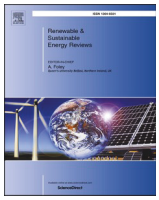
Renewable and Sustainable Energy Reviews 175 (2023) 113171

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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



Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting

Martin János Mayer [a,*], Dazhi Yang [b]

[a] Department of Energy Engineering, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111, Budapest, Hungary
[b] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China

ARTICLE INFO

Keywords:
Probabilistic solar forecasting
Ensemble NWP
Physical model chains
Photovoltaic power
Quantile regression
ECMWF

ABSTRACT

Under the two-step framework of photovoltaic (PV) power forecasting, that is, forecasting first the irradiance and then converting it to PV power, there are two chief ways in which one can account for the uncertainty embedded in the final PV power forecast. One of those is to produce probabilistic irradiance forecast through, for example, ensemble numerical weather prediction (NWP), and the other is to pass the irradiance forecast through a collection of different irradiance-to-power conversion sequences, which are known as model chains. This work investigates, for the first time, into the question: Whether pairing ensemble NWP with ensemble model chain is better than leveraging any individual method alone? Using data from 14 utility-scale ground-mounted PV plants in Hungary and the state-of-the-art global mesoscale NWP model of the European Centre for Medium-Range Weather Forecasts, it is herein demonstrated that the best probabilistic PV power forecast needs to consider both ensemble NWP and ensemble model chain. Furthermore, owing to the higher-quality probabilistic forecasts, the point forecast accuracy is also improved substantially through pairing. Overall, the recommended paring strategy achieves a mean-normalized continuous ranked probability score and a root mean square error of 18.4% and 42.1%, respectively.

1. Introduction

The backbone of state-of-the-art photovoltaic (PV) power forecasting, or energy forecasting in general, is physical modeling [1–3]. The general process of physical PV power forecasting, on intra-day and day-ahead scales, includes, as the first step, modeling the dynamics of the atmosphere with numerical weather prediction (NWP) models, and then, modeling the power output of the PV plant with physical model chains. At the end of these two steps, one should further apply various data-driven (i.e., statistical and machine learning) post-processing methods, which are able to enhance the quality of the physical modeling results, as to achieve bias reduction and calibration [4,5]. There are certainly many aspects of post-processing that still require attention, this work should nonetheless wish to focus on the physical modeling itself, and use just a basic calibration tool.

Forecasts can be either deterministic (also known as point, single-valued, or best-guess) or probabilistic (also known as ensemble or distributional). NWP leverages the laws of physics to predict the state of the atmosphere. In that respect, it is a two-step procedure, first of diagnosis and second of prognosis. The diagnosis step may be viewed as an initial value problem, in which myriads of observations of current weather from heterogeneous and autonomous sources are assimilated, as to arrive at an estimate of the current state of the atmosphere. Then, in the prognosis step, future states of the atmosphere are determined by integrating the governing partial differential equations describing the evolution of the weather. Weather is historically regarded as a deterministic process, in that if the initial conditions and the physical laws are known exactly, the future states can be deterministically calculated. In this regard, each NWP model consists of a set of physical schemes and parameterizations, selected from a large pool of differing options. The most appropriate combination (i.e., the "perfect" model) is usually identified by a group of experienced meteorologists, according to the forecasting situation of concern and after years of tuning and updating. Unfortunately, owing to the chaotic nature of the weather, which was first discovered in the 1960s, even the accuracy of the "perfect" model is frequently found to be unsatisfactory. Ensemble NWP has thence become the mainstream. Simply put, ensemble NWP runs the same NWP model several times, each with a different set of initial conditions, which

* Corresponding author.
E-mail address: mayer@energia.bme.hu (M.J. Mayer).

https://doi.org/10.1016/j.rser.2023.113171
Received 20 August 2022; Received in revised form 3 January 2023; Accepted 6 January 2023

results in as many equally likely weather trajectories, i.e., ensemble weather forecasts [6,7]. Insofar as the present experience can advise, ensemble NWP possesses two distinct advantages over deterministic NWP: (1) it is able to quantify the uncertainty and thus predictability of the forecasting situation; and (2) when forecasts from an ensemble NWP model are summarized into deterministic ones, e.g., by taking the means or medians, they are almost always found to be more accurate than the best-guess forecasts from the "perfect" model. The concept of ensemble NWP is illustrated in Fig. 1 (a).

Physical model chain, as suggested by its name, considers a series of energy meteorology models in cascade, where the output of a preceding model acts as input of a succeeding model. For instance, with the location and time information, solar positioning can be performed, of which an output is the solar zenith angle, which is required by the separation and transposition models. And the output of transposition models, namely, the tilted global irradiance, is needed by the DC power model. Whereas the DC power model outputs the DC power, an inverter model is used to convert that into AC power, considering the topology of the PV system, which includes wiring, fleet spacing, voltage level, among other factors. The reader is referred to Refs. [8–10] for fundamentals of model chains. In short, a model chain converts the initial weather input variables, which include, in the main, global horizontal irradiance (GHI), ambient temperature, and surface wind speed, into PV power, in a step-by-step fashion, using the PV plant's design parameters and specifications. In the language of computer scientists, the model chain is the cyber dual of the physical PV system. Generally speaking, the higher degree to which the design parameters and specifications are known, the better the performance of a model chain would be, although some information is more critical than others [11]. Similar to the physical schemes and parameterizations of NWP, solar energy meteorology models used in the model chain also need to be selected, as there could be as many as hundreds of options for each stage of the model chain. Traditionally, the model chain is also viewed as a deterministic process: One can either select the best-performing model for each stage, or optimize the model chain as a whole—this is because the simple cascade

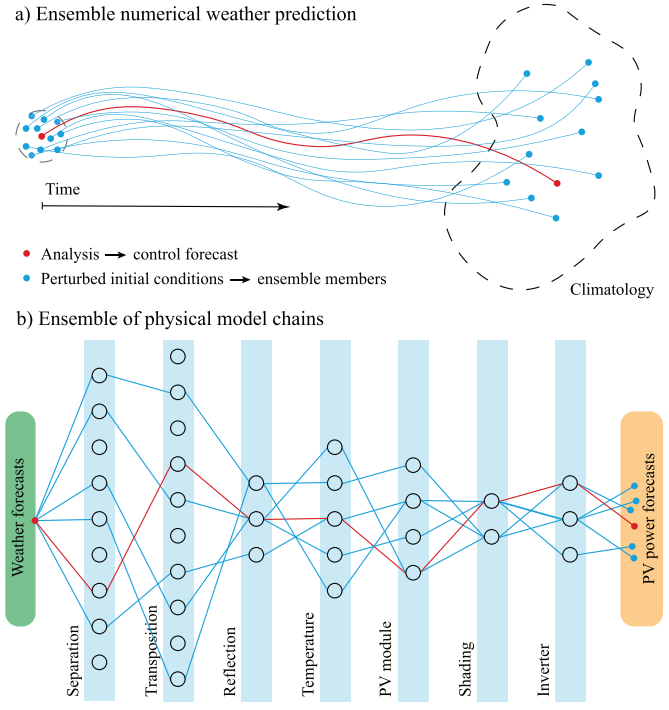


Fig. 1. Schematics of (a) ensemble numerical weather prediction and (b) ensemble of model chains, where each circle represents a component model. In both subfigures, the red line marks the "best-guess" prediction, whereas the blue lines exemplify the member trajectories.

of best single-stage models can rarely lead to the best overall performance [8]. On this point, if each model chain is considered as a probable path of irradiance-to-power conversion, collecting several of those results in an ensemble model chain; the situation is highly analogous to the ensemble of weather trajectories. Fig. 1 (b) shows the schematic diagram of the ensemble model chain.

With the above background information, it is obvious that there are three ways to generate probabilistic PV power forecast: (1) deterministic NWP + ensemble model chain, (2) ensemble NWP + deterministic model chain, and (3) ensemble NWP + ensemble model chain. The relevant question is therefore this: Which option is the most advantageous and why? This work seeks to provide some empirical evidence on that, using data collected at 14 utility-scale ground-mounted PV plants in Hungary, alongside the ensemble NWP forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). As a byproduct, we also investigate the deterministic forecasting performance of various options—recall that probabilistic forecasts can be summarized into deterministic ones—and compare the summarized forecasts to the ones obtained from the traditional deterministic forecasting procedure, that is, deterministic NWP + deterministic model chain. Last but not least, since ensemble forecasts from physical methods are often found to be underdispersed, quantile regression, as a calibration tool, is thought beneficial and therefore included in the analysis.

The remaining part of the paper is organized as follows. Section 2 introduces the data. Worth-noting is that, aside from the PV system and NWP forecast data, we also consider satellite-derived irradiance in this work, which is used to gauge the quality of NWP irradiance forecasts. Section 3 delivers the methods for generating and calibrating the PV power forecasts. The verification of deterministic and probabilistic forecasts resulting from various aforementioned options is depicted in Section 4. Conclusions follow at the end.

