Marco Pierro
Institute for Renewable Energy,
EURAC Research,
Bolzano 39100, Italy;
Department of Enterprise Engineering,
University of Rome Tor Vergata,
Rome 00133, Italy
e-mail: marco.pierro@gmail.com

Francesco Bucci
Department of Enterprise Engineering,
University of Rome Tor Vergata,
Rome 00133, Italy

Matteo De Felice
ENEA,
Casaccia R.C.,
Climate Impacts and Modelling Laboratory,
Rome 00123, Italy

Enrico Maggioni
IDEAM S.r.l.,
Cinisello Balsamo 20092, Italy

Alessandro Perotto
IDEAM S.r.l.,
Cinisello Balsamo 20092, Italy

Francesco Spada
IDEAM S.r.l.,
Cinisello Balsamo 20092,Italy

David Moser
Institute for Renewable Energy,
EURAC Research,
Bolzano 39100, Italy

Cristina Cornaro
CHOSE,
Department of Enterprise Engineering,
University of Rome Tor Vergata,
Rome 00133, Italy

Deterministic and Stochastic Approaches for Day-Ahead Solar Power Forecasting

Photovoltaic (PV) power forecasting has the potential to mitigate some of effects of resource variability caused by high solar power penetration into the electricity grid. Two main methods are currently used for PV power generation forecast: (i) a deterministic approach that uses physics-based models requiring detailed PV plant information and (ii) a data-driven approach based on statistical or stochastic machine learning techniques needing historical power measurements. The main goal of this work is to analyze the accuracy of these different approaches. Deterministic and stochastic models for day-ahead PV generation forecast were developed, and a detailed error analysis was performed. Four years of site measurements were used to train and test the models. Numerical weather prediction (NWP) data generated by the weather research and forecasting (WRF) model were used as input. Additionally, a new parameter, the clear sky performance index, is defined. This index is equivalent to the clear sky index for PV power generation forecast, and it is here used in conjunction to the stochastic and persistence models. The stochastic model not only was able to correct NWP bias errors but it also provided a better irradiance transposition on the PV plane. The deterministic and stochastic models yield day-ahead forecast skills with respect to persistence of 35% and 39%, respectively.
[DOI: 10.1115/1.4034823]

1 Introduction

IEA reports show that in the last decade, the cumulative installed capacity of photovoltaic has grown at an average rate of 49% per year reaching by the end of 2014 a worldwide installed capacity of 177 GW. In 19 countries, the annual PV contribution to electricity demand was estimated to exceed the 1% mark, with Italy leading with at least 7.9% followed by Greece at 7.6% and Germany at 7%. Different IEA scenarios have predicted for 2050 a PV penetration between 6% and 11% of the world electric consumption [1,2].

Electricity load can be affected by high PV generation, introducing a stochastic variability dependent on the meteorological conditions [3]. In particular, on the daily time scale, the PV production increases load ramps so that a greater secondary reserve and ready supply is needed. This, for example, in case of domestic load, is accentuated in the evenings when the rapid reduction of large amounts of PV power is added to an increase in electric consumption.

Thus, the large share of PV power introduces new challenges for the stability of the electrical grid, both at the local and national level, requiring more reserves to ensure electrical balancing and overcome the unpredictability and variability of the electricity demand. Moreover, it implies an increase in costs related to transactions on the day-ahead and intraday energy market and dispatching operations on the real-time energy market. Despite the challenges, grids can sustain high penetration of distributed power generation provided that quality of supply is addressed at connection point through the capabilities of modern power electronics, distributed control, and the use of ancillary services.

The PV power forecast could mitigate the effects of high solar power injection into the electricity grid on grid management and on the energy market. The short-term forecast (intrahour) could be used to predict the power ramps and voltage flickers as well as to better control the operations on the real-time market and dispatching management. The midterm forecasts (intraday and day-ahead) could be used, on the one hand, for load following to control voltage and frequency instability and for transmission scheduling to reduce the secondary reserve. On the other hand, it could be

Contributed by the Solar Energy Division of ASME for publication in the Journal of Solar Energy Engineering: Including Wind Energy and Building Energy Conservation. Manuscript received February 5, 2016; final manuscript received September 6, 2016; published online November 30, 2016. Assoc. Editor: Carlos F. M. Coimbra.

employed for a better match between the intraday and day-ahead market commitment and the real PV production, thus reducing the energy unbalancing costs. For these reasons, the site and regional day-ahead forecast of the solar power generated by large PV producers and distribution system operators is now mandatory in many European and non-European countries (Italy, Germany, Spain, Romania, U.S., Japan, etc.), and consequently, increasing the forecast accuracy becomes more and more important.

For day-ahead forecasts (24–72 h horizon), the numerical weather prediction (NWP) data should be employed to obtain an acceptable accuracy level, as for intrahour and intraday forecast, the use of full sky images and satellite data is essential. The NWP data are generated by global or mesoscale simulation models able to provide the numerical integration of the coupled differential equations describing the dynamics of the atmosphere and radiation transport mechanisms [4]. Moreover, these data are usually corrected by postprocessing algorithms called model output statistics (MOS) that use past ground measurements to partially remove the systematic errors [5–7].

The PV power generation forecast on the horizon of 24–72 h can be achieved through deterministic or data-driven approaches. Figure 1 presents a visual summary of the two approaches.

The first approach is based on physical or semi-empirical models. It uses a transposition method to project the global horizontal irradiance (GHI) predicted by NWP models (eventually corrected by MOS post processing) on the PV tilted surface. In a second step, the predicted global plane of array irradiance (GPOAI) together with other NWP variables is used in a power estimation model to forecast the PV generation. Deterministic models using different NWP data, transposition, and power estimation methods were developed by Lorenz et al. [8,9] and Pelland et al. [10], and were used both for site and regional PV power forecast.

The second approach involves a wide range of machine learning models that can be built using past measurements. By training on historical datasets, these statistic or stochastic models try to reconstruct the relationship between input and output data. They are based on data-driven algorithms that do not require knowledge of the physical laws describing the phenomenon. These algorithms can use a variety of NWP variables to directly provide the PV generation forecast. The hybrid models could be obtained using different models in series.

A variety of machine-learning models using different techniques were developed by various authors. For site day-ahead power forecasts, Yona et al. [11] implemented models based on a multilayer perceptron (MLPNN), radial basis function (RBFNN), and recurrent neural networks (RNN). Chen et al. [12] and Tao et al. [13] used the RBFNN and nonlinear autoregressive exogenous NN (NARX) while Wang et al. [14] coupled the MLPNN with the gray model (GM). Mellit et al. [15] introduced the adaptive feedforward backpropagation network (AFFNN) for the short-term forecasting. Larson et al. [16] make use of a multilinear regression model.

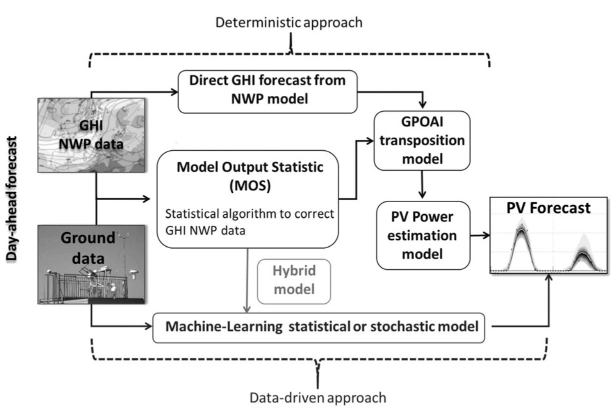

Fig. 1 Schematics of the different approaches for day-ahead PV power forecast

Bacher et al. [17] developed autoregressive with and without exogenous input (AR/ARX) models for intraday and day-ahead power forecast. Mellit and Pavan [18] and Li et al. [19] tested two power forecast models based on the MLPNN and ARMAX algorithms, which make use only of ground measurement data. For site and regional forecast, Zamo et al. [20] compared the accuracy of many different techniques: bagging, random forest, boosting, support vector machine (SVM), and generalized additive model (GAM), concluding that the random forest was the best performing model. For regional forecast, da Silva Fonseca et al. [21] explored an interesting technique based on support vector regression (SVR) coupled with a principal component analysis (PCA) preprocessing, while in Ref. [22], the same authors developed and tested different models based on the SVR. Finally in more recent years, a probabilistic approach extensively employed in wind forecasting has started to be adopted also for PV power predictions. This method is focused on informing about the distribution of potential events through a set of conditional probability density functions [23,24] or ensemble of alternative forecast trajectories obtained through deterministic or data-driven models [25,26]. An overview on PV power forecast techniques can be found in Refs. [27–29].

The deterministic approach does not require past power measurements whereas detailed information on the PV plant setup (geographic position, electrical configuration, etc.) is needed. On the contrary, for the data-driven stochastic approach, power measurements are essential while none or very little plant information is needed. Which of the two is the outperforming method is still under investigation and it is the main goal of this work. Indeed, the aim of the paper is not only to develop and test two original deterministic and stochastic models for site day-ahead PV generation forecast but also to better understand the main sources of error of the different approaches. Moreover, in the paper, a new parameter called "clear sky performance" index was defined. It was used both to develop the stochastic model and to set up a better performing persistence model. This parameter could be considered the equivalent of the clear sky index for the PV power generation.

The deterministic model is composed by a MOS correcting the weather research and forecasting model (WRF) irradiance prediction, an isotropic transposition model, and a high performing power estimation model. The stochastic model uses a probabilistic approach applied to a qualified ensemble of 300 artificial neural network predictions. Four years of monitored weather and production data from a 662 kWp Cadmium Telluride PV plant, located in Bolzano (Italy), were used to train and test the models. First, a preliminary assessment of the deterministic model was developed to assess the quality of the NWP prediction of the global horizontal irradiance and the accuracy on actual data of the used transposition and power estimation methods. Then, the accuracy of the two models together with the persistence model (used as benchmark) was compared and analyzed. Finally, a detailed analysis of the forecast errors was performed. In particular, the error propagation inside model chain used in the deterministic approach was studied. Moreover, the impact of the nominal power degradation of the CdTe modules on the power estimation and forecast accuracy was also evaluated. Then, the main improvements of the stochastic approach with respect to the deterministic one were pointed out.

