



## Progress in regional PV power forecasting: A sensitivity analysis on the Italian case study

Marco Pierro <sup>a</sup>, Damiano Gentili <sup>b</sup>, Fabio Romano Liolli <sup>b</sup>, Cristina Cornaro <sup>b, f</sup>,  
David Moser <sup>a</sup>, Alessandro Betti <sup>c, \*</sup>, Michela Moschella <sup>c</sup>, Elena Collino <sup>d</sup>, Dario Ronzio <sup>d</sup>,  
Dennis van der Meer <sup>e</sup>

<sup>a</sup> EURAC Research, Viale Druso 1, 39100, Bolzano, Italy

<sup>b</sup> University of Rome "Tor Vergata", Department of Enterprise Engineering, Via del Politecnico 1, 00133 Rome, Italy

<sup>c</sup> i-EM S.r.l, Via Lampredi 45, 57121, Livorno, Italy

<sup>d</sup> Ricerca sul Sistema Energetico - RSE S.p.A., Via R. Rubattino 54, 20134, Milano, Italy

<sup>e</sup> Mines ParisTech – PSL, PERSEE Center, Rue Claude Daunesse CS 10207, 06904, SOPHIA ANTIPOLIS CEDEX, France

<sup>f</sup> CHOSE, University of Rome Tor Vergata, Italy

### ARTICLE INFO

#### Article history:

Received 29 July 2021

Received in revised form

19 February 2022

Accepted 5 March 2022

Available online 7 March 2022

#### Keywords:

PV plants

Regional PV power Forecast

Upscaling forecast methods

### ABSTRACT

The increasing penetration of PV generation, driven by climate strategies and objectives, calls for accurate production forecasting to mitigate the negative effects associated with inherent variability, such as overgeneration, grid instability, supplementary reserve request. The regional PV power forecasting is crucial for Transmission and Distribution system operators for a better management of energy flows. In this work many aspects of regional PV power forecasting are investigated, by means of a comparison of six different forecasting models applied to predict the hourly production of the following days on six Italian bidding zones for one year. In particular, the work shows that the forecasting accuracy is mainly affected by the algorithm and its pre and post processing, with a range of 30% in performance accuracy, while it is less impacted by the forecasting horizon. It has been verified that the accuracy in the irradiation prediction, used in input to the power forecasting algorithm, has less impact compared to single plants. The work confirms the performance improvement which can be obtained by increasing the size of the area to which the prediction refers, through a comparison between the forecasting at bidding zone and national level. Finally, we show that the larger the controlled forecast area, the smaller the impact on the forecast accuracy due to the non-uniform spatial and capacity distribution of the PV fleet. This means that as the size of the region increases, the average irradiance progressively becomes the best PV power predictor. We refer to this phenomenon as: "input smoothing effect".

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The penetration of renewable energy sources (RESs) into the global energy sector is constantly increasing. This trend will characterize also the next decades, driven by the climate targets defined in agreements of multiple countries [1]. Despite the

forementioned agreements, photovoltaic (PV) power generation accounted for merely 3% of the global electricity generation in 2019 [2]. However, the instantaneous PV power contribution on the regional or national level can reach up to 100% during sunny weekends or holidays [3], causing increasing residual load/generation scheduling prediction errors that affect demand/supply irradiance both on transmission [4] and distribution [5] levels, voltage and frequency fluctuations due to overgeneration and subsequent reverse power flow events, grid congestions and electric loss, thereby potentially compromising stability and security of the system [6].

In order to mitigate voltage fluctuations and electric losses, it is important to balance supply and demand of electricity by adjusting controllable generators with respect to the expected RES

\* Corresponding author.

E-mail addresses: [marco.pierro@gmail.com](mailto:marco.pierro@gmail.com) (M. Pierro), [damianogentili-96@hotmail.com](mailto:damianogentili-96@hotmail.com) (D. Gentili), [fabio.liolli.FRL@gmail.com](mailto:fabio.liolli.FRL@gmail.com) (F.R. Liolli), [cornaro@uniroma2.it](mailto:cornaro@uniroma2.it) (C. Cornaro), [david.moser@eurac.edu](mailto:david.moser@eurac.edu) (D. Moser), [alessandrobetti53@yahoo.it](mailto:alessandrobetti53@yahoo.it) (A. Betti), [michela.moschella@i-em.eu](mailto:michela.moschella@i-em.eu) (M. Moschella), [Elena.Collino@rse-web.it](mailto:Elena.Collino@rse-web.it) (E. Collino), [Dario.Ronzio@rse-web.it](mailto:Dario.Ronzio@rse-web.it) (D. Ronzio), [dennis.van\\_der\\_meer@mines-paristech.fr](mailto:dennis.van_der_meer@mines-paristech.fr) (D. van der Meer).

generation. However, the variability of solar irradiance—mainly affected by cloud cover evolution—poses challenges when forecasting PV power generation, also because the coarse resolution of numerical weather prediction (NWP) models prohibits these models from resolving small-scale cloud formations [7]. In addition, the current generation of satellites cannot detect individual clouds because their spatial resolution is generally too coarse [8].

The increasing spatial distribution between PV systems, e.g., building-mounted or utility-scale, decreases the correlation between the power outputs and is referred to as the ensemble smoothing effect [9]. When handling regional PV power forecasting, a significant challenge is represented by the fact that small PV plants, contributing to a large share of the installed capacity, are unmetered and their exact location and peak capacity are unknown due to data protection legislation. Therefore, the aggregated power output has to be estimated, which is essential to stakeholders such as transmission system operators (TSOs), distribution system operators (DSOs) and energy traders [10].

The recent PV power forecasting literature gathers a wide variety of forecasting methods, the selection of which depends on the forecasting horizon, the inputs used and the forecasting technique adopted [11]. The PV power forecasting methods can broadly be classified as: (i) physical methods, based on sky- and satellite-imagery or NWP, that require a post-processing step to convert their output to PV power, and (ii) data-driven methods that map exogenous variables to the power output of the PV plant [12]. However, the majority of the proposed forecasting models focuses on the PV plant level [13], whereas analysis on large-scale control areas is rare in the literature [14].

Generally, regional PV power forecasting approaches can be classified as: (i) bottom-up approach, where the PV power generation in the area is predicted by summing the PV power computed for each PV site, and (ii) upscaling approaches, which can follow the so-called Models Output Average or Model Inputs Average methods (as classified in Ref. [10]). Models Output Average is mainly based on upscaling by using clustering, sub-setting of reference PV sites or a probabilistic approach based on a statistical analysis of the fleet of PV systems. Model Inputs Average instead directly forecasts power generation at regional level by means of exogenous NWP inputs aggregated at a lower level than regional, thus modelling the whole area as a single virtual power plant (VPP).

An early attempt using the Models Output Average approach was done in Ref. [15] where two different spatial distribution of the selected representative subsets were tested: a uniform distribution and a distribution that reflect the regional fleet distribution. Then the regional generation was upscaled from the selected subset. In Ref. [16], the same authors derived hourly regional PV power forecasts from solar irradiance forecasts issued by the European Centre for Medium-Range Weather Forecasts (ECMWF) by upscaling the aggregated output of PV systems that were preliminarily clustered based on their geographical location. In Ref. [17], it was assumed that a short time series of all PV systems was available from which representative sites were then chosen through a combination of k-means clustering and Principal Component Analysis (PCA), whose output was finally mapped by a regression method to the overall power generation in the area. The work in Ref. [18] extends [17] by presenting a probabilistic forecast model. Four upscaling methods were proposed in Ref. [19] to accommodate various scenarios of data-availability and plant information, where the representative sites were chosen based on stratified sampling of the installed capacities and locations. In Ref. [20] an upscaling approach based on reference systems was compared to satellite-derived estimates. The performance improvement achievable by combining the two data sources was analyzed and the Inverse Distance Weighting (IDW) method was exploited to

obtain the PV power output using the power output from nearby plants. Interestingly, upscaling resulted in similar results compared to the hybrid approach when 10 or more reference systems were used.

The main shortcoming of clustering and sub-setting methods is the requirement of the knowledge of PV site coordinates and their respective peak capacity or cluster peak capacity, which are often unknown. Therefore, [21] proposed a novel method of spatial clustering of PV systems based on Voronoi Decomposition which needs only the installed capacity of solar PV connected to bulk supply points and their location.

Concerning the Model Inputs Average approach, [22] represented the whole area as a single VPP, forecasting aggregated power generation by means of a machine learning based approach, thus only requiring the overall installed capacity, as well as aggregated historical PV power measurements. In Ref. [23], PV generation of four Japanese regions was predicted by applying a SVR model to NWPs previously preprocessed by PCA to remove redundant information. In Ref. [24], the authors combined a convolutional neural network (CNN) and a long short-term memory (LSTM) neural network to forecast the aggregated PV power output of Germany. By using multi-channel CNN, the authors could pass NWP maps of each variable to extract spatial features that were later passed to the LSTM neural network to extract temporal features. In Ref. [25], a first attempt to directly predict wind and solar generation in each Italian bidding zone has been conducted through an upscaling method similar to the one adopted in this work. Unfortunately, the authors tested the method only on a few months normalizing the errors by the maximum power produced in the month and not by the installed capacity. This prevents a comparison of the accuracies achieved in Ref. [25] and in the present work.

