The added value of combining solar irradiance data and forecasts: a probabilistic benchmarking exercise

Philippe Lauret[a,*], Rodrigo Alonso-Suárez[b], Rodrigo Amaro e Silva[c], John Boland[d], Mathieu David[a], Wiebke Herzberg[e], Josselin Le Gall La Salle[a], Elke Lorenz[e], Lennard Visser[f], Wilfried van Sark[f], Tobias Zech[e]

[a] *University of La Réunion - PIMENT laboratory, 15, avenue René Cassin, 97715 Saint-Denis*
[b] *Laboratorio de Energía Solar (LES), Departamento de Física, CENUR Litoral Norte, Udelar, Uruguay*
[c] *O.I.E. Centre Observation, Impacts, Energy, MINES Paris, PSL Research University, 06904 Sophia Antipolis, France*
[d] *Industrial AI, Centre for Industrial and Applied Mathematics, UniSA STEM, University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA 5095, Australia*
[e] *Fraunhofer Institute for Solar Energy Systems ISE, Heidenhofstr. 2, 79110 Freiburg, Germany*
[f] *Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a, 3584 CB, Utrecht, The Netherlands*

Abstract

Despite the growing awareness in academia and industry of the importance of solar probabilistic forecasting for further enhancing the integration of variable photovoltaic power generation into electrical power grids, there is still no benchmark study comparing a wide range of solar probabilistic methods across various local climates. Having identified this research gap, experts involved in the activities of IEA PVPS T16[1] agreed to establish a benchmarking exercise to evaluate the quality of intra-hour and intra-day probabilistic irradiance forecasts

The tested forecasting methodologies are based on different input data including ground measurements, satellite-based forecasts and Numerical Weather Predictions (NWP), and different statistical methods are employed to generate probabilistic forecasts from these. The exercise highlights different forecast quality depending on the method used, and more importantly, on the input data fed into the models.

In particular, the benchmarking procedure reveals that the association of a point forecast that blends ground, satellite and NWP data with a statistical technique generates high-quality probabilistic forecasts. Therefore, in a subsequent step, an additional investigation was conducted to assess the added value of such a blended point forecast on forecast quality. Three new statistical methods were implemented using the blended point forecast as input.

Overall, skill scores (which quantify the relative improvement of the tested methods over a reference forecast) of methods that use the blended point forecast ranges from 42% to 46% for the intra-hour scenario and 27% to 32% for the intra-day scenario. Conversely, methods

[1]International Energy Agency - Photovoltaic Power Systems - Solar Resource for High Penetration and Large Scale Applications.

that do not use the blended point forecast exhibit skill scores ranging from 33% to 43% for intra-hour forecasts and 8% to 16% for intra-day forecasts.

These results suggest that using a) blended point forecasts that optimally combine different sources of input data and b) a post-processing with a statistical method to produce the quantile forecasts is an effective and consistent way to generate high-quality intra-hour or intra-day probabilistic forecasts.

Keywords: probabilistic solar forecasting, benchmarking exercise, blended point forecast, CRPS, IEA PVPS T16

1. Introduction

Accurate forecasts of solar energy generation play a crucial role in effectively integrating solar power into existing grids and reducing associated expenses (Notton et al., 2018). This is because power output from photovoltaic plants (PV) is greatly influenced by weather conditions, making it highly variable. Consequently, having precise information about future solar power production is essential to minimize the need for costly balancing services and power reserves. Hence, enhancing solar forecasting models to increase the value of solar power generation becomes critically important. This work will focus in Global Horizontal Irradiance (GHI) forecasting since it is deemed one of the main drivers for solar power forecasting (Lorenz et al., 2021).

Today, users are faced with a myriad of forecasting methods. They can vary in the nature of the approach, the kind of inputs fed into the models, the outputs they produce or even in the forecasting horizon under consideration. Review publications aim to identify and structure all these elements, often comparing the performance values reported across the literature (Antonanzas et al., 2016; Sobri et al., 2018; Blaga et al., 2019; Yang et al., 2022). However, benchmark studies aim to do so using a methodology that ensures a fair comparison, i.e. the same locations and training and test periods. This has been done for many different aspects of solar forecasting: the post-processing of numerical weather prediction models (NWP) (Verbois et al., 2022), baseline approaches (Alonso-Suárez et al., 2022), autoregressive statistical learning approaches (Pedro and Coimbra, 2012; Lauret et al., 2015), cloud motion vector techniques (Aicardi et al., 2022), deep learning approaches using sky images as inputs (Paletta et al., 2021), among others. In other words, these benchmark studies ensure a comprehensive understanding and comparable indicators of the benefits

*corresponding author

Email addresses: philippe.lauret@univ-reunion.fr (Philippe Lauret), r.alonso.suarez@gmail.com (Rodrigo Alonso-Suárez), rodrigo.amaro_e_silva@minesparis.psl.eu (Rodrigo Amaro e Silva), John.Boland@unisa.edu.au (John Boland), mathieu.david@univ-reunion.fr (Mathieu David), wiebke.herzberg@ise.fraunhofer.de (Wiebke Herzberg), josselin.le-gal-la-salle@univ-reunion.fr (Josselin Le Gall La Salle), elke.lorenz@ise.fraunhofer.de (Elke Lorenz), l.r.visser@uu.nl (Lennard Visser), w.g.j.h.m.vansark@uu.nl (Wilfried van Sark), tobias.zech@ise.fraunhofer.de (Tobias Zech)

associated with a particular approach. For instance, in the frame of IEA SHC Task 46 (IEA-SHC-T46, 2024), Lorenz et al. (2009) designed a standardized procedure to evaluate the accuracy of day-ahead deterministic irradiance forecasts.

However, all these previous studies focus on deterministic forecasting and dismiss the inherent uncertainty of a forecast. Indeed, when it comes to decision-making for grid operators, utilities, aggregators, balancing responsible parties and others, having not only a point forecast but also an associated uncertainty or prediction interval becomes immensely valuable. In other words, reliable probabilistic predictions can significantly enhance the integration of variable energy sources within the energy network, leading to improved efficiency (Morales et al., 2014). Unlike the mature field of wind power forecasting, where probabilistic forecasting is well-established (Morales et al., 2014; Iversen et al., 2015; Jung and Broadwater, 2014; Pinson et al., 2007), probabilistic solar forecasting is still relatively nascent (Hong et al., 2016; van der Meer et al., 2018; Hong et al., 2020). Consequently, there are considerably less benchmark studies focusing on solar probabilistic forecasts.

A literature review restricted to intra-hour/intra-day solar probabilistic forecasts reveals that a few studies have started to address this gap. Among others, one can cite the following works. Grantham et al. (2016) used a non parametric bootstrapping method for generating prediction intervals of GHI at a forecast horizon of 1h. The bootstrap technique requires a point forecast which is, in their work, delivered by a linear auto-regressive (AR) model. With only past ground data, David et al. (2016) used a combination of point forecast ARMA model and a parametric GARCH model to generate intra-hour (up to 1h ahead with a time step of 10 mins) and intra-day (up to 6h ahead with a time step of 1h) GHI probabilistic forecasts. Lauret et al. (2017) evaluated the quality of three probabilistic models for intra-day solar forecasting. A linear quantile regression technique is used to build three models for generating 1 to 6h ahead probabilistic forecasts. Inputs of the models are either only ground data or ground data with day-ahead forecasts provided by the European Center for Medium-Range Weather Forecast (ECMWF). The results demonstrated that the Numerical Weather Prediction (NWP) exogenous inputs improve the quality of the intra-day probabilistic forecasts. Using only past ground GHI measurements, David et al. (2018) set up a combination of 3 points forecasting methods and 7 probabilistic methods to issue intra-day GHI forecasts. None of the model combinations clearly outperformed the others. However, regardless of the point forecasting method used, linear models in quantile regression, weighted quantile regression and gradient boosting decision trees appear to produce probabilistic forecasts with higher quality than the other proposed methods. In their work, Alonso-Suárez et al. (2020) developed three models aimed at generating intra-day probabilistic GHI forecasts, spanning lead times from 10 minutes to 3 hours with a granularity of 10 minutes. The initial model solely relies on historical ground measurements. The second model enhances the first one by integrating a variability metric derived from these historical ground measurements. The third model incorporates satellite albedo as an additional input. A linear quantile regression is employed to create directly (i.e. without using a point forecast) a range of quantiles summarizing the predictive distributions of the global solar irradiance. The findings demonstrate that the inclusion of satellite data further enhances the quality of the probabilistic forecasts. Mazorra-Aguiar et al. (2021) assessed the performance of two approaches for solar probabilistic forecasting to generate intra-day solar forecasts covering time horizons from 1 hour to 6 hours. The first approach involves a two-step process. Initially, point forecasts are generated for each forecast horizon, followed by the utilization of quantile regression techniques to estimate the prediction intervals. The second methodology directly predicts the quantiles of the predictive distribution using past ground data as input. Yang et al. (2020) benchmarked 5 forecasting intrahour/intraday solar probabilistic methods (including notably an Analog Ensemble method and a linear quantile regression technique) on a standardized dataset set up by (Pedro et al., 2019). All the proposed methods generate directly the quantile forecasts without resorting to a point deterministic forecast. The findings clearly highlight the significance of exogenous inputs in probabilistic solar forecasting, as all methods demonstrate enhanced results upon the integration of exogenous features computed from sky, satellite images and NWP outputs provided by the NAM the North American Mesoscale (NAM) forecast system.

Finally, it must be noted that specific methods based on Cloud-Motion Vector (CMV) approach or combination of sky and satellite images have been recently proposed in the literature. For instance, Carrière et al. (2021) designed a CMV-based probabilistic method which is an extension of the deterministic CMV approach by adding Gaussian noise to the norm and direction of the cloud motion vector estimates. Paletta et al. (2023) used an hybrid deep learning method combining sky images, satellite observations and/or past ground irradiance to generate intra-hour solar forecasts.

Following the previous literature review, the following comments can be made. To evaluate the quantile forecasts two methodologies can be distinguished. The first one leverages on a point deterministic forecast to produce with a specific statistical technique the prediction intervals. The second one generates directly (i.e. without resorting to a point forecast) the quantile forecasts. Regarding the first methodology, no work tries to evaluate the impact of a high quality point forecast on the generation of probabilistic forecasts.

Moreover, to the best of our knowledge, no benchmarking exercise has been conducted to compare classical probabilistic techniques like quantile regression or Analog Ensemble with a CMV-based probabilistic approach on multiple sites experiencing different climate conditions.

Therefore, as part of IEA PVPS T16 (IEA-PVPS-T16, 2024), experts engaged in Activity 3.3 on solar probabilistic forecasts found it essential to complement these previous studies regarding intra-hour and intra-day solar probabilistic forecasts. In other words, it appears important to experts of the IEA PVPS T16 to propose to the solar forecasters community a comprehensive benchmarking exercise related to intra-day and intra-hour solar irradiance forecasts.

To this end, five participants set-up a benchmarking exercise based on a shared ground measurement, satellite and NWP data. Eight European sites with diverse climatic conditions were chosen for this purpose. The proposed benchmarking procedure is implemented to compare 15-min irradiance probabilistic forecasts up to 6 hours issued by each participant with their own forecasting methods. In particular, to fill the gaps highlighted by the literature review, it appeared first important to the IEA PVPS experts to jointly evaluate a CMV probabilistic system with traditional quantile forecasting methods. Second, an assessment of the impact of a high quality point forecast on the quality of the generated probabilistic predictions is also conducted in this work.

To evaluate the quality of the probabilistic predictions, different diagnostic tools and scoring rules can be employed (Lauret et al., 2019). For user convenience, the verification scheme should be kept simple. For that, we propose using the reliability diagram as a visual diagnostic tool and the Continuous Ranked Probability Score (CRPS), as the numerical score. It is commonly adopted by the community in the verification process of solar irradiance probabilistic forecasts.

Further, in this work, unlike most of the previous studies, and in order to better highlight the skill of a forecasting method, we propose the numerical decomposition of the CRPS into the reliability and resolution components as in (Lauret et al., 2019).

The rest of the paper is organized as follows. Firstly, this paper introduces the benchmarking exercise. Section 3 details the data used in the exercise while Section 4 gives an overview of the diverse forecasting methodologies. The verification framework is presented in Section 5. Section 6 gives the main results of the benchmark and Section 7 discusses the impact of combining a blended point forecast with statistical techniques to generate probabilistic solar forecasts. Finally, Section 8 concludes this paper.

2. The benchmarking exercise

In the frame of the IEA PVPS Task 16 (IEA-PVPS-T16, 2024), five participants agreed to set up a benchmarking exercise related to intra-day and intra-hour solar irradiance probabilistic forecasting. Table 1 lists the participants of this benchmark, the code that will be used to identify them throughout the following sections and plots, as well as the forecasting methods used. For this exercise, each participant submitted their forecasts under the form of quantile forecasts (i.e. the quantiles of the predictive distribution).

Together, the participants designed the framework that would guide this benchmark, namely the type of input data that could be fed into the forecasting models, the forecast horizons to be considered, and the selection of the probability levels of the quantile forecasts. Table 2 gives details regarding these decisions. Thus, each participant was responsible for generating 15-min irradiance probabilistic forecasts up to 2 hours ahead (for intra-hour) and up to 6h (for intra-day) for 8 selected European sites (described in Section 3). The verification of the forecasts was conducted blindly by one of the authors of this paper. The MAE, equivalent to the CRPS of a deterministic forecast, of the median of the predictive distribution was included in order to evaluate the improvement of CRPS brought by the probabilistic approaches over their deterministic counterparts.

Participant	Code/Method	Forecasting methodology	Input data
Mines Paris (OIE)	OIE	CMV-based probabilistic approach	Satellite data
University of La Réunion (PIMENT)	PIMENT	Parametric method ARMA-GARCH	Ground data
Fraunhofer (ISE)	ISE	Blended point forecast +Analog Ensemble	Ground + satellite + NWP data
Utrecht University (UU)	UU	Non-linear Quantile Regression Forest	Ground data
Laboratorio de Energía Solar (Udelar)	LES	Linear quantile regression	Ground + satellite data

Table 1: List of participants. The code associated to each participant also identifies the forecasting methodology used by the participant

Type of Input Data	- Ground GHI measurements - Satellite estimates - NWP forecasts - Solar geometry variables (e.g. Solar Zenith Angle (SZA), etc)
Forecast horizon	- Intra-hour: 8 horizons (15 to 120 min, in 15-min steps) - Intra-day: 16 horizons (135 to 360 min, in 15-min steps)
Forecasts specification	15 GHI quantile forecasts with probability levels of $[0, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.975, 1]$
Verification metrics	- Reliability diagram - CRPS - CRPS skill score with the CSD-CLIM model as baseline - see (Le Gal La Salle et al., 2021). - MAE of the median of the predictive distribution

Table 2: Parameters of the benchmarking exercise. Details regarding the verification metrics are provided in section 5

3. Data for the benchmarking exercise

3.1. Ground measurements

It is crucial to use identical data for evaluation when comparing various prediction methods. The selected dataset comprises 15-min measured GHI values from eight locations of Europe. We restricted the evaluation to European sites since the methods of two of the participants used satellite data only covering most of the European domain. The evaluation period spans from January 2017 to December 2018. The year 2017 was chosen for the training set of the different methods described below, while 2018 was used for testing these methods.