This paper could be considered complementary to Ref. [10], where the impact of different transposition and power estimation methods on the deterministic model accuracy was studied.

2 Data

2.1 Actual Data (Ground Measurements). Four years of experimental data were used to train and test the forecast models:

from Jan. 1, 2011 to Dec. 31, 2014. Measurements were performed in the outdoor test facility of Airport Bolzano Dolomiti (ABD) (position ca. 46.46 N, 11.33 E, elevation: 262 m) located in South Tyrol, Italy [30]. The power generation measurements are the hourly average production of a 662 kWp commercial PV plant with cadmium telluride thin film technology. The modules are oriented 8.5 deg west of south and 30 deg tilted. The DC and AC side electrical parameters are measured every 15 min by commercial inverters that assure a good level of accuracy in current ($I_{\rm mpp}$) and voltage ($V_{\rm mpp}$), with an average difference from a dedicated system of less than 5% and less than 2%, respectively, further decreasing at higher irradiance [31]. Moreover, an independent meteorological station provides measurements of global horizontal and diffuse irradiance, global plane of the array irradiance, wind speed, air temperature, and back of the module temperature with a frequency of 1 min, which is then averaged on a 15 min time interval. The sensors are systematically cleaned and periodically calibrated in order to comply with the standard IEC61724:1998 (1998).

Figure 2(*a*) shows the time series of daily reference yield and final yield in the considered period. It should be remarked that irradiance measurements started from Feb. 6, 2011. Figure 2(*b*) reports the monthly average of daily power yield for the considered four years. A large monthly variability of power generation between years due to the different amount of clear sky days is clearly visible.

2.2 Numerical Weather Prediction Data. The NWP model used is weather research and forecasting (WRF–ARW) version 3.6.1 developed by National Center of Atmospheric Research (NCAR). The model is run operationally by the U.S. National Weather Service and, being open-source and easily portable, it is widely used around the world for research and weather forecasts [32]. The details of used WRF physics configuration are reported in the Appendix.

Daily hindcasts were performed for the period considered (2011–2014). The model was initialized at 12 UTC, analyzing the 24 h forecasts starting from the following 00 UTC, which is the typical procedure for the NWP solar day-ahead forecast. The model domain is centered over Italy with a horizontal resolution of 12 km, a higher resolution inner domain is nested centered on the region of interest, with a horizontal resolution of approximately 3 km. This horizontal resolution was necessary because of the complex orography of the region. For the nested domain, a physics configuration suitable for operational forecasting was made, balancing accuracy in the results with computational efficiency (Table 1).,

3 Clear Sky Performance Index

The clear sky index (Eq. (1)) is extensively used in literature on solar irradiance forecast to normalize the measured solar radiation to its theoretical value under perfectly transparent atmosphere

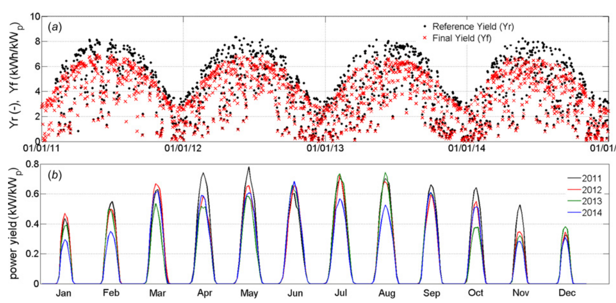

Fig. 2 (*a*) Daily reference yield (Yr) and final yield (Yf) from Jan. 1, 2011 to Dec. 31, 2014 and (*b*) monthly average of daily power yield for all the considered years

Table 1 WRF physics configuration

Microphysics	Eta microphysics scheme [33]
Surface layer	MM5 similarity scheme [34]
Surface	Noah surface model [35]
Boundary layer	YSU [36]
Cumulus scheme	Kain–Fritsch [37]

$$K_{\rm cs} = {\rm GHI}/{\rm GHI}_{\rm cs} \tag{1}$$

where $\rm GHI_{cs}$ is the global horizontal irradiance estimated by a clear sky model. On the other hand, in the photovoltaic sector, the performance ratio (Eq. (2)) is the main index used to evaluate the plant performance

$$\rm PR = \frac{P_m/P_n}{GPOAI/G_0} \tag{2}$$

where P_m is the power output, P_n is the plant nominal power, GPOAI is the global irradiance on the plane of array, and $G_0 = 1000\,{\rm W/m^2}$ is the reference irradiance. It measures the ratio of the real power production for unit of installed peak power (P_m/P_n) and the expected generation with an incident irradiance of GPOAI. Thus, it evaluates the plant performance and relative losses in real operating conditions.

In this paper, a new parameter that could be considered the equivalent of the clear sky index for the PV power generation is defined. It is called "clear sky performance index" ($PK_{\rm cs}$), and it combines the two above-mentioned parameters

$$PK_{\rm cs} = \frac{P_m/P_n}{{\rm GPOAI_{cs}}/G_0} \tag{3}$$

where $\rm GPOAI_{cs}$ is the plane of array clear sky irradiance. As for the irradiance forecast, the clear sky performance index could be used both to develop forecast models and to characterize sky conditions (when no irradiance data are available).

In perfect clear sky conditions, $PK_{\rm cs}$ is equal to PR. On the contrary, in very overcast, $PK_{\rm cs}$ is similar to $K_{\rm cs}$ evaluated on the plane of array.

Figure 3 shows an example of the daily trend of the above-mentioned indexes. In could be observed that the $PK_{\rm cs}$ behaves like PR in clear sky conditions (Aug. 16, 2011) and like $K_{\rm cs}$ in overcast conditions (Sept. 18, 2011).

$PK_{\rm cs}$ can be used as $K_{\rm cs}$ to characterize the sky conditions: overcast days with $PK_{\rm cs} < 0.4$, partially cloudy with $0.4 < PK_{\rm cs} < 0.8$, and clear sky with $PK_{\rm cs} > 0.8$. Nevertheless, the threshold values (0.4 and 0.8) are more site dependent, since the $PK_{\rm cs}$ index depends not only on the used clear sky model but also on the PV plane position and PV modules technology. Thus, it is better to tune these parameters on the specific site.

4 Description of Forecast Models

In this section, we provide a description of all the PV production forecast models developed in this paper and later applied (see Sec. 6) on a specific site. Both deterministic and stochastic approaches are used here. The first (described in Sec. 4.2) is a method based on a chain of three different semi-empirical models derived from physical considerations. The second (described in Sec. 4.3) is a data-driven method based on an ensemble of artificial neural networks. Finally, persistence models used as benchmark are described in Sec. 4.1.

4.1 Persistent Benchmark Models. This is a trivial model that assumes that the weather conditions on the present day and

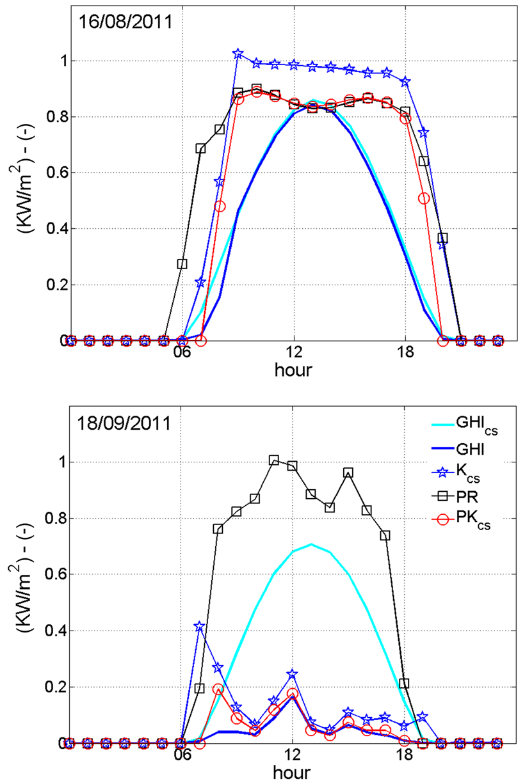

Fig. 3 Daily trend of clear sky index (K_{cs}), performance ratio (PR), and clear sky performance index (PK_{cs}) in a clear sky day and in an overcast day

on the day of the forecast are identical. There are several persistence models used in day-ahead solar irradiance predictions that lead to different root-mean-square error (RMSE) [38]. The "naive" persistence model (PM) assumes the persistence of global horizontal irradiance while other more accurate models (KPM) presume the persistence of the clear sky index (clear sky persistence models). In the literature, regarding PV generation forecast, only the simple persistence is used

$$P_m^{\mathrm{PM}}(h+H) = P_m(h) \tag{4}$$

where h is the hour of the day, H is the forecast horizon (in this case 24 h), $P_m^{\mathrm{PM}}(h+H)$ is the power output predicted by the persistence model for the time $h+H$, and $P_m(h)$ is the actual power generated at time h. In this paper, a new model based on the persistence of the daily PK_{cs} index is defined

$$P_m^{\mathrm{KPM}}(h+H) = PK_{\mathrm{cs}}(dd)\mathrm{GPOAI}_{\mathrm{cs}}(h+H) \tag{5}$$

where $P_m^{\mathrm{KPM}}(h+H)$ is the power output predicted by the new persistence model for the time $h+H$, $\mathrm{GPOAI}_{\mathrm{cs}}(h+H)$ is the clear sky global plane of array irradiance provided by the WRF for the time $h+H$, and $PK_{\mathrm{cs}}(dd)$ is the daily clear sky performance index of the actual day dd

$$PK_{\mathrm{cs}}(dd) = \frac{\sum\limits_{h=1:24} P_m(h)/P_n}{\sum\limits_{h=1:24} \mathrm{GPOAI}_{\mathrm{cs}}(h)/G_0} \tag{6}$$

This model could be considered the equivalent of the irradiance clear sky persistence for PV generation forecast, and it could be adopted as a benchmark model.

