

# Comparing global and regional downscaled NWP models for irradiance and photovoltaic power forecasting: ECMWF versus AROME

Martin János Mayer<sup>a,\*</sup>, Dazhi Yang<sup>b</sup>, Balázs Szintai<sup>c</sup>

<sup>a</sup> Department of Energy Engineering, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Műgyetem rkp. 3, H-1111 Budapest, Hungary

<sup>b</sup> School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China

<sup>c</sup> Hungarian Meteorological Service, Budapest, Hungary

## HIGHLIGHTS

- Global ECMWF outperforms regional AROME for global horizontal irradiance forecasting.
- The advantage of ECMWF is reduced if photovoltaic power forecasting is concerned.
- Both numerical weather prediction models have the same accuracy for regional forecasting.
- Photovoltaic power forecast errors are the highest in winter.

## ARTICLE INFO

### Keywords:

Verification  
Numerical weather prediction  
Solar irradiance  
Photovoltaic power  
ECMWF  
AROME

## ABSTRACT

Inspecting the literature, much effort has been placed on the verification of irradiance forecasts from numerical weather prediction (NWP) models, as such forecasts are thought to have profound implications on the photovoltaic (PV) power forecasts, which in turn affects grid operators' confidence in integrating such power into the electricity grid. However, perhaps due to the proprietary nature of PV plants and lack of access to state-of-the-art NWP model output, only few have had the chance to conduct head-to-head comparisons of global mesoscale and regional downscaled NWP models, in terms of how their irradiance forecast inaccuracies propagate to PV power forecasts. In this regard, this work presents such a study, in which irradiance and PV power forecasts from the European Centre for Medium-Range Weather Forecasts' High-Resolution (HRES) and Météo-France's Application of Research to Operations at Mesoscale (AROME) models are thoroughly verified against the ground-based measurements from 32 research-grade radiometry stations and 94 actual PV plants in Hungary. A wide range of techniques and case studies concerning verification is herein considered, including variance ratio analysis, Murphy–Winkler decomposition, point-versus-areal verification, and seasonal verification. Despite that the results are too numerous to be summarized in a few sentences, the overarching observation from the verification exercise is that the performance of irradiance forecasts can only be used to infer that of PV power forecasts to a certain extent, which contrasts the conventional wisdom.

## 1. Introduction

It is, now, widely accepted that numerical weather prediction (NWP) plays an essential role in photovoltaic (PV) power forecasting, particularly over horizons longer than a few hours, which are required by the grid operators for conducting power system scheduling and dispatch [1]. Stated more precisely, international and national weather centers worldwide have recognized the importance of providing data support, in

the style of disseminating weather forecasts, to PV plant owners, who are responsible for converting the issued weather forecasts to PV power forecasts and submitting the latter to grid operators [2]. The utility of this two-step PV power forecasting framework, namely, first of NWP and then of irradiance-to-power conversion, is believed to be able to persist into a foreseeable future, insofar as the power grid is to be transformed from today's marginal-renewable context into tomorrow's renewable-dominant context.

\* Corresponding author.

E-mail address: [mayer@energia.bme.hu](mailto:mayer@energia.bme.hu) (M.J. Mayer).

<https://doi.org/10.1016/j.apenergy.2023.121958>

Received 8 June 2023; Received in revised form 30 August 2023; Accepted 11 September 2023

Available online 20 September 2023

0306-2619/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

NWP depends for its capacity to issue forecasts upon integrating forward in time a set of differential equations that describes grid-scale motions, whereas using parameterizations to account for the non-physically resolved atmospheric processes in the subgrid scale. The atmosphere is fluid, as such it obeys the laws of fluid mechanics (i.e., conservation of momentum), thermodynamics (i.e., conservation of energy), and conservation of mass and water, which are prognostic in nature. Together with the ideal gas law, which is diagnostic in nature, these primitive equations jointly describe what is known as the model dynamics [3]. On the other hand, the word “physics,” in the weather forecasting community, commonly denotes processes that take place in the atmosphere on smaller scales, such as turbulence, small-scale processes in clouds (i.e., cloud microphysics), friction from subgrid features at the earth’s surface, or rain [4]. Physics in NWP is handled by parameterization schemes, which aim at quantifying the various sources and sinks in the prognostic equations, and thereby closing the set of primitive equations [5]. It should be highlighted that NWP is a difficult subject on its own, and techniques pertaining to it are still subject to an indefinite amount of refinement. To that end, most solar forecasters, such as PV plant owners, have no choice but to remain as end users of weather forecasts.

Suppose the weather forecasts are made available to the plant owners, the next step is to convert them into PV power forecasts based on the plant design parameters and specifications, such as panel orientation, string layout, or inverter sizing. Generally, the higher granularity at which such information is made available to the modeler, the more accurately one is able to perform the conversion. There are two main ways of converting relevant weather variables (namely, irradiance, air temperature, wind speed, and albedo) to PV power, one direct and the other indirect. The direct approach uses statistical and machine-learning methods to establish a correspondence between weather variables and PV power. The indirect approach leverages a set of solar energy meteorology models, including but not limited to the separation model, transposition model, reflection model, DC model, various loss models, and AC model. These models are applied in cascade, where the output of a preceding model is used as input by the succeeding one, in that, the indirect method is also known as the *model chain*. Model chain has been extensively studied during the past few years, and the reader is referred to [6–8] for representative works.

Despite that the two-step framework for PV power forecasting is now commonplace, many technical questions are yet to be attended to. One such question is regarding the choice of the NWP model, that is, how the choice would affect the final quality of PV power forecasts. Broadly, NWP models can be divided, based on their geographical extent, into global and regional (i.e., limited area) models, of which the latter often uses the output of the former as boundary and initial conditions [9]. A key difference between the global and regional models is in terms of the model resolutions in both space and time. Regional models dynamically downscale the output of global models, as such their resolutions are usually higher. One advantageous aspect of regional models is that they often leverage meteorological observations that are not available to the global model, and assimilating such data could lead to an initial state of the atmosphere that is more apt [10]. Notwithstanding, since a regional model uses a different set of parameterization schemes than the global counterpart, it may not always lead to a favorable outcome in terms of predictive accuracy, because the interplay of different parameterization schemes is by far the most challenging aspect of weather modeling, and there is only limited knowledge in regard to which combination of schemes is better for which weather regime. Hence, from a practical viewpoint, comparing NWP models is doubtlessly perceived as an effective means of model selection [11]. This task, in the field of forecasting, is known as comparative verification, which has hitherto been attracting attention.

Indeed, there have been many works that seek to compare two or more NWP models. Because atmospheric predictability varies across geographic regions, seasons, weather regimes, and time scales, among

other factors, this kind of comparative study is, in the main, of interest only to the situations considered, despite the forecast error tabulation and verification procedure are often lengthy. For instance, perhaps the largest collaboration in comparing NWP forecasts, in the field of solar forecasting, is the one conducted by Perez et al. [12], who considered eight NWP models, of which many have several variants. The general observation made therein is that the European Centre for Medium-Range Weather Forecasts (ECMWF), which is a global model, performs substantially better than the limited-area Weather Research and Forecasting (WRF) models initialized by the Global Forecast System (GFS). Whereas Perez et al. [12] compared global and regional models that are great in number, others have approached the problem by emphasizing procedural complexity. Yang et al. [13] took into consideration only three models, namely, the ECMWF model and two rapid refresh models from the National Oceanic and Atmospheric Administration (NOAA), of which the latter two issue new sets of forecasts every hour. The main tool employed therein is the Murphy–Winkler mean square error (MSE) decomposition, which decomposes the MSE between forecasts and observations into numerous indicators of aspects of quality; however, this decomposition necessarily led to many figures and tables, which complicate model ranking. Besides these two works, some have compared post-processed versions of NWP models [14,15]. Be that as it may, one aspect of NWP comparison that has hitherto remained more or less unexplored is how different NWP models would affect PV power forecasts. This problem is thought to be primarily due to the lack of public-domain PV data, as most data of that sort is of a proprietary nature. To that end, this work should wish to present such a study.

This work considers two NWP models, one global and the other regional. The global model of choice is the ECMWF High-Resolution (HRES) model, which has been shown, by references mentioned above and those elsewhere, to be one of (if not) the best global models to date. The regional model of choice is a version of the Application of Research to Operations at Mesoscale (AROME) model employed by the Hungarian Meteorological Service (OMSZ).<sup>1</sup> On top of the NWP forecast verification, major attention has been placed on verifying how the NWP forecast errors would amplify (or reduce) when the NWP forecasts are converted to PV forecasts; this is only possible if irradiance measurement stations collocate with PV plants. On this point, the greatest attraction of this work is perhaps the inclusion of data from 32 meteorological stations and 94 utility-scale PV plants, among which four pairs are intimately located; this scale of verification exercise, to the authors’ best knowledge, is unprecedented.

The contribution of this work does not end there. The verification framework used herein integrates some of the most insightful methods and techniques proposed in the forecast verification literature, enabling a comprehensive analysis of the different aspects of forecast quality. Besides performing the verification for point locations only, a regional forecasting scenario is also investigated, which is one of the most important products of solar forecasting from the perspective of large-scale grid integration. Based on this, further insights are provided on the error reduction potential resulting from spatial smoothing in large geographical areas and its implications for the NWP model selection. All presented results are discussed in great detail, referring back to the theoretical basics of irradiance and PV modeling, e.g., it is thoroughly explained how and why the initial advantage of ECMWF in irradiance forecasting disappears when it comes to regional PV power forecasting.

The remaining pages of this work comprise the following sections. It proceeds with Section 2, in which the two NWP models of concern are briefly introduced, and their pros and cons elaborated. Ground-based data, from both meteorological stations and PV plants, are discussed in Section 3, alongside the verification methodology. In particular, the Murphy–Winkler MSE decomposition is reviewed in Section 3.3,

<sup>1</sup> The abbreviation “OMSZ” reflects the Hungarian names of the Hungarian Meteorological Service, which is Országos Meteorológiai Szolgálat.

whereas the concept and use of the model chain are the subjects of discussion in Section 3.4. Section 4 presents the verification results, which are divided into two parts. In the first, results pertaining to the four collocated station–plant pairs are depicted, whereas in the second, the overall regional performance of the two NWP models is discussed. Conclusions follow at the end.

## 2. Numerical weather prediction models

### 2.1. ECMWF IFS

At the ECMWF, the Integrated Forecasting System (IFS) is developed and run operationally at several resolutions. The HRES model considered here is the highest-resolution model of IFS. The horizontal resolution of the HRES is about 9 km, whereas in the vertical, there are 137 levels. The model is run four times a day (at 00Z, 06Z, 12Z, and 18Z) with a maximum horizon of 10 days. The forecasts issued have an hourly resolution over the first 90 h, and the resolution drops to 3 h over 93–144-h-ahead, and 6 h over 150–240-h-ahead horizons. (The 3 and 6 hourly forecasts are only available for 00Z and 12Z runs.) HRES uses a four-dimensional variational data assimilation system.

