Technology Collaboration Programme
by iea

International Energy Agency
Photovoltaic Power Systems Programme

Task 16 Solar Resource for High Penetration and Large Scale Applications

PVPS

Regional solar power forecasting

2020

Report IEA-PVPS T16-01: 2020

What is IEA PVPS TCP?

The International Energy Agency (IEA), founded in 1974, is an autonomous body within the framework of the Organization for Economic Cooperation and Development (OECD). The Technology Collaboration Programme (TCP) was created with a belief that the future of energy security and sustainability starts with global collaboration. The programme is made up of 6.000 experts across government, academia, and industry dedicated to advancing common research and the application of specific energy technologies.

The IEA Photovoltaic Power Systems Programme (IEA PVPS) is one of the TCP's within the IEA and was established in 1993. The mission of the programme is to "enhance the international collaborative efforts which facilitate the role of photovoltaic solar energy as a cornerstone in the transition to sustainable energy systems." In order to achieve this, the Programme's participants have undertaken a variety of joint research projects in PV power systems applications. The overall programme is headed by an Executive Committee, comprised of one delegate from each country or organisation member, which designates distinct 'Tasks,' that may be research projects or activity areas.

The IEA PVPS participating countries are Australia, Austria, Belgium, Canada, Chile, China, Denmark, Finland, France, Germany, Israel, Italy, Japan, Korea, Malaysia, Mexico, Morocco, the Netherlands, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, and the United States of America. The European Commission, Solar Power Europe, the Smart Electric Power Alliance (SEPA), the Solar Energy Industries Association and the Cop- per Alliance are also members.

Visit us at: www.iea-pvps.org

What is IEA PVPS Task 16?

The objective of Task 16 of the IEA Photovoltaic Power Systems Programme is to lower barriers and costs of grid integration of PV and lowering planning and investment costs for PV by enhancing the quality of the forecasts and the resources assessments.

Authors

- **Main Content:** M. Pierro (marco.pierro@gmail.com), C. Cornaro (cornaro@uniroma2.it), D. Moser (david.moser@eurac.edu), A. Betti (alessandro.betti@i-em.eu), M. Moschella (michela.moschella@i-em.eu), E. Collino (Elena.Collino@rse-web.it), R. Dario (Dario.Ronzio@rse-web.it), D. van der Meer (dennis.vandermeer@angstrom.uu.se), J. Widen (joakim.widen@angstrom.uu.se), L. Visser (l.r.visser@uu.nl), T. AlSkaif (t.a.alskaif@uu.nl), W. van Sark (W.G.J.H.M.vanSark@uu.nl)
- **Corresponding Author:** Alessandro Betti, i-EM srl, **Phone**: +3905860958-20, **Email**: alessandro.betti@i-em.eu
- **Editors:** Elke Lorenz, Fraunhofer ISE,(elke.lorenz@ise.fraunhofer.de), Carmen Köhler, p3r solutions (carmen.koehler@p3r.solutions), Jan Remund, Meteotest AG, Operating Agent IEA PVPS Task 16, (jan.remund@meteotest.ch)

COVER PICTURE

Picture kindly offered by Jan Remund. Sunset at 110 kV line Mühleberg-Wattenwil, CHE. It depicts the main topic of the report: Irradiance incident and electricity grid. Regional solar power forecasts are often used by distribution and transmission grid operators.

ISBN 978-3-906042-88-6: Regional Solar Power Forecasting

INTERNATIONAL ENERGY AGENCY
PHOTOVOLTAIC POWER SYSTEMS PROGRAMME

IEA PVPS
Task 16, Subtask 3 - Activity 3.2

Report IEA-PVPS T16-01:2020
May – 2020

ISBN 978-3-906042-88-6

EXECUTIVE SUMMARY

High levels of photovoltaic (PV) power penetration pose challenges to the operational performance of the power system. Regional forecasts of PV power allow transmission system operators (TSOs) and distribution system operators (DSOs) to take appropriate measures to maintain balance between supply and demand.

In this work, we compare the accuracy of several up-scaling methods for regional PV power forecast from 1 up to 3 days-ahead using three different case studies characterized by different sizes of the area of interest and located in two different countries:

- Forecast of the aggregated PV power output of 150 PV systems in the province of Utrecht (Netherlands) within a controlled area of 1.4 $x10^3$ km^2;
- Forecast of the regional PV generation of six Italian macro zones (market zones) with controlled areas that range from $24x10^3$ to $119x10^3$ km^2;
- Forecast of the regional PV generation of the whole of Italy with a controlled area of 300 $x10^3$ km^2.

The up-scaling approaches are applied to map meteorological forecasts of numerical weather prediction (NWP) models to regional PV power generation. All the up-scaling methods shown here directly predict the regional PV power generation, i.e. they consider the PV power output of the whole PV fleet as if it had been produced by a single “virtual” solar power plant, rather than predicting PV power for representative PV power plants as a basis to predict regional PV power.

In order to improve regional PV power forecasts and to advance the state-of-the-art, it is important that researchers have access to identical datasets, so they may compare their new methods. To that end, this report presents three datasets for Italy and Utrecht, the Netherlands. More specifically, for Italy, the datasets are made of satellite derived global horizontal irradiance data, numerical weather forecasting of some variables affecting PV production and corresponding PV power data. The Utrecht dataset is comprised of NWP forecasts and aggregated PV power measurements of 150 systems. These datasets have been cleaned in order to be suitable to test different PV power forecasting methods.

The focus of this work is on the comparison of different PV power up-scaling methods, that have been performed on the aforementioned datasets. Since accuracy of a PV power forecast model is greatly affected by the input NWP data, the same NWP forecast has been used within each case study. Among the benchmarks methods there are artificial neural networks, analog ensembles, cooperative ensembles of machine learning algorithms, quantile regression forests, random forests and gradient boosted regression trees. These methods use state-of-the-art machine-learning approaches and allow researchers to compare their new models against each other. It should be noted that probabilistic forecast evaluation is not in the scope of this report, although several models presented herein are capable of producing probabilistic forecasts. Besides the description of the benchmark models, the report also describes several feature engineering techniques and pre-processing procedures aimed at improving the accuracy of the benchmarks presented herein. The accuracy of the benchmark models is assessed extensively, both quantitatively and qualitatively, allowing researchers and engineers to thoroughly compare their new methods against these benchmarks.

For the PV power forecasts at the zonal levels of Italy, the analog ensemble model achieves a root mean squared error (RMSE) skill score between 20% and 36% and a mean bias error (MBE) between -1.9% and 1.5% for all forecast horizons (i.e., 1-3 days ahead). It should be noted that the forecasts and measurements are normalized with respect to the installed capacity in the area throughout this report. Considering the 24h ahead forecasts of the entire PV fleet of Italy, the benchmarks achieve an RMSE skill score between 28% and 47%, while the MBE lies between 0.1% and 1.9%. The so called “smoothing effect” has been also investigated over Italy, demonstrating that the RMSE decreases with the root of the area under control following an exponential or hyperbolic trend and decreasing from almost 5.5% of the installed capacity on average over market zones down to 3.6% over Italy. In addition, blends of the models are also applied to this dataset, which improves the RMSE skill score from 47% (the best model) to 51% using a nonlinear blending technique. Finally, the benchmark models achieve RMSE skill scores between 42% and 44% on the PV power production forecasts over the province of Utrecht, the Netherlands, while the MBE lies between 0% and 0.8%.

The comprehensive analysis presented herein shows to the interested reader the importance of carefully tuning all the different up-scaling steps in the PV power forecast chain in order to achieve the highest accuracy.

TABLE OF CONTENTS

FOREWORD

The International Energy Agency (IEA), founded in November 1974, is an autonomous body within the framework of the Organisation for Economic Co-operation and Development (OECD) which carries out a comprehensive programme of energy co-operation among its member countries

The IEA Photovoltaic Power Systems Technology Collaboration Programme (IEA-PVPS) is one of the collaborative R&D agreements established within the IEA and, since 1993, its participants have been conducting a variety of joint projects in the applications of photovoltaic conversion of solar energy into electricity.

The participating countries and organisations can be found on the www.iea-pvps.org website.

The overall programme is headed by an Executive Committee composed of one representative from each participating country or organization, while the management of individual Tasks (research projects / activity areas) is the responsibility of Operating Agents. Information about the active and completed tasks can be found on the IEA-PVPS website www.iea-pvps.org.

This document is the report of Task 16 “Solar Resource for High Penetration and Large Scale Applications”, which includes 20 participating countries from around the world and focuses mainly on two scientific issues: (i) high frequency variability and solar forecasts for managing grids with high penetration pf PV and local storage and (ii) high quality solar resource assessments and forecasts for solar installations notable in the multi MW class.

The report corresponds to Subtask 3 “Evaluation of Current and Emerging Solar Forecasting Techniques” and specifically to Activity 3.2 “Regional Solar Power Forecasting”, which aims to collect and disseminate example datasets to help research activities in regional PV power forecasting and support the development of methods for aggregated regional PV power forecasting. The final goal is facilitating the penetration of distributed renewable systems into the electricity grid.

ACKNOWLEGMENT

This paper received valuable contributions from several IEA-PVPS Task 16 members and other international experts.

First of all, the authors would like to acknowledge Tomas Landelius, Sandra Andersson, Weine Josefsson and Thomas Carlund from SMHI Institute for having provided a dataset.

For their support in reviewing the document and/or participation in the survey, many thanks go to (in alphabetic order):

Carmen Kohler

Elke Lorenz

Jan Remund

This report is supported by:

Netherlands Ministry of Economic Affairs and Climate

Netherlands Organisation for Scientific Research (NWO), as part of the eNErgy intrAneTs (NEAT) project

…

ABBREVIATIONS AND ACRONYMS

AOI	Angle of incidence on an equivalent plane of the array [^{o}]
DOY	Day of Year
GHI_{CS}	Clear Sky Global Horizontal Irradiance
GHI^{FOR}	Forecasted Global Horizontal Irradiance
GHI_{SAT}	Global Horizontal Irradiance from Satellite
IEA	International Energy Agency
K_{CS}^{FOR}	Forecasted Clear Sky Index
KPI	Key Performance Index
ML	Machine Learning
MOS	Model Output Statistic
NOCT	Nominal Operating Cell Temperature
NWP	Numerical Weather Prediction
PCA	Principal Component Analysis
POA	Plane Of Array
P_n	PV capacity installed in a geographical area [MWp]
P_{PV}	Actual PV Power Generation [MWh]
P_{PV}^{CS}	Clear Sky PV Power Generation [MWh]
P_{PV}^{FOR}	Forecasted PV Power Generation [MWh]
PRF	Performance Factor
PVK_{sc}	PV Clear Sky Index
QRF	Quantile Regression Forest
T_{panel}	Panel Temperature
T_{2m}	Air Temperature at 2 m above the ground
TSO / DSO	Transmission and Distribution System Operator

1 INTRODUCTION

Background: Nowadays global warming poses increasing severe risks for the ecosystem, which is already facing impact of climate change, including for example rising sea levels and a higher frequency of extreme weather events. Global greenhouse gas emission is by far due to the intensive use of non-renewable fossil fuel energy sources, which represent today two thirds of greenhouse gas emission.

In order to promptly limit global warming, countries are now adopting actions to facilitate the transition towards clean energy sources. While the sector of renewable energies (REs) is constantly increasing, this is not sufficient: in order to meet climate targets, the total share of RE "must rise from around 15% of the total primary energy supply in 2015 to around two-thirds by 2050" [1].

Photovoltaic (PV) power represents one of the main important shares of the total REs along with wind power generation [1]. However, high penetration of PV power generation poses several challenges for stability of the electrical grid, due to the stochastic variability of the residual electric load, i.e. the difference between the energy need (load) and the distributed PV power generation, dependent on meteorological conditions and the course of the sun. In this context, accurate forecasting of PV power generation is highly required for a better energy scheduling and a proper balance between production and demand. This information is essential for Distribution (DSO) and Transmission System Operators (TSO), as well as aggregators and energy traders [2].

Regional PV power forecasting techniques represent a hot topic in the research field and can be mainly grouped in physical approaches and data-driven models, requiring detailed PV plant information and historical power data for training and inference, respectively. Any combination of two or more of the previous models is often used in order to enhance further performances and is called hybrid model. Furthermore, sometimes the term "statistical method" is used to indicate a PV power generation forecasting technique based on historical data, either by means of a statistical analysis of the different input variables (e.g. ARMA, ARIMA, etc.) or by using more intelligent machine learning algorithms, able to handle also non-linear and non-stationary data patterns [3] [4].

Two main strategies exist in the literature for regional PV power forecasting. The first strategy can be defined as "bottom-up" or accumulation approach, which consists of first predicting the power output for each PV site in the regional area under consideration, and subsequently aggregating the predictions for the whole area. This method is characterized by high computational burden and requires a detailed knowledge of every plant in the area and therefore it is not normally feasible, especially for large areas.

The second strategy is represented by "up-scaling" methods, which may be implemented in two possible ways: models output average and model inputs average. In the first case the representative PV sites of each area (or clusters of PV plants), are sampled, then the PV power production forecast of such plants (or clusters) is rescaled to obtain the regional power prediction according to the total capacity in the area. Also in this case it would be necessary to estimate with high accuracy the production for the representative plants. This approach can be effectively applied only in presence of a significant number of representative plants evenly distributed on the area, otherwise the scaling will gain poor results.

The alternative up-scaling method is based on a direct approach called model inputs average, which is based on the spatial smoothing of the input features. In this case the PV power

generation of the area is considered as a virtual power plant and prediction is made directly at regional level by using a historical data set of power measurements and input meteorological data aggregated at a lower level than regional. The main advantage of this approach is the possibility of obtaining a reliable power forecast without details on the installations, except for the total installed capacity of the area.

In this report approaches corresponding to the method model inputs average are compared: the up-scaling is applied to map meteorological forecasts of NWP prediction models directly to regional PV power generation. Of course, the performances of up-scaling models strongly depend on the quality of the NWP model adopted, whose accuracy can vary depending on the geographical location under control, as well as on weather and climate conditions. Since the present work is focused on the comparison of different up-scaling methods and not on the impact of different NWP models on the final performance, the same NWP data has been used in input to the up-scaling methods for each case study.

Objective: The objective of this work is twofold: (i) Setup of common reference data sets, which were shared among the participants of this study and allowed to compare their models on identical datasets. (ii) Setup a benchmark on the aforementioned datasets to evaluate performances and compare results of different regional PV power forecasting methods. The analysis and the conclusions achieved in the present report facilitates the investigation of new challenges in regional PV power generation forecasting, which will be the main subject of the next report of IEA Task 16 – Activity 3.2.

Structure: The manuscript is organized as follows. Chapter 2 depicts the technical background, with an overview of the physical quantities involved and a brief introduction to the most widely used PV forecasting techniques. The datasets employed in this work are described in Chapter 3, whereas Chapter 4 presents the PV power forecasting models used for the benchmarking activity.

Chapter 5 and 6 show the results of the benchmark for Italy and for the province of Utrecht, The Netherlands, respectively. Finally, Chapter 7 draws conclusions and outlines possible future challenges concerning PV power forecasting at regional level.