4.2 Deterministic Model: Model Output Statistics for Relative Humidity (MOSRH) + Isotropic Transposition Model (IM) + Sandia PV Array Performance Model (SAPM). The deterministic approach follows the chain of three different models:

(1) the model output statistic (MOSRH), which provides a corrected GHI forecast, through physical postprocessing of WRF irradiance output
(2) the Liu–Jordan isotropic transposition model (IM) that projects the predicted GHI on the plane of array, providing the GPOAI forecast
(3) the Sandia PV array performance model (SAPM) that converts the GPOAI prediction into PV power generation

(1) A first version of the MOSRH postprocessing algorithm can be found in Ref. [7]. The final version of this algorithm is here described. The idea behind MOSRH is based on the fact that GHI experiences a reduction at every atmospheric layer with a non-negligible value of water molecules in the liquid phase. This situation occurs when relative humidity is near or at the saturation level. On the contrary, the majority of WRF's built-in radiation schemes behave like a step function in damping GHI, since GHI experiences a considerable damping only if one or more vertical model levels present a water vapor content at the saturation level. This produces a considerable error in the irradiance prediction. The MOSRH does not only take into consideration vertical levels where relative humidity saturates but also all levels where the relative humidity is greater than a predefined threshold. The GHI damping increases linearly for the levels where the relative humidity is higher than its threshold to reach complete absorption only when the relative humidity reaches 100%. Subsequently, the impact of damping is weighed accordingly to the quantity of humidity in the single vertical level, now independently from their altitude. The weights decrease exponentially with decreasing humidity and are derived from the application of the Beer–Lambert law [39]. Subsequently, the weights are normalized in order to obtain a value between 0 and 100, similar to the calculation of the total cloud cover percentage. The result was named pseudocloud cover (PCC)

$$\mathrm{PCC} = \sum_j (\mathrm{RH}_j w_j) / \sum_j w_j \tag{7}$$

where RH_j is the relative humidity value of level j and the sum is calculated from the bottom to the top of the atmosphere; w_j is the weight of the level j and is equal to zero for the levels below their specific threshold. Finally, in this MOSRH version, the resulting corrected GHI at the lowest level of the atmosphere is given by the following equation:

$$\mathrm{GHI}_{\mathrm{for}} = d\mathrm{GHI}_{\mathrm{cs}}(1 - a\mathrm{PCC}^b) + c \tag{8}$$

where $\mathrm{GHI}_{\mathrm{cs}}$ is the clear sky GHI at the lowest level of the atmosphere predicted by the RRTMG radiation scheme; a, b, c, and d are regression coefficients obtained by fitting observational data. To better fit the observational data, three sets of coefficients that exclude one another are used based on the forecast meteorological condition. The clear sky conditions are defined by $\mathrm{PCC} < 0.05$, partly cloudy conditions by $0.05 < \mathrm{PCC} < 0.7$, and overcast conditions by $\mathrm{PCC} > 0.7$. A similar regression for the sky cover predictions from the National Digital Forecast Database was found by Perez et al. [5]. The values of the coefficients a, b, c, and d for Bolzano are shown in Table 2.

(2) The Liu–Jordan isotropic model (IM) [40] is one of the most used transposition methods adopted in forecast applications [8,10], since it requires a small number of parameters and forecast data. There are many different transposition models; nevertheless, Pelland et al. [10] proved that the choice of transposition model has little impact on the PV forecast accuracy.

Table 2 Regression coefficients for Bolzano

Meteorological conditions	a	b	c	d
Clear sky (PCC < 0.05)	0	1	22	0.96
Partly cloudy (0.05 < PCC < 0.7)	0.36	0.88	0	1
Overcast (PCC > 0.7)	0.68	2.33	0	1

Table 3 Accuracy metrics where P_m^{obs} = observed hourly PV power generation (kW), P_m^{for} = forecast hourly PV power generation (kW), P_n = plant nominal power (or plant capacity) (kWp), and n = number of sun hours in the considered period

Name	Acronym and formulas
Forecast error	$e_h = (P_m^{\text{for}} - P_m^{\text{obs}})/P_n$
Root-mean-square error	$\text{RMSE} = \sqrt{\frac{\sum_{h=1}^{n} e_h^2}{n}}$
Mean absolute error	$\text{MAE} = \frac{\sum_{h=1}^{n} \lvert e_h \rvert}{n}$
Mean bias error	$\text{MBE} = \frac{\sum_{h=1}^{n} e_h}{n}$
RMSE skill score	$\text{SS} = 100\left(\frac{\text{RMSE}_{\text{ref}} - \text{RMSE}_{\text{test}}}{\text{RMSE}_{\text{ref}}}\right)$
Energy imbalance	$\text{EI} = \sum_{h=1}^{n} \lvert e_h \rvert$

(3) The Sandia PV array performance model (SAPM) was developed at the Sandia Laboratory [41], and it is one of the most accurate semi-empirical models used to estimate the PV power generation in real operating conditions. The peculiarity of SAPM with respect to other semi empirical models [8–10,42] is the estimation of the module optical losses through the modeling of the effective irradiance. Moreover in this paper, a multiplicative yearly degradation factor ($\Delta Pn = P_{m0}/P_n$) was introduced to take into account the reduction of the plant peak power (P_{m0}) with respect to the nominal power (P_n). Similar technique was used in Ref. [42] for better module characterization. Thus, the present version of SAPM is a very complete power estimation model that takes into account all the main effects that modify the PV plant performance in real operating conditions: spectral and reflection, temperature and irradiance effects, and power degradation.

4.3 Stochastic Model: Relative Humidity Neural Network Model (RHNN). The RHNN is a stochastic model based on a qualified ensemble of artificial neural networks (ANNs) developed by the ESTER laboratory of the University of Rome "Tor Vergata." The ANN is a mathematical model that invokes the structure of biological neural connections [43]. The model consists of a group of neurons, and it processes information using a connectionist approach to computation. This technique is often employed in solving forecasting problems; an extensive review can be found in Ref. [44]. Concerning solar radiation forecast using ANN, a review can be found in Ref. [45]. Several neural networks architectures exist and have been studied; in this paper, the multilayer perceptron (MLPNN) was adopted. The MLPNN has the ability to imitate natural intelligence in its learning from existing sample data, so that the algorithm learns from sample data by constructing input–output connections. The connections are depicted by the weights matrix (**W**) that mimics the strength of the synapse connections between neurons, and the bias vector (**b**) that stands for the neurons activation threshold. In an MLPNN with one hidden layer, the relation between input stimuli (**X**) and the neurons activities (**Y**) is modeled as follows:

$$\mathbf{Y} = f^{(2)}(\mathbf{W}^{(2)} f^{(1)}(\mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)}) \tag{9}$$

where (i = 1, 2) is the layer index and $f^{(i)}$ are transfer functions modeling the intensity of neurons activities. The parameters $\mathbf{W}^{(i)}$ and $\mathbf{b}^{(i)}$ are empirically derived by a training and validation procedure, minimizing the error between the input and the output of a known set of data. The MLPNN is a nonlinear semi-empirical function dependent on a large number of parameters. A stochastic component is introduced by a random choice of the initial condition of the minimization procedure and by a random partition of the training data into training and validation sets. The MLPNN is typically used to model complex relationships between inputs and outputs. Since it establishes a mapping between the input stimuli (**X**) and the output signal (**Y**), it can be simply expressed as

$$\mathbf{Y} = f_{NN}(\mathbf{X}) \tag{10}$$

It has to be stressed that the performance of the MLPNN is strongly dependent on the internal structure of the network, on the variable chosen as input and on the method used for the training. For the RHNN model, an MLPNN with two layers was adopted (see Fig. 4), and a hyperbolic tangent and linear functions were chosen as transfer functions between the layers. The model provides the forecast of the clear sky performance index (defined in Sec. 3), using the seven inputs (**X**) reported in Fig. 4 and described as follows.

The first input is negative if the solar azimuth angle is lower than 180 deg and positive if the solar azimuth angle is greater than 180 deg. It describes the sun position with respect to the plant angle of incidence and the small asymmetry in the PK_{cs} between morning and afternoon of clear sky days due to different water vapor contents (input sign). It also takes into account the impact on the plant performance of reflection and incident spectrum. The second input (NWP of ground air temperature) was introduced to consider the effect of temperature on the plant performance. The other four inputs are the average values and the standard deviation of the relative humidity predicted by WRF for the vertical levels below 775 hPa and between 775 hPa and 400 hPa. The seventh is the relative humidity predicted by WRF at a higher level corresponding to 300 hPa. The last five parameters aim to take into account the cloud formation at different atmosphere levels: low clouds approximately below 2500 m, midclouds between 2500 and 7500 m, and high clouds up to 9000 m. Following the idea behind MOSRH (described in Sec. 4.2), the model RHNN aims to provide a direct forecast of power output, starting from the numerical weather prediction of the relative humidity of atmospheric levels. Nevertheless, we need to remark that also different inputs were tested to develop the ANN model: relative humidity, air temperature, wind speed, and geopotential predicted by WRF for 20 vertical levels. Different inputs combination and average do not effectively improve the accuracy of RHNN model. Thus, this model is the best compromise between input information and model complexity. This also confirms that the major source of error in the WRF irradiance prediction is the capability to estimate the damping of the clear sky irradiance due to the relative humidity of atmospheric levels. A similar approach was used in

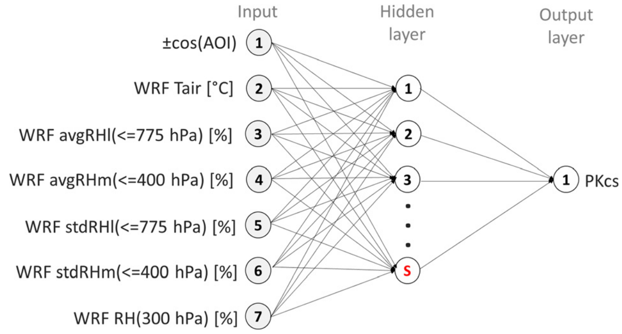

Fig. 4 Inputs (X) and output (Y) of the RHNN model

Ref. [23], where the cloud cover forecast of three atmospheric levels is used as predictor for their random forest model.