Radiation in HRES is parameterized with the ecRad scheme [16], which applies the Rapid Radiative Transfer Model for GCMs (RRTM-G; [17]) both for the shortwave and longwave part of the spectrum. Monthly ozone and aerosol climatologies are derived from the products of the Copernicus Atmospheric Monitoring Service (CAMS). Due to the high computational cost of radiation parameterization, the ecRad scheme is called every hour and on a coarser grid.

Considering the resolution of the HRES model, deep convection is not resolved explicitly and has to be parameterized. For this, originally, the mass-flux formulation described in [18] was used, which was further developed in many aspects in recent years [19,20]. Stratiform cloud microphysics are computed using five prognostic hydrometeors (vapor, cloud water, cloud ice, rain, and snow) as described in [21]. Turbulent diffusion in the boundary layer is described by a first-order K-closure in stable, and the eddy-diffusivity mass-flux approach in unstable regimes [22]. Land surface processes are parameterized by the TESSEL scheme (Tiled ECMWF Scheme for Surface Exchanges over Land), which calculates soil moisture and temperature prognostically on four soil layers [23].

### 2.2. AROME

The AROME project was initiated at Météo-France in 2002 with the aim of developing a non-hydrostatic NWP model running at a 2.5-km horizontal resolution [24]. The temporal resolution of the AROME, which is the timestep of the model which is applied during the integration of hydro-thermodynamic equations, is 60 s, but the outputs are produced with a 15-min temporal resolution. The AROME model is now used in several countries of the ACCORD limited-area modeling consortium. In Hungary, AROME has been running operationally at the OMSZ since 2010 [25]. Currently, the deterministic version of AROME operates over a domain covering the Carpathian Basin with a 2.5-km horizontal resolution and 60 vertical levels, and is run eight times a day (at 00Z, 03Z, 06Z, 09Z, 12Z, 15Z, 18Z, and 21Z) up to two days ahead. AROME applies a three-dimensional variational data assimilation system [26].

AROME uses an earlier version of the ECMWF IFS radiation scheme. For the shortwave part, the Fouquart and Bonnel [27] formulation is used, whereas for the longwave part, the RRTM code [17] is applied. Ozone monthly profiles are given by analytical functions that have been fitted to the U.K. Universities Global Atmospheric Modelling Programme (UGAMP) climatology [28] with three coefficients [29]. Monthly aerosol fields are extracted from the Tegen climatology [30]. Radiation parameterization is called every 15 min on the original model grid.

Due to the relatively high horizontal resolution of the AROME model,

it is assumed that deep convection is resolved explicitly, consequently, no deep convection parameterization is required. Convective and stratiform cloud processes are handled with a single scheme, the so-called ICE3 scheme [31], which carries six prognostic microphysical variables (vapor, cloud water, cloud ice, rain, snow, graupel), and describes the phase change processes among these variables. AROME applies a prognostic turbulent kinetic energy (TKE) scheme for the description of small-scale turbulence [32] and uses the eddy-diffusivity mass-flux approach for shallow convective boundary layer [33]. Surface processes are simulated using the Surface Externalisée (SURFEX) [34] platform. SURFEX uses the tiling approach—each grid point is divided into tiles of four different surface types: sea, inland water, town, and vegetated land. Each tile uses the same atmospheric forcing (air temperature, humidity, wind speed, long and shortwave radiation, pressure, and precipitation), but the parameterizations are different and independent of each other.

AROME forecasts are regularly validated with observations in all the ACCORD countries and experiences are shared about model performance. Although model performance differs slightly depending on the local climate, verification results are generally similar, which suggests that the findings of the present study are transferable to other regions in mid-latitudes outside Hungary. It must be noted though that AROME is a very complex modeling system, and certain settings of physical parameterizations and data assimilation can differ from country to country, and this can also have an influence on model accuracy.

### 2.3. Comparison of the two models

The main advantage of the AROME model as compared to HRES is its increased horizontal resolution, which allows it to explicitly resolve deep convection. This results in a more realistic simulation of cloud cover during summertime convective cases. However, the main difficulty of high-resolution modeling is the so-called “double penalty” problem [35]. Coarse resolution models tend to predict a smoother field of cloud cover and precipitation, whereas high-resolution models can simulate individual convective cells and thus the cloud cover and precipitation fields are more of a binary nature. Consequently, if the location of convective cells is slightly shifted in the forecasts of high-resolution models, it can result in large errors.

The second advantage of high-resolution limited-area models is the flexibility toward end-user applications. For instance, it is possible to produce 15-min output from a limited-area model, whereas any resolution beyond hourly is usually not possible for global models in an operational context due to storage concerns. Furthermore, physical parameterizations of limited-area models can be tuned for the specific climate of the model domain, but global models have to operate with the same settings for the entire globe.

Finally, it is important to note the differences between observational data used by global and limited-area models. In the OMSZ’s version of AROME, GNSS Zenith Total Delay (ZTD) data is assimilated operationally, which adds more information on the upper-air humidity and thus improves cloud cover forecasts. Also, it is planned for meteorological radar data to be assimilated into the AROME model in the near future at OMSZ.

Running limited-area models has difficulties as well. The main challenge regarding the AROME configuration run at OMSZ is the correct handling of lateral boundary conditions with respect to prognostic hydrometeors. Due to the substantial differences between the microphysical schemes of ECMWF HRES and AROME it is not possible to couple hydrometeors on the model boundaries. Consequently, these AROME variables receive a value of zero in the boundary coupling zone, and then AROME develops these quantities from temperature and humidity fields over several time steps. As a result, under-estimations of clouds and precipitation often occur close to the boundaries, however, as the boundary zone is more than 300 km away from the borders of Hungary, this model deficiency does not have an impact on the results

presented in this study.

Based on the different resolutions and physics of the two models, several *a priori* expectations can be made regarding their performance in solar forecasting. The higher spatial and temporal resolution of the AROME model suggests that it can better capture the variability of the solar resource. However, it is not clear whether the higher resolution and the inclusion of a wider range of observations could actually translate to forecast accuracy due to the aforementioned “double penalty” problem. The ECMWF model has been used operationally for a longer time, which has allowed more time for fine-tuning the model and reducing its bias. However, the regional AROME model may be better tailored to the weather conditions of Hungary, leading to lower systemic biases. Overall, even though one can have several prior expectations on the characteristics of the forecasts created by each model, their accuracy can only be judged reliably through detailed quantitative verification.

### 3. Ground-based observations and method

Involved in this work are four datasets. Besides the two NWP forecast datasets, namely, ECMWF and AROME, two datasets containing the verifications (i.e., observations) are considered; they are to be narrated in this section, alongside the method used for forecast verification. One of the ground-based datasets consists of global horizontal irradiance (GHI) measurements collected at 32 meteorological stations, and the other consists of PV power measurements from 94 utility-scale PV plants. Both the meteorological stations and PV plants are distributed quite evenly across Hungary, and thus the results acquired here must be deemed as representative for the country if not for a wider region.

#### 3.1. Meteorological stations and PV plants

OMSZ operates 32 meteorological stations with research-grade GHI measurements in Hungary. The GHI is measured by Kipp & Zonen CM11

and SMP11 pyranometers calibrated every three years. The data is sampled every minute and logged at a 10-min resolution, and the time stamp is placed at the end of each 10-min interval (i.e., a ceiling aggregation scheme is used).

On the other hand, the ground-truth data for PV power forecast verification is the measured power output of 94 utility-scale ground-mounted PV plants in Hungary. The nominal DC power of these plants varies between 0.56 MW to 20.7 MW, and their other design parameters, with a few exceptions due to local factors, follow what is considered optimal in Hungary. The PV modules are oriented toward south and tilted between 20° and 35°. The inverters are undersized by 15% on average with respect to the nominal DC power. The spacing of the parallel PV mounting structure rows roughly follows the convention to avoid shading on the day of the winter solstice at least at the solar noon. The power output of these PV plants is measured and stored with a 15-min resolution.

The geographical locations of the meteorological stations and PV plants are shown in Fig. 1. There are five locations where the distance between a station and the nearest PV plant is less than 5 km, four of which are selected for a more detailed analysis considering the geographic diversity of these sites. These four pairs are encircled in Fig. 1, and their names are also displayed in a station–plant order.

Meteorological stations and PV plants are both located fairly evenly across the country. However, PV plants are slightly more concentrated in the south, probably due to the slightly higher solar resources in those areas. Overall, PV plants are located in areas with friendly conditions for PV plant installation, such as high solar resource availability, easy-to-access flat land surface, or good connectivity to the power grid. In contrast, the meteorological stations are typically installed in places where accurate forecasts have a higher social relevance (near big cities, lakes, and airports) or where the weather conditions are interesting (e.g., mountain tops).