In [10,26,27] hybrid upscaling strategies between the two above mentioned approaches has been tested. In Ref. [26] clustering methods were used for spatial grouping of PV plants and then the power output of each cluster was considered produced by a virtual PV plant and directly predicted by deterministic or machine learning models. In Ref. [10], the authors clustered the PV systems of a region and compared two approaches: (i) similar to Ref. [26], they obtain the regional PV power forecast by averaging the forecasts of each cluster considered as a VPP (Models Output Average on regional scale), and (ii) use the inputs of each centroid to directly forecast the regional PV power (Model Inputs Average), the latter of which resulted in slightly better accuracy. A hybrid, but probabilistic approach, could be found in Ref. [27]. Authors first divided the PV fleet in different clusters on the basis of the plant position, on each centroid produced the forecast considering 22 different possible plane of the array, finally they found the parameters of a multi-linear regression to fit the regional generation combining all these trajectories (#centroids x 22). The best parameters distribution was estimated using a Bayesian statistic approach to provide a probabilistic forecast.

Studies into the source of errors, e.g., the effect of spatial interpolation, have also been performed. For instance Ref. [28], presented a sub-setting based approach using the IDW method for power estimation at test plants, based on reference PV sites. In particular, the source of errors incurred by upscaling was separated into the components caused by spatial interpolation and by the aggregation of the interpolated values. It was shown that the spatial interpolation errors are mainly caused by differences between the parameters of the reference and test PV systems, as well as meteorological conditions. Regarding instead the aggregation of interpolated values, the authors showed that an increasing number of reference systems improves the accuracy of the regional forecasts, although there is a saturation effect where the accuracy increase becomes smaller. A similar study was conducted for Japan in Ref. [29], without any spatial weighting. By repeating the

experiment 10,000 times and randomly selecting 200 out of 2219 PV sites, a highly concentrated accuracy was achieved, indicating thus that spatial distribution becomes less relevant when the number of reference systems is large. In Ref. [30], the average and IDW spatial interpolation were compared, where the latter resulted in slightly better accuracy.

Bottom-up approaches are uncommon in the literature due to the computational overhead and the lack of system-information. In order to overcome this issue, a probabilistic bottom-up approach was proposed in a study conducted in Germany, where the location of PV sites was available together with the corresponding peak capacity [28]. Since the power output of PV systems is significantly affected by tilt and azimuth, they proposed to construct joint distributions of panel orientation binned on the nominal capacity, in order to define weights based on their frequency. The power at any location was finally estimated by using a weighted average of all possible configurations, which represent an ensemble of estimates of PV power production. As aforementioned, [27] proposed a Bayesian method to retrieve the parameters matrix that takes into account the plane of the arrays of the regional fleet.

An alternative approach to reduce the uncertainty related to unknown tilt and azimuth distribution of a regional fleet was proposed in Ref. [31]. The authors introduced the concept of “equivalent” Plane of the Array of a VPP, i.e., the tilt and orientation that provide, by means of a physical-based model, the outperforming estimation of PV distributed generation. They showed that the Italian PV generation can be considered a south oriented VPP with a tilt around  $10^\circ$ .

However, if the systems are known and the VPP is not too large, e.g., a combination of utility-scale PV plants, then a bottom-up approach is feasible as exemplified in Ref. [32]. There, the authors proposed a CNN that features an input layer and a convolution-pooling layer for each PV plant to generate forecasts for the aggregated output of the 10 plants. While this is feasible for a relatively small number of plants, i.e., 10 in Ref. [32], the authors did not comment on the scalability of their forecast model.

In this paper we dive deeply into the problems of regional solar forecasting, studying the impact on the forecast accuracy: (1) of the upscaling method itself; (2) of the prediction horizon (1–2–3 days-ahead); (3) of the quality of the NWP; (4) of the adopted upscaling approach: model outputs average at regional scale vs model inputs average; (5) of the forecast-controlled area; (6) of the irradiance/capacity non-uniform spatial distribution; and (7) of the machine learning training complexity.

In this work different up-scaling methods based on machine learning with VPP approach have been applied to forecast the regional generation of six Italian Market Zones which correspond to controlled areas defined by the Italian TSO, spanning from  $24 \times 10^3$  to  $119 \times 10^3$  km<sup>2</sup> up to three days ahead. All the methods work with no information on single PV plants pertaining the zone, and test a variety of pre-processing, NWP data used as forecast model inputs, aggregation methods adopted for input dimensionality reduction, PV forecast models, as well as post-processing correction.

The study has been carried out under the auspices of the International Energy Agency’s PV Power Systems (PVPS) Task 16 and deepens and extends the preliminary research conducted in Ref. [33]. Indeed, in this work, we enlarge the analysis in Ref. [33] both on market and national scales using additional NWP data and other upscaling methods. Some results confirm the founding reported in Ref. [33], others are completely new and provide useful information on critical aspects of “model input average” approach to regional PV forecasting.

Overall, the work is configured as a comprehensive sensitivity study on relevant aspects that affect the performance of regional PV power forecast obtained by a “model input average” approach,

some of which have been individually investigated in the aforementioned literature. In Ref. [22] authors compared different ML models for regional PV power forecasting, but they did not need any post-processing as they used for training/test the models the sum of high quality production data coming from plants generation monitoring. In contrast, we highlight the importance of post-processing as we used regional PV generation data estimated by TSO (surely affected by additional uncertainty). More generally, we focused our attention on all possible steps of an upscaling process rather than on the forecast performance of the ML models itself. In Ref. [23], the authors showed the advantage of using preprocessing based on PCA to reduce the dimensionality of the NWP input. Similarly, authors of [28,29] studied the amount of reference plants and their spatial distribution that should be upscaled when an output model average approach is adopted. However, these authors did not analyze how this input dimensionality/number of references plants scales with the size of the controlled area, i.e. how the non-homogeneous spatial distribution of irradiance and PV capacities affect forecast performance at different spatial scales. Then, we also proposed a method to quantify the trade-off between the amount of information provided to a ML model and its training complexity. These are one of the most important results of our work. In addition, to the best of the authors’ knowledge, there are no works evaluating the impact of NWP quality on the accuracy of regional PV forecasts and this is the other main goal of our investigation. Finally, as in Ref. [26], we compared the accuracy of the output model average applied on regional scale with the simplest model input average, evaluating the computational cost-benefits of the two approaches. These are the main novelties of the work with respect to the existing literature.

The remainder of this paper is organized as follows: the used data are described in Section 2; the methodology adopted in the investigation is reported in Section 3; the upscaling methods are briefly described in Section 4; the description of the metrics used to evaluate the forecast can be found in Section 5; the results are discussed in Section 6; finally, conclusions in Section 7.

## 2. Data

### a. Regional PV power output

The work has been carried out using PV generation data at bidding zone level for the Italian territory. These data, freely available on the Italian Transmission System Operator (TSO) website, consist of the PV hourly production, based on Terna’s (TSO) measurements and estimates [34]. The dataset covers the period 2016 ÷ 2017. The year 2016 has been used to train the algorithms, while 2017 represents the test dataset. During that time, the Italian Electrical Market consisted of six bidding zones, represented in Fig. 1.

In order to perform the analysis, power data have been normalized to the installed capacity of each area (supposing a linear growth with time).

The Italian bidding zones differ both in terms of size and PV installed capacity. In particular, with reference to the 19.7 GWp of installed capacity at the end of 2017, 44% pertains to Northern Italy, 12% to the CNOR area, 14% to Central-Southern Italy, 19% to the South, 7% to Sicily and 4% to Sardinia.

### b. Numerical weather prediction data

The Numerical Weather Prediction variables (NWPs) that have been ingested into the developed upscaling methods are: the global horizontal irradiance (GHI) and the 2-m air temperature (Tair). The time series covered the period 2016–2017. Two data sets of these



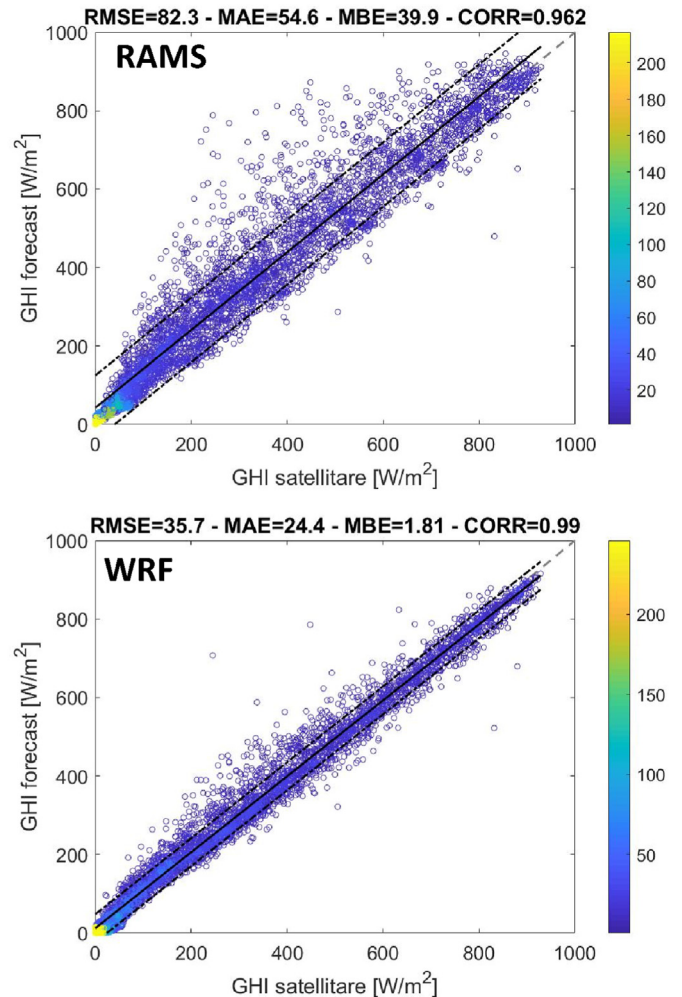
**Fig. 1.** Italian market areas during the period 2016–2017. Each colour represents a single bidding zone. Green: Northern Italy –(NORD), violet: Centre-North (CNOR), yellow: Centre-South (CSUD), orange: South (SUD), light blue: Sardinia (SARD), red: Sicily (SICI).

variables have been used.

