Applied Energy 371 (2024) 123681

Contents lists available at ScienceDirect

Applied Energy

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

Optimal place to apply post-processing in the deterministic photovoltaic power forecasting workflow

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

HIGHLIGHTS

- Post-processing is best performed on PV power instead of GHI
- Model chains change the bias and variance of the input forecasts
- Post-processing methods developed for GHI are transferable to PV power

ARTICLE INFO

Keywords:
Post-processing
Solar forecasting
Photovoltaic power
Model output statistics
Kernel conditional density estimation

ABSTRACT

Is the post-processing of global horizontal irradiance (GHI) forecasts necessary for issuing good photovoltaic (PV) power forecasts? Whenever this question is raised, the instinctive supposition always seems to be "yes," because GHI is the most important weather parameter governing the amount of PV power generated, and surely, the better the GHI forecasts are, the better the PV power forecasts should result. To attend to this question more scientifically and more formally, two classic deterministic-to-deterministic post-processing methods, namely, the model output statistics and kernel conditional density estimation, are applied at various stages of PV power forecasting, resulting in four distinct workflows. These different workflows are trained and tested on three PV plants in Hungary, using data from a four-year (2017–2020) period. Both ground-based GHI and satellite-retrieved GHI are used as the "truth" with which numerical weather prediction (NWP) GHI forecasts are post-processed. A very thorough deterministic forecast verification exercise is conducted following the best practices. It is found that contrary to the common supposition, post-processing GHI only leads to marginal, if that can be quantified at all, benefits, so long as the PV power forecasts are to be post-processed. This ought to be deemed as a very important finding, as it puts into question the "GHI forecasting + post-processing + irradiance-to-power conversion" workflow that has dominated solar forecasting for decades.

1. Introduction

State-of-the-art photovoltaic (PV) power forecasting includes two main steps, of which the first is forecasting the solar irradiance, and the second is converting the irradiance forecasts to power forecasts via a solar power curve [1]. This process is widely referred to as the two-step framework of PV power forecasting, in contrast to the single-step framework, in which the PV power forecasts are directly extrapolated from the historical power output and other supporting data. The preference for the two-step framework is a priori, as the irradiance forecasting procedure, such as camera-based ray tracing, satellite-based cloud motion advection/diffusion, or numerical weather prediction (NWP), is in the main of a physical one that seeks to exploit the physical laws governing the atmospheric processes governing the amount of radiation reaching the earth's surface. On the other hand, the single-step extrapolation is of a data-driven nature [2], incapable of capturing the cloud motion in most cases, and thus is often confined in terms of forecasting performance [3,4]. For this reason, this work should only be concerned with the two-step PV power forecasting workflow.

Inasmuch as irradiance forecasts are to be issued, the methods to do so should be referred to as *base methods*. As mentioned above, base methods leverage information collected by all-sky cameras,

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

https://doi.org/10.1016/j.apenergy.2024.123681
Received 6 December 2023; Received in revised form 13 May 2024; Accepted 9 June 2024

geostationary weather satellites, and numerical weather models, and make forecasts following the underlying physical principles as closely as possible. Given the spatial and temporal resolutions of those data sources, different base methods have long been known to suit forecasting over different horizons [5]. Whereas the reader is referred to [6] for the latest schematic showing the correspondence between base methods and forecast horizon, it is worth noting that NWP-based irradiance forecasting has been widely recognized as the chief method for purposes concerning grid integration of solar energy [7]. Modern NWP is able to issue sub-hourly and hourly irradiance forecasts, with a high refresh rate, over wide geographical regions, and in a probabilistic style [8–10]. As such, NWP-based irradiance forecasts are used as the basis for the present work.

Moving beyond irradiance forecasting, methods to convert them into PV power forecasts are ample. Collectively known as the solar power curves, the irradiance-to-power conversion methods could be broadly categorized into direct, indirect, and hybrid ones. Whereas direct methods seek to construct a mapping between forecast weather variables and PV power via some form of regression [11], e.g., with the currently popular deep learning methods [12,13], the indirect methods utilize the so-called *model chain* for the conversion in a step-by-step fashion [14]. Since most solar engineers should already be well accustomed to the concept and use of model chains, such information is not reiterated here, but those who are unfamiliar with it are referred to the series of papers by Mayer and co-workers [15–19]. As for hybrid solar power curves, the idea is to utilize the model chain to a certain stage and leave the rest to regression, in that, some previous investigations are also available [20]. There are, at the moment, no results that are absolutely conclusive on the relative performance of these three classes of solar power curve modeling methods. However, the general guideline is to consider the data at hand: For newly commissioned PV plants, a model chain is the only possibility, whereas if some historical data are available, one may proceed with direct and hybrid approaches.

Besides irradiance forecasting and solar power curve modeling, a third step of statistical post-processing is also often included in the PV power forecasting workflow to reduce, if not eliminate, the conditional bias in initial forecasts, improve calibration and resolution, and thereby increase the quality of the final forecasts. In this regard, this post-processing-extended process may be called the three-step PV forecasting framework. Most of the post-processing literature deals with correcting the global horizontal irradiance (GHI) forecasts; therefore, the widely accepted sequence of the main steps is: (1) irradiance forecasting, (2) post-processing, and (3) irradiance-to-power conversion [6]. However, insofar as the applicability of the methodology is concerned, post-processing could also be performed on the final PV power output, making it the last step of the workflow. At the moment, the larger number of studies dealing with the post-processing of solar irradiance than that of PV power can be explained, to a large extent, by the better data availability of the former. Whereas GHI observation and forecast data can be obtained from several public sources, see [8,10,21–23] for examples, the public availability of PV power production data has hitherto been scarce due to propriety reasons and the strict data privacy policies of most PV plant owner companies. Another more implicit reason for the higher popularity of GHI forecasting is that solar power curve modeling, which belongs to the subject area of solar energy meteorology, concerns mostly just solar engineers. In contrast, post-processing of weather variables (i.e., GHI in this case) is a general task that could attract combined attention from researchers with diverse backgrounds, e.g., meteorology, engineering, or statistics. That said, most of the methods developed for GHI post-processing can also be applied to PV power forecasts without significant modifications.

The errors, including their bias, of PV power forecasts largely depend on the model chain, which can be constructed in many different ways depending on which component models are used in the main modeling steps. Mayer and Gróf [14] compared 32,400 different model chains for PV power forecasting from NWP forecasts and found that the mean-normalized mean bias error (MBE) of the PV power forecasts varies in a wide range of −11.3% to 10.6% depending on which model chain is used. However, there is no universal best model chain, but it depends on the GHI forecasts and the PV plants. Consequently, the best model chain for the PV plant and data at hand can only be found by a so-called "model chain optimization" process, which relies on the testing of a large number of different model chains [20]. Model chain optimization is important for obtaining the most accurate forecasts, but it further complicates the PV power forecasting workflow, which is a disadvantage in such applications where ease of use is of high importance. The bias introduced by the model chain also depends on the design parameters of the PV plant, most importantly on the tilt angle and inverter sizing factor, as revealed by the error propagation study of Mayer [17]. To that end, even if the GHI forecasts and unbiased, the PV power forecasts calculated by the model chain are likely to be biased to an extent depending on both the initial GHI forecasts, the design parameters of the PV plant, and the component models of the utilized model chain. This suggests that the "GHI forecasting + post-processing + irradiance-to-power conversion" workflow, which has been widely perceived as the best practice by solar forecasters for decades, is not able to provide unbiased forecasts without a final post-processing step applied to the PV power; however, to the best knowledge of the authors, no previous publications have been prepared yet to provide a detailed investigation on this question.

In view of the above introduction, the main aim of this work is to inquire about the optimal way of applying post-processing to NWP-based PV power forecasts. In practical PV power forecasting applications, the optimal sequence of the main forecasting steps should be chosen in a way that ensures the best possible accuracy, respecting the data availability concerns. Given that post-processing can be either applied to GHI or PV power, a total of four workflows result: (1) no post-processing for both GHI and PV power, (2) post-processing GHI but not PV power, (3) post-processing PV power but not GHI, and (4) post-processing both GHI and PV power. This constitutes the first dimension of the investigation—the position of post-processing within the PV power forecasting workflow. The second dimension is the post-processing methods, which are numerous in choice. Since the goal here is to post-process deterministic forecasts into another set of deterministic forecasts, two classic methods, namely, the model output statistics (MOS) and kernel conditional density estimation (KCDE), are considered, which represent the parametric and nonparametric regression techniques, respectively. Moving beyond post-processing techniques, the efficacy of any data-driven method necessarily depends upon the quality and amount of data. Therefore, the third dimension on which forecast verification is to be conducted is the length of training data. Last but not least, it is noted that the ground-based GHI measurement is rare as compared to other meteorological variables, such as temperature or humidity, hence, it is customary to consider the next-best option—satellite-retrieved irradiance—for post-processing; this is the fourth dimension of the verification exercise depicted in this work. The forecast verification procedure used in this work follows the latest recommendations in that, besides the regular error metrics, the different versions of post-processed forecasts are gauged through variance ratio and the statistics resulting from the Murphy–Winkler decompositions.