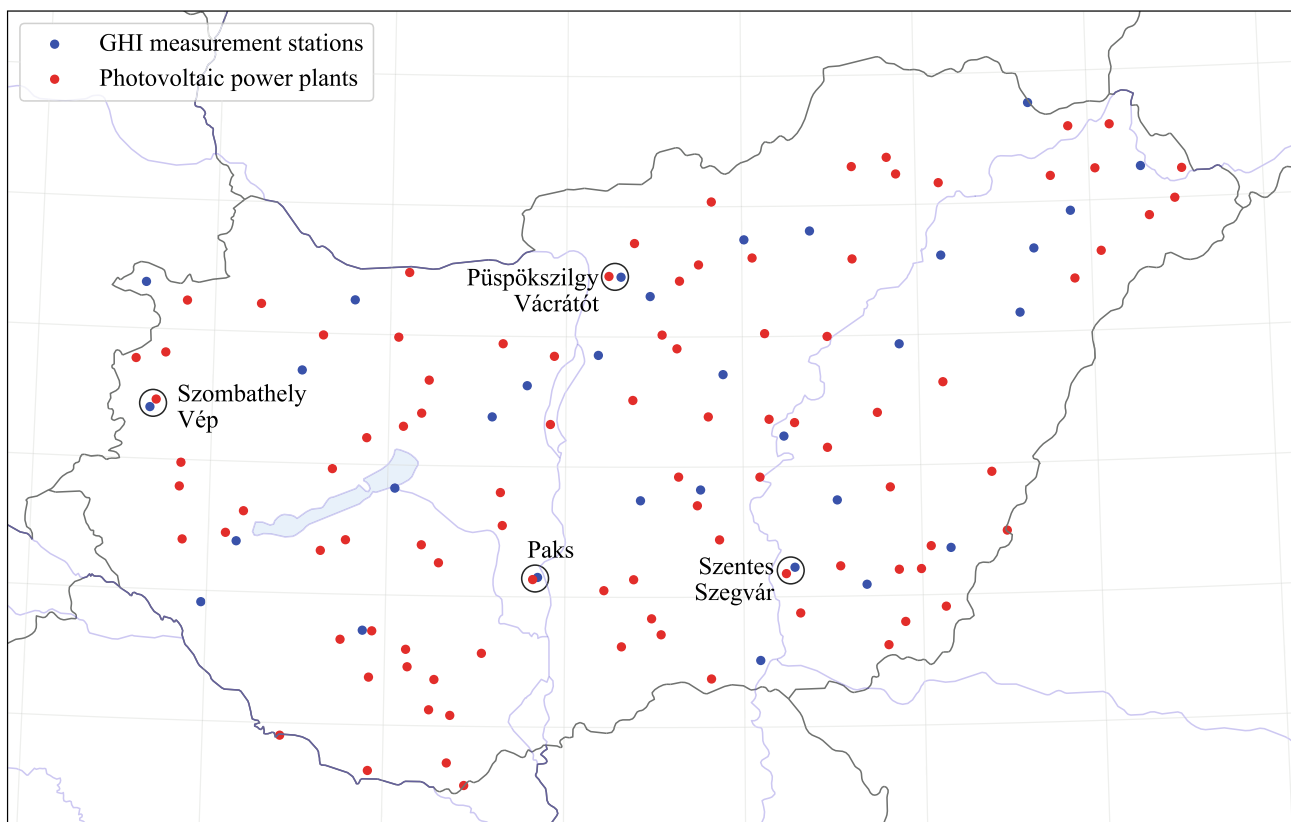


Fig. 1. Locations of the 32 meteorological stations with GHI measurements and the 94 PV power plants in Hungary.



### 3.2. Data preprocessing and quality control

The ECMWF forecasts used in this verification are the deterministic forecasts generated by the HRES model, in which the model output parameter corresponding to GHI is named “surface solar radiation downwards.” The AROME forecasts are issued by the Hungarian implementation of the AROME model. Both the ECMWF and AROME forecasts are provided by the OMSZ in a gridded netCDF format at the native spatial resolution of the respective models. The forecasts for each station and plant are taken from the nearest grid point of the focal location. All forecasts are generated by the 00Z (i.e., 00 UTC) model runs, covering a 48-h forecast horizon since each initialization. Aside from the GHI, forecasts for the 2-m temperature and the two components of the 10-m wind are also included in the files, considering their importance during the physical modeling of the PV power output.

The data from different sources have different temporal resolutions; therefore, they must be resampled to a common resolution before the verification can be performed. Following the imbalance settlement period of the Hungarian electricity market, 15 min is chosen as the common resolution for the verification. To resample the GHI data with different resolutions, the “clear-sky interpolation” is conducted, which considers the diurnal cycle of GHI. This process starts with turning the GHI time series into a clear-sky index time series by dividing it by the clear-sky GHI; then, a linear interpolation is performed in clear-sky index terms, and the interpolated clear-sky index time series is converted back to GHI in the last step. The clear-sky GHI is acquired from the McClear service [36], which is available from the SoDa website and retrievable via the application programming interface offered. The ambient temperature and wind speed forecasts issued by the ECMWF model, as required for the PV power modeling, are downscaled using a simple linear interpolation. During all interpolation procedures, special care must be paid to the time stamps of the data. By default, data for each time stamp stand for the average of the time interval between the given and preceding time stamps in all datasets. First, all time stamps are shifted back to the center of each time interval, and the interpolation is only performed afterward.

To avoid spurious data from entering the comparison and thereby negatively impacting the confidence in the verification results, a simple quality-control procedure is herein used. For GHI measurements, the extremely rare limit, which is recommended by the Baseline Surface Radiation Network as a standard filter, is used. As for PV power, only positive values (i.e., power > 0) are retained, as zero production during the daytime indicates that the PV plant is shut down due to malfunction or maintenance. Lastly, a zenith angle filter of  $Z < 85^\circ$  is applied to both GHI and PV power, to eliminate those low-sun instances that are often highly uncertain yet unimportant to solar power generation. The zenith angle is calculated for the center of each time interval.

### 3.3. Verification of deterministic forecasts

The discussion about the best forecast verification practice dates back several decades [37]. In the context of deterministic solar forecasting, the most comprehensive paper summarizing all the latest recommendations is the one by Yang et al. [38] in 2020. However, even since then, important further additions to this topic have been published, such as those findings related to the role of correlation coefficient and the calibration of the forecasts [39]. In order to perform the most insightful verification that can be useful for the widest range of the possible readership, the most valuable recommendations of these recent works are considered in this paper, alongside the operational verification practices of the OMSZ.

The most common error metrics, which are widely used not only in forecast verification but also in model validation in general, are mean bias error (MBE), mean absolute error (MAE), and root mean square error (RMSE). These metrics are calculated as

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

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |f_i - x_i|, \text{ and} \quad (2)$$

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

where  $f$  and  $x$  denote the forecast and observation, and  $N$  is the number of data points. The mean-normalized forms of these metrics, namely the nMBE, nMAE, and nRMSE, are calculated by dividing the respective metric by the mean of the observations.

The correlation coefficient measures the association between the forecasts and observations. It is calculated as

$$\rho = \frac{\text{cov}(f, x)}{\sigma_f \sigma_x}, \quad (4)$$

where  $\text{cov}(\cdot, \cdot)$  stands for the covariance and  $\sigma$  for the standard deviation. The correlation coefficient reflects the potential skillfulness of forecasts, since it does not depend on the calibration of the forecasts, i.e., under which directive the forecasts are optimized [39]. It also has a disadvantage, in that, the correlation coefficient is independent of the bias and variance of the forecasts; therefore, it should be used along with other metrics that can gauge these properties. The ratio of the variance of the forecasts and that of the observations, which was introduced in [6], is complementary to the correlation coefficient in forecast verification; the variance ratio is defined as

$$F = \frac{\mathbb{V}(f)}{\mathbb{V}(x)}, \quad (5)$$

where  $\mathbb{V}(\cdot)$  is the variance operator. The variance ratio shows what proportion of the variability of the observations is captured by the forecasts, where  $F < 1$  indicates underdispersed forecasts, whereas  $F > 1$  indicates overdispersed forecasts.

Murphy and Winkler [37] proposed a distribution-oriented forecast verification framework that can provide deeper insights into the forecast quality than the classical measure-oriented approach. It relies on the joint distribution of forecasts and observations, and offers two different factorizations of the joint distribution involving the conditional and marginal distributions. These factorizations can be translated into two different decompositions of the MSE, namely, the calibration–refinement factorization,

$$\text{MSE}(f, x) = \mathbb{V}(x) + \mathbb{E}_f[f - \mathbb{E}(x|f)]^2 - \mathbb{E}_f[\mathbb{E}(x|f) - \mathbb{E}(x)]^2, \quad (6)$$

and the likelihood–base rate factorization,

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

where  $\mathbb{E}(\cdot)$  is the expectation operator. The four terms of these equations represent four different aspects of quality:

- $\mathbb{E}_f[f - \mathbb{E}(x|f)]^2$ : type-1 conditional bias (lower is better)
- $\mathbb{E}_f[\mathbb{E}(x|f) - \mathbb{E}(x)]^2$ : resolution (higher is better)
- $\mathbb{E}_x[x - \mathbb{E}(f|x)]^2$ : type-2 conditional bias (lower is better)
- $\mathbb{E}_x[\mathbb{E}(f|x) - \mathbb{E}(f)]^2$ : discrimination (higher is better)

An extensive discussion on the calculation and interpretation of these terms can be found in [38].

In addition to the metrics, which are quantitative, four different kinds of diagrams are used to visually assess the quality of forecasts. The joint distribution of forecasts and observations, which contains all time-

independent information relevant to verifying the quality of the forecasts, are visualized through scatterplots that are colored based on the density of data points. Time series plots of the forecasts and observations for sample days from each month display how forecast errors evolve over time, which facilitates visual judgment on the time-related information pertaining to forecast verification. The remaining two diagrams are inspired by the operational forecast verification practices of the OMSZ. The first of these is a plot of the average forecast errors as a function of the lead time over the 48-h forecast horizon. This diagram is useful to reveal the daily patterns in forecast errors and the deterioration of forecast accuracy over time. The second diagram shows the forecast errors over the year, smoothed by a moving average with a uniform window of 11 days, indicating the variation of forecast accuracy over the year continuously without the need for binning the data into months or seasons. These plots can be created for any error metric that involves averaging; in this paper, the choices include the MBE, MAE, and RMSE.

### 3.4. Photovoltaic power modeling

The PV power forecasts can be derived from the weather forecasts via a physical model chain, as visualized in Fig. 2. The inputs of the model chain are the forecast weather parameters, namely the GHI, ambient temperature, and wind speed, whereas the output is the forecast PV power. The first stage of the model chain is to calculate the zenith and azimuth angles, which describes the position of the sun, based on the time stamps and the geographic coordinates of the PV plant using the Solar Position Algorithm (SPA) of the National Renewable Energy Laboratory (NREL) [40]. Solar positioning is followed by separation modeling, which decomposes GHI into beam and diffuse components. Among the many different models proposed in the literature, the Yang model [41] is used herein, as it is the most accurate separation model according to a recent worldwide comparison study [42]. The next step is to transpose the irradiance from the horizontal onto the tilted plane on which the PV modules lie. The irradiance transposition is done via the

Perez model [43], which is widely recognized as the best transposition model for over three decades [44]. The reflection losses in the beam and diffuse irradiance components, due to the module encapsulation, are calculated separately using the physical air-glass model [45], following the procedure presented in [46]. The cell temperature of the PV modules is estimated as a function of the plane-of-array irradiance, ambient temperature, and wind speed using the model developed by Mattei et al. [47]. The DC power output of the PV modules is calculated by the model proposed by Beyer et al. [48] using the fixed coefficient as implemented in [49]. The degradation losses are also taken into account by a fixed 2% loss for the light-induced degradation and a 0.5% annual degradation loss factor, as presented in [50]. The shading losses are calculated assuming parallel mounting structure sheds with a uniform row spacing, and it also considers the mismatch losses of the PV modules due to the non-uniform irradiance [50]. The masking of the diffuse irradiance, which is also an important contributor to the overall shading loss [51], is calculated by a decreased view factor between the PV modules and the sky [52]. The inverter efficiency is calculated by the power and voltage-dependent formulation of Driesse et al. [53], and the clipping of the power output at the nominal AC power of the inverters is also considered. Finally, the losses on cables, transformers, and other devices are estimated based on the relevant information found in the design documentation of the PV plants.