The original reference database comprises high temporal resolution GHI data (1 minute) that were collected for a benchmarking exercise of modelled solar radiation data (Forstinger et al., 2021). This benchmark exclusively incorporates quality-assured data, meticulously checked using an extensive range of best practices and newly established quality-control procedures (Forstinger et al., 2021). These procedures encompass automated and manual data quality tests along with descriptive quality flagging, conducted by a team of experts of the IEA PVPS T16 subtask 1 (IEA-PVPS-T16, 2024).

The 15-min dataset results from a downsampling of these original 1-min. More precisely, the 1-min raw GHI were averaged at 15 minutes resolution. Also, in case of missing raw data, a linear interpolation is done if the gap is below 1h otherwise the whole day is discarded. Finally, data for which solar elevation $\leq 10°$ have been filtered out and are consequently not taken into account in the evaluation process.

Table 3 gives all the details related to each site. Note that, except for the TAB site provided by CIEMAT/DLR (CIEMAT, 2024), all sites are part of the BSRN (BSRN, 2024).

Site	Code	Latitude (°N)	Longitude (°E)	Altitude (m)	Köppen C.	Source	Mean GHI (W/m^2)
Cabauw	CAB	51.9711	4.9267	0	Cfb	BSRN	315.0
Carpentras	CAR	44.083	5.059	100	Csa	BSRN	411.1
Cener	CEN	42.816	-1.601	471	Cfb	BSRN	381.6
Milan	MIL	45.4761	9.2545	150	Cfa	RSE	394.1
Palaiseau	PAL	48.713	2.208	156	Cfb	BSRN	347.7
Payerne	PAY	46.815	6.944	491	Cfb	BSRN	368.9
Plataforma Solar	TAB	37.0909	-2.3581	500	Bsk	CIEMAT/DLR	499.5
Toravere	TOR	58.254	26.462	70	Dfb	BSRN	318.1

Table 3: Locations and key figures of ground measurements used for the benchmark. Column "Köppen C." lists the Köppen-Geiger climate classification of each site while column "Mean GHI" gives the average GHI of the test dataset.

3.2. Satellite data

In this study, two data sets of GHI estimates based on satellite data are considered. Both are based on images obtained by the SEVIRI instrument onboard the Meteosat Second Generation (MSG) satellite using the MSG 15-minute visible channel with a spatial resolution of approximately 1×2 km at Europe.

The first data set used for the OIE and LES forecasts is derived from the satellite images using the Heliosat-4 model (Qu et al., 2017a). The GHI estimates in the second data set used for the ISE forecasts are based on a modified version of the Heliosat method (Hammer et al., 2003).

Generally, for forecasting purposes, a sequence of satellite images is used to infer cloud motion vectors (CMV), i.e. vectors that describe cloud advection, which can be extrapolated into the future to make a prediction. For a better understanding of the assumptions and limitations of the use of CMV, the reader is directed to (Lorenz et al., 2021).

3.3. NWP forecasts

NWP forecasts are used as input to generate blended point forecasts by ISE. Thereto, ISE includes the atmospheric model high resolution 10-day forecast (HRES) product of the ECMWF IFS forecast with a spatial resolution of 0.125° and a time resolution of 1 h. The forecast of shortwave solar radiation downwards (ssrd) is used at base-times 0:00 UTC and 12:00 UTC. It is spatially smoothed over 9x9 grid points, and upsampled from the original 1 h time resolution to 15 min via clear sky index (see Equation 1) interpolation.

4. Description of the probabilistic methods

Let us recall that all probabilistic methods depicted in this work generate quantile forecasts with the probability levels given in Table 2. However, each participant utilizes their own method to produce this set of quantiles (listed and briefly described above in Table 1).

Three classes of approaches are employed in this benchmark. The first class extends to the probabilistic domain a framework traditionally used to produce deterministic CMV-based forecasts, mainly by adding Gaussian noise to its inputs (similarly to a Monte Carlo approach). That is the case of Mines Paris OIE. The second class corresponds to a two-step approach where a deterministic forecast is generated and then used as input by a statistical technique to generate the quantile forecasts. This is the case of participants PIMENT and ISE. Conversely, the third class produces directly (in one step) the quantile forecasts from a set of input variables. This is the case of participants UU and LES.

Finally, in the field of solar forecasting, it is a standard procedure to detrend the Global Horizontal Irradiance (GHI) time series due to its non-stationary nature, characterized by daily cycles and annual seasonalities (Lauret et al., 2022). This detrending process involves utilizing the output of a clear sky model. Specifically, a new deseasonalized series, known as the clear sky index k_c time series, is derived by employing the following data transformation

$$k_c = \frac{I}{I_c}, \tag{1}$$

where I is the measured global horizontal irradiance and I_c is the output of a clear sky model. All the proposed forecasting models described below make use of the clear sky index k_c time series. However, it should be noted that the choice of the clear sky model may vary depending on the participant.

4.1. Description of Mines Paris OIE model

This approach, proposed originally in Carrière et al. (2021), combines physical and statistical elements and leverages a standard satellite-based solar forecasting framework which is traditionally used for deterministic forecasting. Figure 1 gives an overview of the principle of the method.

First, a 25x25 grid with 0.04° resolution centered in the location of interest is defined. For each grid cell, satellite-derived time series for GHI and its clear-sky expectation (I_c) are obtained from the Copernicus Atmospheric Monitoring Services (CAMS) Radiation product (Qu et al., 2017b) using the pvlib Python interface (Jensen et al., 2023). Note that while the native resolution of this product depends on the distance to nadir, the CAMS Radiation product adjusts to any requested coordinates by means of interpolation. Based on this data, the corresponding clear-sky index k_c grid was calculated according Equation 1 and its spatial resolution was further increased by a factor of 3 through a 2D linear interpolation. Then, the CMV of each downscaled k_c cell is inferred using an Optical Flow technique following the work of Chow et al. (2015) and using an efficient method proposed by (Liu, 2009).

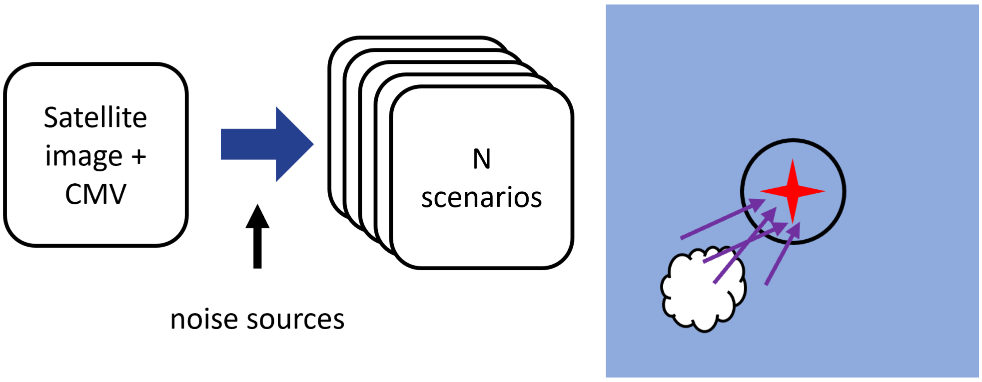
Deterministic CMV-based forecasts are calculated, for example, by extrapolating the k_c grid in space (according to the CMVs and the forecast horizon) and selecting the advected k_c value which is closest to the location of interest. In this approach, an Eulerian spatial extrapolation is considered, where the clouds are assumed to move in a straight-line trajectory.

Here, the probabilistic aspect is enabled by three elements: i) a Gaussian noise distribution relative to the CMV estimates, namely to the norm and direction of a vector; ii) a Gaussian noise distribution relative to the k_c estimates; and iii) a monitoring perimeter. The first two aim to describe the uncertainty associated with the estimation of the satellite-derived variables, whereas the third defines a distance threshold below which an extrapolated grid cell is considered a plausible forecast candidate. This allows the generation of a set of plausible advection scenarios. The third corresponds to a distance threshold below which an extrapolated k_c pixel is considered a plausible forecast candidate. Thus, the combination of the viable candidates from all the generated scenarios constitutes an ensemble of k_c forecasts from which an empirical CDF is built.

Finally, the k_c CDF is converted back to GHI by multiplying it with the I_c obtained from CAMS Radiation (see Equation 1), which considers the McClear clear-sky model, which accounts for water vapor and aerosol effects (Lefèvre et al., 2013). To mitigate potential calibration issues, the forecasts are post-processed by first considering the baseline model CSD-CLIM (Le Gal La Salle et al., 2021) for defining the bounding quantiles (Q0 and Q100) and then implementing a quantile mapping approach for calibration, adjusting the effective probability rate of each quantile to its theoretical rate according to the training data.

More details on this implementation can be found in Carrière et al. (2021), including the parameters assumed for the considered sources of Gaussian noise. Concerning the model implementation, a few remarks:

- It is only tested for the locations covered by the CAMS Radiation service (i.e., covered by the Meteosat Second Generation geostationary satellite);
- It is only tested for the horizon range between 15 minutes and 2 hours;
- In two situations, the baseline approach proposed by (Le Gal La Salle et al., 2021) is considered instead of the CMV-based one: i) when for a given day, forecast time, and horizon, there is yet no available satellite image; ii) when this approach leads to less than 50 k_c candidates, which possibly compromises the representativeness of the produced ensemble.



(a) (b)

Figure 1: Mechanisms leveraging the probabilistic CMV approach: (a) the generation of advection scenarios by inputting noise to the base CMV; (b) the consideration of all advected grid cells that fall inside a monitoring perimeter.

4.2. Description of PIMENT model

The PIMENT model is based on a parametric approach commonly used in the financial domain. It combines a AutoRegressive Moving Average model (ARMA) and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH), which successively generate a point forecast and then its associated uncertainty. This combination in the field of solar energy has been first introduced by David et al. (2016). The model is applied to the clear sky index k_c time series with the McClear model (Lefèvre et al., 2013) selected as the clear-sky model .

The AutoRegressive Moving Average model (ARMA) stands as a prevalent and widely-applied method in time series prediction. Its extensive utilization in forecasting renewable energy has underscored its competitive edge, owed largely to its parsimonious nature. Notably, its application spectrum encompasses the forecasting of solar irradiance among other domains (Bacher et al., 2009; David et al., 2018). A general formulation of an ARMA(p,q) model with p autoregressive (AR) terms and q moving average (MA) terms is given by (Tsay, 2010). Its application to the h-ahead forecast of a variable y is given by the following equation

$$\hat{y}_{t+h} = \alpha_0 + \sum_{i=1}^{p} \alpha_i \times y_{t-i+1} + \sum_{j=1}^{q} \beta_j \times \epsilon_{t-j+1}, \tag{2}$$

with $h = 1, 2, \cdots$ the forecast horizon and $\alpha_0, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q$ the coefficients to be estimated. The error term ϵ is the difference between the previous forecasts and observations as defined in the following equation:

$$\epsilon_t = \hat{y}_t - y_t. \tag{3}$$

The ARCH (AutoRegressive Conditional Heteroskedasticity) models, introduced by Engle (Engle, 1982), is used to model the variance of time series in the financial domain. These models are particularly efficient to predict changes in variance over the time, for instance the error of point forecast generated with an ARMA model (Bollerslev, 1986). PIMENT applied the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model proposed by Bollerslev (1986), which gives a more parsimonious formulation than the simple ARCH model. In GARCH models, the conditional variance is a linear function of lagged squared error terms and also lagged conditional variance terms (Taylor, 2004). The general formulation of a GARCH(r,s) model, with r error terms, s conditional variance terms and an horizon of forecast h, is given by:

$$\hat{y}_{t+h} = \hat{\sigma}_{t+h} \times \varepsilon_t, \tag{4}$$

with ε an uniformly distributed random variable with a null mean and a unitary variance, and $\hat{\sigma}_{t+h}$ the predicted standard deviation given by

$$\hat{\sigma}^2_{t+h} = \gamma_0 + \sum_{i=1}^{r} \gamma_i \times \epsilon_{t-i+1} + \sum_{j=1}^{s} \delta_j \times \sigma^2_{t-j+1}. \tag{5}$$

As for the ARMA models, $\gamma_0, \gamma_1, ..., \gamma_r, \delta_1, ..., \delta_s$ are the coefficients to be estimated. There are numerous methods to estimate these coefficients. The two most widely used are the least squares (LS) and the Maximum Likelihood Estimation (MLE) methods. Here, we propose to implement the Recursive Least Squares (RLS) method, which is a variation of the LS method. This method reduces the computational cost with the coefficient of the model being updated in real-time as new data become available. The RLS method is very efficient in an operational context where forecast have to be timely delivered.

To determine the lag parameters p, q of the ARMA model, PIMENT ran the model on the training year for different combinations of the lag parameters with values varying from 1 to 10. The best combination is the one that minimizes the RMSE of the point forecast. For the probabilistic part, PIMENT used a GARCH(1,1), which is appropriate for the error times series of the point forecast.

4.3. Description of Fraunhofer ISE model

The approach by Fraunhofer ISE consists of two steps. In a first step, blended point forecasts are derived from different input data (Section 4.3.1). In a second step, quantile forecasts are generated from these blended point forecasts using the Analog Ensemble (AnEn) method (Section 4.3.2).

4.3.1. Blended forecasts by Fraunhofer ISE

Deriving blended forecasts from several distinct input forecasts using statistical or machine learning methods is a common approach in deterministic forecasting (see e.g. Lorenz et al., 2021).

Here, GHI forecasts are generated by blending three different types of forecasts

- a persistence forecast,
- a satellite-based forecast,
- and the deterministic ECMWF IFS forecast (see Section 3.3).