2 BACKGROUND INFORMATION

The sun is the single most abundant source of energy [5]. It is possible to harvest part of this energy through the photovoltaic (PV) effect by converting irradiance (in Wm^{-2}) into electricity. However, this poses challenges due to the fact that the irradiance that reaches the earth's surface is variable, due to the presence of clouds and atmosphere components affecting the air transparency (aerosols), and because of the deterministic daily and seasonal course of the sun caused by the rotation of the earth and its orbit around the sun. Indeed, high levels of PV power penetration introduce a dependence on meteorological conditions into the net load curve, i.e. demand minus PV supply, which is the residual electric demand that should be covered by dispatchable energy sources, or by means of storage systems [6]. In many energy markets the peak demand occurs close to the sunset. As a consequence, the resulting net load shows a sharp increase at the beginning of the evening and resembles the shape of a duck. The term duck-curve [7] has been coined for the first time in 2012 by the California Independent System Operator and illustrates the mismatch between electricity consumption and PV power production.

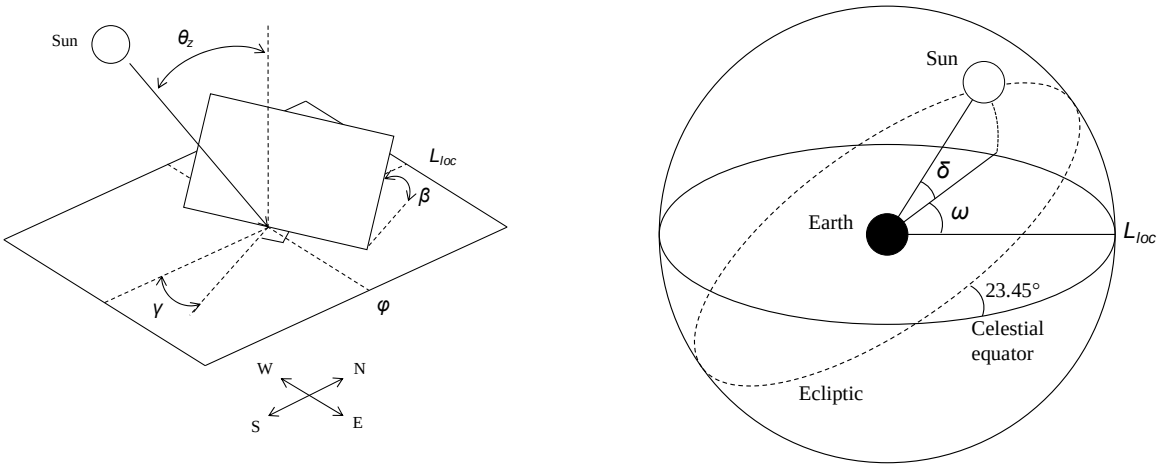
PV power forecast is a cost-effective method useful to enable the integration of the solar power into the electricity grid. In particular it can be exploited for a better management of transmission scheduling in order to reduce energy imbalance and related cost of penalties. A regional power forecasting model is a particular type of solar prediction since it benefits from the so-called "ensemble smoothing effect", i.e. the variability of spatially aggregated PV systems decreases with the size of the area under control, which in parallel increases the accuracy of the forecasts [8]. Furthermore regional PV power forecast requires a different strategy with respect to the classical power forecast for single plants, either extrapolating a set of representative PV systems to the entire regional PV fleet, or using data driven methods, in order to forecast the total generation of an area by means of the existing relationships between weather conditions in the area and the corresponding PV power production.

2.1 Irradiance and its Effect on Production, Consumption and Net Load

Irradiance is defined as the radiant energy that reaches a surface per unit of time [9]. The term "irradiance" used here is synonymous with the global horizontal irradiance (GHI), which consists of direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI), and which is defined as the total solar radiation measured in the horizontal plane. More specifically, GHI is defined as in Eq. (1):

$$GHI = DNI \cdot cos(\theta_z) + DHI \qquad (1)$$

where θ_z represents the solar zenith angle. While it is possible to establish a statistical relationship between GHI and PV power production, another approach is to calculate the in-plane irradiance (or Plane of array POA irradiance) from which the PV power may be estimated using system information. Figure 1(a) and Figure 1(b) present the relevant angles (plane tilt β, azimuth angle γ, latitude ϕ, longitude L_{loc}, zenith angle θ_z, hour angle ω and solar declination δ) that can be used to calculate the angle of incidence θ, which is necessary to calculate the in-plane irradiance. The interested reader is referred to [10] for the mathematical derivation.



(a) (b)

Figure 1: In (a), the relevant angles (plane tilt β, azimuth angle γ, latitude ϕ, longitude L_{loc} and zenith angle θ_z) that define the location and orientation of the tilted plane. In (b), the relevant angles (hour angle ω and solar declination δ) defining earth's position relative to the sun that constitute the deterministic variability. Combined, these can be used to calculate the angle of incidence θ and subsequently the plane irradiance. Copied with permission of the copyright owners [10].

GHI is affected by two types of variability: (1) the deterministic rotation of the earth and its orbit around the sun, and (2) the stochastic atmospheric processes such as clouds and wind. These types of variability play a crucial role when integrating PV power into the power system. Deterministic variability requires seasonal planning and diurnal ramping of traditional power plants.

The deterministic variability component of the GHI time series can be strongly reduced by dividing GHI by the clear-sky irradiance, i.e. the horizontal irradiance without any influence from (2) (see [9] for the mathematical derivation). The result is an approximately stationary time series that has roughly constant mean and variance over time and space and that is typically more straightforward to forecast. Stochastic variability, on the other hand, requires accurate forecasts in order to schedule traditional power plants appropriately, or necessitates backup power plants capable of high ramp rates. Figure 2 presents electricity consumption and net load profiles from the California Independent System Operator (CAISO) [11], where the net load is defined as electricity consumption minus solar and wind power generation. The figure clearly shows the deterministic variability caused by the rotation of the earth and the stochastic variability due to clouds and other atmospheric phenomena.

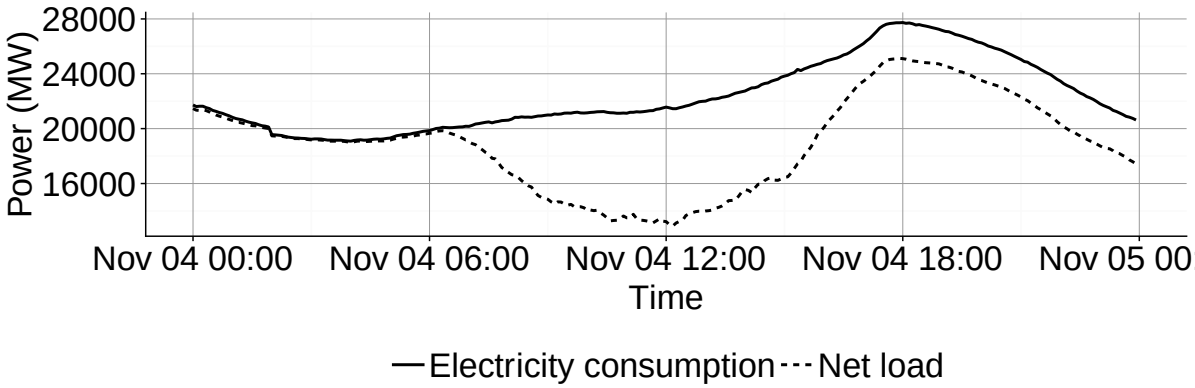


Figure 2: Electricity consumption and net load from the California Independent System Operator. The net load clearly displays the deterministic variability between sunrise and sunset, and the stochastic variability can also be observed. Date accessed: 2018-11-04.

2.2 Forecast Techniques and Up-scaling

Forecast requirements depend on the spatio-temporal scale of the purpose or application. For instance, irradiance can be forecasted up to 30 minutes ahead at a single site using all sky-imagers (ASI) with high accuracy, up to six hours ahead for an area using satellite images or satellite derived irradiance, or several days ahead for an area using numerical weather prediction (NWP) [12]. Figure 3 presents the spatial and temporal scales for the aforementioned forecasting techniques.

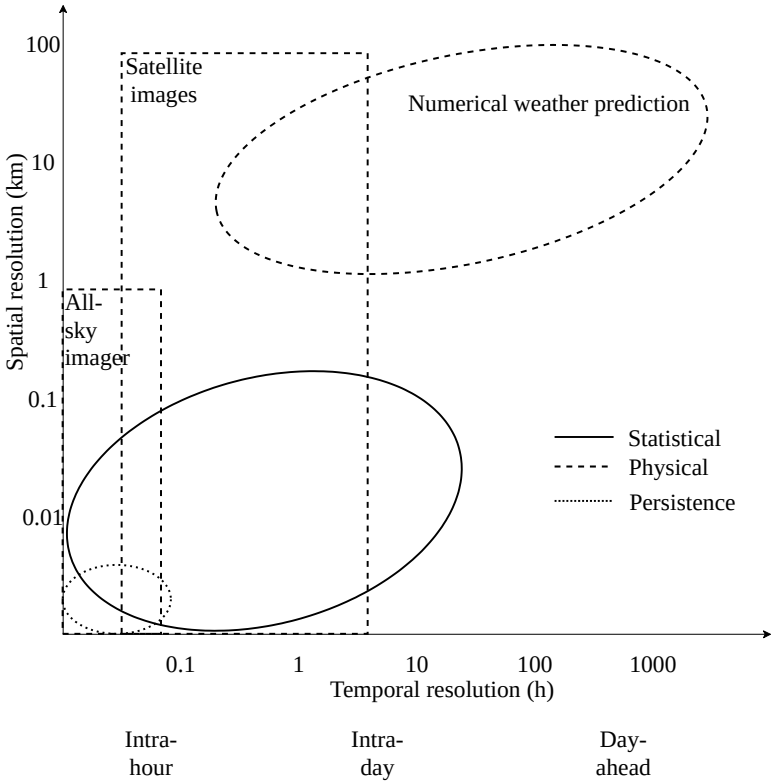


Figure 3: The spatial and temporal resolutions of various forecasting techniques. Adapted with permission from [13] and originally inspired by [14].

Considering single power plants, in order to map an irradiance forecast to a PV power forecast, it is common to use physical based models or data driven models. Physical based models describe the PV system in detail through its location and orientation in combination with manufacturer specifications, whereas data driven models learn the relationship between irradiance and PV power directly from historical data [15]. In the former case, it is common to add a temperature forecast due to its relationship with the efficiency of PV panels, whereas any related variable (e.g., past measurements, satellite derived irradiance, various NWP products, etc.) can be included in the latter case.

Data driven models can also play an important role when a regional PV power forecast is required. In this case, they can spatially aggregate a large number of input variables to directly predict the regional solar power generation, providing an effective method to upscale the PV generation forecast from the single plant level to the regional level. Indeed usually PV power production data from only a small share of total installed capacity in an area is available. Furthermore the details on the PV plants in the area are often unknown. In order to produce regional PV power forecasts from this subset of PV systems, up-scaling is necessary. Local variability plays a minor role in regional forecasts due to the geographical smoothing effect, i.e. spatial averaging. In contrast, regional weather phenomena require an accurate representation of the spatial distribution of the PV systems in order to forecast their impact accurately [16]. Currently, many grid operators use measurements and forecasts of the power of some reference systems to estimate the regional power production [17]. Recently, however, researchers have investigated more sophisticated methods to estimate regional PV power production. For instance, [16] proposed to select a representative set based on plant location, technology and installed capacity to improve the resemblance to the entire set of PV systems. [18] included the module orientation. In a later publication, [17] compared various up-scaling strategies differing in consideration of module orientation, spatial interpolation and calibration using a loss factor. [2] clustered PV systems that acted as representative sets. Another up-scaling method was proposed by [19], in which the PV power generation of the area was considered as a virtual PV power plant which can be directly forecast by mapping NWP forecasts to the output of the virtual power plant. It is worth noting that such up-scaling approach model inputs average may be applied easily to a PV fleet growing in time due to the poor information required, i.e. the metering time series and the total installed capacity in the area.

3 DATA SETS

Chapter 3 describes in detail the datasets used for benchmarking.

Before proceeding, let us introduce some useful definitions: endogenous variable is the tag forecasted by the model (in this case power generation) and corresponding to the model output. Exogenous variables are instead those determined outside the model and fed into it as input.

The datasets used in the benchmarking study differ for exogenous variables, target geographical area under consideration, forecast horizon of NWP data, as well as spatial and temporal aggregation.

After a proper evaluation, three reference datasets were selected for benchmarking activities: two are referred to Italy and one to the province of Utrecht, the Netherlands.

Concerning Italy, the following datasets have been selected for benchmarking:

- data provided by UNIROMA2/EURAC: it is exploited for benchmarking at country level;
- data provided by RSE: it is used for benchmarking at market zone level.

Regarding evaluation of model performances over Utrecht's area, the following dataset has been selected:

- data provided by Utrecht University: used for benchmarking at a regional level.

Below a detailed description of the reference datasets used for benchmarking is presented.

3.1 UNIROMA2/EURAC Data

A brief description of the provided data is reported in the following sections:

- Section 3.1.1, satellite retrieved Global Horizontal Irradiance (GHI_{SAT});
- Section 3.1.2, day ahead forecast of Global Horizontal Irradiance (GHI^{FOR}) and 2 meters air temperature (T_{2m}^{FOR});
- Section 3.1.3, PV power generation (P_{PV}) and PV capacity (P_n) over Italy ;

The Italian surface was divided into 1'325 grid points with a spatial resolution of 12 km (see Figure 4). The GHI and T_{2m} data consists in 1'325 time series of three years (2014 / 2015 /2016) with an hourly resolution. Each time series is the irradiance or the temperature of a specific point of the grid.

Figure 4: Grid points over Italy.

3.1.1 GHI Satellite Data

The satellite derived irradiance comes from the geostationary radiative fluxes products, under Météo-France responsibility. It was obtained by OSI SAF SSI algorithm[1] applied to the satellite images provided by METEOSAT-9 (MSG-3) at 0° longitude, covering 60S-60N and 60W-60E, at 0.05° latitude-longitude.

- **Area of interest**: Italy
- **Spatial aggregation**: Italy
- **Unit of measurement**: W/m^2 (average on the previous hour)
- **Time zone**: Europe/Rome
- **Period**: 2014/2015/2016
- **Granularity**: 1 hour
- **Spatial resolution**: 12 km

[1] http://www.osi-saf.org/lml/doc/osisaf_cdop2_ss1_pum_geo_flx.pdf

3.1.2 GHI and Temperature Day Ahead Numerical Weather Prediction Data

NWP forecasts were generated by the Weather Research and Forecasting (WRF–ARW 3.8) mesoscale model developed by National Center of Atmospheric Research (NCAR). Data were provided to UNIROMA2/EURAC by the company IDEAM[2].

- **Area of interest**: Italy
- **Spatial aggregation**: Italy
- **Unit of measurement**: W/m^2 (for GHI) and ^{o}C (for temperature), average on the previous hour
- **Time zone**: Europe/Rome
- **Period**: 2014/2015/2016
- **Initial and boundary data for model initialization**: GSF model
- **Radiation scheme**: “Rapid Radiative Transfer Model” (RRTM)
- **Forecast horizon**: 24 hour
- **Granularity**: 1 hour
- **Spatial resolution**: 12 km

Global Horizontal Irradiance (GHI) provided by WRF was post-processed with an original Model Output Statistic.