Thus, Eq. (10) becomes

$$PK_{cs}^{for} = \mathrm{RHNN}(\mathbf{X}) \tag{11}$$

so that the PV power generation forecast 24 h ahead can be calculated as follows:

$$P_m^{for} = P_n PK_{cs}^{for}(\mathrm{GPOAI_{cs}}/G_0) \tag{12}$$

where $\mathrm{GPOAI_{cs}}$ is the clear sky plane of array irradiance predicted by the WRF model, P_m^{for} is the hourly power generation forecast, and P_n is the plant nominal power.

Finally, the RHNN model was generated by a master optimization procedure, described in detail in Ref. [46] and here briefly summarized.

The Levenberg–Marquardt algorithm was used to minimize the mean square error (MSE) function using 60% of one year data for training and 40% for validation. The net structure was identified through an optimization process that provided the best number of neurons in the hidden layer (S) through a further MSE minimization procedure. Once the best number of hidden neurons was identified, 500 ANNs were generated using the repeated random sample validation procedure. Subsequently, a qualified ensemble was selected (around 300 ANNs), choosing all the ANNs with the MSE lower than the average MSE of the 500 networks. Finally, the forecast was obtained by averaging the ensemble outputs. Since the relationship between the predicted weather variables (i.e., RH and temperature) and the clear sky performance index, modeled by RHNN, should not depend on time, it remains stationary during the considered years (2011–2014) making each year independent from each other. For this reason, the accuracy of ANNs ensemble should not depend on the year used for the training and validation procedure. Thus, two different ensembles of ANNs were generated: one trained and validated on the year 2011 with seven neurons in the hidden layer and the second trained and validated on the year 2012 with six hidden neurons. The ANN ensemble trained on 2011 was used for the forecast of the years 2012, 2013, and 2014, while the ANN ensemble trained on 2012 was used to forecast the year 2011.

5 Metrics to Evaluate Forecast Accuracy

According to the solar forecast literature, the main metrics used to evaluate the models accuracy are reported in Table 3.

The RMSE accentuates the greater forecasting errors while the MAE is exactly the measure of the unbalanced power. The MBE describes systematic deviation of the forecast. All the metrics are calculated using one year of observational data to provide a reliable information on the forecast performance. The normalized error indexes: NRMSE, NMAE, and NMBE are calculated dividing by the yearly mean generated power. All the performance indexes above are calculated excluding the night values (when the plane of array clear sky irradiance provided by WRF is equal to zero). The forecast performance on a fixed horizon essentially depends on site and year. To compare the accuracy obtained by different models in different sites or years, a reference model is used. If the reference model has a similar performance (with respect to a specific metric) in two different sites or years, then the two weather conditions can be considered comparable. Thus, also the accuracy of different forecast models calculated by the same metric can be compared. The skill score (SS) is less site and year dependent, and it allows to evaluate which forecasting model outperforms.

The most common reference model used in the solar forecast sector is the persistence model (PM), as defined in Sec. 4. The accuracy of the PM can be considered as a measurement of the forecast difficulty for a specific site and period so that the skill score with respect to the PM states the quality of the adopted forecast model. An example of the use of the skill score to evaluate irradiance variability can be found in Ref. [47]. Finally, the energy imbalance (EI) measures the difference between the forecast and the real energy production in the considered period.

6 Results

6.1 Deterministic Model Preliminary Assessment

6.1.1 NWP Irradiance Assessment. Lorenz et al. [48] presented a benchmarking of different approaches to predict the GHI measured by different weather stations in four European countries. Figure 5 shows minimum, maximum, and average values of RMSE obtained by these different approaches together with the RMSE achieved by the MOSRH(WRF) irradiance prediction for the site of Bolzano. These forecast errors are plotted as a function of the RMSE of the persistence model used to measure the irradiance variability at different locations (see Ref. [47]).

It can be observed that accuracy reached by the MOSRH(WRF) forecast is inside the European performance range, below the Austria and Switzerland average values, even though the RMSE of the persistence model is higher in Bolzano than in the other locations. Indeed, the persistence error indicates greater difficulties in irradiance forecasting. Thus, the GHI numerical weather prediction data used in the deterministic approach show a "state-of-the-art" quality level.

6.1.2 Transposition Model Assessment on Actual Data. Figure 6 shows the errors of the isotropic transposition model (IM) in different typologies of days identified by the daily clear sky index: overcast days with $K_{cs} < 0.4$, partially cloudy with $0.4 < K_{cs} < 0.8$, and clear sky with $K_{cs} > 0.8$. In the figure, also yearly accuracy values are reported. From Fig. 6, it appears that the maximum underestimation is achieved in clear sky conditions with variable irradiance ($0.6 < K_{cs} < 0.8$) when the reflected irradiance incident on the plane of array can have higher values than the horizontal one. This confirms the result in Ref. [49]. Indeed, the author found that the Liu–Jordan isotropic transposition model provides a good fit to empirical data under overcast skies but underestimates the amount of solar radiation incident on tilted surfaces under clear and partly cloudy conditions.

As a whole, the models achieve an NRMSE of 7.5% (with an averaged GPOAI of 340.3 W/m^2), values that are not so far from the pyranometer uncertainty limits, confirming the results found in Ref. [50].

6.1.3 Power Estimation Model Assessment on Actual Data. The Sandia PV array performance model (SAPM) provides a very

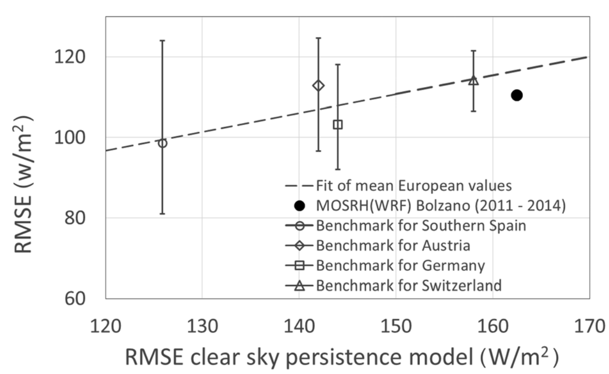

Fig. 5 Minimum, maximum, and average values of RMSE obtained by different irradiance forecast approaches reported in Ref. [44] and the RMSE achieved by the MOSRH(WRF) solar irradiance prediction for Bolzano site versus the RMSE of the clear sky persistence model

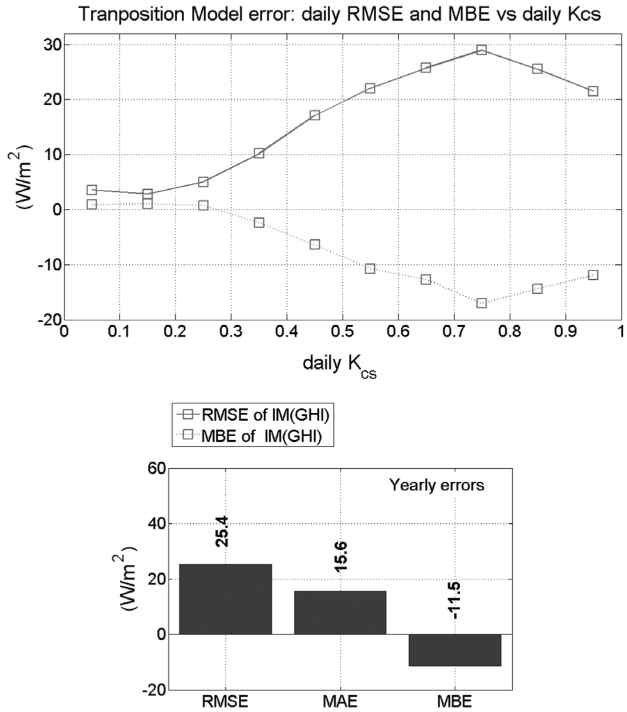

Fig. 6 Liu–Jordan isotropic transposition errors in different typologies of days identified by the daily clear sky index

accurate description of all the main sources of loss or gain that affect the PV power generation. Nevertheless, the variation of the nominal power of the plant due to metastability of modules material (degradation or seasonal fluctuation) is not taken into account. The impact of this effect on the performance of crystalline modules is usually negligible, while it could be very important for thin films devices, see Refs. [30,42,51]. Indeed, since in this case the plant is based on CdTe thin film technology, a fine-tuning of the module degradation brings to a remarkable improvement of the model accuracy. If this effect is neglected, the SAPM yields an RMSE of 5.2% of P_n, and it provides a notable overestimation of the generated power (MBE = 3.2% of P_n) because the real PV plant power is degrading over the years. On the contrary, considering an initial degradation of 5% of P_n ($\Delta Pn = 95\%$ of P_n) and a yearly degradation rate of −1.3% ($\Delta Pn = 98.7\%$ of P_n), as measured by Belluardo et al. [30], the model exhibits an RMSE and MBE of 2.8% and 1.2% of P_n. Nevertheless, it should be pointed out that the estimation of the degradation rate is rarely possible in operative forecasting. Thus, the prediction accuracy reached by the deterministic model reported in this paper could be easily considered as the higher limit of the deterministic forecast performance. Figure 7 shows the good agreement between the measured power and performance ratio and estimated ones. It should be remarked that the higher errors in the PR estimation early in the morning and late in the afternoon are related to the shadowing by mountains that cannot be estimated by the SAPM.