The persistence forecast is created by deriving a clear sky index from the latest measurement, which is then extrapolated into the future. Fraunhofer ISE employs the clear-sky model from Dumortier (1995) and the turbidity model from Bourges (1992) to compute clear sky irradiance and the clear sky index, not only for persistence, but for all modeling steps described in this section. The satellite-based forecasts are based on CMVs derived from MSG satellite images (see section 3.2) following Kühnert et al. (2013). The CMVs are computed using a block matching algorithm. Future images are obtained by repetitive application of the cloud motion vectors. Finally, smoothing filters depending on forecast lead times are applied to the future images.

The three different forecast types are blended using a set of linear regression models, fitted for each forecast horizon and time of the day. This allows to adjust the regression weights to the varying performance of the different input forecasts in dependence of the forecast horizon (see Figure 2). The resulting blended forecast can be written as

$$I_{\text{blend}}^{h,\tau} = c_{\text{pers}}^{h,\tau} \cdot I_{\text{pers}}^{h,\tau} + c_{\text{cmv}}^{h,\tau} \cdot I_{\text{cmv}}^{h,\tau} + c_{\text{nwp}}^{h,\tau} \cdot I_{\text{nwp}}^{h,\tau}, \tag{6}$$

where I_X are the GHI of the blended, persistence, CMV or NWP forecast, c_X are the corresponding regression weights, and h and τ denote the index of the forecast horizon and time of the day respectively. When determining the regression weights $c_X^{h,\tau}$ for a forecast horizon h, data from $h \pm 2$ horizons were included in the fit, to enable the generation of a smooth blended output forecast from the different linear models. The regression weights are trained for the year 2017 and applied to generate the forecasts for 2018. With multiple sites of observation and forecast data, either one set of the blending parameters for all sites or separate sets for each site can be derived. It was decided for the single site training, to obtain close-to-bias-free blended forecasts for each site.

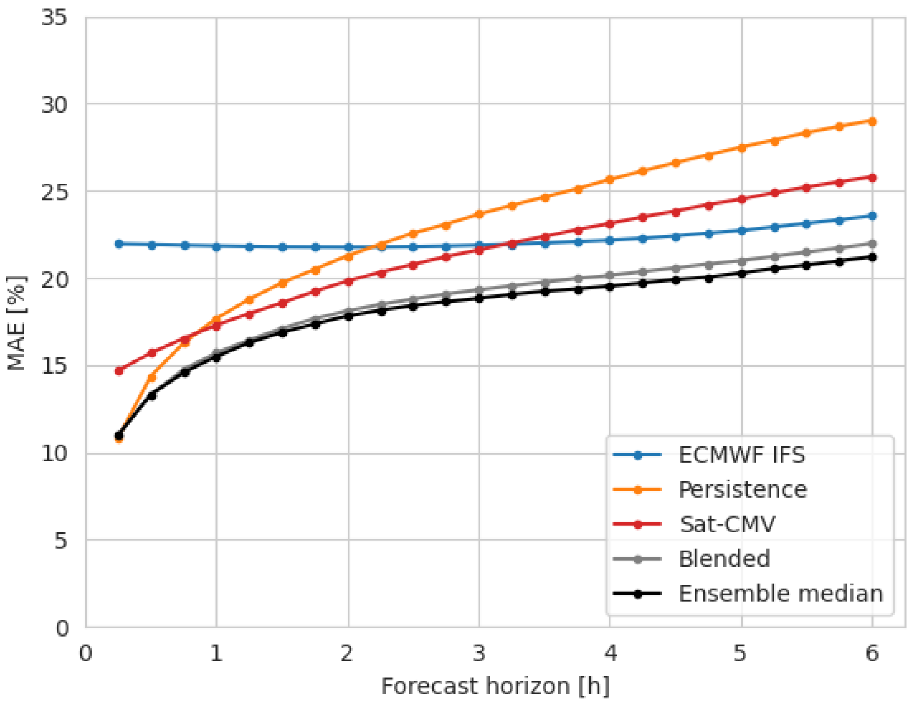


Figure 2: Relative MAE vs. forecast horizon, computed for all eight sites together for the test year of 2018. The MAE is normalized with the average GHI, which differs between 351 W/m^2 and 397 W/m^2 depending on filtering. Filtering such that all component forecasts for blending are available; the deviations of the ensemble median here compared to the benchmark (e.g. Figure 4) result from differences in the filtering.

When generating the blended forecasts, missing data is handled in the following way: If one of the component forecasts is missing, a simple mean of the remaining component forecasts is calculated instead of using the regression weights. For horizons or times of the day for which no regression weights could be determined, NWP data is returned. It should be noted here that persistence as well as satellite based forecasts can only be calculated after sunrise and before sunset, which impacts availability of these forecasts in the early morning hours depending also on the forecast horizon. For example, if the earliest satellite based forecast could be calculated at 7:00 am, forecasts for horizons of 4 hours ahead are available only from 11:00 am onwards.

The performance of the different forecasts, quantified in terms of MAE, is shown as a function of forecast horizon in Figure 2. Here the different availabilities of the forecasts discussed above have to be considered. For reasons of comparability, calculation of MAE includes only data points for which all displayed forecasts are available. The Figure shows that the persistence forecasts are best performing for short horizons up to 45 minutes, satellite-based forecasts are best performing for intermediate horizons up to about 3 hours, and the NWP performs best for even larger horizons. The blended forecasts always exhibit a lower MAE than any of the individual input forecasts. They form the basis to derive probabilistic forecasts in a next step.

4.3.2. Analog Ensemble

The Analog Ensemble (AnEn) is a non-parametric ensemble prediction method (Delle Monache et al., 2013; Junk et al., 2015). The method is based on the evaluation of historic observations and deterministic forecasts. Past forecasts are compared to the current forecast and the observations corresponding to the most similar forecast situations, called analogs, then form an ensemble of possible future values. The quantities, which are used to measure similarity, are called predictors. Alessandrini et al. (2015) constructed an AnEn for PV power prediction and used the GHI, solar elevation and azimuth, cloud cover, and ambient temperature (T2M) from the deterministic ECMWF IFS forecast as predictors. Here, an AnEn to predict the GHI instead of the PV power is created. Furthermore, just one predictor quantity is used for the AnEn, namely the forecasted clear sky index.

To set up the Analog Ensemble, in a first step, the clear sky index values of the blended forecasts and the measurements are computed. Situations with measurement-based clear sky index values above 1.2 are excluded. To identify the analogs for a current forecast value with a forecast horizon h_0, similarity to past forecasts is evaluated using the Euclidean distance of the forecast values of five horizons $\{h_0 - 30\text{min}, h_0 - 15\text{min}, h_0, h_0 + 15\text{min}, h_0 + 30\text{min}\}$, centered around h_0. This window helps to reduce fluctuations in the distance measure and to improve meteorological similarity between situations. The measurement-based clear sky indices corresponding to the 40 most similar situations are taken to form the AnEn.

Similar to a k-nearest-neighbor regression, the AnEn has no explicit training phase, instead the analogs are selected from a search space at the time a prediction is made. The analog search is performed separately for each forecast horizon and each time of the day, reflecting different uncertainties in dependence of the forecast horizon (see Fig. 2) and the time of the day. The search space for the analogs is composed using a rolling window of the last 180 days and integrating all eight European sites together, resulting in 1440 historic data points from which the 40 analogs are selected. Integrating the different sites increases the search space and thus the reliability of the ensemble. It is made possible by using the clear sky index as a predictor instead of GHI and close-to-bias-free forecasts for the different sites. The ensemble of clear sky indices is, then, transformed back to GHI values by multiplication with clear sky irradiance. The quantiles are finally obtained by a linear interpolation of the ensemble members, see method 7 of (Hyndman and Fan, 1996).

4.4. Description of Utrecht University (UU) model

Quantile regression forest (QRF) is a nonlinear ensemble model that is based on the random forest regression (RF) model (Koenker, 2005; Meinshausen and Ridgeway, 2006). Similar to a RF model, QRF is made up of a predefined set of decision trees that exist of a number of layers (t_n) and decision nodes ($2t_n$). The trees are constructed independently from each other by considering bootstrap samples from the training dataset in the training stage. The nodes are constructed by selecting a random subset of the predictor variables and optimizing the decision node on a preset loss function, e.g the mean squared error. In contrast to a RF model, the QRF model predicts a conditional distribution function (or weighted distribution of observations). Hence, given a set of predictor variables, each tree in the QRF model predicts the conditional quantiles of the target variable, i.e., GHI. Finally, a post-processing step is added in which each quantile value is set to be equal or higher than zero and equal or lower than the clear sky irradiance estimate (I_c).

The UU forecast model follows the approach described in Visser et al. (2023) to find the optimal hyperparameter settings. Hence, the training set is split into several training and validation subsets, using k-fold cross-validation with $k = 8$ (Raschka and Mirjalili, 2019). Once the optimal hyperparameter settings are found, the QRF model is trained considering the entire training set, i.e., one year (2017) and then applied to predict the GHI for the test year (2018).

The QRF model considered in this study operates endogenously. This implies that the model only relies on historical observations of the target variable, i.e., the GHI (I), as well as variables that are available at any time, i.e., the clear sky irradiance I_c. Using I and I_c, we construct two additional variables: the clear-sky index k_c (see Equation 1) and the expected GHI using a clear sky-based smart persistence model, similar as discussed in Section 4.3.

From these four main variables, a large set of predictor variables can be generated by simply considering a multitude of lagged values. In this study, UU optimizes the number of historical values by means of an iterative process. Hence, starting with a base model, at each iteration one lagged value is added, where after an evaluation if the addition leads to a significant performance improvement is made. The final set of variables considers 18 predictor variables, including: the previous eleven GHI measurements $(I(t), (...), I(t-10))$, the clear-sky irradiance $(I_c(t+k))$, the clear-sky index $(k_c(t))$ and the persistence forecast considering the three most recent irradiance measurements $(I_{pers}(t+k,t), I_{pers}(t+k,t-1), I_{pers}(t+k,t-2))$.

4.5. Description of Udelar LES model

The LES forecast is an adaptation of the methodology proposed by Alonso-Suárez et al. (2020). This approach utilizes lagged ground measurements and geostationary satellite data as inputs for a Linear Quantile Regression (LQR) model, as described by Koenker and Bassett (1978). The LQR model is used to predict quantiles of the clear-sky index (k_c). These quantiles are then converted to quantiles of the Global Horizontal Irradiance (GHI) using the McClear clear-sky model (see Equation 1). While the mathematical formulation is relatively simple, the crucial aspect lies in the predictors' selection. The forecasting model incorporates the present time and the six preceding k_c values, along with four other predictors derived from either the past k_c values or a satellite space cell that surrounds the specific location. The first additional predictor is the local short-term variability (σ_c), which is calculated as the standard deviation of the last six changes in k_c. For more in-depth information on how σ_c is calculated, please refer to Alonso-Suárez et al. (2020). The remaining three predictors are derived from the Heliosat-4 satellite estimates in a 25×25 px space cell provided by the Copernicus Atmosphere Monitoring Service (CAMS). By employing the McClear model, the clear-sky index is calculated for each pixel in the satellite cell. From this index, the average, standard deviation, and cloud coverage are computed and utilized as input variables in the LQR method. The cloud coverage is estimated as the fraction of pixels in the cell with $k_c < 0.85$. In summary, the inputs for the LQR model are six lagged k_c values, the local short-term variability, and four variables related to the current time. These current time variables consist of the measured k_c, the space average and standard deviation of the satellite-derived k_c, and the satellite-estimated cloud coverage within the cell. Different LQR parameters are trained for each site, forecast horizon, and quantile.

5. Proposed evaluation framework

Visual diagnostic tools and quantitative scores are used to assess the quality of the probabilistic forecasts (i.e. the correspondence between ground truth and the forecasts). Diagnostic tools are used to visually assess the quality of probabilistic forecasts, while numerical scores are used to quantify the skills of a forecasting system and to rank competing prediction methods.

In this study, following the recommendation of Lauret et al. (2019), we adopt the CRPS as the scoring rule to assess the overall performance of the forecasting method. Moreover, to gain a deeper understanding of the forecast skill of each forecasting method, we further decompose the CRPS into two components: reliability and resolution. In case of predictive distributions summarized by discrete quantile forecasts, Lauret et al. (2019) proposed specific formulae to compute the CRPS and its related decomposition. The interested reader is referred to (Lauret et al., 2019) for more details regarding the computation of this CRPS decomposition. Another useful assessment is whether a prediction system outperforms a trivial baseline model. To this end, we compute the CRPS skill score with the climatological model CSD-CLIM (Le Gal La Salle et al., 2021) as the reference model.

Finally, in this work, we use reliability diagrams to visually evaluate reliability of the different forecasts.

5.1. Visual assessment with reliability diagrams

The reliability diagram serves as a graphical tool for assessing the reliability of probabilistic forecasts. In this paper, we follow the methodology established by (Pinson et al., 2010), which is tailored for predictive distributions summarized by quantile forecasts. Specifically, quantile forecasts are considered reliable when their stated probabilities match the observed proportions. In essence, over a sufficiently large evaluation dataset, the disparity between observed and nominal probabilities should be minimized (Pinson et al., 2010).

One of the advantages of this representation is that it allows deviations from perfect reliability, represented by the diagonal line, to be readily visualized (Pinson et al., 2010). However, it's important to acknowledge that due to the finite sample of observation/forecast pairs and potential serial correlation in the sequence of forecasts and observations, observed proportions may not align precisely along the diagonal, even if the forecasts are perfectly reliable (Pinson et al., 2010). In other words, reliability diagrams can sometimes be misleading because even for perfectly reliable forecasts, deviations from the ideal diagonal case can be observed.

To address the limitations arising from the finite number of observation/forecast pairs, (Bröcker and Smith, 2007) introduced reliability diagrams with consistency bars. Additionally, Pinson et al. (2010) has proposed consistency bars that consider the combined effects of serial correlation and limited data. In this work, consistency bars are calculated according to (Pinson et al., 2010). When interpreting reliability diagrams with consistency bars, it becomes clear that one cannot reject the hypothesis of the quantile forecasts being perfectly reliable if the observed proportions fall within the consistency bars. In practice, incorporating consistency bars into reliability diagrams can provide additional support for the user's, possibly subjective, assessment of the reliability of the different models.

