Title
Operational solar forecasting for the real-time market

Journal
International Journal of Forecasting, 35(4)

ISSN
0169-2070

Authors
Yang, Dazhi
Wu, Elynn
Kleissl, Jan

Publication Date
2019-10-01

DOI
10.1016/j.ijforecast.2019.03.009

Peer reviewed

Operational solar forecasting for the real-time market

Abstract

Despite the significant progress made in solar forecasting over the last decade, most of the proposed models cannot be readily used by independent system operators (ISOs). This article proposes an operational solar forecasting algorithm that is closely aligned with the real-time market (RTM) forecasting requirements of the California ISO (CAISO). The algorithm first uses the North American Mesoscale (NAM) forecast system to generate hourly forecasts for a 5-h period that are issued 12 h before the actual operating hour, satisfying the lead-time requirement. Subsequently, the world's fastest similarity search algorithm is adopted to downscale the hourly forecasts generated by NAM to a 15-min resolution, satisfying the forecast-resolution requirement. The 5-h-ahead forecasts are repeated every hour, following the actual rolling update rate of CAISO. Both deterministic and probabilistic forecasts generated using the proposed algorithm are empirically evaluated over a period of 2 years at 7 locations in 5 climate zones.

Keywords: Solar forecasting, Ensemble, Numerical weather prediction, Operational forecasting, Real-time market

1. Introduction

Integrating variable solar energy into the power grid requires forecasting of solar irradiance or power output of a photovoltaic (PV) or concentrating solar power (CSP) plant. While many innovative algorithms have been published in the solar forecasting literature, issues related to the implementation in an actual power system operational environment are generally not discussed. This trend has also been observed in load forecasting (Hong and Fan, 2016). Implementational issues are important to promoting energy forecasting to practitioners and to satisfy the ultimate goal of doing forecasting research which is to create knowledge for industrial applications (Hong and Fan, 2016). Examples of implementational issues include:

1. What is needed to build a database that is suitable for storing and retrieving data used in operational solar forecasting?
2. Is the algorithm fast enough for real-time wide-area operation. For example this could be a concern for computationally demanding data-driven methods are used?
3. How does the lead time—time between forecast submission and the start of an operating hour—affect the solar forecast error?
4. How to manipulate data to comply with forecasting resolution requirements? For example, how to convert hourly satellite-based forecasts to 15-min or 5-min forecasts that are required by the system operators in a way that maintains the mean and the variance of the raw forecast?

With the growing maturity of solar forecasting methods in recent years, some of the above-mentioned issues have started to draw attention from solar forecasters. For instance, Pedro et al. (2018) noticed the need to advance solar forecasting to a production stage and discussed the implementation of a solar forecasting MySQL database. Cervone et al. (2017) investigated the scalability of several data-driven methods, and confirmed the necessity of using supercomputers and parallel computing for operational applications. However, the time (referring here to lead time, horizon, and resolution) requirements in operational solar forecasting have been discussed less. This article discusses time requirements and illustrates their application through an operational solar forecasting method for the real-time market (RTM). More specifically, a state-of-the-art pattern-matching algorithm (PMA) is combined with hourly post-processed numerical weather prediction (NWP) forecasts, to produce deterministic and probabilistic forecasts at a higher time resolution that can directly be used by an independent system operator (ISO).

Nomenclature

Abbreviations

AnEn Analog Ensemble

ANN Artificial Neural Network

ARIMA AutoRegressive Integrated Moving Average

CAISO CAlifornia Independent System Operator

CSI Clear-Sky Index

ETS ExponenTial Smoothing

FFT Fast Fourier Transform

GHI Global Horizontal Irradiance

kNN k-Nearest Neighbor

MASS Mueen's Algorithm for Similarity Search

MOS Model Output Statistics

NAM North American Mesoscale

NWP Numerical Weather Prediction

PeEn Persistence Ensemble

QR Quantile Regression

RTED Real-Time Economic Dispatch

RTM Real-Time Market

RTUC Real-Time Unit Commitment

STL Seasonal and Trend decomposition using Loess

STUC Short-Term Unit Commitment

SURFRAD SURFace RADiation budget network

TBATS Trigonometric, Box–cox transform, Arma errors, Trend, and Seasonal

Terminologies for the similarity-search algorithm

$\boldsymbol{\Sigma}$ cector of moving sum-of-squares

history length-n history time series, i.e., n hours of historical ground measurements

l $l = n - m + 1$

m length of *query*

n length of *history*

query length-m query time series, i.e., m hours of NWP forecasts

Datasets and methods

Ens ensemble NAM forecasts (1-h resolution)

Nam raw NAM forecasts (1-h resolution)

Oracle oracle NAM forecasts (1-h resolution)

Pers smart persistence (15-min resolution)

Sarima seasonal ARIMA forecasts (15-min resolution)

Surfrad15 15-min aggregated ground-based measurements

Surfrad60 60-min aggregated ground-based measurements

1.1. Time-related issues in operational forecasting

For different grid operations in the day-ahead market and RTM, the forecasting requirements are also different in terms of *forecast horizon*. In the literature, there is a strong consensus on the choice of forecasting method for a given horizon (Inman et al., 2013). For day-ahead forecasting, NWP is almost always used, whereas satellite-based and statistical-learning methods are well-suited for a few hours ahead forecasting. Lastly, sky-camera-based forecasting has demonstrated its capability for a horizon shorter than 15 min. The reader is referred to a recent review for an overview of solar forecasting (Yang et al., 2018). The 6–8 h-ahead forecasting required by the RTM lies at the transition between satellite and NWP: while satellite data is often used for intra-day forecasting (e.g., Aguiar et al., 2016; Nonnenmacher and Coimbra, 2014), 6-h-ahead forecasts errors are typically double the error of 1-h-ahead forecasts, and the forecast horizon usually does not extend beyond 6 h (Perez et al., 2010). Therefore NWP is more suited to cover the full horizons required by the RTM.

Most NWP (and satellite) models only produce forecasts with an hourly resolution,[1] which is not granular enough for RTM applications. These mismatches in *forecast resolution* are rarely discussed in the literature. In statistical

[1]Most NWP models are capable of producing forecasts with higher temporal resolutions as the native time step is on the order of minutes, but due to data storage concerns the output is typically only hourly.

and machine-learning forecasting, the data resolution needs to match the forecast resolution. For example, when the phrase "hourly forecasting" is mentioned, most forecasting models would end up generating one forecast value per hour (1-step-ahead forecasting using 1-h aggregated data) (e.g., Bae et al., 2017; Shakya et al., 2017). On the contrary, what the grid operators need is in fact a series of high-resolution forecasts with smaller intervals, e.g., 5-min (Makarov et al., 2011). Therefore, for NWP applications to the RTM, raw 1-step-ahead forecasts with a 1-h resolution need to be *downscaled* to smaller intervals. Downscaling introduces additional forecast errors, hence, it is important to understand the propagation of errors in an actual operational scenario. Additional complications due to forecast resolution requirements are discussed in Appendix A.

The third time-related issue is *forecast lead time*.[2] In power systems research, the term "lead time" commonly refers to the time needed by the system operators to perform generator scheduling, unit commitment, and economic dispatch (Chen et al., 2017); in this article, for clarity, the forecast lead time is differentiated from forecast horizon. For example, the California Independent System Operator (CAISO) requires the day-ahead load forecasts to be submitted before 10:00 on the day prior to the operating day (Makarov et al., 2011), which corresponds to a lead time of 14 h. In a recent study, Yang and Dong (2018) showed that adding the lead time to the forecast horizon results in higher forecast errors, simply because it is harder to predict further into the future. Therefore, it is necessary to consider lead time when interpreting forecast error metrics, so that the operators has more realistic expectation for the uncertainty of the submitted forecasts. This distinction is rarely discussed in the solar forecast literature.

The last complication involved in operational forecasting is the *forecasting rolling update rate*. Although the forecasting requirement may state "5-h-ahead," this does not mean that the forecasts are produced every 5 h. Instead, the forecasts are usually produced in an hourly rolling manner (Kaur et al., 2016). For example, suppose forecasts for 9:00–14:00 were submitted at 7:45, the next submission will be at 8:45, for the period of 10:00–15:00 and the forecasts from 10:00–14:00 therefore are produced twice at different issue times and similarly six different forecasts are produced for every hour. Owing to this rolling nature of operational forecasting, the evaluation procedure is somewhat complicated, since there are multiple forecasts issued at different times apply to each timestamp. Although including a rolling update rate simply means a change in the forecast horizon, such a forecast setup is rarely demonstrated, which may lead to some ambiguity. For example, suppose the 5-h-ahead forecasting is run for 10 hours. If the rolling update rate is 5 h, there are 2 forecasts made for each forecast horizon. On the other hand, if the rolling update rate is 1 h, there are 10 forecasts made for each forecast horizon. This will directly affect the forecast evaluation and the reported metrics. These different forecasts made for the same timestamp need to be validated separately.

1.2. An overview of the proposed algorithm

Based on the above discussions, it can be concluded that there is a gap in the discussion and exemplification of operational solar forecasting models in the academic literature. In this paper, we present an operational forecast example and discuss the related implementation issues. An operational RTM forecast algorithm needs to have the following characteristics:

1. Sufficient stability for forecasting algorithm within the 5-h forecast horizon is desirable. Stability refers to homogeneity of the forecast error variance, i.e., constant or near constant root-mean-square errors across the different forecast horizons. Better stability implies higher confidence at far-away horizons, and thus reduces the bullwhip effect[3] in unit commitment.
2. The forecasting algorithm should be able to generate forecasts with granular resolutions. More specifically, some forecast downscaling methods are useful, when the raw forecasts are in an hourly resolution.
3. A distinction between the lead time and forecast horizon should be made, and no information *after* the forecast submission time should be used. In other words, all forecasts covering the lead time and forecast horizon need

[2] Lead time can be considered as part of the total forecast horizon. In other words, a lead time t simply means that the forecasts generated up to t are irrelevant.

[3] This is a phenomenon seen in supply-chain management; it refers to increasing swings in the inventory in response to shifts in customer demand. Supply-chain entities further up, such as the manufacturer, are more affected. In the present case, if each nodal-level forecast is over-dispersed, such conservative planning strategy may cascade to a very large required reserve at the power system level, which will be difficult for the ISO to meet.

to be prepared strictly before the submission time.[4]

4. Given the difference between the forecast horizon h and rolling update rate r, $\lceil h/r \rceil$ forecasts would be made *for each* timestamp, at different forecast submission times. Furthermore, the evaluation should be done $\lceil h/r \rceil$-times, based on these different forecasts made for the same timestamps.

To that end, an NWP-based data-driven algorithm, based on *pattern matching*, is thus proposed in this article to close the gap.

First of all, NWP is chosen due to its ability to model and assimilate the atmospheric physics in continuous time. Physically-based methods have the distinct advantage over satellite-based or statistical-learning methods in capturing the complex evolution of weather throughout a day up to several days ahead. More specifically, the North American Mesoscale (NAM) forecast system, a major weather model run by the National Centers for Environmental Prediction (NCEP), is used. However, NAM only produces forecasts with a 1-h resolution, which is not sufficient for RTM. To comply with the forecast-resolution requirement, these 1-h forecasts are downscaled to a shorter timescale (15 min in this case). This downscaling is achieved using a similarity-search algorithm (Mueen et al., 2017), by matching a length-m forecast time series at 1 h resolution to all length-m sub-series from a historical ground-based irradiance measurement time series (aggregated to 1 h resolution). Since the best-matched hourly sub-series has a corresponding 15 min series, this high-resolution time series is used as the final forecasts. This circumvents the need to synthetically generate the high-frequency forecasts. Fig. 1 illustrates this procedure. In addition, if multiple good matches can be found, this group of high-resolution time series can be used to construct an ensemble, and thus generate probabilistic forecasts, which is another desirable forecast property (van der Meer et al., 2018).

This proposed algorithm has several variations, since the hourly forecasts used for pattern matching can vary, e.g., using the raw NAM forecasts, or using the post-processed NAM forecasts. Hence, to differentiate these variations, the pattern-matching-based algorithm itself is denoted using Pma hereafter, whereas the data input to pattern matching is denoted with an additional version name, e.g., Pma+Nam, if the raw NAM forecasts are used.

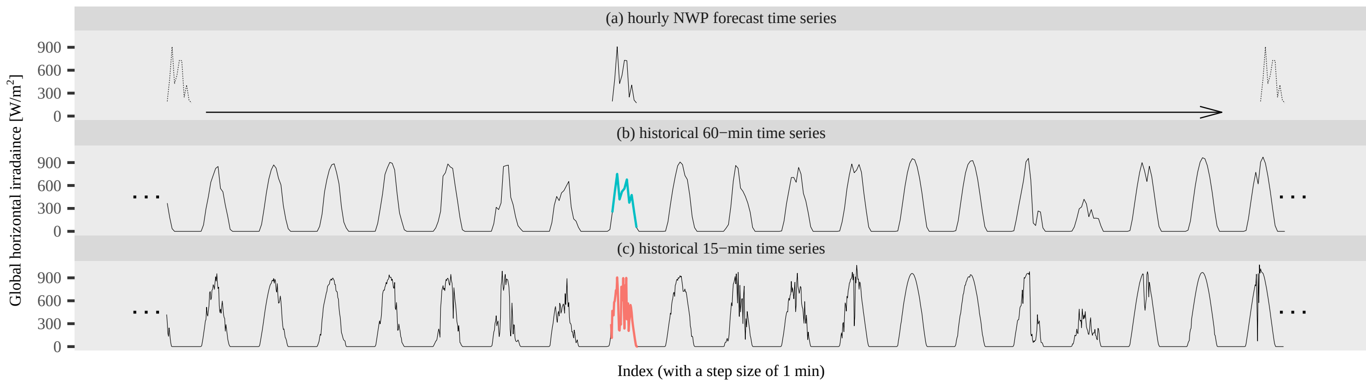

Figure 1: An illustration of the proposed forecasting concept. The short length-m *query* time series (i.e., NWP forecasts with a 1-h resolution) sweeps through the long length-n ($n \gg m$) *history* time series (historical ground measurements aggregated to 1-h resolution), and is compared to each sub-series. After the best match (shown in turquoise) is found, the corresponding high-frequency *history* sub-series (ground measurements aggregated to 15-min resolution, with a length of $4m$, shown in Indian red), is used as the downscaled forecasts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1.3. A brief review on NAM-based forecasting methods

The NAM model operates over the continental United States. The core of the model is based on the non-hydrostatic version of the Weather Research and Forecasting (WRF). The horizontal resolution for NAM is 12 km, and the vertical

[4]There is some major confusion on this issue in the literature, especially when Kalman filtering, an algorithm that adjust the forecasts sequentially, is involved. For example, in Diagne et al. (2014), although the paper appears to describe a day-ahead forecasting scenario, when hourly Kalman filtering was used, the forecasting is in fact "hour-ahead".

coordinate includes 60 hybrid sigma-level terrain-following grids. The NAM forecast is run four times a day at 0:00, 6:00, 12:00, and 18:00 UTC. The output is available hourly out to 36 h then 3-hourly from 36 to 84 h. GHI is computed using the geophysical fluid dynamics laboratory short wave (GFDL-SW) (Wang, 1976) radiation transfer model (RTM). Changes in GHI are based on the weather conditions in each atmospheric column because GFDL-SW is an one-dimensional model. While the spatial and temporal resolution are not as high as some of the other operational weather models—e.g., 3 km horizontal resolution for the High-Resolution Rapid Refresh (HRRR), or hourly-update in the Rapid Refresh (RAP)—the NAM has a consistency advantage than the HRRR and the RAP forecast because the latter constantly undergo major updates. This means that the errors in the NAM are more consistent over the years and could be corrected for in a simpler way.