3.1.3 PV Power Data

The PV power generation data of the whole of Italy were downloaded from the website [20] of the Italian Transmission System Operator (TSO), i.e. TERNA.

Terna does not directly measure the PV power generation but just estimates it with a method which is not published. The Electrical Service company (GSE) that manages the distributed Italian renewable energy generation collects the real solar production data, provided by the PV meters from a large numbers of PV plants. GSE each year publishes statistical reports[3] on the PV generation in Italy.

The Terna PV power generation curves have been pre-processed to make PV capacity and monthly solar generation coherent with the values reported by GSE.

Finally the pre-processed curves have been synchronized with the satellite derived irradiance to remove time shift between the time series.

The results of this process are high quality PV generation data, coherent with the statistical production values (GSE), well synchronized with the irradiance and temperature forecast obtained from the NWP model.

- **Area of interest**: Italy
- **Spatial aggregation**: Italy
- **Unit of measurement**: MW (average on the previous hour)

[2] http://www.ideameteo.com/

[3] https://www.gse.it/Dati-e-Scenari_site/statistiche_site

- **Time zone**: Europe/Rome
- **Period**: 2014/2015/2016
- **Granularity**: 1 hour

3.2 RSE Data

In the following sections an overview of the RSE dataset is provided. The dataset is made up of observed power data for each Italian market zone, global horizontal irradiation (GHI) estimated from satellite data and GHI forecasted by means of a regional NWP model.

The data refer to the years 2016 and 2017. The dataset is available to perform the benchmark of different solar forecasting methods on the Italian market zones.

Data format is detailed in the next sections.

3.2.1 PV Power Data

Power metering data derive from Italian TSO website [20], which provides free access to fundamental data and information on the actual generation for different energy sources, including also production from PV plants.

The power values represent the solar production on the zone during the reference hour.

The data have the following specifications:

- **Area of interest**: Italy
- **Spatial aggregation**: Italian Market Zone
- **Granularity**: 1 hour
- **Unit of measurement**: MW (average on the previous hour)
- **Time zone**: UTC (Universal Time Coordinated).
- **Period**: 2015/2016

Each market zone is identified by a name, according to the TSO terminology.

The market zone names are listed in Table 5, whereas Figure 5 shows the geographical distribution. Note that Italian market zones are macro-areas defined by TERNA (Italy's TSO), for security requirements of the electric power system, due to physical limits of energy exchange with other bordering areas.

Table 1: Italian Market Zones highlighted with the colours reported in Figure 5.

Zone Name	Zone description
NORD	Northern Italy
CNOR	Central-Northern Italy
CSUD	Central-Southern Italy
SUD	Southern Italy
SICI	Sicily
SARD	Sardinia

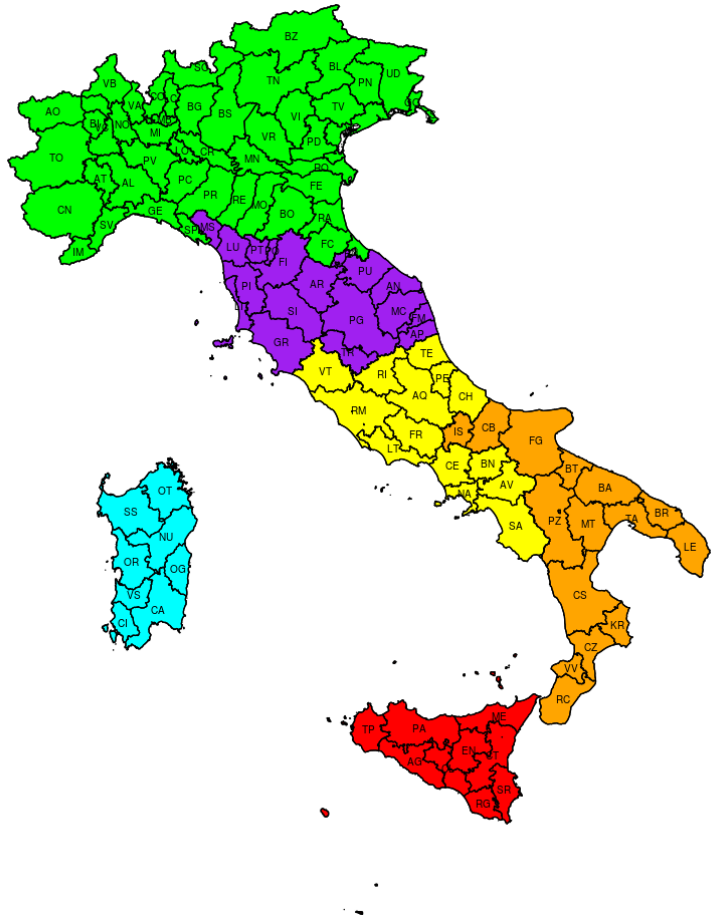


Figure 5: Geographical distribution of the Italian provinces (given by black borders) and market zones (different colours).

3.2.2 Irradiance Satellite Data

Irradiance data are estimates of the mean hourly irradiance (GHI), derived from Meteosat Second Generation (MSG) satellite data. The algorithm to derive the GHI has been developed by RSE.

Irradiance data have the following specifications:

- **Area of interest**: Italy
- **Spatial aggregation**: Italian Provinces (overall number: 110)
- **Granularity**: mean hourly irradiance, related to the previous hour with respect to the reference time
- **Unit of measurement**: W/m^2
- **Time zone**: UTC
- **Data provider**: RSE [21]
- **Period**: 2015/2016

3.2.3 Irradiance Forecast Data

GHI forecast data derive from the NWP Model RAMS, with lead time between +12 and +84 hours, driven by the global European model IFS of ECMWF.

GHI forecast data have the following specifications:

- **Area of interest**: Italy
- **Spatial aggregation**: Italian Province (110)
- **Granularity**: mean hourly data, related to the previous hour with respect to the reference time.
- **Time horizon**: three days
- **Unit of measurement**: W/m^2
- **Time zone:** UTC.
- **Numerical Weather Model**: regional model RAMS run at 12 UTC with lead time between +12 and +84 hours, driven by the global European model IFS of ECMWF.
- **Period**: 2015/2016

3.3 Utrecht University Data

In this section three dataset descriptions are given. In the first the dataset that consists of forecasts of a NWP model operated by ECMWF [22] is discussed. Thereafter the collection of additional variables is presented. The third dataset holds the aggregated PV output power measurements of the UPP sensor network in Utrecht [23]. All datasets hold hourly values for the period February 2014 until February 2017.

3.3.1 NWP

The NWP forecasts are obtained from the global European model IFS of ECMWF. By means of interpolation, a spatial resolution of approximately 9 kilometers is obtained. In order to keep low complexity, the weather predictions are collected for one grid point corresponding to KNMI

weather station in De Bilt, with coordinates 52.10N 5.18E. The considered archived weather predictions have a time horizon up to 24 hours, with a time step of 3 hours. Hourly values are obtained by means of Akima interpolation [24]. The 12 UTC model run is used to obtain the day-ahead weather predictions for 00:00-11:00. Weather predictions for 12:00-23:00 are collected from the 00 UTC model run.

Consequently, weather predictions have a time horizon varying from 12 to 24 hours ahead. The weather prediction variables available in the dataset can be found in Table 2.

Table 2: Weather variables available from NWP forecasts by ECMWF.

Weather prediction variables	Unit	Abbreviation
Mean sea level pressure	Pa	msl
Ambient temperature at 2 meters	K	t2m
Dewpoint temperature at 2 meters	K	d2m
Zonal wind vector at 10 meters	m s-1	u10
Meridional wind vector at 10 meters	m s-1	v10
GHI	J m-2	ssrd
Low cloud cover	(0-1)	lcc
Mid cloud cover	(0-1)	mcc
High cloud cover	(0-1)	hcc
Total cloud cover	(0-1)	tcc
Total precipitation	m	tp

- **Area of interest**: Utrecht, the Netherlands (roughly an area of 60x60 km^2)
- **Spatial aggregation**: Utrecht, the Netherlands (one grid point corresponding to KNMI weather station)
- **Granularity**: 1 hour
- **Time horizon:** 24 hours
- **Time zone**: UTC
- **Unit of measurement**: see Table 2, (average on the successive hour)
- **Numerical Weather Model**: global European model IFS of ECMWF
- **Period**: February 2014/February 2017

3.3.1 Seasonal and Diurnal Variations

In addition to the input variables defined above, the Clear Sky Irradiance (CSI) and solar zenith angle (SZA) are collected. These variables are included in order to account for seasonal and diurnal variations in solar irradiance. The CSI is obtained from the Ineichen-Perez Clear Sky Model [25].

- **Area of interest**: Utrecht, the Netherlands
- **Spatial aggregation**: Utrecht, the Netherlands
- **Granularity**: 1 hour
- **Time zone**: UTC
- **Unit of measurement**: W/m^2 and degrees (average on the successive hour)
- **Model**: Ineichen-Perez Clear Sky Model
- **Period**: February 2014/February 2017

3.3.2 PV Power Measurements

PV power measurements are extracted from the PV sensor network in Utrecht (UPP). The UPP network consists of 202 PV-systems that are distributed throughout the province of Utrecht, the Netherlands [26]. All these PV systems are rooftop mounted residential installations, and individual system sizes range from 0.5 up to 6.8 kWp [26].

PV power measurements expressed in W with a one-minute time resolution are extracted for all PV-systems. After filtering on data availability, 150 PV-systems are selected (see Figure 6). Subsequently, minute power measurements are converted to hourly power production values by averaging. The total production in kWh/kWp is calculated by:

$$X_t = \sum_{k=1}^{n} \frac{p_{k,t}}{p_{c,k}}, \qquad (2)$$

where $p_{k,t}$ is the production of system k at time t, $p_{c,k}$ is the installed capacity of system k and n denotes the number of PV-systems with a power measurement at time t.

- **Area of interest**: Utrecht, the Netherlands
- **Spatial aggregation**: 150 PV-systems in Utrecht (see Figure 6)
- **Granularity**: 1 hour
- **Time zone**: UTC
- **Unit of measurement**: W (mean value in the successive hour)
- **Period**: February 2014/February 2017

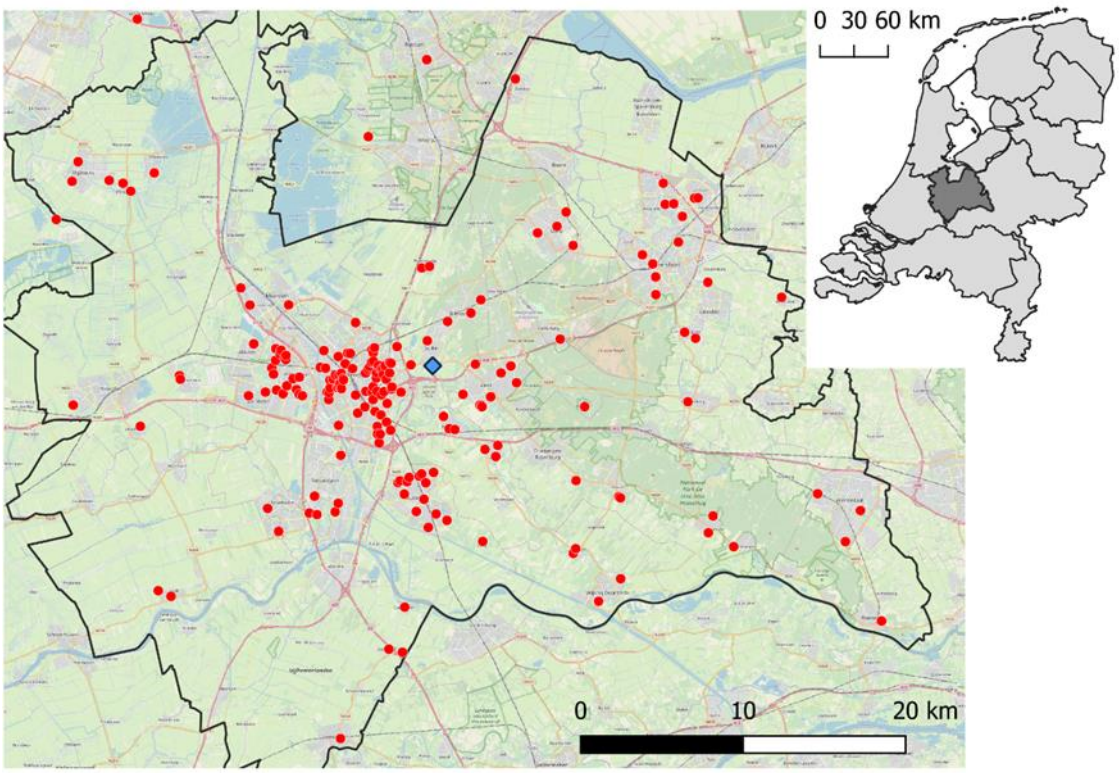


Figure 6: Distribution of PV systems (red dots) and the KNMI weather station (blue diamond) in Utrecht, the Netherlands [23].

4 UP-SCALING METHODS FOR PV POWER PRODUCTION FORECASTING AT REGIONAL LEVEL

Chapter 4 gives an overview of the different approaches adopted for forecasting the PV power generation at regional level.

Five partners contributed with a model (Figure 7):

- UNIROMA2/EURAC
- RSE
- i-EM
- Uppsala University
- Utrecht University

The models differ for the aggregation level of inputs and predicted output, forecast horizon, target geographical area considered (Italy and the Netherlands), as well as the algorithmic core used.

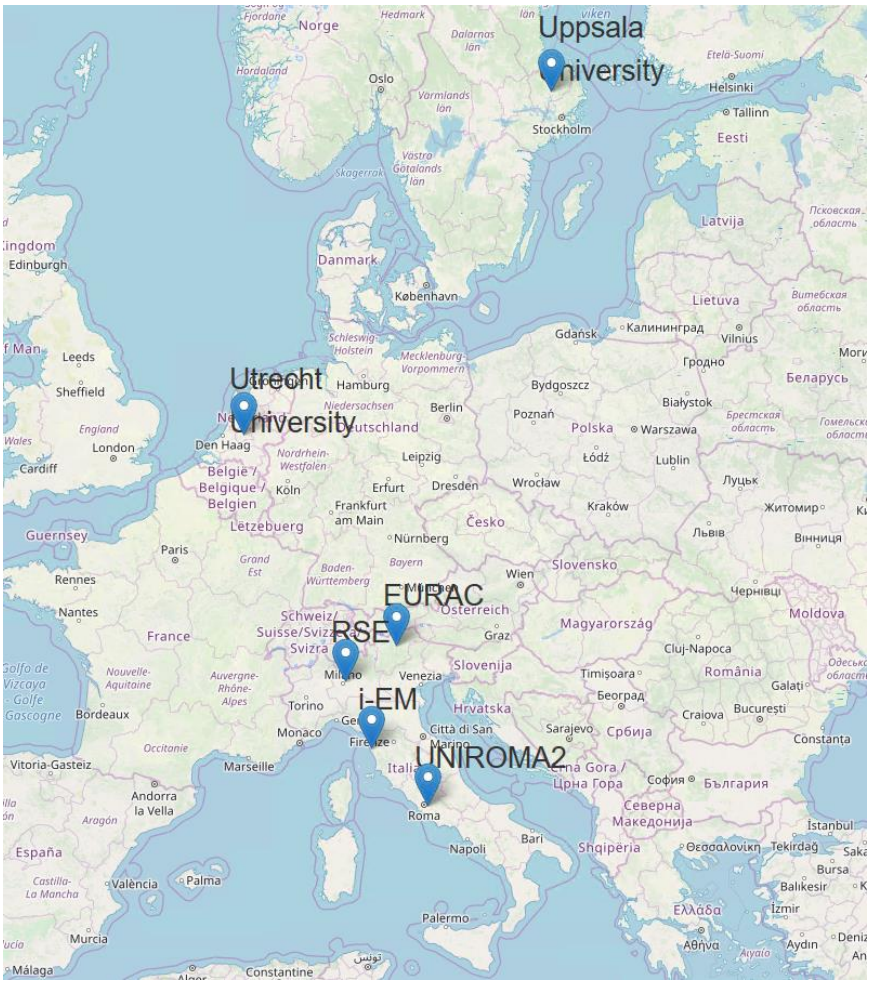


Figure 7: Partners contributing with a model.