The detailed physical modeling of PV plants requires extensive knowledge of their design parameters, which are typically available from their design documentation. However, the physical model chains can also be used when some design data are unknown, in that, a decent accuracy may still be achieved even if only the DC and AC nominal power and the tilt and azimuth angles of the PV plants are known [7]. Since these basic parameters are commonly accessible for utility-scale PV plants, design data availability does not constitute a limiting factor in forecast verification for PV power.

In the last step, before the verification results are presented, the forecast and measured PV power data are both normalized to a 1-kWp installed capacity, i.e., this is done by dividing the actual power by the nominal DC power of each PV plant. The reason for this is to ensure that all PV plants contribute to the averaged error metrics with equal weights during regional verification, such that the results are not dominated by those of the bigger PV plants.

## 4. Results and discussion

The verification results are split into two parts. Among the meteorological stations and PV plants, there are four pairs that are closely located, namely, Paks–Paks, Püspökszilgy–Vácrátót, Szentes–Szegvár, and Szombathely–Vép (see Fig. 1). Hence, the first part of this section revolves around these four pairs, as to check whether or not there is any consistency between irradiance and PV power forecasts in terms of their verification outcome. In the second part, verification results for all stations and plants over the country are comprehensively depicted.

### 4.1. Results for the four station–plant pairs

The error metrics calculated for the GHI and PV power forecasts of the AROME and ECMWF models are summarized in Table 1 for the four collocated station–plant pairs. All metrics representing the overall errors, which are MAE, RMSE, and correlation coefficient, indicate that the ECMWF model is more accurate than the AROME model for all locations and for both GHI and PV power forecasts. The errors of the PV power forecasts are higher than the GHI forecasts due to the error amplification of the model chains, which has been discussed in great detail in [49]. From the MBE and nMBE entries, it can be seen that a smaller bias in GHI forecasts does not necessarily translate to a smaller bias in PV power forecasts. More specifically, ECMWF GHI forecasts have a smaller bias than AROME GHI forecasts for all four station–plant pairs, yet the biases in their respective PV power forecasts are

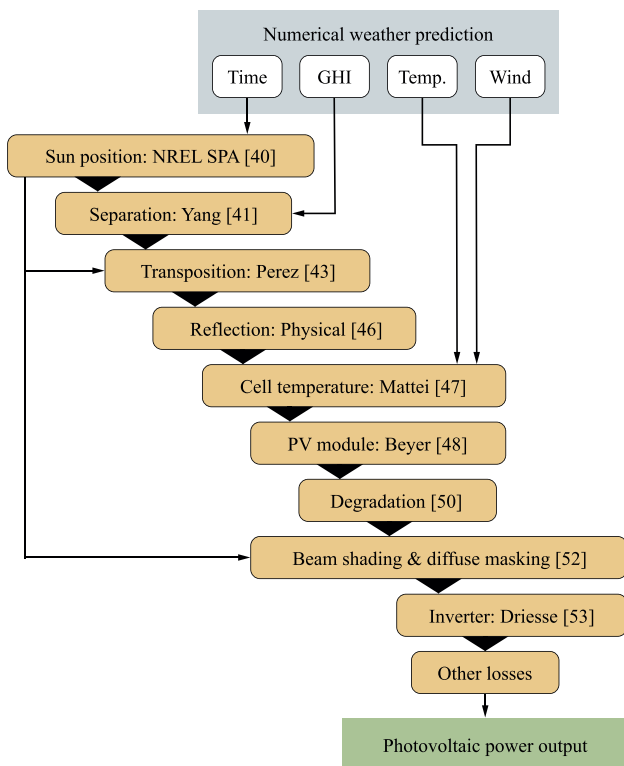


Fig. 2. Schematic of the physical model chain used for the PV power calculation.

**Table 1**

Error metrics for the four station–plant pairs. The metrics for the GHI and PV power are in  $\text{W/m}^2$  and  $\text{W/kWp}$ , respectively.

	Model	MBE	MAE	RMSE	nMBE	nMAE	nRMSE	Corr. coeff.	Var. ratio	Sqrt. type-1 CB	Sqrt. res.	Sqrt. type 2 CB	Sqrt. disc.
Szombathely, GHI	ECMWF	− 4.8	78.1	115.9	1.4%	22.8%	33.8%	0.898	74.2%	16.1	236.1	61.8	203.8
	AROME	21.4	84.9	129.0	6.3%	24.8%	37.6%	0.875	81.3%	26.5	230.1	61.0	207.6
Vép, PV power	ECMWF	23.2	85.2	130.7	6.7%	24.7%	37.9%	0.863	88.9%	31.9	217.2	55.9	205.6
	AROME	0.8	92.8	144.7	0.2%	26.9%	41.9%	0.831	96.6%	39.4	209.3	49.4	206.3
Paks, GHI	ECMWF	6.0	79.5	117.0	1.7%	23.1%	34.0%	0.892	81.8%	12.8	231.3	52.5	209.5
	AROME	19.4	86.1	127.1	5.6%	25.1%	37.0%	0.876	85.6%	25.5	227.0	54.3	210.2
Paks, PV power	ECMWF	31.9	86.0	139.8	9.6%	25.9%	42.1%	0.859	99.5%	50.2	221.2	52.4	221.1
	AROME	− 3.5	89.3	144.9	1.1%	26.9%	43.7%	0.840	99.2%	43.0	216.5	44.7	215.6
Szentes, GHI	ECMWF	− 7.1	77.7	115.5	2.0%	22.0%	32.6%	0.903	77.1%	13.8	241.9	58.1	212.8
	AROME	33.0	85.5	126.3	9.3%	24.2%	35.7%	0.890	80.9%	34.5	238.6	63.9	214.8
Szegvár, PV power	ECMWF	27.2	84.6	137.3	8.2%	25.6%	41.5%	0.865	96.6%	42.9	226.9	52.3	223.8
	AROME	− 7.4	88.0	143.1	2.2%	26.6%	43.2%	0.849	97.8%	39.5	222.9	45.4	220.5
Püspökszilgy, GHI	ECMWF	− 3.7	80.9	117.8	1.1%	23.3%	33.9%	0.897	74.2%	14.3	237.7	63.2	205.4
	AROME	25.0	88.8	132.9	7.2%	25.6%	38.3%	0.871	78.9%	28.2	230.9	67.1	205.5
Vácraót, PV power	ECMWF	10.6	90.0	142.3	3.0%	25.5%	40.3%	0.860	89.1%	28.8	236.1	58.1	223.8
	AROME	14.0	97.3	156.4	4.0%	27.5%	44.2%	0.833	91.9%	40.3	228.9	60.7	220.0

significantly higher, despite the RMSEs of ECMWF-based PV power forecasts are still lower, which is an interesting observation on its own account, motivating in-depth analysis of the other more revealing measures of quality.

The variance ratios show that all forecasts are underdispersed, but the variances of the ECMWF forecasts are even lower than those of the AROME forecasts. For the ECMWF forecasts, the underdispersion could be partly attributed to the downscaling of the hourly forecasts to a 15-min resolution. However, the verification was also performed on hourly aggregated data, showing the exact same trends, which means that the underdispersion is inherited from the model itself, and the variance is only slightly influenced by the downscaling. The higher variances of the AROME forecasts as compared to those of the ECMWF can be clearly attributed to the higher spatial and temporal resolutions of the AROME model. The underdispersion of the GHI is also apparent from the joint distribution scatterplots shown in Fig. 3, where it is realized as a positive bias in the low, and a negative bias in the mid-to-high, GHI ranges. Another representation of this phenomenon is the lack of high GHI forecasts, as the forecast GHI never exceeds  $900 \text{ W/m}^2$ , even though among the observations GHI values greater than  $1000 \text{ W/m}^2$  are not all rare. A similar underdispersion of the ECMWF forecasts has also been reported for sites in the United States, and found to be a deficiency of specific to the ECMWF model that is not shared by other NWP models, such as the NOAA models [13]. Nonetheless, since AROME is down-scaled from ECMWF, it inherits such property to a large extent. That said, as it can be proven theoretically [39], forecasts calibrated for the lowest MSE are always underdispersed. Hence, such underdispersion is not necessarily a deficiency, but it might be an intended characteristic that serves for the lowest possible MSE of the forecasts.

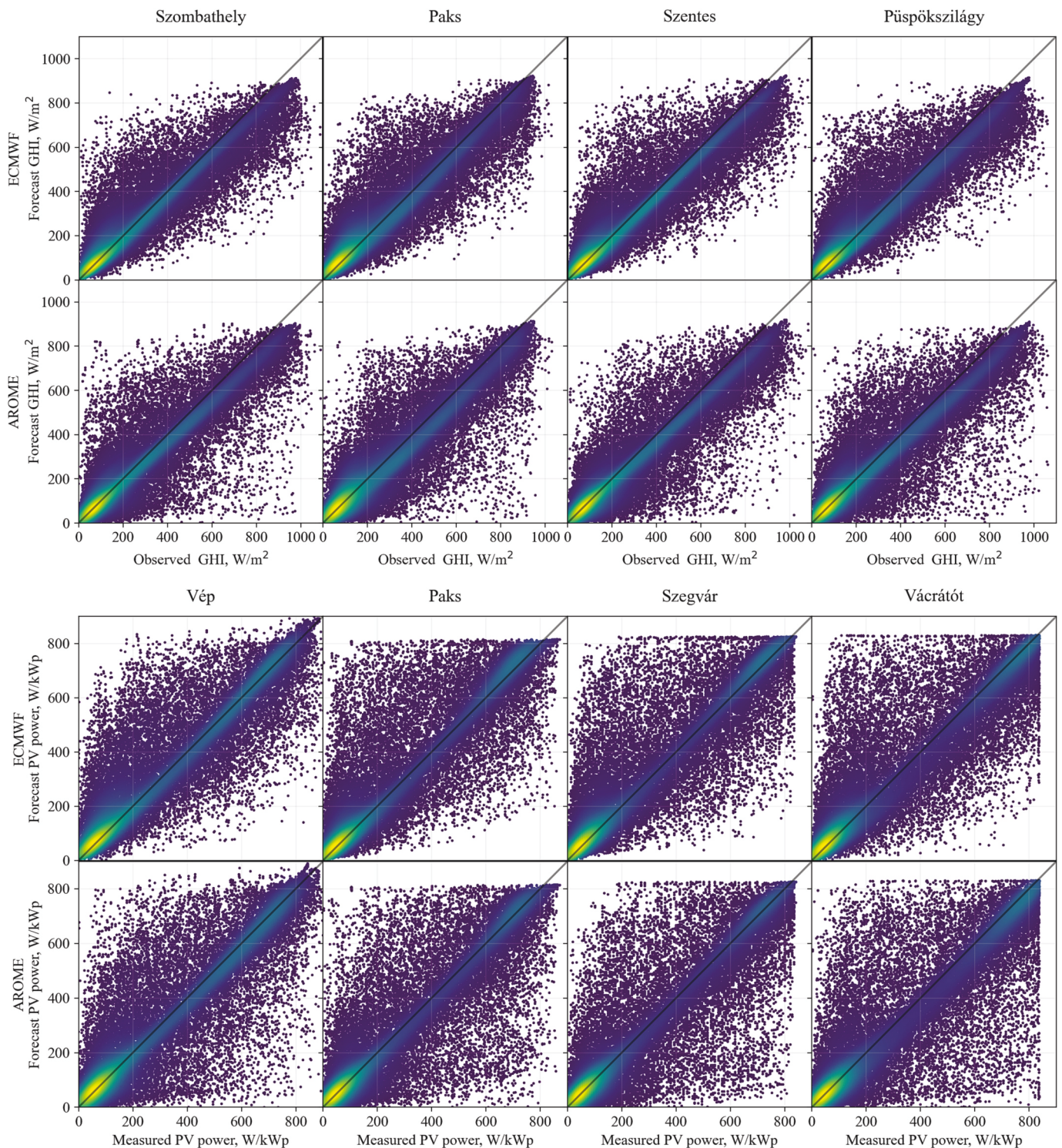
The last four columns of Table 1 show the measure of quality resulting from the Murphy–Winkler MSE decomposition. The pattern there does not seem to be able to lead to any easy conclusion. For instance, type-1 conditional bias is a measure of calibration (or reliability), as it suggests the departure between the mean observation given a particular forecast, i.e.,  $\mathbb{E}(x|f)$ , and the conditioning forecast,  $f$ . Whereas the ECMWF is found to be better than AROME for most cases, the type-1 conditional bias of the PV forecasts from the former is higher than that from the latter at Paks and Szegvár. The results for type-2