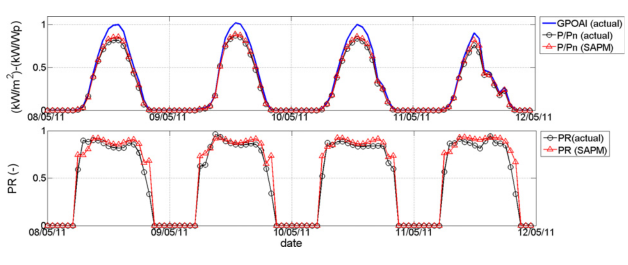

Fig. 7 Hourly trend of power and performance ratio, measured and estimated by the SAPM

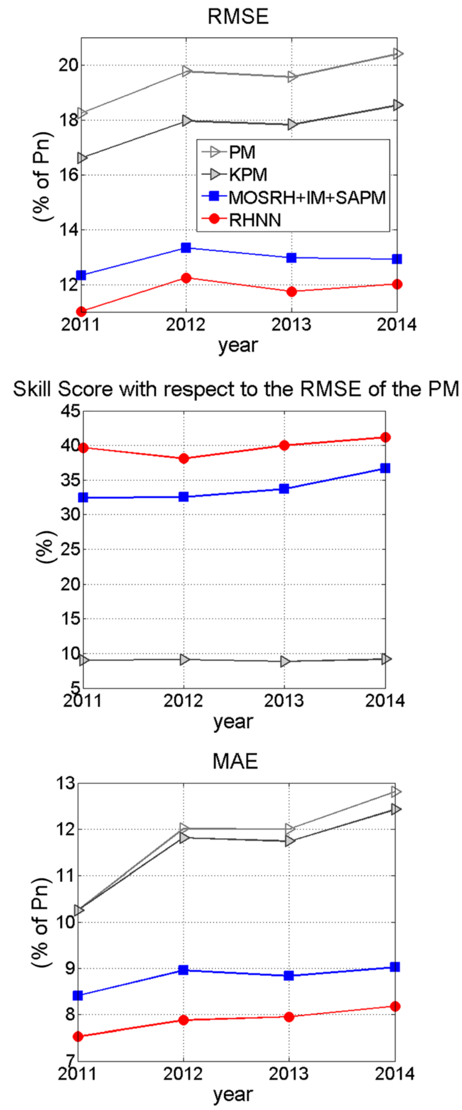

Fig. 8 Accuracy comparison of PV power generation forecast (yearly trends)

6.2 Forecast Models Accuracy Analysis

6.2.1 Accuracy Comparison. In Figs. 8 and 9, the yearly accuracy values of the deterministic and stochastic models are reported together with the reference persistence models. From Fig. 8, it can be pointed out that for all the models, the skill score is almost constant for all the years. For this reason, this metric is usually used to compare forecast methods in different years or sites.

In the irradiance forecast community, usually the persistence model used as reference is the one obtained by the clear sky index since it achieves an RMSE lower than the simple persistence [38]. In the same way, for the power forecast, the persistence calculated using the clear sky performance index (KPM) shows a 10% of

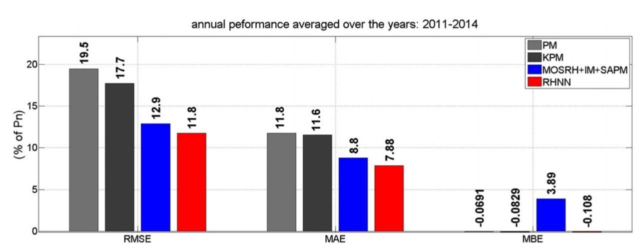

Fig. 9 Accuracy comparison of PV power generation forecast

skill score with respect to the simple persistence model (PM) (see Fig. 8), so that it could be used as a better benchmark model. Nevertheless, since in the literature on the PV generation forecast, the simple persistence is adopted as a reference model, the skill scores reported in this section are calculated with respect to the RMSE of the simple PM.

The deterministic model (MOSRH + IM + SAPM) achieves an RMSE and MAE of 12.9% and 8.8% of P_n, showing a systematic power overestimation confirmed by an MBE of 4% of P_n. It obtains a skill score of 35% with respect to the RMSE of PM.

The stochastic model (RHNN) exhibits an unbiased forecast with RMSE and MAE of 11.8% and 7.8% of P_n, obtaining a skill score of 39%.

Figure 10 shows the behavior of residuals of the two approaches. The probability to obtain a forecast error inside the range of $\pm 5\%$ of P_n is 46.5% for the deterministic model and 50% for the stochastic one. Indeed, the stochastic approach reduces the overestimation at low generation level ($P_m/P_n < 450$ kW/kWp) even if it slightly increases the underestimation at high generation ($P_m/P_n > 950$ kW/kWp), see Figs. 10(*a*2) and 10(*b*2).

Moreover, this model obtains a more accurate power distribution forecast, bringing the distance of the Kolmogorov–Smirnov test (KS) from 0.13 to 0.078, see Figs. 10(*a*3) and 10(*b*3).

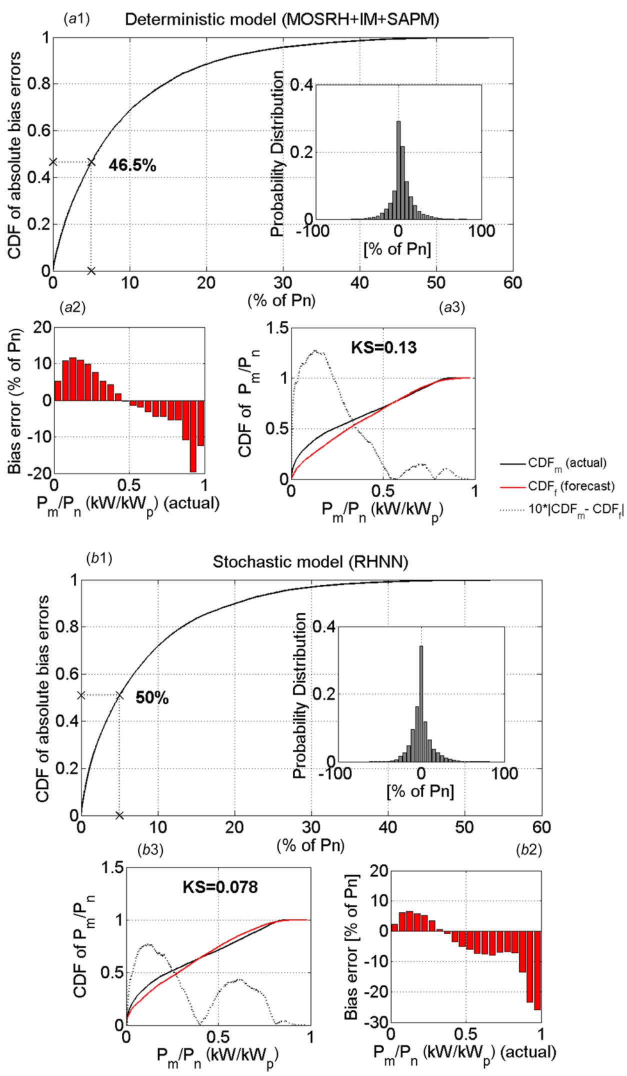

Fig. 10 Behavior of residuals of the deterministic and stochastic models: (1) power distribution of hourly errors, (2) bias errors versus rated power (P_m/P_n), and (3) Kolmogorov–Smirnov test

Furthermore, the stochastic model provides a remarkable improvement of the number of days with lower daily mean absolute error with respect to the deterministic one. It appears that the rate of days with a daily MAE less than 5–10% of P_n obtained through the two forecast models is, respectively, 19–66% (deterministic) and 29–76% (stochastic). In addition to this, the first method leads to a yearly average energy imbalance of 378 kWh/kWp corresponding to 29% of the produced energy. The second method achieves an energy imbalance of 338 kWh/kWp, corresponding to 26% of the produced energy.

Thus, the stochastic model outperforms the deterministic one. The reason of this accuracy improvement will be explained in Sec. 6.3.2.

Similar results were found by Huang et al. [52], where they obtained an RMSE of 12.45% for their deterministic model and 10.5% for the stochastic one (based on MLPNN architecture). For that specific study, the one-diode model used for power estimation is less accurate than the SAPM, obtaining on actual data an RMSE of 5.5% of P_n.

Unfortunately, there are no European accuracy benchmarks readily available in the literature on PV power generation forecasts, such as Ref. [48] for the GHI prediction. Thus, it is difficult to understand the quality of the reported forecast methods. In spite of this, Lorentz [53] using a deterministic approach reports for Germany an average value of RMSE of 12.8% (one day-ahead forecast on single PV plant). Considering that the RMSE of GHI forecast (achieved by persistence) for Germany is lower than the RMSE of the Bolzano site (see Fig. 5), the obtained accuracy of 12.9% could be considered as an interesting result. Always using a deterministic approach with GEM postprocessed NWP, Pelland et al. [10] found, for three Canadian PV plants, an RMSE range from 6.38% to 9.17% corresponding to a skill score between 36.7% and 64% with respect the RMSE of PM. This could prove that the postprocessed GEM prediction of solar irradiance outperforms both the one used in this work and the ECMWF used by Lorenz. Moreover, da Silva Fonseca et al. [21] mainly report some accuracy values for regional PV generation forecasts in Japan. Nevertheless, from the study, the single site RMSE obtained by the support vector regression model could be deduced: 9–14% of P_n. For a single plant in France counties, Zamo et al. [23] experiment different forecast approaches. They lead to a skill score between 29.4% and 47%. Again, the best result was achieved by the stochastic random forest model. Thus, the RHNN with an RMSE of 11.8% and a skill score of 39% is perfectly inside the range of accuracy reported in the literature.