The method used for the benchmarking activity is schematically summarized in Figure 8 and Figure 9.

The activity was conducted for three PV power datasets based on the data available and up-scaling models developed by the participants. The schematics of the methodology for the benchmarks are shown in Figure 8 and Figure 9.

Regional forecast was applied to two countries, Italy and The Netherlands. In particular, for Italy, two benchmarking studies were performed: one considering the forecast all over Italy (Figure 8a) and one considering the forecast at market zones level (Figure 8b). For the whole of Italy also a blending of the various models tested was performed.

Figure 9 shows an overview on the benchmarking study by Utrecht University to forecast PV power production of aggregated rooftop PV systems all located in the province of Utrecht. In Table 3 the complete list of models available is reported.

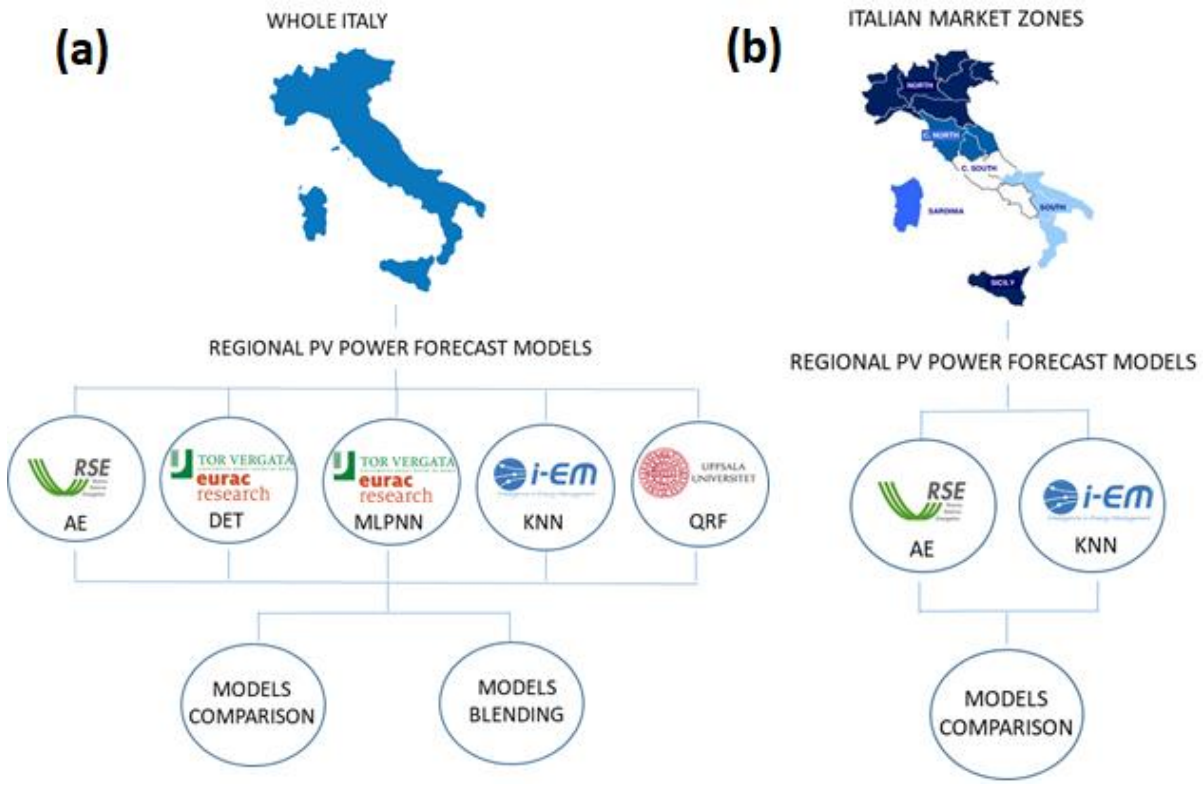


Figure 8: (a) schematic overview of the benchmarking study for the regional forecast all over Italy; (b) schematic of benchmarking study for the regional forecast at market zones level for Italy (for models abbreviations see Table 3 and sections 4.1, 4.2, 4.3, and 4.4).

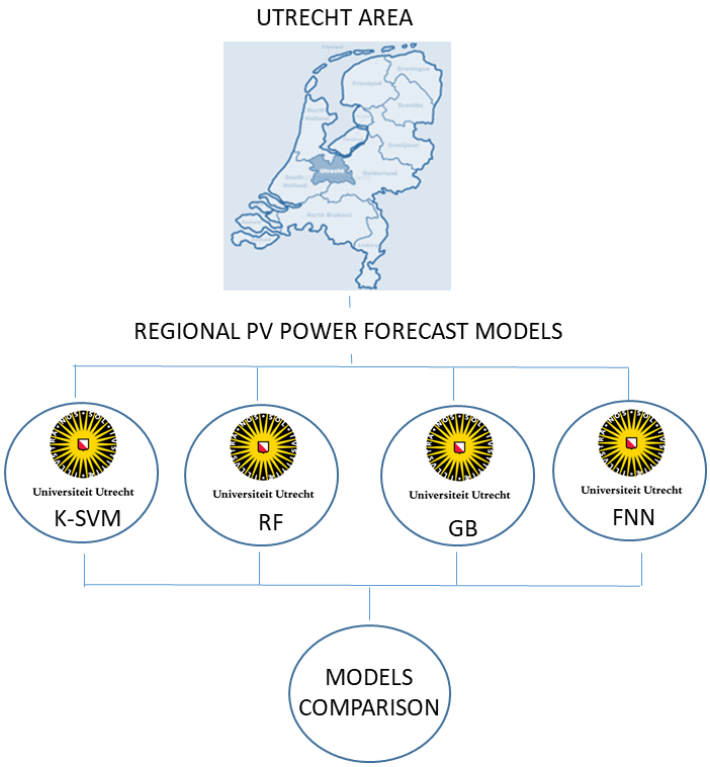


Figure 9: Schematic of the benchmarking study for the Utrecht area forecast (for models abbreviations see Table 3 and sections 4.1, 4.2, 4.3, and 4.4).

Table 3: Overview of models participating to the benchmarking.

Model's Author	Country Area	Target Area	Model Name
i-EM	Italy	Market Zones / Whole of Italy	KNN
RSE	Italy	Market Zones / Whole of Italy	AE
UNIROMA2/EURAC	Italy	Whole of Italy	MLPNN
UNIROMA2/EURAC	Italy	Whole of Italy	DET
UPPSALA University	Italy	Whole of Italy	QRF
Utrecht University	The Netherlands	Utrecht	K-SVM
Utrecht University	The Netherlands	Utrecht	RF
Utrecht University	The Netherlands	Utrecht	GB
Utrecht University	The Netherlands	Utrecht	FNN

For the all over Italy benchmark (Figure 8 a) a comparison between the models set up by i-EM, RSE, UNIROMA2 and Uppsala University was carried out for 1 day-ahead forecasting. The different PV power up-scaling methods of the partners are applied to the same NWP forecast data set of GHI and temperature (section 3.1.2 WRF with MOS). The data from the Italian TSO TERNA for 2015 was used to train the models, while 2016 data was used as a test set. Data was provided by UNIROMA2 (see section 3.1).

At market zones level (Figure 8 b) a comparison between i-EM KNN and RSE - AE models that forecast PV production up to three days ahead was carried out. The same NWP data corresponding to model RAMS were used (section 3.2.3). Data was provided by RSE (section 3.2), where 2016 was used as training dataset and 2017 was chosen to test the models.

Utrecht University compared the performances of different machine learning algorithms based on IFS forecasts by ECMWF for 1 day-ahead forecasting of 150 PV systems in the province of Utrecht (Figure 9), The Netherlands. Data for the period February 2014 until January 2016 was used as training, whereas the period from February 2016 to February 2017 was used for testing.

A detailed description of the models is provided below.

4.1 UNIROMA2/EURAC Models: an Up-scaling Method Based on a Deterministic Physical Model and Artificial Neural Networks

UNIROMA2/EURAC developed two different up-scaling methods: (1) method based on chain of physical semi-empirical models (Direct Insolation Simulation Code [27], Isotropic Transposition model [28] and Sandia Array Performance model [29]) that, for sake of brevity, it is called SAPM; (2) hybrid method based on the parallel working of SAPM and Multi-Layer Perceptron Neural Network model, namely MLPNN.

Here, the physical approach is called "Deterministic"[4] method because, according to the definition of deterministic process, it is based on equations in which no randomness is involved, so that the same input variables produce always the same output. Both the methods make use of pre-processing and post-processing procedures.

Figure 10 depicts a scheme of the two up-scaling methods that have been used for day ahead forecasting of the Italian PV generation.

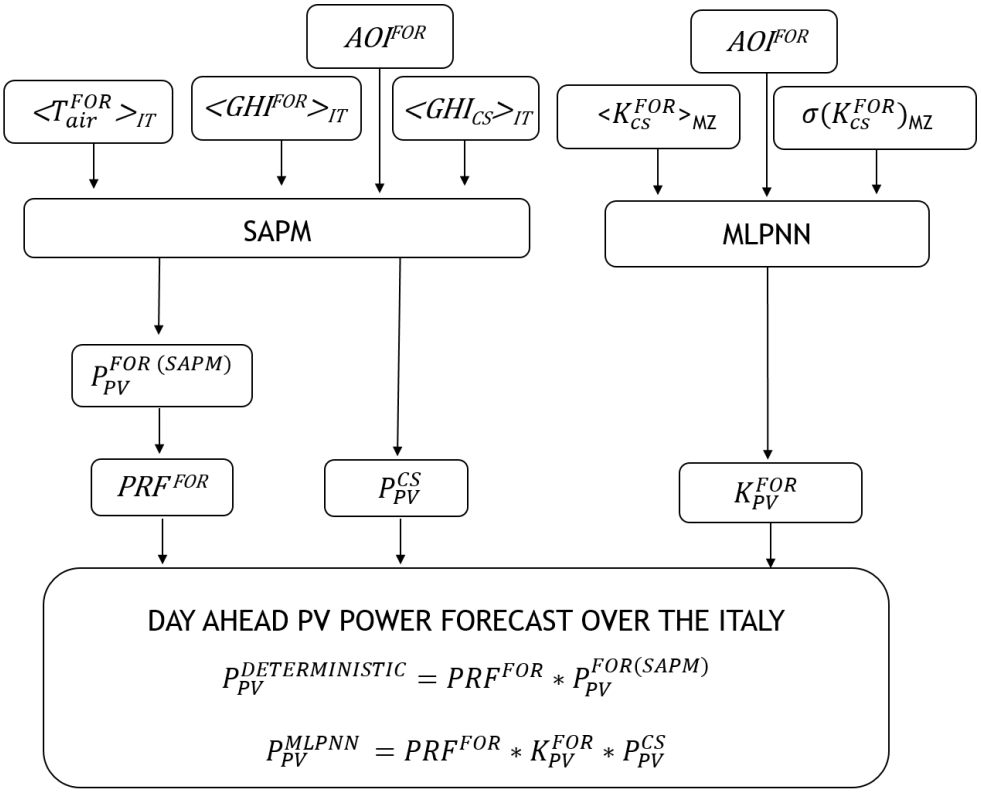


Figure 10: Overview of PV power generation forecasting methods.

[4] It is worth remarking that the same terminology “deterministic” is widely used in the forecast community to indicate non-probabilistic forecast or “point”/”best” forecast.

4.1.1 Pre-processing Procedure

The regional PV generation is considered as if it is produced by a single "virtual" power plant.

The pre-processing procedure aims to retrieve the optimal "equivalent" plane of the array (POA) of this "virtual" PV system. We use the model SAPM to transform the satellite derived GHI averaged all over Italy into the power generated by a virtual PV system with a specific POA. Then, we optimize the POA minimizing the RMSE between the SAPM output and the measured Italian PV generation [30]. The optimal "equivalent" tilted and orientation angles are retrieved using the training year data and then used to predict the equivalent Angle Of Incidence during the test year (AOI^{FOR}).

4.1.2 Deterministic Forecast Up-scaling Model

The deterministic method is just a rescaling of the physical model output (SAPM). In this case, SAPM makes use of the GHI and Tair from the numerical weather predictions (section 3.1.2 WRF with MOS) and of the equivalent AOI forecast.

Therefore, the input and output features are the following:

- Input data:
  - Forecast of the equivalent angle of incidence (retrieved by the pre-processing procedure using the satellite data of the training year): AOI^{FOR};
  - Italy mean values of the NWP model forecasts of the irradiance and temperature: $<GHI^{FOR}>_{IT}$ and $<T_{air}^{FOR}>_{IT}$;
  - Italy mean value of the clear sky irradiance: $<GHI_{CS}>_{IT}$.
- Output data:
  - The Sandia Array Performance Model power forecast over Italy normalized by the PV capacity: $P_{PV}^{FOR(SAPM)}$;
  - Performance correction factor: PRF^{FOR}. This factor was obtained by a SAPM post-processing procedure described in the following subsection 4.1.4;
  - Italy mean value of the PV clear sky generation obtained by SAPM, normalized by the PV capacity: P_{PV}^{CS}.

The deterministic forecast of the PV power output normalized by the installed capacity, will be:

$$\boldsymbol{P_{PV}^{DETERMINISTIC}} = PRF^{FOR} * \boldsymbol{P_{PV}^{FOR(SAPM)}} \qquad (3)$$

This model ingests only the Italian spatial average of the horizontal irradiance and air temperature as input, hence it does not take into account the spatial variability of the irradiance and temperature or the spatial non uniform distribution of the PV capacity.

4.1.3 MLPNN Forecast Model

The hybrid method makes use of the performance correction factor (*PRF*) of the clear sky PV generation obtained by SAPM and of an ensemble of artificial neural networks to predict the PV clear sky index K_{PV} [31], defined as:

$$\boldsymbol{K_{PV} = P_{PV}/(PRF * P_{PV}^{CS})} \quad (4)$$

The artificial neural network architecture is a Multi-Layer Perceptron with one hidden layer (MLPNN). The number of neurons of the hidden layer and the number of model included in the ensemble are obtained by a Master Optimization Procedure described in [32].