2. Data

The analyses performed in this study rely on data from three sources, which are introduced in the following subsections. The ensemble NWP weather forecast data of the ECMWF, which serve as a basis for calculating the PV power forecasts, is described in Section 2.1. The specifications and characteristics of the Copernicus Atmosphere Monitoring Service (CAMS) radiation service (CAMS-Rad) satellite-derived irradiance, which is used as verifications for NWP forecasts, are detailed in Section 2.2. The design data of the PV plants, which is required for their physical modeling, and their power production data, which is used for the verification, are summarized in Section 2.3.

2.1. Numerical weather prediction data

The NWP forecasts used in this paper come from the ECMWF's Integrated Forecasting System (IFS). IFS is a mesoscale NWP system covering the entire globe and hosts a range of meteorological and oceanographic products. Two products are of relevance here, which are "Set I – Atmospheric Model high resolution 10-day forecast (HRES)" and "Set III – Atmospheric model ensemble 15-day forecast (ENS)," with the former issuing deterministic forecasts and the latter probabilistic. Both products are operational and disseminate forecasts to licensed users four times a day, at 00Z, 06Z, 12Z, and 18Z, where "Z" denotes Zulu time. In each run, HRES and ENS produce 240-h and 360-h forecasts, respectively. Since these products are operational, they undergo changes regularly, not just in terms of the physics packages and parameterizations used in the model, but also in terms of output variable and output resolution.

HRES is the highest-resolution configuration of IFS. At present, it has a horizontal resolution of 9 km with 137 vertical layers. Temporally, HRES forecasts are hourly for the first 90 h, 3 hourly for 93–144 h, and 6 hourly for 150–240 h. In contrast, ENS has a lower horizontal resolution of 18 km, but with the same temporal resolution, except that it issues 6-

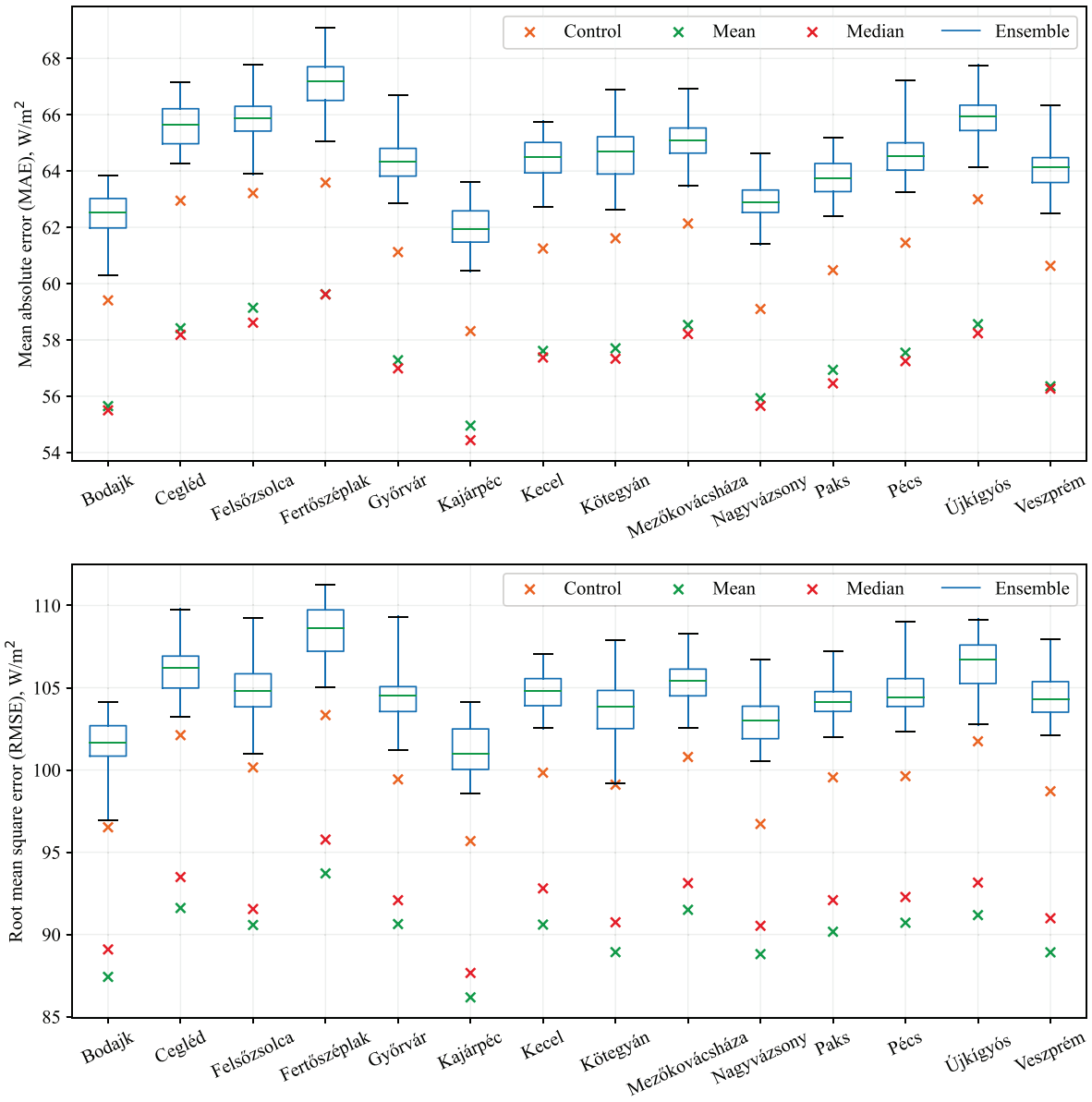


Fig. 2. Mean absolute errors (MAE) and root mean square errors (RMSE) of the 50 ensemble members (shown by boxplots), the control forecast, and the mean and median of the ensemble members.

hourly forecasts for horizons ranging from 150 to 360 h. Hence, to isolate the effect of spatial-scale mismatch [12,13] as far as possible, we do not mix HRES deterministic forecasts with ENS ensemble forecasts in this work. Instead, the control (i.e., best-guess) run of ENS is used as the deterministic version.

Indeed, ENS, as an ensemble model, repeats the HRES model 51 times, among which there are 1 control run and 50 perturbed runs. Whereas the control run leverages the best-possible initial conditions, each of the 50 perturbed runs uses a slightly different set of initial conditions, which are obtained using a method combining both the singular vector (SV) and ensemble data assimilation (EDA), following:

$$\text{ENS member } i = \text{HRES} + (\text{EDA member } i - \text{EDA mean}) + \text{SV Perturbation } i \quad (1)$$

The reader is referred to Bonavita et al. [14] and Diaconescu and Laprice [15], as well as the official documentation,[1] for details on EDA and SV, respectively. In any case, as reported by Buizza et al. [16], the combination of EDA and SV is novel and more effective than using any one method alone.

ENS forecasts can be obtained in real-time from the ECMWF's Meteorological Archival and Retrieval System (MARS). In this work, forecasts from the 00Z run, over a period of two years (2019–2020), over a region covering all PV systems of interest, are downloaded from MARS at the original resolution without any post-processing. More specifically, three parameters, namely, GHI, ambient temperature at 2 m, and surface wind speed (converted from the U- and V-component wind), are acquired, which are necessary for meaningful model-chain implementation. To comply to the actual operational day-ahead solar forecasting requirements, 25–48-h-ahead forecasts are selected for each 00Z run; this factors in both the initial run time of the NWP model (i.e., the first

[1] https://confluence.ecmwf.int/display/FUG/5.1+Generation+of+the+Ensemble.

few hours of forecasts are not disseminated on time) and the grid-side submission lead time (i.e., forecasts are to be submitted a few hours before the operating day).

Notwithstanding, given the fact that the PV power data has a resolution of 15 min (see below), the hourly NWP needs to be downscaled. Here, a simple clear-sky interpolation is used for GHI forecasts, whereas simple linear interpolation is used for temperature and wind speed forecasts. It should be emphasized that clear-sky interpolation of irradiance is essential due to the presence of the diurnal cycle in irradiance time series. Stated differently, the interpolation must be done first in clear-sky index terms and then back-transformed into irradiance using the corresponding clear-sky GHI values at the interpolated time stamps. The clear-sky irradiance model used to retrieve the clear-sky index from GHI is McClear [17,18], which has been recommended as suitable for solar forecasting applications [19,20].

2.2. Satellite-derived irradiance data

Since the PV power forecast accuracy depends on that of irradiance, it is of interest to verify the NWP irradiance forecasts. For that purpose, this work considers the satellite-derived irradiance from the CAMS-Rad database. The verification of forecasts, at least in the solar energy meteorology context, is conventionally based upon ground-based measurements, for data of this kind is of a higher quality as compared to satellite-derived irradiance. Whereas this preference still holds, it suffers from a major drawback—high-quality long-term radiometry stations are exceedingly few due to their high equipment and maintenance costs. Therefore, meteorologists have long been leveraging remote-sensed observations as verifications (i.e., the "true" values), especially when forecast verification needs to be done over a region [21,22].