The remaining part of the paper is organized as follows. Section 2 introduces the datasets, including the PV power data from three plants in Hungary, the ground-based and satellite-derived GHI data at locations near those PV plants, and the operational NWP forecasts issued by HMS. Section 3 outlines six different post-processing workflows, which differ from one another in terms of the place(s) at which post-processing is conducted, and the source of the GHI "truth," which could be from either ground-based measurement or satellite-derived GHI database. Section 4 presents the results and discussions pertaining to various versions of post-processed forecasts. Conclusions follow at the end.

2. Data

The assessment and analysis presented in this paper are based on the deterministic power forecasts of three utility-scale PV plants in Hungary. The data related to the PV plants, the GHI observations, and the NWP forecasts are presented respectively in the following three subsections.

2.1. PV power measurements

PV power forecasting is to be performed for three utility-scale, gound-mounted PV plants in Hungary. The name, location, and main design information of these PV plants are summarized in Table 1. The three sites are located fairly close to each other (they fit into a circle with a diameter of 180 km), and thus exhibit similar climatic conditions. The general layout of the PV plants is also similar, i.e., the PV modules are mounted on parallel south-facing sheds, but their main design parameters, including the nominal power, inverter sizing factor, and tilt angle, are different. The detailed design parameters of the plants are also available, enabling the possible highest accuracy of the physical PV power modeling [19].

The power output of these PV plants is measured at a 15-min resolution and available for the four full calendar years from 2019 to 2022. The reliability and accuracy of the measurement are good in general, as it also serves as the basis of the financial settlements, however, there are several daytime zero production data due to the maintenance of the plants, which are excluded from the analysis. Following a common convention, a zenith angle filter of $\Theta_z < 85°$ is also applied to the dataset in order to retain only those reliable daytime data samples. The number of valid samples, also considering the availability of the GHI data, are listed for each year in Table 1.

2.2. GHI observations/retrievals

The ground-based GHI observations are obtained from the meteorological stations operated and maintained by the HMS. The stations are located close to, but not exactly at, the selected PV plants; the distances between the PV plants and their respective closest meteorological stations are <4.2 km (see Table 1), similar to the spatial resolution of the NWP model and the pixel size of the satellite images. Therefore, the spatial mismatch is not expected to have any considerable negative influence on the results. Ground-based GHI is measured by Kipp & Zonen CM11 and SMP11 pyranometers, which are calibrated every three years. Even though the overall data quality is good, a simple quality-control routine is performed using the extremely rare limit filter of the Baseline Surface Radiation Network (BSRN), to get rid of the spurious data. The GHI observation data is logged at a 10-min resolution, and it is converted into 15-min-resolution data using a clear-sky interpolation, in that, the linear interpolation is performed not directly on the GHI but on the clear-sky index instead. The clear-sky GHI is obtained from the McClear service [24].

Besides using ground-based GHI as the observation, satellite-retrieved irradiance is further considered. This is because the latest remote-sensing irradiance retrievals are available for all mid-latitutde locations, whereas the ground-based irradiance monitoring stations could be rare, so it would be interesting to inquire about the potential of using satellite-derived irradiance to replace ground-based observations during forecast calibration and verification. The satellite-retrieved GHI is acquired from the Copernicus Atmosphere Monitoring Service (CAMS) radiation service (CAMS-Rad). CAMS-Rad uses the Heliostat-4 to retrieve GHI from the images taken by the Meteosat Second Generation (MSG) satellites. The CAMS-Rad GHI retrievals, along with the McClear clear-sky GHI, were downloaded at a 15-min resolution using the web service of the SoDa-pro website.[1] The accuracy of the satellite-derived CAMS-Rad GHI data versus the ground-based measurements for all three locations is displayed in the top three rows of Table 2, using metrics presented later in Section 3.5. The satellite-derived GHI retrievals have a small positive bias at two out of three locations, otherwise, its accuracy is roughly similar in all three locations.

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

2.3. Numerical weather prediction

The forecasts of the weather parameters required for the PV power modeling, i.e., GHI, ambient temperature, and wind speed, are issued by the operational version of the Application of Research to Operations at Mesoscale (AROME) model implemented and run by the HMS. AROME is a regional, non-hydrostatic, mesoscale NWP model with a 15-min temporal and 2.5-km spatial resolution. Further details on the Hungarian implementation of the AROME model can be found in [25,26], while for detailed verification of the AROME GHI and the derived PV power forecasts, the reader is referred to [27].

The forecasts used in this study are issued by the 00Z (00 UTC) model run over the 0–24-h horizon, which is suitable for intra-day PV power forecasting. The utilized forecasts include GHI, 2-m ambient temperature, and 10-m wind speed. The 15-min resolution of the forecasts aligns with that of the PV power measurements and complies with the requirements of the grid, therefore, no resampling or interpolation is required.

The raw AROME GHI forecasts are verified against both the ground-measured and satellite-derived observations using the metrics presented in Section 3.5, and the results for all three PV plants are listed in Table 2. It can be seen that the raw forecasts have a negative bias in all locations, as evidenced by the normalized mean bias error (nMBE) column, and they are also underdispersed, as evidenced by their small variance ratios. The verification against the satellite-derived GHI exaggerates the goodness of the forecasts, since they seem to be more accurate in terms of normalized root mean square error (nRMSE), normalized mean absolute error (nMAE), and correlation coefficient in all PV plants. The only exception is the nMBE, since the CAMS-Rad and AROME forecast GHI are biased in a different direction; therefore, the apparent bias of the forecasts compared to the satellite data is higher. These results are consistent with the findings of Yang and Perez [28], who performed a similar analysis using a different dataset for several locations in the United States. The overall conclusion is that satellite-derived data are not sufficient to properly gauge the goodness of forecasts but may be suitable for the comparison of different sets of forecasts.

3. Post-processing, irradiance-to-power conversion, and verification methods

This part of the paper first describes the different post-processing workflows and their implementation details, which constitute the main inquiry of this paper. The description of the two applied regression methods, namely, MOS and KCDE, is provided in Sections 3.2 and 3.3, respectively. Section 3.4 introduces the physical model chains that serves the PV power simulations, whereas Section 3.5 details the verification process.

3.1. Implementation of post-processing

The post-processing used in this study aims to reduce the model-led bias and thus the overall errors of the deterministic NWP forecasts, and as such, it falls into the regression thinking tool of the deterministic-to-deterministic (D2D) category, following the typology proposed by Yang and van der Meer [29].

The GHI forecasts can either be post-processed or passed directly forward to the model chains, and similarly, the calculated PV power can also either be post-processed or used directly as the final forecasts. From the combination of these possibilities, four different workflows arise, which are identified by a two-digit code, where the first digit

Table 1
Metadata of the three utility-scale ground-mounted PV plants selected for the analysis, their distance from the closest meteorological stations, and the yearly number of valid samples.

Location	Coordinates	P_{DC} kWp	P_{AC} kW	Tilt angle	Met. st. dist., km	Number of valid samples			
						2019	2020	2021	2022
Paks	46.56° N, 18.82° E	20,680	17,244	35°	2.3	15,863	15,777	15,961	15,780
Vép	47.23° N, 16.68° E	3802	3600	20°	4.2	15,096	15,058	15,054	15,137
Kaposhomok	46.36° N, 17.93° E	590	498	30°	4.0	16,042	15,931	16,041	16,018

Table 2
Error metrics for the satellite-derived GHI verified against the ground-measured observations, and for the raw AROME GHI forecasts verified against both the ground-measured and satellite-derived reference data. The unit of the values in the last four columns is W/m^2.

Verified dataset	Reference dataset	PV plant	nMBE	nRMSE	nMAE	Corr. coeff.	Var. ratio	Sqrt. type-1 CB	Sqrt. res.	Sqrt. type 2 CB	Sqrt. disc.
Satellite	Ground	Paks	2.8%	22.4%	14.5%	0.955	98.4%	17.5	248.9	18.9	246.8
Satellite	Ground	Vép	0.0%	21.6%	13.8%	0.955	94.2%	11.0	249.4	21.2	242.1
Satellite	Ground	Kaposhomok	4.4%	21.8%	14.0%	0.961	95.6%	18.3	254.8	24.0	249.2
Forecast	Ground	Paks	-4.0%	34.3%	23.3%	0.890	85.9%	19.1	231.8	49.3	215.0
Forecast	Ground	Vép	-5.0%	38.5%	25.8%	0.853	82.2%	27.6	223.1	63.6	202.4
Forecast	Ground	Kaposhomok	-4.6%	35.2%	23.6%	0.891	80.9%	19.3	236.4	56.6	212.7
Forecast	Satellite	Paks	-6.7%	31.9%	21.9%	0.901	87.3%	28.0	232.9	48.9	217.6
Forecast	Satellite	Vép	-5.1%	36.4%	24.0%	0.863	87.2%	30.5	219.1	53.6	204.4
Forecast	Satellite	Kaposhomok	-8.5%	32.1%	21.9%	0.903	84.6%	33.5	234.3	54.4	215.4

corresponds to GHI and the second to the PV power post-processing, and each digit can be 1 or 0 depending on whether or not post-processing is involved in the given stage. More specifically, the four possible workflows are:

- 00: no post-processing,
- 10: post-processing of GHI but not the PV power,
- 01: post-processing of the PV power but not GHI,
- 11: post-processing of both GHI and the PV power.

The workflows that involve GHI post-processing are further divided into two cases, depending on the source of the GHI truth data used for fitting the post-processing models, which can be either ground-measured or satellite-derived, which is distinguished by letter “G” or “S” at the end of the code, respectively. A graphical explanation of these six workflows is also given in Fig. 1.