The input and output features of the MLPNN model are the following:

- Input data:
  - Forecast of the equivalent angle of incidence (retrieved by the pre-processing procedure using the satellite data of the training year): AOI^{FOR};
  - Average on the Italian market zones of the day ahead forecast of the clear sky index: $<K_{CS}^{FOR}>_{MZ}$, derived from NWP GHI: $<K_{CS}^{FOR}>_{MZ}=<GHI^{FOR}/GHI_{CS}>_{MZ}$
  - Standard deviation on the Italian market zones of the day ahead forecast of the clear sky index: $\sigma(K_{SC}^{FOR})_{MZ}$.
- Output data:
  - Forecast of the PV power generation clear sky index: K_{PV}^{FOR}.

Finally, the PV power generation of the whole of Italy is predicted by the equation:

$$\boldsymbol{P_{PV}^{MLPNN} = PRF^{FOR} * K_{PV}^{FOR} * P_{PV}^{CS}} \quad (5)$$

This model takes into account both the spatial variability of the irradiance (since it inputs both the mean and the standard deviation of the clear sky index of each market zones) and the non-uniform distribution of the PV capacity (since the model will automatically weight market zones inputs according to PV capacity of each zones). The model is trained on the year 2015 and tested on the year 2016.

4.1.4 Post-processing Procedure

The post-processing procedure aims to predict a Performance Correction Factor (PRF^{FOR}) that rescales the SAPM output. This daily rescaling factor accounts for errors in capacity estimation, degradation and AC losses not included in the physical model SAPM and for NWP bias errors.

The performance correction factor is computed using the SAPM forecast ($P_{PV}^{FOR(SAPM)}$) and the PV observed generation (P_{PV}^{OBS}) of the current day (dd):

$$\boldsymbol{PRF^{FOR}(dd+1) = \sum_{h=1,24} P_{PV}^{OBS}(h|dd) \Big/ \sum_{h=1,24} P_{PV}^{FOR(SAPM)}(h|dd)} \quad (6)$$

4.2 RSE Model: the Statistical Method Analog Ensemble

RSE method provides the hourly solar power production on each Italian market zone with a time horizon of three days. The forecast system is based on the statistical technique Analog Ensemble (AE) [33]. The method estimates the probability distribution of the future generation using the measured production registered in the past when the weather scenario looked very similar to the forecasted one. An overview-scheme of the regional forecasting method is shown in Figure 11.

The AE uses as input the forecast of the Global Horizontal Irradiation and the forecast of the module Temperature (T_{panel}).

The GHI derives from the NWP Model Regional Atmospheric Modeling System (RAMS), with lead time between +12 and +84 hours, driven by the global European model Integrated Forecast System (IFS) of ECMWF (see also section 3.2.3). The T_{panel} is calculated on the basis of the air temperature at 2m above the ground and the GHI, both forecasted by RAMS, according to the formula:

$$\boldsymbol{T_{panel} = T_{2m} - 1 + GHI\left(\frac{NOCT - 20}{800}\right)} \qquad (7)$$

where T_{2m} is the air temperature at 2 metres above the ground, and *NOCT* is the Nominal Operating Cell Temperature, set to 45°C.

The RAMS point forecast is averaged on each Italian province (see Figure 5).

The forecasted GHI is processed in order to improve the quality of the GHI forecast.

An AE is applied to the RAMS-GHI, using as observation the GHI estimated from satellite data (see section 0).

The AE, used to forecast the solar power production is trained with the hourly forecast of the corrected GHI and T_{panel}, together with the corresponding solar power measurements on each Italian market zone, referred to the year 2016, whereas the 2017 data set is used to verify the method (see RSE dataset in section 3.2).

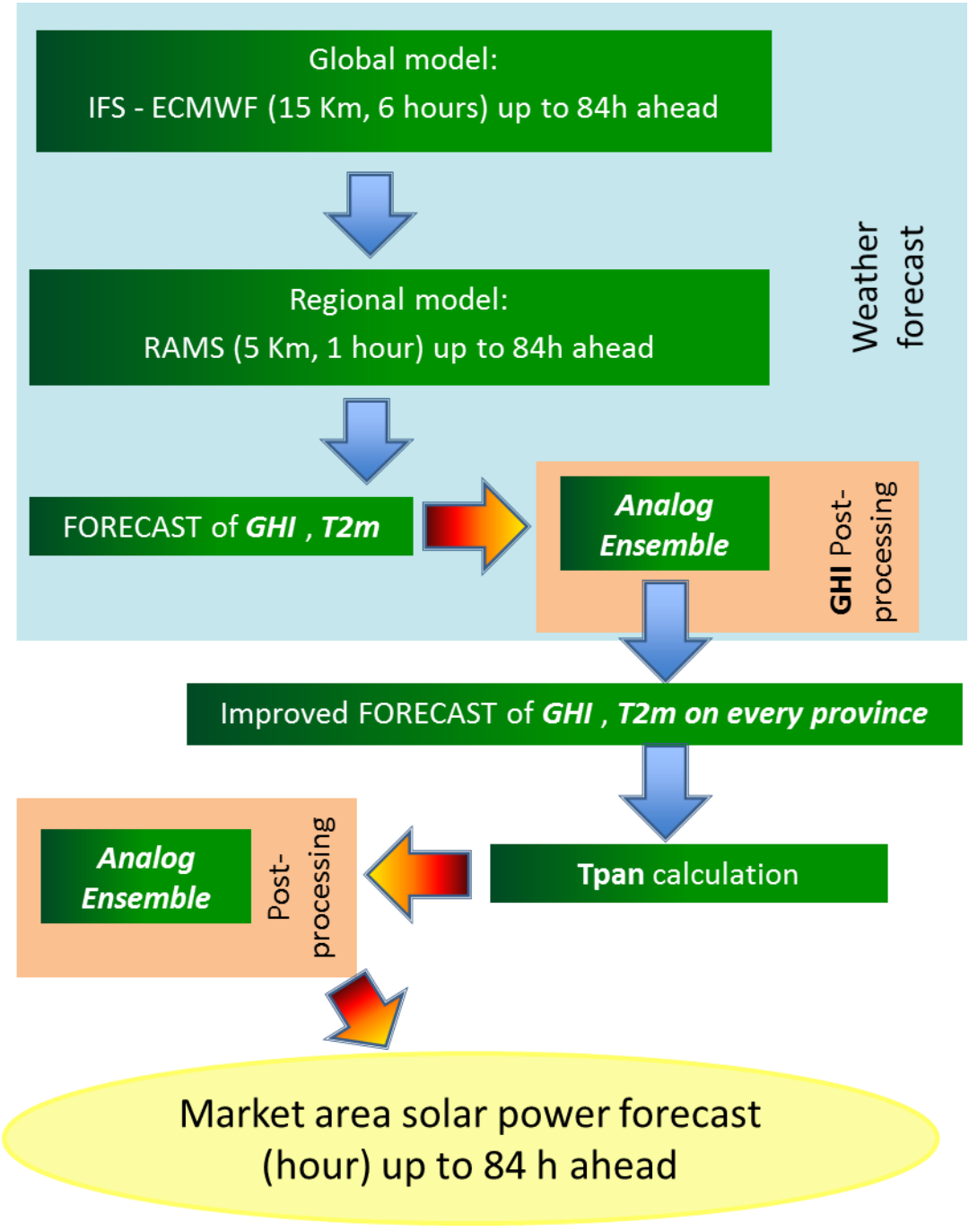


Figure 11: Scheme of RSE method.

The same up-scaling method has been used also to forecast the PV generation of the whole of Italy. In this case, the NWP model input data are the ones provided by the University of Rome "Tor Vergata" described in section 3.1.

In both case studies the point forecast is extracted from the probability distribution of the future generation by using the average.

4.3 i-EM Model: a Cooperative Ensemble of Machine Learning Algorithms

The i-EM PV forecast model is inspired by [34] and aims to predict PV power generation at market zone level of the Italian country, with a forecast horizon up to 4 days ahead.

More in detail, by means of a data-driven approach, it exploits input data aggregated at Province level to forecast the power generation future values for the Italian market zones (see Figure 5).

4.3.1 Datasets

The input and output data of the model have the following features:

- Input data:
  - Clear sky irradiance (GHI_{CS}) and global horizontal irradiance (*GHI*) data aggregated by provinces of Italy (ISTAT 2016 boundaries[5], see Figure 5). In particular, values retrieved by satellite data (see sections 3.1.1 and 3.2.2) are used to train the algorithmic core of the model, whereas NWP model forecasts (see sections 3.1.2 and 3.2.3) are exploited to predict the future power generation of Italian market zones.
  - Historical data of past power generation from market zones, used during the training stage of the model (see section 3.2.1).
- Output data are future power data, aggregated by market zones of Italy.

4.3.2 Model Workflow

The workflow of the model is illustrated in Figure 12, where the main processing steps are shown.

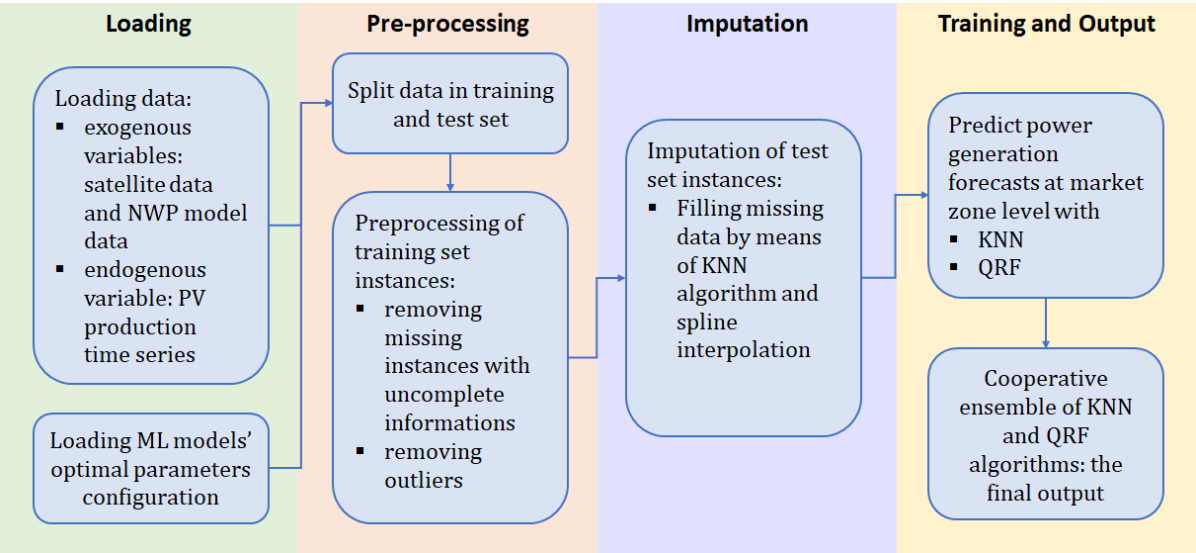


Figure 12: workflow of i-EM forecast model.

The data import phase is a computationally demanding stage, since we have to load a huge and heterogeneous amount of data, and to parse them in a suitable form easy to access.

[5] http://www4.istat.it/it/archivio/209722

The pre-processing is the second stage of the algorithm and concerns training instances, that is the irradiance value from satellite data and the historical values of power generation; it is fundamental to remove outliers and prevent errors from learning wrong patterns in the training set. Outliers are identified by considering the mean irradiance satellite value $\overline{GHI}$ (averaged over all provinces of a specific market zone) with respect to the power generation P of the related area. A linear approximation is considered for the relation between $\overline{GHI}$ and P, hence outliers are identified as those instances are too much discordant from this type of relation (see Figure 13 We also remove timestamps with missing data, because is preferable to do not consider these samples instead of imputing, i.e. estimating, missing data with the possibility of introducing wrong information during the training stage.

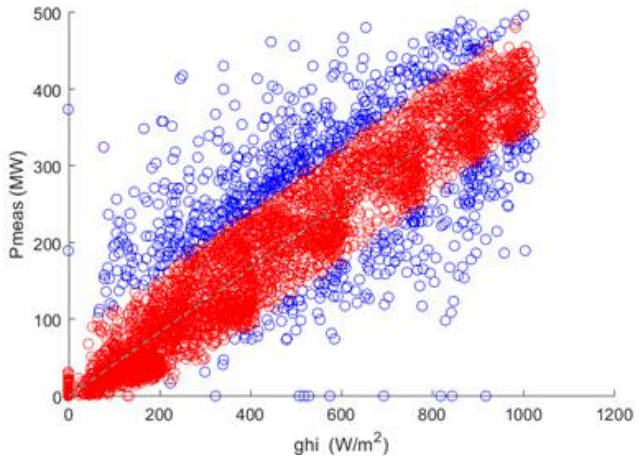


Figure 13: Example of the implementation of the pre-processing procedure: the blue points are related to the outliers' instance, whereas the red ones represent the good samples of the training set.

The core of the model is represented by pure data-driven methodologies: k-Nearest Neighbours (KNN) and Quantile Regression Forest (QRF). These algorithms are trained independently, and both of them take as input NWP model data, and produce power generation forecast values. Finally, a cooperative ensemble of these two algorithms produces the final output of the model. More in detail, the final power value is obtained as a convex combination of the outputs of KNN and QRF algorithms, with an optimal configuration of weights. In the next sections, this ensemble will be shortly denoted as KNN (since QRF is also used by another model involved in the benchmarking study).

Data used for PV power generation forecasting at market zone level are those provided by RSE (see section 3.2). The same up-scaling method has been used also to forecast the PV generation of the whole of Italy: in this case the NWP input data are the one provided by the University of Rome "Tor Vergata" described in section 3.1.

4.4 Uppsala's University Model: Quantile Regression Forests and the Principal Component Analysis

This section describes a dimension reduction technique and a probabilistic forecasting technique that Uppsala University applied to the data provided by the colleagues of UNIROMA2/EURAC. This data set contains numerical weather prediction (NWP) forecasts of the GHI for 24 hours ahead for the entirety of Italy and consists of 1'325 grid points (see Section 3.1). The dimensionality of this data set requires some form of dimensionality reduction technique and principal component analysis (PCA) is a relatively straightforward way to address the aforementioned issue.

It is important to note that although the focus of this section lies on probabilistic forecasting by means of quantile regression forests (QRFs), it is still possible to acquire a point forecast, i.e. a unique "best" forecast trajectory. Herein, we evaluate the mean as the deterministic forecast although other statistics are also possible, e.g., the median.

4.4.1 Principal Component Analysis

In order to avoid overfitting and to reduce training time, Uppsala University exploits PCA to reduce the dimensionality of the data matrix $\boldsymbol{X}$. PCA transforms the data matrix to a new coordinate system such that the greatest variance can be found on the first axis, which is referred to as the first principal component [35]. The subsequent axes, which are referred to as the subsequent principal components, are projected orthogonally on the preceding principal component and are selected such that each principal component maximizes the variance found in the data. This procedure results in a series of vectors that form an uncorrelated orthogonal basis set [35]. Readers interested in the technicalities may refer to [35]. A suitable number of components is determined with cross-validation, see Section 4.4.3.