Verification of irradiance forecasts using satellite-derived irradiance is but a very recently thought of strategy. The earliest known work is the one by Perez et al. [23], who noticed that the forecast root mean square errors (RMSEs) of some forecasts against ground-based measurements and that against SolarAnywhere irradiance are comparable. Nonetheless, it is well known now that accuracy is just one aspect of forecast quality, and many other aspects, such as bias, calibration, refinement, or discrimination, do contribute to the overall judgment and confidence of forecasters and forecast users. In this regard, Yang and Perez [24] revisited the problem, presented a holistic extension to the initial work, and documented the distinctions in NWP forecast verification results, using both ground-based measurements and satellite-derived irradiance; the Murphy–Winkler verification framework was used, which is able to quantify the different aforementioned aspects of forecast quality. The main conclusion made therein is that satellite-derived irradiance, based on the latest generation of geostationary weather satellites and cloud-to-irradiance algorithms, is adequate for forecast-verification purposes.

Coming back to CAMS-Rad, it is derived using the Heliosat-4 method from the images acquired by the Meteosat Second Generation (MSG) satellites, and has been shown to be more advantageous than the Surface Solar Radiation Data Set – Heliosat (SARAH-2) [25], which is another satellite-derived irradiance product based on the MSG satellites. Spatially, CAMS-Rad covers a disk region centered at the Meridian and the Equator, extending out to ±66° in latitude and similar in longitude. Temporally, CAMS-Rad irradiance can be obtained in 1-, 15-, 30-, and 60-min resolutions. In this work, 15-min CAM-Rad irradiance is downloaded using the web services from the SoDa-pro website,[2] over the two-year (2019–2020) period and for each PV system location. It is worth mentioning that, unlike other satellite-derived irradiance products, such as SARAH-2 or the National Solar Radiation Database (NSRDB), of which the spatial resolution is fixed to that of the raw satellite imagery, CAMS-Rad uses interpolation. For that reason, it is able to alleviate spatial-scale mismatch to some extent, and thus may better resemble ground-based measurements.

The verification of the NWP GHI forecasts is performed by calculating the mean absolute error (MAE) and root mean square error (RMSE) of the control forecasts, all ensemble members, and the mean and median of the ensemble. The MAE and RMSE are shown separately for each PV plant location in Fig. 2. The results reveal that the control forecasts lower errors than any of the 50 individual ensemble members in terms of both metrics in all locations. However, the mean and median of the ensemble is even considerably more accurate than the "best guess" control forecasts, which suggests that ensemble forecasting has a significant added value even if the ultimate goal is deterministic forecasting. In terms of MAE, the median of the ensemble is the more accurate, while in terms of RMSE, the mean is the most accurate, which empirically demonstrates the statistical fact that the MAE is minimized by the median, and the MSE is minimized by the mean.

2.3. Photovoltaic power plant and production data

The power forecasting methods proposed in this paper are tested for 14 utility-scale ground-mounted PV plants in Hungary. These PV plants were installed in 2018, and they are operated by MVM Green Generation Ltd. The modules are mounted on parallel rows of mounting structures with a south-facing orientation and 20–35° tilt angles, depending on the plants. The power generation data of these PV plants cover the whole years of 2019 and 2020 with a 15-min temporal resolution.

The preprocessing of the power production data includes the removal of all nighttime values, which are identified by a zenith angle >90° filter. Moreover, the daytime periods with no power output are also removed, as those discontinuities are due to errors or maintenance of the PV plants. The metadata of the PV plant and generation data is listed in Table 1.

Table 1
Name, location, and design data of the 14 ground-mounted PV power plants included in this paper. P_{DC}: installed DC capacity, P_{AC}: nominal AC power.

Name	Coordinates	P_{DC} kW	P_{AC} kW	Nr. Of valid data points	
				2019	2020
Bodajk	47.33° N, 18.22° E	590	498	17,496	17,547
Cegléd	47.19° N, 19.80° E	590	498	17,035	17,015
Felsőzsolca	48.12° N, 20.89° E	20,038	16,776	17,300	17,337
Fertőszéplak	47.61° N, 16.84° E	590	498	17,499	17,583
Győrvár	46.99° N, 16.83° E	590	498	17,467	17,437
Kajárpéc	47.51° N, 17.62° E	590	498	17,339	17,574
Kecel	46.53° N, 19.22° E	590	498	17,056	17,084
Kötegyán	46.74° N, 21.48° E	590	498	16,921	17,102
Mezőkovácsháza	46.40° N, 20.90° E	590	498	17,052	16,973
Nagyvázsony	46.98° N, 17.69° E	590	498	17,513	17,543
Paks	46.57° N, 18.82° E	20,680	17,244	17,179	17,188
Pécs	46.06° N, 18.26° E	10,044	10,097	16,993	17,090
Újkígyós	46.60° N, 20.99° E	590	498	17,061	16,924
Veszprém	47.10° N, 17.87° E	590	498	17,517	17,533

3. Methods

The general concepts of creating deterministic and probabilistic power forecasts are explained in Section 3.1. Afterward, the methods responsible for the generation and calibration of the power forecasts from the raw NWP forecasts, namely, physical PV model chains and the quantile regression, are described in Sections 3.2 and 3.3, respectively. Finally, Sections 3.4 and 3.5 present the applied verification practices, respectively, for the deterministic and probabilistic forecasts.

[2] https://www.soda-pro.com/web-services/radiation/cams-radiation-service

3.1. Concepts for photovoltaic power forecasting

The general procedure of NWP-based physical PV power forecasting is to convert the GHI, ambient temperature, and wind speed forecasts to power output forecasts by simulating the behavior of the PV plants by a physical model chain.[3] The model chain itself is a general concept, which can be realized in many different ways, depending on how many modeling steps are included and which models are used in these steps. Different model chains map the weather forecasts to PV power differently; therefore, the selection of the model chain can affect the power forecast errors by more than 10% [8]. To that end, the power conversion with an arbitrarily selected model chain is often suboptimal, and thus it prevents reliably judging the potential accuracy of any forecasting method that involves model chains. A possible solution, as proposed in Ref. [8], is to optimize the model chain by calculating the power forecasts with many different model chains, then selecting the one that performs the best in terms of a chosen metric. This process, which can only issue deterministic forecasts, is referred to as *method 0* in this paper.

Model chain was traditionally considered a deterministic tool, and it was not until the method recently proposed by Mayer and Yang [26] that the utilization of the model chain was extended for probabilistic irradiance-to-power conversion. The idea behind this extension is the poor man's ensemble (or model ensemble), where the same input data are converted to different outputs by different models. In the context of irradiance-to-power conversion, it means that the deterministic NWP forecasts are converted to PV power by a set of different model chains, which are then interpreted as an ensemble, describing a probability distribution of the power output. This approach is implemented in this paper and is referred to as *method 1* henceforth; without loss of generality, we randomly selected 50 model chains for method 1, as to match the number of ensemble members of the ECMWF ENS forecasts.

In methods 2 and 3, the irradiance-to-power conversion is performed similarly as in methods 0 and 1, respectively, with the difference that methods 2 and 3 are based on probabilistic instead of deterministic NWP forecasts. In the case of the ensemble members, it must be noted that they cannot be regarded as models, since the members for each run are subject to a random initialization independent of previous runs. In other words, members having the same index for two consecutive days (e.g., member 1 today and member 1 tomorrow) are unrelated and thus should not be grouped together during training. What should be done instead is to first translate the ensemble members into the quantiles of the predictive density by assuming equal probability for each member, which can be done simply by sorting the GHI member forecasts for each timestep. Afterward, instead of the original members, the sorted quantiles are used in all further calculations. In *method 2*, the GHI, ambient temperature, and wind speed forecasts of all 50 ENS quantiles are converted individually to PV power by a single model chain, namely, the one that was found to be the best for deterministic forecasting in method 0. In *method 3*, the NWP forecasts of the 50 ensemble quantiles are converted to PV power with 50 different, randomly selected model chains (1-to-1 paring, i.e., one model chain for each ENS quantile).

Raw ensemble forecasts are typically underdispersed, as they are unable to fully account for all sources of uncertainty during their creation [27]. Therefore, the reliability of the ensemble forecasts can only be ensured by a data-driven calibration procedure. In this paper, the calibration is achieved by performing quantile regression, which requires no assumption on the shape of the distribution of the power forecast error. Calibration requires historical data; therefore, it cannot be performed for newly commenced PV plants. Considering the calibration step, methods 1, 2, and 3 can all be further separated into two cases, distinguished by letters "R" and "C," which stand for the raw and calibrated ensemble, respectively. In the methods relying on the raw ensemble, the ensemble members are translated to quantiles using a uniform spacing of the cumulative probabilities [28].

The above-presented methods are visualized in Fig. 3. As mentioned in the introduction, the NWP forecasts and the irradiance-to-power conversion can both be deterministic or probabilistic, which results in four distinct combinations. Based on this, method 0 yields deterministic forecasts with no uncertainty quantification, while the forecasts created by method 1 only include the uncertainty of the PV plant modeling, the forecasts of method 2 only include the uncertainty of the weather prediction, while the forecasts of method 3 include the uncertainty of both sources. Methods 1R, 2R, and 3R, when paired with quantile regression, lead to three new methods (i.e., 1C, 2C, and 3C), making a total of seven methods.