The first data set derives from the Regional Atmospheric Modelling System (RAMS) [45]. This local area model has been initialized with boundary conditions supplied by the global area model Integrated Forecast System (IFS), developed by the European Centre for Medium Range Weather Forecasts, UK (ECMWF) [35]. For this study, RAMS has been run once a day, starting at 12 UTC. Each run has provided the forecast for the following three days, with 4.5 km of spatial resolution and 1 h of time resolution over a domain covering the Italian territory. The RAMS point forecast has been aggregated at province level. The representation of the Italian provinces is shown in Fig. 1.

The second data set was obtained by the Weather Research and Forecasting model (WRF-ARW 3.8) [36] with an initialization at 12 UTC of the current day, to produce the 24 h forecasts starting from the following 00 UTC of day to be predicted (as for the RAMS data). The initial contour data for model initialization come from the GFS model output. The spatial resolution is of 12 km × 12 km covering the whole Italian territory (1325 grid points). The irradiance forecast systematic errors generated by the WRF Radiative Transfer Scheme have been corrected by a physical based Model Output Statistic that is the last upgrade of the one described in Ref. [37].

Fig. 2 compares the scatterplots of day-ahead irradiance forecast vs the satellite derived irradiance averaged over the Northern market zone (in which the 44% of the PV capacity is installed) generated by RAMS and WRF models. It has to be specified that the WRF irradiance predictions were post-processed, while the RAMS predictions are the simple model output.



**Fig. 2.** Scatter plots of day-ahead irradiance forecast vs the satellite derived irradiance averaged over the Northern market zone.

### 3. Methodology

To assess the range of accuracy that can be achieved by different upscaling methods that make use of the same NWP, six different upscaling methods have been compared. All the methods have ingested RAMS NWP prediction at Italian province level with one day-ahead horizon. In addition, for the two best performing methods, the forecast up to three days in advance was analyzed to investigate performance variations as a function of the forecast horizon. Subsequently, we have investigated how much the power forecast accuracy is affected by the quality of the NWP, comparing the accuracy of the outperforming model fed with RAMS or WRF. These analyses have been conducted for each individual market area.

For what concerns the identification of the outperforming method between model inputs average and model outputs average, we have also computed the day-ahead PV power forecast at national level following the two different approaches. To this end, we have built a new model fed with WRF irradiation prediction on the 1325 model grids points, to directly forecast the whole Italian solar generation (model inputs average), compared with another one where the generation derives from a weighted average of the market areas solar predictions previously obtained by WRF (models output average). Then, we have verified how the forecast accuracy can be improved by enlarging the prediction footprint from the

market zones to the whole Italy i.e., we have evaluated the forecast smoothing effect. Finally, we have investigated the impact on the forecast accuracy of the irradiance/PV capacity non-uniform spatial distribution depending on the size of the controlled area, which is what we refer to as input smoothing effect. To this end, we have computed the accuracy of the outperforming method by varying spatial input granularity of WRF irradiance and the size of the controlled area.

At the same time, we assessed what we defined the “machine learning (ML) complexity limit”, also known as the number of intra-correlated input features that can be ingested by a ML algorithm without definitively compromising the forecast accuracy because of the increasing complexity of the training phase and redundant features. For this purpose, we have compared the accuracy reached by our ML outperforming method with the accuracy achievable by a statistical data-driven method that can virtually ingest an infinite number of features without losing accuracy.

Fig. 3 summarizes the objectives of our study and the size of the forecast-controlled areas over which our investigation was carried out.

#### 4. Upscaling methods

While PV power forecast at the single plant level is essentially based on a forecasting model (such as physical based or data-driven models) that maps NWP outputs into power generation, the structure of an upscaling method could be much more complex (Fig. 4). It may involve different steps and items: NWP prediction provided at a given spatial/temporal resolution and horizon, data pre-processing, the power forecasting model and output post-processing.

NWP models are always the essential element for the short-term forecast (with a horizon greater than 4–5 h ahead). Data pre-processing could be needed to reduce NWP bias by Model Output Statistic algorithms (MOS), to filter solar generation outliers or unrealistic data and, most importantly, to decrease the number of input features by aggregating the NWP output with different techniques (average, clustering, PCA, etc.). The forecast model itself that must map the NWP into PV regional generation plays a much more important role than in the case of the single site prediction. Finally, the power output post-processing can correct PV power estimation errors (Italian TSOs/DSOs do not meter the solar generation of all the plants in the controlled area directly), due to missing information regarding the PV installed capacity (which, in the regional case, is no longer constant but grows with time and is usually updated monthly/annually), missing information on fleet location, plane of array, etc. and finally to correct the bias of the forecast model output. The post-processing procedure is usually

most effective when applied in real time by means of a moving window involving the most recent historical generation data.

We developed six different up-scaling methods, briefly described as follows:

##### a. Analog Ensemble

Analog Ensemble (AE) is not a machine learning algorithm, but a statistical technique based on the principle of similarity. The technique compares past NWP forecasts (the analogs) to the NWP forecast. The observed power associated with the most similar analogs is then used to build a probability distribution. Consequently, the assumption is that similar weather patterns and subsequent power generation repeat themselves. The similarity is evaluated by minimizing the distance between the current and the past predictors. Two different AE forecast upscaling methods have been developed. What follows is a description of their main characteristics.

##### a1. Analog Ensemble 1

The first AE model compares weather forecasts from the RAMS model. In this case, we only aggregate the features at the province levels. Furthermore, we apply the MOS correction to the RAMS model output, similar to the technique used in Ref. [38]. Afterwards, it is possible to compare the current forecast with the analogs using:

$$d_{Abs}(t, t') = \sum_{v=1}^{N_v} \frac{1}{\sigma_v} |F_{v,t} - A_{v,t'}| \tag{1}$$

where  $v$  represents each variable of the set of  $N_v$  predictors,  $F_{v,t}$  the  $v$ -th component of the array of the current forecast at time  $t$ , and  $A_{v,t'}$  the  $v$ -th component of the past forecasts on the same temporal horizon of  $F_{v,t}$  and  $\sigma_v$  is the standard deviation of each variable calculated in the historical period. The forecast model is similar to the one proposed by Alessandrini et al. [39]. In this case, the predictors are the GHI and the cell temperature derived by the standard equation:  $T_{cell} = T_{air} + \frac{NOCT - 20^\circ}{800} GTI$  (where NOCT is the Nominal Operating Cell Temperature specific to the module type). Analog ( $A_{v,t'}$ ) are sought in the historical year ( $t' \in 2016$ ) selecting past situations with a zenith angle of  $15^\circ$  around the zenith angle of the current forecast. Then, the observed power corresponding to the first 15 analogs with the smallest distance  $d_{Abs}(t, t')$  are selected in order to build the probability density function (pdf) of future power. For this work the mean of the pdf has been used as the final power forecast:



forecast controlled area	objectives of the investigation
 <p>PV power forecast at market zones level</p>	<ul style="list-style-type: none"> <li>a. Accuracy benchmark of different upscaling methods</li> <li>b. Impact of forecast horizons (1/2/3 days ahead)</li> <li>c. Impact of NWP accuracy</li> </ul>
 <p>PV power forecast at National level</p>	<ul style="list-style-type: none"> <li>d. Comparison of different approaches for PV power prediction</li> <li>e. Smoothing effect of the output of the upscaling method</li> <li>f. Input smoothing effect and machine learning complexity limit</li> </ul>

Fig. 3. Forecast controlled area and objectives of the investigation.

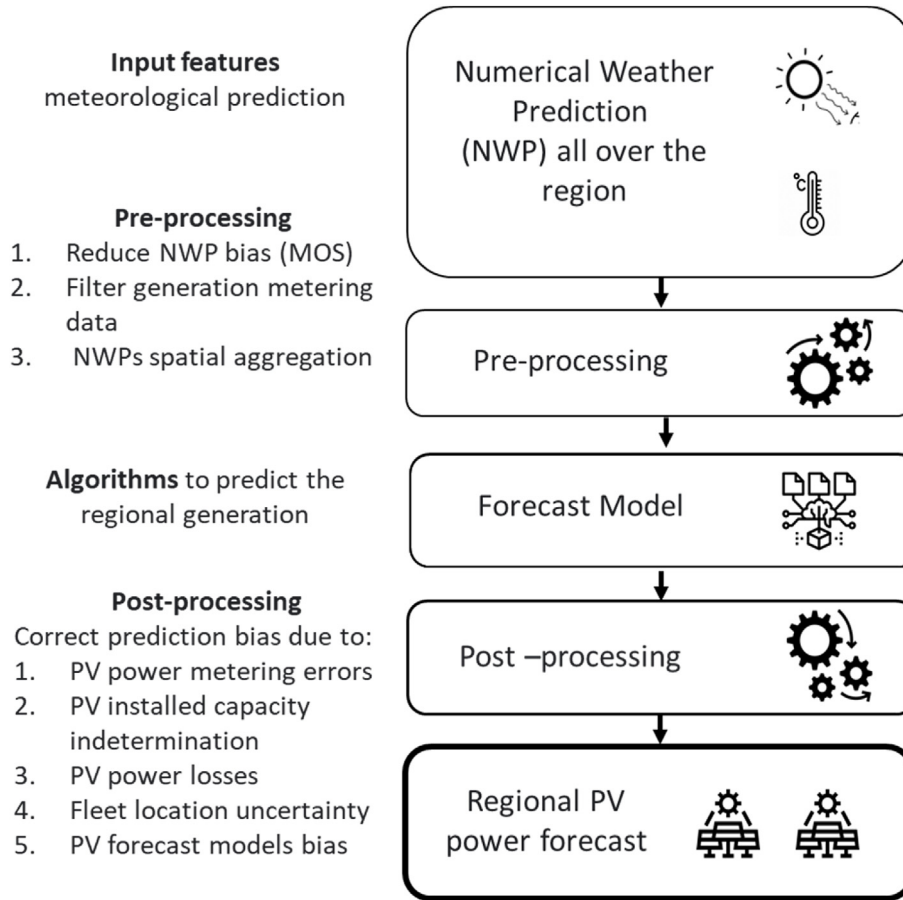


Fig. 4. Structure of forecast up-scaling method.

