1063/1.5124711.

[143] Bentley JL. Multidimensional binary search trees used for associative searching. Commun ACM 1975;18(9):509–17. http://dx.doi.org/10.1145/361002.361007, 9.

[144] Arya S, Mount D, Kemp SE, Jefferis G. RANN: Fast nearest neighbour search (wraps ANN library) using L2 metric. 2019, R package version 2.6.1. URL https://CRAN.R-project.org/package=RANN.

[145] Mount DM. ANN programming manual. Maryland, United States: University of Maryland, College Park; 2010, URL https://www.cs.umd.edu/~mount/ANN/Files/1.1.2/ANNmanual_1.1.pdf.

[146] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat. Methodol. 2007;69(2):243–68. http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2007.00587.x.

[147] David M, Luis MA, Lauret P. Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data. Int J Forecast 2018;34(3):529–47. http://dx.doi.org/10.1016/j.ijforecast.2018.02.003, URL http://www.sciencedirect.com/science/article/pii/S0169207018300384.

[148] Pinson P, Kariniotakis G. Conditional prediction intervals of wind power generation. IEEE Trans. Power Syst. 2010;25(4):1845–56. http://dx.doi.org/10.1109/TPWRS.2010.2045774.

[149] Yang D. Reconciling solar forecasts: Probabilistic forecast reconciliation in a nonparametric framework. Solar Energy 2020;210:49–58. http://dx.doi.org/10.1016/j.solener.2020.03.095, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20303418.

[150] Athanasopoulos G, Gamakumara P, Panagiotelis A, Hyndman RJ, Affan M, et al. Hierarchical forecasting. In: Working paper 02/19. Department of Econometrics and Business Statistics. Monash University; 2019.

[151] Grantham A, Gel YR, Boland J. Nonparametric short-term probabilistic forecasting for solar radiation. Solar Energy 2016;133:465–75. http://dx.doi.org/10.1016/j.solener.2016.04.011, URL http://www.sciencedirect.com/science/article/pii/S0038092X16300342.

[152] Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. Wind Energy 2004;7(2):119–32. http://dx.doi.org/10.1002/we.114, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/we.114.

[153] Wasserman L. All of statistics: a concise course in statistical inference. Springer Science & Business Media; 2013.

[154] Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, Bastiani FD. Flexible regression and smoothing: The GAMLSS packages in R. Chapman and Hall/CRC; 2015.

[155] Brabec M, Paulescu M, Badescu V. Tailored vs black-box models for forecasting hourly average solar irradiance. Solar Energy 2015;111:320–31. http://dx.doi.org/10.1016/j.solener.2014.11.003, URL http://www.sciencedirect.com/science/article/pii/S0038092X14005350.

[156] Bakker K, Whan K, Knap W, Schmeits M. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. Solar Energy 2019;191:138–50. http://dx.doi.org/10.1016/j.solener.2019.08.044, URL http://www.sciencedirect.com/science/article/pii/S0038092X19308242.

[157] Koenker R. Quantile regression. Econometric society monographs, Cambridge University Press; 2005, http://dx.doi.org/10.1017/CBO9780511754098.

[158] Nagy GI, Barta G, Kazi S, Borbély G, Simon G. GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. Int J Forecast 2016;32(3):1087–93. http://dx.doi.org/10.1016/j.ijforecast.2015.11.013, URL http://www.sciencedirect.com/science/article/pii/S0169207015001521.

[159] Murphy AH, Winkler RL. A general framework for forecast verification. Mon. Weather Rev. 1987;115(7):1330–8. http://dx.doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2.

[160] Rosenblatt M. Remarks on a multivariate transformation. Ann Math Stat 1952;23(3):470–2. http://dx.doi.org/10.1214/aoms/1177729394.

[161] Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 2005;133(5):1155–74. http://dx.doi.org/10.1175/MWR2906.1.

[162] Gneiting T, Raftery AE, Westveld AH, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon. Weather Rev. 2005;133(5):1098–118. http://dx.doi.org/10.1175/MWR2904.1.