The NAM has been used extensively in solar forecasting, whether as the initial and boundary condition for a higher resolution mesoscale model (Mathiesen et al., 2013), as a member of a blended ensemble forecast (Perez et al., 2014), or as an input to utilize machine learning techniques for improved accuracy (Lu et al., 2015). It is shown that with some post-processing, solar forecast utilizing NAM can achieve higher accuracy. To this end, techniques to improve NAM forecast accuracy will be described in more details in Section 3.

1.4. A brief review on pattern-matching-based forecasting methods

The pattern-matching-based method is not a new concept in weather forecasting. It can be traced to at least 1969, when Lorenz coined the term *analogs*, for two or more states of the atmosphere that resemble each other (Lorenz, 1969). In the recent years, the method is regaining popularity in solar forecasting, largely due to the increasing amount of ground-based measurements and satellite-derived irradiance data. Since many solar forecasting papers of this kind adopt very primitive[5] ways of pattern matching (e.g., Akarslan and Hocaoglu, 2017; Wang et al., 2017), only several representative and innovative works are reviewed here.

In Alessandrini et al. (2015), one of the earliest pattern-matching-based solar forecasting papers, analog ensemble (AnEn) is used to forecast the PV output of three plants in Italy, for a forecast horizon of 0–72 h. The particular matching strategy used in the paper is performed over five NWP output parameters, namely, GHI, total cloud cover, air temperature, solar azimuth and elevation angles. More specifically, the similarity between the current forecast, F_t, and an analog, A_t, is given by:

$$||F_t, A_t|| = \sum_{i=1}^{5} w^{(i)} \sqrt{\sum_{j=1}^{3} \left(F^{(i)}_{t+j-2} - A^{(i)}_{t+j-2}\right)^2}, \tag{1}$$

where i is indexing the 5 weather variables, j is indexing the time around t, and $w^{(i)}$ are the weights of the weather variables, which need to be trained from data. To construct the AnEn, 20 analogs are used. AnEn is compared to quantile regression (QR) and persistence ensemble (PeEn). It was found that AnEn is similar to QR, and both methods outperforms PeEn. It is worth noting that PeEn is a commonly used benchmarking model for probabilistic solar forecasting. Although there are several variants to it, the particular form that was used in Alessandrini et al. (2015) is given by:

$$\text{PeEn} = \{\text{GHI}_{t-24\times i} : i = 1, \dots, 20\}. \tag{2}$$

In other words, PeEn is made of the most recent available 20 measured GHI values at the same hour.

Using Alessandrini et al. (2015) as a foundation, the same group of researchers later extended their work in two directions: (1) combining artificial neural network (ANN) with AnEn; and (2) analyzing and evaluating the computational efficiency of the methodology (Cervone et al., 2017). In their new paper, ANN-based regression models are used to generate deterministic forecasts based on the NWP output. Subsequently, the 5-parameter AnEn model is modified to a 6-parameter AnEn model, with the ANN forecast as the 6$^{\text{th}}$ parameter; in other words, the ANN post-processed NWP output is included in the ensemble. Including the post-processed NWP forecasts into the AnEn, the AnEn performance improves. Aside from the ANN–AnEn hybrid modeling, a computation speed analysis is also

[5]The word “primitive” refers to several things: (1) the matching is based on brute-force search algorithms, (2) only a single match is considered, i.e., point forecasting, (3) the query length is arbitrarily chosen without proper motivation and analysis.

conducted (Cervone et al., 2017). It was found that Eq. (1) contributes 84% of the computational time, whereas the analog sorting and selection only contributes 16%.

Whereas Alessandrini et al. (2015); Cervone et al. (2017) used NWP output data and the matching was perform in time only, Ayet and Tandeo (2018) demonstrated a similar method on satellite-derived data with spatio-temporal matching. For a given location and time, the analogs are selected using a k-nearest neighbor (kNN) algorithm. The kNN is performed in a 4-dimensional feature space[6] compressed from satellite-derived cloud-index images.

1.5. Contributions of this work

The first and foremost contribution is that this work takes all time parameters involved in RTM operational forecasting into consideration. Such fundamental requirements are typically overlooked, or deemed unimportant, during solar forecasting research. Even though there are thousands of forecasting papers in the literature, it is believed that this work is the *first* one that shows a correct and completely realistic demonstration of intra-day operational solar forecasting. Section 2 elaborates the various time-related considerations in detail. Since these considerations add major difficulties in terms of implementation and design of forecasting experiments, partial data and code[7] are provided as supplementary materials to clarify potential confusions and ambiguities.

The second contribution of this work is a state-of-the-art NWP—time-series ensemble; this is used to improve the day-ahead NAM forecast accuracy. In the literature, NWP forecasts are often adjusted through post-processing techniques such as model output statistics (MOS), Kalman filtering, or machine-learning-based correction. According to Ren et al. (2015), post-processing can be considered as a cooperative ensemble approach. Alternative to the cooperative ensemble, competitive ensemble (e.g., perturbing the NWP initial conditions, or forecast combination) is also frequently used to boost the forecast accuracy. In this regard, this article uses both cooperative and competitive ensembles. More specifically, MOS is used to post-process the raw NWP output, whereas seasonal time series models are used as alternatives and thus compete with NWP forecasts through forecast combination. This contribution is described in Section 3 of the article.

Thirdly, the scalability—in terms of computational speed—of the proposed solar forecasting problem is enhanced through adopting a state-of-the-art pattern-matching algorithm. Brute-force searches, i.e., using for-loops to compute Euclidean distances, are ubiquitously used in weather applications. This is no doubt inefficient, and very little has been done algorithmically. Fortunately, there is a large number of fast search algorithms in the field of computer science that are suitable for the present application. Hence, an ultra-fast similarity-search algorithm based on fast Fourier transform (FFT) is used. FFT-based distance calculation is usually used to compute the z-normalized Euclidean distance (Mueen et al., 2017) and this article modifies the FFT distance calculation to allow the fast computation of unnormalized Euclidean distance. The relationship between Euclidean distance computation and FFT is derived mathematically, and the proposed similarity-search algorithm is discussed in Section 4.

Lastly, and most importantly, this article shows empirically that by using Pma, the accuracy of intra-day forecasting highly correlates with that of the day-ahead NWP forecasts. In other words, improvements in day-ahead NWP forecasts carry through to boost performance in 15-min 6–8-h-ahead forecasts. This suggests that future research in solar forecasting should focus on improving the NWP forecasts, the remaining tasks, namely, downscaling, creating ensemble, generating deterministic and probabilistic forecasts for the RTM, can be handled by Pma with decent accuracies.

Besides the above-mentioned sections, the remaining part of the article is as follows. Section 5 presents a case study to demonstrate the proposed operational forecasting algorithm in detail. Both deterministic and probabilistic forecasting results are presented with a suite of evaluation metrics. Section 6 discusses advantages, disadvantages, as well as several possible variations to the proposed algorithm. Conclusions follow at the end.

2. Forecasting requirements in CAISO RTM, design of case study, forecasting models, and evaluation metrics

The CAISO real-time market has three major scheduling processes, namely, real-time unit commitment (RTUC), short-term unit commitment (STUC), and real-time economic dispatch (RTED) (Makarov et al., 2011). In all of these

[6]These 4 features are: cloud fraction, cloud spread, clear sky intensity, and cloud intensity.

[7]The complete data and code is over 1.5 GB, and can be obtained from the corresponding author.

Table 1: An illustration of hourly rolling 5-h-ahead forecasting. A total of twenty 15-min forecasts spanning the next 5 h are generated 75 min prior to each operating hour.

submission time	operating hour	forecast timestamps +1 h	+2 h	+3 h	+4 h	+5 h
⋮	⋮			⋮		
07:45	09:00	09:15	10:15	11:15	12:15	13:15
		09:30	10:30	11:30	12:30	13:30
		09:45	10:45	11:45	12:45	13:45
		10:00	11:00	12:00	13:00	14:00
08:45	10:00	10:15	11:15	12:15	13:15	14:15
		10:30	11:30	12:30	13:30	14:30
		10:45	11:45	12:45	13:45	14:45
		11:00	12:00	13:00	14:00	15:00
09:45	11:00	11:15	12:15	13:15	14:15	15:15
		11:30	12:30	13:30	14:30	15:30
		11:45	12:45	13:45	14:45	15:45
		12:00	13:00	14:00	15:00	16:00
⋮	⋮			⋮		

operations, four time parameters are involved: (1) forecast horizon, the time span that the forecasts need to cover; (2) forecast resolution, the time interval of each submitted forecast; (3) forecast lead time, the time needed prior to the operating hour or day; and (4) forecast update rate, the frequency for the forecasts to be refreshed. A quadruplet can be used to denote these time parameters, i.e., $(\mathcal{H}, \mathcal{R}, \mathcal{L}, \mathcal{U})$ denote forecast **h**orizon, **r**esolution, **l**ead time and **u**pdate rate, respectively. For example, the submission requirement for RTED is $(\mathcal{H}^{65\text{min}}, \mathcal{R}^{5\text{min}}, \mathcal{L}^{7.5\text{min}}, \mathcal{U}^{5\text{min}})$ (Makarov et al., 2011). In other words, a total of thirteen 5-min forecasts need to be submitted 7.5 min prior to the operating hour, this process repeats every 5 min. For STUC, the submission requirement is $(\mathcal{H}^{5\text{h}}, \mathcal{R}^{15\text{min}}, \mathcal{L}^{75\text{min}}, \mathcal{U}^{1\text{h}})$, or twenty 15-min forecasts need to be submitted 75 min prior to the operating hour, and the process repeats every hour (Makarov et al., 2011).

In view of the above requirements, a timeline can be drawn to illustrate the CAISO's requirement for STUC $(\mathcal{H}^{5\text{h}}, \mathcal{R}^{15\text{min}}, \mathcal{L}^{75\text{min}}, \mathcal{U}^{1\text{h}})$, which is the target of this article. Fig. 2 depicts an example timeline, assuming the operating hour starts at 9:00 on an arbitrary day. Based on Fig. 2, the forecasting case study can be designed. Firstly, for each forecast submission, forecasts over a 5-h period, with a 15-min resolution, are generated. Although a 75-min lead time is needed, during actual operation, any lead time longer than that is acceptable. Since the NWP forecasts have an hourly resolution, this article extends the lead time to 2 h.[8] Next, given the forecast update rate of 1 h, the 5-h-ahead forecast needs to be updated every hour. This process is exemplified in Table 1. It is noted that a *complete* forecast time series (columns in Table 1) can be formed for each forecast horizon, ranging from 1- to 5-h-ahead. Hence, the forecast evaluation is performed for each hourly forecast horizon as exemplified in Fig. 3.

2.1. Models for deterministic forecasting

This article considers three methods: (1) clear-sky persistence, (2) the family of seasonal auto-regressive integrated moving average (SARIMA) models, and (3) the proposed Pma, for deterministic forecasting.

2.1.1. Clear-sky persistence

The persistence model takes the most recent available measurement as the forecast. The performance of this raw persistence model can be improved by considering the diurnal cycle in the solar irradiance, namely, the clear-sky expectation. The *clear-sky* persistence model assumes the forecast clear-sky index (CSI) is equal to the most recent available CSI measurement. The forecast CSI is then adjusted using the current clear-sky expectation. Given the

[8]Since the lead time of NWP forecasting accuracy only depends weakly on lead time, a longer lead time does not complicate the time consideration here.

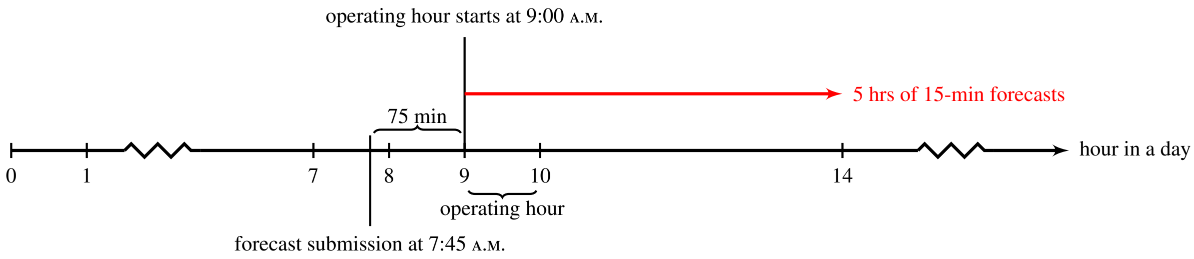

Figure 2: Real-time market operation in CAISO. For an operating hour starting at 9:00 A.M., 5 hours of 15-min forecasts need to be submitted at 7:45 A.M., i.e., 75 min prior to the operating hour.

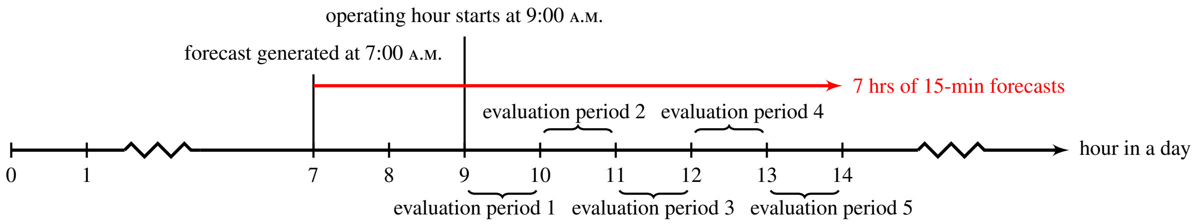

Figure 3: Forecast evaluation design in this article. For each operating hour, 7 hours of 15-min forecasts are generated 2 h prior to the operating hour. The forecasts are evaluated over five hourly periods, separately.

time parameters $(\mathcal{H}^{5\text{h}}, \mathcal{R}^{15\text{min}}, \mathcal{L}^{75\text{min}}, \mathcal{U}^{1\text{h}})$, the clear-sky persistence model used in this article takes the single most recent *non-zero* CSI value prior to the submission deadline as the forecast CSI across the entire forecast horizon. For example, for an operating hour starts at 9:00, the CSI value at 7:45 (if it is a non-zero value), will be used for 9:15, 9:30, . . . , 13:45, 14:00 (all 20 timestamps). This model is denoted as PERS.