$$P_{PV}^{for}(t) / P_n(t) = \langle P_{PV}(t' | \min\{d_{Abs}\}) / P_n(t') \rangle, \quad (2)$$

where  $P_{PV}$  is the regional power output and  $P_n$  is the regional installed capacity.

a2. Analog Ensemble 2 (AE2)

The second AE model compares weather forecasts from the RAMS/WRF model. Furthermore, we pre-process the irradiance forecast using an upgraded version of MOS described in Ref. [37]. In addition, we test several kinds of feature aggregations depending on the used NWP forecasts and the forecast control area (see Tables 2 and 3). Since features can have varying importance, we adjust the analog ensemble distance metric as follows:

$$d_W(t, t') = \sum_{p=1}^{N_p} W_p \| \hat{\mathbf{F}}_{p,t} - \hat{\mathbf{A}}_{p,t'} \|, \quad (3)$$

where  $p$  indexes the predictor,  $\hat{\mathbf{F}}_{p,t}$  is a vector composed of the current forecast (at time  $t$ ) of the  $p$ -th predictor type (normalized by min/max values of the historical data set), and  $\hat{\mathbf{A}}_{p,t'}$  is the normalized past forecasts vector (at time  $t'$ ) of predictor  $p$  and  $W_p$  is the optimal weight associated to predictor  $p$  ( $\sum_p W_p = 1$ ). In this case, the predictors are the GHI,  $T_{air}$  and sun elevation and the

Table 1 Metrics for accuracy assessment.

acronym	Forecast technique
Forecast error [W/m <sup>2</sup> or % of P <sub>n</sub> ]	$e_h = (X^{FOR}(h) - X^{OBS}(h))$
Pearson correlation [-]	$CORR = \frac{COV(X^{FOR}, X^{OBS})}{\sigma_X^{FOR} \sigma_X^{OBS}}$
Root mean square error [W/m <sup>2</sup> or % of P <sub>n</sub> ]	$RMSE = \sqrt{\frac{\sum_{h=1}^n e_h^2}{n}}$
Mean absolute error [W/m <sup>2</sup> or % of P <sub>n</sub> ]	$MAE = \frac{\sum_{h=1}^n  e_h }{n}$
Mean bias error [W/m <sup>2</sup> or % of P <sub>n</sub> ]	$MBE = \frac{\sum_{h=1}^n e_h}{n}$
Skill score [%]	$SS = 100 \left( 1 - \frac{RMSE^{forecastmodel}}{RMSE^{referencemodel}} \right)$

optimal weights are found by means of trial and error: 0.8, 0.1, 0.1, respectively. The historical period has a moving window of the past 30 days. At each time  $t$ , the first 25 analogs which have the lowest distance  $d_W$  are selected and then the power forecast is computed as:

$$P_{PV}^{for}(t) / P_n(t) = \frac{1}{25} \sum_{i=1}^{25} b_n P_{PV}(t^i | d_W(t, t^i)) / P_n(t^i) \quad (4)$$

where  $b_1 = 0.2$  for the analog with the smallest distance and  $b_n = (1 - b_1) / 24$  for the remaining analogs.

**Table 2**  
Methods main features.

acronym	Forecast technique	Input
PM	Simple persistence	Past regional PV power output
DM	Deterministic physical based model	RAMS prediction of GHI and Tair and AOI averaged over the market zone
kNN/QRF	Blending of k-nearest neighbors and Quantile Regression Forest	RAMS prediction of GHI or Kcs and Tair and sun elevation aggregated at province level
AE1	Analog Ensemble	
AE2	Analog Ensemble	
DM/NN	Hybrid method based on physical/NN models	

**Table 3**  
Methods main features.

acronym	Forecast approach	Input
1DM/NN	Model input average	WRF NWP aggregated in 20 clusters
6DM/NN	Model output average	WRF NWP aggregated by market zones provinces

**b. Deterministic method**

The third model is a deterministic model that uses RAMS/WRF weather forecasts as input features. Two types of pre-processing procedures are used: (1) using an upgraded version of MOS as previously [37], and (2) areal averages of the weather forecasts over the NWP grid points of the target area. The optimal equivalent POA is near south and the optimal equivalent tilt is 10° as tested on training data [31]. The forecast model itself is a power estimation model described in Ref. [40]. It is a parametric model that maps GHI,  $T_{air}$  and the equivalent angle of incidence (AOI) into the regional PV power output:

$$P_{PV}^{for0}(t) / P_n(t) = f_{PV}^{DM}(AOI, GHI^{for}, T_{air}^{for}) \quad (5)$$

Since the predictors are the simple regional averages, the model does not take into account any input spatial variability and non-uniformity of capacity distribution. Finally, we correct the forecasts using a performance factor (PRF), which is computed on 15 days past data (15 days moving window) as the average of the ratio between daily solar generated energy and the daily energy forecast:

$$PRF(dd + H) = \frac{1}{15} \sum_{dd'=1}^{15} \left( \frac{\sum_{t'=1}^{24} P_{PV}(t' | dd - dd')}{\sum_{t'=1}^{24} P_{PV}^{for0}(t' | dd - dd')} \right) \quad (6)$$

where  $dd$  is the current day, while  $dd + H$  is the day to predict. This factor accounts for the TSO's errors in estimating daily solar generation, outages, power losses and plant degradation. The regional generation forecast is then obtained as:

$$P_{PV}^{for}(t) / P_n(t) = PRF \cdot P_{PV}^{for0}(t) / P_n(t) \quad (7)$$

**c. Hybrid method based on deterministic and MLP neural network models**

Similar as the third model, the fourth model uses weather forecasts from RAMS/WRF as input features, as well as the same pre-processing step, i.e., areal averages of the weather forecasts over the NWP grid points of the target area. The forecast model is similar to the model described in Refs. [31,33], thus it is a hybrid forecast method based on several steps: (1) the correction factor PRF is computed by the previously mentioned deterministic model and post-processing step; (2) the deterministic model is applied to

the clear sky irradiance  $GHI^{CS}$  [41], subsequently corrected by the PRF to obtain the clear sky regional generation (8) and the PV clear sky index (9):

$$P_{PV}^{CS}(t) / P_n(t) = PRF \cdot f_{PV}^{DM}(AOI, GHI^{CS}, T_{air}^{for}) \quad (8)$$

$$K_{PV}(t) = P_{PV}(t) / P_{PV}^{CS}(t) \quad (9)$$

$K_{PV}$  represents the changes in PV generation while removing the deterministic trend that occurs during clear sky conditions; (3) An ensemble of neural networks [42] (single-layer perceptron- MLP-) is used to forecast  $K_{PV}$  starting from the predicted clear sky index ( $K_{cs}^{for} = GHI^{for} / GHI^{CS}$ ) with different spatial aggregation (depending on the controlled area). Therefore, the final regional prediction results from:

$$P_{PV}^{for}(t) / P_n(t) = [P_{PV}^{CS} ANN(K_{cs}^{for})] / P_n(t) \quad (10)$$

The NN ensemble has been trained on the 2016 data set and tested on the 2017 data set. In this case, the prediction model takes into account the non-uniform distribution of irradiance and photovoltaic capacity because the NNs will automatically select, throughout the controlled area, the most representative  $K_{cs}^{for}$  values.

**d. Linear blending of K-nearest neighbors and Quantile Regression Forest models (k-NN/QRF)**

The fifth and final advanced forecast model uses RAMS weather forecasts as input features. The historical samples of the training set were pre-processed according to two main steps. (i) First, samples with missing information are removed from the training set; (ii) then, outliers are removed by making a least-squares fitting in the plane defined by ( $\langle GHI_{sat} \rangle, P$ ), where  $\langle GHI_{sat} \rangle$  is the average satellite irradiance over each control area, and  $P$  is the related hourly power, and discarding all samples falling out of a safety cone centred on the linear fitting. Furthermore, there may be missing information regarding the exogenous variables in the testing set. For this reason, a two-step imputation is conducted. Specifically: (1) first, a k-NN is exploited to fill missing values for those records that contains at least one variable's value available; (2) residual missing data are imputed according to a cubic spline interpolation.

The forecast model is a linear blending of k-NN and Quantile Regression Forest (QRF) [25]. k-NN is a non-parametric similarity method which looks for past instances with similar weather conditions, selecting the k most similar historical samples and combining the corresponding historical powers with a weight depending on the degree of similarity. We use the Euclidean metric as distance measure and the hyperbolic kernel for the combination of k nearest neighbors. In addition, only the past k nearest neighbors close to the hour and month of the specific power instance to be predicted are selected for power forecasting.

The QRF algorithm is a variant of the well-known Random Forest (RF) introduced by Breiman [43] which provides the full conditional

distribution of the response variable. Specifically, assuming that  $Y$  is the target variable and  $X$  the vector of predictors, then the final goal is finding the relationship between  $X$  and  $Y$ . Supposing a certain level  $\alpha$  ( $0 < \alpha < 1$ ), QRF estimates the conditional quantile

$$Q_{\alpha}(x) = \inf\{y: F(y|X = x) \geq \alpha\}$$

where  $F(y|X = x)$  is the conditional distribution function of  $Y$  given  $X = x$ .