6.3 Error Analysis

6.3.1 Error Propagation in the Deterministic Approach. To better understand the reason for the lower accuracy of the deterministic forecast, it is useful to analyze the error propagation of all the models used in this approach: GHI forecast model, transposition model, and power estimation model. The PV power generation can be described by

$$P_m = P_n \mathrm{PR}(\mathrm{GPOAI}/1000) = P_n \mathrm{PR}(\mathrm{TF})(\mathrm{GHI}/1000) \quad (13)$$

where PR is the performance ratio defined in Sec. 3, and TF is the transposition factor TF = GPOAI/GHI

Thus, to study the impact of each model on the overall performance of the deterministic approach, we present the comparison of (i) the accuracy reached using only the GHI numerical weather prediction, (ii) using the GHI forecast and the transposition, and (iii) using all the three models in cascade

$$P_m(\mathrm{MOSRH}) = P_n\, \mathrm{PR}(\mathrm{actual})\, \mathrm{TF}(\mathrm{actual})\, (\mathrm{GHI}(\mathrm{MORH})/1000) \times (\mathrm{kW}) \quad (14)$$

$$P_m(\text{MOSRH}+\text{IM}) = P_n\,\text{PR}(\text{actual})\,(\text{GPOAI}(\text{MORH})/1000)\,(\text{kW}) \quad (15)$$

$$P_m(\text{MOSRH}+\text{IM}+\text{SAPM}) = P_n\,\text{PR}(\text{SAPM})\,(\text{GPOAI} \times (\text{MORH})/1000)\,(\text{kW}) \quad (16)$$

Figure 11(*a*) shows the accuracy of the three forecasts, and thus the error propagation. In the box (Fig. 11(*b*)), the performance of the SAPM is also reported when the nominal power degradation effect is neglected.

The errors of the first model (MOSRH) obviously reflect the accuracy of the NWP forecast data (see Fig. 5) reaching an RMSE of 10% of P_n. Then, the isotropic transposition model (IM) brings the RMSE to 13% of P_n. Finally, using an accurate evaluation of the degradation effect, the power estimation model (SAPM) does not introduce any further error. As previously mentioned, in operative forecasting, the degradation is not as easily predictable. Figure 11(*b*) shows that the power estimation model could increase the RMSE up to 14% of P_n and the bias up to 6% (because the forecast overestimation is increased by the nominal power degradation). In any case, the main source of accuracy loss, right after the GHI prediction, is due to the transposition model.

It is possible to generalize these results on the deterministic approach. The impact of the GHI forecast on the overall RMSE could go from a minimum of 80% in case of optimal PV position (as in this case) to a maximum of 95–100% in case of a horizontal PV plant. On the contrary, the contribution of the transposition model to the RMSE could range from 25% to 0% of P_n (optimal PV position–horizontal position). The accuracy of SAPM depends on the PV technology, and it could impact the forecast errors from a minimum of 0% in case of crystalline modules (very weak power degradation) to a maximum of 5% in case of thin films modules (remarkable initial performance loss and annual degradation of plant nominal power).

6.3.2 Deterministic Versus Stochastic Approach: An In-Depth Error Analysis. From the error propagation analysis of the deterministic model, it is possible to understand the main contribution of the stochastic approach in improving the forecast performance. Figure 12(*a*) shows the error on actual data of the isotropic transposition model (RMSE_{G0}), the errors of GHI and GPOAI forecast (RMSE_{G1} and RMSE_{G2}), and the forecast error due to the transposition model ($\text{RMSE}_d = \text{RMSE}_{G2} - \text{RMSE}_{G1}$) versus the daily clear sky index (overcast days with $K_{cs} < 0.4$, partially cloudy with $0.4 < K_{cs} < 0.8$, and clear sky with $K_{cs} > 0.8$).

It appears that the error of the isotropic model on the actual data (RMSE_{G0}) is lower than the error of the isotropic model on the forecast data (RMSEd). The first reason of this difference is related to albedo: the nearer the tilt is to 90 deg, the greater could be the contribution of the reflected component on the global incident irradiance. Since the reflected irradiance cannot be forecast by numerical weather prediction models, the GPOAI forecast accuracy decreases with respect to the GHI in all the typologies of days. The second main reason is related to the NWP errors in the

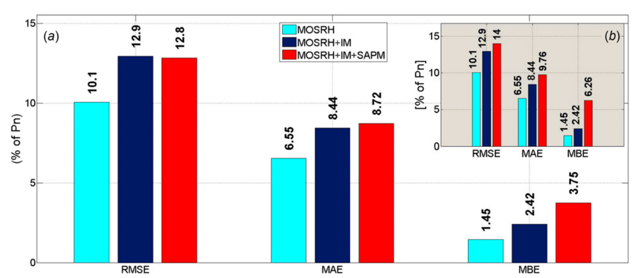

Fig. 11 (*a*) Error propagation in the deterministic model, in box (*b*) error propagation not considering the nominal power degradation of the PV modules

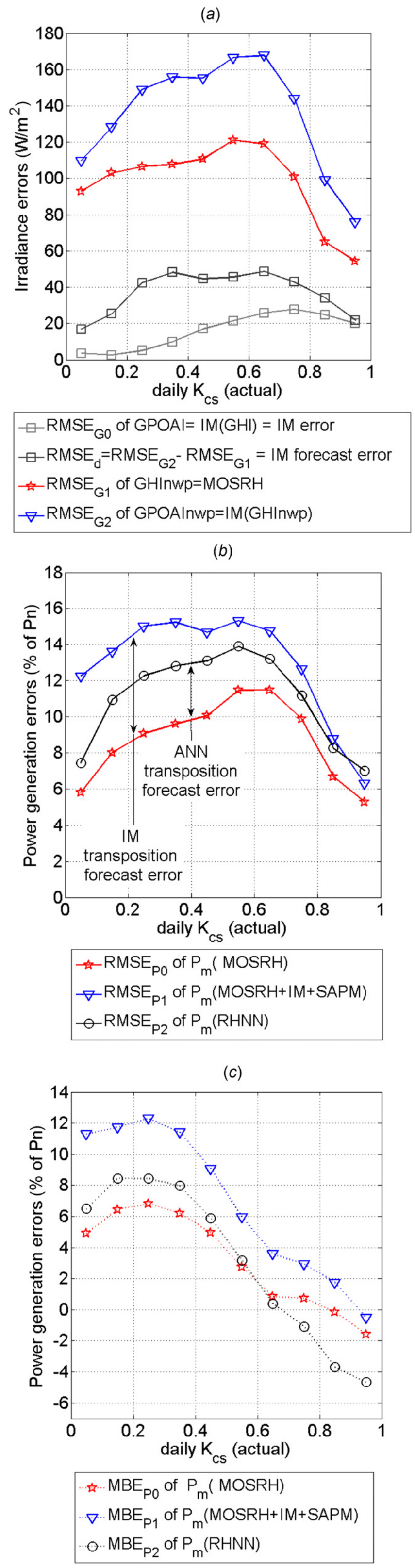

Fig. 12 (*a*) RMSE of the isotropic transposition model on actual GHI data (RMSE_{G0}), RMSE of GHI and GPOAI prediction (RMSE_{G1} and RMSE_{G2}), and the forecast error due to the transposition model ($\text{RMSE}_d = \text{RMSE}_{G2} - \text{RMSE}_{G1}$) ((*b*) and (*c*)) RMSE and MBE comparison of the deterministic and stochastic power forecast in different typologies of days (RMSE_{P0}–MBE_{P0} of the model P_m(MOSRH) considering only the GHI forecast; RMSE_{P1}–MBE_{P1} of the full deterministic model P_m(MOSRH + IM + SAPM), and RMSE_{P2}–MBE_{P2} of the stochastic model P_m(RHNN))

prediction of direct normal irradiance (DNI). Indeed, the impact of DNI on the global irradiance is higher as the plane of array moves from horizontal to its optimal position: south orientation and 30 deg of tilt angle (for Italy). So that the error in DNI prediction will affect more the GPOAI than the GHI. For this reason, the transposition error on forecast data (RMSE_d) reaches its maximum in partially cloudy days when the DNI prediction has higher probability of failure. Thus, the transposition model amplifies the NWP errors of the GHI forecast.

Figures 12(*b*) and 12(*c*) report the RMSE_{P0} and MBE_{P0} for the deterministic power forecast P_m(MOSRH) considering only the GHI forecast, the RMSE_{P1} and MBE_{P1} of the deterministic power forecast P_m(MOSRH + IM + SAPM) considering all the models chain, and the RMSE_{P2} and MBE_{P2} of the stochastic power forecast P_m(RHNN) in different typologies of days.

The RMSE of the deterministic models clearly reflects the errors of the irradiance prediction. Indeed, the power forecast errors (RMSE_{P0} and RMSE_{P1}) show similar trend of GHI and GPOAI forecast errors (RMSE_{G0} and RMSE_{G1}). The stochastic model (RHNN) achieved lower RMSE and MBE with respect to the deterministic one (MOSRH + IM + SAPM) for almost all the different typologies of days. The difference between the RMSE_{P1} and RMSE_{P0} is the isotropic transposition forecast error while the difference between RMSE_{P2} and RMSE_{P0} is the ANN transposition forecast error. It can be observed that the ANN model halve the transposition error, bringing the MBE almost at the same level of P_m(MOSRH) model. Thus, the stochastic approach is outperforming mainly because it is able to provide a better transposition of the GHI forecast.

The RHNN not only corrected NWP bias errors of the GHI prediction (avoiding the MOSRH post processing) but also improved the accuracy of the GPOAI forecast.

It should be remarked that Pelland et al. [10] tested different transposition models for their deterministic PV power forecast. The authors proved that even using outperforming transposition methods, the forecast accuracy did not show remarkable improvement. Indeed, the GHI prediction errors are much higher than the GHI transposition errors (see RMSE_{G0} and RMSE_{G1} in Fig. 12(*a*)) so that an improvement of transposition model performance has a weak effect on the GPOAI forecast accuracy. In the deterministic approach, the transposition method essentially amplifies the GHI prediction errors. On the contrary, the stochastic approach is able to reduce both the forecast and the transposition error. Gulin et al. [54] proved that stochastic models based on ANN outperform the most used isotropic and anisotropic transposition models when applied on actual data. In particular, as in the present case, the ANN models reduce the RMSE of 50% with respect to RMSE of the Liu–Jordan isotropic model.