2.1.2. Multi-step-ahead time series model

Most time series models, such as autoregressive integrated moving average model (ARIMA) or exponential smoothing state space mode, have the capability of modeling the seasonal component, in this case, the diurnal cycle. In many recent studies, various time series models have been compared, and their performance are mostly similar (Yang and Dong, 2018; Yang et al., 2015b). To that end, seasonal ARIMA, or SARIMA, is used to represent multi-step-ahead time series models.

In the present case, the SARIMA model is used to generate 25-step-ahead forecasts using 15-min ground data, covering the 5-h horizon with a lead time of 75 min. The training length is set to be 5 days (a length-480 time series) prior to the submission deadline. The process order and model parameters of the SARIMA model are re-trained every hour to comply with the rolling forecast submission required by the RTM.

The above SARIMA model has a seasonal period of 96, i.e., number of 15-min data points in a day. The high seasonal frequency causes the parameter estimation to be time consuming, and it requires a large amount of memory. Although this should not pose any problem during the actual operational forecasting, speeding up the run time is nevertheless desired. In this regards, based on a discussion by Rob Hyndman,[9] a Fourier series de-seasonality approach is used:

$$y_t = \text{const.} + \sum_{k=1}^{K} \left[\alpha_k \sin\left(\frac{2\pi kt}{96}\right) + \beta_k \cos\left(\frac{2\pi kt}{96}\right) \right] + N_t, \tag{3}$$

where y_t is the GHI time series, and N_t is an ARIMA process. The value of K is chosen to be 3 since the unimodal diurnal cycle do not require a large K. For each N_t model, the Akaike information criterion is used for model selection

[9] Rob Hyndman is the main author of the famous *forecast* package in R. See, `https://robjhyndman.com/hyndsight/longseasonality/` for his discussion on long seasonal period.

with an ARIMA process order up to ($p = 3$, $d = 0$, $q = 3$), where p, d, and q are the orders for the autoregressive, differencing, and moving average parts, respectively. This model is referred to as Sarima hereafter.

2.1.3. Pma

The previous two benchmarking methods operate on 15-min data directly, whereas Pma first generates forecasts with an hourly resolution and then downscales them to a 15-min resolution. In this regard, three variations are used to exemplify the procedure.

The first model uses the raw NAM forecasts without any correction. For each operating hour, 8 hourly forecasts are used as query for pattern matching. For example, if the operating hour starts at 9:00, NAM forecasts for 7:00, 8:00, . . . , 14:00 are used, see Fig. 3. These 8 numbers are compared to all length-8 sub-series in the hourly historical measured data, through Pma. After the best-matched sub-series (in terms of unnormalized Euclidean distance) is found, the corresponding 15-min measurements from the same historical period are used as the final forecasts. However, it should be noted that length-8 hourly sub-series corresponds to a length-32 15-min series. Therefore, only those 20 data points relevant to the 5-h-ahead forecasting are recorded. This process repeats every hour, so that the forecasts can be evaluated based on the evaluation periods, see Fig. 3.

The second model has the exact same setup as the first one, except that the NAM forecasts are corrected and ensembled prior to the pattern matching. This is to investigate whether improved hourly forecasts can lead to better 15-min forecasts. Of course, this is likely to be the case, therefore, a more relevant question is: how much of the hourly forecast improvements can be carried to the 15-min forecasts? As mentioned earlier, both cooperative (MOS correction) and competitive (time series) ensembles will be used to improve the raw NAM forecasts.

The last model is designed to study the extreme case of having perfect hourly forecasts. Since both the NWP forecasting step and the downscaling step contribute to the final error, isolating the downscaling error is of interest. By assuming the hourly NWP forecasts are 100% accurate, i.e., the hourly measurements from the forecast hours are used directly, any remaining error solely comes from the downscaling step. This type of models is usually called "oracle model" in forecasting works (Yang and Dong, 2018). In what follows, these three models are denoted as Pma+Nam, Pma+Ens, and Pma+Oracle, respectively.

2.2. *Models for probabilistic forecasting*

Since all three above-mentioned deterministic forecasting methods can be extended to probabilistic forecasting, the probabilistic forecasting portion of the article adopts the same three methods.

2.2.1. *Clear-sky persistence ensemble*

Whereas Pers discussed in Section 2.1.1 considers the most recent available CSI values as forecast CSI, the clear-sky PeEn takes the CSI values recorded at N most recent *non-zero* 15-min timestamps to create an ensemble. Following Alessandrini et al. (2015), the value of N is set to 20 in this article. For example, consider the forecasting scenario depicted previously: instead of only assigning CSI at 7:45 to 9:15, 9:30, . . . , 13:45, 14:00, 20 CSI values are assigned to each of these 20 timestamps. More explicitly, suppose the daylight hour starts at 7:00 and ends at 19:00, these 20 CSI values come from: today 7:45, . . . , 7:00, and yesterday 19:00, 18:45, . . . , 15:30, 15:15.

2.2.2. *SARIMA with normal prediction interval*

In a previous contribution by Yang (2017), it has been shown that by fitting a SARIMA model to hourly irradiance time series, the residual follows a normal distribution—as least for the case of the experimental data therein used. Hence, normal prediction interval is assumed in this work. More specifically, if the standard deviation of an h-step-ahead forecast, $\hat{\sigma}_h$, is known or can be estimated, the prediction interval can be formed. Mathematically, the intervals are given as:

$$\left(\hat{y}^U_{t+h}, \hat{y}^L_{t+h}\right) = \left(\hat{y}_{t+h} + c\hat{\sigma}_h, \hat{y}_{t+h} - c\hat{\sigma}_h\right), \tag{4}$$

where the multiplier c depends on the coverage probability, e.g., $c = 1.96$ for the 95% prediction interval. However, the estimation of σ_h is not always straightforward, especially for $h > 1$. For different time series models, the closed-forms of $\hat{\sigma}_h$ are also different; sometimes, the closed-form is not available and an approximation needs to be used. In this article, the most well-developed forecasting toolbox (Hyndman et al., 2018) is used, and the σ_h estimates of the SARIMA models are readily available.

2.2.3. Pma with multiple analogs that form an ensemble

As compared to the previous two methods, it is much easier to form ensembles using Pma. Based on a given *query*, instead of finding and recording one analog, the top N analogs can be recorded. The ranking of analogs is based on the unnormalized Euclidean distance. The value of N is again taken to be 20 in this article.

2.3. Evaluation metrics

2.3.1. Metrics for deterministic forecasts

Three metrics are used throughout the article to evaluate the deterministic forecasts made by various models, namely, the normalized mean bias error (nMBE), normalized root-mean-square error (nRMSE), and forecast skill. Whereas nMBE is used to access the systematic bias in the forecasts, nRMSE is used to access whether the forecasts contain large errors. Finally, forecast skill is used to determine the improvement of each model over the reference model, in this case, the clear-sky persistence. These metrics are given as:

$$\text{nMBE} = \frac{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)}{\frac{1}{n}\sum_{t=1}^{n} y_t} \times 100, \tag{5}$$

$$\text{nRMSE} = \sqrt{\frac{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{\frac{1}{n}\sum_{t=1}^{n} y_t^2}} \times 100, \tag{6}$$

$$s = \left(1 - \frac{\text{nRMSE}_{\text{model}}}{\text{nRMSE}_{\text{reference}}}\right) \times 100, \tag{7}$$

where $\hat{y}_t$ and y_t are the forecast and measurement at time t. All three metrics are expressed in percentage. It should be noted that another frequently used way to compute nRMSE is $\frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}}{\frac{1}{n}\sum_{t=1}^{n} y_t} \times 100$. However, this different formulation of nRMSE does not change the forecast skill.

2.3.2. Metrics for probabilistic forecasts

To evaluate the probabilistic forecasts, the Brier score (BS), continuous ranked probability score (CRPS), and CRPS skill score are used. The Brier score is given by:

$$\text{BS} = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}(p_{ti} - o_{ti})^2, \tag{8}$$

where p_{ti} is the probability that the forecast at time t falls in category i, and o_{ti} takes the value of 0 or 1 according to whether or not the event occurred in category i. In this article, a bin width of 100 W/m^2 is used. In this way, a total of 14 bins are formed for irradiance ranging from 0 to 1400 W/m^2, i.e., $m = 14$ in Eq. (8).

The CRPS is given by:

$$\text{CRPS} = \frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\infty}\left(F^{\hat{y}_t}(x) - \mathbf{1}(x - y_t)\right)^2 dx, \tag{9}$$

where $F^{\hat{y}_t}$ is the CDF of the forecast $\hat{y}_t$ and $\mathbf{1}(x - y_t)$ is the Heaviside step function shifted to y_t.

Lastly, the CRPS skill score is given by:

$$s = \left(1 - \frac{\text{CRPS}_{\text{model}}}{\text{CRPS}_{\text{reference}}}\right) \times 100. \tag{10}$$

The clear-sky PeEn is used as the reference model to evaluate the CRPS skill score of the probabilistic forecasts. The Brier skill score (BSS) could also be used instead of the CRPS. However, since BS depends on the number of defined classes, so does the BSS. This allows one to tune the score, which is undesirable.

3. Data

Two sets of data are involved in the empirical part of this article. For the ground-based measurements, 20 years (1998–2017) of research-grade data from a SURFRAD station is used, whereas for the NWP data, 2 years (2016–2017) of hourly NAM forecasts are considered.

3.1. SURFRAD data

The surface radiation budget network (SURFRAD) was established in 1993 by the National Oceanic and Atmospheric Administration to collect long-term high-resolution radiation measurements and support climate research. There are a total of 7 stations. Whereas the results for all stations are provided in Appendix C, the algorithm performance is demonstrated in details at the station Desert Rock (DRA), Nevada, due to its geographical proximity to California. While DRA is not in California, it is close to several solar power plants that are outside California yet deliver their energy to CAISO. DRA started collecting data in March 1998, and only GHI data is of interest here. Prior to 2009, the station collected 3-min data; since 2009-01-01, 1-min data have been collected. Ground data needs to be quality checked and averaged. The original SURFRAD quality control (QC) is basic, and the primary goal of this QC is to eliminate physically impossible GHI values. Even though more advanced and stricter QC sequences exist, for forecasting applications, the original QC should suffice.

The 1-min SURFRAD data is first aggregated to the nearest 15-min timestamp using the *ceiling* operator; this data frame is referred to as SURFRAD15 hereafter. Next, to match the "snapshot" nature of the NAM data, SURFRAD15 is aggregated to hourly data using the *round* operator, e.g., 11:45, 12:00, 12:15, and 12:30 are averaged to the 12:00 timestamp. This is equivalent to averaging 1-min SURFRAD data from 11:31 to 12:30. The resultant hourly data frame is denoted with SURFRAD60. A graphical representation of this averaging scheme is shown in Table 2. A similar scheme is used for 3-min data. It is noted that data aggregation is a processing issue that is constantly being overlooked. Due to the diurnal cycle of GHI, one should be careful in aligning the timestamps of different datasets. Miss-aligned datasets can cause higher errors; this is typified by the discussion in Yang (2018a).

After the first aggregation, SURFRAD15 has a total of 694,176 of 15-min records, for which 1.1% are missing. This rather small percentage of missing values are replaced with their corresponding clear-sky expectations, calculated via the Ineichen–Perez model. Subsequently, SURFRAD15 is aggregated to SURFRAD60, which has a total of 173,544 records.

3.2. NAM data

GHI computed from the NAM forecast is used for this work. As briefly described in Section 1.3, changes in GHI are based on the weather conditions in each atmospheric column. Variables such as solar zenith angle, clouds, aerosols, and water vapor concentration all contribute to changes in GHI. Of particular importance is cloud optical thickness, which is parameterized based on prognostic variables such as liquid and ice water mixing ratio, cloud temperature, and pressure (Stephens, 1978). Additionally, NAM uses climatological tables for aerosols (GFDL Global Atmospheric Model Development Team, 2004), often resulting in a systematic clear sky bias from the ground observation. The following section describes ways to account for these biases.

NAM is run 4 times per day, starting from 00:00, 06:00, 12:00, and 18:00 UTC. In this work, the 12–35 hours-ahead forecasts generated by the 12:00 runs are used.[10] For example, for the NAM run starts at 2015-12-31 12:00, 24 point forecasts for timestamps 2016-01-01 00:00, . . . , 2016-01-01 23:00 are saved. The next run starts at 2016-01-01 12:00, and the forecasts span 2016-01-02 00:00, . . . , 2016-01-02 23:00. This procedure repeats until the forecasts over 2017-12-31 00:00, . . . , 2017-12-31 23:00 are generated. As a result, two full years of NAM 12-h-ahead forecasts are obtained. Fig. 4 plots a one-day time series plot of SURFRAD and NAM data. The two data sources show good temporal alignment.

[10]The CAISO STUC requires an hourly rolling update rate. Since the NWP forecast accuracy does not degrade with forecast horizon for the first 24 to 48 hours (Perez et al., 2013), these 24-h-rolling NAM forecasts do not affect the analyses below. Furthermore, starting 2017-02-01, the NAM output has been archived hourly, which could be used for actual operational forecasting.