The exogenous inputs for k-NN and QRF are  $\{GHI, GHI_{CS}\}$  and  $\{GHI, GHI_{CS}, \text{Month}, \text{Hour}\}$ , respectively. The weights of k-NN and QRF in the models blending (0.5 and 0.5, respectively) are tuned according to a trial-and-error approach on a validation dataset not used in the present work.

e. Reference model

We use a simple persistence model to evaluate the forecast performance improvement of the aforementioned forecast models. The persistence model is defined as

$$P_{PV}^{for}(t) / P_n(t) = P_{PV}^{for}(t | dd) / P_n(t | dd) \tag{11}$$

5. Accuracy metrics

Table 1 reports the main metrics used in forecast literature to evaluate the non-probabilistic forecast accuracy, where  $X$  is the variable that should be predicted (irradiance or PV power output),  $n$  is the number of yearly sun hours.

6. Results

In this section we report and discuss the results obtained studying different topics that affect the accuracy of upscaling methods for regional PV power forecast.

In this work, the developed upscaling methods were applied 65 times by varying the NWP, input features, and feature aggregation according to the specific analysis, while the technologies are the ones described in section 4. For this reason, at the beginning of each following sub-section, if needed, we provide a table reporting the acronym, the forecast technique and the input description.

a. Accuracy benchmark of different upscaling methods.

First of all, we wanted to understand how the forecast accuracy was affected by the upscaling method itself, thus, all the developed methods (listed in Table 2) were fed with the same NWP: day-ahead RAMS forecasting aggregated at province level. The investigation was performed for each market zone.

Fig. 5 shows the accuracy values of each method averaged over all the market zones.

As expected, all the methods outperform the persistence model with an improvement (skill score) that ranges between 15% and 36%. The maximum average RMSE (6.65%) was reached by the KNN/QRF method that indeed is strongly biased (2.75% of MBE). The minimum average RMSE (5%) is obtained by the ANN hybrid method.

The physical based method (DM) and the analog ensemble (AE1) approaches perform similarly with a RMSE between 5% and 5.5%.

It is worth noting that the RMSE of the forecasts can vary up to 30% according to the used upscaling approach, therefore, pre-processing, forecast models and post-processing should be suitably tuned to reach the best regional PV power forecast.

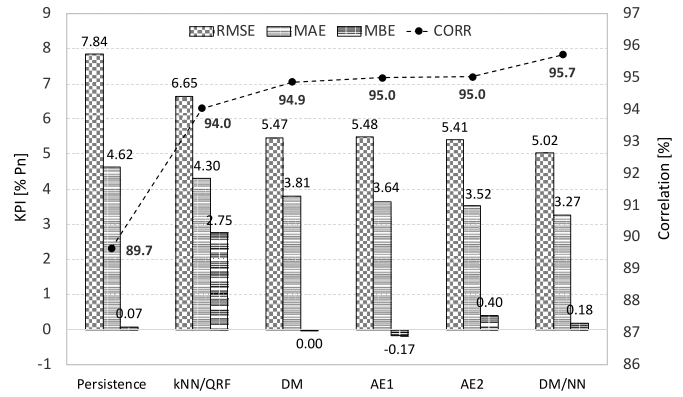


Fig. 5. Market zones average accuracy of the different upscaling methods.

For example, we must point out that the methods DM, AE2 and DM/NN make use of a moving window post-processing and this could be the reason why DM/NN, in average, outperforms KNN/QRF and AE1.

The same 30% of RMSE variability (from 5% to 3.7%) has been obtained in a similar accuracy benchmark of the whole 2016 Italian generation forecast [33]. This confirms that in regional PV power prediction, there is wide room for performance improvement related to the upscaling methods used.

b. Impact of forecast horizons (1/2/3 days ahead)

To assess which is the role that the prediction horizon plays in regional short-term PV forecasting, we applied the methods reported in Table 2 using the two and three day-ahead RAMS prediction.

Fig. 6 shows that the NWP horizon does not significantly affect the accuracy of the regional forecast in any of the market areas.

c. Impact of the NWP

To understand how much the accuracy of regional forecast could depend on the quality of the ingested NWP, we applied the same outperforming method DM/NN, previously fed with RAMS NWP, also to WRF 1 day-ahead predictions.

In the single PV system, incident irradiance accounts for 80% of the energy production, so, for example, improving NWP performance by 50% should produce around 40% increase in the accuracy of PV output prediction. In contrast, different forecasting models that map irradiance prediction into PV generation could account for around the 15% of RMSE improvement [44]. Therefore, in a single-site forecast, the quality of NWP takes the lion's share of the achievable accuracy.

Fig. 7 shows that in regional PV power forecasting, an average improvement of about 50% in NWP accuracy translates, on average, into an increase of only 20% in the performance of the resulting market zones PV generation forecasting. On the other hand, we have just demonstrated that more effective upscaling algorithms could lead up to a 30% improvement in regional PV forecast accuracy.

These differences between site and regional solar forecast depend on the additional uncertainties introduced by: PV power output estimation errors, shutdown/under-performance of PV plants, long term power degradation (that on regional scale could be much important than for a single site), distributed power loss, missing information on the PV installed capacity, missing information on fleet location/technology, plane of array, etc. as well as irradiance spatial variability and non-uniform capacity distribution that directly affect the NWP aggregation methods.



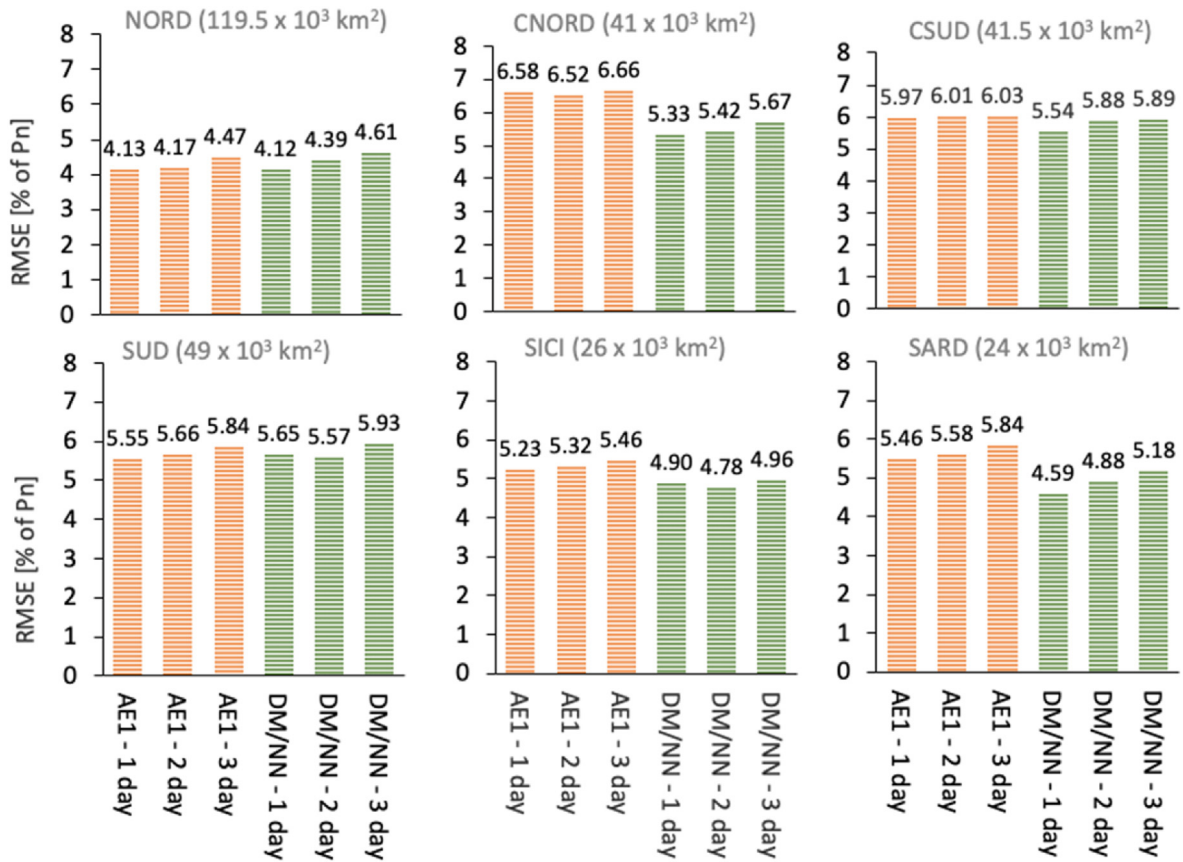


Fig. 6. RMSE of the AE1 and DM/NN forecasts from 1 to 3 days ahead.