4.4.2 Quantile Regression Forests

Uppsala University exploits QRFs to quantify the uncertainty by means of the predictive distribution, proposed by Meinshausen [36]. It is an extension of random forests (RFs) developed by Breiman [37]. A feature of regression trees is that they produce predictions with low bias but high variance. The aforementioned issue can be improved by bootstrap aggregation where the average of many trees results in the final prediction [35]. In order to reduce the variance, each tree is grown on a random subset of the explanatory variables. This allows RFs to predict with comparable accuracy as boosted trees, but more straightforward to train and tune [35]. The interested reader is referred to [36] for technical details. The R package quantregForest [38] is used to produce the probabilistic forecasts.

4.4.3 Cross Validation

Herein, cross validation is used to select the appropriate number of principal components to include in the design matrix $\boldsymbol{X}$. In addition, the number of trees and number of randomly sampled variables at each split are considered in the cross-validation. To this end, training data is divided into *K* sequential train and test folds and performances are evaluated on each test fold by means of a numeric score. This implies that Uppsala University iterates over an increasing number of principal components $p \in \{1, \dots, P\}$ and over $k \in \{1, \dots, K\}$ folds. This results in a score—averaged over *K* folds—for each set of principal components $p \in \{1, \dots, P\}$. Figure 14 presents the workflow of the model described herein. It should be noted that the cross-validation depicted in Figure 14 is first performed for the principal components and then for the hyperparameters (i.e., the number of trees and the number of randomly sampled variables) of the QRF model. The number of principal components *P* resulting in the lowest

score is 21, while the hyperparameters had relatively little impact on the accuracy of the model.

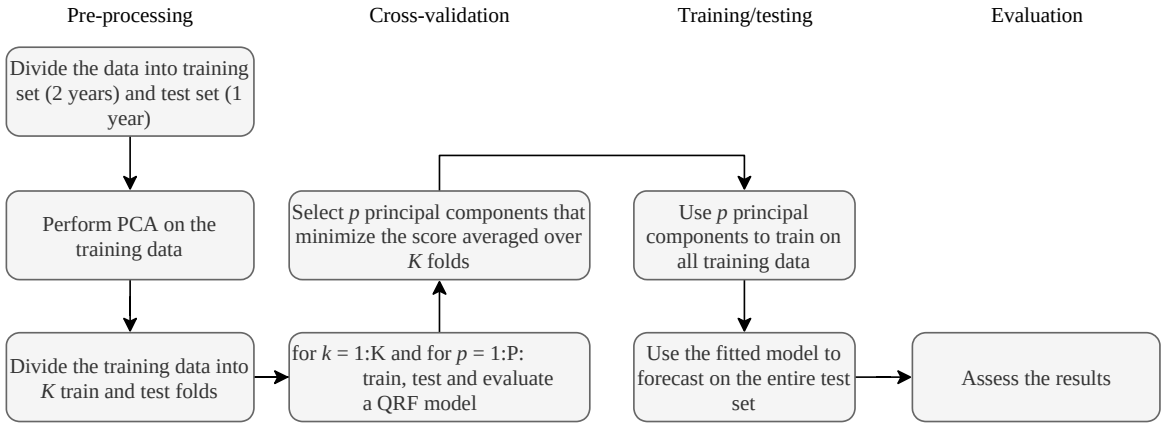


Figure 14: Workflow of the model, including data pre-processing, cross validation and training/testing.

4.5 Utrecht University (UU) Model: Machine Learning Method for Aggregated PV-systems

The UU Regional PV forecast model generates hourly day ahead predictions of the PV power production of 150 PV systems that are distributed throughout the province of Utrecht, the Netherlands, and covering an area of 38 by 54 kilometers (see Figure 6). The models are data-driven and can therefore operate online without any interference of the user. Also, no user inputs are needed to initialize the models.

The regional PV power forecast models take as an input 1) weather predictions extracted from the NWP model IFS of the ECMWF [22] and 2) values that account for predictable seasonal and diurnal variations in the solar irradiance (see section 3.3).

It is important to note that if the total installed PV capacity for the province of Utrecht is known, the models can be configured accordingly in order to forecast the total PV power production in the province. In addition, the models can be applied to other locations by changing all input values according to the region of interest. In order to forecast the day-ahead solar power production for 150 aggregated PV-systems in Utrecht several steps are taken. The model workflow is shown in Figure 15 and discussed in the following subsections.

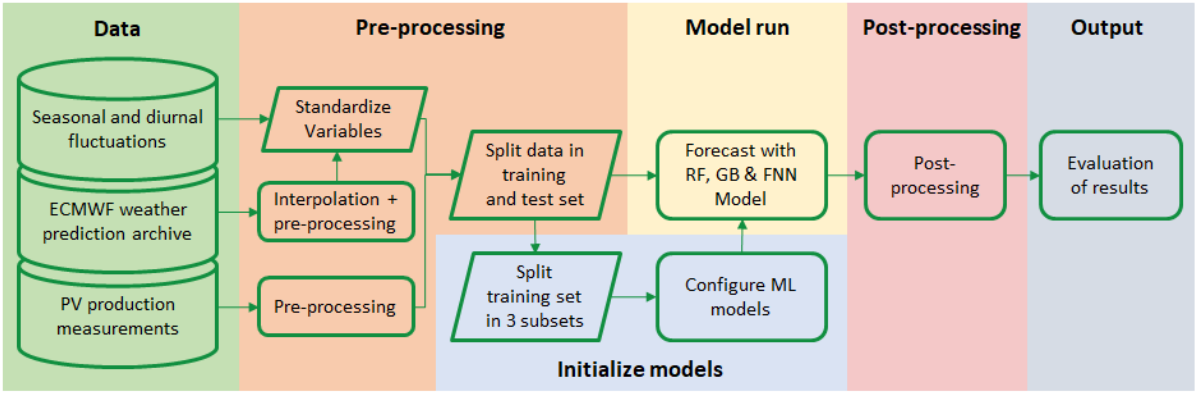


Figure 15: Workflow of the UU regional solar power forecast model.

4.5.1 Data Collection

All variables of interest are described in section 3.3.

4.5.2 Pre-Processing

Pre-processing of the data is done in several steps. First, since weather predictions are only available for 3 hourly steps, hourly values are obtained by interpolating the original values with Akima interpolation [24]. Afterwards, outliers are removed by considering predefined minima and maxima per variable, which are extracted from the original values.

The second step considers standardization of all input variables, see Eq. (8).

$$v_{i,t} = \frac{x_{i,t} - \mu_i}{\sigma_i}, \tag{8}$$

where $x_{i,t}$ is the value of variable i at time t, μ_i and σ are respectively the mean and standard deviation of variable i.

Furthermore, the collected PV power measurements are pre-processed. First, the one-minute power measurement samples are converted into hourly production values by averaging these minute samples for each hour. Subsequently, outliers are removed while considering the installed capacity of the reviewed PV systems. From the dataset missing values are also removed. Thereafter, the PV power production is aggregated for the 150 individual sites.

Finally, the input variables and PV power measurements are merged together and split into a separate training dataset containing two years of data, which considers the period February 2014 until January 2016. Subsequently, the consecutive year is taken as the test dataset, from February 2016 until January 2017.

4.5.3 Model Configuration

The models considered are Support Vector Machine (SVM), Random Forest (RF), Gradient Bosting (GB) and Feed-Forward Neural Network (FNN). After completing the pre-processing steps, the optimal configuration of the just mentioned models' parameters are determined. For this, a k-fold cross validation method with K=3 is used. This step implies that the training dataset is split into three separate subsets. Since the test dataset is not considered here, each subset holds 8 months of data. The optimal parameter settings for the forecasting models are found by combining two of these subsets as a training set while validating the results on the third set, for all three possible combinations. The result of the k-fold cross validation step determines the optimal parameter setting for each model, which are used to initialize the forecast model by tuning all of the model parameters [26].

4.5.4 Model Run

The models simulated include the already introduced SVM, RF, GB and FNN. After these models have been initialized, the input data of the test dataset is fed into the forecasting models. In addition, a persistence model is added such that the results can be benchmarked.

4.5.5 Post-processing

In this stage the output of the PV power production forecast models is post-processed. In this step outliers are removed from the predictions by considering minimum and maximum production values that could be obtained under clear sky conditions during the hour of interest.

4.5.6 Output

The last step in the process considers evaluation of the forecast accuracy. The machine learning models generates aggregated day-ahead solar power forecasts for a total of 150 PV-systems in the province of Utrecht, the Netherlands (see Figure 6). The models produce forecasts with an hourly resolution.

4.6 Overview of Models and Benchmark Studies

An overview of the analyzed models in three different benchmarking studies is given inTable 4.

Table 4: Overview of benchmarking studies and models.

FORECASTING STEPS / ACCURACY / CASE STUDIES	(1) DATA PRE-PROCESSING	(2) NWP	(3) INPUTS AGGREGATION	(4) PV POWER UP-SCALING MODELS	(5) POST-PROCESSING	SIZE OF THE AREA [10^3 km^2]	UP-SCALING METHOD NAME
PROVINCE OF UTRECHT (test period from February 2016 to February 2017)	SPATIAL AND TEMPORAL INTERPOLATION OF NWP TIME AGGREGATION AND EXCLUSION OF OUTLIERS FROM POWER MEASUREMENTS	IFS OF ECMWF	-	SUPPORT VECTOR MACHINE	REMOVE OUTLIERS FOR POWER PREDICTIONS	1.4	K-SVM
				RANDOM FOREST			RF
				GRADIENT BOOSTING			GB
				FEED FORWARD NEURAL NETWORK (MLP)			FNN
MARKET ZONES OF ITALY (test period 2017)	EXCLUSION OF OUTLIERS FROM POWER MEASUREMENTS	MESOSCALE MODEL (RAMS) + MOS	AGGREGATION BY PROVINCES	ANALOG ENSEMBLE	-	50 (zones average)	AE
				KNN + QUANTILE REGRESSION FORESTS	POWER OUTPUT RESCALING		KNN
ITALY (test period 2016)	-	MESOSCALE MODEL (WRF) + MOS	PRINCIPAL COMPONENT ANALYSIS	QUANTILE REGRESSION FORESTS	-	300	QRF
	-		AGGREGATION BY PROVINCES	KNN + QUANTILE REGRESSION FORESTS	POWER OUTPUT RESCALING		KNN
	RETRIEVE EQUIVALENT POA		AGGREGATION BY ITALY	DETERMINISTIC (SAPM)	PERFORMANCE RESCALING		DETERMINISTIC
	-		AGGREGATION BY PROVINCES	ANALOG ENSEMBLE	-		AE
	RETRIEVE EQUIVALENT POA		AGGREGATION BY MARKET ZONES	FEED FORWARD NEURAL NETWORK ENSEMBLE (MLPNN)	PER-FORMANCE RESCALING		MLPNN

4.7 Models Blending

Models blending methods or Multi-model approach is a method to properly combine different forecasts. Recent papers show that it can improve forecast accuracy [39, 40, 41, 42, 43]. Nevertheless, this method is not fully investigated.

The simplest method is the linear blending:

$$P_{blend}^{for} = \sum_{i=1}^{N} W_i * P_i^{for} \quad (9)$$

where, N is the number of the predictions in the ensemble, P_i^{for} is the forecast of the model (i) of the ensemble and W_i is the weight associated to this model according to the condition: $\sum_{i=1}^{N} W_i = 1$.

If the models are bias free (i.e. if MBE_i= 0 for all models i), it can be easily proven that:

$$RMSE_{blend} \cong \sqrt{\sum_{i=1,N}\sum_{j=1,N}\left(W_i * W_j * RMSE_i * RMSE_j * \rho_{ij}\right)} \quad (10)$$

where, $RMSE_i$ is the root mean square error of the prediction (i) and ρ_{ij} is the correlation between the error of the model (i) and of the model (j). Thus, the accuracy of a linear blending depends on the accuracy (RMSE) of the models included in the ensemble and on the error correlation together with the assigned weights. If $W_i = \frac{1}{N}(\text{with i} = 1, \text{N})$, the linear blend in Eq. (9) is the simple average of the predictions.

More complex methods are based on a non-linear function of the forecast trajectories: $P_{blend}^{for} = f(P_i^{for})$. In this case, the accuracy of these methods cannot be computed a priori.

In this work, we develop different methods to blend day-ahead predictions of the PV generation all over Italy.

The first method is the simple forecast average (avg):

$$P_{avg\ blend}^{for} = \sum_{i=1}^{N} P_i^{for} \Big/ N \quad (11)$$

The second method is a smart linear blend (SL) in which the weights depend on K_{PV}^{for} predicted by the best forecast of the ensemble:

$$P_{SL\ blend}^{for} = \left(\sum_{i=1}^{N} W_i(K_{PV}^{best\ for}) * K_{PV}^{for(i)}\right) * P_{cs} \quad (12)$$

where, $K_{PV}^{best\ for}$ is the forecast of the PV clear sky index obtained by the MLPNN model, $K_{PV}^{for(i)}$ is the PV clear sky index predicted by the model (i) and P_{cs} is the clear sky PV power output (average all over Italy). For each $K_{PV}^{best\ for}$ interval ($< 0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, > 0.8$) the weights $W_i(K_{PV}^{best\ for})$ have been computed through a brute force minimization that searches the minimum $RMSE$ of the blend over 20^6 possible W_i values. Therefore, globally the brute force algorithm explores $5x20^6$ possible W_i values.

The third method (NN) is based on an artificial neural network non-linear function of the prediction included into the ensemble:

$$\begin{cases} P_{NN\ blend}^{for} = K_{PV}^{NN\ blend} * P_{cs} \\ K_{PV}^{NN\ blend} = f^{(2)}(W^{(2)} f^{(1)}(W^{(1)} K_{PV}^{for} + b^{(1)}) + b^{(2)}) \end{cases} \quad (13)$$

where (i= 1, 2) is the layer index, $f^{(i)}$ are transfer functions, ($W^{(i)}$) are the weights matrices, ($b^{(i)}$) are the bias vectors and K_{PV}^{for} is the vector of the predicted PV clear sky index (used as NN input).

Since only one year of forecast data was available for test (year 2016), for the training, validation and test of the SL and NN blending methods, a K-folder cross validation procedure was used (with K= 10). Therefore, data have been divided in 10 distinct sets and for ten times, 9 sets were used for training and validation of the model (with proportions 70% - 30%, respectively) and 1 set for test. In this way, all the 10 test sets represent the blending forecast of the whole year 2016 obtained by 10 different models (always trained and validated on a data set different from the test set).

5 RESULTS OF BENCHMARK FOR ITALY

5.1 Performance Metrics

In order to evaluate the PV power forecasting results independently of the installed capacity, the forecasts and measurements are normalized to the installed capacity in the area.

The error at a given hour and in a specific geographical area is obtained as:

$$e_i = \frac{P_i^{for} - P_i}{P_n}, \quad [MW/MWp] \tag{14}$$

where P_i^{for} and P_i are the forecasted and measured power production corresponding to i-th hourly instance, respectively, and P_n is the hourly installed capacity in the area.