Six out of the seven methods generate probabilistic forecasts; however, these can also be summarized into deterministic forecasts by assigning a statistical functional, e.g., the mean or the median, of the probability density function for each timestep. A sample time series plot of how the original probabilistic and the resulting deterministic forecasts are related to each other is shown in Fig. 4. All seven methods are capable of creating deterministic forecasts as well; therefore, they are also evaluated as deterministic forecasting methods, which enables one to quantify the added benefit of the ensemble NWP not only for probabilistic but also for deterministic PV power forecasting.

3.2. Physical photovoltaic model chains

Conversion of weather forecasts into PV power forecasts can be done by a successive set of models, which are constituents of a model chain. After solar positioning, which is essential for obtaining the Sun-position-related information such as zenith and incidence angles, the calculation starts with the separation of the GHI into the beam and diffuse components, which is then followed by transposing the horizontal irradiance components onto the tilted module plane. Afterward, the cell temperature is estimated from the incident irradiance, ambient temperature, and wind speed, and the power output of the PV modules is calculated by taking into account the dependence of the DC power output on the irradiance and cell temperature. Above these basic modeling steps, more sophisticated model chains also account for the reflection losses from the module surface, the inter-row shading, and the inverter, cable, and other losses.

The model chain framework used in this paper follows the implementation first proposed in Ref. [8], which has benefited several later studies [11,29]. As shown in Fig. 5, this model chain framework includes nine modeling steps, seven of which are considered as the main steps, where multiple component model options are included. The list of the component models of choice is summarized in Table 2. In summary, the nine separation, ten transposition, three reflection, five cell temperature, four PV efficiency, two shading, and three inverter models, result in a total of 32,400 different model chains. During the optimization of the model chain, the power forecasts are calculated by all these model chains for the year 2019, and the model chains that lead to the lowest mean absolute error (MAE) and root mean square error (RMSE) are selected as the two best options corresponding to these two verification directives. In the methods with randomly selected model chains, the used model chains are picked from these 32,400 variants with a uniform probability.

3.3. Quantile regression

Quantile regression (QR) maps the relationship between the predictors and a selected quantile of a probability distribution. To do this, the loss function of quantile regression is the pinball loss, which is calculated as

[3] For high-latitude region, surface albedo constitutes another essential input parameter, as excessive surface brightness due to white sand and snow can affect the ground-reflected irradiance substantially. But considering the geography of Hungary, surface albedo is less important a parameter, and thus is not considered.

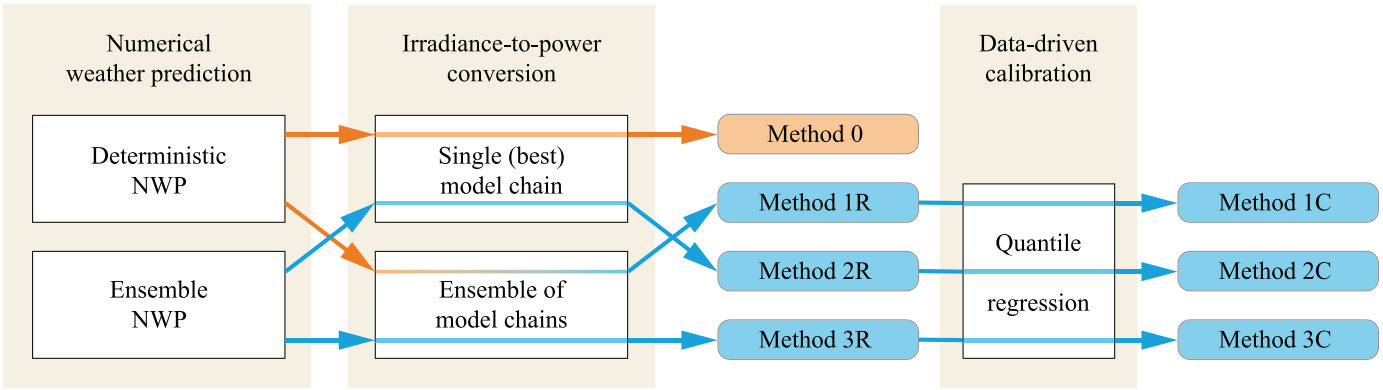


Fig. 3. Schematic of the different methods compared in this paper for PV power forecasting based on NWP. Orange and blue colors indicate deterministic and probabilistic forecasts, respectively.

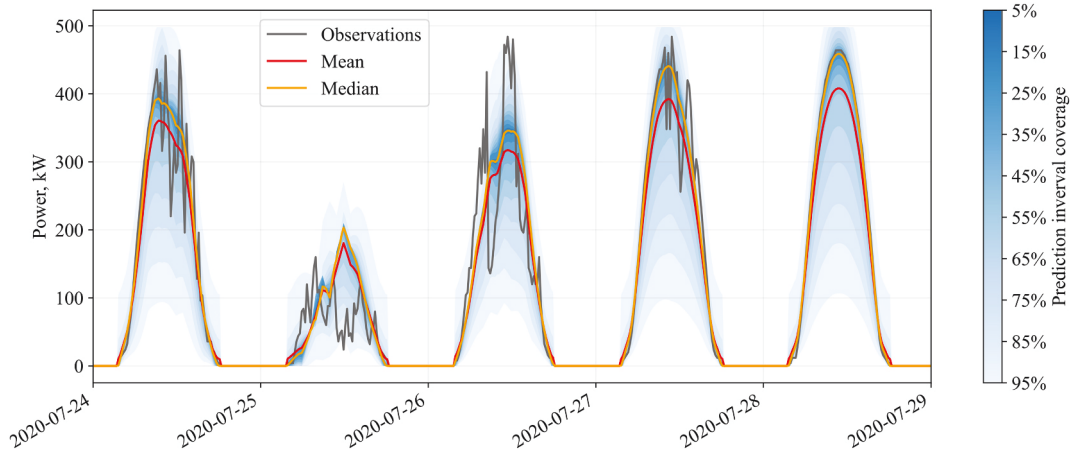


Fig. 4. Time series plot of the observations, the probabilistic forecasts created by the method 3C, and the deterministic forecasts created as the mean and the median of the predictive distribution for a sample period for the Kecel PV plant.

$$\psi_\tau = \begin{cases} \tau(y - x_\tau) & \text{if } x_\tau < y \\ (1-\tau)(x_\tau - y) & \text{if } x_\tau > y \end{cases}, \tag{2}$$

where y is the observation and x_τ is the forecast for quantile τ. QR can either be linear or nonlinear; however, as the model chains can account for the nonlinear dependence of the PV power on the weather forecasts, the linear regression is thought to be sufficient for calibration purposes. Moreover, nonlinear QR models can also learn the uncertainty from the historical data, while a linear QR only scales the spread of the ensemble. Considering this, linear QR better aligns with the main aim of the present paper, which is to gauge the capabilities of physical methods in probabilistic forecasting.

The predictors of the QR are the 50 PV power forecast ensemble members in the case of each method. The calibration is performed by calculating the 20 quantiles from 0.025 to 0.975 with 0.05 increments using the *QuantileRegressor* class of the *scikit-learn* Python package. The coefficients of the linear quantile regression models are fitted using the data from 2019. As a final post-processing step, the calculated quantiles are limited to the range of the physically possible power output of the PV plants, which is between zero and the nominal installed AC power.

3.4. Verification of deterministic power forecasts

In this paper, except for method 0, the deterministic forecasts are summarized from the predictive distribution of probabilistic forecasts either by using the mean or the median of the distribution as point forecasts. These two statistical functionals, however, correspond to two distinct directives. The mean value minimizes the mean square error (MSE), while the median minimizes the mean absolute error (MAE) of the forecasts [60]. These two directives are to create MSE-optimized forecasts and MAE-optimized forecasts, respectively. In forecast verification, the rule of consistency means that the forecasts should be evaluated using the same score as they are optimized for [61,62]. In that, considering these two directives, the two main error metrics used for deterministic forecast verification are the MAE and RMSE, calculated as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|f_i - x_i|, \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - x_i)^2}, \tag{4}$$

where f and x denote the forecast and observation, and N is the number of daytime data points.