[163] Sperati S, Alessandrini S, Delle Monache L. An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting. Solar Energy 2016;133:437–50. http://dx.doi.org/10.1016/j.solener.2016.04.016, URL http://www.sciencedirect.com/science/article/pii/S0038092X1630041X.

[164] Yang D. Ensemble model output statistics as a probabilistic site-adaptation tool for satellite-derived and reanalysis solar irradiance. J. Renew. Sustain. Energy 2020;12(1):016102. http://dx.doi.org/10.1063/1.5134731.

[165] Yang D. Ensemble model output statistics as a probabilistic site-adaptation tool for solar irradiance: A revisit. J. Renew. Sustain. Energy 2020;12(3):036101. http://dx.doi.org/10.1063/5.0010003.

[166] Quan H, Yang D. Probabilistic solar irradiance transposition models. Renew Sustain Energy Rev 2020;125:109814. http://dx.doi.org/10.1016/j.rser.2020.109814, URL http://www.sciencedirect.com/science/article/pii/S136403212030109X.

[167] Yang D, Gueymard CA. Ensemble model output statistics for the separation of direct and diffuse components from 1-min global irradiance. Solar Energy 2020;208:591–603. http://dx.doi.org/10.1016/j.solener.2020.05.082, URL http://www.sciencedirect.com/science/article/pii/S0038092X2030582X.