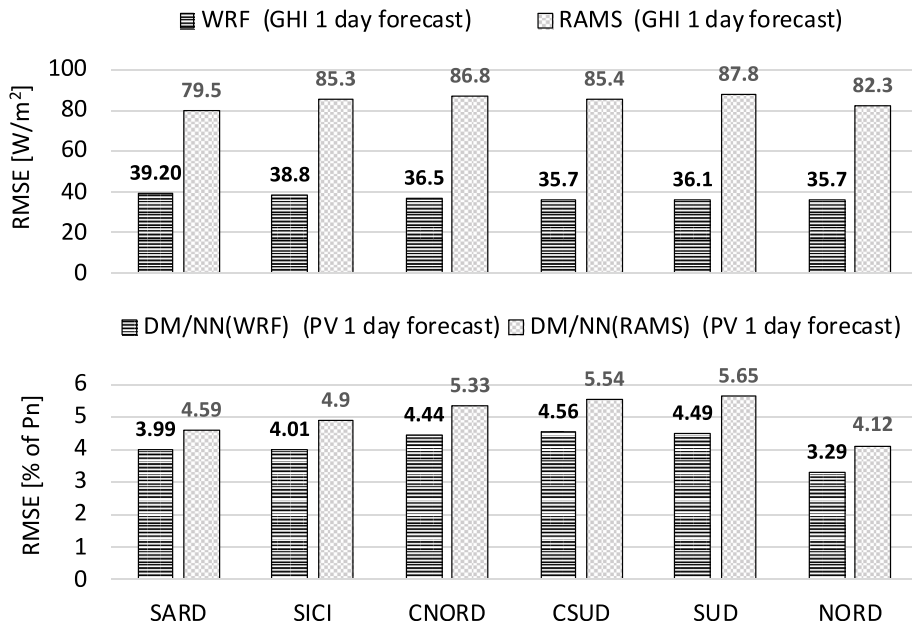


Fig. 7. Comparison of the accuracy of the two GHI NWP (forecast vs satellite irradiance) and the accuracy of the PV generation forecast of each market zone obtained by the DM/NN method ingesting the two different NWP. In the figure, the market zones are listed in order of increasing area.

Thus, to reduce these input/output uncertainties, in addition to the prediction model itself, data pre/post-processing could play an essential role.

d. Comparison of different approaches for PV power prediction at national level

In this work we have also analyzed if, to predict the national solar generation, it is more effective to use 6 forecasts of the market zones productions (model output average at regional level) rather than a single forecast that directly predicts the whole Italian production (model input average). For this purpose, on one hand, we computed the average (weighted with the capacity) of the 6 zonal forecasts (DM/NN) previously provided to obtain the Italian solar power prediction (namely 6DM/NN). On the other hand, we have input into the DM/NN model the WRF predictors over the whole Italy to predict directly the national generation (namely 1DM/NN). To set up this upscaling method, as in Ref. [28], we tested different types of spatial aggregation of the 1325 predictor time series (uniformly dispersed across the country). In this case, we found that grouping the irradiance in 20 different clusters provides the most accurate prediction. Table 3 summarizes the main features of the methods.

As it was found in Ref. [10] on smaller area, the two approaches provided predictions with very similar accuracy, although the model output average (6DM/NN) slightly outperforms the model input average (1DM/NN) with 5% of SS. Nevertheless, the latter approach is more straightforward and easier to implement (Table 4).

e. Smoothing effect of the output of the upscaling method

It is well known that enlarging the forecast-controlled area produces an accuracy improvement. This phenomenon, known as forecast smoothing effect, depends on the increasing decorrelation of the forecast errors which is achieved enlarging the prediction footprint [9,33]. However, expanding the area controlled by the

forecast means improving transmission capacity i.e., removing network constraints/congestions between neighboring areas.

Fig. 8 shows the benefit in terms of regional forecast accuracy that can be achieved by advanced upscaling methods and national transmission grid reinforcement in the Italian case.

- (1) Considering the market zone, on average, passing from a simple power persistence prediction to the DM/NN forecast the RMSE reduces from 7.84% to 4.13% of the installed capacity. Therefore, advanced regional forecasting techniques can improve prediction accuracy up to 47.5%. In this case, the RMSE results from RMSE realized in each market zone, weighted with the installed capacity, that corresponds to the forecast that can be achieved on a controlled area of  $50 \times 10^3 \text{ km}^2$  (market zones average surface).
  - (2) Ideally, removing transmission capacity bottlenecks between market areas, thus allowing for the forecasting of solar generation of the entire country ( $300 \times 10^3 \text{ km}^2$ ), would reduce the RMSE from 4.13% to 2.22%. Consequently, strengthening the national transmission grid could lead to an additional 46.5% of improvement in forecast accuracy.
- f. Input smoothing effect and machine learning complexity limit

The last part of our work aims to investigate two different issues, both related to the number of input features of the forecast method:

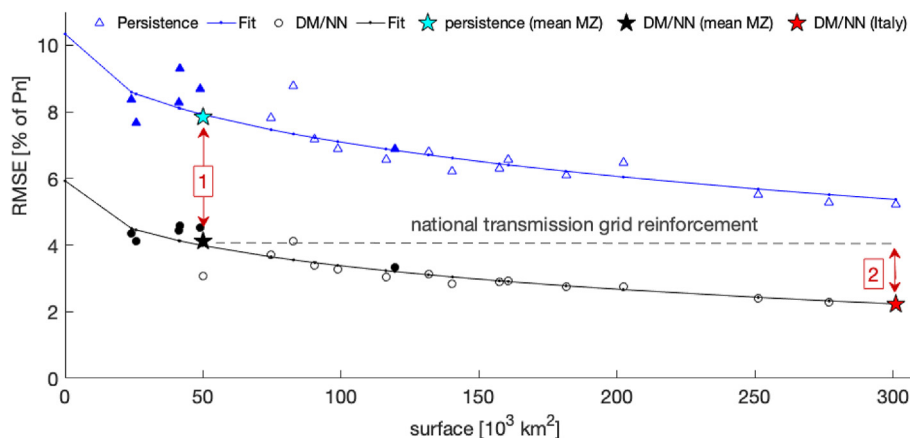
1. To evaluate the impact on forecast accuracy of irradiance spatial variability and PV capacity distribution depending on the region size;
2. To assess when, in a machine learning prediction method, the accuracy benefits of adding input information are completely overridden by the increasing complexity of the training phase.

For these purposes we applied the AE2 and DM/NN hybrid methods by using the WRF irradiance predictions at different spatial aggregation level and region size (Table 5), and we analyzed the evolution of the resulting forecast accuracy.

Regarding the first issue, Fig. 9 shows that in all the considered regions, the most accurate prediction was always obtained by the DM/NN method. In Sardinia, which is the smallest market area ( $24 \times 10^3 \text{ km}^2$ ), the outperforming DM/NN ingests the irradiance predicted on all the 109 NWP grid points pertaining to the region.

**Table 4**  
Accuracy of the two upscaling approaches.

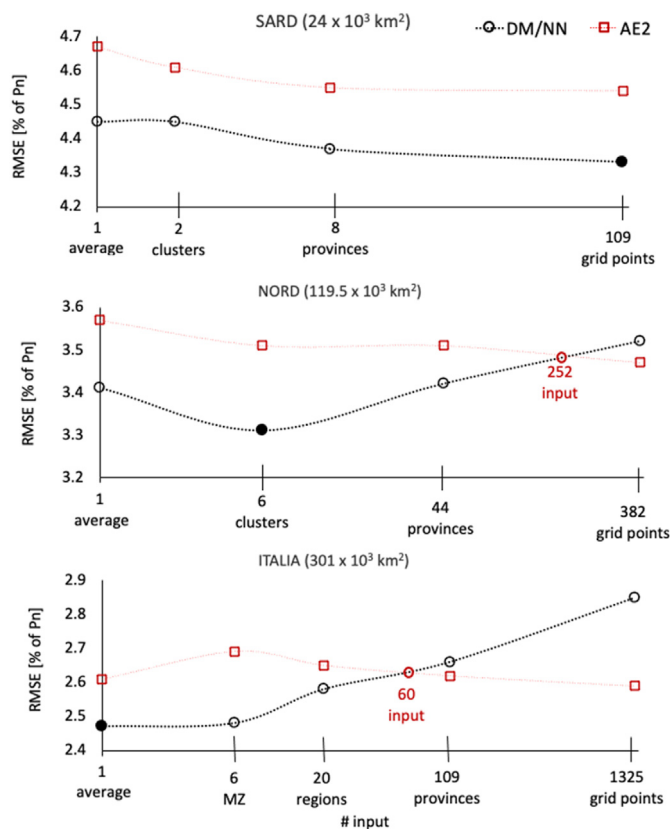
acronym	CORR [-]	RMSE [% of Pn]	MAE [% of Pn]	MBE [% of Pn]	SS [%]
1DM/NN	0.990	2.34	1.61	-0.37	54.5
6DM/NN	0.991	2.22	1.55	0.28	56.8



**Fig. 8.** Margin of performance improvement of regional PV power prediction related to more accurate forecast techniques (1) and grid reinforcement (2). Black dots are the accuracy in the generation forecast of a single market area. Empty dots are the accuracy of the generation forecast in a cluster of two or more of adjacent market zones between which National Transmission Grid (NTG) constraints have been removed. The lines result from a slight modification of the Perez/Lorenz hyperbolic/exponential fitting curves [33].

**Table 5**  
Methods main features.

acronym	Forecast technique	Input
AE2	Analog ensemble	Average sun elevation and WRF GHI day-ahead-prediction
DM/NN	Hybrid method based on physical/NN models	Average sun elevation and WRF Kcs day-ahead-prediction



**Fig. 9.** RMSE trend of the methods AE2 and DM/NN when varying the number of inputs for 3 regions of different extension. Plots are in semi-log scale.