The correlation coefficient reflects the association between the observations and the forecasts, and it can be calculated as

$$\rho = \frac{\mathrm{cov}(f,x)}{\sigma_f \sigma_x}, \tag{5}$$

where cov $(\cdot,\cdot)$ stands for the covariance and σ is the standard deviation. The correlation coefficient can also be seen as a measure of the potential skillfulness of forecasts [63]. On the one hand, this means it is less affected by the directive for which the forecasts are calibrated; it can be used as a more universal indicator of the goodness of forecasts [64]. On the other hand, the correlation coefficient does not carry any

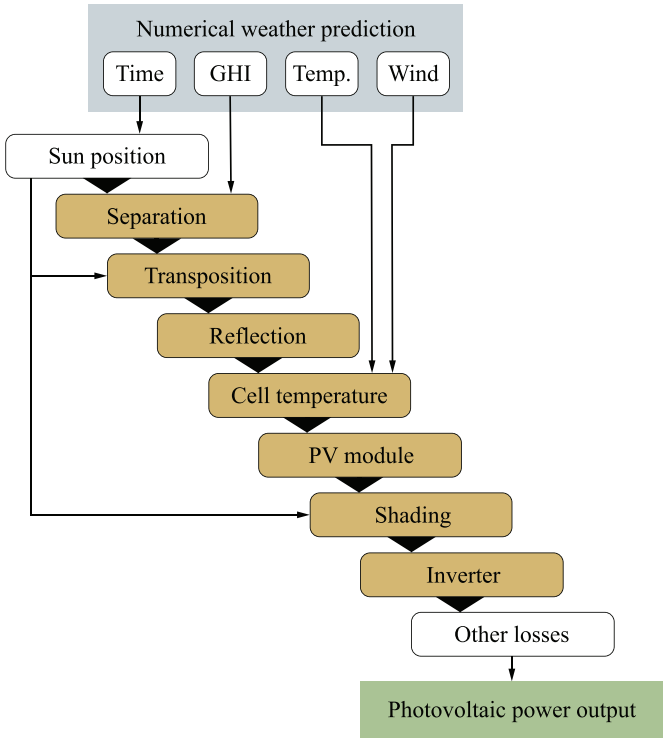


Fig. 5. The component models of the physical PV model chains for converting NWP into PV power output [8]. Yellow boxes represent the seven main modeling steps where multiple model variants are compared.

information about the bias and variance of the forecasts; it should be used along with other metrics that reflect the actual calibration of the forecasts. The unconditional bias between the observations and forecasts can be measured by the mean bias error (MBE), calculated as

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}(f_i - x_i). \tag{6}$$

The dispersion of the forecasts can be reflected by the variance ratio, as introduced in Ref. [8], which is the ratio of the variance of the forecasts and observations,

$$F = \frac{\mathbb{V}(f)}{\mathbb{V}(x)}, \tag{7}$$

where V (·) is the variance operator. A variance ratio lower and higher than 1 indicates underdispersion and overdispersion of the forecasts, respectively. Underdispersed forecasts are less capable of covering extremely low and high power outputs. However, as shown in Ref. [64], MSE-optimized forecasts are always underdispersed. Therefore, underdispersion, as long as it is done consciously, should not be considered a sign of bad forecasting performance.

Finally, perhaps the most general recommendation for deterministic solar forecast verification is to use the RMSE skill score to measure the overall skillfulness of the forecasts [19]. The skill score is calculated as the relative RMSE reduction compared to a naïve reference method,

$$s_{cp} = 1 - \frac{\mathrm{RMSE}_f}{\mathrm{RMSE}_{cp}}, \tag{8}$$

where the *cp* subscript stands for the CLIPER reference method, which is the optimal convex combination of climatology and persistence [65]. The RMSE skill score is only calculated for the MSE-optimized forecasts, as it is not consistent with the directive of the MAE-optimized forecasts.

Table 2
Modeling steps and component models of the physical model chains considered in this study [11].

Modeling steps	Component models	Ref.
Separation	Erbs	[30]
	Skartveit-Olseth	[31]
	DISC	[32]
	DIRINT	[33]
	DIRINDEX	[34]
	BRL	[35]
	Engerer	[36]
	Starke	[37]
	Abreu	[38]
Transposition	Liu-Jordan	[39]
	Steven	[40]
	Hay	[41]
	Willmot	[42]
	Skartveit-Olseth	[43]
	Gueymard	[44]
	Muneer	[45]
	Klucher	[46]
	Perez	[47]
	Reindl	[48]
Reflection	None	
	Martin-Ruiz	[49]
	Physical	[50]
Temperature	NOCT	[51]
	King	[52]
	Faiman	[53]
	Mattei	[54]
	Skoplaki	[55]
PV module	Evans	[56]
	Huld	[57]
	Single diode 4 par.	[58]
	Single diode 5 par.	[58]
Shading	None	
	Beam shading	[10]
Inverter	Constant	
	Quadratic	[59]
	Driesse	[59]

3.5. Verification of probabilistic power forecasts

The two main attributes that characterize the quality of probabilistic forecasts are reliability and sharpness. Reliability (or calibration) indicates the statistical consistency between the observations and the forecasts, which means that the difference between the nominal and observed probabilities should be small. A visual tool to evaluate reliability is a reliability diagram, which shows the proportions of the observations as the function of the nominal probability for all quantiles (see Fig. 6). For perfectly reliable forecasts, all points of the reliability

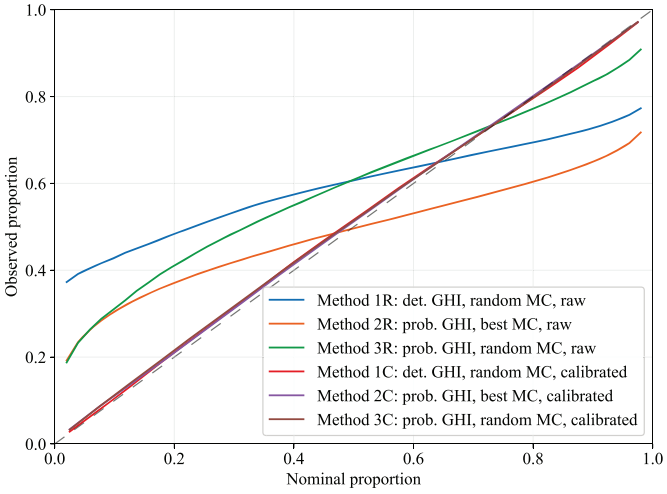


Fig. 6. Reliability diagram for all six methods. MC: model chain.

diagram should fall on the diagonal line. A commonly used metric to measure reliability is the prediction interval coverage probability (PICP). Nevertheless, its drawback is that it can suggest perfect reliability even if both quantiles defining the prediction interval are biased. To overcome this drawback, a new metric called mean absolute reliability error (MARE) is proposed in Ref. [26], which is calculated as the mean absolute error between the observed and nominal proportions for all quantiles, i.e.,

$$\mathrm{MARE} = \frac{1}{K}\sum_{k=1}^{K}\left|q^{k}_{\mathrm{obs}} - q^{k}_{\mathrm{nom}}\right|, \tag{9}$$

where K is the number of the calculated quantiles, and q^{k}_{obs} is the observed and q^{k}_{nom} is the nominal proportion at the k^{th} calculated quantile. Visually, MARE is the MAE of the calibration line compared to the diagonal in the reliability diagrams (see Fig. 6).

Sharpness expresses how informative the forecasts are, which is reflected by the concentration of the predictive distribution. Sharpness is commonly measured by the prediction interval average widths (PIAW), calculated as

$$\mathrm{PIAW} = \frac{1}{N}\sum_{i=1}^{N}(U_i - L_i), \tag{10}$$

where U_i and L_i are the upper and lower bounds of the prediction interval of interest. PIAW is calculated and averaged for all symmetric prediction intervals defined by the calculated quantiles, resulting in a $\overline{\mathrm{PIAW}}$. Moreover, the sharpness of all symmetric prediction intervals separately can be visualized by a sharpness diagram, which shows the PIAW as a function of the nominal coverage of the prediction intervals.

The overall performance of probabilistic forecasts is measured by scoring rules. The consistency, in this case, can be ensured by using strictly proper scoring rules, which assign the best score if the forecasts correspond to the best judgment of the forecasters [27]. The most commonly used strictly proper score is the continuous ranked probability score (CRPS), calculated as

$$\mathrm{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left[F^{y_i}(x) - \mathbf{1}(x - y_i)\right]^2 dx, \tag{11}$$

where $F^{y_i}(x)$ is the predictive distribution of y_i, $\mathbf{1}(x - y_i)$ is the Heaviside step function, and N is the number of data points. The CRPS can be decomposed into reliability (REL), uncertainty (UNC), and resolution (RES) components [66], that is,

$$\mathrm{CRPS} = \mathrm{REL} + \mathrm{UNC} - \mathrm{RES}. \tag{12}$$

The UNC term depends only on the variability of the observations; therefore, it cannot be affected by the forecaster. The REL component quantifies the contribution of reliability to the CRPS, with a lower value representing more reliable forecasts. The RES term expresses the ability to issue case-dependent forecasts, i.e., different distributions are forecast for different conditions. Even though their definition is different, the resolution is closely related to (but not the same as) sharpness for reliable forecasts, since the more case-dependent the forecasts are, the more concentrated their distribution can be [28].

Quantile score (QS) measures the accuracy of a specific quantile, and it is calculated for each quantile as the average of the pinball loss for the whole dataset.

$$\mathrm{QS}_{\tau} = \frac{1}{N}\sum_{i=1}^{N}\psi_{\tau} \tag{13}$$

The weighted sum of the QS for all quantiles is equal to half of the CRPS [67]; therefore, a diagram of QS as a function of the quantiles can reveal how the estimations of the different quantiles contribute to the CRPS.

4. Results and discussion

All results presented herein are calculated for the 15-min forecasts and observations covering the whole year of 2020. The generation and verification of all forecasts are performed individually for each of the 14 PV plants. The relative performance of all methods is very similar at each PV plant. For that reason, the results are not presented for the individual plants, but averaged over all of them to increase their overall reliability. As three PV plants have significantly higher nominal power compared to others, all metrics that are proportional to the nominal power (namely the MAE, RMSE, MBE, PIAW, CRPS, REL, RES, and QS) are normalized to the mean daytime power output of the PV plants before averaging to ensure that all PV plants equally contribute to the presented metrics.