[168] Hollands KGT, Suehrcke H. A three-state model for the probability distribution of instantaneous solar radiation, with applications. Solar Energy 2013;96:103–12. http://dx.doi.org/10.1016/j.solener.2013.07.007, URL http://www.sciencedirect.com/science/article/pii/S0038092X13002727.
[169] Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E. Calibrated probabilistic forecasting at the Stateline Wind Energy Center. J. Amer. Statist. Assoc. 2006;101(475):968–79. http://dx.doi.org/10.1198/016214506000000456.
[170] Yagli GM, Yang D, Srinivasan D. Ensemble solar forecasting using data-driven models with probabilistic post-processing through GAMLSS. Solar Energy 2020;208:612–22. http://dx.doi.org/10.1016/j.solener.2020.07.040, URL http://www.sciencedirect.com/science/article/pii/S0038092X20307738.
[171] Alessandrini S, Sperati S, Pinson P. A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. Appl Energy 2013;107:271–80. http://dx.doi.org/10.1016/j.apenergy.2013.02.041, URL http://www.sciencedirect.com/science/article/pii/S0306261913001499.
[172] Wang X, Bishop CH. Improvement of ensemble reliability with a new dressing kernel. Q. J. R. Meteorol. Soc. 2005;131(607):965–86. http://dx.doi.org/10.1256/qj.04.120, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.04.120.
[173] Fortin V, Favre A-C, Saïd M. Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. Q. J. R. Meteorol. Soc. 2006;132(617):1349–69. http://dx.doi.org/10.1256/qj.05.167, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.167.
[174] Aryaputera AW, Verbois H, Walsh WM. Probabilistic accumulated irradiance forecast for Singapore using ensemble techniques. In: 2016 IEEE 43rd photovoltaic specialists conference. 2016, p.1113–8.
[175] Doubleday K, Jascourt S, Kleiber W, Hodge B. Probabilistic solar power forecasting using Bayesian model averaging. 2020, p. 1. http://dx.doi.org/10.1109/TSTE.2020.2993524,
[176] Bremnes JB. Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. Mon. Weather Rev. 2004;132(1):338–47. http://dx.doi.org/10.1175/1520-0493(2004)132<0338:PFOPIT>2.0.CO;2.
[177] Ben Bouallègue Z. Statistical postprocessing of ensemble global radiation forecasts with penalized quantile regression. Meteorol Z 2017;26(3):253–64. http://dx.doi.org/10.1127/metz/2016/0748.
[178] Cannon AJ. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Comput Geosci 2011;37(9):1277–84. http://dx.doi.org/10.1016/j.cageo.2010.07.005, URL http://www.sciencedirect.com/science/article/pii/S009830041000292X.
[179] Huber PJ. Robust regression: asymptotics, conjectures and monte carlo. Ann Stat 1973;1(5):799–821. http://dx.doi.org/10.1214/aos/1176342503.
[180] Huber PJ. Robust estimation of a location parameter. Ann Math Stat 1964;35(1):73–101. http://dx.doi.org/10.1214/aoms/1177703732.
[181] Massidda L, Marrocu M. Quantile regression post-processing of weather forecast for short-term solar power probabilistic forecasting. Energies 2018;11(7):1763. http://dx.doi.org/10.3390/en11071763, URL https://www.mdpi.com/1996-1073/11/7/1763.
[182] Meinshausen N. Quantile regression forests. J. Mach. Learn. Res. 2006;7:983–99, URL http://www.jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf.
[183] Taillardat M, Mestre O, Zamo M, Naveau P. Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. Mon. Weather Rev. 2016;144(6):2375–93. http://dx.doi.org/10.1175/MWR-D-15-0260.1.