5.2. Continuous Rank Probability Score (CRPS) and its decomposition

The Continuous Ranked Probability Score (CRPS) is a numerical score that quantifies the difference between the predicted and observed cumulative distribution functions (CDF) (Hersbach, 2000). It is formulated as follows:

$$\text{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left[\hat{F}^{i}_{fcst}(x) - F^{i}_{x_{obs}}(x)\right]^{2} dx, \tag{7}$$

where $\hat{F}_{fcst}(x)$ is the predictive CDF of the variable of interest x (e.g. GHI) and $F_{x_{obs}}(x)$ is a cumulative-probability step function that jumps from 0 to 1 at the point where the forecast variable x equals the observation x_0 (i.e. $F_{x_{obs}}(x) = 1_{\{x \geq x_{obs}\}}$). The squared difference between the two CDFs is averaged over the N observation/forecast pairs.

The CRPS score is designed to reward forecasts that concentrate their probabilities around the step function located at the observed value, promoting accuracy and precision in forecast predictions (Wilks, 2009). Put simply, the CRPS serves as a penalty for both insufficient resolution in predictive distributions and biased forecasts. It is worth noting that the CRPS is oriented negatively, meaning that smaller values indicate better performance, and it has the same unit as the forecast variable.

As previously mentioned and in accordance with its nature as a proper scoring rule (Gneiting and Raftery, 2007), the CRPS can be decomposed into two fundamental aspects of probabilistic forecasts: reliability and resolution. This decomposition of the CRPS yields the following equation:

$$\text{CRPS} = \text{REL+UNC-RES}. \tag{8}$$

The reliability REL component of the CRPS provides an assessment of forecast biases, while the resolution RES component quantifies the improvement achieved by issuing case-dependent probability forecasts. The uncertainty component UNC, on the other hand, is inherent to the observations and cannot be influenced by the forecast system; it depends solely on the variability of the observed data (Wilks, 2009).

Given that the CRPS is negatively oriented, the objective of a forecast system is to minimize the reliability component as much as possible, while also maximising the resolution component. By employing this decomposition of the CRPS, a detailed evaluation of the forecast performance of different forecasting methods can be obtained.

Besides, in the case of deterministic forecasts, the CRPS reduces to the Mean Absolute Error (MAE). This characteristic enables a direct comparison between the performance of a probabilistic model and a deterministic one, or equivalently, it allows for assessing the additional value provided by a probabilistic approach (Ben Bouallègue, 2015). In this study, we calculate the Mean Absolute Error (MAE) of the forecast distribution's median to evaluate the extent to which the probabilistic approach enhances (or fails to enhance) the overall quality of the forecasts over its deterministic counterpart.

5.3. The CSD-CLIM model and the associated CRPS skill score

Probabilistic scores do not allow fair comparisons between different sites or datasets. To do so, it is a good practice to consider the relative performance against reference models (Gneiting et al., 2023) through skill scores. Over the past years, several benchmark models for probabilistic forecasting were introduced in the literature (Doubleday et al., 2020; Gneiting et al., 2023). For this work, the CSD-CLIM model has been selected. For each site, the measurements of the training dataset are gathered according to a set of bins of clear-sky irradiance. Then, empirical CDFs are built independently for each bin. In the test period, the clear-sky model is used to select the appropriate bin and the associated forecasting CDF. Thus, the CSD-CLIM approach is climatological in the sense that it only uses historical data and a clear-sky model. More details about theory and implementation are available in (Le Gal La Salle et al., 2021). In this work, the McClear clear-sky model (Lefèvre et al., 2013) has been chosen with 30 clear-sky irradiance bins.

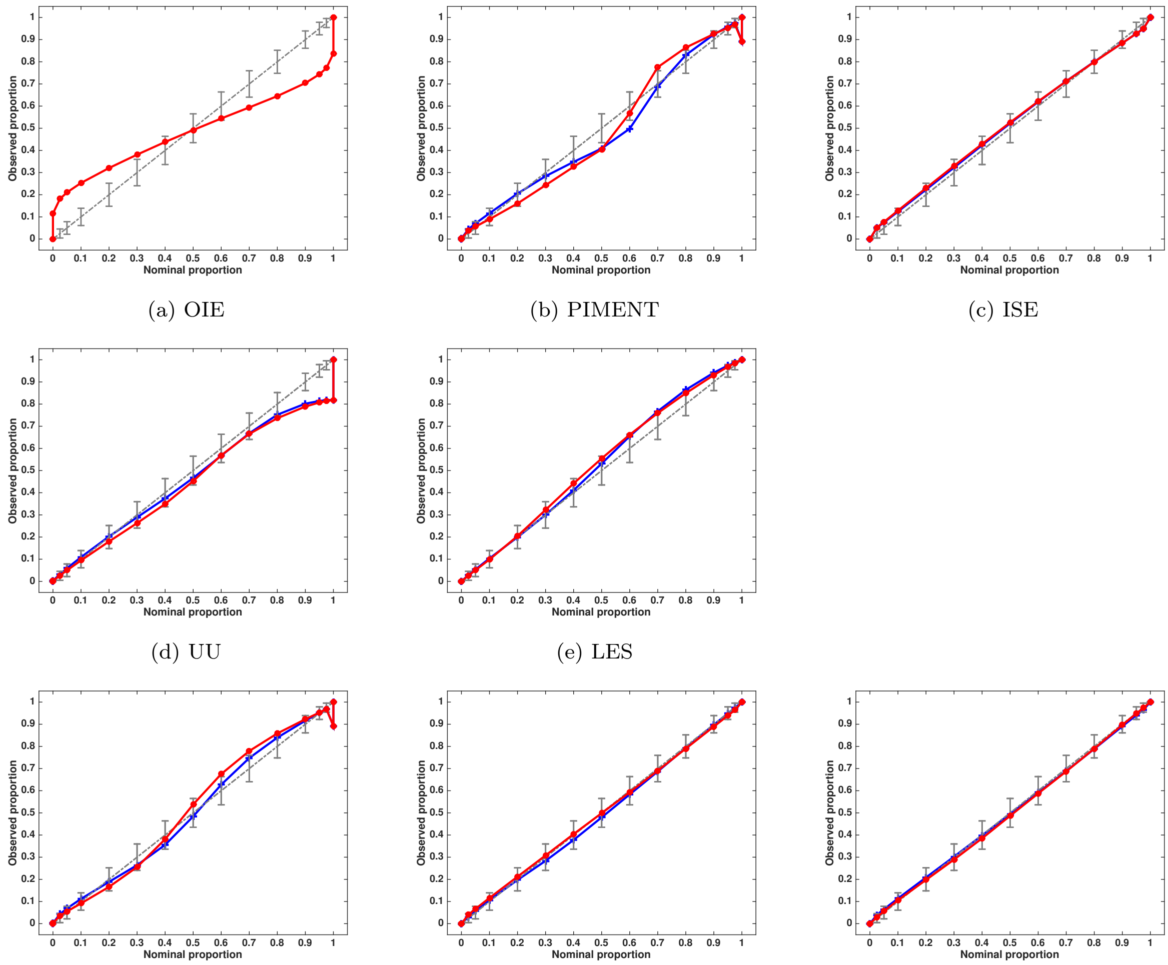
A skill score represents the degree of improvement of a forecasting model compared to the reference baseline model. The CRPS skill score (CRPSS) reads as

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{\mathrm{model}}}{\mathrm{CRPS}_{\mathrm{reference}}}. \tag{9}$$

6. Results

6.1. Reliability diagrams

Reliability diagrams related to each scenario (intra-hour or intra-day) for each participant are given in Figure 3, with each of the plots averaged for all sites. Note that the averaging for all sites is performed by aggregating GHI observations and forecasts for each site into two distinct time series. This procedure will be also used to calculate the overall CRPS results in section 6.2.1. Consistency bars for a 90% confidence level are individually computed for each nominal proportion. Note that comments related to Figures 3f, 3g and 3h will be made in Section 7.



(a) OIE (b) PIMENT (c) ISE

(d) UU (e) LES

(f) BLEND-GARCH (g) BLEND-LQR (h) BLEND-QRF

Figure 3: Reliability diagrams for all sites averaged over all the forecast horizons related to each participant. Consistency bars for a 90% confidence level around the ideal line are individually computed for each nominal proportion. The red curves stand for intra-hour forecasts while the blue one are for intra-day forecasts.

Irrespective of the scenario (intra-hour or intra-day), the visual analysis shows that only probabilistic forecasts derived from ISE (Figure 3c) possibly has a high reliability. All the other forecasts are possibly non reliable namely those generated by OIE (Figure 3a), PIMENT (Figure 3b), UU (Figure 3d) and to a lesser extent LES (Figure 3e). More specifically, forecasts provided by OIE appear to be clearly non reliable and forecasts issued by LES, UU and PIMENT experience high deviations from the ideal line for high nominal proportions.

For the ISE forecasts for which the observed proportions lie within the consistency bars, while this does not confirm the perfect reliability of the quantile forecasts, it also does not allow us to confidently assert their lack of reliability at a 10% significance level.

6.2. CRPS and its decomposition

6.2.1. Overall results for all sites

Following Yang et al. (2020), Table 4 and Table 5 report the overall results (i.e. computed from the aggregation of forecasts and data of all sites) obtained by each metric in terms of "mean ± standard deviation". More precisely, for intra-hour forecasting, the mean and standard deviation are computed from the 8 values of the metric corresponding to the 8 horizons while for intra-day forecasting, the mean and standard deviation are computed from the 16 values related to the 16 horizons (see Table 2). We recall here that intra-hour forecasts correspond to 15-120 min-ahead forecasting at 15mins timesteps while the intra-day forecasts are for 120-360 min-ahead forecasting at 15mins timesteps. Also, in this section, we will comment on the results obtained by the five participants with their original proposed method (see Table 1). The methods BLEND-GARCH, BLEND-LQR and BLEND-QRF will be presented and discussed later in Section 7.

Regarding intra-hour forecasts (see Table 4), the best performer is ISE regardless of the metric while the worst one is PIMENT. In terms of skill score, the forecast skill of ISE is (in average) 46.6% while PIMENT exhibits a skill score of 32.9%. It appears clearly that the better performance of ISE originates from its better resolution and reliability. In line with the reliability diagrams of Figure 3a, the CMV-based probabilistic method of OIE leads to poor results notably in terms of reliability. More generally, the quantitative reliability (REL) component of the CRPS confirms the visual diagnosis provided by the reliability diagrams in Figure 3. Finally, it should be noted that the linear LQR method proposed by LES, fed with ground and satellite data, achieves similar results to the nonlinear QRF method of UU, which uses only ground data.

As shown by Table 5, for intra-day forecasts, the same comments made above for intra-hour forecasts still hold. However, except for ISE, the only participant integrating NWP forecasts, one can observe a strong decrease in forecast skill particularly for PIMENT for which a decrease of 25 points in the mean skill score is noted. Similar to deterministic forecasts (see Figure 2), with increasing forecast horizon, the positive impact of integrating NWP forecasts in the modelling process is clearly demonstrated.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	53.8 ± 5.6	34.6 ± 6.7	4.5 ± 0.3	101.8 ± 5.4	67.9 ± 8.3
PIMENT	55.2 ± 10.3	32.9 ± 12.5	3.9 ± 0.7	99.7 ± 9.6	72.9 ± 15.4
ISE	43.9 ± 6.6	46.6 ± 8.1	0.8 ± 0.2	107.9 ± 6.5	59.4 ± 9.4
UU	48.3 ± 8.4	41.3 ± 10.2	1.1 ± 0.2	103.8 ± 8.3	64.1 ± 12.2
LES	47.1 ± 8.6	42.8 ± 10.4	1.5 ± 0.3	105.5 ± 8.3	62.7 ± 12.1
BLEND-GARCH	47.7 ± 6.9	42.1 ± 8.3	2.0 ± 0.3	105.4 ± 6.6	61.6 ± 9.2
BLEND-LQR	47.3 ± 6.4	42.6 ± 7.8	2.1 ± 0.2	105.9 ± 6.2	61.4 ± 9.2
BLEND-QRF	44.2 ± 5.9	46.3 ± 7.2	1.1 ± 0.1	107.9 ± 5.8	59.5 ± 8.6

Table 4: Intra-hour forecasts overall results. For each method, the metrics are presented as "mean ± standard deviation," calculated over all forecast horizons. For these overall results, the CRPS of the CSD-CLIM is 82.3 W.m^{-2} and the uncertainty component UNC of the CRPS is 150.8 W.m^{-2}

.

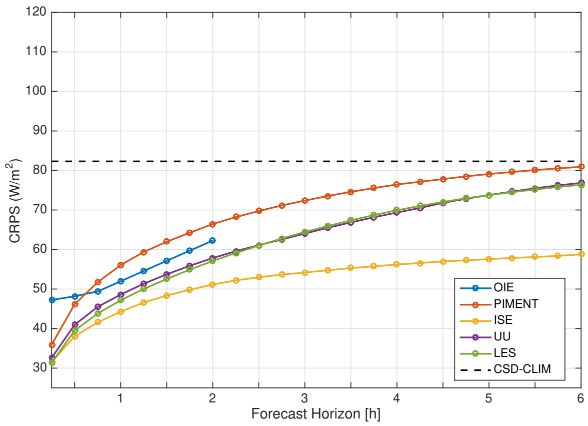
Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	76.0 ± 4.0	7.7 ± 4.9	4.7 ± 0.1	79.8 ± 3.9	105.3 ± 6.5
ISE	56.0 ± 2.0	31.9 ± 2.5	1.3 ± 0.2	96.3 ± 1.9	76.6 ± 2.7
UU	69.3 ± 5.6	15.8 ± 6.8	1.6 ± 0.2	83.3 ± 5.4	96.3 ± 8.9
LES	69.4 ± 5.5	15.6 ± 6.7	2.7 ± 0.4	84.3 ± 5.1	94.5 ± 8.3
BLEND-GARCH	60.4 ± 2.2	26.6 ± 2.6	2.9 ± 0.3	93.6 ± 1.8	79.4 ± 3.0
BLEND-LQR	59.2 ± 2.0	28.0 ± 2.4	3.1 ± 0.4	94.9 ± 1.6	79.1 ± 2.9
BLEND-QRF	55.5 ± 2.0	32.5 ± 2.5	1.7 ± 0.2	97.2 ± 1.8	75.7 ± 2.8

Table 5: Intra-day forecasts overall results. For each method, the metrics are presented as "mean ± standard deviation," calculated over all forecast horizons. OIE has "NA" values since OIE method is limited to intra-hour forecasting.