Table 2: The data averaging scheme used in this article.

$$\begin{array}{ccc}
\hline
\text{Time} & \textsc{Surfrad}15 & \textsc{Surfrad}60 \\
\hline
\vdots & \vdots & \vdots \\
\left.\begin{array}{c}11{:}31\\ \vdots \\ 11{:}45\end{array}\right\} & 11{:}45 & \\
\left.\begin{array}{c}11{:}46\\ \vdots \\ 12{:}00\end{array}\right\} & 12{:}00 & \\
\left.\begin{array}{c}12{:}01\\ \vdots \\ 12{:}15\end{array}\right\} & 12{:}15 & \\
\left.\begin{array}{c}12{:}16\\ \vdots \\ 12{:}30\end{array}\right\} & 12{:}30 & \left.\right\} 12{:}00 \text{ (11:31–12:30)} \\
\vdots & \vdots & \vdots \\
\hline
\end{array}$$

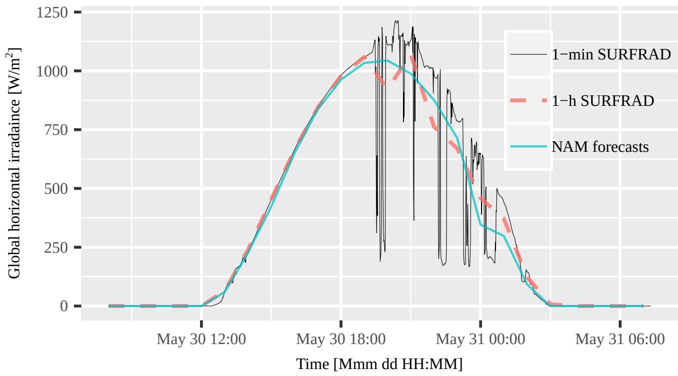

Figure 4: A one-day time series plot of SURFRAD and NAM data.

3.3. Improving the NAM forecast accuracy

Using the previously discussed time-parameter notation, the raw NAM forecasts can be denoted using Nam with $(\mathcal{H}^{24\text{h}}, \mathcal{R}^{1\text{h}}, \mathcal{L}^{12\text{h}}, \mathcal{U}^{24\text{h}})$. By comparing Nam to Surfrad60, a nRMSE of 18.91% is observed. The corresponding day-ahead persistence model results in a 25.68% nRMSE. Although there is a positive forecast skill, it is known, *a priori*, that more accurate day-ahead hourly forecasts will lead to more accurate intra-day 15-min forecasts, i.e., the error in Nam will propagate to the pattern-matching step later. To that end, time series ensembles Yang and Dong (2018) are used to improve the accuracy of Nam. Before the ensemble methods are elaborated, the component models are described below.

3.3.1. Component model 1: MOS-corrected NAM

MOS is perhaps the most well-accepted way of post-processing the NWP forecasts. The choice of MOS herein used follows Mathiesen and Kleissl (2011); Lorenz et al. (2009), namely, the bias correction through a fourth-degree polynomial:

$$\text{bias}_t = a_1 \cos^4 Z_t + a_2 \hat{k}_t^4 + a_3 \cos^3 Z_t + \cdots + a_8 \hat{k}_t, \tag{11}$$

where Z_t is the zenith angle at time t, and $\hat{k}_t$ is the forecast clear-sky index at time t. Using this equation, the model-led bias of a new forecast can be estimated once the regression coefficients are obtained. The regression coefficients are fitted by season and by year. More specifically, the coefficients fitted using data from 2016 January to March are used to correct the NAM forecasts from 2017 January to March. This procedure is applied to other quarters of the year. Similarly, the coefficients fitted using data from 2017 are used to correct the NAM forecasts from 2016. Through this cross validation, true out-of-sample MOS can be applied to all data points. This correction leads to a smaller nRMSE of 17.47%.

3.3.2. Component model 2: The family of seasonal ETS models

The family of exponential smoothing (ETS) models contains a total of 30 models, among which 20 are seasonal models. These models have been extensively studied for solar forecasting applications (Yang and Dong, 2018; Yang et al., 2015a; Dong et al., 2013). The R package "forecast" (Hyndman et al., 2018) is herein used to perform ETS forecasting. To align with Nam, a 12-h lead time is considered. Following Yang and Dong (2018), the training period is set to be 14 days. For example, to generate the forecasts for 2016-01-01 00:00, . . . , 2016-01-01 23:00, Surfrad60 data from 2015-12-17 12:00 to 2015-12-31 11:00 (336 hourly data points) are used. Given $\mathcal{U}^{24\text{h}}$, the ETS model selection and parameter estimation is performed every 24 h, and the Akaike information criterion is used in model selection. Since ETS is a time series method, it does not consider any physical evolution of the atmosphere. Hence the nRMSE is 20.39%, which is worse than Nam but better than persistence.

3.3.3. Component model 3: STL decomposition

The number of parameters in a SARIMA or ETS model is quite large. To reduce the computational burden, data-driven decomposition method is often used. The seasonal and trend decomposition using loess (STL) is a mature procedure rooted in time series forecasting. In solar engineering, it has been shown to be useful in separating the variable solar time series component from the clear-sky component (Yang, 2017; Yang et al., 2012). Therefore, STL decomposition is used as a component model in this article. The time series setup of STL decomposition follows the ETS setting exactly. Its nRMSE is 20.50%, which is similar to ETS, but with an improved computational speed.

3.3.4. Component model 4: TBATS

The abbreviation "TBTAS" is constructed using the initials of five phrases, namely, trigonometric, Box–Cox transform, ARMA errors, trend, and seasonal, that jointly describe the nature of the model. TBATS is evolved from the linear version of the Holt–Winter additive seasonal exponential smoothing:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \tag{12a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \tag{12b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{12c}$$

$$s_t = s_{t-m} + \gamma\varepsilon_t, \tag{12d}$$

where ε is the white noise; m is the period of the seasonal cycle; ℓ, b and s represent the level, growth and seasonal components of the time series $\{y_t\}$; and α, β and γ are the smoothing parameters to be fitted. TBATS improves over the Holt–Winter model in several aspects. Firstly, it uses a Box–Cox transformed time series instead of the original time series, which may be non-stationary. TBATS also models the error component, i.e., ε_t in Eq. (12), with an ARMA process:

$$\varepsilon_t = \sum_{i=1}^{p} \varphi_i \varepsilon_{t-i} + \sum_{i=1}^{q} \theta_i a_{t-1} + a_t. \tag{13}$$

Lastly, TBATS has the capability of modeling multiple seasonal components with different cycles. For the i^{th} seasonal component, $s_t^{(i)}$, the trigonometric representation is given by:

$$
\begin{aligned}
s_t^{(i)} &= \sum_{j=1}^{k_i} s_{j,t}^{(i)}, && (14a)\\
s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} \varepsilon_t, && (14b)\\
s_{j,t}^{*(i)} &= -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} \varepsilon_t, && (14c)\\
\lambda_j^{(i)} &= 2\pi j/m_i, && (14d)
\end{aligned}
$$

where k_i is the number of harmonics required for the ith seasonal component; $s_{j,t}^{(i)}$ and $s_{j,t}^{*(i)}$ are the stochastic level and growth of the ith seasonal component. Owing to its elaborate modeling procedure, TBATS has previously been shown to outperform most time series models (Yang and Dong, 2018). For the present dataset, it leads to an nRMSE of 20.11%, which is the smallest among the three time series models.

3.3.5. Time series ensemble models

The reason for having ensembles is to reduce the data, parameter, and modeling uncertainties. In the present case, the same datasets are used for the component models, and there is no parameter perturbation involved. Hence, the ensemble mainly contributes in terms of reducing modeling uncertainty. The results from the five component models, namely, uncorrected NAM, MOS, ETS, STL, and TBATS, are used to generate ensembles. The forecast-generating mechanisms of these component models are different, which is a common prerequisite for the ensembles to be effective, i.e., to prevent underdispersed ensembles.

The choice of ensemble methods employed in this article follows Yang and Dong (2018), in which several regression-based combination methods were introduced. In a companion paper, the exact methods have been extended to spatial prediction problems (Yang, 2018b). Both works showed that by combining predictions, the risk of forecast busts can be reduced.

The first ensemble is constructed through simple averaging; it is denoted as Avg. Given the forecasts made for time t using the i^{th} component model, $\hat{y}_t^{(i)}$, where $i = 1, \ldots, 5$, the final ensemble forecast is simply:

$$
\hat{y}_t = \frac{1}{5} \sum_{i=1}^{5} \hat{y}_t^{(i)}. \tag{15}
$$

This approach does not require any training, and each component forecast has the same contribution to the final forecast. Since some of the component models are more accurate than others, it is logical to assign a larger weight to a model accurate model. One of the intuitive ways of weight assignment is by considering the mean squared error (MSE):

$$
\hat{y}_t = \sum_{i=1}^{5} \frac{\frac{1}{\text{MSE}_i}}{\sum_{i=1}^{5} \frac{1}{\text{MSE}_i}} \hat{y}_t^{(i)}, \tag{16}
$$

where MSE_i is the observed MSE for the i^{th} component model. This method is referred to as Var, i.e., averaging through variance-based weighting. Besides Var, regressions can be used to estimate the combining weights:

$$
\hat{y}_t = \sum_{i=1}^{5} \hat{\beta}^{(i)} \hat{y}_t^{(i)} + \hat{\beta}_0. \tag{17}
$$

In this setting, the regressand is the observed GHI, and the regressors are the forecasts made using the component models. The regression parameters, $\hat{\beta}_0$ and $\hat{\beta}^{(i)}$, can be estimated using any regression technique. Ordinary least squares, least absolute deviations, and lasso are used to exemplify this class of methods; they are denoted with Ols, Lad, and Lasso, respectively. The reader is referred to Yang (2018b); Yang and Dong (2018) for the details of the regression-based ensemble construction.

Aside from Avg, the other ensemble schemes require training the weights. On this point, the cross validation procedure used earlier for MOS is applied here. In other words, for each quarter in each year, the weights are estimated using data from the same quarter in the other year. The nRMSEs for Avg, Var, Ols, Lad, and Lasso are 17.61%, 17.18%, 16.74%, 17.10%, and 16.81%, respectively. The scatter plots of all the forecasts described in this section are shown in Fig. 5. As compared to Nam, the ensemble models are effective in reducing the number of severely underpredicted cases (i.e., fewer blue points below the identity line).

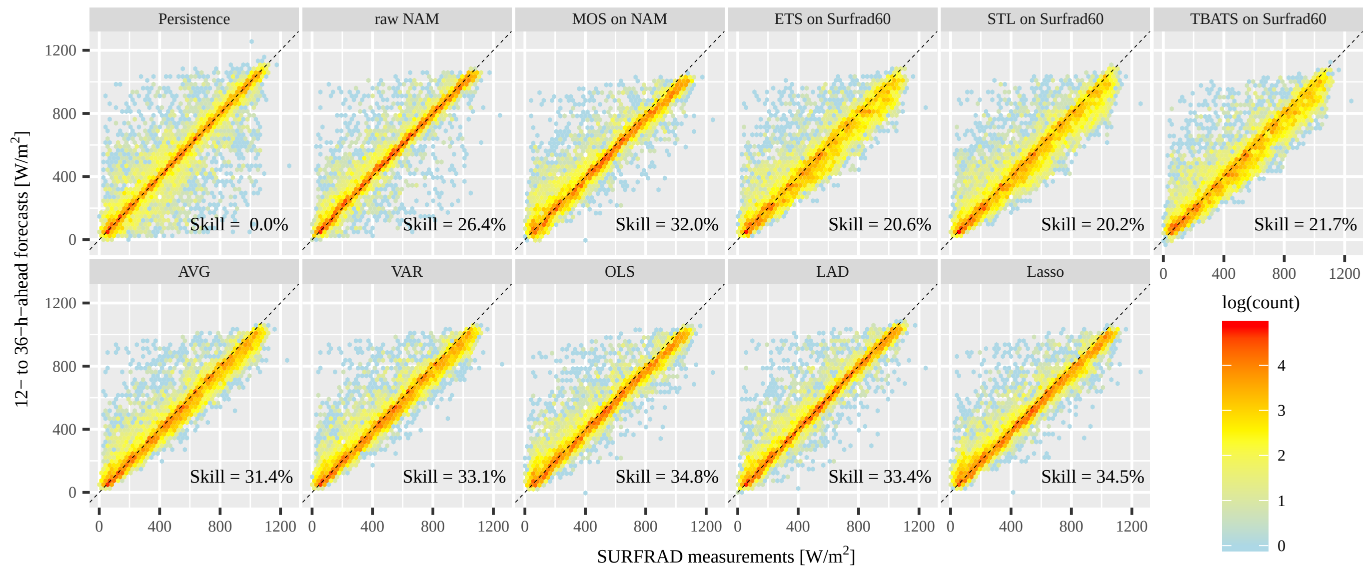

Figure 5: The forecast ($\mathcal{H}^{24\text{h}}, \mathcal{R}^{1\text{h}}, \mathcal{L}^{12\text{h}}, \mathcal{U}^{24\text{h}}$) versus measured GHI at Desert Rock ($-116.02°$, $36.62°$). The component models are arranged in the top row, whereas the ensembles are in the bottom row. Hexagon binning is used for visualization. For a higher contrast, the color scheme is based on the logarithm of bin frequency.

Based on this posterior observation, Ols forecasts are used hereafter as queries for pattern-matching, i.e., the hourly forecasts used in Pma+Ens comes from Ols. However, it should be noted that in a real-time environment, the best ensemble model might be unknown to the forecasters. Nevertheless, in most cases, the ensemble performance dominates that of the component models. Hence, opting for an ensemble model is less risky than choosing any component model alone.

4. An ultra-fast Euclidean distance sweeping algorithm

As mentioned in Section 1, the main step to downscale the hourly forecasts to 15-min forecasts is to perform a similarity search. For that, a similarity metric is required. In contrary to the literature, where z-normalized Euclidean distance is preferred, this article favors the *unnormalized* Euclidean distance. The reason is illustrated with an example. Consider two GHI time series, each with three elements: $\{100, 200, 300\}$ and $\{200, 400, 600\}$ W/m^2. The z-normalized Euclidean distance between these two series is zero. In other words, when the z-normalized Euclidean distance is used, the matching results may be far from the actual irradiance levels. To mitigate this issue, Alessandrini et al. (2015) considered a metric that requires 5 weather variables, recall Eq. (1), among which solar elevation angle and azimuth angle are jointly used to constrain the matching. Nevertheless, one can simply circumvent the above-mentioned issue by using the unnormalized Euclidean distance.

Besides the choice of similarity metric, another issue is the computational time of the search. In weather applications, the computational time for a single Euclidean distance is manageable. However, when the history gets long, or the number of distance computations is large, brute-force computation is no longer feasible. Such scalability issues have been discussed in Cervone et al. (2017), and a super-computer is used in that work. While leveraging strong computational power is one approach, the other approach is to examine the construction of Euclidean distance, and improve the speed in terms of algorithm design. On this point, Mueen's algorithm for similarity search (Mueen et al., 2017) is perhaps the world's fastest similarity search algorithm under Euclidean distance. Notwithstanding, that algorithm is designed for the z-normalized Euclidean distance, and some modifications are required if the unnormalized Euclidean distance is used. The modified algorithm is discussed next.