[184] Bremnes JB. Constrained quantile regression splines for ensemble postprocessing. Mon. Weather Rev. 2019;147(5):1769–80. http://dx.doi.org/10.1175/MWR-D-18-0420.1.
[185] Bremnes JB. Ensemble postprocessing using quantile function regression based on neural networks and Bernstein polynomials. Mon. Weather Rev. 2020;148(1):403–14. http://dx.doi.org/10.1175/MWR-D-19-0227.1.
[186] Stone M. The opinion pool. Ann Math Stat 1961;32(4):1339–42. http://dx.doi.org/10.1214/aoms/1177704873.
[187] Winkler RL, Grushka-Cockayne Y, Lichtendahl KC, Jose VRR. Probability forecasts and their combination: A research perspective. Decis Anal 2019;16(4):239–60. http://dx.doi.org/10.1287/deca.2019.0391.
[188] Jose VRR, Grushka-Cockayne Y, Lichtendahl KC. Trimmed opinion pools and the crowd's calibration problem. Manage. Sci. 2014;60(2):463–75. http://dx.doi.org/10.1287/mnsc.2013.1781.
[189] Grushka-Cockayne Y, Jose VRR, Lichtendahl KC. Ensembles of overfit and overconfident forecasts. Manage. Sci. 2017;63(4):1110–30. http://dx.doi.org/10.1287/mnsc.2015.2389.
[190] Wang Y, Zhang N, Tan Y, Hong T, Kirschen DS, Kang C. Combining probabilistic load forecasts. IEEE Trans. Smart Grid 2019;10(4):3664–74. http://dx.doi.org/10.1109/TSG.2018.2833869.
[191] Gaba A, Tsetlin I, Winkler RL. Combining interval forecasts. Decis Anal 2017;14(1):1–20. http://dx.doi.org/10.1287/deca.2016.0340.
[192] Grushka-Cockayne Y, Jose VRR. Combining prediction intervals in the M4 competition. Int J Forecast 2020;36(1):178–85. http://dx.doi.org/10.1016/j.ijforecast.2019.04.015, M4 Competition. URL http://www.sciencedirect.com/science/article/pii/S0169207019301141.
[193] Taleb NN. Foreword to the M4 Competition. Int J Forecast 2020;36(1):1–2. http://dx.doi.org/10.1016/j.ijforecast.2019.05.003, M4 Competition. URL http://www.sciencedirect.com/science/article/pii/S0169207019300883.
[194] Petropoulos F, Makridakis S. The M4 competition: Bigger. Stronger. Better. Int J Forecast 2020;36(1):3–6. http://dx.doi.org/10.1016/j.ijforecast.2019.05.005, M4 Competition. URL http://www.sciencedirect.com/science/article/pii/S0169207019301116.
[195] Hong T. Forecasting with high frequency data: M4 competition and beyond. Int J Forecast 2020;36(1):191–4. http://dx.doi.org/10.1016/j.ijforecast.2019.03.013, M4 Competition. URL http://www.sciencedirect.com/science/article/pii/S0169207019300846.
[196] Winkler RL. The consensus of subjective probability distributions. Manage. Sci. 1968;15(2):B61–75, URL http://www.jstor.org/stable/2628853.
[197] Mitchell J, Hall SG. Evaluating, comparing and combining density forecasts using the KLIC with an Application to the Bank of England and NIESR 'fan' charts of inflation. Oxford Bull Econ Stat 2005;67(s1):995–1033. http://dx.doi.org/10.1111/j.1468-0084.2005.00149.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.2005.00149.x.
[198] Garratt A, Mitchell J, Vahey SP, Wakerly EC. Real-time inflation forecast densities from ensemble Phillips curves. North Am J Econ Finance 2011;22(1):77–87. http://dx.doi.org/10.1016/j.najef.2010.09.003, Symposium on Nowcasting and Model Combination. URL http://www.sciencedirect.com/science/article/pii/S1062940810000410.
[199] Li T, Wang Y, Zhang N. Combining probability density forecasts for power electrical loads. IEEE Trans. Smart Grid 2020;11(2):1679–90.