First, the results for probabilistic forecasting are presented in Section 4.1, followed by the verification results for deterministic forecasting in Section 4.2.

4.1. Probabilistic PV power forecasting

The averaged metrics used for the verification of probabilistic forecasts are presented in Table 3. First, the reliability of the forecasts created by the different methods is assessed. Both the REL component of the CRPS and the MARE shows that the raw ensembles are unreliable, which is also supported by the reliability diagram in Fig. 6. Method 1R, which only covers the uncertainty of the irradiance-to-power conversion, is the least reliable, followed by method 2R, that only accounts for the uncertainty of the NWP, while among the raw ensembles, the most reliable is method 3R, which includes the uncertainty from both sources. This supports the belief that the more sources of uncertainty are included in the ensemble, the more reliable it can be [27].

The linear calibration largely improves the reliability of the forecasts. Based on the MARE, the three methods are almost equally reliable, while based on the REL component, method 1C is slightly more reliable than the others. The reliability diagram in Fig. 6 fully supports the conclusion drawn from the metrics. The calibration curves for methods 1C, 2C, and 3C largely overlap, which suggests that QR, based on a full year of fitting data, is an effective tool for calibration, regardless of the reliability of the raw ensemble. For all three calibrated methods, there is a small positive bias on the quantiles up to around 0.8 cumulative probability, which supposedly originates from the inter-annual differences in the statistical characteristics of the solar resource, and it is expected to decrease as the length of historical data increases.

The sharpness, as measured by the $\overline{\mathrm{PIAW}}$, varies inversely with the reliability in the case of five out of these six methods (the exception being method 3C), i.e., the more reliable forecasts are generally less sharp. Gneiting's paradigm for probabilistic forecasting is to maximize sharpness subject to calibration [27], which suggests that reliability is a prerequisite for good forecasts, whereas sharpness can be used to identify the best among the reliable forecasts. Following this train of thought, only the three calibrated methods can be accepted to be suitable for probabilistic forecasting, and among these three, method 3C is the sharpest, thus the most recommended. A more detailed picture of the sharpness of the forecasts created by the different methods can be seen in the sharpness diagram in Fig. 7. The curves for the presented six methods do not cross each other, which means that the forecasts that are sharper on average are also sharper on all prediction intervals. The resolution, as measured by the RES component, of the forecasts correlates well with their average sharpness.

The forecasts with the lowest CRPS are created by method 3C, which is a clear consequence of its good reliability and its best resolution among the calibrated methods. It is followed by methods 2C and 1C with a 2.2–3.5% relative increase in the CRPS, respectively. The relatively good performance of method 1C is remarkable, as it achieves this good performance without the need for ensemble NWP forecasts [26], which

Table 3
Verification metrics of the probabilistic forecasts created with the six different methods averaged for all PV plants. Bold fonts stand for the column-wise best values.

Method	QR	NWP	Model chain	CRPS	REL	RES	MARE	$\overline{\text{PIAW}}$
1R	Raw	Point	Random	26.8%	21.8%	**45.2%**	0.176	**8.4%**
2R		Ensemble	Single best	20.6%	7.0%	36.5%	0.145	24.9%
3R			Random	21.3%	6.7%	35.6%	0.120	28.4%
1C	Calibrated	Point	Random	19.1%	**0.6%**	31.7%	0.014	53.7%
2C		Ensemble	Single best	18.8%	0.8%	32.1%	**0.013**	51.5%
3C			Random	**18.4%**	0.7%	32.4%	0.014	50.7%

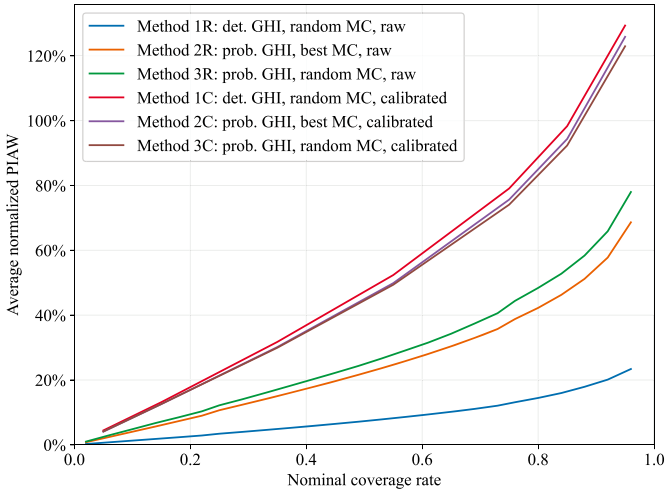


Fig. 7. Normalized PIAW, averaged for all PV plants, as a function of the PI nominal coverage rate for all six methods. MC: model chain.

carries much significance in practical applications where probabilistic irradiance forecasts are not available. Among the raw forecasts, method 2R has the lowest CRPS, which is (in relative terms) 12.0% higher than the overall lowest CRPS (method 3C). This suggests that, even if the calibration is not possible due to the lack of historical data, the ensemble NWP carries enough information about the expected uncertainties to make probabilistic forecasts with a relatively low CRPS—even though these forecasts, as shown above, are highly unreliable.

The QS, as a function of the quantiles, is shown in Fig. 8. The overall quadratic shape of the curve with the highest values around the median is typical for the QS [28], and thus this observation does not require further explanation. Method 3C has the lowest score for all quantiles,

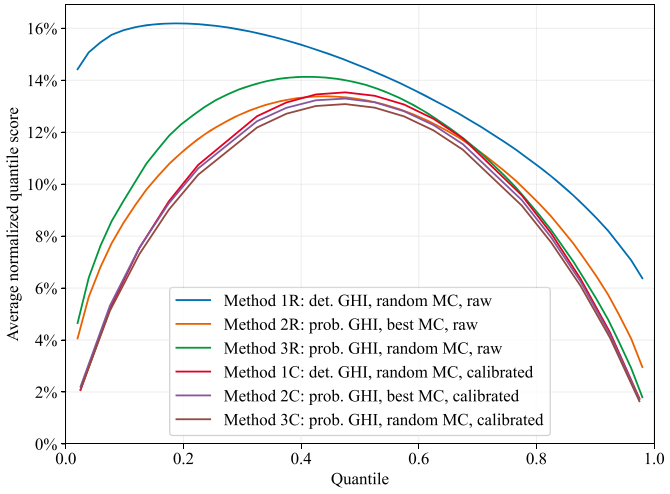


Fig. 8. Normalized quantile score, averaged for all PV plants, as a function of the quantiles for all six methods. MC: model chain.

which is in line with its lowest CRPS. All curves show some degree of asymmetry, which is more pronounced for the raw methods, reflecting that the forecasts for the higher quantiles are slightly more accurate than the forecasts for the lower quantiles. The reason for this should be sought in the propagation of the errors through the irradiance-to-power conversion process, as analyzed in detail in Ref. [68]. Particularly, the power outputs higher than the nominal AC power of the PV plants are clipped by the inverter, which means that the errors in the high irradiance domain are effectively reduced or even eliminated by the PV plant, making it easier to estimate the higher than the lower quantiles.

4.2. *Deterministic PV power forecasting*

The verification results for the seven methods for deterministic forecasting are summarized in Table 4, in parallel for the MAE-optimized and MSE-optimized forecasts. For all methods, the MAE-optimized forecasts have lower MAEs, and the MSE-optimized forecasts have lower MSEs compared to the forecasts optimized for the other directive, which confirms the expectations and justifies the importance of the optimization. In all cases, it is important to keep the rule of consistency in mind, which means that the predictive distribution should always be summarized by the elicitable functional that corresponds to the directive, i.e., the median and mean for the MAE and MSE-optimized forecasts, respectively.

The best deterministic forecasts, optimized either for MAE or MSE, are created by method 3C, the one that also provides the most accurate probabilistic forecasts. Among the uncalibrated methods, method 2R also offers a decent accuracy with only a 1.8% higher MAE and 1.3% higher RMSE compared to method 3C, which might suggest that calibration is only of marginal importance to deterministic forecasting. However, it must be seen that method 3R has significantly higher errors, which means that using random model chains is only advised if calibration is possible; otherwise, it is better to stay with the single optimized model chain for all ensemble NWP members. The added value of the ensemble NWP in the deterministic PV power forecasting is in line with the GHI forecast verification results, shown in Fig. 2, which revealed that both the mean and the median of the ensemble are more accurate than the control forecast.

If only deterministic NWP is available, method 1C can still offer a 4.5% MAE and 4.6% RMSE reduction as compared to the base case of method 0. The calibration is an important contributor to the good performance of this method, since its uncalibrated version (method 1R) has the highest errors in all respects, which is coherent with the findings of [26]. The calibration also decreases the quite significant positive bias of the raw forecasts, and it also helps to set the variance ratio close to its optimal value. For example, the optimal variance ratio of an MSE-optimized forecast with a 0.883 correlation coefficient is around 78% (which is the square of the correlation coefficient), which is indeed close to the actual variance ratio of the forecasts created with method 3C. For the best methods, the RMSE skill scores over the CLIPER reference exceed 40%, which is even higher than the skill of other state-of-the-art PV power forecasting methods, e.g., the hybridization of physical model chains with machine learning [29], which can be attributed to the known good accuracy of the ECMWF GHI forecasts [69], the added value of the ensemble NWP, and the effectiveness of the

Table 4
Verification metrics of the deterministic forecasts created with the seven different methods averaged for all PV plants. Bold fonts stand for the column-wise best values. MBE, MAE, and RMSE are presented as a percentage of the mean of the observations.