In the North region, which is the largest market area ( $109 \times 10^3$  km<sup>2</sup>), to achieve the most accurate forecast, the 382 irradiance forecasts across the region had to be aggregated into six spatial clusters before being fed into the DM/NN model. For the entire Italy ( $301 \times 10^3$  km<sup>2</sup>), in order to obtain the best prediction accuracy, it was sufficient to supply the DM/NN model the average of 1325 irradiance predictions. These results show that the larger the size of the controlled area, the smaller the effect on prediction accuracy of spatial variability of irradiance and non-uniform installed solar capacity. We refer to this phenomenon as “input smoothing effect”. This effect is strictly related to distributed nature of the solar generation. In fact, since photovoltaic systems are widespread throughout the country, from a national perspective they appear to be uniformly distributed in both space and capacity. Therefore, in this case the spatially averaged irradiance is a very good predictor since it reflects the national solar generation quite well. As the size of the forecast footprint decreases, the non-uniformity of spatial and capacity distributions becomes increasingly important, thus, in order to achieve the most accurate forecast, the spatial resolution of the irradiance predictions must be increased.

This result is coherent with the findings of Umizaki et al. [29] and Saint-Drenan et al. [28]. Unlike our work, in which we deal only

with the aggregated regional data, Umizaki et al. managed more than 2000 single plants, randomly selecting a growing number of reference plants used to deduce the overall generation. The authors showed that if the number of PV system is large, upscaling the generation forecast of a 5% of the fleet randomly sampled is enough to obtain the maximum accuracy. In contrast, as was found in Ref. [28], when the number of systems is small (350 systems), the upscaling forecast error will depend on the spatial distribution of the fleet, so that the reference plants should be carefully selected. We can add that what it was found in Refs. [28,29] could not only depends on the number of systems but also on the size of the controlled areas (since the first area [29] appears to be, at least, four time wider than the second [28]).

Regarding the second issue of the investigation, one of the basic rules dealing with machine learning techniques (ML) is that the input features should be correlated with the output, but uncorrelated to each other as much as possible. However, on one hand, increasing the number of input features will, in any case, provide further information that can help to improve ML performance. On the other hand, additional features increase the dimension of the parameters space and the number of local minima of the cost-function appearing during the training phase. As a consequence, the probability of under/over fitting solutions increases. Therefore, to optimize ML performance, a tradeoff between additional information and training complexity must be found. To investigate on this topic, we compared the performance of DM/NN base model with the one achieved by the statistical Analog Ensemble method. Indeed, adding features in the AE input does not increase the dimension of the model parameters that should be optimized during the training, thus does not increase the complexity of the training. For this reason, we used AE performance to evaluate the effectiveness of our DM/NN model depending on the number of irradiance inputs and regional size.

Fig. 9 shows that AE always benefit of additional information regardless the number of intra-correlated input. In contrast, DM/NN model shows both the optimal number of features (black dots), and what we call “complexity limit” (red empty dot), i.e., the number of inputs beyond which AE becomes more performant. Both these numbers depend on the size of the region (as previously explained) and obviously on the ML technology and architecture.

g. Comparison of the forecast accuracy obtained in this study with the accuracy reported in literature

It is very challenging to compare the accuracy obtained by other upscaling methods reported in literature as the method performance depends on local climatic condition, year of test, forecast horizon, size of the controlled area (see section 6e), data quality of the aggregated regional generation to be predicted and available data on the PV fleet [19] (positions, capacity, plane of the array, modules technology, etc.). The only way for real performance comparison is to apply upscaling methods to predict the same regional PV generation, as in this paper or in Refs. [16,19,23,27]. For these reasons, the comparison between the forecast performances reported in this work and in other papers (Table 6) can be considered as indicative.

Table 6 shows that the achieved RMSE is comparable with the other works that consider controlled areas of similar size. In contrast, except of the whole Italy, the skill score is significantly/ slightly lower depending on the used NWP. However, it is worth noting that, unlike the works cited in the table, the current investigation was conducted without any information on the regional PV fleet, other than the total annual installed capacity in the region (at the start/end of the test year), and this can have a large impact on forecast performance. Another factor that could explain the lower

**Table 6**

Comparison between the day-ahead prediction accuracy of our outperforming upscaling method (for both the NWP) with the accuracy obtained by the most accurate methods reported in literature according to similar size of the controlled area.

Forecast controlled area	Regions size [103 km <sup>2</sup> ]	RMSE [% of P <sub>n</sub> ]	Skill score [%]
Italy (RAMS)	300	3.6	30
Italy (WRF)		2.2	56.8
Germany zone [16]	214	4.1	48
<b>Italy NORD zone (RAMS)</b>	<b>119.5</b>	<b>4.1</b>	<b>36.1</b>
<b>Italy NORD zone (WRF)</b>		<b>3.3</b>	<b>48.9</b>
Germany zone [16]	104	4.3	52.8
Germany zones [27]	140/109	3.82/4.23	52/56.4
<b>Italy CNORD/CSUD/SUD zones (RAMS)</b>	<b>41/41.5/49</b>	<b>5.3/5.5/5.6</b>	<b>35.7/36.7/34.8</b>
<b>Italy CNORD/CSUD/SUD zones (WRF)</b>		<b>4.4/4.6/4.5</b>	<b>46.4/47.9/47.2</b>
Germany zone [27]	73.1	4.51	54
Japan zone [23,29]	72.6	5.1	66
<b>Italy SICIL/SARD zones (RAMS)</b>	<b>26/24</b>	<b>4.9/4.6</b>	<b>30.8/42.3</b>
<b>Italy SICIL/SARD zones (WRF)</b>		<b>4/4</b>	<b>43.5/49.9</b>
Japan zones [23,29]	32.4/33.1/36.7	5.2/5.2/5.1	67.5/66/66

skill score is the RMSE of the persistence model (day-ahead variability) which for the Japanese zones is around 15%, for the German zones is around 9.15%, while for the Italian zones is 7.8%. On the one hand, in a low variable weather conditions (as for Italy) the PV generation is easier to predict, on the other hand the persistence model is much more accurate (in clear conditions it is the outperforming model), which results in lower skill. Indeed, it could be noted in Table 6 that when the RMSE of persistence is higher, the forecast skill score is also higher: 66% for Japanese zones, 54% for Germany zones, 47% for Italian zones (WRF input), while the RMSE of the Italian forecast is slightly lower than the German and Japan forecast error: 5.15% Japan, 4.2% Germany, 3.9% Italy (WRF input). For these reasons, the best performing developed methods can be considered to be at the “state of the art” level when fed with high-quality NWP forecasts.

## 7. Conclusion

In this work, we have analyzed several relevant issues that affect the accuracy of upscaling methods used for short term regional PV generation forecasting. The investigation was conducted using Italy as case study, in particular the six different market zones in which Italy was divided. In the study we made use of regional PV power generation estimated by the Italian TSO, two different kinds of Numerical Weather Predictions (NWP), and several upscaling methods. The work is part of the research activity carried out within the IEA-PVPS task 16: Solar Resource for High Penetration and Large Scale Applications.

The main results can be summarized as following:

1. We have shown that in regional power forecasting the upscaling technique plays a significant role, while the quality of the NWP forecasts becomes less relevant. In contrast, it has been shown in the literature that the forecasting accuracy of individual PV systems is driven mainly by the performance of the NWP forecasts. Indeed, in our case study, we found that the RMSE of regional forecasts based on the same NWPs can be reduced by 30% with a more accurate upscaling method, while feeding the same upscaling method with a 50% better performing NWPs reduces the RMSE by only 20%. In contrast, for single-site power prediction, 50% better performing NWPs should result in, at a minimum, a 40% improvement in accuracy, while a more accurate prediction model can reduce the RMSE of the prediction by about 15%.
2. We also showed that the accuracy of the 1-2-3 day ahead regional forecast are very similar to each other, so the short-

- term forecasts could have been slightly affected by the forecast horizon also because of the above mentioned reasons.
3. We have shown that merging sub-regional PV power forecasts to predict national solar generation (average model output approach on sub-regional level) only slightly improves the accuracy of the prediction of the whole country's generation (average model input approach). Nevertheless, the second approach is much more straightforward and therefore preferred.
4. We quantified the margin of improvement in prediction accuracy that can be achieved by more accurate upscaling methods and by enlarging the areas controlled by the prediction, i.e., removing grid constraints between subregions. We estimated that, for the Italian case, each of these strategies (to be applied sequentially) can bring up to 50% improvement in the accuracy of national PV generation forecasting.
5. We showed that, dealing with a VPP approach and ML forecast model (model input average), the aggregation of NWP input features should increase with the forecast footprint to reach the outperforming prediction. Thus, large area forecast is not only more accurate (forecast smoothing) but also more straightforward to implement since fewer degrees of freedom are required (we call this effect “input smoothing”). Our finding is coherent with the results achieved in recent researches [28,29] for other upscaling approaches that aim to detect the relevant plants whose generation better reflect the regional production (output models average). Nevertheless, our result could indicate that forecast uncertainty due to this upscaling approach could depend not only on the number of selected reference plants but also on the forecast footprint. In other words, the smaller the controlled area the higher should be the density of reference plants needed to obtain a given upscaling forecast uncertainty. This is a counterintuitive result as it implies that the larger the controlled area, the lower the accuracy impact of the non-uniform irradiance/capacity distribution and certainly needs further investigation.
6. We have shown and quantified the maximum amount of correlated information that can be ingested into our machine learning (ML) forecast method before it becomes worse than a simpler statistical method (used as a benchmark). This number, which we called: “complexity limit”, depends on the size of the controlled area and, of course, on the ML algorithm itself. This means that, coherently with the aforementioned results, dealing with ML based upscaling method, GHI spatial aggregation becomes more and more important as the forecast footprint increases. In particular, we have shown that although 109 GHI

inputs for our NN based method appears to be a large number of features, they are below the complexity limit for a small area such as Sardinia and still provide relevant information. In contrast, for the whole of Italy the complexity limit is reduced to 60 input features, which means that more information (intra-correlated) ingested into the model completely compromises its accuracy.