[200] Bracale A, Carpinelli G, De Falco P. A probabilistic competitive ensemble method for short-term photovoltaic power forecasting. IEEE Trans. Sustain. Energy 2017;8(2):551–60. http://dx.doi.org/10.1109/TSTE.2016.2610523.
[201] Thorey J, Chaussin C, Mallet V. Ensemble forecast of photovoltaic power with online CRPS learning. Int J Forecast 2018;34(4):762–73. http://dx.doi.org/10.1016/j.ijforecast.2018.05.007, URL http://www.sciencedirect.com/science/article/pii/S016920701830089X.
[202] Hora SC. Probability judgments for continuous quantities: Linear combinations and calibration. Manage. Sci. 2004;50(5):597–604. http://dx.doi.org/10.1287/mnsc.1040.0205.
[203] Ranjan R, Gneiting T. Combining probability forecasts. J. R. Stat. Soc. Ser. B Stat. Methodol. 2010;72(1):71–91. http://dx.doi.org/10.1111/j.1467-9868.2009.00726.x.
[204] Dawid A, DeGroot M, Mortera J, Cooke R, French S, Genest C, et al. Coherent combination of experts' opinions. TEST 1995;4(2):263–313. http://dx.doi.org/10.1007/BF02562628, URL https://link.springer.com/article/10.1007/BF02562628.
[205] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 2007;102(477):359–78. http://dx.doi.org/10.1198/016214506000001437.
[206] Bracale A, Carpinelli G, De Falco P. Developing and comparing different strategies for combining probabilistic photovoltaic power forecasts in an ensemble method. Energies 2019;12(6):1011. http://dx.doi.org/10.3390/en12061011, URL https://www.mdpi.com/1996-1073/12/6/1011.
[207] Winkler RL. A decision-theoretic approach to interval estimation. J. Amer. Statist. Assoc. 1972;67(337):187–91, URL http://www.jstor.org/stable/2284720.
[208] Ni Q, Zhuang S, Sheng H, Kang G, Xiao J. An ensemble prediction intervals approach for short-term PV power forecasting. Solar Energy 2017;155:1072–83. http://dx.doi.org/10.1016/j.solener.2017.07.052, URL http://www.sciencedirect.com/science/article/pii/S0038092X17306370.
[209] Pinson P, Tastu J. Discussion of "Prediction intervals for short-term wind farm generation forecasts" and "Combined nonparametric prediction intervals for wind power generation". IEEE Trans Sustain Energy 2014;5(3):1019–20. http://dx.doi.org/10.1109/TSTE.2014.2323851.
[210] Möller A, Groß J. Probabilistic temperature forecasting based on an ensemble autoregressive modification. Q. J. R. Meteorol. Soc. 2016;142(696):1385–94. http://dx.doi.org/10.1002/qj.2741, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2741.
[211] Möller A, Groß J. Probabilistic temperature forecasting with a heteroscedastic autoregressive ensemble postprocessing model. Q. J. R. Meteorol. Soc. 2020;146(726):211–24. http://dx.doi.org/10.1002/qj.3667, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3667.
[212] Fatemi SA, Kuh A, Fripp M. Parametric methods for probabilistic forecasting of solar irradiance. Renew Energy 2018;129:666–76. http://dx.doi.org/10.1016/j.renene.2018.06.022, URL http://www.sciencedirect.com/science/article/pii/S0960148118306530.
[213] Ghalanos A, Theussl S. Rsolnp: General non-linear optimization using augmented Lagrange multiplier method. 2015, R package version 1.16..
[214] Wickramasuriya SL, Athanasopoulos G, Hyndman RJ. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. J. Amer. Statist. Assoc. 2019;114(526):804–19. http://dx.doi.org/10.1080/01621459.2018.1448825.
[215] Pinson P. Wind energy: Forecasting challenges for its operational management. Statistical Science 2013;28(4):564–85. http://dx.doi.org/10.1214/13-STS445.