Method	QR	NWP	Model chain	MAE-optimized					MSE-optimized					
				MAE	MBE	RMSE	ρ	F	RMSE	MBE	MAE	ρ	F	s_{cp}
0	Raw	Point	Single best	27.8%	5.2%	45.7%	86.7%	89.6%	44.8%	2.7%	29.1%	86.8%	77.6%	36.6%
1R			Random	29.1%	11.1%	48.0%	86.0%	92.5%	48.1%	11.2%	29.2%	86.0%	92.2%	32.2%
2R		Ensemble	Single best	26.5%	5.6%	43.5%	87.9%	88.1%	42.7%	4.7%	27.0%	88.2%	83.5%	39.8%
3R			Random	27.8%	11.3%	45.7%	87.4%	91.0%	44.8%	10.8%	28.1%	87.7%	86.1%	36.9%
1C	Calib-rated	Point	Random	27.0%	2.6%	44.5%	87.2%	89.5%	43.5%	−0.3%	29.2%	87.6%	73.2%	38.8%
2C		Ensemble	Single best	26.5%	1.2%	43.3%	87.8%	86.9%	42.5%	−0.7%	28.5%	88.1%	72.3%	40.1%
3C			Random	**26.1%**	1.5%	43.1%	88.0%	89.1%	**42.1%**	−0.5%	27.9%	88.3%	74.8%	40.6%

presented methods.

5. Conclusion

The main aim of this paper is to discover the most effective probabilistic photovoltaic (PV) power forecasting method relying on physical modeling approaches, namely, numerical weather prediction (NWP) for forecasting solar irradiance and other relevant weather variables, and model chains to simulate the operation of PV plants and covert the weather forecasts to PV power. In this two-step general framework, the uncertainty information can be drawn from either (or both) an ensemble NWP or an ensemble of model chains.

The performance of the raw and calibrated versions of three distinct methods are evaluated empirically at 14 utility-scale ground-mounted PV plants in Hungary, based on the ensemble NWP forecasts of the ECMWF. The results revealed that the best approach for probabilistic forecasting, in terms of the lowest CRPS, is to convert all ensemble NWP quantiles into PV power by randomly selected model chains and calibrate the resulting ensemble by a linear quantile regression (QR). Alternatively, if no ensemble NWP is available, the method proposed in Ref. [26], which includes the conversion of the deterministic NWP into power by an ensemble of randomly selected model chains and the calibration by a QR, is also favorable, with only a 3.5% CRPS increase compared to the best approach. Alternatively, if the calibration is not possible due to the lack of at least a one-year-long historical dataset, ensemble NWP forecasts could still be converted to PV power forecasts by an optimized model chain, yielding a CRPS 12% higher than the best case. However, the resulting power forecasts would be underdispersed, as the reliability of the probabilistic forecasts cannot be ensured without calibration, even if the uncertainties originated from both the NWP and the model chains are included.

Since any probabilistic forecasts can be easily turned into deterministic forecasts by summarizing their predictive distributions, the presented methods are also evaluated for their deterministic forecasting performance. The results reveal that the method found to be the best for probabilistic is also the most accurate for deterministic forecasting. An advantage of creating probabilistic forecasts first, even if the ultimate goal is deterministic forecasting, is that it does not depend on a directive, e.g., on whether MAE-optimized or MSE-optimized forecasts are required. Stated differently, the directive needs not to be known in advance, but it can be decided in the last step, in which the predictive distributions are summarized. A clear benefit of this strategy is that the forecaster does not require to take care of the directives of the users, but it is possible to disseminate the same probabilistic forecasts to many different users, who can then convert them to such deterministic forecasts that correspond to their own directives.

Credit author statement

Martin János Mayer: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Software, Data curation, Writing – original draft preparation, Visualization. **Dazhi Yang**: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft preparation.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgment

The authors thank Norbert Péter from the MVM Green Generation Ltd. for the PV plant design and production data. This paper was supported by the National Research, Development and Innovation Fund, project no. OTKA-FK 142702, ÚNKP-22-5-BME-305, and TKP2021 grant no. BME-NVA-02, and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

European Centre for Medium-range Weather Forecast (ECMWF) operational forecast and analysis data used in this study were downloaded from the ECMWF Meteorological Archival and Retrieval System (MARS) in 2022. Access to the ECMWF archived data was provided by ECMWF's Data Services. Special thanks go to Emma Pidduck, Ruth Coughlan, and Ilaria Parodi from the ECMWF's Data Services team, for their swift communication regarding data access.

References

[1] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: towards carbon neutrality. Renew Sustain Energy Rev 2022;161:112348. https://doi.org/10.1016/j.rser.2022.112348.
[2] Yang D, Wang W, Xia X. A concise overview on solar resource assessment and forecasting. Adv Atmos Sci 2022;39:1239–51. https://doi.org/10.1007/s00376-021-1372-8.
[3] Hong T, Pinson P, Wang Y, Weron R, Yang D, Zareipour H. Energy forecasting: a review and outlook. IEEE Open Access J Power Energy 2020;7:376–88. https://doi.org/10.1109/OAJPE.2020.3029979.
[4] Yang D, van der Meer D. Post-processing in solar forecasting: ten overarching thinking tools. Renew Sustain Energy Rev 2021;140:110735. https://doi.org/10.1016/j.rser.2021.110735.
[5] Schulz B, El Ayari M, Lerch S, Baran S. Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting. Sol Energy 2021;220:1016–31. https://doi.org/10.1016/j.solener.2021.03.023.
[6] Leutbecher M, Palmer TN. Ensemble forecasting. J Comput Phys 2008;227:3515–39. https://doi.org/10.1016/j.jcp.2007.02.014.
[7] Du J, Berner J, Buizza R, Charron M, Houtekamer P, Hou D, et al. Ensemble methods for meteorological predictions. Handb. Hydrometeorol. Ensemble forecast. Berlin, Heidelberg: Springer Berlin Heidelberg; 2019. p. 99–149. https://doi.org/10.1007/978-3-642-39925-1_13.
[8] Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. Appl Energy 2021;283:116239. https://doi.org/10.1016/j.apenergy.2020.116239.
[9] Mayer MJ. Design optimization and power forecasting of photovoltaic power plants. Budapest University of Technology and Economics; 2020. http://hdl.handle.net/10890/15112.

[10] Mayer MJ, Gróf G. Techno-economic optimization of grid-connected, ground-mounted photovoltaic power plants by genetic algorithm based on a comprehensive mathematical model. Sol Energy 2020;202:210–26. https://doi.org/10.1016/j.solener.2020.03.109.

[11] Mayer MJ. Influence of design data availability on the accuracy of physical photovoltaic power forecasts. Sol Energy 2021;227:532–40. https://doi.org/10.1016/j.solener.2021.09.044.

[12] Sun X, Yang D, Gueymard CA, Bright JM, Wang P. Effects of spatial scale of atmospheric reanalysis data on clear-sky surface radiation modeling in tropical climates: a case study for Singapore. Sol Energy 2022;241:525–37. https://doi.org/10.1016/j.solener.2022.06.001.

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

[14] Bonavita M, Isaksen L, Hólm E. On the use of EDA background error variances in the ECMWF 4D-Var. Q J R Meteorol Soc 2012;138:1540–59. https://doi.org/10.1002/QJ.1899.

[15] Diaconescu EP, Laprise R. Singular vectors in atmospheric sciences: a review. Earth Sci Rev 2012;113:161–75. https://doi.org/10.1016/J.EARSCIREV.2012.05.005.

[16] Buizza R, Leutbecher M, Isaksen L. Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. Q J R Meteorol Soc 2008;134:2051–66. https://doi.org/10.1002/QJ.346.

[17] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos Meas Tech 2013;6:2403–18. https://doi.org/10.5194/AMT-6-2403-2013.

[18] Gschwind B, Wald L, Blanc P, Lefèvre M, Schroedter-Homscheidt M, Arola A. Improving the McClear model estimating the downwelling solar radiation at ground level in cloud-free conditions – McClear-v3. Meteorol Z 2019;28:147–63. https://doi.org/10.1127/metz/2019/0946.

[19] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, et al. Verification of deterministic solar forecasts. Sol Energy 2020;210:20–37. https://doi.org/10.1016/j.solener.2020.04.019.

[20] Yang D. Choice of clear-sky model in solar forecasting. J Renew Sustain Energy 2020;12:026101. https://doi.org/10.1063/5.0003495.

[21] Gilleland E, Ahijevych DA, Brown BG, Ebert EE. Verifying forecasts spatially. Bull Am Meteorol Soc 2010;91:1365–76. https://doi.org/10.1175/2010BAMS2819.1.

[22] Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. Intercomparison of spatial forecast verification methods. Weather Forecast 2009;24:1416–30. https://doi.org/10.1175/2009WAF2222269.1.