Two D2D post-processing methods are tested and compared in this paper in order to discover whether or not the optimal workflow is affected by the selection of the post-processing method. The first method is the MOS formulation of Lorenz et al. [30], which estimates the modelled bias with a bivariate fourth-degree polynomial regression (see Section 3.2), and it is widely accepted as a baseline method for deterministic post-processing [29]. The second method is the KCDE, first introduced in the context of solar forecasting by Yang [31]. KCDE is conceptionally different from Lorenz’s MOS, since it is a nonparametric regression, which provides greater flexibility in mapping the relationship between the bias and the predictor variables (see Section 3.3). The predictors of Lorenz’s MOS are the cosine of the zenith angle and the forecast clear-sky index, therefore, for better comparability, the predictors of the KCDE are also limited to these two variables.

The four-year dataset is divided into a training set and a testing set. Data from 2022 is used for testing of all cases, to ensure the comparability and reliability of the results. As for the training data, three scenarios are tested, each with a different training data length:

- 1 year: data from 2021 are used for training,
- 2 years: data from 2020 and 2021 are used for training,
- 3 years: data from 2019, 2020, and 2021 are used for training.

The general procedure is that the post-processing models are fitted using the data from the training years, and then the fitted models are used to correct the forecasts from the testing year. However, if both the GHI and PV power are post-processed (workflows 11S and 11G), it is important to first correct the GHI forecasts for the training years before converting them to power forecasts, and then use the power forecasts based upon the corrected GHI forecasts to fit the PV power post-processing model, as to avoid the double-correction of the same error patterns.

The simplest usage of the post-processing models is to fit a single model for the whole training period for each location, which is referred to as yearly fit in what follows, as it treats the whole yearly data together. Alternatively, quarterly and monthly fits are also possible, in which separate model fits are used for each quarter or each month of the year, respectively. For instance, in a monthly fit with three years of training data, the data from all three Januaries are used to fit the post-processing model, which is then used to correct the forecast from the January of the testing year, and the same process is repeated for all months.

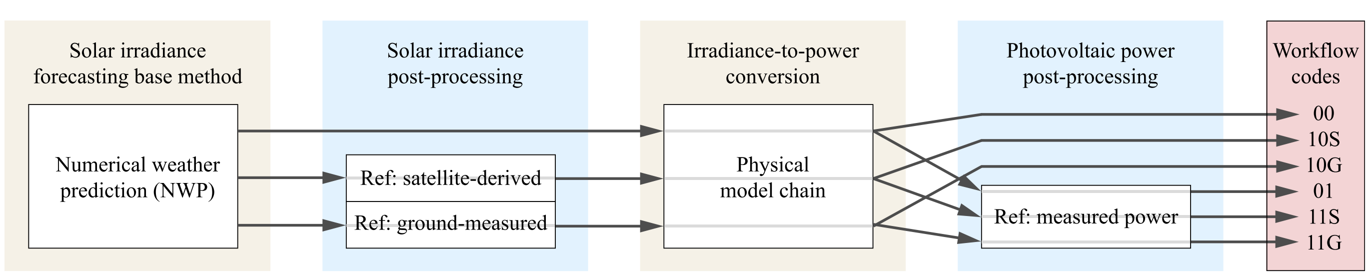

Fig. 1. Schematic of the six compared post-processing workflows.

3.2. Lorenz's MOS

Lorenz's MOS approximates the bias of NWP forecasts as a fourth-degree polynomial of two predictors, namely the cosine of the zenith angle (Θ_z) and the forecast clear-sky index (k_{cs}) [30]:

$$\text{bias} = a_0 + a_1 cos^4\Theta_z + a_2 k_{cs}^4 + a_3 cos^3\Theta_z + \ldots + a_8 k_{cs} \quad (1)$$

where $a_0, \ldots, a_8$ are the regression coefficients to be fitted, e.g., via least squares. The clear-sky index is the ratio of the forecast GHI to the clear-sky GHI, which is estimated by a clear-sky model. The McClear model is used in this study, however, the choice of the clear-sky model is not expected to have a significant impact on the accuracy in solar forecasting applications [32].

The bias is the difference between the forecast and observed values, either the GHI or the PV power, depending on which stage of the power forecasting workflow the post-processing is applied. The biases obtained from the training data are used for fitting the regression coefficients, and then the corrected forecasts are created by adding onto the raw forecasts those biases estimated for the testing period.

3.3. Kernel conditional density estimation

KCDE estimates directly from the historical data the conditional density of the predictand for a given set of values of the predictor variables, and the expectation of the estimated density can be used for the deterministic correction. The whole process of the KCDE is thoroughly explained in [31], therefore, only the most important details related to its implementation are reiterated here.

The two predictor variables, similarly to Lorenz's MOS, are the cosine of the zenith angle and the forecast clear-sky index, both used with a Gaussian kernel. The bandwidth values are selected using a grid-search method, and the selected values are 0.05 for yearly, 0.08 for quarterly, and 0.12 for monthly fits for both predictors. These values are found to work well with all three training data lengths. The predictand, similarly to Lorenz's MOS, is the bias, either of the GHI or of the PV power forecasts, depending on which stage the post-processing is applied to.

As a nonparametric method, KCDE is not fitted for the training data, but training data are directly used for computing the bias correction. However, this makes no significant difference on the conceptual level, therefore, the training data terminology is also used for the KCDE for the sake of consistency, even though no actual fitting or training is performed in this case.

3.4. Physical PV model chains

The power output of the PV plants can be calculated from the weather parameters using a detailed physical model chain. The term "model chain" implies that the calculation process relies on applying a set of component models successively, where the inputs of each component model are the outputs of the preceding one [14]. The detailed model chain used in this study consists of ten modeling steps, which are visualized in Fig. 2. The main steps of the implemented physical model chain are described in the following paragraphs, while interested readers can find more detailed descriptions of this topic in [17,33,34].

First, the position of the sun respective to an observer on earth can be fully specified by the zenith and azimuth angles, which are calculated from the timestamp and the location of the plant using the solar position algorithm (SPA) of the National Renewable Energy Laboratory (NREL) [35]. GHI is decomposed into beam and diffuse components using the temporal-resolution cascade Yang model [36], which has been claimed and thoroughly verified to be the most accurate separation model to date [37]. The horizontal irradiance components are converted to those irradiance components on the tilted module plane by the Perez

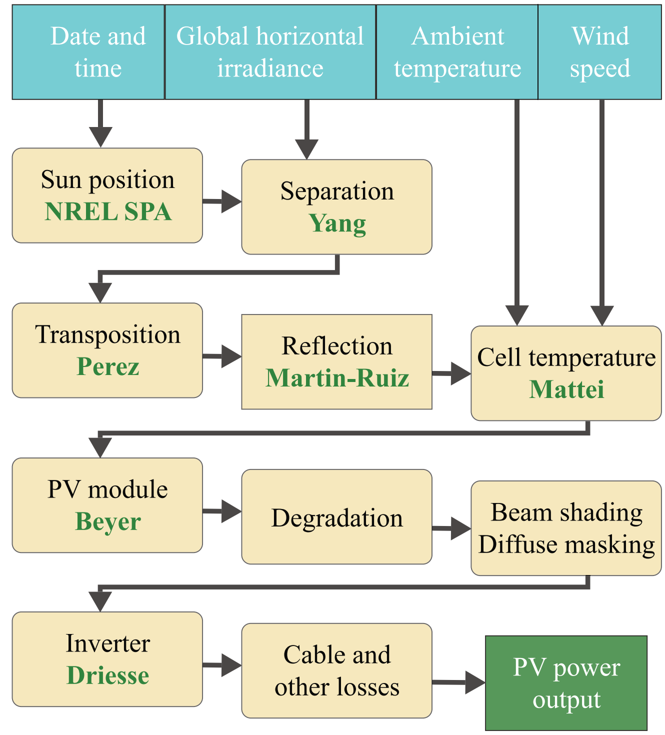

Fig. 2. Schematic of the physical PV model chain implemented in this study for irradiance-to-power conversion.

transposition model [38], which is widely acknowledged as the most accurate transposition model since it was developed more than three decades ago [39]. The reflections from the module surface are accounted for by the Martin–Ruiz model, which offers separate angular loss factors for the beam, sky-diffuse, and ground-reflected irradiance components [40]. At this stage, the so-called effective irradiance is obtained.

The cell temperature of the PV modules is calculated by the Mattei model, covering both the radiative heating by irradiance and the convective heat removal by wind [41]. The efficiency and maximum DC power output of the PV modules are estimated by the temperature and irradiance-dependent Beyer model [42] with the fixed parameters proposed in [17]. The degradation of the PV modules is considered by a 2% initial light-induced degradation loss factor and a further 0.5% per annum degradation rate [33]. The losses due to the shading of the beam and masking of the diffuse irradiance components by the adjacent module rows are calculated per the method described in [33,43]. The DC to AC conversion efficiency of the inverter is calculated as a function of the DC power and voltage with the formula proposed by Driesse et al. [44], and the clipping losses are included by maximizing the AC output power at the nominal power of the inverters. Finally, the electric losses on the DC and AC cables, transformers, and components are also estimated on a level corresponding to the relevant information available in the design documentation of the PV plants.