conditional bias and discrimination are similar, in that, neither model consistently outperforms the other. The notable exception here is the result of resolution, under which the ECMWF-based forecasts dominate those of AROME—ECMWF has better (i.e., higher) resolution than AROME, for both GHI and PV power forecasts. Recall that the resolution marks the squared difference between the condition and unconditional mean observation, i.e.,  $\mathbb{E}_f[\mathbb{E}(x|f) - \mathbb{E}(x)]^2$ . This indicates that ECMWF can generate forecasts that are more insightful, in the sense that different forecasts are followed by different observations. (If the conditional and unconditional means are the same, it means the observations materialize with disregard for what forecasts are being generated, which is clearly not a desirable trait.)

Comparing the spread of the scatters visually, one may conclude that large errors are more common in the AROME forecasts than in the ECMWF ones, which explains the higher values of the error metrics, especially in terms of the RMSE, which is more sensitive to large errors. The spread of the scatterplot for the PV power forecasts is even wider than for the GHI. The most noticeable difference between the scatterplots for the GHI and PV power is that, for the latter, the high forecast and observations values are both clipped by the undersized inverters. As a consequence, the errors and, most importantly, the significant negative bias in the high GHI range cannot propagate into the PV power forecasts. On the one hand, this is the main reason why there are more positive biases in the PV power forecasts than in the GHI forecasts. On the other hand, it also explains the lower underdispersion of the power forecasts.

The time series displayed in Fig. 4 reveals that the regional AROME forecasts are not simply a downscaled version of the global ECMWF forecasts, but due to the differences in the model physics, the forecast weather trajectories also often differ. The most prominent example of this is the 25th of February: On that day, the ECMWF model predicted clear-sky conditions (which happened to be the case), whereas AROME predicted cloudiness for most of the country. Beyond this, significantly different forecasts can also be seen on other days, e.g., on the 20th of April, 18th of July, and 20th of September. On clear days, like the 10th of May and 10th of August, both models under-estimate the GHI around the solar noon, indicating a negative bias at high GHI values, which is a projection of their underdispersion. However, due to inverter clipping,





**Fig. 3.** Joint distribution scatterplots of the GHI observations and forecasts (top), and PV power measurements and forecasts (bottom), issued by the ECMWF and AROME models, for the four selected meteorological stations and PV plants. Lighter colors indicate a higher density of data points.

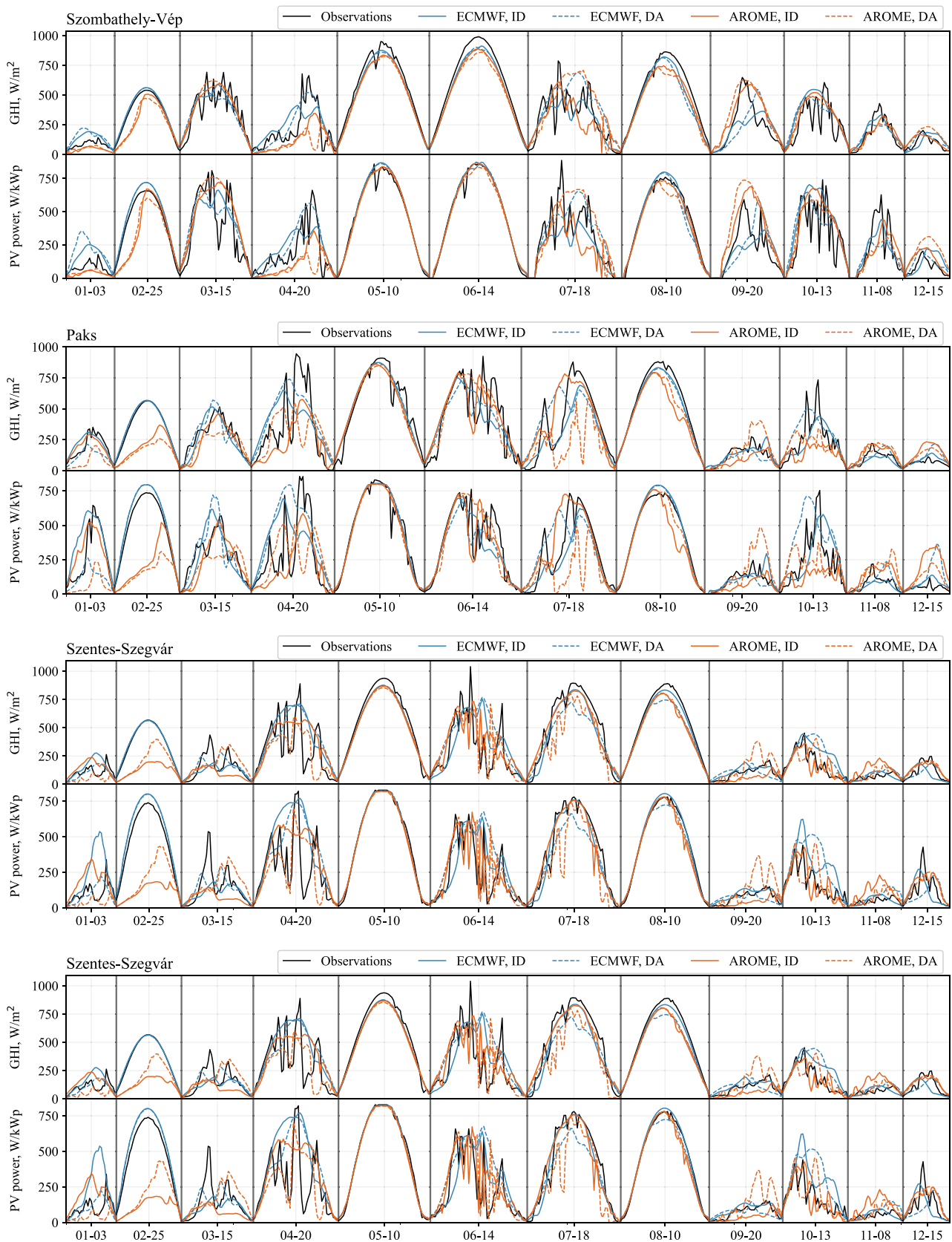
these errors do not appear in the PV power forecasts, which are quite accurate based on both NWP models. On days with variable sky conditions, none of the NWP models is able to properly capture the fluctuations of the GHI. On most days, ECMWF forecasts are smoother and AROME forecasts are the more fluctuating, which aligns with their higher variance ratio and the fact that the ECMWF forecasts are interpolated from the hourly resolution series, whereas AROME has higher spatial and temporal resolutions. This smoother transient of ECMWF

allows it to avoid larger errors, as also evidenced in the joint distribution scatterplots in Fig. 3.

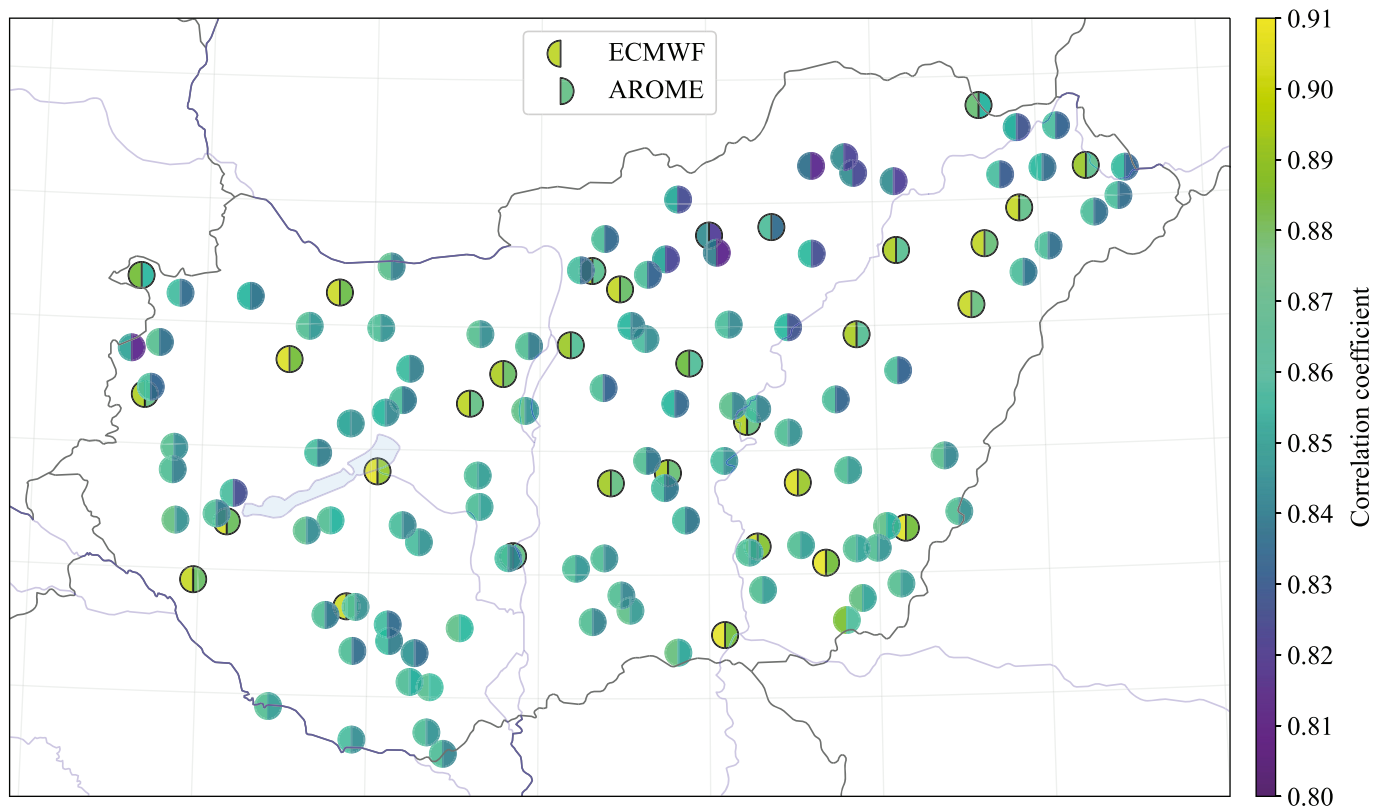
#### 4.2. Results for all stations and plants

The correlation coefficient, which provides a good overview of the potential accuracy of forecasts, is overlaid onto the map of Hungary in Fig. 5, for all meteorological stations and PV plants, and for both the





**Fig. 4.** Time series plots of the GHI and PV power observations and forecasts for the four station–plant pairs. ID (intraday) and DA (day-ahead) refer to the 0–24 h and 24–48 h lead time part of the forecasts, respectively.