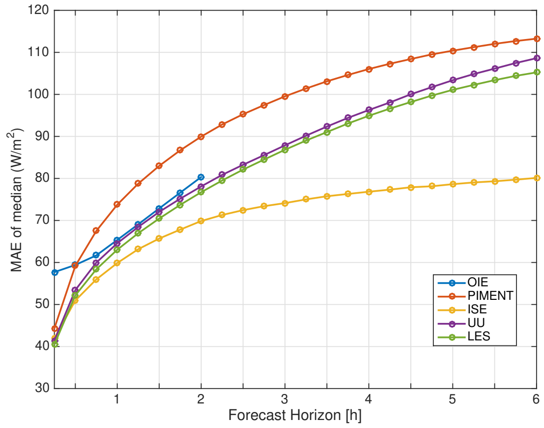
We complement the above quantitative metrics analysis summarized for intra-day and day-ahead by plotting over all the forecast horizons the numerical scores selected for this benchmarking exercise namely CRPS, CRPS reliability, CRPS resolution, MAE and CRPS skill score (see Figure 4). It should also be noted that we deliberately use the same Y-scale for the CRPS and MAE plots to emphasize the improvement in quality brought by the probabilistic approach. Indeed, as shown by Figures 4a and 4b, the CRPS (i.e., the MAE) of the median of the predictive distribution is clearly worse than the CRPS of the entire predictive distribution.

As shown by all the plots, the highest overall skill of the ISE forecasts is clearly demonstrated irrespective of the forecast horizon. In terms of CRPS skill score, over the whole range of forecast horizons, the best performer is ISE with skill scores between 60% and almost 30% while PIMENT forecasts lead to the worst forecasting results with a CRPS skill score ranging between 55% and almost 0%

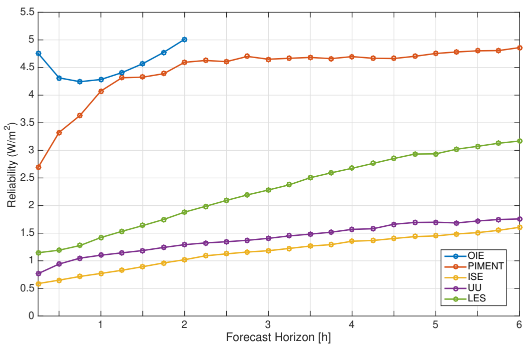
Again, the specific method developed by OIE does not outperform the other methods. The forecasts issued by this technique are clearly non reliable and confirms the visual inspection of the related reliability diagram.



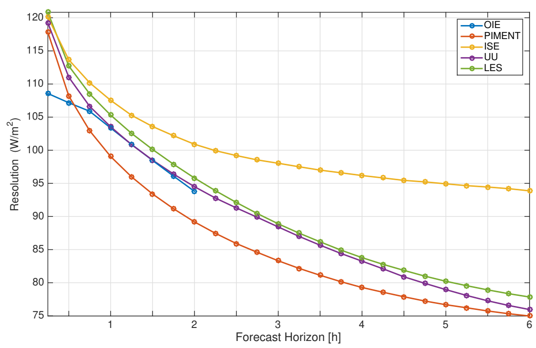
(a) CRPS



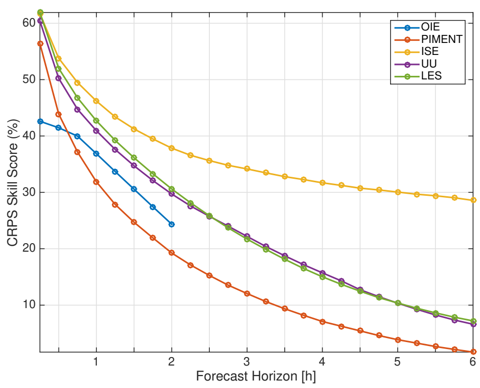
(b) MAE



(c) Reliability (REL)



(d) Resolution (RES)



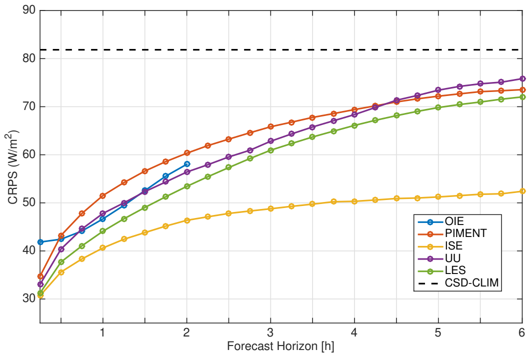
(e) CRPS skill score

Figure 4: CRPS and its associated decomposition for all stations. Average GHI for all sites is 382.4 W.m^{-2} and can be used to calculate the relative counterparts of the different metrics. Notice that the same Y-scale is used for the CRPS and MAE plots to highlight the improvement brought by the probabilistic approach.

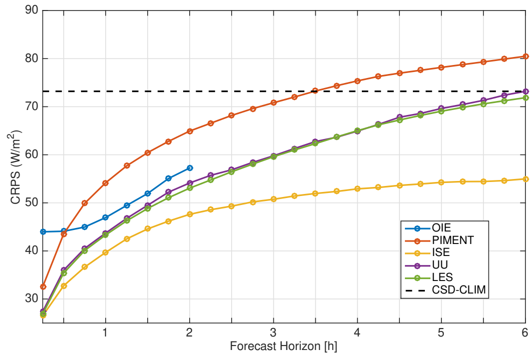
6.2.2. *Results for each site*

To gain a deeper insight in the performances of the different methods, Figure 5 focuses on the CRPS values obtained by each contestant for each of the eight sites. As shown, again, the best CRPS values are obtained by ISE whatever the location. Interestingly, the parametric model of PIMENT cannot beat the climatological model CSD-CLIM for some sites namely CAR, MIL and TAB for certain forecast horizons.

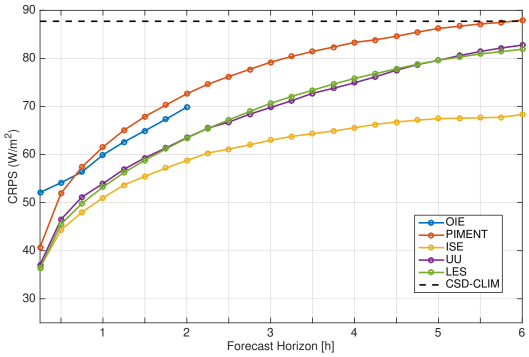
For the site TAB which is located in the South of Spain and corresponds to a semi-arid climate (see Koppen-Geiger classification provided in Table 3) and which experiences a high share of clear skies PIMENT is outerperformed by CSD-CLIM at 1h lead time while at 2h lead time, this is the case of LES and UU methods. Notice again the similar CRPS behavior of LES and UU models.



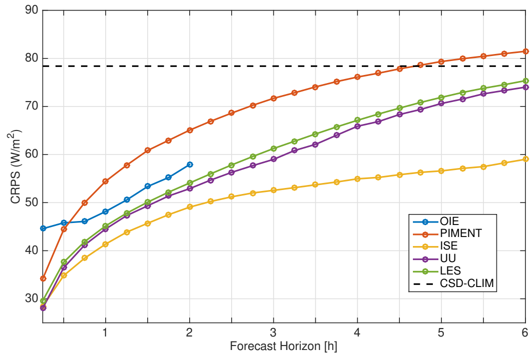
(a) CAB



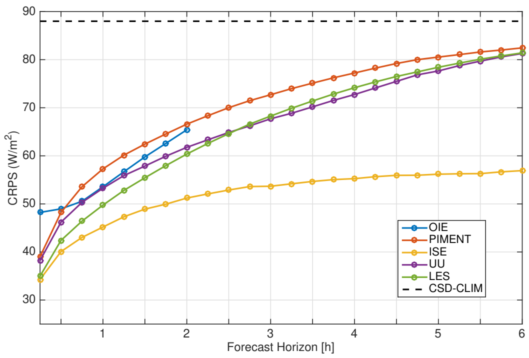
(b) CAR



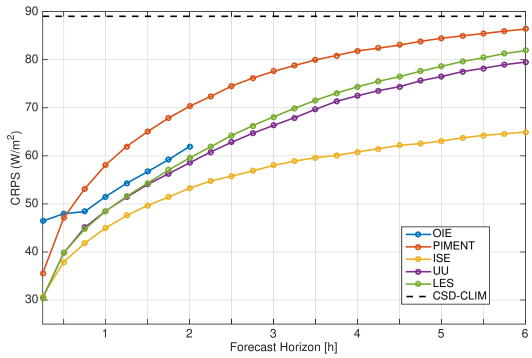
(c) CEN



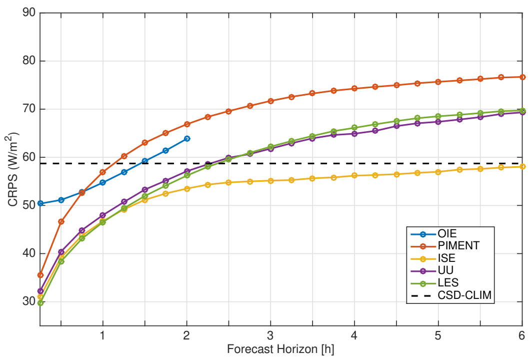
(d) MIL



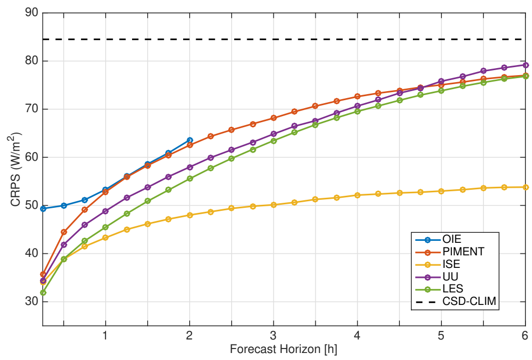
(e) PAL



(f) PAY



(g) TAB



(h) TOR

Figure 5: CRPS for the different locations. Table 3 lists the average GHI of each site that can be used to compute the relative CRPS

Finally, the interested reader is directed to Appendix 10 where results related to the 8 sites are tabulated.

7. Impact of a high quality point forecast on the skills of the probabilistic methods

In the previous Section 6, we observed that the ISE forecasting methodology clearly outperforms the other methods. We hypothesize that the skill of the ISE method comes from the blended point forecasts which are used by the Analog Ensemble technique and not necessarily by the Analog technique by itself.

Method	Forecasting technology	Input data
BLEND-GARCH	ISE blended point fcst + GARCH	ground data+SAT+NWP
BLEND-LQR	ISE blended point fcst + LQR technique	ground data+SAT+NWP
BLEND-QRF	ISE blended point fcst + QRF technique	ground data+SAT+NWP

Table 6: New proposed forecasting methods

In order to confirm our assumption, we use the same blended point forecasts as inputs to three other approaches to generate probabilistic forecasts, including the PIMENT parametric GARCH approach, applied to ARMA point forecasts before. This new forecast is denoted BLEND-GARCH. In addition, we designed two other models based respectively on the LQR and QRF technique that use as input the blended ISE forecasts. These 2 new models are denoted BLEND-LQR and BLEND-QRF. Table 6 lists the new combinations of the ISE blended point forecast with the different techniques employed to generate the quantile forecasts. The results of the newly proposed methods are listed in the last 3 lines of Table 4 and Table 5.

The combination BLEND-GARCH clearly improves the original PIMENT method for all the considered metrics. In particular, for the intra-hour scenario, the gain in average skill score is 9 points while for intra-day forecast, the gain in average forecast skill is 19 points. The decomposition of the CRPS permits to highlight the improvement in resolution brought by the BLEND-GARCH combination.

Regardless of the scenario (intra-hour or intra-day forecasts), we can state that the BLEND-GARCH and BLEND-LQR exhibit similar performances. The same statement is also valid for BLEND-QRF and ISE forecasts.

Specifically, for intra-day forecasts, the BLEND-QRF slightly outperforms the original ISE method in terms of forecast skill.

Again, for a better inspection of the results, we provide the visual display of the metrics. Here also, in Figure 6 the metrics are computed from the aggregation of forecasts and GHI data of all sites while Figure 7 plots the CRPS obtained on each site. Notice that, for sake of comparison, the metrics related to the ISE and PIMENT previous methods are also plotted.

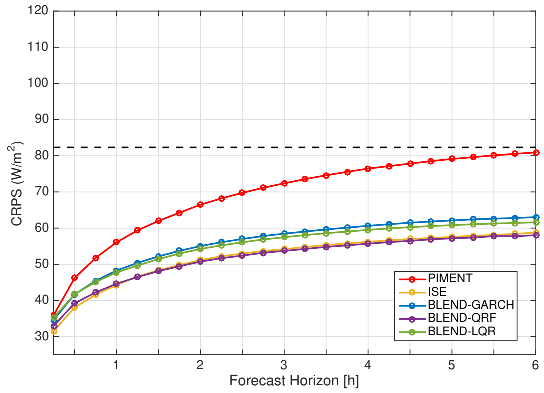
As shown by Figure 6, irrespective of the forecast horizon and metric, a clear improvement is brought by using the blended point forecasts. For instance, Figure 6e shows that the CRPS skill score of the new BLEND-GARCH now ranges from 58% to 22% instead of 56% to 1% obtained by the previous PIMENT method. The improvement is more pronounced at higher forecasting horizons and at the last forecast horizon the skill score of the previous PIMENT method based on ground measurements only is increased by almost 22 percentage points.

Also, combining the blended forecasts with techniques like LQR or QRF improves the skills of the probabilistic forecasts. From the decomposition of the CRPS into REL and RES, it appears that the BLEND-QRF slighlty outperforms the ISE method in terms of resolution. However, the ISE method is still the best performer in terms of reliability.

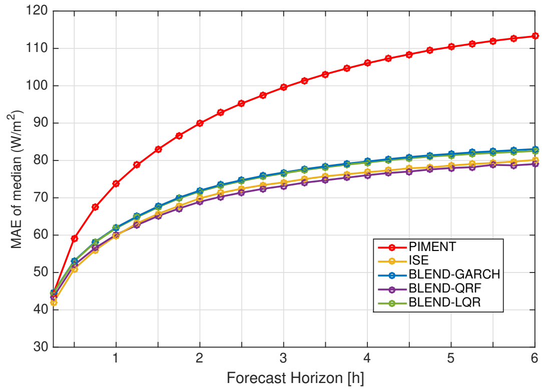
Figure 7 displays the CRPS obtained by the new combinations for each site under study. Now, all the new proposed forecasting techniques beat the CSD-CLIM model but for the site TAB (see Figure 7g) the CSD-CLIM outperforms the BLEND-GARCH from a 2h forecast horizon. Note that for the site TAB, the BLEND-QRF exhibits the best skill, a considerable improvement compared to the Analog Ensemble is found form two hour on-wards.