3.5. Verification of deterministic forecasts

An insightful forecast verification exercise entails using a good collection of different evaluation metrics and graphical tools to provide an as complete as possible picture of the forecast quality [45]. However, consistency is another important prerequisite of meaningful verifications; it means that the forecasts should be evaluated with such metrics that are consistent with the directives that they are optimized for [46]. Lorenz's MOS is fitted by the least squares method, while for the KCDE, the bias is determined from the expected value (i.e., mean) of the estimated conditional density, therefore, both methods post-process the forecasts under the directive of minimizing the mean square error

(MSE). Consequently, the the mean bias error (MBE) and the root mean square error (RMSE), which are both consistent with the MSE, are the main metrics used for the evaluation of the different cases. These metrics are defined as

$$\text{MBE} = \frac{1}{N}\sum_{i=1}^{N}(f_i - x_i), \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - x_i)^2}, \tag{3}$$

where f and x stand for the forecast and observation, and N is the number of samples entering the verification.

Even though these two metrics are sufficient to identify the best forecasts, they do not provide enough information to understand why one set of forecasts is better than another. To that end, further error metrics are also calculated, which together can provide a better explanation of the differences between the examined cases. The mean absolute error (MAE) is another commonly used metric that measures the overall errors of the forecasts; it is calculated as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|f_i - x_i|. \tag{4}$$

The correlation coefficient reflects the association between forecasts and observations, which can be viewed as an indicator of the potential skillfulness of the forecasts that can be achieved if their bias and variance are perfectly calibrated [47]. It is calculated as

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

where $\text{cov}(\cdot,\cdot)$ stands for the covariance and σ for the standard deviation. The dispersion of the forecasts can be best expressed by an F variance ratio, proposed by Mayer and Gróf [14] as

$$F = \frac{\mathbb{V}(f)}{\mathbb{V}(x)}, \tag{6}$$

where $\mathbb{V}(\cdot)$ is the variance operator. A variance ratio lower and higher than 1 indicates under- and over-dispersed deterministic forecasts, respectively.

The distribution-oriented verification framework proposed by Murphy and Winkler [48] is an effective tool for detailed assessments of the calibration of forecasts [45]. Relying on the marginal and conditional distributions of forecasts and observations, the MSE can be decomposed in two different ways, which are referred to as the calibration–refinement factorization,

$$\text{MSE}(f,x) = \mathbb{V}(x) + \mathbb{E}_f[f - \mathbb{E}(x|f)]^2 - \mathbb{E}_f[\mathbb{E}(x|f) - \mathbb{E}(x)]^2, \tag{7}$$

and the likelihood–base rate factorization,

$$\text{MSE}(f,x) = \mathbb{V}(f) + \mathbb{E}_x[x - \mathbb{E}(f|x)]^2 - \mathbb{E}_x[\mathbb{E}(f|x) - \mathbb{E}(f)]^2, \tag{8}$$

where $\mathbb{E}(\cdot)$ is the expectation operator. There are four terms in these two equations that can be used to describe the different aspects of forecast quality:

- Type-1 conditional bias: $\mathbb{E}_f[f - \mathbb{E}(x|f)]^2$,
- Resolution: $\mathbb{E}_f[\mathbb{E}(x|f) - \mathbb{E}(x)]^2$,
- Type-2 conditional bias: $\mathbb{E}_x[x - \mathbb{E}(f|x)]^2$,
- Discrimination: $\mathbb{E}_x[\mathbb{E}(f|x) - \mathbb{E}(f)]^2$.

The conditional bias terms are negatively oriented (i.e., lower values are better), whereas the resolution and discrimination terms are positively oriented. For a more detailed description of the calculation and meaning of these terms, readers are referred to [28,45,48].

The three PV plants have different installed capacities, therefore, the calculated metrics have to be normalized to ensure that each plant contributes to the presented results with equal weights. The MBE, RMSE, and MAE are divided by the mean of the observations and denoted by nMBE, nRMSE, and nMAE after this normalization. The four terms of the Mourphy–Winkler factorizations are calculated by rescaling the PV power output to a 1-kWp installed capacity.

Aside from the numeric metrics, the joint distribution of forecast and observation is also visualized using scatterplots colored by the density of points. These plots contain all time-independent information related to the quality of the forecasts [45].

4. Result and discussion

The results of the performed analyses are summarized and discussed in this section. First, the verification of the raw and post-processed GHI forecasts is presented in Section 4.1, followed by the verification of the PV power forecasts created by all six post-processing workflows in Section 0. The results for the three PV plants show highly similar trends, therefore, they are averaged, and only the mean values are presented to keep the dimensionality of the presented results under control. The average of the metrics for the three PV plants is a better representation of the goodness that can be expected in any new unseen location. Section 4.3 provides further discussion and recommendations for the most effective means of applying post-processing under different practical circumstances.

4.1. Correction of GHI forecasts

The metrics calculated for the raw and corrected GHI forecasts are summarized in Table 3. The raw forecasts have a slight negative bias of -4.5% on average and are underdispersed (with a variance ratio of 83.0%), which seems to be a general tendency of the AROME forecasts in Hungary [27]. This negative bias is turned into a smaller positive one by all post-processing methods and cases. The magnitude of the remaining bias after post-processing depends mostly on the source of the GHI truth data. For ground-measured GHI, the bias is $<$1% in most cases, whereas for satellite-derived irradiance, the nMBE is around 3% due to the positive bias in the satellite-derived GHI (see Table 2). The post-processing also improves the nRMSE, nMAE, and correlation coefficient values. It is known that MSE and MAE can be conflicting error metrics in the sense that they are optimized by forecasts with different statistical characteristics [49]. In the presented cases, however, the decrease of RMSE is typically accompanied by the decrease of MAE as well, which suggests that the RMSE reduction is not only the result of calibrating the forecasts for MSE [47], but it indicates the improvement of the overall accuracy of the forecasts.

The variance ratio of the post-processed forecasts is similar to that of the raw forecasts, i.e., the corrected forecasts are still underdispersed. However, this is not a deficiency of the methods but a general property of MSE-optimized post-processing [47]. It can be proved theoretically through the bias–variance decomposition of the MSE that the optimal variance ratio is $F = \rho^2$ for MSE-optimized forecasts [47]. The optimal variance ratio for a correlation coefficient of 0.89 is 79.2%, which is well captured by the post-processed forecasts. The variance ratio is lower for the forecasts fitted to the satellite-derived GHI as these data are already underdispersed as compared to the ground-measured GHI.

The type-1 conditional bias and the resolution terms of the Murphy–Winkler factorization are improved by all post-processing, but the improvement is higher for the ground-measured truth data. In contrast, the type-2 conditional bias and the discrimination terms only show a consistent improvement for the ground-measured GHI. If satellite-derived data are used as the reference, they either improve or deteriorate depending on the forecasting method, training time, and periods used. Overall, post-processing using the satellite-derived GHI as the

Table 3
Summary of error metrics for the GHI forecasts. The unit of the values in the last four columns is W/m^2. Bold letters indicate the best results for the given method and training data length.

Method	Training data length	Training period	Truth data	nMBE	nRMSE	nMAE	Corr. coeff.	Var. ratio	Sqrt. type-1 CB	Sqrt. res.	Sqrt. type 2 CB	Sqrt. disc.
Raw forecasts				-4.5%	36.0%	24.2%	0.878	83.0%	22.4	230.5	56.8	210.1
MOS	1 year	Y	Gr.	1.1%	**34.1%**	23.3%	0.890	83.3%	13.5	233.5	51.3	213.4
			Sat.	2.7%	34.2%	23.8%	0.889	79.0%	16.1	233.4	57.5	207.6
		Q	Gr.	1.0%	34.1%	23.5%	0.890	83.0%	12.1	233.4	51.7	212.9
			Sat.	2.5%	34.4%	23.8%	0.888	79.4%	14.2	233.0	57.2	207.9
		M	Gr.	1.3%	34.7%	23.9%	0.886	83.4%	13.8	232.4	52.3	212.5
			Sat.	2.9%	34.9%	24.2%	0.884	79.8%	15.8	232.1	57.7	207.6
	2 years	Y	Gr.	0.4%	**34.0%**	23.4%	0.890	84.7%	14.8	233.7	49.4	215.1
			Sat.	2.7%	34.2%	23.4%	0.889	81.5%	16.7	233.5	54.2	210.8
		Q	Gr.	0.6%	34.0%	23.5%	0.890	83.1%	11.8	233.5	51.4	213.1
			Sat.	2.8%	34.2%	23.4%	0.889	80.6%	14.6	233.4	55.4	209.6
		M	Gr.	0.3%	34.0%	23.5%	0.890	83.1%	11.8	233.5	51.3	213.2
			Sat.	2.6%	34.2%	23.5%	0.889	80.6%	13.9	233.2	55.4	209.5
	3 years	Y	Gr.	0.8%	34.0%	23.4%	0.890	84.5%	14.7	233.7	49.6	214.9
			Sat.	3.0%	34.2%	23.5%	0.889	81.2%	17.1	233.5	54.7	210.5
		Q	Gr.	1.0%	33.9%	23.4%	0.891	82.8%	12.3	233.7	51.4	212.9
			Sat.	3.2%	34.1%	23.4%	0.890	80.2%	15.7	233.5	55.8	209.4
		M	Gr.	0.8%	**33.9%**	23.4%	0.891	82.9%	11.7	233.8	51.3	213.1
			Sat.	3.0%	34.1%	23.4%	0.890	80.2%	14.8	233.6	55.7	209.3
KCDE	1 year	Y	Gr.	1.1%	**33.7%**	22.9%	0.892	80.8%	11.2	234.0	54.2	210.7
			Sat.	2.9%	33.9%	23.3%	0.892	76.9%	15.5	233.9	60.1	205.4
		Q	Gr.	0.9%	33.7%	23.0%	0.892	80.6%	10.0	233.9	54.4	210.3
			Sat.	2.5%	34.0%	23.4%	0.890	77.2%	13.4	233.6	59.6	205.7
		M	Gr.	1.1%	34.0%	23.3%	0.890	80.6%	10.6	233.5	54.7	209.9
			Sat.	2.6%	34.2%	23.7%	0.889	76.9%	14.2	233.2	60.4	204.8
	2 years	Y	Gr.	0.5%	**33.6%**	22.9%	0.893	81.0%	9.8	234.2	53.5	211.1
			Sat.	3.0%	33.8%	23.0%	0.892	78.8%	14.2	234.0	57.4	208.0
		Q	Gr.	0.5%	33.7%	23.1%	0.892	80.5%	9.1	234.0	54.3	210.2
			Sat.	2.8%	33.9%	23.1%	0.892	78.5%	13.0	233.8	57.7	207.6
		M	Gr.	0.3%	33.7%	23.2%	0.892	80.2%	9.5	233.9	54.6	209.8
			Sat.	2.6%	34.0%	23.3%	0.891	78.0%	12.9	233.6	58.5	206.6
	3 years	Y	Gr.	0.8%	33.6%	22.9%	0.893	80.7%	10.5	234.2	53.8	210.8
			Sat.	3.1%	33.8%	23.0%	0.892	78.3%	14.9	234.1	58.0	207.4
		Q	Gr.	0.9%	33.6%	23.0%	0.893	80.3%	10.2	234.2	54.3	210.1
			Sat.	3.1%	33.8%	23.1%	0.892	78.1%	14.6	234.1	58.1	207.2
		M	Gr.	0.7%	**33.5%**	23.1%	0.893	79.9%	9.8	234.2	54.7	209.7
			Sat.	2.9%	33.8%	23.2%	0.892	77.5%	14.5	234.1	58.8	206.3