The following error metrics are used as Key Performance Indicators (KPIs) for the evaluation of the forecasting performance. They are the Pearson Correlation (CORR), the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE) and the Skill Score (SS) here:

$$CORR = \frac{COV(X^{FOR}, X^{OBS})}{\sigma_{X^{FOR}}\, \sigma_{X^{OBS}}} \; [-] \tag{15}$$

$$RMSE = 100 \times \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2}, \quad [\%\ of\ Pn] \tag{16}$$

$$MAE = 100 \times \frac{1}{N}\sum_{i=1}^{N} |e_i|, \quad [\%\ of\ Pn \tag{17}$$

$$BE = 100 \times \frac{1}{N}\sum_{i=1}^{N} e_i, \quad [\%\ of\ Pn] \tag{18}$$

$$Skill\ Score = 100 \times \left(\frac{RMSE(ref) - RMSE(for)}{RMSE(ref)}\right), \quad [\%] \tag{19}$$

where N is the number of daylight hours in a year. Night values with no irradiance have been discarded from the evaluation, as common in the literature, since the only effect they produce is a decrease of the error metrics. The persistence obtained by propagating the last day available prior to the test day for the whole forecast horizon is considered as the reference (*ref*) forecasting method.

5.2 Benchmark Results of the Day-ahead Forecasts of the PV Generation of all the Italian Market Zones

Two models are compared for the Italian Market zones here: the Analog Ensemble (AE) model of RSE and the ensemble of kNN and QRF (shortly denoted as KNN here) by i-EM.

The Italian TSO divided Italy in six market zones (Figure 16a) characterized by bottlenecks of the transmission grid. For this reason, these zones are often subjected to power congestions and constraints on power capacity.

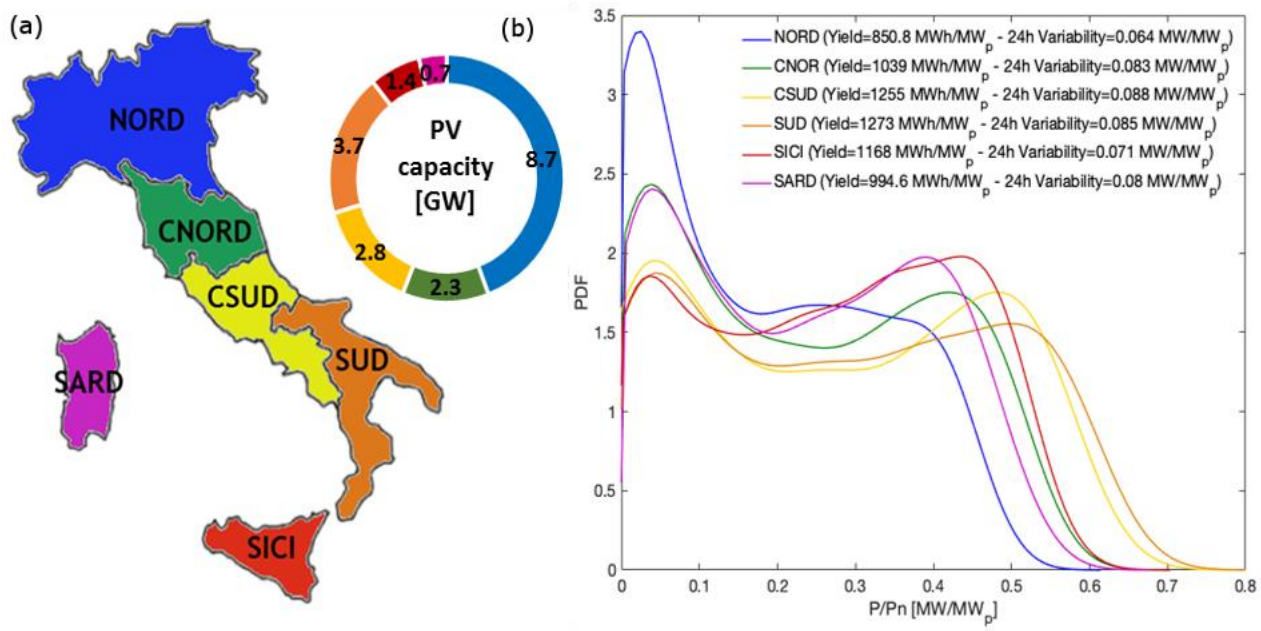


Figure 16: Italian market zones and PV capacity at 2017 (a); Distribution of the PV power generation in each market zone (b). For definition of variability see Eq. (20).

In 2017 the total PV installed capacity in Italy was 19.7 GWp, almost 45% is installed in the North (NORD), 12% and 14% in the centre North and South (CNORD and CSUD), 19% in the South and 7% and 4% in the islands (SICI and SARD) . The minimum energy yield is in the Northern zone with 851 MWh/MWp during the year 2017, where a flat generation probability (that occurs during the mid-day hours) appears between 0.25 and 0.4 MW/MWp (Figure 16b). The maximum yield is in the Southern zone with 1273 MWh/MWp and a prevalent PV generation between 0.4 and 0.55 MW/MWp. The lowest weather variability conditions are found in the North and in Sicily with 6.4% and 7.1% of the PV installed capacity, respectively (Figure 16b). The first because there are many subsequent days with overcast conditions, while the second because of many subsequent days with clear sky conditions. The highest PV generation variability is found in the Centre South zone (8.8 % of the PV installed capacity) where a more frequent change of weather conditions occurs from one day to the next.

In order to analyse the performance of the different methods in forecasting the hourly production on each Italian market zone, a suitable graph is the Taylor diagram.

The Taylor diagram in Figure 17 shows, for each market zone, the Pearson correlation between forecasts and observations and the standard deviation of the observation and of the corresponding forecast. The standard deviations are all between 0.15 and 0.2 MW/MWp, in agreement with Figure 16b, the minimum is in the North (low maximum power generation) and the maximum is in the South (high maximum power generation). The correlation between forecasts provided by the two models AE and KNN and observations range between 0.92 and 0.96. The closer the points in the diagram are to the observations point of the corresponding market zones, the more accurate are the predictions. The two forecasts always outperform the persistence. The Analog Ensemble (AE) model provides the more accurate prediction in all the zones with the exception of the Centre North (CNORD) area. Nevertheless, this model underestimates the variability of the PV generation both in the North and in the South, predicting a standard deviation lower than the observed one.

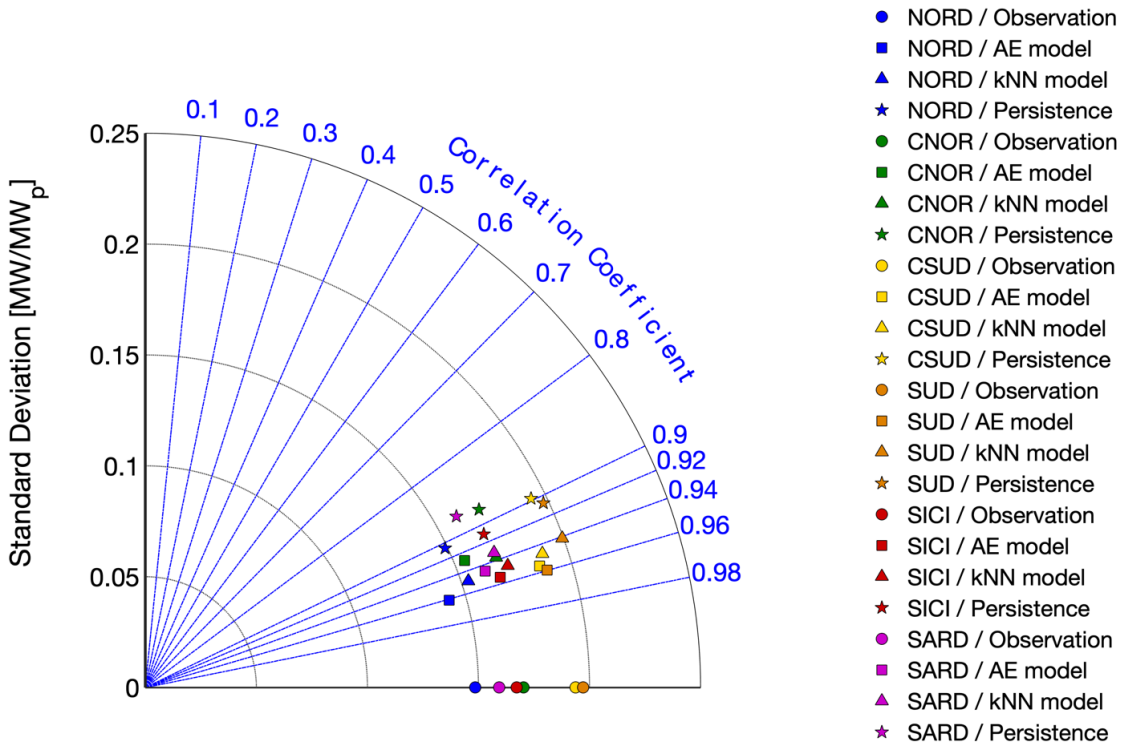


Figure 17: Taylor diagram of the performance of the forecast models AE and KNN in each market zone.

The scatter plots in Figure 18 show a remarkable pattern in the forecast errors of the KNN model in all the market zones. The model predicts a constant power out at high PV generation levels. A similar behaviour is found in Sicily also for the AE model but the upper limit is higher than for the KNN model. Generally, the scatter is larger for the KNN model than for the AE model.

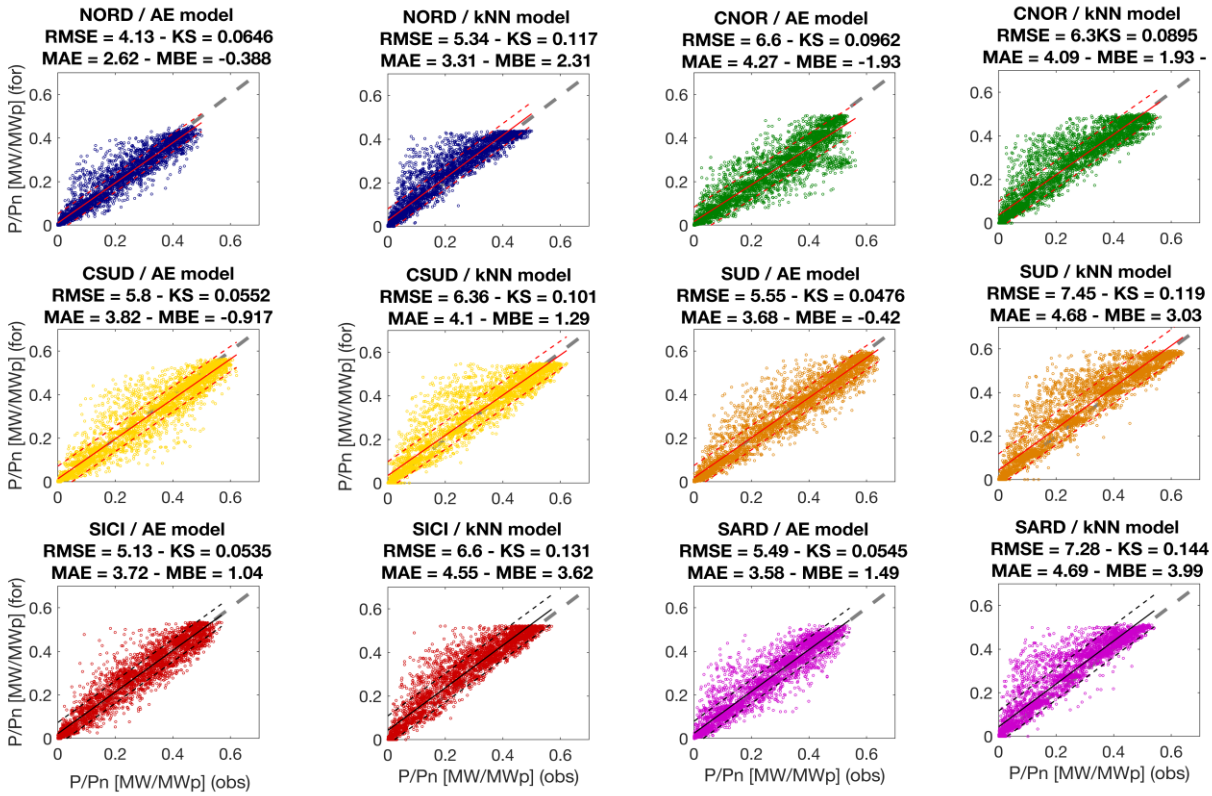


Figure 18: Scatter plots of the 1 day-ahead PV power forecasts of each market zone obtained by the AE and KNN models. The metrics RMSE, MAE, MBE and Kolmogorov-Smirnov (KS) index are also shown.

The KNN forecast overestimates the observations with a MBE between 1.2% (CSUD) and 4% (SARD) of the PV capacity and the RMSE ranges between 5.3% (NORD) and 7.4% (SUD). The AE forecast obtains a MBE and RMSE that range from -1.9% (CNOR) to 1.5% (SARD) and from 4.1% (NORD) to 6.6% (CNORD), respectively (Figure 18).

With the exception of the Centre North (CNOR) zone, the AE forecast always shows a lower RMSE and correspondingly also reaches a better skill score (Figure 19).

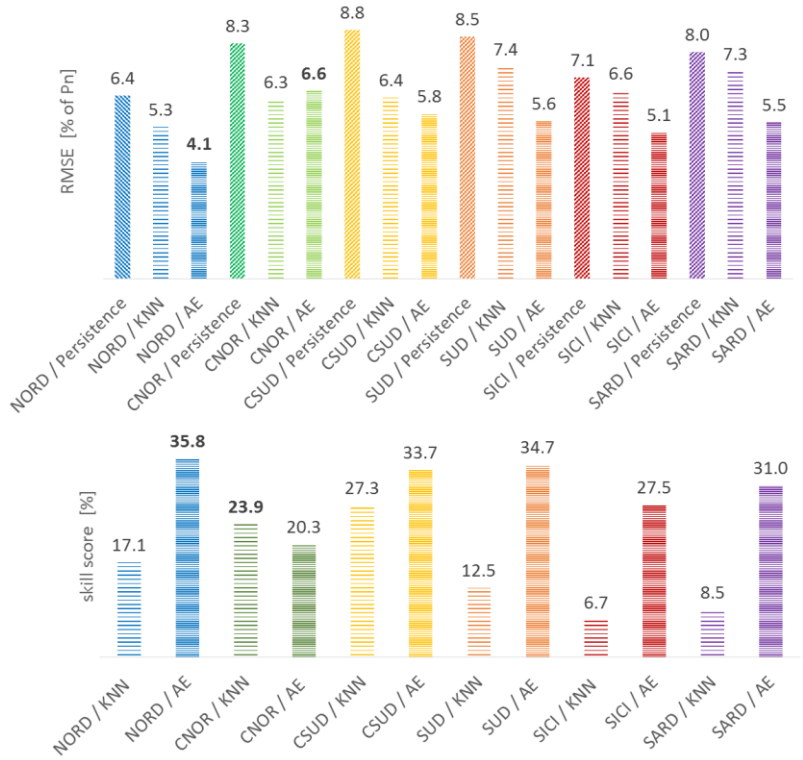


Figure 19: RMSE of the forecasts (top) and skill score (bottom) with respect to the persistence.