Given a length-m *query* time series:

$$\boldsymbol{Q} = \{q_1, q_2, \dots, q_m\}, \tag{18}$$

and a length-n *history* time series:

$$\boldsymbol{H} = \{h_1, h_2, \dots, h_n\}, \tag{19}$$

the total number of Euclidean distance to be calculated is $l = n - m + 1$. More specifically, if the sub-series of $\boldsymbol{H}$ from the i^{th} element to j^{th} element is denoted as $\boldsymbol{H}[i : j]$, the first distance is computed between $\boldsymbol{Q}$ and $\boldsymbol{H}[1 : m]$, the second distance is computed between $\boldsymbol{Q}$ and $\boldsymbol{H}[2 : (m + 1)]$, and until the last distance is computed between $\boldsymbol{Q}$ and $\boldsymbol{H}[l : n]$. Mathematically, the distances are given as:

$$\begin{aligned}
d_1(\boldsymbol{H}[1 : m], \boldsymbol{Q}) &= \sqrt{\sum_{i=1}^{m} (h_i - q_i)^2} \\
d_2(\boldsymbol{H}[2 : (m + 1)], \boldsymbol{Q}) &= \sqrt{\sum_{i=1}^{m} (h_{i+1} - q_i)^2} \\
&\vdots \\
d_l(\boldsymbol{H}[l : n], \boldsymbol{Q}) &= \sqrt{\sum_{i=1}^{m} (h_{i+l-1} - q_i)^2}.
\end{aligned} \tag{20}$$

By expanding the summations, Eq. (20) becomes:

$$\begin{aligned}
d_1(\boldsymbol{H}[1 : m], \boldsymbol{Q}) &= \sqrt{\sum_{i=1}^{m} h_i^2 + \sum_{i=1}^{m} q_i^2 - 2\sum_{i=1}^{m} h_i q_i} \\
d_2(\boldsymbol{H}[2 : (m + 1)], \boldsymbol{Q}) &= \sqrt{\sum_{i=1}^{m} h_{i+1}^2 + \sum_{i=1}^{m} q_i^2 - 2\sum_{i=1}^{m} h_{i+1} q_i} \\
&\vdots \\
d_l(\boldsymbol{H}[l : n], \boldsymbol{Q}) &= \sqrt{\sum_{i=1}^{m} h_{i+l-1}^2 + \sum_{i=1}^{m} q_i^2 - 2\sum_{i=1}^{m} h_{i+l-1} q_i}.
\end{aligned} \tag{21}$$

It can be observed that the $\sum_{i=1}^{m} q_i^2$ term does not change for each distance; it only needs to be calculated once. On the other hand, for each subsequent distance, the first summation is only differed by one element, i.e., in d_1, the summation is over $h_1^2, h_2^2, \dots, h_m^2$, whereas in d_2, the summation is over $h_2^2, h_3^2 \dots, h_{m+1}^2$. Based on this characteristic, the first sum-of-squares term can be calculated with a single pass of the *history* time series, i.e., calculated simultaneously when reading the array. Therefore, the only term left to be computed is the last summation term.

To better understand the computational trick, a simpler example is used. Let $n = 5$, $m = 3$, Eqs. (18) and (19) become:

$$\boldsymbol{Q} = \{q_1, q_2, q_3\}, \tag{22}$$
$$\boldsymbol{H} = \{h_1, h_2, h_3, h_4, h_5\}. \tag{23}$$

By reversing $\boldsymbol{Q}$ and padding the result with zeros, i.e.,

$$\boldsymbol{Q}_{\downarrow} = \{q_3, q_2, q_1, 0, 0\}, \tag{24}$$

the convolution between $\boldsymbol{H}$ and $\boldsymbol{Q}_{\downarrow}$ is given by:

$$\boldsymbol{H} \circledast \boldsymbol{Q}_{\downarrow} = \begin{pmatrix} h_1q_3 \\ h_1q_2 + h_2q_3 \\ h_1q_1 + h_2q_2 + h_3q_3 \\ h_2q_1 + h_3q_2 + h_4q_3 \\ h_3q_1 + h_4q_2 + h_5q_3 \\ h_4q_1 + h_5q_2 \\ h_5q_1 \\ 0 \\ 0 \end{pmatrix}^{\top}. \tag{25}$$

It is evident that the third to fifth elements of the convolved vector correspond to the last summation terms in Eq. (21). This ingenious convolution step was proposed in Mueen et al. (2017); however the current algorithm applies convolution to the unnormalized $\boldsymbol{Q}_{\downarrow}$, and above mathematical derivation is distinct from that shown in Mueen et al. (2017). Since the convolution does not require any loop, the algorithm is ultra-fast[11] in terms of sweeping all-pair Euclidean distances. Lastly, it is well-known that convolution in the time domain equals to point-wise multiplication in the frequency domain. The convolution is thus computed via the fast Fourier transform (FFT) and inverse FFT. To summarize the section, the ultra-fast Euclidean distance computation (UFEDC) procedure is depicted in Algorithm 1.

Algorithm 1 Ultra-fast Euclidean distance computation

procedure UFEDC(*history*, *query*)
  $n \leftarrow$ `len`(*history*)
  $m \leftarrow$ `len`(*query*)
  $\boldsymbol{\Sigma} \leftarrow$ `mvss`(*history*) ▹ Moving sum-of-squares
  $\boldsymbol{Q}_{\downarrow} \leftarrow$ `rev`(*query*) ▹ Reverse *query*
  $\boldsymbol{Q}_{\downarrow}[m+1:n] \leftarrow 0$ ▹ Pad the reversed *query* with 0's
  $dots \leftarrow$ `ifft(fft(`*history*`)` $*$ `fft(`$\boldsymbol{Q}_{\downarrow}$`))` ▹ Conv. between *history* and $\boldsymbol{Q}_{\downarrow}$
  $result \leftarrow \texttt{sqrt}\left(\texttt{sum}\left(\boldsymbol{Q}_{\downarrow}^2\right) + \boldsymbol{\Sigma} - 2 * dots[m:n]\right)$ ▹ Eq. (21)
  return *result*
end procedure

5. Empirical study

The empirical validation for $(\mathcal{H}^{5\text{h}}, \mathcal{R}^{15\text{min}}, \mathcal{L}^{75\text{min}}, \mathcal{U}^{1\text{h}})$ using the five models discussed in Section 2 is presented in this section. The validation period spans two full years, namely, 2016 and 2017. The total number of 15-min data is 70,176, i.e., $(365+366)\times24\times4$. After applying a zenith angle filter of $Z < 85°$, 32,642 data points remain. Therefore, the error metrics for *each* evaluation period shown in Table 1 and Fig. 3 are computed over 32,642 forecasts.

To ensure that the forecasts can cover the full 2-year period, Pers and Sarima use a small portion of data from December 2015, so that the first forecasts can fall on 2016-01-01 00:00. On the other hand, for the pattern-matching

[11] A similar algorithm—sweeping using normalized Euclidean distance—is tested again the current implementation in the National Center for Atmospheric Research (R code courtesy of Stefano Alessandrini), the speed of the convolution-based algorithm is approximately two orders of magnitude faster than the default PeEn implementation.

Table 3: Forecast evaluation for deterministic forecasting over a 2-year period. The five evaluation periods correspond to 1–5-h into the operating hour, with a lead time of 75 min and a forecast resolution of 15 min.

Evaluation period	PERS	SARIMA	PMA+NAM	PMA+ENS	PMA+ORACLE
			nMBE [%]		
1	-0.86	0.14	3.69	0.40	0.41
2	-2.07	0.10	3.79	0.56	0.33
3	-3.60	-0.02	3.81	0.62	0.23
4	-5.25	-0.09	3.88	0.60	0.21
5	-6.85	-0.08	3.81	0.26	0.11
			nRMSE [%]		
1	20.24	19.91	20.77	19.04	12.07
2	22.33	21.10	20.71	19.21	12.17
3	24.24	21.70	20.81	19.14	12.18
4	26.26	21.99	20.86	19.15	12.07
5	28.27	22.11	21.00	19.42	12.10
			Forecast skill [%]		
1	0.00	1.63	-2.63	5.91	40.37
2	0.00	5.54	7.28	13.99	45.53
3	0.00	10.48	14.12	21.01	49.74
4	0.00	16.23	20.54	27.06	54.03
5	0.00	21.78	25.72	31.31	57.21

models, the *history* time series is extracted from SURFRAD60; it starts from 1998-03-16 00:00 and ends at 2015-12-31 23:45. Although during the actual operation, the length of *history* increases as more data becomes available, i.e., after 2016-01-01 is forecast, it can be used as part of the *history* to forecast 2016-01-02, this article fixes the length of *history* throughout the empirical study.

5.1. Deterministic forecasting

The results for deterministic forecasting are shown in Table 3. The following observations can be made. In terms of nMBE, only PMA+NAM shows a sizable positive bias, and NWP–time-series ensemble—PMA+ENS in this case—is effective in removing such bias. In terms of nRMSE, PERS and SARIMA show increasing errors as the forecast horizon increases, whereas the PMA models have relatively “flat” errors across the 5 evaluation periods. In terms of forecast skill, all models yield positive skills. Among these models, it is evident that PMA+ENS (besides PMA+ORACLE of course) has the highest skills for all evaluation periods. The performance of PMA+ORACLE reveals that the downscaling step leads to a ≈12% nRMSE, whereas the nRMSE of ENS is about ≈19%. This means the hourly day-ahead forecasting error (recall Section 3, this error is about 17%) and the downscaling error do not stack.

5.2. Probabilistic forecasting

The error metrics of the probabilistic forecasts from the five models are shown in Table 4. Unlike the case of deterministic forecasting, these results are rather disappointing. Besides PMA+ORACLE, all other models have shown worse performance—over one or more evaluation periods—than the baseline model, PeEn, in terms of all metrics. It is now clear that good deterministic forecasting does not guarantee good performance in probabilistic forecasting. In this regard, it confirms the necessity to check both the deterministic and probabilistic performance of a model, in a forecasting study.

To investigate the cause, the probabilistic forecasts over a 7-day period are plotted in Fig. 6. The 95% and 80% prediction intervals are plotted as light and dark gray ribbons. This sequence of days consists of 4 clear days and 3 cloudy days. Quite a number of observations can be made from this simple plot.

Firstly, observations on PeEn are discussed. Given the model assumption (i.e., CSI from 20 most recent 15-min timestamps), the PeEn forecasts rely largely on the variability of the previous hours/day. It is evident from the plot of day 7 that if the previous day is cloudy, and thus has low CSI values, the prediction interval in the morning will be large. This leads to a wide interval width, and thus the coverage of PeEn is quite good. Since the natural bound of probabilistic forecasts is always $\pm\infty$, which ensures 100% coverage rate, good coverage does not imply good forecasts. The interval width is also important.

Table 4: Forecast evaluation for probabilistic forecasting over a 2-year period. The five evaluation periods correspond to 1–5-h into the operating hour, with a lead time of 75 min and a forecast resolution of 15 min. The last column will be discussed in Section 6.1.

Evaluation period	PeEn	SARIMA	PMA+NAM	PMA+ENS	PMA+ORACLE	Interval averaging
				Brier score		
1	0.52	0.63	0.54	0.70	0.30	0.51
2	0.55	0.65	0.54	0.69	0.30	0.52
3	0.56	0.65	0.54	0.69	0.29	0.53
4	0.57	0.66	0.55	0.68	0.29	0.54
5	0.57	0.66	0.55	0.69	0.29	0.55
				CRPS [W/m^2]		
1	47.83	55.87	50.27	54.55	20.69	43.68
2	52.04	59.81	50.31	54.18	20.85	44.95
3	55.24	61.62	50.52	54.08	20.75	46.04
4	57.65	62.57	51.12	54.27	20.44	47.08
5	59.54	63.10	51.72	54.65	20.10	47.91
				CRPS skill score [%]		
1	0.00	-16.81	-5.10	-14.04	56.75	8.68
2	0.00	-14.93	3.32	-4.12	59.93	13.63
3	0.00	-11.56	8.55	2.09	62.44	16.65
4	0.00	-8.54	11.32	5.86	64.55	18.32
5	0.00	-5.98	13.14	8.22	66.24	19.54

For SARIMA, it is observed that the interval width on the consecutive clear days (days 1, 2, and 3) decreases through time. This implies that the confidence of SARIMA depends on the training error standard deviation—multiple clear days lead to a smaller standard deviation, and thus a narrower prediction interval. Next, the effect of Fourier modeling on prediction interval is also apparent, see the interval variation during the nighttime in Fig. 6. However, since the nighttime forecasts are irrelevant, it does not affect the performance of SARIMA.

PMA+ORACLE gives narrow intervals with good coverage. This is expected. On the other hand, the performance of PMA+NAM and PMA+ENS depends highly on whether the NWP model is able to forecast the hourly variability. In days 4 and 5, PMA+NAM and PMA+ENS have very similar intervals to those of PMA+ORACLE, indicating that the NWP was successful in predicting the irradiance variability for these days. However, for day 6, despite the varying 15-min pattern, PMA+NAM and PMA+ENS do not reflect much deviation in their ensemble members (i.e., small interval width). The reason can be traced to the NWP forecasts—when the NWP forecasts a clear sky day, the ensemble members most likely come from other clear days. Lastly, it is observed that PMA+ENS is somewhat inaccurate near solar noon during a clear day. This is because of the MOS adjustment, see Fig. 5. The MOS correction applied in this article tends to move GHI towards the average GHI observed for a given predicted CSI and solar zenith angle; therefore the forecast tends to underpredict on clear days and overpredict on cloudy days. However, developing better MOS models is not within the scope of this work.

6. Discussion

6.1. How to improve the poor probabilistic forecasting performance?

Given the good deterministic forecasting performance of the proposed pattern-matching method, the present focus is on improving its probabilistic forecasting performance. It should be clear now that the poor performance of PMA+NAM and PMA+ENS is owing to the poor coverage. In other words, due to the high similarity among the ensemble members, PMA+NAM and PMA+ENS generate prediction intervals that are too narrow.

To diversify the ensemble members, several actions can be taken: (1) increase the *query* length m, (2) decrease the *history* length n, and (3) increase the number of ensemble members N. By increasing m, the Euclidean distance will have more degrees-of-freedom, and thus the analogs are more diversified. By decreasing n, the choice of candidates is reduced, and thus less similar candidates will be added. Lastly, the aim of increasing N is also to loosen the selection criterion, and thus include some less similar analogs. There is no doubt that one could iterate these settings and somewhat identify a best approach, see Appendix B for additional empirical results. Nevertheless, from a data science perspective, the empirically identified "best choice" is only suitable for the current dataset, which may not apply to other scenarios. A more general solution is preferred.