reference brings almost the same improvement in the RMSE as if ground-measured GHI is used; however, using ground-based measurements results in better overall forecast quality, indicated by the lower conditional biases and the higher resolution and discrimination terms. The KCDE achieves a higher reduction in type-1 conditional bias but only at the cost of a lower reduction in the type-2 conditional bias, compared to the MOS.

The different post-processing methods and cases are compared by the RMSE since this metric is consistent with the directive that the forecasts are optimized for. The KCDE, in accordance with a previous paper [31], outperforms Lorenz's MOS in all cases, probably due to its higher flexibility resulting from its nonparametric nature. The longer training dataset improves the accuracy but only to a very small extent. However, the optimal training period clearly depends on the length of the full dataset. For one year of training data, fitting the same model for the whole year has the lowest RMSE, closely followed by the quarterly fit, and the monthly fit results in a substantially lower improvement. For two years of training data, there is almost no difference between the three options, but the yearly fit still maintains a small advantage. However, with three years of training data, the monthly fit yields the lowest RMSE, followed by the quarterly fit. These trends are the same for both MOS and KCDE, which suggests that it is fairly independent of the choice of post-processing method. Suggested by the results is that the irradiance data exhibits some seasonality, which could benefit from monthly fitting of post-processing models, however, the benefit is only obvious if the monthly fitting is supported with sufficient samples over multiple years.

Overall, the best approach for GHI post-processing among the examined options is monthly KCDE relying on three years of historical data, which reduces the average nRMSE to 33.5% from the 36.0% of the raw forecasts, which corresponds to a 6.9% relative RMSE reduction.

4.2. Correction of PV power forecasts

The verification results of the PV power forecasts are summarized in Table 4 and Table 5 for MOS and KCDE, respectively. The main conclusions of the GHI verification in Section 4.1 also apply here, e.g., KCDE is more effective than MOS, longer training time results in slightly lower RMSE, or shorter training periods are optimal for longer data lengths.

The forecasts of the purely physical approach (workflow 00 with no post-processing) are almost unbiased, which means that the original negative bias is compensated by the model chain of concern. This means, however, that if the original GHI forecasts were almost unbiased due to an effective GHI post-processing (workflow 10G), then the particular model chain used here would introduce an unwanted bias into the final PV forecasts. In workflow 10S, when the post-processed GHI forecasts have already inherited a positive bias from the satellite-derived observations, the final PV forecasts have an even higher nMBE, exceeding 7% in some cases. This model chain also increases the variance ratio of the PV forecasts compared to the input GHI forecasts.

The bias introduced by the model chains, however, can be effectively reduced by applying post-processing on the PV output power (workflow 01 and 11). This post-processing also pulls the variance ratio from its higher value increased by the model chain down to a level close to the optimum for MSE-optimized forecasts. The MBE, RMSE, and type-1 conditional bias of the PV forecasts created by following workflows 01

Table 4
Summary of error metrics for the PV power output forecasts with the MOS post-processing method. The unit of the values in the last four columns is W/kWp. Bold letters indicate the best results for the given training data length.

Training data length	Training period	Workflow code	nMBE	nRMSE	nMAE	Corr. coeff.	Var. ratio	Sqrt. type-1 CB	Sqrt. res.	Sqrt. type 2 CB	Sqrt. disc.
Physical		00	-0.7%	43.4%	27.4%	0.835	94.3%	39.2	216.3	52.9	210.1
		01	0.7%	40.7%	28.2%	0.848	81.3%	19.5	219.8	63.6	198.9
		10G	4.6%	41.7%	27.0%	0.849	92.5%	40.0	220.4	54.5	211.9
	Y	10S	7.0%	41.9%	27.4%	0.847	88.4%	39.4	219.8	60.7	206.7
		11G	1.0%	**40.5%**	28.1%	0.849	79.2%	19.0	220.0	65.8	196.3
		11S	0.9%	40.8%	28.4%	0.847	80.1%	20.9	219.7	64.9	197.0
		01	0.6%	40.7%	28.1%	0.848	79.4%	17.7	219.7	66.1	196.5
		10G	4.8%	42.1%	27.1%	0.847	93.4%	41.7	219.7	54.6	212.4
1 year	Q	10S	6.9%	42.4%	27.4%	0.844	90.1%	41.1	219.1	60.0	207.9
		11G	0.7%	40.6%	28.1%	0.848	77.3%	16.7	219.7	68.8	193.8
		11S	0.6%	40.7%	28.3%	0.847	77.3%	17.3	219.6	68.8	193.7
		01	0.9%	41.2%	28.5%	0.844	79.1%	18.6	218.6	67.6	195.2
		10G	4.9%	42.3%	27.3%	0.845	93.2%	41.2	219.2	55.6	211.8
	M	10S	7.1%	42.7%	27.7%	0.842	90.0%	40.8	218.4	61.0	207.5
		11G	1.0%	40.9%	28.2%	0.846	77.0%	16.5	219.1	69.9	193.0
		11S	1.1%	41.1%	28.3%	0.845	76.8%	17.9	218.9	70.3	192.5
		01	1.2%	40.7%	27.9%	0.850	84.5%	23.9	220.2	59.9	202.8
		10G	3.4%	41.6%	27.0%	0.850	93.6%	40.3	220.7	51.9	213.2
	Y	10S	6.8%	42.0%	27.1%	0.848	91.2%	42.1	220.1	57.6	209.9
		11G	1.1%	40.4%	27.8%	0.851	81.2%	21.5	220.5	63.3	199.0
		11S	1.2%	40.7%	28.0%	0.849	83.0%	24.1	220.2	61.5	200.8
		01	1.2%	40.6%	27.8%	0.849	81.0%	19.7	220.0	64.4	198.6
		10G	3.9%	41.9%	26.9%	0.848	93.4%	39.7	219.9	53.3	212.6
2 years	Q	10S	6.8%	42.2%	27.0%	0.847	91.3%	41.4	219.6	58.2	209.9
		11G	1.3%	40.5%	27.8%	0.849	78.5%	17.6	220.0	67.4	195.7
		11S	1.2%	40.6%	27.9%	0.849	78.8%	17.9	219.9	67.1	195.9
		01	1.0%	40.6%	27.7%	0.850	80.6%	19.3	220.1	64.9	198.3
		10G	3.8%	41.8%	26.8%	0.848	93.3%	38.8	219.8	53.7	212.6
	M	10S	6.7%	42.2%	27.0%	0.846	91.3%	40.6	219.3	58.6	209.9
		11G	1.0%	**40.4%**	27.6%	0.850	78.1%	16.4	220.2	67.7	195.4
		11S	0.9%	40.4%	27.7%	0.850	78.2%	16.6	220.1	67.7	195.4
		01	1.9%	40.7%	28.0%	0.850	85.1%	25.6	220.4	59.5	203.7
		10G	3.9%	41.5%	27.1%	0.850	93.4%	40.2	220.8	52.4	213.1
	Y	10S	7.3%	42.0%	27.2%	0.848	90.8%	42.3	220.1	58.6	209.6
		11G	1.9%	40.4%	27.8%	0.851	81.8%	22.8	220.6	63.0	199.8
		11S	1.9%	40.7%	28.1%	0.849	83.6%	25.2	220.2	61.1	201.6
		01	2.1%	40.5%	27.7%	0.851	81.8%	21.4	220.3	63.6	199.9
		10G	4.6%	41.7%	26.8%	0.849	93.3%	39.7	220.2	53.6	212.9
3 years	Q	10S	7.5%	42.1%	27.0%	0.848	91.0%	42.0	219.9	59.1	209.9
		11G	2.2%	40.4%	27.7%	0.851	79.5%	19.2	220.4	66.5	197.2
		11S	2.1%	40.5%	27.8%	0.850	79.8%	19.6	220.2	66.1	197.4
		01	2.0%	40.4%	27.6%	0.852	81.3%	20.8	220.5	64.1	199.5
		10G	4.4%	41.6%	26.6%	0.850	93.3%	38.9	220.4	53.6	213.2
	M	10S	7.3%	42.0%	26.8%	0.848	91.1%	41.1	219.9	59.0	210.2
		11G	2.0%	**40.2%**	27.4%	0.853	78.9%	17.8	220.8	66.8	196.9
		11S	1.9%	40.2%	27.5%	0.852	78.9%	17.7	220.6	66.8	196.8