[216] Hong T, Xie J, Black J. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. Int J Forecast 2019;35(4):1389–99. http://dx.doi.org/10.1016/j.ijforecast.2019.02.006.
[217] Yang D, Quan H, Disfani VR, Liu L. Reconciling solar forecasts: Geographical hierarchy. Solar Energy 2017;146:276–86. http://dx.doi.org/10.1016/j.solener.2017.02.010, URL http://www.sciencedirect.com/science/article/pii/S0038092X17301020.
[218] Athanasopoulos G, Ahmed RA, Hyndman RJ. Hierarchical forecasts for Australian domestic tourism. Int J Forecast 2009;25(1):146–66. http://dx.doi.org/10.1016/j.ijforecast.2008.07.004.
[219] Hyndman RJ, Ahmed RA, Athanasopoulos G, Shang HL. Optimal combination forecasts for hierarchical time series. Comput Stat Data Anal 2011;55(9):2579–89. http://dx.doi.org/10.1016/j.csda.2011.03.006.
[220] Hyndman RJ, Lee AJ, Wang E. Fast computation of reconciled forecasts for hierarchical and grouped time series. Comput Stat Data Anal 2016;97:16–32. http://dx.doi.org/10.1016/j.csda.2015.11.007.
[221] Yagli GM, Yang D, Srinivasan D. Reconciling solar forecasts: Sequential reconciliation. Solar Energy 2019;179:391–7. http://dx.doi.org/10.1016/j.solener.2018.12.075, URL http://www.sciencedirect.com/science/article/pii/S0038092X18312726.
[222] Yagli GM, Yang D, Srinivasan D. Reconciling solar forecasts: Probabilistic forecasting with homoscedastic Gaussian errors on a geographical hierarchy. Solar Energy 2020;210:59–67. http://dx.doi.org/10.1016/j.solener.2020.06.005, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20306137.
[223] Yagli GM, Yang D, Srinivasan D, Monika. Solar forecast reconciliation and effects of improved base forecasts. In: 2018 IEEE 7th world conference on photovoltaic energy conversion (wcpec) (a joint conference of 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC). 2018, p. 2719–23. http://dx.doi.org/10.1109/PVSC.2018.8547846.
[224] Marzban C, Sandgathe S. Optical flow for verification. Weather Forecast. 2010;25(5):1479–94. http://dx.doi.org/10.1175/2010WAF2222351.1.
[225] Gilleland E, Ahijevych DA, Brown BG, Ebert EE. Verifying forecasts spatially. Bull Am Meteorol Soc 2010;91(10):1365–76. http://dx.doi.org/10.1175/2010BAMS2819.1.
[226] Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. Intercomparison of spatial forecast verification methods. Weather Forecast. 2009;24(5):1416–30. http://dx.doi.org/10.1175/2009WAF2222269.1.
[227] Ebert EE. Neighborhood verification: A strategy for rewarding close forecasts. Weather Forecast. 2009;24(6):1498–510. http://dx.doi.org/10.1175/2009WAF2222251.1.
[228] Yang D. A correct validation of the National Solar Radiation Data Base (NSRDB). Renew Sustain Energy Rev 2018;97:152–5. http://dx.doi.org/10.1016/j.rser.2018.08.023, URL http://www.sciencedirect.com/science/article/pii/S1364032118306087.
[229] Urraca R, Gracia-Amillo AM, Huld T, de Pison FJM, Trentmann J, Lindfors AV, et al. Quality control of global solar radiation data with satellite-based products. Solar Energy 2017;158:49–62. http://dx.doi.org/10.1016/j.solener.2017.09.032, URL http://www.sciencedirect.com/science/article/pii/S0038092X17308046.
[230] Perez R, Schlemmer J, Kankiewicz A, Dise J, Tadese A, Hoff T. Detecting calibration drift at ground truth stations a demonstration of satellite irradiance models' accuracy. In: 2017 IEEE 44th photovoltaic specialist conference. 2017, p. 1104–9. http://dx.doi.org/10.1109/PVSC.2017.8366469.
[231] Nguyen H, Katzfuss M, Cressie N, Braverman A. Spatio-temporal data fusion for very large remote sensing datasets. Technometrics 2014;56(2):174–85. http://dx.doi.org/10.1080/00401706.2013.831774.
[232] Nguyen H, Cressie N, Braverman A. Spatial statistical data fusion for remote sensing applications. J. Amer. Statist. Assoc. 2012;107(499):1004–18. http://dx.doi.org/10.1080/01621459.2012.694717.
[233] Yang D, Gueymard CA. Producing high-quality solar resource maps by integrating high- and low-accuracy measurements using Gaussian processes. Renew Sustain Energy Rev 2019;113:109260. http://dx.doi.org/10.1016/j.rser.2019.109260, URL http://www.sciencedirect.com/science/article/pii/S136403211930468X.
[234] Zhang Q, Deng X, Qian PZG, Wang X. Spatial modeling for refining and predicting surface potential mapping with enhanced resolution. Nanoscale 2013;5:921–6. http://dx.doi.org/10.1039/C2NR33603K.
[235] Xu M, Pinson P, Lu Z, Qiao Y, Min Y. Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting. Wind Energy 2016;19(12):2321–36. http://dx.doi.org/10.1002/we.1985, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1985.
[236] Lee G, Ding Y, Genton MG, Xie L. Power curve estimation with multivariate environmental factors for inland and offshore wind farms. J. Amer. Statist. Assoc. 2015;110(509):56–67. http://dx.doi.org/10.1080/01621459.2014.977385.
[237] Jeon J, Taylor JW. Using conditional kernel density estimation for wind power density forecasting. J. Amer. Statist. Assoc. 2012;107(497):66–79. http://dx.doi.org/10.1080/01621459.2011.643745.
[238] Pinson P, Madsen H. Ensemble-based probabilistic forecasting at Horns Rev. Wind Energy 2009;12(2):137–55. http://dx.doi.org/10.1002/we.309, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/we.309.