[23] Perez R, Schlemmer J, Hemker K, Kivalov S, Kankiewicz A, Dise J. Solar energy forecast validation for extended areas & economic impact of forecast accuracy. Conf Rec IEEE Photovolt Spec Conf 2016:1119–24. https://doi.org/10.1109/PVSC.2016.7749787. 2016-November.

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

[25] Yang D, Bright JM. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: a preliminary evaluation and overall metrics for hourly data over 27 years. Sol Energy 2020;210:3–19. https://doi.org/10.1016/j.solener.2020.04.016.

[26] Mayer MJ, Yang D. Probabilistic photovoltaic power forecasting using a calibrated ensemble of model chains. Renew Sustain Energy Rev 2022;168:112821. https://doi.org/10.1016/j.rser.2022.112821.

[27] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 2007;102:359–78. https://doi.org/10.1198/016214506000001437.

[28] Lauret P, David M, Pinson P. Verification of solar irradiance probabilistic forecasts. Sol Energy 2019;194:254–71. https://doi.org/10.1016/j.solener.2019.10.041.

[29] Mayer MJ. Benefits of physical and machine learning hybridization for photovoltaic power forecasting. Renew Sustain Energy Rev 2022;168:112772. https://doi.org/10.1016/j.rser.2022.112772.

[30] Erbs DG, Klein SA, Duffie JA. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Sol Energy 1982;28:293–302. https://doi.org/10.1016/0038-092X(82)90302-4.

[31] Skartveit A, Olseth JA. A model for the diffuse fraction of hourly global radiation. Sol Energy 1987;38:271–4. https://doi.org/10.1016/0038-092X(87)90049-1.

[32] Maxwell EL. A quasi-physical model for converting hourly global horizontal to direct normal insolation. Golden, CO. 1987.

[33] Perez RR, Ineichen P, Maxwell EL, Seals RD, Zelenka A. Dynamic global-to-direct irradiance conversion models. ASHRAE Trans 1992;98:354–69.

[34] Perez R, Ineichen P, Moore K, Kmiecik M, Chain C, George R, et al. A new operational model for satellite-derived irradiances: description and validation. Sol Energy 2002;73:307–17. https://doi.org/10.1016/S0038-092X(02)00122-6.

[35] Ridley B, Boland J, Lauret P. Modelling of diffuse solar fraction with multiple predictors. Renew Energy 2010;35:478–83. https://doi.org/10.1016/j.renene.2009.07.018.

[36] Engerer NA. Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia. Sol Energy 2015;116:215–37. https://doi.org/10.1016/j.solener.2015.04.012.

[37] Starke AR, Lemos LFL, Boland J, Cardemil JM, Colle S. Resolution of the cloud enhancement problem for one-minute diffuse radiation prediction. Renew Energy 2018;125:472–84. https://doi.org/10.1016/j.renene.2018.02.107.

[38] Abreu EFM, Canhoto P, Costa MJ. Prediction of diffuse horizontal irradiance using a new climate zone model. Renew Sustain Energy Rev 2019;110:28–42. https://doi.org/10.1016/j.rser.2019.04.055.

[39] Liu B, Jordan R. Daily insolation on surfaces tilted towards equator. ASHRAE J 1961;3:53–9.

[40] Steven MD, Unsworth MH. The diffuse solar irradiance of slopes under cloudless skies. Q J R Meteorol Soc 1979;105:593–602.

[41] Hay JE, Davies JA. Calculation of the solar irradiance incident on an inclined surface. In: Hay JE, Won TK, editors. First can. Sol. Radiat. Data work. Toronto, Ontario: Canada; 1980. p. 59–72.

[42] Willmott CJ. On the climatic optimization of the tilt and azimuth of flat-plate solar collectors. Sol Energy 1982;28:205–16. https://doi.org/10.1016/0038-092X(82)90159-1.

[43] Skartveit A, Olseth JA. Modelling slope irradiance at high latitudes. Sol Energy 1986;36:333–44. https://doi.org/10.1016/0038-092X(86)90151-9.

[44] Gueymard C. An anisotropic solar irradiance model for tilted surfaces and its comparison with selected engineering algorithms. Sol Energy 1987;38:367–86. https://doi.org/10.1016/0038-092X(87)90009-0.

[45] Muneer T. Solar radiation model for Europe. Build Serv Eng Technol 1990;11:153–63. https://doi.org/10.1177/014362449001100405.

[46] Klucher TM. Evaluation of models to predict insolation on tilted surfaces. Sol Energy 1979;23:111–4. https://doi.org/10.1016/0038-092X(79)90110-5.

[47] Perez R, Ineichen P, Seals R, Michalsky J, Stewart R. Modeling daylight availability and irradiance components from direct and global irradiance. Sol Energy 1990;44:271–89. https://doi.org/10.1016/0038-092X(90)90055-H.

[48] Reindl DT, Beckman WA, Duffie JA. Evaluation of hourly tilted surface radiation models. Sol Energy 1990;45:9–17. https://doi.org/10.1016/0038-092X(90)90061-G.

[49] Martin N, Ruiz JM. Calculation of the PV modules angular losses under field conditions by means of an analytical model. Sol Energy Mater Sol Cells 2001;70:25–38. https://doi.org/10.1016/S0927-0248(00)00408-6.

[50] Marion B. Numerical method for angle-of-incidence correction factors for diffuse radiation incident photovoltaic modules. Sol Energy 2017;147:344–8. https://doi.org/10.1016/j.solener.2017.03.027.

[51] Ross AG. Flat-Plate photovoltaic module and array engineering. In: Proc. 1982 annu. Meet. Am. Sect. Int. Sol. Energy soc.; 1982. p. 4321–4. Houston, Texas.

[52] King DL, Boyson WE, Kratochvil JA. Photovoltaic array performance model. Sandia Rep No 2004-3535 2004;8:1–19. doi:10.2172/919131.

[53] Faiman D. Assessing the outdoor operating temperature of photovoltaic modules. Prog Photovoltaics Res Appl 2008;16:307–15. https://doi.org/10.1002/pip.813.

[54] Mattei M, Notton G, Cristofari C, Muselli M, Poggi P. Calculation of the polycrystalline PV module temperature using a simple method of energy balance. Renew Energy 2006;31:553–67. https://doi.org/10.1016/j.renene.2005.03.010.

[55] Skoplaki E, Boudouvis AG, Palyvos JA. A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting. Sol Energy Mater Sol Cells 2008;92:1393–402. https://doi.org/10.1016/j.solmat.2008.05.016.

[56] Evans DL, Florschuetz LW. Cost studies on terrestrial photovoltaic power systems with sunlight concentration. Sol Energy 1977;19:255–62. https://doi.org/10.1016/0038-092X(77)90068-8.

[57] Huld T, Friesen G, Skoczek A, Kenny RP, Sample T, Field M, et al. A power-rating model for crystalline silicon PV modules. Sol Energy Mater Sol Cells 2011;95:3359–69. https://doi.org/10.1016/j.solmat.2011.07.026.

[58] De Soto W, Klein SA, Beckman WA. Improvement and validation of a model for photovoltaic array performance. Sol Energy 2006;80:78–88. https://doi.org/10.1016/j.solener.2005.06.010.

[59] Driesse A, Jain P, Harrison S. Beyond the curves: modeling the electrical efficiency of photovoltaic inverters. In: 2008 33rd IEEE photovolatic spec. Conf. IEEE; 2008. p. 1–6. https://doi.org/10.1109/PVSC.2008.4922827.

[60] Gneiting T. Making and evaluating point forecasts. J Am Stat Assoc 2011;106:746–62. https://doi.org/10.1198/jasa.2011.r10138.

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

[62] Kolassa S. Why the "best" point forecast depends on the error or accuracy measure. Int J Forecast 2020;36:208–11. https://doi.org/10.1016/j.ijforecast.2019.02.017.

[63] Murphy AH, Epstein ES. Skill scores and correlation coefficients in model verification. Mon Weather Rev 1989;117:572–82. https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2.

[64] Mayer MJ, Yang D. Calibration of deterministic NWP forecasts and its impact on verification. Int J Forecast 2022. https://doi.org/10.1016/j.ijforecast.2022.03.008.

[65] Yang D. Standard of reference in operational day-ahead deterministic solar forecasting. J Renew Sustain Energy 2019;11:053702. https://doi.org/10.1063/1.5114985.

[66] Hersbach H. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast 2000;15:559–70. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

[67] Bröcker J. Evaluating raw ensembles with the continuous ranked probability score. Q J R Meteorol Soc 2012;138:1611–7. https://doi.org/10.1002/qj.1891.

[68] Mayer MJ. Impact of the tilt angle, inverter sizing factor and row spacing on the photovoltaic power forecast accuracy. Appl Energy 2022;323:119598. https://doi.org/10.1016/j.apenergy.2022.119598.

[69] Yang D, Wang W, Bright JM, Voyant C, Notton G, Zhang G, et al. Verifying operational intra-day solar forecasts from ECMWF and NOAA. Sol Energy 2022;236:743–55. https://doi.org/10.1016/j.solener.2022.03.004.