7 Summary and Conclusion

In this paper, a detailed analysis of the accuracy of one original deterministic and one original stochastic model for day-ahead PV power forecast is reported. Four years of measurements of PV power plant production, irradiance, and temperature from a 662 kWp cadmium telluride PV plant located in Bolzano, Italy, were used to train and test the models. Moreover, a new index called clear sky performance index (PK_{cs}) was defined. This could be considered as the equivalent of the clear sky index for PV power generation and, as for the irradiance forecast, it could be used both to develop an outperforming persistence model and for site characterization (when no irradiance data are available). The PK_{cs} was also used to build a stochastic model. The deterministic approach consists in applying a transposition model and a power estimation model to the forecast of the GHI coming from an NWP model. In this work, the WRF irradiance forecast postprocessed by model output statistic algorithm was presented and used. Moreover, the Liu–Jordan isotropic model and the Sandia PV array performance model were adopted to transpose the GHI on the plane of array and to estimate the power generation. A preliminary assessment of the NWP data and of the transposition and estimation models used in the deterministic approach was developed. The MOSRH postprocessing used to refine the WRF irradiance forecast obtains an RMSE of 110 W/m^2, perfectly inside the European performance range [48]. In the same way, also the IM and SAPM models tested on the actual data exhibit a high accuracy level with an RMSE of 25 W/m^2 and 2.8% of P_n. Thus, the performance of the deterministic forecast could be considered as a benchmark for this kind of approach. Furthermore, the error propagation inside the model chain used in the deterministic approach (GHI prediction, transposition, and power estimation) was studied. It was proved that 80% of the overall RMSE between actual and predicted power is due to the GHI forecast while the other 20% is due to the IM. Indeed, the transposition amplifies the GHI prediction errors. Moreover, the impact of the nominal power degradation of the CdTe modules on the forecast accuracy was evaluated. Without considering the degradation effect inside SAPM, the RMSE of the deterministic model grows from 12.8% to 14% of P_n. The stochastic approach is based on artificial learning algorithms that directly provide the PV power generation forecast using various meteorological variables coming from NWP tools. A stochastic model (RHNN) based on an ensemble of ANNs was developed. The RHNN used the relative humidity of different vertical levels predicted by WRF to forecast the clear sky performance index. Obviously, the deterministic model errors were strictly related to the GHI prediction used as input and to the transposition model. On the contrary, the stochastic model not only was able to correct NWP bias errors (avoiding the MOSRH post processing) but also to provide a better irradiance transposition. Thus, the stochastic approach outperformed the deterministic one. As a whole, the deterministic approach leads to an RMSE and MAE of 12.9% and 8.8% of P_n, obtaining a skill score of 35% with respect to the simple persistence model. The stochastic approach provides an RMSE and MAE of 11.8% and 7.9% of P_n, with a skill score of 39%. The annual energy imbalance (defined in Table 3) of the deterministic and stochastic forecast are, respectively, of 29–26% of the yearly produced energy.

Acknowledgment

D.M. and M.P. are indebted to the funding received from the EURAC internal project "Regional Energy Modelling:RegEn Mod."

Nomenclature

ANN = artificial neural network
GHI = global horizontal irradiance
GPOAI = global plane of array irradiance
IM = isotropic transposition model
KPM = clear sky persistence model
MLPNN = multilayer perceptron neural network
MOS = model output statistics
MOSRH = model output statistics for relative humidity
NWP = numerical weather prediction
PCC = pseudocloud cover
PM = persistence model
PV = photovoltaics
RHNN = relative humidity neural network model
RRTM = rapid radiation transfer model
SAPM = Sandia PV array performance model
WRF = weather research and forecasting

Appendix: WRF Physics Configuration

The WRF model radiation schemes provide atmospheric heating due to radiative flux divergence and downward surface long and short radiation wavelengths for the ground heat budget. The long wavelength radiation (above 4000 nm) includes infrared or thermal radiation absorbed and emitted by gases and surfaces. The short wavelength radiation (between 300 and 4000 nm) includes visible and surrounding wavelengths that make up the solar spectrum. Absorption, reflection, and scattering processes in the atmosphere and on the surface are simulated. Upward shortwave radiation is the reflection due to surface albedo [32]. The shortwave radiation in the WRF model corresponds to GHI.

The shortwave (SW) radiation scheme chosen is the rapid radiative transfer model (RRTMG) [55–57] because of its skill to describe subgrid cloud variability through a Monte Carlo independent column approximation and due to its ability to distinguish near-infrared, visible, and UV wavelengths. Radiation absorption and scattering by other gases and aerosols are calculated through various parameterizations. Concentration and chemical composition come from parametric values taken from prebuilt tables. To obtain more accurate results on aerosol concentration and chemical composition of the atmosphere, the WRF model could be coupled with an atmospheric chemistry model (WRF-Chem distribution) [58]. Typically, coupling is not performed for daily operational weather forecasts because of the computational time it requires and the lack of accurate source data needed for the entirety of the domain being considered. For these reasons, no coupled chemical model was run for these simulations.

References

[1] IEA, 2014, "2014 Snapshot of Global PV Markets," IEA PVPS, International Energy Agency, Paris, Technical Report No. IEA PVPS T1-26:2015.

[2] IEA, 2014, "Technology Roadmap Solar Photovoltaic Energy: 2014 Edition," IEA Renewable Energy Division, Paris.

[3] Alet, P. J., Baccaro, F., De Felice, M., Efthymiou, V., Mayr, C., Graditi, G., Juel, M., Moser, D., Petitta, M., Tselepis, S., and Yang, G., 2015, "Photovoltaics Merging With the Active Integrated Grid: A White Paper of the European PV Technology Platform," European PV Technology Platform, Brussels, Belgium.

[4] Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Knowe, G. V., Hemker, K., Heinemann, D., Remund, J., Müller, S. C., Traunmüller, W., Steinmauer, G., Pozo, D., Ruiz-Arias, J. A., Lara-Fanego, V., Ramirez-Santigosa, L., Gaston-Romero, M., and Pomares, L. M., 2013, "Comparison of Numerical Weather Prediction Solar Irradiance Forecasts in the US, Canada and Europe," Sol. Energy, **94**, pp. 305–326.

[5] Perez, R., Moore, K., Wilcox, S., Renné, D., and Zelenka, A., 2007, "Forecasting Solar Radiation—Preliminary Evaluation of an Approach Based Upon the National Forecast Database," Sol. Energy, **81**(6), pp. 809–812.

[6] Lorenz, E., Hurka, J., Heinemann, D., and Beyer, H. G., 2009, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., **2**(1), pp. 2–10.

[7] Pierro, M., Bucci, F., Cornaro, C., Maggioni, E., Perotto, A., Pravettoni, M., and Spada, F., 2015, "Model Output Statistics Cascade to Improve Day Ahead Solar Irradiance Forecast," Sol. Energy, **117**, pp. 99–113.

[8] Lorenz, E., Hurka, J., Karampela, G., Heinemann, D., Beyer, H. G., and Schneider, M., 2008, "Qualified Forecast of Ensemble Power Production by Spatially Dispersed Grid-Connected PV Systems," 23rd European Photovoltaic Solar Energy Conference and Exhibition, Valencia, Spain, Sept. 1–5, pp. 3285–3291.

[9] Lorenz, E., Scheidsteger, T., Hurka, J., Heinemann, D., and Kurz, C., 2010, "Regional PV Power Prediction for Improved Grid Integration," Prog. Photovoltaics: Res. Appl., **19**(7), pp. 757–771.

[10] Pelland, S., Galanis, G., and Kallos, G., 2011, "Solar and Photovoltaic Forecasting Through Post-Processing of the Global Environmental Multiscale Numerical Weather Prediction Model," Prog. Photovoltaics: Res. Appl., **21**(3), pp. 284–296.

[11] Yona, A., Senjyu, T., Saber, A. Y., Funabashi, T., Sekine, H., and Kim, C.-H., 2008, "Application of Neural Network to 24-Hour-Ahead Generating Power Forecasting for PV System," IEEE Power and Energy Society General Meeting—Conversion and Delivery of Electrical Energy in the 21st Century, Institute of Electrical and Electronics Engineers (IEEE).

[12] Chen, C., Duan, S., Cai, T., and Liu, B., 2011, "Online 24-h Solar Power Forecasting Based on Weather Type Classification Using Artificial Neural Network," Sol. Energy, **85**(11), pp. 2856–2870.

[13] Tao, C., Shanxu, D., and Changsong, C., 2010, "Forecasting Power Output for Grid-Connected Photovoltaic Power System Without Using Solar Radiation Measurement," 2nd International Symposium on Power Electronics for Distributed Generation Systems, Institute of Electrical and Electronics Engineers (IEEE), June 16–18, pp. 773–777.

[14] Wang, S., Zhang, N., Zhao, Y., and Zhan, J., 2011, "Photovoltaic System Power Forecasting Based on Combined Grey Model and BP Neural Network," International Conference on Electrical and Control Engineering, Sept. 16–18, Institute of Electrical and Electronics Engineers (IEEE), pp. 4623–4626.

[15] Mellit, A., Pavan, A. M., and Lughi, V., 2014, "Short-Term Forecasting of Power Production in a Large-Scale Photovoltaic Plant," Sol. Energy, **105**, pp. 401–413.

[16] Larson, D. P., Nonnenmacher, L., and Coimbra, C. F. M., 2016, "Day-Ahead Forecasting of Solar Power Output From Photovoltaic Plants in the American Southwest," Renewable Energy, **91**, pp. 11–20.