and 11 are almost identical, which means that if the PV power is post-processed, the prior post-processing of the GHI has no significant added value. In this respect, however, the two methods behave differently: with MOS, the lowest RMSE is achieved by workflow 11G, whereas with KCDE, workflow 01 is the overall best. The possible reason for this is that the less versatile structure of MOS can benefit from double post-processing, but single post-processing is enough for the KCDE due to its higher flexibility. Considering that the state-of-the-art post-processing methods are more complex and flexible than MOS, and accepting the simplicity as an advantage, there is no need to post-process the GHI forecasts but only the PV power forecast.

The type-2 conditional bias and the discrimination of the post-processed forecasts are worse compared to the basic workflow 00, which does not involve post-processing, in the majority of the cases. This can be explained based on the discussion presented by Mayer and Yang [47], revealing that the dispersion (measured by the variance ratio) of the forecasts has a significant effect on the conditional bias and discrimination terms. Post-processing fitted for minimizing the MSE inevitably results in underdispersed forecasts, and if the original forecasts are less underdispersed (higher variance ratio), post-processing may deteriorate the type-2 conditional bias and discrimination terms, which is precisely what can be seen here. However, since the directive in this case is to create MSE-optimized forecasts, reduction in metrics that are not consistent with the MSE must not be seen as a shortcoming of the selected method [46].

The overall best method is the quarterly KCDE with three years of historical data and only PV power post-processing, resulting in an nRMSE of 39.7%, which is 8.5% lower than the 43.4% of the physical forecasts. The relative improvement in the RMSE of PV power forecasts is even higher compared to the GHI forecasts.

The joint distribution of different forecasts and observations are shown in Fig. 3, in the form of scatter plots, for the Kaposhomok PV plant for the best post-processing method, which is the quarterly grouped KCDE trained with three years of historical data. Fig. 3 (a) and (b) show the raw and the post-processed GHI forecasts, respectively, (c) and (d) show the PV power forecasts created by converting the forecasts of (a) and (b) to power using the model chain, whereas (e) and (f) are the post-processed version of the forecasts on (c) and (d), respectively. The most apparent deficiency of the raw GHI forecasting is the inability to issue high values, which explains both its negative bias and low variance (see Table 2). The post-processing of GHI partly corrects this negative bias and shifts the region with the highest density of samples closer to

Table 5

Summary of error metrics for the PV power output forecasts with the KCDE post-processing method. The unit of the values in the last four columns is W/kWp. Bold letters indicate the best results for the given training data length.

Training data length	Training period	Workflow code	nMBE	nRMSE	nMAE	Corr. coeff.	Var. ratio	Sqrt. type-1 CB	Sqrt. res.	Sqrt. type 2 CB	Sqrt. disc.
Physical		00	-0.7%	43.4%	27.4%	0.835	94.3%	39.2	216.3	52.9	210.1
1 year	Y	01	0.7%	40.1%	27.3%	0.852	77.7%	13.6	220.6	67.8	195.5
		10G	4.9%	41.4%	26.4%	0.850	90.5%	36.9	220.4	57.3	210.0
		10S	7.2%	41.6%	26.9%	0.849	86.3%	36.4	220.0	63.8	204.9
		11G	0.7%	40.2%	27.3%	0.852	78.1%	14.8	220.5	67.7	195.8
		11S	0.7%	40.3%	27.4%	0.851	78.8%	15.3	220.3	66.9	196.5
	Q	**01**	0.5%	**40.0%**	27.3%	0.853	75.6%	12.2	220.7	70.3	192.9
		10G	4.7%	41.6%	26.5%	0.849	90.5%	37.3	220.1	57.5	209.5
		10S	6.9%	41.9%	27.0%	0.847	87.0%	37.1	219.5	63.1	205.1
		11G	0.6%	40.1%	27.3%	0.852	76.3%	13.3	220.6	69.7	193.8
		11S	0.6%	40.1%	27.4%	0.852	76.3%	13.1	220.6	69.7	193.7
	M	01	0.7%	40.3%	27.5%	0.851	75.8%	13.9	220.3	70.5	192.6
		10G	4.8%	41.8%	26.7%	0.847	90.3%	37.4	219.7	58.1	209.1
		10S	7.1%	42.0%	27.1%	0.845	86.8%	37.1	219.1	63.8	204.5
		11G	0.8%	40.3%	27.4%	0.850	77.0%	14.7	220.2	69.2	194.2
		11S	0.8%	40.3%	27.5%	0.850	76.9%	14.8	220.2	69.4	194.0
2 years	Y	01	1.1%	40.1%	27.0%	0.853	79.6%	16.2	220.8	65.7	198.0
		10G	4.0%	41.2%	26.2%	0.852	91.1%	35.2	220.7	55.2	211.0
		10S	7.2%	41.7%	26.6%	0.849	89.0%	38.9	220.1	60.9	208.0
		11G	1.1%	40.0%	26.9%	0.853	79.5%	16.5	220.9	65.8	197.9
		11S	1.1%	40.3%	27.2%	0.852	80.4%	17.7	220.5	65.1	198.6
	Q	**01**	0.9%	**39.9%**	27.0%	0.854	76.9%	13.4	221.1	68.7	194.9
		10G	3.8%	41.4%	26.3%	0.850	90.6%	34.9	220.2	56.0	210.0
		10S	6.8%	41.7%	26.6%	0.848	88.8%	37.8	219.9	60.8	207.5
		11G	1.0%	39.9%	27.0%	0.854	77.4%	14.2	221.0	68.3	195.4
		11S	1.0%	40.0%	27.1%	0.853	77.4%	13.9	220.9	68.4	195.4
	M	01	0.9%	40.0%	27.1%	0.853	76.7%	14.2	221.0	69.1	194.4
		10G	3.7%	41.4%	26.4%	0.849	90.2%	34.4	220.1	56.5	209.4
		10S	6.7%	41.8%	26.7%	0.848	88.3%	37.3	219.7	61.4	206.9
		11G	1.0%	40.0%	27.0%	0.853	77.4%	14.7	220.8	68.5	195.2
		11S	0.9%	40.0%	27.1%	0.853	77.3%	14.6	220.8	68.6	195.1
3 years	Y	01	1.7%	40.1%	27.0%	0.854	80.2%	17.8	220.9	65.3	198.7
		10G	4.4%	41.1%	26.2%	0.852	90.9%	35.1	220.8	55.6	211.0
		10S	7.6%	41.7%	26.6%	0.850	88.6%	38.9	220.2	61.7	207.7
		11G	1.7%	40.0%	26.9%	0.854	80.1%	17.8	221.0	65.4	198.7
		11S	1.7%	40.3%	27.2%	0.852	81.1%	19.0	220.6	64.5	199.6
	Q	**01**	1.8%	**39.7%**	26.9%	0.856	77.5%	15.3	221.4	68.2	195.9
		10G	4.4%	41.1%	26.2%	0.851	90.4%	34.6	220.6	56.2	210.2
		10S	7.3%	41.6%	26.6%	0.850	88.4%	38.0	220.3	61.5	207.5
		11G	1.9%	39.8%	26.9%	0.855	77.9%	15.6	221.4	67.8	196.4
		11S	1.8%	39.8%	27.0%	0.855	78.0%	15.7	221.3	67.8	196.5
	M	01	1.7%	39.7%	27.0%	0.856	77.1%	15.5	221.4	68.6	195.4
		10G	4.2%	41.1%	26.2%	0.852	90.0%	33.9	220.6	56.5	209.8
		10S	7.2%	41.6%	26.6%	0.850	88.1%	37.5	220.2	61.7	207.1
		11G	1.9%	39.8%	26.8%	0.855	77.8%	16.0	221.4	67.9	196.3
		11S	1.8%	39.8%	27.0%	0.855	77.8%	15.9	221.3	68.0	196.2

the diagonal, but even the corrected GHI fails to account for the extremely high values caused by cloud-enhancement events—although such events do not persist as long as 15 min, their occurrence over a shorter duration would still have an effect on the 15-min aggregated data [18]. The errors in the high-irradiance region, however, do not propagate into the PV power forecasts due to the clipping of the high power output values by the inverters [17]. The spread of the scatterplot, expressed as the number of points far from the diagonal, is higher for the PV power forecasts than for the GHI forecasts. These high-error forecasts are created when the state of the sky (i.e., clear or overcast) is misidentified by the NWP model, and the effect of such errors is amplified by the tilt angle of the PV modules [17]. Post-processing helps to reduce the error of these forecasts, but it does not shift the highest density part of points to the diagonal line but results in a positive bias in the low and a negative bias in the high PV power regime. Statistically speaking, such conditional bias is the realization of an underdispersion of the forecasts, which is an inevitable result of the MSE optimization used during the fitting of the post-processing models [47]. Finally, there is hardly any difference between Fig. 3 (e) and (f), which further demonstrates that the prior post-processing of GHI is unimportant as long as the PV power forecasts are post-processed in the final step.