[239] Yang D. Solar radiation on inclined surfaces: Corrections and benchmarks. Solar Energy 2016;136:288–302. http://dx.doi.org/10.1016/j.solener.2016.06.062, URL http://www.sciencedirect.com/science/article/pii/S0038092X16302432.
[240] Gueymard CA, Ruiz-Arias JA. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Solar Energy 2016;128:1–30. http://dx.doi.org/10.1016/j.solener.2015.10.010, URL http://www.sciencedirect.com/science/article/pii/S0038092X15005435.
[241] Nobre AM, Severiano CA, Karthik S, Kubis M, Zhao L, Martins FR, et al. PV power conversion and short-term forecasting in a tropical, densely-built environment in Singapore. Renew Energy 2016;94:496–509. http://dx.doi.org/10.1016/j.renene.2016.03.075, URL http://www.sciencedirect.com/science/article/pii/S0960148116302592.
[242] Yang D, Boland J. Satellite-augmented diffuse solar radiation separation models. J. Renew. Sustain. Energy 2019;11(2):023705. http://dx.doi.org/10.1063/1.5087463.
[243] Perez R, Ineichen P, Seals R, Michalsky J, Stewart R. Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy 1990;44(5):271–89. http://dx.doi.org/10.1016/0038-092X(90)90055-H, URL http://www.sciencedirect.com/science/article/pii/0038092X9090055H.
[244] Golestaneh F, Gooi HB, Pinson P. Generation and evaluation of space–time trajectories of photovoltaic power. Appl Energy 2016;176:80–91. http://dx.doi.org/10.1016/j.apenergy.2016.05.025.
[245] van der Meer D, Yang D, Widén J, Munkhammar J. Clear-sky index space-time trajectories from probabilistic solar forecasts: Comparing promising copulas. J. Renew. Sustain. Energy 2020;12(2):026102. http://dx.doi.org/10.1063/1.5140604.
[246] Toubeau J, Bottieau J, Vallée F, De Grève Z. Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. IEEE Trans. Power Syst. 2019;34(2):1203–15. http://dx.doi.org/10.1109/TPWRS.2018.2870041.
[247] Ammar M, Sharaf AM. Optimized use of PV distributed generation in voltage regulation: A probabilistic formulation. IEEE Trans. Ind. Inf. 2019;15(1):247–56. http://dx.doi.org/10.1109/TII.2018.2829188.
[248] Kabir M, Mishra Y, Bansal R. Probabilistic load flow for distribution systems with uncertain PV generation. Appl Energy 2016;163:343–51. http://dx.doi.org/10.1016/j.apenergy.2015.11.003, URL http://www.sciencedirect.com/science/article/pii/S0306261915014440.
[249] Zhou L, Zhang Y, Lin X, Li C, Cai Z, Yang P. Optimal sizing of PV and BESS for a smart household considering different price mechanisms. IEEE Access 2018;6:41050–9. http://dx.doi.org/10.1109/ACCESS.2018.2845900.
[250] Tavakoli M, Shokridehaki F, Akorede MF, Marzband M, Vechiu I, Pouresmaeil E. CVaR-based energy management scheme for optimal resilience and operational cost in commercial building microgrids. Int J Electr Power Energy Syst 2018;100:1–9. http://dx.doi.org/10.1016/j.ijepes.2018.02.022, URL http://www.sciencedirect.com/science/article/pii/S0142061517321415.
[251] Li W, Srinivasan D, Zhang J, Yang D. Preface of progress in solar energy special issue: Grid integration. Solar Energy 2020;210:1–2. http://dx.doi.org/10.1016/j.solener.2020.08.093, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20309324.
[252] Li C, Disfani VR, Pecenak ZK, Mohajeryami S, Kleissl J. Optimal OLTC voltage control scheme to enable high solar penetrations. Electr Power Syst Res 2018;160:318–26. http://dx.doi.org/10.1016/j.epsr.2018.02.016, URL http://www.sciencedirect.com/science/article/pii/S0378779618300592.
[253] Camal S, Teng F, Michiorri A, Kariniotakis G, Badesa L. Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications. Appl Energy 2019;242:1396–406. http://dx.doi.org/10.1016/j.apenergy.2019.03.112, URL http://www.sciencedirect.com/science/article/pii/S0306261919305203.
[254] Li B, Zhang J. A review on the integration of probabilistic solar forecasting in power systems. Solar Energy 2020;210:68–86. http://dx.doi.org/10.1016/j.solener.2020.07.066, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20307982.
[255] Murphy AH. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Weather Rev. 1988;116(12):2417–24. http://dx.doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.
[256] Murphy AH. Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather Forecast. 1992;7(4):692–8. http://dx.doi.org/10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2.
[257] Yang D. Standard of reference in operational day-ahead deterministic solar forecasting. J. Renew. Sustain. Energy 2019;11(5):053702. http://dx.doi.org/10.1063/1.5114985.
[258] Yang D. Making reference solar forecasts with climatology, persistence, and their optimal convex combination. Solar Energy 2019;193:981–5. http://dx.doi.org/10.1016/j.solener.2019.10.006, URL http://www.sciencedirect.com/science/article/pii/S0038092X19309880.
[259] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, et al. Verification of deterministic solar forecasts. Solar Energy 2020;210:20–37. http://dx.doi.org/10.1016/j.solener.2020.04.019, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20303947.