[17] Bacher, P., Madsen, H., and Nielsen, H. A., 2009, "Online Short-Term Solar Power Forecasting," Sol. Energy, **83**(10), pp. 1772–1783.

[18] Mellit, A., and Pavan, A. M., 2010, "A 24-h Forecast of Solar Irradiance Using Artificial Neural Network: Application for Performance Prediction of a Grid-Connected {PV} Plant at Trieste, Italy," Sol. Energy, **84**(5), pp. 807–821.

[19] Li, Y., Su, Y., and Shu, L., 2014, "An ARMAX Model for Forecasting the Power Output of a Grid Connected Photovoltaic System," Renewable Energy, **66**, pp. 78–89.

[20] Zamo, M., Mestre, O., Arbogast, P., and Pannekoucke, O., 2014, "A Benchmark of Statistical Regression Methods for Short-Term Forecasting of Photovoltaic Electricity Production, Part I: Deterministic Forecast of Hourly Production," Sol. Energy, **105**, pp. 792–803.

[21] da Silva Fonseca, J. G., Jr., Oozeki, T., Ohtake, H., Shimose, K.-i., Takashima, T., and Ogimoto, K., 2014, "Regional Forecasts and Smoothing Effect of Photovoltaic Power Generation in Japan: An Approach With Principal Component Analysis," Renewable Energy, **68**, pp. 403–413.

[22] da Silva Fonseca, J. G., Jr., Oozeki, T., Ohtake, H., Takashima, T., and Ogimoto, K., 2015, "Regional Forecasts of Photovoltaic Power Generation According to Different Data Availability Scenarios: A Study of Four Methods," Prog. Photovoltaics: Res. Appl., **23**(10), pp. 1203–1218.

[23] Zamo, M., Mestre, O., Arbogast, P., and Pannekoucke, O., 2014, "A Benchmark of Statistical Regression Methods for Short-Term Forecasting of Photovoltaic Electricity Production, Part II: Probabilistic Forecast of Daily Production," Sol. Energy, **105**, pp. 804–816.

[24] Almeida, M. P., Perpiñán, O., and Narvarte, L., 2015, "PV Power Forecast Using a Nonparametric PV Model," Sol. Energy, **115**, pp. 354–368.

[25] Davò, F., Alessandrini, S. S., Sperati, S., Monache, L. D., Airoldi, D., and Vespucci, M. T., 2016, "Post-Processing Techniques and Principal Component Analysis for Regional Wind Power and Solar Irradiance Forecasting," Sol. Energy, **134**, pp. 327–338.

[26] Sperati, S., Alessandrini, S., and Delle Monache, L., 2016, "An Application of the ECMWF Ensemble Prediction System for Short-Term Solar Power Forecasting," Sol. Energy, **133**, pp. 437–450.

[27] Paulescu, M., Paulescu, E., Gravila, P., and Badescu, V., 2012, *Weather Modeling and Forecasting of PV Systems Operation*, Springer Science+Business Media, Berlin.

[28] Kleissl, J., 2013, *Solar Energy Forecasting and Resource Assessment*, Academic Press, Cambridge, MA.

[29] IEA, 2013, "Photovoltaic and Solar Forecasting: State of the Art," IEA PVPS, International Energy Agency, Paris, Technical Report No. IEA-PVPS T14-01: 2013.

[30] Belluardo, G., Ingenhoven, P., Sparber, W., Wagner, J., Weihs, P., and Moser, D., 2015, "Novel Method for the Improvement in the Evaluation of Outdoor Performance Loss Rate in Different PV Technologies and Comparison With Two Other Methods," Sol. Energy, **117**, pp. 139–152.

[31] Bertani, D., Guastella, S., Belluardo, G., and Moser, D., 2015, "Long Term Measurement Accuracy Analysis of a Commercial Monitoring System for Photovoltaic Plants," IEEE Workshop on Environmental, Energy and Structural Monitoring Systems (EESMS), July 9–10, pp. 84–89.

[32] Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G., 2008, "A Description of the Advanced Research WRF Version 3," NCAR, Boulder, CO, Technical Note NCAR/TN-4751STR.

[33] Rogers, E., Black, T., Ferrier, B., Lin, Y., Parrish, D., and DiMego, G., 2001, "Changes to the NCEP Meso Eta Analysis and Forecast System: Increase in Resolution, New Cloud Microphysics, Modified Precipitation Assimilation, Modified 3DVAR Analysis," Office of Metrology, National Weather Service, Silver Spring, MD.

[34] Paulson, C. A., 1970, "The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer," J. Appl. Meteorol., **9**(6), pp. 857–861.

[35] Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R., 2004, "Implementation and Verification of the Unified NOAH Land Surface Model in the WRF Model," 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, pp. 11–15.

[36] Hong, S.-Y., Noh, Y., and Dudhia, J., 2006, "A New Vertical Diffusion Package With an Explicit Treatment of Entrainment Processes," Mon. Weather Rev., **134**(9), pp. 2318–2341.

[37] Kain, J. S., 2004, "The Kain–Fritsch Convective Parameterization: An Update," J. Appl. Meteorol., **43**(1), pp. 170–181.

[38] Beyer, H. G., Polo Martinez, J., Suri, M., Torres, J. L., Lorenz, E., Müller, S. C., Hoyer-Klick, C., and Ineichen, P., 2009, "Report on Benchmarking of Radiation Products," Management and Exploitation of Solar Resource Knowledge (MESOR), Sixth Framework Programme, Contract No. 038665.

[39] Houghton, J., 2002, *The Physics of Atmospheres*, Cambridge University Press, Cambridge, UK.

[40] Liu, B., and Jordan, R., 1961, "Daily Insolation on Surfaces Tilted Towards Equator," ASHRAE J., **10**, p. 5047843.

[41] King, D. L., Kratochvil, J. A., and Boyson, W. E., 2004, "Photovoltaic Array Performance Model," Sandia National Laboratories, Albuquerque, NM, Report No. SAND2004-3535.

[42] Pierro, M., Bucci, F., and Cornaro, C., 2014, "Full Characterization of Photovoltaic Modules in Real Operating Conditions: Theoretical Model Measurement Method and Results," Prog. Photovoltaics: Res. Appl., **23**(4), pp. 443–461.

[43] Basheer, I., and Hajmeer, M., 2000, "Artificial Neural Networks: Fundamentals, Computing, Design, and Application," J. Microbiol. Methods, **43**(1), pp. 3–31.

[44] Zhang, G., Patuwo, B. E., and Hu, M. Y., 1998, "Forecasting With Artificial Neural Networks: The State of the Art," Int. J. Forecasting, **14**(1), pp. 35–62.
[45] Mellit, A., 2008, "Artificial Intelligence Technique for Modelling and Forecasting of Solar Radiation Data: A Review," Int. J. Artif. Intell. Soft Comput., **1**(1), pp. 52–76.
[46] Cornaro, C., Pierro, M., and Bucci, F., 2015, "Master Optimization Process Based on Neural Networks Ensemble for 24-h Solar Irradiance Forecast," Sol. Energy, **111**, pp. 297–312.
[47] Marquez, R., and Coimbra, C. F., 2013, "Proposed Metric for Evaluation of Solar Forecasting Models," ASME J. Sol. Energy Eng., **135**, p. 0110161.
[48] Lorenz, E., Remund, J., Müller, S. C., Traunmüller, W., Steinmaurer, G., Pozo, D., Ruiz-Arias, J. A., Fanego, V. L., Ramirez, L., Romeo, M. G., and Kurz, C., 2009, "Benchmarking of Different Approaches to Forecast Solar Irradiance," 24th European Photovoltaic Solar Energy Conference, Hamburg, Germany, Vol. 21, p. 25.
[49] Klucher, T. M., 1979, "Evaluation of Models to Predict Insolation on Tilted Surfaces," Sol. Energy, **23**(2), pp. 111–114.
[50] Gueymard, C. A., 2008, "From Global Horizontal to Global Tilted Irradiance: How Accurate are Solar Energy Engineering Predictions in Practice?," SOLAR 2008, San Diego, CA, American Solar Energy Society, pp. 1434–1456.
[51] Pierro, M., Bucci, F., and Cornaro, C., 2015, "Impact of Light Soaking and Thermal Annealing on Amorphous Silicon Thin Film Performance," Prog. Photovoltaics: Res. Appl., **23**(11), pp. 1581–1596.
[52] Huang, Y., Lu, J., Liu, C., Xu, X., Wang, W., and Zhou, X., 2010, "Comparative Study of Power Forecasting Methods for PV Stations," International Conference on Power System Technology, Oct. 24–28, Institute of Electrical and Electronics Engineers (IEEE).
[53] Lorenz, E., 2015, "PV Production Forecast of Balance Zones in Germany," IEA PVPS and SHC Workshop at EUPVSEC 2013: Solar Resource and Forecast Data for High PV Penetration, EA PVPS Task 14 and SHC Task 46, Paris.
[54] Gulin, M., Vašak, M., and Baotic, M., 2013, "Estimation of the Global Solar Irradiance on Tilted Surfaces," 17th International Conference on Electrical Drives and Power Electronics (EDPE 2013), pp. 334–339.
[55] Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A., 1997, "Radiative Transfer for Inhomogeneous Atmospheres: RRTM, a Validated Correlated-k Model for the Longwave," J. Geophys. Res.: Atmos., **102**, pp. 16663–16682.
[56] Fu, Q., and Liou, K., 1992, "On the Correlated k-Distribution Method for Radiative Transfer in Nonhomogeneous Atmospheres," J. Atmos. Sci., **49**(22), pp. 2139–2156.
[57] Oreopoulos, L., and Barker, H. W., 1999, "Accounting for Subgrid-Scale Cloud Variability in a Multi-Layer 1D Solar Radiative Transfer Algorithm," Q. J. R. Meteorol. Soc., **125**(553), pp. 301–330.
[58] Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B., 2005, "Fully Coupled 'Online' Chemistry Within the WRF Model," Atmos. Environ., **39**(37), pp. 6957–6975.