4.3. Discussion and recommendations

The overall conclusion of the presented results is that post-processing is best performed on the PV power forecasts, while correcting the GHI before feeding it into the model chain is seemingly unimportant. On the one hand, it calls into question the common interpretation of the three-step PV forecasting framework where post-processing precedes the irradiance-to-power conversion. On the other hand, the majority of the post-processing papers published in the context of solar forecasting apply the correction on the GHI, presumably because of the better worldwide availability of historical GHI measurements and forecasts. However, the methods developed for the post-processing of GHI can be effectively used, in most cases even without modifications, for post-processing the PV power. In this study, the correction of the PV power is performed exactly in the same way as the correction of the GHI by only changing the predictand from the bias of GHI to that of the PV power. Moreover, the post-processing method and settings found to be the best for correcting the GHI are also close to the best for correcting the PV power output. To that end, developing and testing post-processing techniques on GHI remains an appropriate and useful avenue for research and practice.

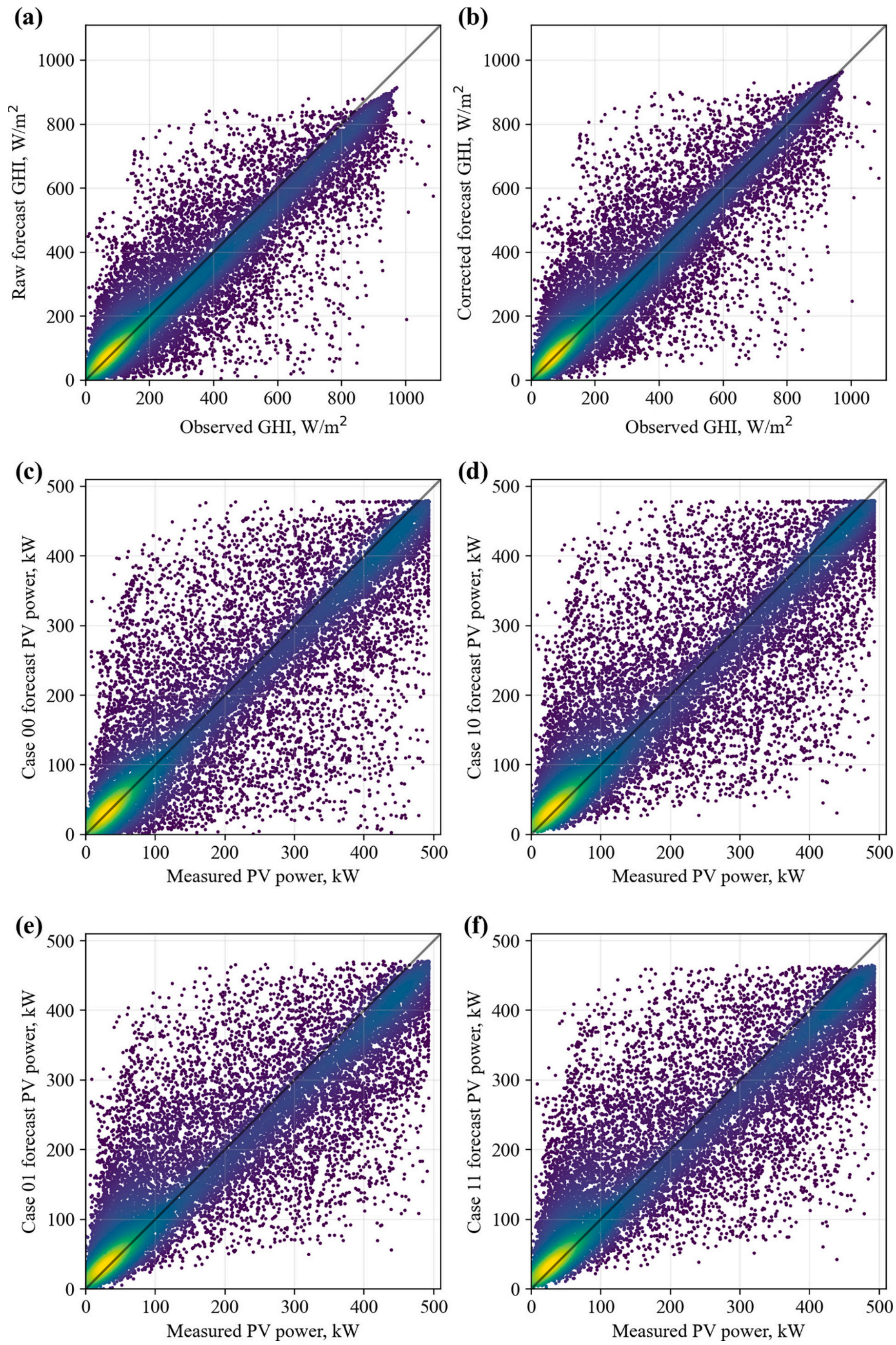

Fig. 3. Joint distribution scatterplots of the PV power measurements and forecasts for Kaposhomok for the quarterly grouped KCDE with three years of historical data. Brighter color indicates a higher density of points.

A longer training dataset improves the accuracy of the forecasts, but this improvement is only marginal in the presented cases. The reason for this might be that both methods rely on only two predictors, namely the clear-sky index and zenith angle, and the relationship between the conditional bias and these basic predictors can be effectively identified from the data of only a single year. More complex post-processing methods relying on more predictors are expected to require a longer dataset to properly map the more complex relationships between the predictors and the predictand, therefore, their accuracy improves to a greater extent as the length of training data increases. A good example of this is the hybrid irradiance-to-power conversion method developed by Mayer [20], where the outputs of a model chain are corrected by an artificial neural network model. In the most complex models with up to 12 predictors, a substantial error decrease is achieved using two years of training data instead of a single year only. This suggests that the increasing length of historical datasets over time puts a greater emphasis on more complex post-processing methods relying on a larger number of relevant predictors and such state-of-the-art machine learning methods that can effectively learn the complex relationships from historical data [50].

In practical applications, historical PV power data is collected only from the commissioning of the plant, therefore, for new PV installations, it takes a long time to obtain multiple years of training data. However, archived historical forecasts can be requested if not purchased from most weather forecasting service providers, and satellite-derived GHI observations covering all regions within $\pm 60^{\circ}$ latitudes can be downloaded for free from several different sources, see Chapter 6 of [51] for detail. To that end, it is typically easier to obtain a longer training dataset for GHI post-processing than for PV power post-processing. In such cases, the two-fold post-processing is recommended, where the GHI is post-processed using the longest possible historical data, and after the irradiance-to-power conversion, the PV power is also post-processed to eliminate the newly introduced bias and optimize the variance (i.e., workflows 11G or 11S but with different training data lengths). Moreover, all papers developing new post-processing methods are encouraged to perform a sensitivity analysis on how the training data length affects the achieved error reduction, in order to find the optimal data requirement of the given method and to help select the most appropriate post-processing method in such practical applications where the amount of historical data is fixed.

5. Conclusion

This paper presents a formal investigation on whether the post-processing in photovoltaic (PV) power forecasting should be applied, following the general belief, on the global horizontal irradiance (GHI) or on the PV power instead. To make such an inquiry, six possible post-processing workflows are compared for the power forecasting of three PV plants in Hungary, using two classic deterministic-to-deterministic post-processing methods, namely, model output statistics (MOS) and kernel conditional density estimation (KCDE), three different fitting strategies, namely, yearly, quarterly, and monthly fittings, and different training data lengths, that is one, two, and three years.

The main finding of this study is that the statistical post-processing is best carried out after the irradiance-to-power conversion since the PV model chains tend to re-introduce bias even to the unbiased GHI forecasts. To that end, unbiased PV power forecasts with the lowest overall errors can only be assured if post-processing is performed directly on the PV power forecasts. The relative reductions in root mean square error (RMSE) are up to 5.3% and 8.5% if the post-processing is applied only on GHI and only on the PV power, respectively. Moreover, if the PV power forecasts are post-processed, only marginal, if any, further error reduction can be achieved by also correcting the GHI forecasts beforehand. It is also shown that satellite-based GHI retrievals can also be used as observations for training, but the achieved error reduction is slightly less than if the training relies on ground-based GHI observations from well-maintained meteorological stations.

The post-processing methods that work well on GHI are also found to be effective when applied on PV power. The best method for correcting GHI is the monthly KCDE trained with three years of historical data, resulting in a 6.9% relative RMSE reduction in terms of GHI, whereas the same method is also among the best ones for correcting the PV power, leading to an even higher 8.5% relative RMSE reduction. This suggests that the common practice of developing post-processing methods for GHI is useful in general, as the results of previous studies of this kind can be easily transferred to PV power post-processing.