[260] Yang D. A universal benchmarking method for probabilistic solar irradiance forecasting. Solar Energy 2019;184:410–6. http://dx.doi.org/10.1016/j.solener.2019.04.018, URL http://www.sciencedirect.com/science/article/pii/S0038092X19303457.

[261] Doubleday K, Van Scyoc Hernandez V, Hodge B-M. Benchmark probabilistic solar forecasts: Characteristics and recommendations. Solar Energy 2020;206:52–67. http://dx.doi.org/10.1016/j.solener.2020.05.051, URL http://www.sciencedirect.com/science/article/pii/S0038092X20305429.

[262] Yang D, van der Meer D, Munkhammar J. Probabilistic solar forecasting benchmarks on a standardized dataset at Folsom, California. Solar Energy 2020;206:628–39. http://dx.doi.org/10.1016/j.solener.2020.05.020, URL http://www.sciencedirect.com/science/article/pii/S0038092X20305090.

[263] Antonanzas J, Perpinan-Lamigueiro O, Urraca R, Antonanzas-Torres F. Influence of electricity market structures on deterministic solar forecasting verification. Solar Energy 2020;210:44–6. http://dx.doi.org/10.1016/j.solener.2020.04.017, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X20303923.

[264] Kehler J, McMullen M, Blatchford J. ISO perspective and experience with integrating wind power forecasts into operations. In: IEEE PES general meeting. 2010, p. 1–5. http://dx.doi.org/10.1109/PES.2010.5590109.

[265] Kaur A, Nonnenmacher L, Pedro HTC, Coimbra CFM. Benefits of solar forecasting for energy imbalance markets. Renew Energy 2016;86:819–30. http://dx.doi.org/10.1016/j.renene.2015.09.011, URL http://www.sciencedirect.com/science/article/pii/S0960148115302901.

[266] Yang D. Comment: Operational aspects of solar forecasting. Solar Energy 2020;210:38–40. http://dx.doi.org/10.1016/j.solener.2020.04.014, Special Issue on Grid Integration. URL http://www.sciencedirect.com/science/article/pii/S0038092X2030390X.

[267] Luoma J, Mathiesen P, Kleissl J. Forecast value considering energy pricing in California. Appl Energy 2014;125:230–7. http://dx.doi.org/10.1016/j.apenergy.2014.03.061, URL http://www.sciencedirect.com/science/article/pii/S0306261914003018.

[268] Notton G, Nivet M-L, Voyant C, Paoli C, Darras C, Motte F, Fouilloy A. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. Renew Sustain Energy Rev 2018;87:96–105. http://dx.doi.org/10.1016/j.rser.2018.02.007, URL http://www.sciencedirect.com/science/article/pii/S1364032118300327.

[269] Brancucci Martinez-Anido C, Botor B, Florita AR, Draxl C, Lu S, Hamann HF, Hodge B-M. The value of day-ahead solar power forecasting improvement. Solar Energy 2016;129:192–203. http://dx.doi.org/10.1016/j.solener.2016.01.049, URL http://www.sciencedirect.com/science/article/pii/S0038092X16000736.

[270] Almeida MP, Perpiñán O, Narvarte L. PV power forecast using a nonparametric PV model. Solar Energy 2015;115:354–68. http://dx.doi.org/10.1016/j.solener.2015.03.006, URL http://www.sciencedirect.com/science/article/pii/S0038092X15001218.

[271] Antonanzas J, Pozo-Vázquez D, Fernandez-Jimenez LA, de Pison FJM. The value of day-ahead forecasting for photovoltaics in the Spanish electricity market. Solar Energy 2017;158:140–6. http://dx.doi.org/10.1016/j.solener.2017.09.043, URL http://www.sciencedirect.com/science/article/pii/S0038092X17308307.

[272] Allen MR, Stainforth DA. Towards objective probabalistic climate forecasting. Nature 2002;419(6903):228.