**Fig. 5.** Correlation coefficient between forecasts and observations for all 32 meteorological stations (circles with black contour) and all 94 PV plants (circles without contour). For each circle, the lighter-colored half-circle corresponds to better forecasts.

ECMWF and AROME forecasts. For each circle in that plot, the left half circle corresponds to the ECMWF forecasts, whereas the right half corresponds to the AROME forecasts; brighter colors denote better forecasts. Comparing the left and right halves of each circle, it becomes clear that the ECMWF forecasts are more accurate at every single stations and plants than the AROME forecasts. Moreover, despite the slight regional differences, the PV power forecasts are less accurate at every PV plant than the GHI forecasts at the closest meteorological station. This confirms that the main findings of Section 4.1 pertaining to the relative accuracies of different forecasts, which are not only valid for the four

selected locations but for the whole country in general.

As for the regional differences, the forecasts are less accurate in the northern and western parts of the country, where the geography is mountainous. The highest forecast accuracy is attained in the southeast, which is part of the Great Hungarian Plain. The lower forecast errors in the lowland areas of Hungary in comparison to the mountains have also been reported in [6]. Apart from this general trend, however, there are no large local outliers. To enhance the reliability of the verification results, the forecast errors are presented averaged for the whole country in the following subsections.

**Table 2**

Summary of error metrics for the country-wide verification. The metrics for the GHI and PV power are in  $W/m^2$  and  $W/kWp$ , respectively. ID (intraday) and DA (day-ahead) rows correspond to forecasts of 0–24-h-ahead and 24–48-h-ahead horizons.

	Model	Hor.	MBE	MAE	RMSE	nMBE	nMAE	nRMSE	Corr. coeff.	Var. ratio	Sqrt. type-1 CB	Sqrt. res.	Sqrt. type 2 CB	Sqrt. disc.
GHI	ECMWF	ID	− 3.0	76.4	112.7	0.8%	22.3%	32.9%	0.904	76.4%	16.5	238.0	58.6	208.5
		DA	0.4	81.6	120.6	0.2%	23.8%	35.1%	0.889	77.5%	15.1	234.1	60.5	206.5
	AROME	ID	− 27.6	85.3	128.4	8.0%	24.8%	37.4%	0.880	81.7%	32.7	231.7	63.3	209.6
		DA	− 25.3	89.2	133.5	7.3%	26.0%	38.9%	0.868	81.7%	32.0	228.7	65.1	206.7
PV, point	ECMWF	ID	23.7	82.4	131.7	7.2%	24.9%	39.7%	0.874	94.0%	37.2	228.8	51.6	222.4
		DA	28.2	90.3	144.1	8.5%	27.2%	43.5%	0.851	94.8%	45.7	222.8	58.6	217.7
	AROME	ID	− 6.3	88.8	143.5	1.9%	26.8%	43.3%	0.848	97.0%	41.0	222.3	48.3	218.9
		DA	− 5.0	94.6	151.7	1.5%	28.5%	45.8%	0.830	96.4%	44.4	217.5	53.6	213.7
PV, regional	ECMWF	ID	23.0	39.1	61.5	7.2%	12.2%	19.2%	0.971	104.8%	26.5	224.8	25.1	230.3
		DA	27.4	45.3	71.7	8.6%	14.2%	22.4%	0.961	106.1%	32.2	222.5	30.1	229.5
	AROME	ID	− 5.8	40.6	63.1	1.8%	12.7%	19.7%	0.963	101.2%	16.3	223.4	11.1	224.5
		DA	− 4.6	43.8	69.4	1.4%	13.7%	21.7%	0.956	101.8%	17.4	221.6	11.4	223.4

The country-wide averages of all error metrics are presented in Table 2. The rows denoted with “GHI” and “PV, point” contain the average of the error metrics calculated for each station and plant, respectively, and as such, they indicate the forecast accuracy for an average point location. However, in countries and regions with a strong grid where transmission congestion is not of concern, such as Hungary, the PV-related net imbalance that operating reserves must cover is the difference between aggregated forecast and production of all PV plants. To echo this fact, the PV power forecasts and measurements are summed for all 94 plants (then normalized back to a 1-kWp installed capacity, which technically turns summation into averaging), and the error metrics calculated for these forecasts are placed beside the “PV, regional” label.

#### 4.2.1. GHI verification

The average error metrics of Table 2 are quite similar to those shown in Table 1 for individual stations, therefore, the conclusions drawn in Section 4.1 are valid for the overall performance of the two NWP models over the entire country.

The MBE, MAE, and RMSE of the GHI forecasts as a function of the forecast lead time are shown in Fig. 6. The shape of the curves follows the diurnal periodicity of solar radiation with a peak around solar noon, when the GHI is also the highest on average. In terms of both MAE and RMSE, the ECMWF GHI forecasts are more accurate over all lead times than those of AROME. On the other hand, the AROME forecasts suffer from significant negative biases during most hours of a day, whereas the biases of the ECMWF forecasts are much smaller, with an asymmetric daily pattern with slightly positive values in the morning and negative

values in the afternoon. The error difference between the intra-day and day-ahead horizons (i.e., 0–24 and 24–48 h, respectively) is relatively small, i.e., less than 10%, which suggests that the accuracy of the NWP models only deteriorates mildly over the first two days. (This deterioration nevertheless continues to amplify over the next few days, which are not shown here, until a point where NWP forecasts are no better than climatology forecasts, which typically occurs beyond five days.)

The annual trend of the GHI forecast errors, with an 11-day moving average, is shown in Fig. 7. The AROME model is more accurate in January, but ECMWF takes the lead for the rest of the year. In most times of the year, the variations of the MAE and RMSE of the two NWP models follow a similar course, which means that the time periods when one model struggles with making accurate forecasts are also hard for the other model. In other words, the accuracy of GHI forecasts depends more on the weather conditions and other circumstances than the NWP model choice. The only significant exception took place near the end of February, where AROME has a huge negative bias, which translates to a surge in both the MAE and RMSE; however, the ECMWF forecasts are not outstandingly bad over that period of time. These errors are caused by four consecutive clear-sky days, when the sky conditions were properly predicted by the ECMWF, but for some reason, AROME predicted strong cloud presence for most of the country (one of these days, the 25th of February, is shown in Fig. 4). The AROME model is not able to outperform the ECMWF model in the summer months, which suggests that the supposedly better description of the convective cloud formation in summer has no added value in GHI forecasting at point locations.

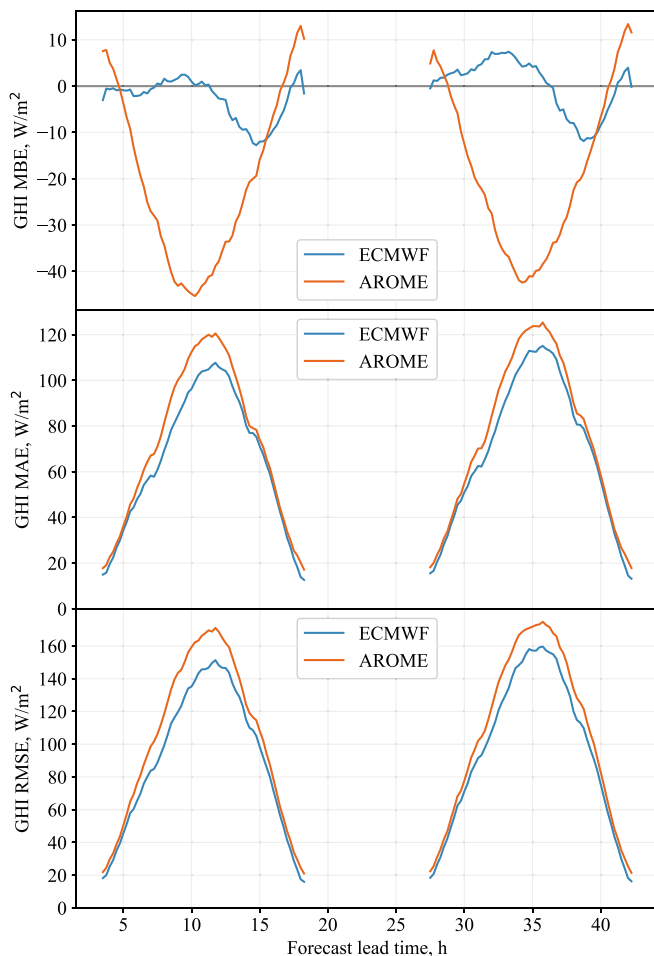


Fig. 6. MBE, MAE, and RMSE of GHI averaged for the 32 meteorological stations as a function of the lead time for the 00Z NWP runs.