From the previous results, we can conclude that the use of the blended point forecasts of participant ISE improves substantially the PIMENT parametric approach forecasting models. Further, the improvements are slighty better when the blended point forecasts are inputted to a nonlinear machine learning technique such as QRF.

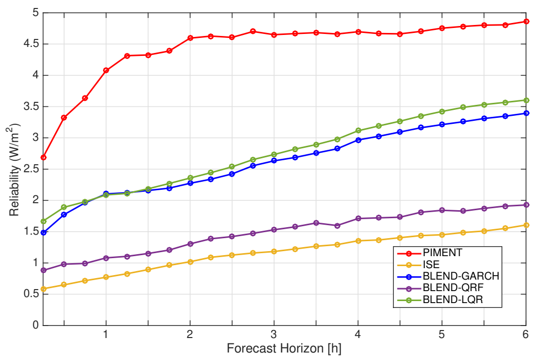
In terms of reliability diagrams, the situation is also clearly improved when one compares Figure 3f against 3b albeit it seems that BLEND-GARCH intra-hour forcecasts still suffer from a lack of reliability at high nominal proportions. Moreover, Figures 3g and 3h reveal that the new proposed method BLEND-LQR and BLEND-QRF generate reliable forecats.



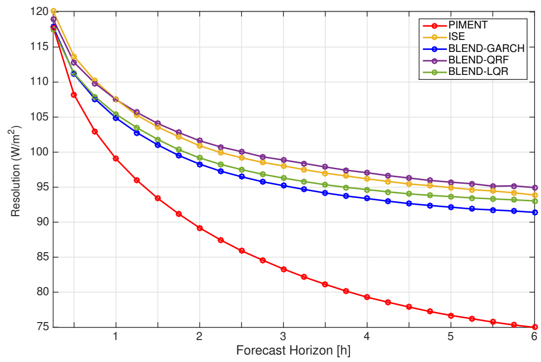
(a) CRPS



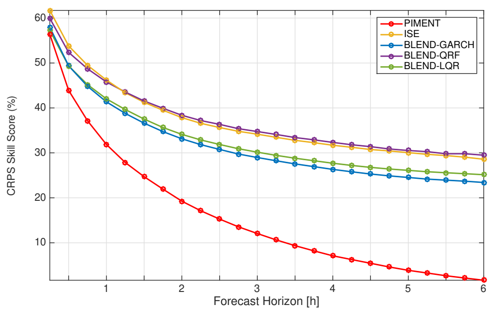
(b) MAE



(c) Reliability (REL)

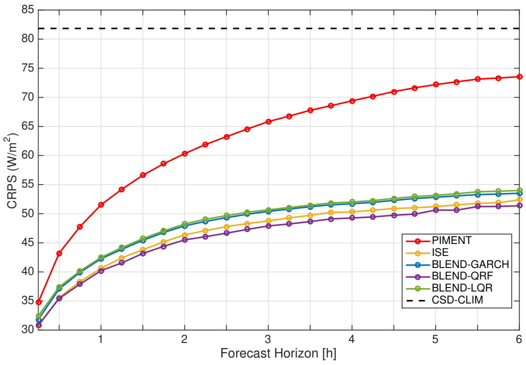


(d) Resolution (RES)

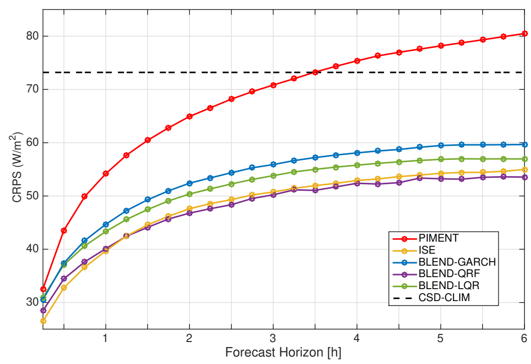


(e) CRPS skill score

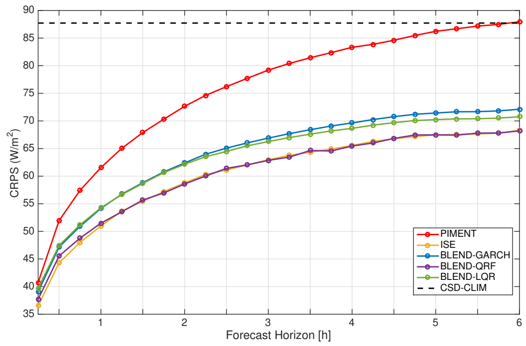
Figure 6: CRPS and its associated decomposition for all stations and for the new proposed models



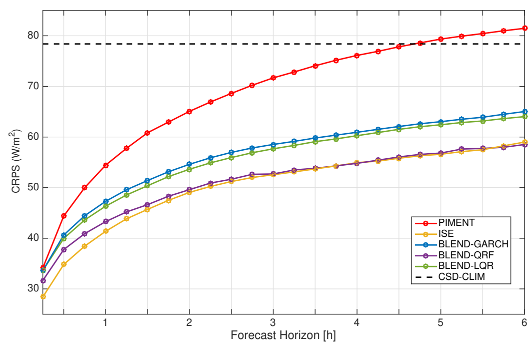
(a) CAB



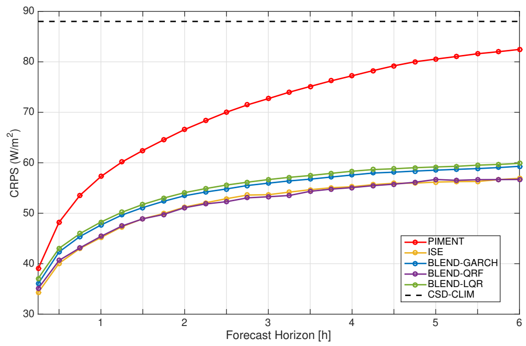
(b) CAR



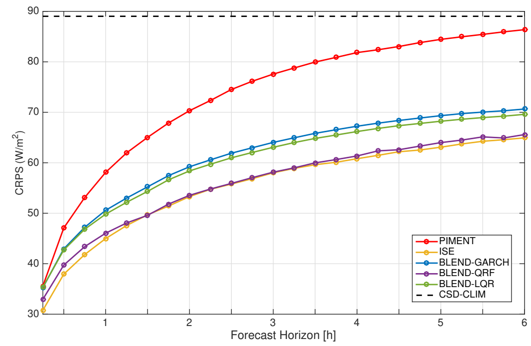
(c) CEN



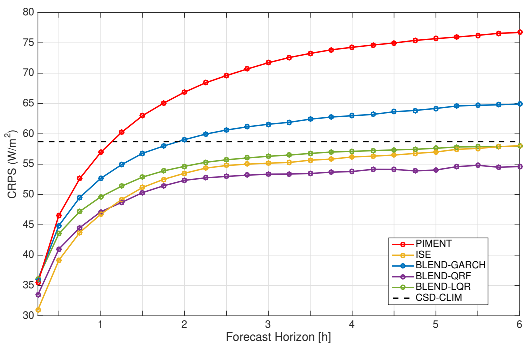
(d) MIL



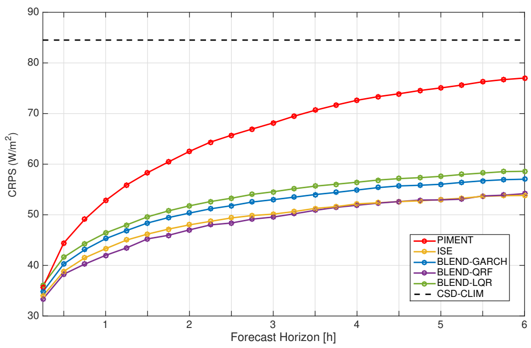
(e) PAL



(f) PAY



(g) TAB



(h) TOR

Figure 7: Improved CRPS by using the blending deterministic forecasts of Fraunhofer ISE for the different locations.

8. Summary and conclusion

A benchmarking procedure set up by a group of experts of the IEA PVPS Task 16 was implemented to assess the performance of intra-hour and intra-day probabilistic solar irradiance forecasts. This procedure was utilized to evaluate eight distinct forecasting algorithms.

In the initial stage, the benchmarking exercise involved evaluating probabilistic forecasts submitted directly by five different participants. This initial comparison of forecasts using different input data and methods revealed a significant variation in performance. Particularly, during this first step, a probabilistic forecast that utilized a blended point forecast outperformed the other methods. In a second analysis, to better understand the impact of input data versus methodology on forecast quality, we combined the well-performing blended point forecast from the first comparison with different probabilistic approaches.

As mentioned earlier, the first stage of this work revealed that the initially proposed methods exhibit varying levels of forecast quality. In particular, the satellite-based method recently developed by OIE, which directly generates the set of quantile forecasts from a CMV model, suffers from a lack of reliability that significantly impacts its overall performance. However, a calibration technique could be employed to enhance this attribute. As expected, a parametric approach, like the one proposed by PIMENT is not suitable to provide high quality probabilistic solar forecast, even with the high performing point forecast as input.

The second stage of this study demonstrated that a high quality point forecast (that blends measurements, satellite-based and NWP forecasts) used in combination with a statistical technique is able to generate probabilistic forecasts with high quality. Overall, the skill scores of methods employing the blended point forecast vary between 42% and 46% for the intra-hour scenario and between 27% and 32% for the intra-day scenario. In contrast, methods that do not utilize the blended point forecast but are based on measurements and/or satellite data only exhibit skill scores ranging from 33% to 43% for intra-hour forecasts and from 8% to 16% for intra-day forecasts.

Besides a good forecast skill, the methodology that consists in generating probabilistic forecasts in a two step approach has the advantage that it is easy to implement in combination with blended deterministic forecasts that are well understood and used operationally. It allows to benefit from high quality deterministic point forecasts with comparatively simple probabilistic techniques applied in a post-processing step.

An alternative to blending the deterministic forecasts before applying the probabilistic techniques, would be to directly use the three deterministic forecasts as input to these techniques. LQR or QRF can be applied to generate quantile forecasts from different inputs in one step. For the AnEn method the different inputs can be combined with predictor weighting. Setting up these more complex models will be subject of future investigations.

Ongoing efforts by members of the IEA PVPS Task 16 involve the continuing development of solar probabilistic methods. Consequently, the evaluation and comparison of probabilistic forecasts will persist, and further analysis will be conducted using recent ground measurement data, satellite or NWP data. This continuing research aims to enhance the quality of solar irradiance probabilistic forecasts.

9. ACKNOWLEGDEMENTS

This research paper contributes to the IEA PVPS task 16 "Task 16 Solar resource for high penetration and large-scale applications," which is a collaborative effort between the International Energy Agency's. The study also received support from the TwInSolar project funded by the European Union's Horizon Europe research and innovation program grant number 101076447.

R. Alonso-Suárez acknowledges financial support from the CSIC Group's Program, Universidad de la República, Uruguay.

The research contribution of Tobias Zech, Wiebke Herzberg, Elke Lorenz (Fraunhofer ISE) received funding by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the SOLREV project (Grant Agreement no. 03EE1010) on the basis of a decision by the German Bundestag.

The authors express their gratitude to Anna Forstinger for providing the ground datasets.

10. Appendices

Appendix A Intra-hour and intra-day forecasts results site CAB

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	48.8 ± 6.1	40.3 ± 7.5	4.7 ± 0.4	81.8 ± 5.7	61.1 ± 8.5
PIMENT	50.9 ± 8.6	37.8 ± 10.5	2.1 ± 0.2	77.1 ± 8.4	69.2 ± 13.2
ISE	40.4 ± 5.3	50.7 ± 6.5	0.9 ± 0.2	86.4 ± 5.1	55.4 ± 7.6
UU	47.4 ± 7.8	42.1 ± 9.5	1.8 ± 0.4	80.3 ± 7.4	63.7 ± 11.4
LES	44.3 ± 7.4	45.9 ± 9.1	1.7 ± 0.5	83.2 ± 7.0	59.8 ± 10.2
BLEND-GARCH	41.9 ± 5.4	48.8 ± 6.6	1.9 ± 0.3	85.8 ± 5.1	55.0 ± 7.6
BLEND-LQR	42.2 ± 5.3	48.4 ± 6.5	2.1 ± 0.4	85.7 ± 4.9	54.7 ± 7.3
BLEND-QRF	39.9 ± 4.9	51.3 ± 6.0	1.1 ± 0.2	87.0 ± 4.7	54.3 ± 7.0

Table 7: Same as for Table 4 (intra-hour) but for site CAB. For site CAB, the CRPS of the CSD-CLIM is 81.8 W.m^{-2} and the uncertainty component UNC of the CRPS is 125.9 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	69.1 ± 3.8	15.6 ± 4.6	2.4 ± 0.2	59.1 ± 3.6	98.2 ± 6.1
ISE	50.2 ± 1.6	38.7 ± 1.9	1.0 ± 0.1	76.6 ± 1.6	69.4 ± 2.4
UU	68.4 ± 5.9	16.5 ± 7.2	3.7 ± 0.8	61.2 ± 5.1	96.3 ± 9.6
LES	65.6 ± 5.3	19.9 ± 6.4	3.7 ± 0.7	64.0 ± 4.6	89.8 ± 8.2
BLEND-GARCH	51.7 ± 1.5	36.9 ± 1.8	2.2 ± 0.1	76.4 ± 1.6	69.5 ± 2.2
BLEND-QRF	49.2 ± 1.6	39.9 ± 2.0	1.3 ± 0.1	78.0 ± 1.7	67.8 ± 2.5
BLEND-LQR	52.0 ± 1.5	36.5 ± 1.9	2.7 ± 0.0	76.6 ± 1.6	68.5 ± 2.4
BLEND-QRF	49.2 ± 1.6	39.9 ± 2.0	1.3 ± 0.1	78.0 ± 1.7	67.8 ± 2.5