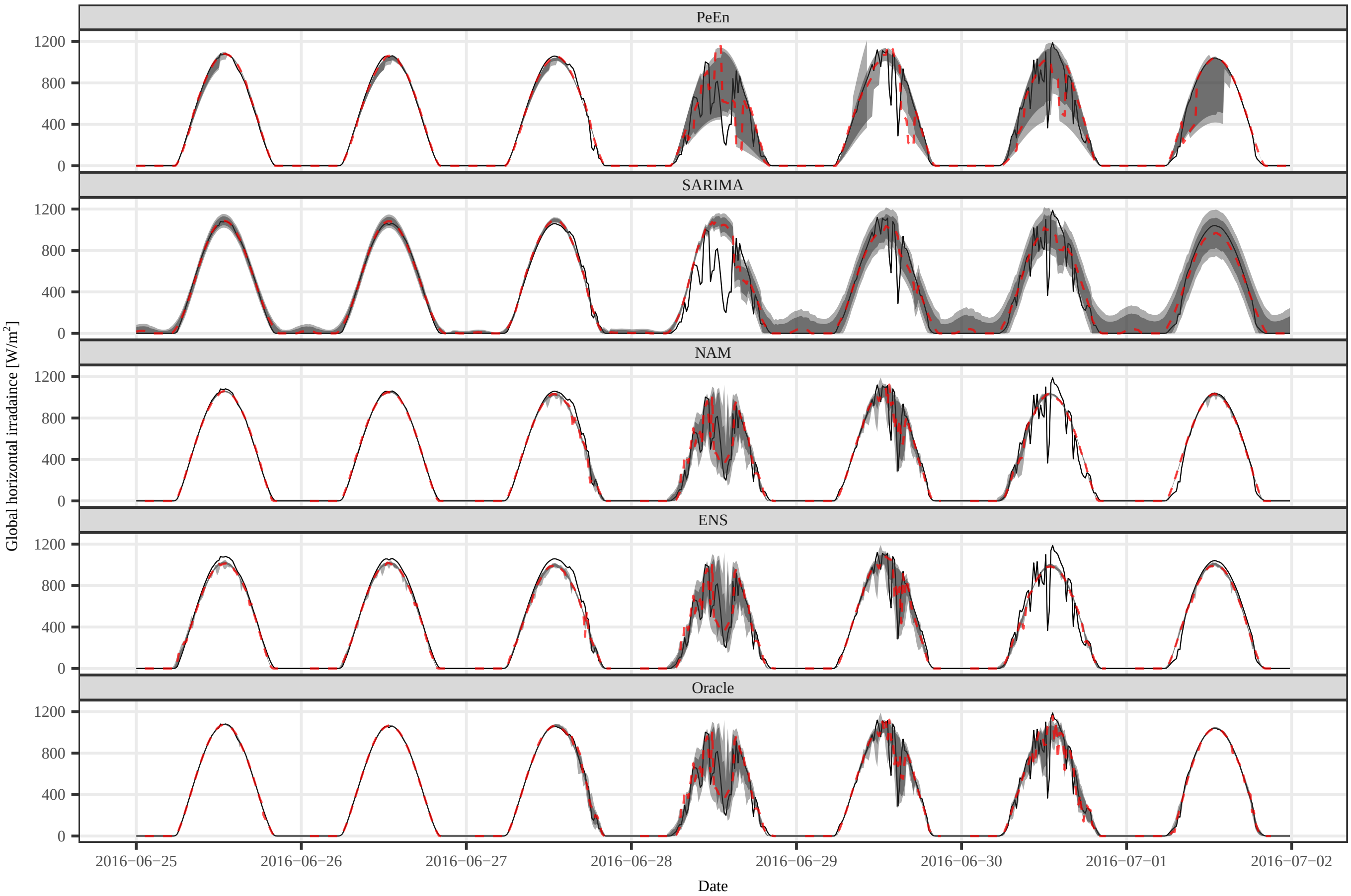

Figure 6: Probabilistic forecasting results over a week in 2016. The solid black lines plots the measurement from Surfrad15, whereas the dashed red lines are the deterministic forecasts. The dark and light ribbons show 80% and 95% prediction intervals, respectively. The time is shifted from UTC to local time for visualization.

Since PeEn has good coverage but wide intervals, whereas Pma+Nam and Pma+Ens do not have enough coverage but their prediction intervals are narrow, the most intuitive approach is to even out the intervals generated by different methods. Although this approach appears too ad hoc at the first glance, it aligns with the well-accepted framework of forecast-ensemble calibration (Raftery et al., 2005). Moreover, in reality, such simple combination of predictions often lead to desirable outcome (Yang and Dong, 2018; Yang, 2018b). To that end, the three sets of forecasts generated by PeEn, Pma+Nam and Pma+Ens are combined. For each model, the 20 forecasts are first sorted. Subsequently, the forecasts made by different models are averaged, following the sorted order. With the 20 newly combined forecast, a new prediction interval can be formed. The performance of this new model is shown in the last column of Table 4. Positive skills are now observed for all evaluation periods.

6.2. Extending the pattern-matching routine to a multivariate case

As mentioned in the introduction, AnEn often select analogs based on the weighted sum of Euclidean distances between several meteorological variables, see Eq. (1). Therefore, extending the current pattern-matching routine to a multivariate case is trivial—one can simply iterate the algorithm several times, and sum the distances. Although the convolution step needs to be repeated N times, the resulting computational speed is still faster than a standard implication by an order of magnitude.[12]

6.3. *The impacts of* Pma *on solar forecasting research*

The case study in Section 5 reveals a series of positive impacts of Pma that could potentially advance the field of solar forecasting. Firstly, the Pma+Oracle, i.e., Pma with perfect day-ahead forecasts, demonstrated extraordinary results in both deterministic and probabilistic forecasting. Hence, it can be concluded that better NWP forecasts would lead to better downscaled forecasts at the 6–8-h horizon. This implies that future solar forecasting research should place a high priority on improving the NWP models.

Secondly, the forecast skill and CRPS skill score of Pma increase with forecast horizon. Although at the 1-h-ahead horizon, Pma slightly underperforms, one can use a regime-switching approach to separate the forecasting tasks based on forecast horizon, i.e., 1-h-ahead forecasting can be replaced by a more suitable algorithm.

Thirdly, Pma complements the traditional way of generating ensemble forecasts using NWP by running the NWP model multiple times; Pma is comparatively computationally cheaper to implement.

To confirm the above-mentioned impacts, the case study is extended to all SURFRAD stations, which covers 5 different climate zones according to the Köppen-Geiger climate classification. The additional deterministic and probabilistic forecasting results are provided in Appendix C. Consistent conclusions can be drawn from the extensive empirical results, confirming the universality of the proposed algorithm.

6.4. *Future works*

Whereas this work provides a framework for operational solar forecasting in the RTM, there are several potential issues that need to be investigated in the future. Firstly, since better NWP forecasts can lead to better intra-hour forecasts, improving the accuracy of the raw NWP forecasts is beneficial. In this regard, the various research versions of WRF developed by the Center for Renewable Resources and Integration, University of California, San Diego (Wu et al., 2018; Sahu et al., 2018; Zhong et al., 2017), can be tested in the future. Besides improving the raw NWP forecasts, better post-processing techniques, such as Rincón et al. (2018), can be involved. Lastly, the topic of prediction interval ensemble in the form of Raftery et al. (2005), can be further explored for solar forecasting.

One interesting features of the pattern-matching based algorithms is that the *history* time series need not come from the same location as the hourly forecasts. In other words, as long as the *history* comes from a location within a same climate zone or with similar latitude (so that the zenith angle can match), the proposed algorithm will most likely suffice. Since NWP forecasts are available throughout the continental US, the present downscaling approach provides a unique solution to high-resolution forecasting, without local measurements.

7. Conclusion

A pattern-matching-based algorithm is proposed to generate solar forecasts for short-term unit commitment in the CAISO real-time market. Unlike previous solar forecasting publications, this work follows the CAISO RTM requirements exactly. All time parameters including forecast horizon, resolution, lead time, and update rate are considered. More specifically, 5-h-ahead forecasts in 15-min intervals are generated 75 min prior to an operating hour, and the forecasts are updated every hour.

The algorithm has three major steps. Firstly, the 12–35-h-ahead NAM forecasts are improved using a state-of-the-art ensemble time series technique. Next, the 1-h resolution forecasts are matched to an 18-year historical hourly GHI series measured at a SURFRAD station, using the world's fastest similarity search algorithm. The best-matched analogs are then downscaled to a 15-min resolution. Lastly, to improve the model performance in probabilistic forecasting, an ensemble of prediction intervals is formed. The algorithm is validated using two years of data. For deterministic forecasting, the proposed model results in a forecasting skill of 5–31%, whereas for the probabilistic forecasting, the proposed model results in a CRPS skill score of 8–20%.

[12] The algorithm has been tested against the R code provided by Stefano Alessandrini, who is a major contributor of the AnEn solar forecasting, and has authored tens of AnEn forecasting papers. The present algorithm has been transferred to the National Center for Atmospheric Research (NCAR), so that a faster Fortran version can be eventually used in NCAR's operational forecasting.

This article focuses on GHI forecasting. However, in actual power system operations, solar-generated power is of interest. Hence, in addition to the method proposed in this work, some irradiance-to-power conversion methods are required. For example, for flat-surface PV systems, it usually takes a three-step procedure: (1) separating diffuse horizontal irradiance component from the GHI forecast (see Gueymard and Ruiz-Arias, 2016, for a review on separation modeling); (2) transposing the horizontal irradiance components to tilted surface (see Yang, 2016, for a review on transposition modeling); and (3) a PV performance model to convert the in-plane irradiance to power (see Skoplaki and Palyvos, 2009, for a review on temperature dependence during power conversion). Since each of these steps would introduce some new errors, it is unclear how the GHI forecast errors reported in this work would propagate to the eventual power forecast error. Therefore, further studies on this subject are needed.

Appendix A. Data aggregation and forecast consistency

With the exception of physically-based forecasting, where weather variables are integrated in time in multiple small steps, the majority of statistical and machine-learning solar forecasting models are limited to the data-aggregation resolution. For example, if the 1-min raw data are aggregated to a 10-min resolution, the forecasts made will be in 10-min steps. In other words, 1-step-ahead forecasting corresponds to 10-min-ahead forecasting, whereas 2-step-ahead forecasting corresponds to 20-min-ahead forecasting. However, there are other ways to generate such 10-min-ahead forecasts. For instance, one can aggregate the 1-min raw data to a 5-min resolution and perform a 2-step-ahead forecasting to obtain a 10-min-ahead forecast. Alternatively, one can also use 2-min data with 5-step-ahead forecasting, or use 1-min data with 10-step-ahead forecasting. Due to the modeling error, each of the above-mentioned forecasting scheme will produce different forecasts that are very unlikely to be *aggregate consistent*, namely, the 5 forecasts made using 2-min data will not add up to the single forecast made using 10-min data. Hence, the question "which scheme should be used?" needs to be addressed. In fact, such discussion has been around since at least (Dong et al., 2013), but has not attracted significant attention from the academicians.

Of course, one simple way to address the question is to test all possible schemes, as seen in Dong et al. (2013), and to contrast the results. Nevertheless, it is time consuming, and conclusions may vary across different datasets. It was not until a recent publication by Athanasopoulos et al. (2017) that this problem is properly addressed. The temporal reconciliation method therein proposed can *unify* all forecasts produced using different horizon–resolution combinations. Furthermore, it improves the forecast accuracy, owing to the cancellation of modeling errors. Such reconciliation has also been applied to solar forecasting (Yang et al., 2017). Unfortunately, neither publication received sizable echo from solar forecasters, for unknown reasons.

Appendix B. Effect of model parameters on PMA

In Section 6.1, several potential approaches—without using interval averaging—to improve the probabilistic forecasting performance of PMA are reasoned. These approaches aim at diversifying the ensemble members by (1) increasing m, (2) decreasing n, and (3) increasing N. This appendix extends the PMA+ENS case study, by perturbing these model parameters.

The results shown in Table 4 are generated using $m = 8$, $n = 18$ years, and $N = 20$. Firstly, the value of m is gradually increased to 24, while n and N are kept unchanged. It is observed that the $m = 24$ case has the smallest CRPS. Next, by fixing $m = 24$ and $n = 18$ years, the number of ensemble members, N, is gradually increased up to 300. Further reduction in CRPS is observed as N goes to 300. On the other hand, reducing the history length n to 5 years seems to have a negative impact on forecast accuracy. These results are tabulated in Table B.5.

It is noted that the approach used here is not practical for two main reasons: (1) the choice of parameters would vary across geographical locations, and (2) the ISOs would rarely have the luxury to fine tune the model parameters for every forecasting task. Hence, interval averaging appears to be a more appropriate way to ensure a satisfactory probabilistic forecasting performance.

Appendix C. Performance of PMA under other climate zones

In this appendix, the performance PMA is further validated at locations in other climate zones that are covered by SURFRAD, see Table C.6 for a summary. The complete procedure including NWP post-processing and various

Table B.5: Effect of model parameters, m, n, and N, on the probabilistic forecasting performance of PMA. The first three columns are identical to Table 4, reprint here for easy referencing.

Evaluation period	PeEn	SARIMA	PMA+ENS $m = 8$, $n = 18$ yr, $N = 20$	$m = 24$, $n = 18$ yr, $N = 20$	$m = 24$, $n = 18$ yr, $N = 300$	$m = 24$, $n = 5$ yr, $N = 300$
			Brier score			
1	0.52	0.63	0.70	0.64	0.58	0.67
2	0.55	0.65	0.69	0.64	0.58	0.67
3	0.56	0.65	0.69	0.64	0.58	0.67
4	0.57	0.66	0.68	0.64	0.58	0.67
5	0.57	0.66	0.69	0.64	0.58	0.67
			CRPS [W/m^2]			
1	47.83	55.87	54.55	51.84	49.93	56.80
2	52.04	59.81	54.18	51.85	49.89	56.76
3	55.24	61.62	54.08	51.84	49.90	56.75
4	57.65	62.57	54.27	51.81	49.89	56.80
5	59.54	63.10	54.65	51.86	49.89	56.93
			CRPS skill score [%]			
1	0.00	-16.81	-14.04	-8.39	-4.38	-18.74
2	0.00	-14.93	-4.12	0.36	4.13	-9.06
3	0.00	-11.56	2.09	6.16	9.67	-2.75
4	0.00	-8.54	5.86	10.12	13.45	1.47
5	0.00	-5.98	8.22	12.90	16.21	4.39

Table C.6: Metadata of the SURFRAD network and their corresponding Köppen-Geiger climate classification.