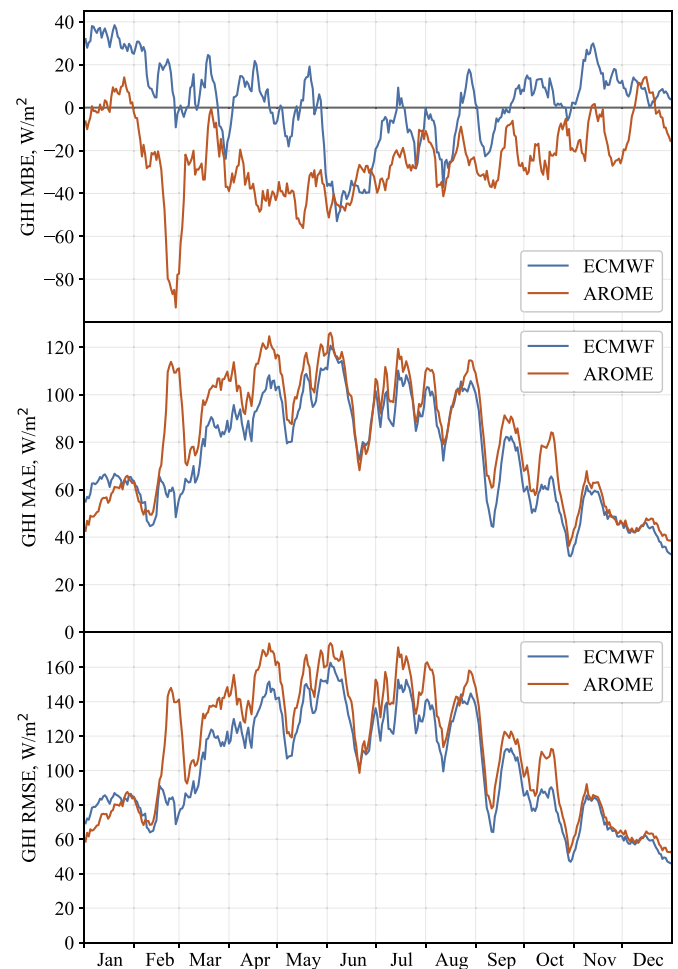


Fig. 7. 11-days moving average MBE, MAE, and RMSE of GHI averaged for the 32 meteorological stations over the year.

#### 4.2.2. PV power verification

The country-wide metric-based verification results for the point and regional PV power forecasts are summarized in Table 2. As is the case of the GHI, the trends for the PV power forecasts for the entire country are also similar to those found for the four sample locations in Section 4.1. The difference in the performance of the two NWP models is smaller for the PV power than for the GHI, which means that the added error of the irradiance-to-power conversion is bigger for the ECMWF than for the AROME. Such a difference is not surprising, considering that the main source of this added error is the amplification of the errors that already existed in the GHI forecasts and not the inaccuracy of the model chains [49]. The error amplification through the PV modeling depends on the design parameters of the PV plants, but for the most commonly used plant design, the errors of the GHI forecasts are amplified in the medium irradiance range and reduced for low and high irradiances [49]. Due to the smoother nature of the ECMWF forecasts, most of its errors probably fall into the medium irradiance range, which justifies the higher added errors of the PV power modeling. Overall, due to this effect, even though the ECMWF model is still more accurate than the AROME for PV power forecasting for point locations, its real advantage is smaller than that suggested by the GHI verification.

The MAE and RMSE of the regional PV power forecasts are less than half of the average MAE and RMSE of the point locations, which is due to the well-known geographical smoothing effect [54,55]. The error reduction due to geographical smoothing is stronger for the AROME forecasts than that for the ECMWF forecasts, which suggests that the former has a higher spatial variability, which clearly follows from its higher resolution. As a result, the initial advantage of ECMWF in GHI forecasting vanishes when it comes to regional PV power forecasting; as for this application, the performance of the forecasts based on both models is roughly the same. In contrast to both the GHI and point PV power forecasts, the regional PV power forecasts are overdispersed, as indicated by their variance ratios being higher than 100%. The explanation for this is that the spatial variability of the forecasts is smaller compared to the observations, therefore, the forecasts are less smoothed by the aggregation than the observations are.

PV power forecast errors as functions of the forecast lead time are shown in Fig. 8. The daily pattern of the bias in the PV forecasts is largely different from the bias of the GHI, where the originally almost unbiased ECMWF forecasts got a significant positive bias. The reason for this is the clipping of the PV power at the high irradiance range by the inverters, which eliminates the negative bias of the forecasts in this range and shifts the unconditional bias to the positive direction, as explained in Section 4.1. However, the bias of the PV power forecasts also depends largely on the model chain. Indeed, the choice of models in the different stages of a physical model chain can influence the overall bias by more than 20% [6], which highlights the importance of opting the most accurate models in all steps, or in a better case, using the most accurate combinations of models. Moreover, the accurate modeling of the losses and degradation is also crucial, as neglecting these factors can also introduce a positive bias. Having said that, even a perfect<sup>2</sup> model chain used on unbiased GHI data can introduce bias depending on the PV plant design parameters, due to the different error propagation in the different irradiance ranges [49]. As a joint consequence of the aforementioned factors, the bias in the GHI forecasts cannot be inferred from the bias of the PV power forecasts, and vice versa. Finally, a bias correction step after the irradiance-to-power conversion is always useful, for even unbiased GHI forecasts and the possible best model chain can lead to biased power forecasts.

The daily pattern of the PV power forecast errors, for both point

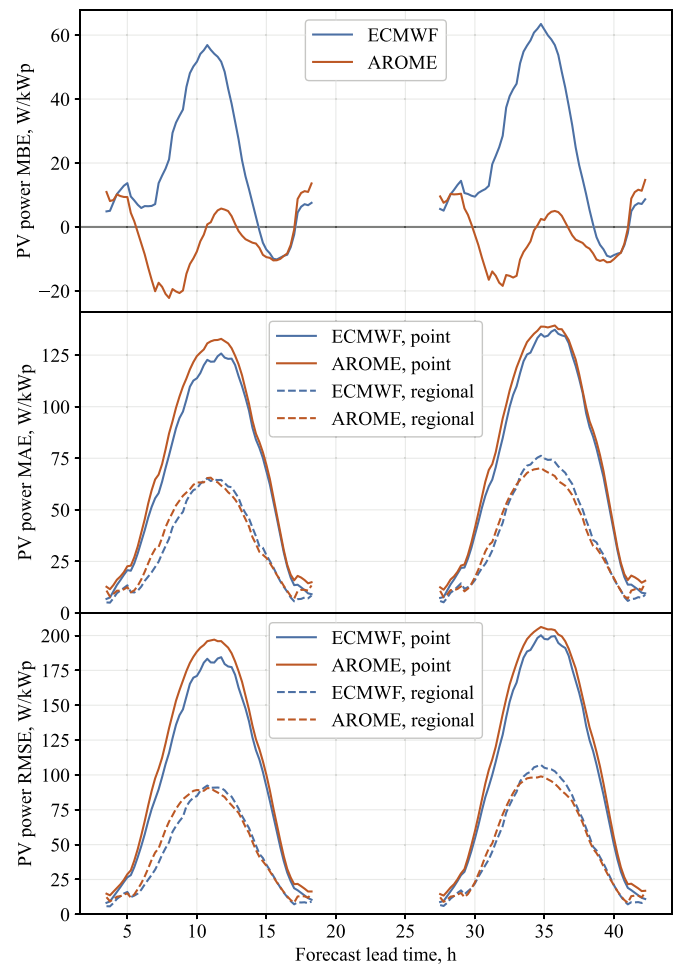


Fig. 8. MBE, MAE, and RMSE of PV power averaged for all PV plants (solid lines) and calculated from the aggregated forecasts and observations (dashed lines) as a function of the lead time for the 00Z NWP runs.

locations and the entire region, is similar to that of the GHI. In terms of MAE and RMSE, the ECMWF is better for point locations for all forecast lead times, but for regional forecasts, it is only better in the morning, whereas AROME takes the lead for the rest of the day, especially over the day-ahead time horizon.

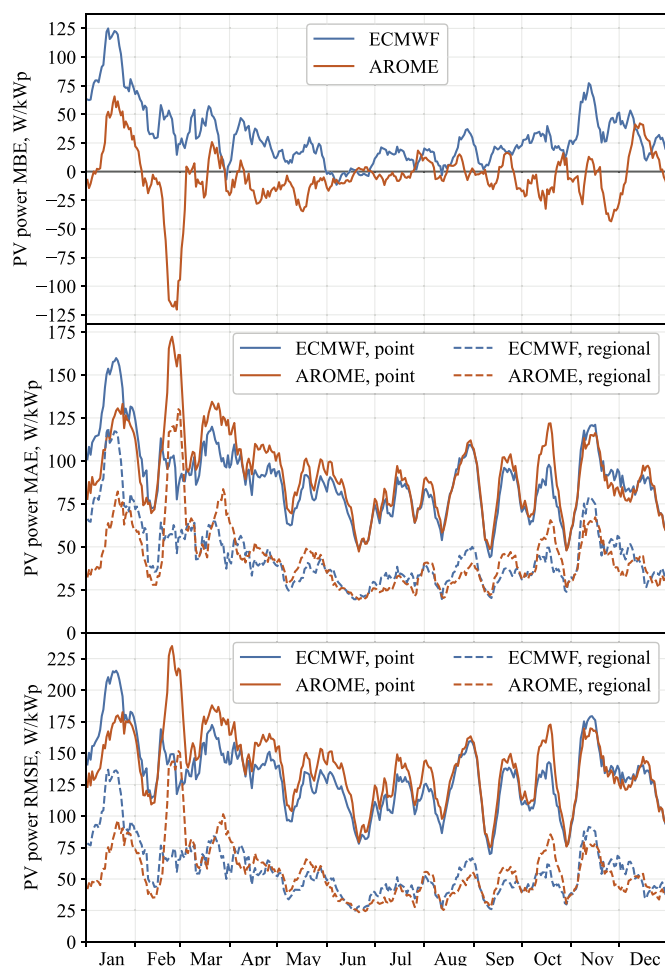
The 11-day moving average of the PV forecast errors can be seen in Fig. 9. Apart from the large AROME errors at the end of February, the error curves of both NWP models run close to each other, and especially for the regional forecasts, it often changes which model is better. Comparing Fig. 7 and Fig. 9, the main inter-annual trends of the GHI and PV power forecast errors are largely different. GHI forecast errors are the highest during summer, whereas PV power forecast errors are almost the same throughout the year, and they are even slightly larger in winter. This could be attributed to the tilt angle of the PV modules, which exaggerates both the irradiance and its errors during winter, see discussions in [49]. From the perspective of PV power forecasting, the inaccuracies in winter are actually more critical than it first seems from the GHI verification, which can also affect the proper fitting of post-processing routines.

## 5. Conclusion

This work has dealt with the verification of solar forecasts, which encompasses both solar irradiance (GHI) forecasts and PV power forecasts, in a typical setting concerning grid integration of solar energy, in which forecasts over the 0–48-h horizons are required. In particular, two

<sup>2</sup> In [49], both the observed and forecast PV power are modeled from GHI with the same model chain, therefore, the inaccuracies in the model chain did not contribute to the errors and the model chain could be considered perfectly accurate.





**Fig. 9.** 11-days moving average MBE, MAE, and RMSE of PV power averaged for all PV plants (solid lines) and calculated from the aggregated forecasts and observations (dashed lines) over the year.