Table 8: Same as for Table 5 (intra-day) but for site CAB

Appendix B Intra-hour and intra-day forecasts results site CAR

Method	CRPS	CRPSS	REL	RES	MAE
OIE	49.2 ± 5.1	32.8 ± 7.0	4.3 ± 0.4	108.5 ± 5.1	63.0 ± 7.4
PIMENT	53.3 ± 10.9	27.2 ± 14.9	5.3± 0.7	105.4 ± 10.2	68.8 ± 16.3
ISE	39.6 ± 7.2	46.0 ± 9.9	1.1 ± 0.3	114.9 ± 7.0	53.3 ± 10.0
UU	43.8 ± 8.9	40.2 ± 12.2	1.1 ± 0.2	110.7 ± 8.8	58.0 ± 12.9
LES	43.1 ± 8.7	41.1 ± 12.0	1.7 ± 0.3	112.0 ± 8.5	57.3 ± 12.6
BLEND-GARCH	44.3 ± 7.4	39.5 ± 10.2	2.9 ± 0.4	112.0± 7.0	55.8 ± 9.5
BLEND-LQR	43.1± 6.6	41.2 ± 9.0	2.9 ± 0.2	113.2 ± 6.3	55.0± 9.6
BLEND-QRF	40.0 ± 6.2	45.4 ± 8.5	1.3 ± 0.2	114.7 ± 6.1	53.2 ± 8.9

Table 9: Same as for Table 4 (intra-hour) but for site CAR. For site CAR, the CRPS of the CSD-CLIM is 73.2 W.m^{-2} and the uncertainty component UNC of the CRPS is 152.3 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	74.9 ± 4.4	-2.3 ± 6.0	6.3 ± 0.2	84.9 ± 4.2	102.2 ± 7.0
ISE	52.6 ± 2.0	28.2 ± 2.7	1.8 ± 0.2	102.7± 1.8	71.0 ± 2.6
UU	65.2 ± 5.6	10.9 ± 7.7	1.3 ± 0.1	89.5 ± 5.7	90.8 ± 8.9
LES	64.7 ± 5.5	11.6 ± 7.5	2.5 ± 0.2	91.1 ± 5.3	88.1 ± 8.0
BLEND-GARCH	57.7 ± 2.0	21.2 ± 2.8	3.7 ± 0.2	99.4 ± 1.8	73.7 ± 2.4
BLEND-LQR	55.3 ± 1.8	24.4 ± 2.5	4.1 ± 0.4	102.2 ± 1.4	73.6 ± 2.6
BLEND-QRF	51.7 ± 1.9	29.4 ± 2.6	2.1 ± 0.3	103.8 ± 1.6	69.7 ± 2.6

Table 10: Same as for Table 5 (intra-day) but for site CAR

Appendix C Intra-hour and intra-day forecasts results site CEN

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	60.9 ± 6.4	30.6 ± 7.3	5.5 ± 0.3	96.2 ± 6.5	78.4 ± 10.1
PIMENT	60.9 ± 10.6	30.5 ± 12.1	4.0 ± 0.7	94.6 ± 9.9	82.5 ± 16.2
ISE	50.6 ± 7.4	42.3 ± 8.5	1.0 ± 0.2	102.0 ± 7.2	69.4 ± 10.9
UU	53.7 ± 8.7	38.7 ± 9.9	0.9 ± 0.1	98.7 ± 8.6	73.1 ± 13.2
LES	53.1 ± 9.0	39.5 ± 10.2	1.2 ± 0.3	99.7 ± 8.7	72.2 ± 13.0
BLEND-GARCH	53.8 ± 7.8	38.7 ± 8.9	1.9 ± 0.3	99.7 ± 7.5	72.0 ± 11.0
BLEND-LQR	53.8 ± 7.5	38.6 ± 8.6	2.4 ± 0.5	100.1 ± 7.1	71.4 ± 11.3
BLEND-QRF	51.0 ± 6.9	41.8 ± 7.9	0.8 ± 0.1	101.4 ± 6.8	70.3 ± 10.4

Table 11: Same as for Table 4 (intra-hour) but for site CEN. For site CEN, the CRPS of the CSD-CLIM is 87.7 W.m^{-2} and the uncertainty component UNC of the CRPS is 153.1 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	82.8 ± 4.2	5.6 ± 4.8	5.3 ± 0.1	74.1 ± 4.1	116.6 ± 6.6
ISE	65.2 ± 2.6	25.6 ± 2.9	1.7 ± 0.2	88.1 ± 2.4	90.6 ± 3.5
UU	75.1 ± 5.6	14.4 ± 6.4	0.8 ± 0.1	77.3 ± 5.7	107.9 ± 9.4
LES	75.4 ± 5.3	14.1 ± 6.0	2.1 ± 0.2	78.2 ± 5.0	105.8 ± 8.6
BLEND-GARCH	69.2 ± 2.6	21.1 ± 3.0	2.7 ± 0.2	85.0 ± 2.4	94.0 ± 3.5
BLEND-LQR	68.3 ± 2.3	22.2 ± 2.7	3.2 ± 0.2	86.5 ± 2.1	94.0 ± 3.7
BLEND-QRF	65.2 ± 2.6	25.7 ± 3.0	1.1 ± 0.1	87.4 ± 2.6	91.9 ± 3.9

Table 12: Same as for Table 5 (intra-day) but for site CEN

Appendix D Intra-hour and intra-day forecasts results site MIL

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	50.2 ± 4.9	36.0 ± 6.2	4.2 ± 0.4	109.6 ± 4.6	63.8 ± 7.4
PIMENT	53.7 ± 10.5	31.5 ± 13.3	6.3 ± 1.1	108.2± 9.4	66.5 ± 15.2
ISE	41.1 ± 7.0	47.5 ± 8.9	1.4 ± 0.4	115.8 ± 6.6	54.6 ± 9.7
UU	43.9 ± 8.4	44.0 ± 10.7	0.9 ± 0.1	112.6 ± 8.3	58.3 ± 12.2
LES	44.8 ± 8.2	42.8 ± 10.4	3.8 ± 0.5	114.6 ± 7.7	56.1 ± 11.3
BLEND-GARCH	46.8 ± 7.0	40.2 ± 9.0	4.3 ± 0.6	113.1 ± 6.4	57.2 ± 9.2
BLEND-LQR	46.0 ± 6.7	41.3 ± 8.6	4.4 ± 0.8	114.0 ± 5.9	58.8 ± 9.3
BLEND-QRF	42.9 ± 6.0	45.2 ± 7.6	2.6 ± 0.5	115.3 ± 5.5	56.3 ± 8.2

Table 13: Same as for Table 4 (intra-hour) but for site MIL. For site MIL, the CRPS of the CSD-CLIM is 78.4 W.m^{-2} and the uncertainty component UNC of the CRPS is 154.6 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	75.7 ± 4.6	3.4 ± 5.8	7.8 ± 0.1	87.8 ± 4.5	100.2 ± 7.2
ISE	54.9 ± 2.6	30.0 ± 3.3	3.2 ± 0.6	104.0 ± 2.0	73.4 ± 3.2
UU	65.4 ± 6.4	16.5 ± 8.2	0.9 ± 0.1	91.1 ± 6.3	91.0 ± 10.0
LES	67.0 ± 6.2	14.5 ± 7.9	5.5 ± 0.6	94.1 ± 5.6	86.5 ± 9.0
BLEND-GARCH	61.0 ± 2.8	22.2 ± 3.5	6.5 ± 0.7	101.1 ± 2.1	77.4 ± 4.3
BLEND-LQR	60.2 ± 2.8	23.2 ± 3.6	8.0 ± 1.3	103.4 ± 1.6	79.9 ± 4.6
BLEND-QRF	55.1 ± 2.4	29.8 ± 3.0	4.3 ± 0.4	104.8 ± 2.0	72.7 ± 3.1

Table 14: Same as for Table 5 (intra-day) but for site MIL

Appendix E Intra-hour and intra-day forecasts results site PAL

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	55.7 ± 6.5	36.7 ± 7.3	5.3 ± 0.5	92.4 ± 6.0	69.3 ± 9.1
PIMENT	56.5 ± 9.2	35.8 ± 10.5	2.9 ± 0.4	89.3 ± 8.9	76.9 ± 14.3
ISE	45.0 ± 5.7	48.9 ± 6.5	1.2 ± 0.2	99.0 ± 5.5	61.3 ± 8.2
UU	52.9 ± 7.8	39.9 ± 8.9	2.0 ± 0.3	92.0 ± 7.6	70.3 ± 11.6
LES	50.0 ± 8.5	43.2 ± 9.6	1.7 ± 0.4	94.6 ± 8.1	68.0 ± 11.9
BLEND-GARCH	47.3 ± 5.8	46.3 ± 6.6	1.8 ± 0.2	97.4 ± 5.7	62.3 ± 8.1
BLEND-LQR	47.9 ± 5.7	45.5 ± 6.5	2.2 ± 0.3	97.1± 5.4	62.2 ± 8.1
BLEND-QRF	45.2 ± 5.4	48.6 ± 6.1	1.3 ± 0.2	98.9 ± 5.2	61.4 ± 7.8

Table 15: Same as for Table 4 (intra-hour) but for site PAL. For site PAL, the CRPS of the CSD-CLIM is 88.0 W.m^{-2} and the uncertainty component UNC of the CRPS is 140.5 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	76.9 ± 4.5	12.6 ± 5.1	3.6 ± 0.2	69.6 ± 4.3	109.4 ±7.3
ISE	55.1 ± 1.4	37.4 ± 1.6	1.4 ± 0.1	89.2 ± 1.5	75.9 ± 2.0
UU	73.1 ± 5.8	16.9 ± 6.6	2.7 ± 0.3	72.5 ± 5.5	102.6 ± 9.8
LES	73.7 ± 6.0	16.2 ± 6.8	3.4 ± 0.6	72.5 ± 5.4	102.3 ± 9.4
BLEND-GARCH	57.3 ± 1.6	34.9 ± 1.8	2.2 ± 0.2	87.8 ± 1.4	76.9 ± 2.3
BLEND-LQR	58.0 ± 1.5	34.1 ± 1.8	3.2 ± 0.3	88.0 ± 1.2	76.9 ±2.3
BLEND-QRF	54.9 ± 1.7	37.6 ± 1.9	1.8 ± 0.2	89.7 ± 1.5	75.4 ± 2.3

Table 16: Same as for Table 5 (intra-day) but for site PAL

Appendix F Intra-hour and intra-day forecasts results site PAY

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	53.3 ± 5.7	40.1 ± 6.4	4.7± 0.3	105.3 ± 5.8	68.4 ± 9.0
PIMENT	57.4 ± 11.7	35.5 ± 13.1	4.6 ± 0.6	101.2 ± 11.1	75.2 ± 17.7
ISE	44.7 ± 7.6	49.8 ± 8.5	0.9 ± 0.1	110.1 ± 7.4	61.1 ± 11.0
UU	48.0 ± 9.3	46.1 ± 10.5	1.3 ± 0.3	107.2 ± 9.1	64.0 ± 13.3
LES	48.2 ± 9.7	45.8 ± 10.9	2.0 ±0.5	107.7 ± 9.2	64.3 ± 13.8
BLEND-GARCH	50.1 ± 8.0	43.7 ± 9.0	3.0 ± 0.3	106.8 ± 7.8	64.9 ± 11.2
BLEND-LQR	49.5 ± 7.7	44.4 ± 8.6	3.4 ± 0.6	107.8 ± 7.1	64.6 ± 11.2
BLEND-QRF	45.6 ± 6.8	48.7 ± 7.6	1.8 ± 0.2	110.1 ± 6.6	61.7 ± 10.1

Table 17: Same as for Table 4 (intra-hour) but for site PAY. For site PAY, the CRPS of the CSD-CLIM is 89.0 W.m^{-2} and the uncertainty component UNC of the CRPS is 151.6 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	81.1 ± 4.3	8.8 ± 4.8	5.5 ± 0.1	78.2 ± 4.2	112.0 ± 6.7
ISE	60.7 ± 3.2	31.8 ± 3.6	1.4 ± 0.2	94.6 ± 3.0	84.7 ± 4.5
UU	71.9 ± 6.0	19.2 ± 6.7	1.9 ± 0.2	83.9 ± 5.8	100.0 ± 9.4
LES	73.8 ± 6.3	17.1 ± 7.1	4.6 ± 1.1	84.7 ± 5.2	101.2 ± 9.4
BLEND-GARCH	66.8 ± 3.2	24.9 ± 3.6	4.1 ± 0.4	91.2 ± 2.8	88.4 ± 4.3
BLEND-LQR	65.8 ± 3.1	26.1 ± 3.5	5.4 ± 0.6	93.5 ± 2.5	88.4 ± 4.4
BLEND-QRF	61.2 ± 3.4	31.2 ± 3.9	2.7 ± 0.3	95.4 ± 3.2	84.3 ± 4.8

Table 18: Same as for Table 5 (intra-day) but for site PAY

Appendix G Intra-hour and intra-day forecasts results site TAB

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	56.3 ± 4.9	4.1 ± 8.4	6.2 ± 0.8	111.3 ± 4.2	70.2 ± 7.1
PIMENT	55.9 ± 10.6	4.9 ± 18.1	6.0 ± 1.2	111.5 ± 9.5	72.7 ± 16.0
ISE	45.9 ± 7.7	21.9 ± 13.1	1.6 ± 0.5	117.2 ± 7.2	59.9 ± 10.5
UU	47.7 ± 8.3	18.7 ± 14.2	1.4 ± 0.3	115.1 ± 8.0	61.2 ± 11.5
LES	46.2 ± 8.9	21.4 ± 15.1	1.2 ± 0.2	116.4 ± 8.7	61.2 ± 12.6
BLEND-GARCH	51.5 ± 7.8	12.3 ± 13.3	3.7 ± 0.9	113.7 ± 6.9	63.9 ± 9.6
BLEND-LQR	48.7 ± 6.3	17.1 ± 10.7	3.1 ± 0.3	115.8 ± 6.0	63.6 ± 9.1
BLEND-QRF	46.1 ± 6.3	21.5 ± 10.8	1.6 ± 0.3	116.9 ± 6.1	60.5 ± 8.8

Table 19: Same as for Table 4 (intra-hour) but for site TAB. For site TAB, the CRPS of the CSD-CLIM is 58.7 W.m^{-2} and the uncertainty component UNC of the CRPS is 160.0 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	73.8 ± 2.6	-25.7 ± 4.3	7.8 ± 0.4	95.5 ± 2.2	101.2 ± 4.2
ISE	56.2 ± 1.2	4.2 ± 2.0	3.4 ± 0.6	108.6 ± 0.5	73.9 ± 1.2
UU	64.9 ± 3.3	-10.5 ± 5.7	1.7 ± 0.1	98.2 ± 3.4	84.1 ± 4.2
LES	65.5 ± 3.7	-11.6 ± 6.4	2.1 ± 0.3	98.0 ± 3.4	87.4 ± 4.6
BLEND-GARCH	63.0 ± 1.6	-7.2 ± 2.7	6.4 ± 0.9	104.9 ± 0.7	80.1 ± 2.5
BLEND-LQR	57.0 ± 0.8	2.9 ± 1.4	4.1 ± 0.5	108.5 ± 0.4	76.6 ± 1.3
BLEND-QRF	53.8 ± 0.6	8.3 ± 1.1	2.3 ± 0.2	109.9 ± 0.5	71.7 ± 0.8

Table 20: Same as for Table 5 (intra-day) but for site TAB

Appendix H Intra-hour and intra-day forecasts results site TOR

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	55.4 ± 5.3	34.5 ± 6.3	7.3 ± 0.4	79.9 ± 4.9	67.3 ± 7.6
PIMENT	52.4 ± 9.0	38.0 ± 10.7	2.6 ± 0.2	78.1 ± 8.8	70.0 ± 13.4
ISE	43.0 ± 4.8	49.1 ± 5.6	0.9 ± 0.0	85.8 ± 4.8	59.3 ± 7.0
UU	48.8 ± 7.8	42.2 ± 9.3	2.3 ± 0.6	81.4 ± 7.3	64.4 ± 11.4
LES	45.9 ± 7.9	45.7 ± 9.3	1.7 ± 0.5	83.7 ± 7.4	62.1 ± 10.7
BLEND-GARCH	44.8 ± 5.2	47.0 ± 6.2	1.7 ± 0.1	84.8 ± 5.1	60.0 ± 7.2
BLEND-LQR	46.1 ± 5.2	45.5 ± 6.2	2.8 ± 0.4	84.7 ± 4.9	59.4 ± 7.3
BLEND-QRF	41.9 ± 4.6	50.4 ± 5.4	1.4 ± 0.1	87.4 ± 4.5	56.9 ± 6.9

Table 21: Same as for Table 4 (intra-hour) but for site TOR. For site TOR, the CRPS of the CSD-CLIM is 84.5 W.m^{-2} and the uncertainty component UNC of the CRPS is 125.9 W.m^{-2}.