### CRedit authorship contribution statement

**Marco Pierro:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Damiano Gentili:** Methodology, Software, Formal analysis. **Fabio Romano Liolli:** Methodology, Software, Formal analysis. **Cristina Cornaro:** Results discussion, Supervision, Writing – review & editing. **David Moser:** Results discussion, Supervision, Writing – review & editing. **Alessandro Betti:** Software, Results discussion, Supervision, Writing – review & editing. **Michela Moschella:** Software, Results discussion, Supervision, Writing – review & editing. **Elena Collino:** Software, Results discussion, Supervision, Writing – review & editing. **Dario Ronzio:** Software, Results discussion, Supervision, Writing – review & editing. **Dennis van der Meer:** Results discussion, Supervision, Writing – review & editing.

### 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.

### Acknowledgment

This work has been partly financed by Research Fund for the Italian Electrical System with the Decree of April 16, 2018. This work has been conceived in the framework for the IEA PVPS Task16, Solar resources for high penetration and large-scale applications. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano for covering the Open Access publication costs. We also thank the forecast provider Ideam Srl for the supply of the WRF NWP used for this study.

### References

- [1] IRENA, Global Energy Transformation - A Roadmap to 2050, 2019.
- [2] IEA, Solar PV [Online]. Available: <https://www.iea.org/reports/solar-pv>, 2020.
- [3] H. Wirth, Recent Facts about Photovoltaics in Germany, Fraunhofer ISE, 2020.
- [4] M. Pierro, R. Perez, M. Perez, D. Moser, C. Cornaro, Italian protocol for massive solar integration: imbalance mitigation strategies, *Renew. Energy* 153 (2020) 725–739.
- [5] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, C. Cornaro, Photovoltaic generation forecast for power transmission scheduling: a real case study, *Sol. Energy* 174 (2018) 976–990.
- [6] F. Katiraei, J.R. Agüero, Solar pv integration challenges, *IEEE Power Energy Mag.* 9 (3) (2011) 62–71.
- [7] V. Larson, Chapter 12, forecasting solar irradiance with numerical weather prediction models, in: *Solar Energy Forecasting and Resource Assessment*, Academic Press, Boston, 2013, pp. 299–318.
- [8] R. Perez, T. Cebeauer, M. Sürri, Chapter 2 - semi-empirical satellite models, in: J. Kleissl (Ed.), *Solar Energy Forecasting and Resource Assessment*, Academic Press, Boston, 2013.
- [9] R. Perez, T.E. Hoff, Chapter 6 - solar resource variability, in: *Solar Energy Forecasting and Resource Assessment*, Academic Press, Boston, 2013.
- [10] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, C. Cornaro, Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data, *Sol. Energy* 158 (2017) 1026–1038.
- [11] R. Ahmed, V. Sreeram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization, *Renew. Sustain. Energy Rev.* 124 (2020) 109792.
- [12] U. Das, K. Tey, M. Seydmahmoudia, Forecasting of photovoltaic power generation and model optimization: a review, *Renew. Sustain. Energy Rev.* 81 (2018) 912–928.
- [13] L. Gigoni, A. Betti, E. Crisostomi, F. A. M. Tucci, D. Mucci, Day-ahead hourly forecasting of power generation from photovoltaic plants, *IEEE Trans. Sustain. Energy* 9 (2) (2018) 831–842.
- [14] E. Nuño, M. Koivisto, N.A. Cutululis, P. Sørensen, On the simulation of aggregated solar PV forecast errors, *IEEE Trans. Sustain. Energy* 9 (4) (2018) 1889–1898.
- [15] E. Lorenz, J. Hurka, G. Karampela, D. Heinemann, H.S. Beyer, Qualified forecast of ensemble power production by spatially dispersed gri-connected PV systems, in: *23rd EU PVSEC Section 5AO.8.6*, 2008.
- [16] E. Lorenz, T. Scheidsteiger, J. Hurka, D. Heinemann, C. Kurz, Regional PV power prediction for improved grid integration, *Prog. Photovoltaics Res. Appl.* 19 (2011) 757–771.
- [17] H. Shaker, H. Zareipour, D. Wood, A data-driven approach for estimating the power generation of invisible solar sites, *IEEE Trans. Smart Grid* 7 (5) (2016) 2466–2476.
- [18] H. Shaker, D.Z.H. Manfre, Forecasting the aggregated output of a large fleet of small behind-the-meter solar photovoltaic sites, *Renew. Energy* 147 (2020) 1861–1869.
- [19] J.O.T. Fonseca, H. Ohtake, T. Takashima, K. Ogimoto, Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods, *Prog. Photovoltaics Res. Appl.* 23 (10) (2015) 1203–1218.
- [20] J.M. Bright, S. Killinger, D. Lingfors, N.A. Engerer, Improved satellite-derived PV power nowcasting using real-time power data from reference PV systems, *Sol. Energy* 168 (2018) 118–139.
- [21] L. Lorente, X. Liu, D.J. Morrow, Spatial aggregation of small-scale photovoltaic generation using Voronoi decomposition, *IEEE Trans. Sustain. Energy* 11 (4) (2020) 2677–2686.
- [22] M. Zamo, O. Mestre, P. Arbogast, O. Pannekoucke, A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production part I: deterministic forecast of hourly production, *Sol. Energy* 105 (2014) 792–803.
- [23] J. Fonseca, T. Oozeki, H. Ohtake, K. Shimose, T.K. Takashima, Regional forecasts and smoothing effect of photovoltaic power generation in Japan: an approach with principal component analysis, *Renew. Energy* 68 (2014) 403–413.
- [24] P. Aillaud, J. Lequeux, J. Mathé, L. Huet, C. Lallemand, O. Liandrat, N. Sebastien, F. Kurzrock, N. Schmutz, Day-ahead forecasting of regional photovoltaic production using deep learning, in: *47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020.
- [25] M. Moschella, M. Tucci, E. Crisostomi, A. Betti, A machine learning model for long-term power generation forecasting at bidding zone level, in: *IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Budapest, 2019, 2019.
- [26] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, D. Heinemann, Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data, *Sol. Energy* 135 (2016) 197–208.
- [27] Y. Saint-Drenan, S. Vogt SKillinger, J. Bright, R. Fritz, R. Potthast, Bayesian parameterisation of a regional photovoltaic model – application to forecasting, *Sol. Energy* 188 (2019) 760–774.
- [28] Y. Saint-Drenan, G. Good, M. Braun, T. Freisinger, Analysis of the uncertainty in the estimates of regional PV power generation evaluated with the upscaling method, *Sol. Energy* 135 (2016) 536–550.
- [29] M. Umizaki, F. Uno, T. Oozeki, Estimation and forecast accuracy of regional photovoltaic power generation with upscaling method using the large monitoring data in Kyushu, Japan, *IFAC-PapersOnLine* 52 (28) (2018) 582–585.
- [30] S. Killinger, B. Müller, B. Wille-Haussmann, R. McKenna, Evaluating different up-scaling approaches to derive the actual power of distributed PV systems, in: *44th Photovoltaic Specialist Conference, (PVSC)*, Washington, DC, 2017.
- [31] M. Pierro, D. Moser, C. Cornaro, Chapter 8, Machine learning-based PV power forecasting methods for electrical grid management and energy trading, in: *MACHINE LEARNING AND DATA SCIENCE IN THE POWER GENERATION INDUSTRY*, Elsevier, 2020, pp. 165–192.
- [32] Y. Yu, M. Wang, F. Yan, M. Yang, J. Yang, Improved convolutional neural network-based quantile regression for regional photovoltaic generation probabilistic forecast, *IET Renew. Power Gener.* 14 (14) (2020) 2712–2719.
- [33] IEA PVPS Task 16, Solar resource for high penetration and large-scale applications, "Regional solar power forecasting, Tech. rep. (2020).
- [34] Terna, Terna download center [Online]. Available: <https://www.terna.it/it/sistema-elettrico/transparency-report/download-center>.
- [35] ECMWF, "IFS Documentation CY47R1, Part III: Dynamics and Numerical Procedures," 2020.
- [36] W. Skamarock, J. Klemp, J. Dudhia, D.B.D.A. Gill, Description of the Advanced Research WRF Version 3. NCAR Tech; Note NCAR/TN-4751STR, Technical Report NCAR, Boulder, CO, USA, 2008.
- [37] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, C. Cornaro, Model output statistics cascade to improve day ahead solar irradiance forecast, *Sol. Energy J.* 117 (2015) 99–113.
- [38] L. Delle Monache, F. Eckel, D. Rife, B. Nagarajan, K. Searight, Probabilistic weather prediction with an analog ensemble, *Mon. Weather Rev.* 141 (2013) 3498–3516.

- [39] S. Alessandrini, L. Delle Monache, S. Sperati, G. Cervone, An analog ensemble for short-term probabilistic solar power forecast», *Appl. Energy* 157 (2015) 95–110.
- [40] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, C. Cornaro, Deterministic and stochastic approaches for day-ahead solar power forecasting, *J. Sol. Energy Eng.* 139 (2017) 21010.
- [41] P. Ineichen, R. Perez, A New airmass independent formulation for the Linke turbidity coefficient, *Sol. Energy* 73 (2002) 151–157.
- [42] E. Fix, J. Hodges, Discriminatory analysis : nonparametric discrimination, consistency properties, Randolph Field, Tex. USAF School of Aviation Medicine (1951).
- [43] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [44] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, C. Cornaro, Multi-Model Ensemble for day ahead prediction of photovoltaic power generation, *Sol. Energy* 134 (2016) 132–146.
- [45] R.A. Pielke, A comprehensive meteorological modeling system—RAMS, *Meteorol. Atmos. Phys.* 49 (1) (1992). Art. n. 1, mar. 1992, vol. 49, no. 1, p. Art. n.1.