NWP models, one global (ECMWF HRES) and one regional (AROME), which provide forecasts of weather variables that influence the production of PV power, are herein considered. GHI and PV power forecasts have been verified against the ground-based measurements from 32 research-grade radiometry stations and 94 actual PV plants, among which four station–plant pairs are closely located, serving as a basis for studying the error propagation from the initial irradiance forecasts to the eventual PV power forecasts. Verification has also been conducted for regional PV forecasts, which is most relevant to grid integration, where the overall fluctuations from a cluster of PV systems connected to the same power system node are of concern. In summary, the findings are numerous, among which the major ones are:

- The higher spatial and temporal resolutions of the AROME model allow it to better capture the variability of GHI, but ECMWF issues smoother GHI forecasts, which may be more desirable considering the “double penalty” problem.
- ECMWF is clearly more accurate than AROME for deterministic GHI forecasting for point locations, but its advantage reduces for PV power forecasting and completely disappears for regional forecasting.
- GHI forecasts are more accurate than PV power forecasts, therefore, the error metrics calculated for one or another, even in their normalized form, must not be compared directly.
- Bias and variance of the GHI forecasts cannot be judged based on the PV power due to the error propagation through the model chains,

since the sensitivity of PV forecasts errors to the GHI errors is different over different irradiance ranges [49].

- The GHI errors peak during summer, while PV power errors are higher in the winter. Based on GHI, summer seems to be the most critical period where the improvement of the model will yield the highest gain, but if one considers the end-use, which is PV power forecasting, it is more important to focus on accuracy improvement in winter.

#### CRedit authorship contribution statement

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

#### 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 AROME and ECMWF numerical weather prediction data. This paper was supported by the National Research, Development and Innovation Fund, project no. OTKA-FK 142702 and ÚNKP-22-5-BME-305, and by the Hungarian Academy of Sciences through the Sustainable Development and Technologies National Programme (FFT NP FTA) and the János Bolyai Research Scholarship.

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

#### 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, Kleissl J. Summarizing ensemble NWP forecasts for grid operators: Consistency, elicibility, and economic value. *Int J Forecast* 2023;39:1640–54. <https://doi.org/10.1016/j.ijforecast.2022.08.002>.
- [3] Bubnová R, Hello G, Bénard P, Geleyn J-F. Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Mon Weather Rev* 1995;123:515–35. [https://doi.org/10.1175/1520-0493\(1995\)123<0515:IOTFEE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<0515:IOTFEE>2.0.CO;2).
- [4] Smagorinsky J. General circulation experiments with the primitive equations. *Mon Weather Rev* 1963;91:99–164. [https://doi.org/10.1175/1520-0493\(1963\)091<0099:GCEWTP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2).
- [5] Bougeault P. A simple parameterization of the large-scale effects of cumulus convection. *Mon Weather Rev* 1985;113:2108–21. [https://doi.org/10.1175/1520-0493\(1985\)113<2108:ASPOTL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<2108:ASPOTL>2.0.CO;2).
- [6] 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>.
- [7] 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>.
- [8] 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>.

- [9] Davies HC. A lateral boundary formulation for multi-level prediction models. *Q J Roy Meteorol Soc* 1976;102:405–18. <https://doi.org/10.1002/qj.49710243210>.
- [10] Brousseau P, Seity Y, Ricard D, L  ger J. Improvement of the forecast of convective activity from the AROME-France system. *Q J Roy Meteorol Soc* 2016;142:2231–43. <https://doi.org/10.1002/qj.2822>.
- [11] Ebert E, Wilson L, Weigel A, Mittermaier M, Nurmi P, Gill P, et al. Progress and challenges in forecast verification. *Meteorol Appl* 2013;20:130–9. <https://doi.org/10.1002/met.1392>.
- [12] Perez R, Lorenz E, Pelland S, Beauharnois M, Van Knowe G, Hemker K, et al. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Sol Energy* 2013;94:305–26. <https://doi.org/10.1016/j.solener.2013.05.005>.
- [13] 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>.
- [14] Bakker K, Whan K, Knap W, Schmeits M. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. *Sol Energy* 2019;191: 138–50. <https://doi.org/10.1016/J.SOLENER.2019.08.044>.
- [15] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. *Renew Sustain Energy Rev* 2022;154:111768. <https://doi.org/10.1016/j.rser.2021.111768>.
- [16] Hogan RJ, Bozzo A. A flexible and efficient radiation scheme for the ECMWF model. *J Adv Model Earth Syst* 2018;10:1990–2008. <https://doi.org/10.1029/2018MS001364>.
- [17] Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J Geophys Res Atmos* 1997;102:16663–82. <https://doi.org/10.1029/97JD00237>.
- [18] Tiedtke M. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon Weather Rev* 1989;117:1779–800. [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSP>2.0.CO;2).
- [19] Bechtold P, K  hler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, et al. Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Q J Roy Meteorol Soc* 2008;134:1337–51. <https://doi.org/10.1002/qj.289>.
- [20] Bechtold P, Semane N, Lopez P, Chaboureaud J-P, Beljaars A, Bormann N. Representing equilibrium and nonequilibrium convection in large-scale models. *J Atmos Sci* 2014;71:734–53. <https://doi.org/10.1175/JAS-D-13-0163.1>.
- [21] Forbes Richard, Tompkins A. An improved representation of cloud and precipitation. *ECMWF Newsl* 2011;129. <https://doi.org/10.21957/nfgulzhe>. 13–8.
- [22] K  hler M, Ahlgrim M, Beljaars A. Unified treatment of dry convective and stratocumulus-topped boundary layers in the ECMWF model. *Q J Roy Meteorol Soc* 2011;137:43–57. <https://doi.org/10.1002/qj.713>.
- [23] Balsamo G, Beljaars A, Scipal K, Viterbo P, van den Hurk B, Hirschi M, et al. A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *J Hydrometeorol* 2009; 10:623–43. <https://doi.org/10.1175/2008JHM1068.1>.
- [24] Seity Y, Brousseau P, Malardel S, Hello G, B  nard P, Bouttier F, et al. The AROME-France convective-scale operational model. *Mon Weather Rev* 2011;139:976–91. <https://doi.org/10.1175/2010MWR3425.1>.
- [25] Szintai B, Sz  cs M, Randriamampianina R, Kullmann L. Application of the AROME non-hydrostatic model at the Hungarian Meteorological Service: Physical parameterizations and ensemble forecasting. *Id  j  r  s* 2015;119:241–65.
- [26] T  th H, Homonnai V, Mile M, V  rkonyi A, Kocsis Z, Szanyi K, et al. Recent developments in the data assimilation of AROME/HU numerical weather prediction model. *Id  j  r  s* 2021;125:521–53. <https://doi.org/10.28974/idojaras.2021.4.1>.
- [27] Fouquart Y, Bonnel B. Computations of solar heating of the earth's atmosphere—A new parameterization. *Beitrag Zur Phys Der Atmosphere* 1980;53:35–62.
- [28] Li D, Shine KP. A 4-dimensional ozone climatology for UGAMP models 1995. *UGAMP Internal Rep* 2023;35.
- [29] Bouteloup Y, Toth H. Refinements in the parameterisation of radiative exchanges. *ALADIN Newsl* 2003;23:178–83.
- [30] Tegen I, Hollrig P, Chin M, Fung I, Jacob D, Penner J. Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *J Geophys Res Atmos* 1997;102:23895–915. <https://doi.org/10.1029/97JD01864>.
- [31] Pinty J-P, Jabouille P. A mixed-phased cloud parameterization for use in a mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitation. In: *Proc. Conf. Cloud Phys.* Everett, WA: American Meteorological Society; 1998. p. 217–20.
- [32] Cuxart J, Bougeault P, Redelsperger J-L. A turbulence scheme allowing for mesoscale and large-eddy simulations. *Q J Roy Meteorol Soc* 2000;126:1–30. <https://doi.org/10.1002/qj.49712656202>.
- [33] Pergaud J, Masson V, Malardel S, Couvreux F. A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Bound Layer Meteorol* 2009;132:83–106. <https://doi.org/10.1007/s10546-009-9388-0>.
- [34] Le Moigne P. SURFEX scientific documentation. M  t  o-France, CNRM, Toulouse, France: Groupe de M  t  orologie a Moyenne Echelle; 2012.
- [35] van der Plas E, Schmeits M, Hooijman N, Kok K. A comparative verification of high-resolution precipitation forecasts using model output statistics. *Mon Weather Rev* 2017;145:4037–54. <https://doi.org/10.1175/MWR-D-16-0256.1>.
- [36] Lefevre 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>.
- [37] 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:AGFFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFV>2.0.CO;2).
- [38] 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>.
- [39] 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>.
- [40] 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>.
- [41] Yang D. Temporal-resolution cascade model for separation of 1-min beam and diffuse irradiance. *J Renew Sustain Energy* 2021;13:056101. <https://doi.org/10.1063/5.0067997>.
- [42] 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 Sustain Energy Rev* 2022;159:112195. <https://doi.org/10.1016/j.rser.2022.112195>.
- [43] 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](https://doi.org/10.1016/0038-092X(90)90055-H).
- [44] Reda I. Solar radiation on inclined surfaces: Corrections and benchmarks. *Sol Energy* 2016;136:288–302. <https://doi.org/10.1016/j.solener.2016.06.062>.
- [45] Sjerps-Koomen EA, Alsema EA, Turkenburg WC. A simple model for PV module reflection losses under field conditions. *Sol Energy* 1996;57:421–32. [https://doi.org/10.1016/S0038-092X\(96\)00137-5](https://doi.org/10.1016/S0038-092X(96)00137-5).
- [46] 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>.
- [47] 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>.
- [48] 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 application in a procedure for the performance check of grid connected systems. In: *Proc. 19th Eur. Photovoltaic Sol. Energy Conf*; 2004. p. 5.
- [49] 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>.
- [50] Mayer MJ, Gr  f 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>.
- [51] 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>.
- [52] Maor T, Appelbaum J. View factors of photovoltaic collector systems. *Sol Energy* 2012;86:1701–8. <https://doi.org/10.1016/j.solener.2012.03.017>.
- [53] Driesse A, Jain P, Harrison S. Beyond the curves: Modeling the electrical efficiency of photovoltaic inverters. In: *2008 33rd IEEE Photovoltaic Spec. Conf. IEEE*; 2008. p. 1–6. <https://doi.org/10.1109/PVSC.2008.4922827>.
- [54] Lave M, Kleissl J, Arias-Castro E. High-frequency irradiance fluctuations and geographic smoothing. *Sol Energy* 2012;86:2190–9. <https://doi.org/10.1016/J.SOLENER.2011.06.031>.
- [55] Marcos J, Marroyo L, Lorenzo E, Garc  a M. Smoothing of PV power fluctuations by geographical dispersion. *Prog Photovolt Res Appl* 2012;20:226–37. <https://doi.org/10.1002/PIP.1127>.