Method	CRPS (W/m^2)	CRPSS (%)	REL (W/m^2)	RES (W/m^2)	MAE (W/m^2)
OIE	NA	NA	NA	NA	NA
PIMENT	72.0 ± 4.1	14.8 ± 4.8	2.8 ± 0.1	58.8 ± 4.0	100.9 ± 6.8
ISE	51.8 ± 1.6	38.7 ± 1.9	1.1 ± 0.1	77.2 ± 1.6	72.0 ± 2.3
UU	70.7 ± 6.3	16.3 ± 7.4	5.1 ± 1.0	62.3 ± 5.3	98.3 ± 10.9
LES	69.1 ± 6.1	18.3 ± 7.2	4.6 ± 1.0	63.5 ± 5.1	93.7 ± 9.4
BLEND-GARCH	54.7 ± 1.9	35.3 ± 2.2	2.5 ± 0.3	75.7 ± 1.6	72.6 ± 2.3
BLEND-LQR	56.2 ± 1.9	33.4 ± 2.2	4.2 ± 0.4	75.9 ± 1.5	72.4 ± 2.2
BLEND-QRF	51.6 ± 2.0	39.0 ± 2.4	2.2 ± 0.2	78.5 ± 1.8	70.2 ± 2.6

Table 22: Same as for Table 5 (intra-day) but for site TOR

References

Aicardi, D., Musé, P., Alonso-Suárez, R., 2022. A comparison of satellite cloud motion vectors techniques to forecast intra-day hourly solar global horizontal irradiation. Solar Energy 233, 46–60.

Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G., 2015. An analog ensemble for short-term probabilistic solar power forecast. Applied Energy 157, 95–110.

Alonso-Suárez, R., Aicardi, D., Marchesoni-Acland, F., 2022. Analysis of persistence-based solar irradiance forecasting benchmarks. 2203.13819.

Alonso-Suárez, R., David, M., Branco, V., Lauret, P., 2020. Intra-day solar probabilistic forecasts including local short-term variability and satellite information. Renewable Energy 158, 554–573.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de Pison, F., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. Solar Energy 136, 78–111.

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. Solar Energy 83, 1772–1783.

Ben Bouallègue, Z., 2015. Assessment and added value estimation of an ensemble approach with a focus on global radiation forecasts. MAUSAN , 541–550.

Blaga, R., Sabadus, A., Stefu, N., Dughir, C., Paulescu, M., Badescu, V., 2019. A current perspective on the accuracy of incoming solar energy forecasting. Progress in Energy and Combustion Science 70, 119–144.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bourges, B. (Ed.), 1992. Climatic Data Handbook for Europe. 13537, Europäische Kommission, Dordrecht: Kluwer Acad. Publ. Available online at http://bookshop.europa.eu/en/-pbEUNA13537/.

Bröcker, J., Smith, L.A., 2007. Increasing the Reliability of Reliability Diagrams. Weather and Forecasting 22, 651–661.

BSRN, 2024. World radiation monitoring center (wrmc), the central archive of the baseline surface radiation network (bsrn). https://bsrn.awi.de. Accessed: 22 January 2024.

Carrière, T., Amaro E Silva, R., Zhuang, F., Saint-Drenan, Y.M., Blanc, P., 2021. A New Approach for Satellite-Based Probabilistic Solar Forecasting with Cloud Motion Vectors. Energies 14, 4951.

Chow, C.W., Belongie, S., Kleissl, J., 2015. Cloud motion and stability estimation for intra-hour solar forecasting. Solar Energy 115, 645–655.

CIEMAT, 2024. Ciemat plataforma solar de almería - europe's biggest test center for concentrating solar power (csp). https://www.dlr.de/sf/en/desktopdefault.aspx/. Accessed: 22 January 2024.

David, M., Mazorra Aguiar, L., Lauret, P., 2018. Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data. International Journal of Forecasting 34, 529–547.

David, M., Ramahatana, F., Trombe, P., Lauret, P., 2016. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. Solar Energy 133, 55–72.

Delle Monache, L., Eckel, F., Rife, D., Nagarajan, B., Searight, K., 2013. Probabilistic weather prediction with an analog ensemble. Monthly Weather Review 141, 3498–3516.

Doubleday, K., Van Scyoc Hernandez, V., Hodge, B., 2020. Benchmark probabilistic solar forecasts: Characteristics and recommendations. Solar Energy 206, 52–67.

Dumortier, D., 1995. Modelling global and diffuse horizontal irradiances under cloudless skies with different turbidities. Final Report Vol. 2. Daylight II, JOU2-CT92-0144.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1007. Publisher: [Wiley, Econometric Society].

Forstinger, A., Wilbert, S., Jensen, A., Kraas, B., Peruchena, C., Gueymard, C., Ronzio, D., Yang, D., Collino, E., Martinez, J., Ruiz-Arias, J., Hanrieder, N., Blanc, P., Saint-Drenan, Y., 2021. Expert quality control of solar radiation ground data sets, in: Proceedings of SWC 2021: ISES Solar World Congress, International Solar Energy Society. pp. 1037–1048. SWC 2021: ISES Solar World Congress ; Conference date: 25-10-2021 Through 29-10-2021.

Gneiting, T., Lerch, S., Schulz, B., 2023. Probabilistic solar forecasting: Benchmarks, post-processing, verification. Solar Energy 252, 72–80.

Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association 102, 359–378.

Grantham, A., Gel, Y.R., Boland, J., 2016. Nonparametric short-term probabilistic forecasting for solar radiation. Solar Energy 133, 465–475.

Hammer, A., Heinemann, D., Hoyer-Klick, C., Kuhlemann, R., Lorenz, E., Müller, R., Beyer, H.G., 2003. Solar energy assessment using remote sensing technologies. Remote Sensing of Environment 86, 423–432.

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather and Forecasting 15, 559–570.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J., 2016. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting 32, 896–913.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H., 2020. Energy forecasting: A review and outlook. IEEE Open Access Journal of Power and Energy 7, 376–388.

IEA-PVPS-T16, 2024. Solar resource for high penetration and large scale applications. https://iea-pvps.org/research-tasks/solar-resource-for-high-penetration-and-large-scale-applications/. Accessed: 22 January 2024.

IEA-SHC-T46, 2024. Solar resource assessment and forecasting. https://task46.iea-shc.org. Accessed: 22 January 2024.

Iversen, E.B., Morales, J.M., Møller, J.K., Madsen, H., 2015. Short-term probabilistic forecasting of wind speed using stochastic differential equations. International Journal of Forecasting .

Jensen, A.R., Anderson, K.S., Holmgren, W.F., Mikofski, M.A., Hansen, C.W., Boeman, L.J., Loonen, R., 2023. pvlib iotools,- open-source python functions for seamless access to solar irradiance data. Solar Energy 266, 112092.

Jung, J., Broadwater, R.P., 2014. Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews 31, 762–777.

Junk, C., Delle Monache, L., Alessandrini, S., Cervone, G., Bremen, L., 2015. Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. Meteorologische Zeitschrift 24, 361–379.

Koenker, R., 2005. Quantile Regression. volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Koenker, R., Bassett, G., 1978. Regression Quantiles. Econometrica 46, 33–50.

Kühnert, J., Lorenz, E., Heinemann, D., 2013. Satellite-based irradiance and power forecasting for the german energy market. Solar Energy Forecasting and Resource Assessment , 267–295.

Lauret, P., Alonso-Suárez, R., Le Gal La Salle, J., David, M., 2022. Solar forecasts based on the clear sky index or the clearness index: Which is better? Solar 2, 432–444.

Lauret, P., David, M., Pedro, H., 2017. Probabilistic Solar Forecasting Using Quantile Regression Models. Energies 10, 1591.

Lauret, P., David, M., Pinson, P., 2019. Verification of solar irradiance probabilistic forecasts. Solar Energy 194, 254–271.

Lauret, P., Voyant, C., Soubdhan, T., David, M., Poggi, P., 2015. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Solar Energy 112, 446–457.

Le Gal La Salle, J., David, M., Lauret, P., 2021. A new climatology reference model to benchmark probabilistic solar forecasts. Solar Energy 223, 398–414.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.J., 2013. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Measurement Techniques 6, 2403–2418.

Liu, C., 2009. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Ph.D. thesis. Massachusetts Institute of Technology. Cambridge, MA, USA.

Lorenz, E., Remund, J., Müller, S.C., Traunmüller, W., Steinmaurer, G., Pozo, D., Ruiz-Arias, J.A., Fanego, V.L., Ramirez, L., Romeo, M.G., Kurz, C., Pomares, L.M., Geijo Guerrero, C., 2009. Benchmarking of different approaches to forecast solar irradiance, in: 24th European photovoltaic solar energy conference, Hamburg Germany. pp. 21–25.

Lorenz, E., Ruiz-Arias, J.A., Martin, L., Wilbert, S., Köhler, C., Fritz, R., Betti, A., Lauret, P., David, M., Huang, J., Perez, R., Kazantzidis, A., Wang, P., Saint-Drenan, Y.M., 2021. Forecasting Solar Radiation and Photovoltaic Power. NREL/TP-5D00-77635, National Renewable Energy Laboratory, Golden, CO.

Mazorra-Aguiar, L., Lauret, P., David, M., Oliver, A., Montero, G., 2021. Comparison of Two Solar Probabilistic Forecasting Methodologies for Microgrids Energy Efficiency. Energies 14, 1679.

Meinshausen, N., Ridgeway, G., 2006. Quantile regression forests. Journal of Machine Learning Research 7, 983–999.

Morales, J.M., Conejo, A.J., Madsen, H., Pinson, P., Zugno, M., 2014. Integrating Renewables in Electricity Markets. volume 205 of *International Series in Operations Research & Management Science*. Springer US, Boston, MA. DOI: 10.1007/978-1-4614-9411-9.

Notton, G., Nivet, M.L., Voyant, C., Paoli, C., Darras, C., Motte, F., Fouilloy, A., 2018. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. Renewable and Sustainable Energy Reviews 87, 96–105.

Paletta, Q., Arbod, G., Lasenby, J., 2021. Benchmarking of deep learning irradiance forecasting models from sky images ,Äì an in-depth analysis. Solar Energy 224, 855–867.

Paletta, Q., Arbod, G., Lasenby, J., 2023. Omnivision forecasting: Combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions. Applied Energy 336, 120818.

Pedro, H.T., Coimbra, C.F., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 86, 2017–2028.

Pedro, H.T.C., Larson, D.P., Coimbra, C.F.M., 2019. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. Journal of Renewable and Sustainable Energy 11, 036102.

Pinson, P., McSharry, P., Madsen, H., 2010. Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. Quarterly Journal of the Royal Meteorological Society 136, 77–90.

Pinson, P., Nielsen, H.A., Møller, J.K., Madsen, H., Kariniotakis, G.N., 2007. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy 10, 497–516.

Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., Wald, L., 2017a. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The heliosat?4 method. Meteorologische Zeitschrift 26, 33–57.

Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., Wald, L., 2017b. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The heliosat?4 method. Meteorologische Zeitschrift 26, 33–57.

Raschka, S., Mirjalili, V., 2019. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing Ltd, Birmingham. third edition.

Sobri, S., Koohi-Kamali, S., Rahim, N.A., 2018. Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management 156, 459–497.

Taylor, J.W., 2004. Volatility forecasting with smooth transition exponential smoothing. International Journal of Forecasting 20, 273–286.

Tsay, R.S., 2010. Analysis of Financial Time Series. Wiley Series in Probability and Statistics, Wiley. 1 edition.

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.

Verbois, H., Saint-Drenan, Y.M., Thiery, A., Blanc, P., 2022. Statistical learning for nwp post-processing: A benchmark for solar irradiance forecasting. Solar Energy 238, 132–149.

Visser, L., AlSkaif, T., Hu, J., Louwen, A., van Sark, W., 2023. On the value of expert knowledge in estimation and forecasting of solar photovoltaic power generation. Solar Energy 251, 86–105.

Wilks, D.S., 2009. Statistical methods in the atmospheric sciences. Number 91 in International geophysics series, Elsevier [u.a.], Amsterdam. 2. ed., [nachdr.] edition. OCLC: 845720508.

Yang, D., van der Meer, D., Munkhammar, J., 2020. Probabilistic solar forecasting benchmarks on a standardized dataset at folsom, california. Solar Energy 206, 628–639.

Yang, D., Wang, W., Gueymard, C.A., Hong, T., Kleissl, J., Huang, J., Perez, M.J., Perez, R., Bright, J.M., Xia, X., van der Meer, D., Peters, I.M., 2022. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renewable and Sustainable Energy Reviews 161, 112348.