Abbrv.	Station	Latitude	Longitude	Time zone	Köppen-Geiger	Climate description
BON	Bondville, Illinois	40.05192° N	88.37309° W	Central	Dfa	Hot-summer humid continental
DRA	Desert Rock, Nevada	36.62373° N	116.01947° W	Pacific	BWk	Cold desert
FPK	Fort Peck, Montana	48.30783° N	105.10170° W	Mountain	BSk	Cold semi-arid (steppe)
GWN	Goodwin Creek, Mississippi	34.25470° N	89.87290° W	Central	Cfa	Humid subtropical
PSU	Penn. State Univ., Pennsylvania	40.72012° N	77.93085° W	Eastern	Dfb	Warm-summer humid continental
SXF	Sioux Falls, South Dakota	43.73403° N	96.62328° W	Central	Dfa	Hot-summer humid continental
TBL	Table Mountain, Boulder, Colorado	40.12498° N	105.23680° W	Mountain	BSk	Cold semi-arid (steppe)

versions of PMA are repeated. Without loss of generality, the PMA setting herein used is $m = 8$, $n = 18$ years, and $N = 20$, except for the Sioux Falls station, South Dakota, which was established in 2003 with $n = 14$ years. Even though the other SURFRAD stations are outside of CAISO, the CAISO operational requirements are used for illustration purposes. The deterministic and probabilistic forecasting results for these additional empirical studies are shown in Tables C.7–C.18.

Based on these extensive empirical studies using data from different climate zones, the universality of the proposed algorithm can be confirmed. All previously discussed issues can be transferred to these new case studies. For clarity, they are re-iterated here:

1. It is necessary to post-process the raw NWP output, since PMA+ENS outperforms PMA+NAM at all stations;
2. The performance of PMA+ORACLE is extraordinary at all stations, indicating that a better hourly forecast would lead to a better 15-min forecasts;
3. The advantages of the proposed algorithm becomes more apparent at 3–5-h-ahead horizons; and
4. The averaging of prediction interval is an effective way of improving the accuracies of probabilistic forecasting.

References

Aguiar, L.M., Pereira, B., Lauret, P., Díaz, F., David, M., 2016. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. Renewable Energy 97, 599 – 610. URL: http://www.sciencedirect.com/science/article/pii/S0960148116305390, doi:https://doi.org/10.1016/j.renene.2016.06.018.

Akarslan, E., Hocaoglu, F.O., 2017. A novel method based on similarity for hourly solar irradiance forecasting. Renewable Energy 112, 337 – 346. URL: http://www.sciencedirect.com/science/article/pii/S0960148117304445, doi:https://doi.org/10.1016/j.renene.2017.05.058.

Table C.7: Same as Table 3, but for Bondville, Illinois (40.05192° N, 88.37309° W).

Evaluation period	Pers	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle
			nMBE [%]		
1	-0.52	0.19	1.72	2.04	-0.07
2	-1.24	0.28	1.64	2.25	-0.21
3	-2.06	0.21	1.73	2.15	-0.17
4	-2.82	0.05	1.74	2.01	-0.12
5	-3.38	-0.07	1.73	1.78	-0.28
			nRMSE [%]		
1	31.48	31.85	34.47	32.03	17.89
2	35.79	35.33	34.42	32.20	17.97
3	39.62	37.38	34.72	32.26	18.12
4	42.96	38.47	34.36	32.16	18.15
5	45.65	39.06	34.84	32.50	18.28
			Forecast skill [%]		
1	0.00	-1.16	-9.49	-1.73	43.18
2	0.00	1.29	3.83	10.04	49.80
3	0.00	5.66	12.37	18.58	54.27
4	0.00	10.47	20.03	25.15	57.75
5	0.00	14.43	23.68	28.81	59.97

Table C.8: Same as Table 4, but for Bondville, Illinois (40.05192° N, 88.37309° W).

Evaluation period	PeEn	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle	Interval averaging
				Brier score		
1	0.68	0.79	0.75	0.92	0.41	0.72
2	0.72	0.82	0.75	0.91	0.41	0.73
3	0.74	0.83	0.75	0.91	0.41	0.74
4	0.76	0.84	0.76	0.91	0.40	0.75
5	0.77	0.84	0.76	0.90	0.40	0.76
				CRPS [W/m^2]		
1	74.32	82.26	79.56	83.97	29.26	68.45
2	82.56	92.76	79.33	83.52	29.26	70.55
3	89.10	98.85	79.54	83.44	29.24	72.39
4	94.29	102.03	80.29	83.53	28.87	73.99
5	98.31	103.85	81.16	83.60	28.55	75.32
				CRPS skill score [%]		
1	0.00	-10.68	-7.05	-12.98	60.62	7.90
2	0.00	-12.36	3.91	-1.16	64.56	14.54
3	0.00	-10.94	10.73	6.36	67.19	18.75
4	0.00	-8.21	14.84	11.41	69.38	21.53
5	0.00	-5.63	17.45	14.96	70.96	23.39

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

Athanasopoulos, G., Hyndman, R.J., Kourentzes, N., Petropoulos, F., 2017. Forecasting with temporal hierarchies. European Journal of Operational Research 262, 60 – 74. URL: http://www.sciencedirect.com/science/article/pii/S0377221717301911, doi:https://doi.org/10.1016/j.ejor.2017.02.046.

Ayet, A., Tandeo, P., 2018. Nowcasting solar irradiance using an analog method and geostationary satellite images. Solar Energy 164, 301 – 315. URL: http://www.sciencedirect.com/science/article/pii/S0038092X18301993, doi:https://doi.org/10.1016/j.solener.2018.02.068.

Bae, K.Y., Jang, H.S., Sung, D.K., 2017. Hourly solar irradiance prediction based on support vector machine and its error analysis. IEEE Transactions on Power Systems 32, 935 – 945. doi:https://doi.org/10.1109/TPWRS.2016.2569608.

Cervone, G., Clemente-Harding, L., Alessandrini, S., Monache, L.D., 2017. Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble. Renewable Energy 108, 274 – 286. URL: http://www.sciencedirect.com/science/article/pii/S0960148117301386, doi:https://doi.org/10.1016/j.renene.2017.02.052.

Chen, R., Wang, J., Botterud, A., Sun, H., 2017. Wind power providing flexible ramp product. IEEE Transactions on Power Systems 32, 2049 – 2061. doi:https://doi.org/10.1109/TPWRS.2016.2603225.

Diagne, M., David, M., Boland, J., Schmutz, N., Lauret, P., 2014. Post-processing of solar irradiance forecasts from WRF model at Reunion Island. Solar Energy 105, 99 – 108. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14001509, doi:https:

Table C.9: Same as Table 3, but for Fort Peck, Montana (48.30783° N, 105.1017° W).

Evaluation period	Pers	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle
			nMBE [%]		
1	-1.14	0.25	8.89	2.51	-0.02
2	-2.40	0.33	8.85	2.52	-0.13
3	-3.82	0.35	8.91	2.66	-0.08
4	-5.26	0.32	8.81	2.45	0.00
5	-6.61	0.26	8.89	2.13	-0.35
			nRMSE [%]		
1	29.27	29.34	32.83	30.40	16.71
2	33.25	32.52	32.66	30.54	16.72
3	36.72	34.24	32.70	30.53	16.68
4	39.30	35.22	32.85	30.58	16.74
5	41.53	35.76	32.95	30.63	16.84
			Forecast skill [%]		
1	0.00	-0.24	-12.17	-3.87	42.90
2	0.00	2.20	1.78	8.16	49.71
3	0.00	6.75	10.95	16.85	54.58
4	0.00	10.39	16.43	22.20	57.40
5	0.00	13.88	20.65	26.25	59.45

Table C.10: Same as Table 4, but for Fort Peck, Montana (48.30783° N, 105.1017° W).

Evaluation period	PeEn	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle	Interval averaging
				Brier score		
1	0.67	0.74	0.76	0.94	0.37	0.71
2	0.71	0.77	0.76	0.94	0.37	0.73
3	0.73	0.79	0.77	0.94	0.37	0.74
4	0.74	0.79	0.77	0.94	0.36	0.75
5	0.75	0.80	0.77	0.94	0.36	0.75
				CRPS [W/m^2]		
1	65.25	69.11	72.65	77.29	24.69	61.70
2	71.84	77.73	72.66	77.15	24.65	63.64
3	76.40	82.18	72.97	77.08	24.51	65.00
4	79.40	84.61	73.51	77.39	24.16	66.03
5	81.46	85.95	74.10	77.74	23.94	66.77
				CRPS skill score [%]		
1	0.00	-5.93	-11.34	-18.46	62.16	5.43
2	0.00	-8.20	-1.14	-7.39	65.69	11.41
3	0.00	-7.57	4.49	-0.88	67.92	14.92
4	0.00	-6.57	7.42	2.54	69.57	16.84
5	0.00	-5.52	9.04	4.56	70.62	18.04

//doi.org/10.1016/j.solener.2014.03.016.

Dong, Z., Yang, D., Reindl, T., Walsh, W.M., 2013. Short-term solar irradiance forecasting using exponential smoothing state space model. Energy 55, 1104 – 1113. URL: http://www.sciencedirect.com/science/article/pii/S0360544213003381, doi:https://doi.org/10.1016/j.energy.2013.04.027.

GFDL Global Atmospheric Model Development Team, 2004. The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations. Journal of Climate 17, 4641 – 4673. doi:https://doi.org/10.1175/JCLI-3223.1.

Gueymard, C.A., Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Solar Energy 128, 1–30. URL: http://www.sciencedirect.com/science/article/pii/S0038092X15005435, doi:https://doi.org/10.1016/j.solener.2015.10.010. special issue: Progress in Solar Energy.

Hong, T., Fan, S., 2016. Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting 32, 914 – 938. URL: http://www.sciencedirect.com/science/article/pii/S0169207015001508, doi:https://doi.org/10.1016/j.ijforecast.2015.11.011.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., 2018. forecast: Forecasting functions for time series and linear models. URL: http://pkg.robjhyndman.com/forecast. r package version 8.4.

Inman, R.H., Pedro, H.T., Coimbra, C.F., 2013. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science 39, 535 – 576. URL: http://www.sciencedirect.com/science/article/pii/S0360128513000294, doi:https://doi.org/10.1016/j.pecs.2013.06.002.

Table C.11: Same as Table 3, but for Goodwin Creek, Mississippi (34.2547° N, 89.8729° W).

Evaluation period	Pers	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle
	nMBE [%]				
1	-0.82	0.98	5.68	1.65	-0.26
2	-1.77	1.04	5.74	1.80	-0.13
3	-2.77	0.86	5.68	1.75	0.03
4	-3.66	0.59	5.83	1.67	-0.23
5	-4.36	0.39	5.66	1.37	-0.18
	nRMSE [%]				
1	31.07	32.45	35.67	32.34	18.41
2	35.02	35.99	35.47	32.39	18.38
3	38.76	38.24	35.60	32.33	18.33
4	42.08	39.65	35.95	32.50	18.36
5	45.13	40.45	36.15	32.76	18.49
	Forecast skill [%]				
1	0.00	-4.45	-14.81	-4.10	40.73
2	0.00	-2.76	-1.28	7.52	47.53
3	0.00	1.33	8.14	16.57	52.71
4	0.00	5.76	14.56	22.76	56.36
5	0.00	10.37	19.90	27.40	59.03

Table C.12: Same as Table 4, but for Goodwin Creek, Mississippi (34.2547° N, 89.8729° W).

Evaluation period	PeEn	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle	Interval averaging
	Brier score					
1	0.69	0.80	0.73	0.92	0.41	0.70
2	0.72	0.83	0.73	0.91	0.41	0.72
3	0.74	0.84	0.73	0.91	0.41	0.73
4	0.76	0.85	0.74	0.91	0.40	0.74
5	0.77	0.86	0.75	0.91	0.40	0.74
	CRPS [W/m^2]					
1	78.12	87.82	83.05	85.85	29.58	70.29
2	87.03	98.96	82.81	85.21	29.71	72.64
3	94.38	105.96	83.09	85.15	29.67	74.78
4	100.41	110.44	84.12	85.52	29.41	76.78
5	105.28	113.11	85.34	86.14	29.04	78.45
	CRPS skill score [%]					
1	0.00	-12.42	-6.32	-9.90	62.13	10.01
2	0.00	-13.71	4.85	2.09	65.86	16.53
3	0.00	-12.28	11.96	9.77	68.57	20.76
4	0.00	-9.99	16.22	14.82	70.71	23.53
5	0.00	-7.43	18.94	18.18	72.42	25.49

Kaur, A., Nonnenmacher, L., Pedro, H.T., Coimbra, C.F., 2016. Benefits of solar forecasting for energy imbalance markets. Renewable Energy 86, 819 – 830. URL: http://www.sciencedirect.com/science/article/pii/S0960148115302901, doi:https://doi.org/10.1016/j.renene.2015.09.011.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2, 2 – 10. doi:https://doi.org/10.1109/JSTARS.2009.2020300.

Lorenz, E.N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences 26, 636 – 646.

Lu, S., Hwang, Y., Khabibrakhmanov, I., Marianno, F.J., Shao, X., Zhang, J., Hodge, B., Hamann, H.F., 2015. Machine learning based multi-physical-model blending for enhancing renewable energy forecast - improvement via situation dependent error correction, in: 2015 European Control Conference (ECC), pp. 283 – 290. doi:https://doi.org/10.1109/ECC.2015.7330558.

Makarov, Y.V., Etingov, P.V., Ma, J., Huang, Z., Subbarao, K., 2011. Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures. IEEE Transactions on Sustainable Energy 2, 433 – 442. doi:https://doi.org/10.1109/TSTE.2011.2159254.

Mathiesen, P., Collier, C., Kleissl, J., 2013. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. Solar Energy 92, 47 – 61. URL: http://www.sciencedirect.com/science/article/pii/S0038092X13000832, doi:https://doi.org/10.1016/j.solener.2013.02.018.

Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy 85, 967 – 977. URL: http://www.sciencedirect.com/science/article/pii/S0038092X11000570, doi:https://doi.org/

Table C.13: Same as Table 3, but for Penn. State Univ., Pennsylvania (40.72012° N, 77.93085° W).