The results of the present study challenge the common belief that in the three-step PV power forecasting framework, post-processing should be applied on the GHI before converting it to PV power. However, the post-processing used in the paper falls into the regression thinking tool of the deterministic-to-deterministic category, while post-processing can be used for a wide range of different purposes, e.g., changing the resolution of forecasts, converting deterministic forecasts into probabilistic ones, or calibrating probabilistic forecasts [29]. To that end, similar studies should be conducted in the future for other post-processing purposes to find if the superiority of PV power post-processing is a general trend or if it is only true in this particular kind of application.

CRediT authorship contribution statement

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

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 MVM Green Generation Ltd. for the PV plant design and production data and Gabriella Szépszó from the Hungarian Meteorological Service for providing the GHI observations and the AROME numerical weather prediction data.

Martin J. Mayer is supported by the National Research, Development and Innovation Fund, project no. OTKA-FK 142702 and ÚNKP-22-5-BME-305 and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Dazhi Yang is supported by the National Natural Science Foundation of China (project no. 42375192) and the China Meteorological Administration Climate Change Special Program (CMA-CCSP; project no. QBZ202315).

References

[1] 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.

[2] Wang H, Yi H, Peng J, Wang G, Liu Y, Jiang H, et al. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Convers Manag 2017;153:409–22. https://doi.org/10.1016/J.ENCONMAN.2017.10.008.

[3] Sweeney C, Bessa RJ, Browell J, Pinson P. The future of forecasting for renewable energy. WIREs Energy and Environ 2020;9:1–18. https://doi.org/10.1002/wene.365.

[4] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101. https://doi.org/10.1016/j.solener.2017.11.023.

[5] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39:535–76. https://doi.org/10.1016/j.pecs.2013.06.002.

[6] 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 Sust Energ Rev 2022;161:112348. https://doi.org/10.1016/j.rser.2022.112348.

[7] Yang D, Kleissl J. Summarizing ensemble NWP forecasts for grid operators: Consistency, elicitability, and economic value. Int J Forecast 2022. https://doi.org/10.1016/j.ijforecast.2022.08.002.

[8] Wang W, Yang D, Hong T, Kleissl J. An archived dataset from the ECMWF Ensemble Prediction System for probabilistic solar power forecasting. Sol Energy 2022;248:64–75. https://doi.org/10.1016/j.solener.2022.10.062.

[9] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. Renew Sust Energ Rev 2022;154:111768. https://doi.org/10.1016/j.rser.2021.111768.

[10] Yang D, Wang W, Hong T. A historical weather forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for energy forecasting. Sol Energy 2022;232:263–74. https://doi.org/10.1016/j.solener.2021.12.011.

[11] Markovics D, Mayer MJ. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew Sust Energ Rev 2022;161:112364. https://doi.org/10.1016/j.rser.2022.112364.

[12] Sabri M, El Hassouni ME. Predicting photovoltaic power generation using double-layer bidirectional long short-term memory-convolutional network. International Journal of Energy and Environmental Engineering 2023;14:497–510. https://doi.org/10.1007/s40095-022-00530-4.

[13] Sabri M, El Hassouni M. Photovoltaic power forecasting with a long short-term memory autoencoder networks. Soft Comput 2023;27:10533–53. https://doi.org/10.1007/S00500-023-08497-Y/TABLES/11.

[14] 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.

[15] Mayer MJ, Yang D. Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting. Renew Sust Energ Rev 2023;175:113171. https://doi.org/10.1016/j.rser.2023.113171.
[16] Mayer MJ, Yang D. Probabilistic photovoltaic power forecasting using a calibrated ensemble of model chains. Renew Sust Energ Rev 2022;168:112821. https://doi.org/10.1016/j.rser.2022.112821.
[17] 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.
[18] Mayer MJ. Effects of the meteorological data resolution and aggregation on the optimal design of photovoltaic power plants. Energy Convers Manag 2021;241:114313. https://doi.org/10.1016/j.enconman.2021.114313.
[19] 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.
[20] Mayer MJ. Benefits of physical and machine learning hybridization for photovoltaic power forecasting. Renew Sust Energ Rev 2022;168:112772. https://doi.org/10.1016/j.rser.2022.112772.
[21] 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.
[22] Yang D. SolarData package update v1.1: R functions for easy access of Baseline Surface Radiation Network (BSRN). Sol Energy 2019;188:970–5. https://doi.org/10.1016/j.solener.2019.05.068.
[23] Yang D. SolarData: An R package for easy access of publicly available solar datasets. Sol Energy 2018;171:A3–12. https://doi.org/10.1016/j.solener.2018.06.107.
[24] 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.
[25] Szintai B, Szűcs M, Randriamampianina R, Kullmann L. Application of the AROME non-hydrostatic model at the Hungarian meteorological service: Physical parameterizations and ensemble forecasting. Időjárás 2015;119:241–65.
[26] Tóth Z, Nagy Z, Szintai B. Verification of global radiation fluxes forecasted by numerical weather prediction model AROME for Hungary. Időjárás 2017;121:189–208.
[27] János M, Yang D, Szintai B. Comparing global and regional downscaled NWP models for irradiance and photovoltaic power forecasting: ECMWF versus AROME. Appl Energy 2023:1–27. submitted.
[28] Yang D, Perez R. Can we gauge forecasts using satellite-derived solar irradiance? Journal of Renewable and Sustainable Energy 2019;11:023704. https://doi.org/10.1063/1.5087588.
[29] Yang D, van der Meer D. Post-processing in solar forecasting: Ten overarching thinking tools. Renew Sust Energ Rev 2021;140:110735. https://doi.org/10.1016/j.rser.2021.110735.
[30] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J Sel Top Appl Earth Obs Remote Sens 2009;2:2–10. https://doi.org/10.1109/JSTARS.2009.2020300.
[31] Yang D. Post-processing of NWP forecasts using ground or satellite-derived data through kernel conditional density estimation. Journal of Renewable and Sustainable Energy 2019;11:026101. https://doi.org/10.1063/1.5088721.
[32] Yang D. Choice of clear-sky model in solar forecasting. Journal of Renewable and Sustainable Energy 2020;12:026101. https://doi.org/10.1063/5.0003495.
[33] 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.
[34] Yang D, Xia X, Mayer MJ. A tutorial review of the solar power curve: Regressions, model chains, and their hybridization and probabilistic extensions. Adv Atmos Sci 2024;41(6):1023–67. https://doi.org/10.1007/s00376-024-3229-4.
[35] Reda I, Andreas A. Solar position algorithm for solar radiation applications. Sol Energy 2004;76:577–89. https://doi.org/10.1016/j.solener.2003.12.003.
[36] Yang D. Temporal-resolution cascade model for separation of 1-min beam and diffuse irradiance. Journal of Renewable and Sustainable Energy 2021;13:056101. https://doi.org/10.1063/5.0067997.
[37] Yang D. Estimating 1-min beam and diffuse irradiance from the global irradiance: A review and an extensive worldwide comparison of latest separation models at 126 stations. Renew Sust Energ Rev 2022;159:112195. https://doi.org/10.1016/j.rser.2022.112195.
[38] 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.
[39] Yang D. Solar radiation on inclined surfaces: Corrections and benchmarks. Sol Energy 2016;136:288–302. https://doi.org/10.1016/j.solener.2016.06.062.
[40] 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.
[41] 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.
[42] Beyer HG, Bethke J, Drews A, Heinemann D, Lorenz E, Heilscher G, et al. Identification of a general model for the MPP performance of PV modules for the applicationin a procedure for the performance check of grid connected systems. Proceedings of the 19th European Photovolatic Solar Energy Conference. 2004. p. 5.
[43] Varga N, Mayer MJ. Model-based analysis of shading losses in ground-mounted photovoltaic power plants. Sol Energy 2021;216:428–38. https://doi.org/10.1016/j.solener.2021.01.047.
[44] Driesse A, Jain P, Harrison S. Beyond the curves: Modeling the electrical efficiency of photovoltaic inverters. 2008 33rd IEEE Photovolatic specialists conference. IEEE 2008:1–6. https://doi.org/10.1109/PVSC.2008.4922827.
[45] 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.
[46] 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.
[47] Mayer MJ, Yang D. Calibration of deterministic NWP forecasts and its impact on verification. Int J Forecast 2023;39:981–91. https://doi.org/10.1016/j.ijforecast.2022.03.008.
[48] Murphy AH, Winkler RL. A general framework for forecast verification. Mon Weather Rev 1987;115:1330–8. https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2.
[49] Mayer MJ, Yang D. Potential root mean square error skill score. Journal of Renewable and Sustainable Energy 2024;16. https://doi.org/10.1063/5.0187044.
[50] Wang H, Liu Y, Zhou B, Li C, Cao G, Voropai N, et al. Taxonomy research of artificial intelligence for deterministic solar power forecasting. Energy Convers Manag 2020;214:112909. https://doi.org/10.1016/J.ENCONMAN.2020.112909.
[51] Yang D, Kleissl J. Solar Irradiance and Photovoltaic Power Forecasting. CRC Press; 2024. https://doi.org/10.1201/9781003203971.