Evaluation period	Pers	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle
			nMBE [%]		
1	-0.70	0.73	4.53	1.51	-0.23
2	-1.05	0.86	4.51	1.44	-0.14
3	-1.07	0.79	4.45	1.38	-0.24
4	-0.65	0.62	4.48	1.04	-0.25
5	0.22	0.42	4.48	0.78	-0.37
			nRMSE [%]		
1	35.74	35.82	39.48	36.19	20.74
2	40.82	39.83	39.34	36.27	20.43
3	45.98	42.33	39.58	36.21	20.21
4	50.85	43.66	39.64	36.49	20.63
5	55.05	44.34	39.75	36.90	20.85
			Forecast skill [%]		
1	0.00	-0.21	-10.44	-1.26	41.97
2	0.00	2.41	3.61	11.14	49.93
3	0.00	7.92	13.91	21.23	56.04
4	0.00	14.14	22.05	28.24	59.44
5	0.00	19.46	27.78	32.96	62.12

Table C.14: Same as Table 4, but for Penn. State Univ., Pennsylvania (40.72012° N, 77.93085° W).

Evaluation period	PeEn	Sarima	Pma+Nam	Pma+Ens	Pma+Oracle	Interval averaging
				Brier score		
1	0.73	0.79	0.82	0.96	0.43	0.77
2	0.77	0.83	0.82	0.96	0.43	0.78
3	0.79	0.84	0.82	0.96	0.43	0.79
4	0.81	0.85	0.82	0.95	0.42	0.80
5	0.83	0.86	0.81	0.95	0.42	0.81
				CRPS [W/m^2]		
1	82.24	86.93	89.21	90.81	30.51	75.12
2	91.95	98.77	88.99	90.61	30.60	77.56
3	99.95	105.84	89.33	90.61	30.39	79.62
4	106.33	109.70	89.84	90.89	30.18	81.28
5	111.26	111.70	90.18	91.06	30.09	82.39
				CRPS skill score [%]		
1	0.00	-5.71	-8.48	-10.42	62.90	8.66
2	0.00	-7.42	3.21	1.45	66.72	15.65
3	0.00	-5.89	10.62	9.34	69.60	20.34
4	0.00	-3.17	15.51	14.53	71.61	23.56
5	0.00	-0.40	18.94	18.15	72.95	25.95

10.1016/j.solener.2011.02.013.

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

Nonnenmacher, L., Coimbra, C.F., 2014. Streamline-based method for intra-day solar forecasting through remote sensing. Solar Energy 108, 447 – 459. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14003752, doi:https://doi.org/10.1016/j.solener.2014.07.026.

Pedro, H.T., Lim, E., Coimbra, C.F., 2018. A database infrastructure to implement real-time solar and wind power generation intra-hour forecasts. Renewable Energy 123, 513 – 525. URL: http://www.sciencedirect.com/science/article/pii/S0960148118301873, doi:https://doi.org/10.1016/j.renene.2018.02.043.

Perez, R., Kankiewicz, A., Schlemmer, J., Hemker, K., Kivalov, S., 2014. A new operational solar resource forecast model service for PV fleet simulation, in: 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC), pp. 69 – 74. doi:https://doi.org/10.1109/PVSC.2014.6925204.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D., Hoff, T.E., 2010. Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy 84, 2161 – 2172. URL: http://www.sciencedirect.com/science/article/pii/S0038092X10002823, doi:https://doi.org/10.1016/j.solener.2010.08.014.

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Knowe, G.V., Hemker, K., Heinemann, D., Remund, J., Müller, S.C., Traunmüller, W., Steinmauer, G., Pozo, D., Ruiz-Arias, J.A., Lara-Fanego, V., Ramirez-Santigosa, L., Gaston-Romero, M., Pomares, L.M., 2013. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Solar Energy 94, 305 – 326. URL: http://www.

Table C.15: Same as Table 3, but for Sioux Falls, South Dakota (43.73403° N, 96.62328° W).

Evaluation period	PERS	SARIMA	PMA+NAM	PMA+ENS	PMA+ORACLE
			nMBE [%]		
1	-0.46	0.44	5.10	3.14	-0.02
2	-0.63	0.61	5.06	3.19	-0.05
3	-0.67	0.54	5.11	3.32	-0.03
4	-0.49	0.43	5.12	3.21	-0.08
5	-0.05	0.32	5.17	2.88	-0.22
			nRMSE [%]		
1	30.10	31.30	34.14	31.72	15.77
2	34.50	35.37	34.10	31.91	15.81
3	38.63	37.88	34.18	32.00	15.53
4	42.47	39.35	34.20	31.91	15.65
5	45.80	40.14	34.41	31.90	15.59
			Forecast skill [%]		
1	0.00	-3.98	-13.41	-5.38	47.61
2	0.00	-2.52	1.15	7.51	54.17
3	0.00	1.94	11.52	17.16	59.81
4	0.00	7.36	19.48	24.86	63.14
5	0.00	12.37	24.87	30.36	65.96

Table C.16: Same as Table 4, but for Sioux Falls, South Dakota (43.73403° N, 96.62328° W).

Evaluation period	PeEn	SARIMA	PMA+NAM	PMA+ENS	PMA+ORACLE	Interval averaging
				Brier score		
1	0.66	0.77	0.74	0.90	0.37	0.70
2	0.71	0.81	0.74	0.89	0.37	0.72
3	0.74	0.82	0.74	0.89	0.37	0.73
4	0.76	0.83	0.74	0.89	0.37	0.75
5	0.78	0.84	0.75	0.89	0.37	0.76
				CRPS [W/m^2]		
1	70.23	76.86	75.45	79.25	24.38	64.71
2	79.43	88.68	75.20	78.96	24.56	67.10
3	87.25	95.67	75.33	78.99	24.41	69.26
4	93.70	99.85	75.96	79.13	24.00	71.17
5	98.75	102.34	76.85	79.43	23.67	72.62
				CRPS skill score [%]		
1	0.00	-9.45	-7.43	-12.84	65.29	7.86
2	0.00	-11.64	5.32	0.59	69.09	15.53
3	0.00	-9.65	13.66	9.47	72.02	20.62
4	0.00	-6.56	18.94	15.55	74.39	24.05
5	0.00	-3.63	22.18	19.57	76.03	26.46

sciencedirect.com/science/article/pii/S0038092X13001886, doi:https://doi.org/10.1016/j.solener.2013.05.005.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review 133, 1155 – 1174. doi:https://doi.org/10.1175/MWR2906.1.

Ren, Y., Suganthan, P., Srikanth, N., 2015. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renewable and Sustainable Energy Reviews 50, 82 – 91. URL: http://www.sciencedirect.com/science/article/pii/S1364032115003512, doi:https://doi.org/10.1016/j.rser.2015.04.081.

Rincón, A., Jorba, O., Frutos, M., Alvarez, L., Barrios, F.P., González, J.A., 2018. Bias correction of global irradiance modelled with weather and research forecasting model over Paraguay. Solar Energy 170, 201 – 211. URL: http://www.sciencedirect.com/science/article/pii/S0038092X18304948, doi:https://doi.org/10.1016/j.solener.2018.05.061.

Sahu, D.K., Yang, H., Kleissl, J., 2018. Assimilating observations to simulate marine layer stratocumulus for solar forecasting. Solar Energy 162, 454 – 471. URL: http://www.sciencedirect.com/science/article/pii/S0038092X18300185, doi:https://doi.org/10.1016/j.solener.2018.01.006.

Shakya, A., Michael, S., Saunders, C., Armstrong, D., Pandey, P., Chalise, S., Tonkoski, R., 2017. Solar irradiance forecasting in remote microgrids using markov switching model. IEEE Transactions on Sustainable Energy 8, 895 – 905. doi:https://doi.org/10.1109/TSTE.2016.2629974.

Skoplaki, E., Palyvos, J., 2009. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. Solar Energy 83, 614–624. URL: http://www.sciencedirect.com/science/article/pii/S0038092X08002788, doi:https://doi.org/10.1016/j.solener.2008.10.008.

Table C.17: Same as Table 3, but for Table Mountain, Boulder, Colorado (40.12498° N, 105.2368° W).

Evaluation period	PERS	SARIMA	PMA+NAM	PMA+ENS	PMA+ORACLE
			nMBE [%]		
1	-0.46	0.44	5.10	3.14	-0.02
2	-0.63	0.61	5.06	3.19	-0.05
3	-0.67	0.54	5.11	3.32	-0.03
4	-0.49	0.43	5.12	3.21	-0.08
5	-0.05	0.32	5.17	2.88	-0.22
			nRMSE [%]		
1	30.10	31.30	34.14	31.72	15.77
2	34.50	35.37	34.10	31.91	15.81
3	38.63	37.88	34.18	32.00	15.53
4	42.47	39.35	34.20	31.91	15.65
5	45.80	40.14	34.41	31.90	15.59
			Forecast skill [%]		
1	0.00	-3.98	-13.41	-5.38	47.61
2	0.00	-2.52	1.15	7.51	54.17
3	0.00	1.94	11.52	17.16	59.81
4	0.00	7.36	19.48	24.86	63.14
5	0.00	12.37	24.87	30.36	65.96

Table C.18: Same as Table 4, but for Table Mountain, Boulder, Colorado (40.12498° N, 105.2368° W).

Evaluation period	PeEn	SARIMA	PMA+NAM	PMA+ENS	PMA+ORACLE	Interval averaging
				Brier score		
1	0.66	0.77	0.74	0.90	0.37	0.70
2	0.71	0.81	0.74	0.89	0.37	0.72
3	0.74	0.82	0.74	0.89	0.37	0.73
4	0.76	0.83	0.74	0.89	0.37	0.75
5	0.78	0.84	0.75	0.89	0.37	0.76
				CRPS [W/m^2]		
1	70.23	76.86	75.45	79.25	24.38	64.71
2	79.43	88.68	75.20	78.96	24.56	67.10
3	87.25	95.67	75.33	78.99	24.41	69.26
4	93.70	99.85	75.96	79.13	24.00	71.17
5	98.75	102.34	76.85	79.43	23.67	72.62
				CRPS skill score [%]		
1	0.00	-9.45	-7.43	-12.84	65.29	7.86
2	0.00	-11.64	5.32	0.59	69.09	15.53
3	0.00	-9.65	13.66	9.47	72.02	20.62
4	0.00	-6.56	18.94	15.55	74.39	24.05
5	0.00	-3.63	22.18	19.57	76.03	26.46

Stephens, G.L., 1978. Radiation profiles in extended water clouds. II: Parameterization schemes. Journal of the Atmospheric Sciences 35, 2123 – 2132.

van der Meer, D., Widén, J., Munkhammar, J., 2018. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renewable and Sustainable Energy Reviews 81, 1484 – 1512. URL: http://www.sciencedirect.com/science/article/pii/S1364032117308523, doi:https://doi.org/10.1016/j.rser.2017.05.212.

Wang, W.C., 1976. A parameterization for the absorption of solar radiation by water vapor in the Earth's atmosphere. Journal of Applied Meteorology (1962-1982) 15, 21 – 27. URL: http://www.jstor.org/stable/26177550, doi:https://doi.org/10.2307/26177550.

Wang, Z., Koprinska, I., Rana, M., 2017. Solar power forecasting using pattern sequences, in: Lintas, A., Rovetta, S., Verschure, P.F., Villa, A.E. (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2017, Springer International Publishing. pp. 486 – 494.

Wu, E., Clemesha, R.E., Kleissl, J., 2018. Coastal stratocumulus cloud edge forecasts. Solar Energy 164, 355 – 369. URL: http://www.sciencedirect.com/science/article/pii/S0038092X18302056, doi:https://doi.org/10.1016/j.solener.2018.02.072.

Yang, D., 2016. Solar radiation on inclined surfaces: Corrections and benchmarks. Solar Energy 136, 288 – 302. URL: http://www.sciencedirect.com/science/article/pii/S0038092X16302432, doi:https://doi.org/10.1016/j.solener.2016.06.062.

Yang, D., 2017. On adding and removing sensors in a solar irradiance monitoring network for areal forecasting and PV system performance evaluation. Solar Energy 155, 1417 – 1430. URL: http://www.sciencedirect.com/science/article/pii/S0038092X17306461, doi:https://doi.org/10.1016/j.solener.2017.07.061.

Yang, D., 2018a. A correct validation of the National Solar Radiation Data Base (NSRDB). Renewable and Sustainable Energy Reviews 97, 152 – 155. URL: http://www.sciencedirect.com/science/article/pii/S1364032118306087, doi:https://doi.org/10.1016/

j.rser.2018.08.023.

Yang, D., 2018b. Spatial prediction using kriging ensemble. Solar Energy 171, 977 – 982. URL: http://www.sciencedirect.com/science/article/pii/S0038092X1830656X, doi:https://doi.org/10.1016/j.solener.2018.06.105.

Yang, D., Dong, Z., 2018. Operational photovoltaics power forecasting using seasonal time series ensemble. Solar Energy 166, 529 – 541. URL: http://www.sciencedirect.com/science/article/pii/S0038092X18301282, doi:https://doi.org/10.1016/j.solener.2018.02.011.

Yang, D., Jirutitijaroen, P., Walsh, W.M., 2012. Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy 86, 3531 – 3543. URL: http://www.sciencedirect.com/science/article/pii/S0038092X12003039, doi:https://doi.org/10.1016/j.solener.2012.07.029. solar Resources.

Yang, D., Kleissl, J., Gueymard, C.A., Pedro, H.T., Coimbra, C.F., 2018. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Solar Energy 168, 60 – 101. URL: http://www.sciencedirect.com/science/article/pii/S0038092X17310022, doi:https://doi.org/10.1016/j.solener.2017.11.023. Advances in Solar Resource Assessment and Forecasting.

Yang, D., Quan, H., Disfani, V.R., Rodríguez-Gallegos, C.D., 2017. Reconciling solar forecasts: Temporal hierarchy. Solar Energy 158, 332 – 346. URL: http://www.sciencedirect.com/science/article/pii/S0038092X17308423, doi:https://doi.org/10.1016/j.solener.2017.09.055.

Yang, D., Sharma, V., Ye, Z., Lim, L.I., Zhao, L., Aryaputera, A.W., 2015a. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. Energy 81, 111 – 119. URL: http://www.sciencedirect.com/science/article/pii/S0360544214013528, doi:https://doi.org/10.1016/j.energy.2014.11.082.

Yang, D., Ye, Z., Lim, L.H.I., Dong, Z., 2015b. Very short term irradiance forecasting using the lasso. Solar Energy 114, 314 – 326. URL: http://www.sciencedirect.com/science/article/pii/S0038092X15000304, doi:https://doi.org/10.1016/j.solener.2015.01.016.

Zhong, X., Sahu, D.K., Kleissl, J., 2017. WRF inversion base height ensembles for simulating marine boundary layer stratocumulus. Solar Energy 146, 50 – 64. URL: http://www.sciencedirect.com/science/article/pii/S0038092X17301147, doi:https://doi.org/10.1016/j.solener.2017.02.021.