Figure 20 shows the forecast errors of the two models at different forecast horizons: 1, 2 and 3 days ahead. It shows that the forecast errors do not notably change when the forecast horizon changes from 1 to 3 days. Thus, the value of the one day-ahead prediction accuracy can be considered representative of the forecast performance in the whole horizon range.

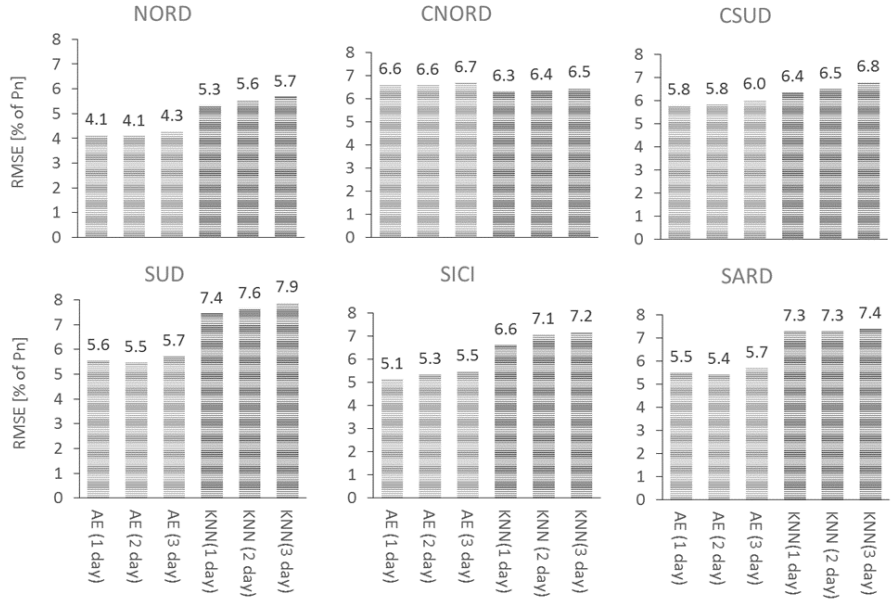


Figure 20: RMSE of the PV power forecasts of each market zone obtained by the AE and KNN models at different horizons.

5.2.1 Dependency of Forecast Errors on Area Size

A high forecast accuracy can be reached by simply enlarging the foot print of forecast controlled area [44, 45, 46, 47, 48]. This phenomenon is called “smoothing effect” and it is related to the correlation of the forecast errors of different locations. The more the regions include a wide range of locations with different irradiance variability, the more the solar forecast errors will be uncorrelated and the higher will be the accuracy of the regional PV power forecast. Figure 21 depicts the day-ahead forecast accuracy that can be reached predicting the PV generation of different control areas. The RMSE of the persistence model represents the PV power variability in each zone (as defined in [48]):

$$\boldsymbol{\sigma(\Delta P_{\Delta t})} = \sqrt{\langle(\Delta \boldsymbol{P_{\Delta t}} - \langle \Delta \boldsymbol{P_{\Delta t}}\rangle)^2\rangle} = \sqrt{\langle(\langle \boldsymbol{e^{PM}}\rangle - \boldsymbol{e^{PM}})^2\rangle} = \sqrt{\langle(\boldsymbol{e^{PM}})^2\rangle} = \boldsymbol{RMSE(persistence)} \quad (20)$$

where Δt is the ramp rate (in this case 24 hours), $P(t)$ is the PV power output at the time t (in this case hour), $P^{PM}(t) = P(t - \Delta t)$ is the PV power forecast of the Persistence Model (PM) with horizon Δt and $\Delta P_{\Delta t} = P(t) - P(t - \Delta t) = P(t) - P^{PM}(t) = -e^{PM}(t)$. In Eq. (20) it is assumed $\langle e^{PM}\rangle = 0$.

Both the variability and the forecast errors decrease with the size of the region. They can be well fitted either by a hyperbolic function (similar to one proposed by Perez et al. in [48]) or by an exponential function (similar to the one proposed by Lorenz et al. in [46] [49]):

$$\boldsymbol{RMSE(Surf) \cong a/(1 + b\sqrt{Surface})} \quad (21)$$

$$\boldsymbol{RMSE(Surf) \cong ae^{-b\sqrt{Surface}}} \quad (22)$$

Enlarging the footprint of the forecast passing from the prediction of the PV generation in each market zones to the prediction of the PV generation all over Italy, will decrease the RMSE from 5.5% (market zones average) to 3.6% for the AE model.

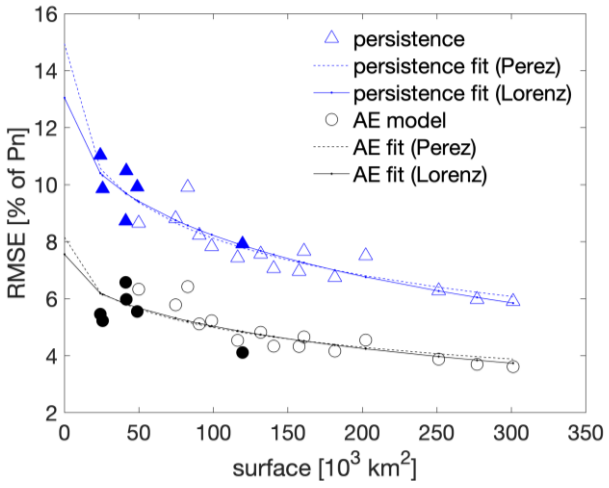


Figure 21: Smoothing effect all over Italy (full dots are the RMSE of the AE forecast of the PV generation of each market zone, while empty dots are the RMSE that could be reached merging two or more adjacent market zones).

Therefore, the expansion of the transmission grid to manage the power generation of whole of Italy not only reduces congestions and constraints on production capacity between the market zones (as required by the National Energy Strategy plan) but also increases the forecast accuracy by 34% (relative value). For this reason, it is important to develop models able to predict directly the solar generation of the whole Italian PV fleet.

5.3 Benchmark Results of the Day-ahead Forecasts of the PV Generation all over Italy

For the benchmarking study for all of Italy five different PV Power up-scaling models are compared, all applied to the same NWP data set (WRF with MOS). The models include again the AE model of RSE and the kNN model by i-EM, and additionally the MLPNN and the "Deterministic"model by Uni Roma2/Eurac and the QRF model by Uni Uppsala.

The Taylor diagram in Figure 22 shows the Pearson correlation between forecasts and observations and the standard deviation of each forecast normalized by the standard deviation of the observations. The outperforming forecast model is the MLPNN followed by the AE. Both the models provided forecasts with 0.98 of correlation and a standard deviation almost equal to the observations. The least accurate model again is persistence that is used as reference.

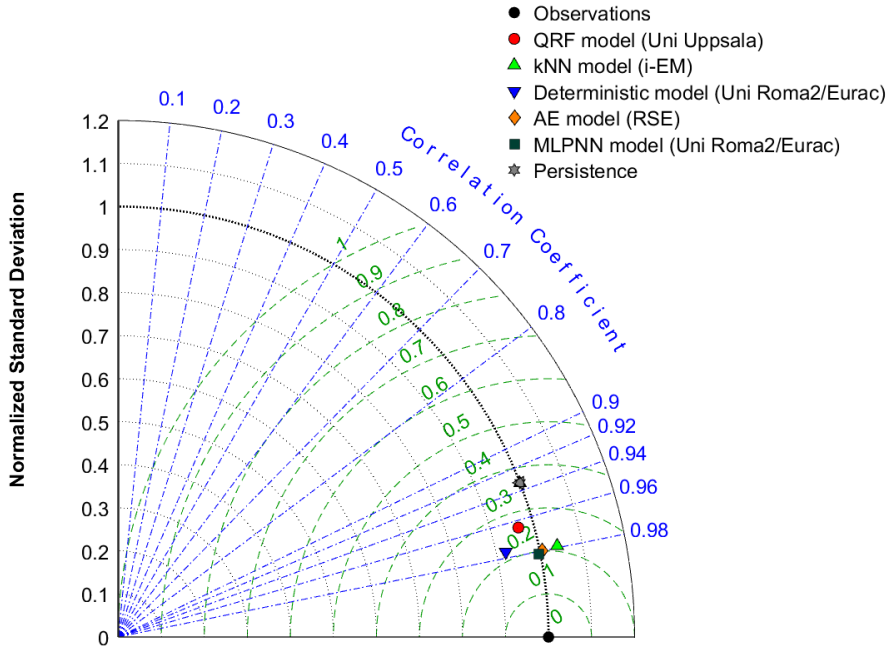


Figure 22: Taylor diagram of the forecast models performance. Green dashed lines represent the standard deviation of the forecast errors (expressed as fraction of the standard deviation of the observed data).

Except for KNN and Deterministic, the other models produce a forecast with very similar positive and negative errors distributions (Figure 23). All predictions are almost unbiased apart from KNN. KNN model provides a remarkably over forecast with a positive mean bias error of 1.9% of the nominal capacity (P_n). On the other hand, it should be possible to improve the forecast accuracy of this model just removing the systematic error.

The QRF, KNN and deterministic models obtain a RMSE ranging between 5% and 4% of P_n and a MAE between 3.6% and 3% of P_n (Figure 24). KNN is less accurate than the deterministic model in terms of RMSE, but is outperforming in terms of MAE. This indicates

that the KNN forecast has more outliers than the deterministic forecast. The cumulative density function (CDF) of the QRF forecast differs from the observed CDF at different PV generation levels with a maximum absolute difference of 0.036 (Kolmogorov-Smirnov index (KS) [50]).

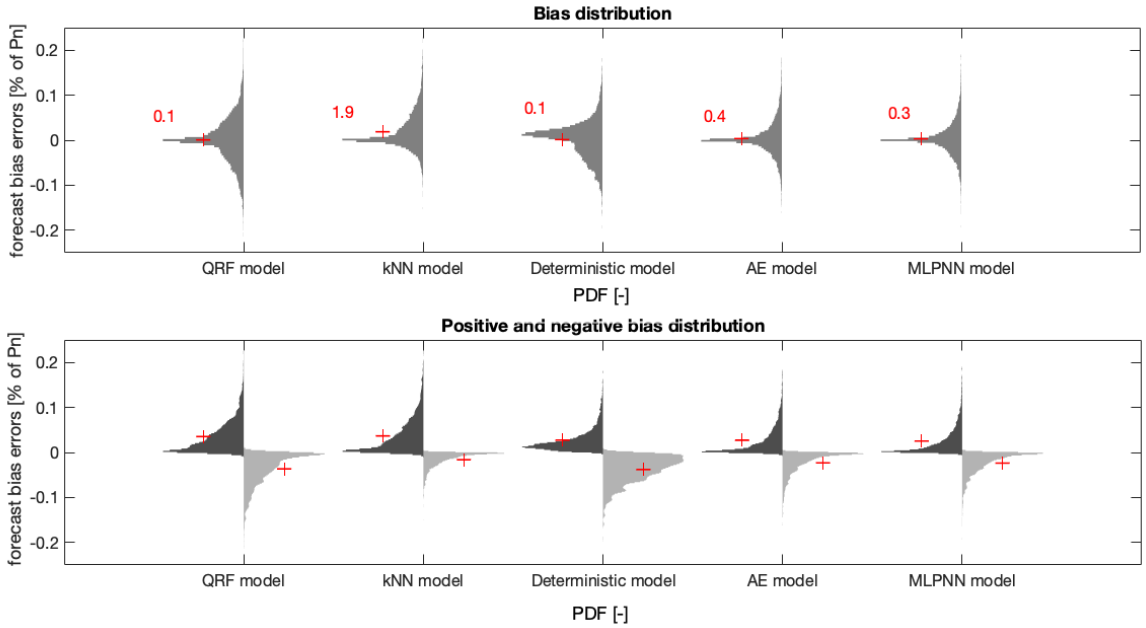


Figure 23: Probabilistic Density Distribution (PDF) of the forecast bias errors. In the top subplot the MBE value (red) is also shown in % of the installed capacity.

The CDF of KNN forecast is constantly underestimating with a KS of 0.048. Indeed, since the KNN model provides a systematic over-forecast, the probability to find a predicted PV generation lower than the specific value is always lower than the observed probability.

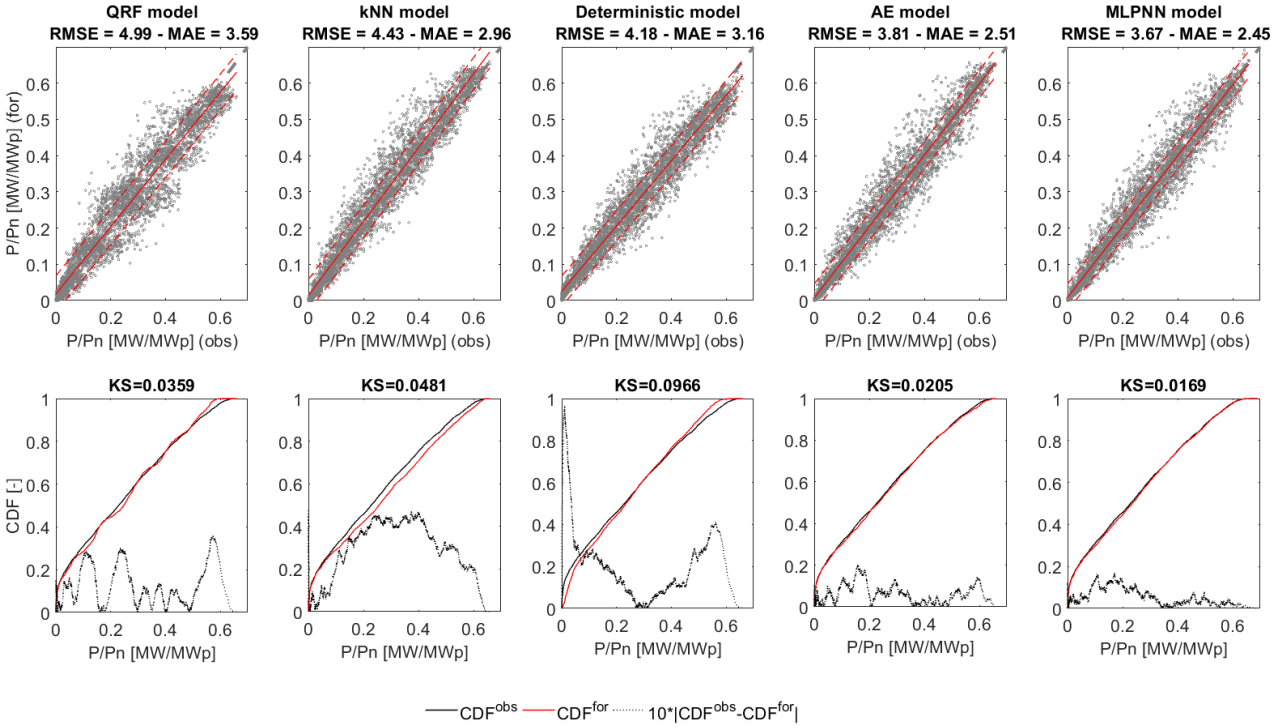


Figure 24: Scatter plots (top) and Cumulative Density Function (CDF, bottom) of the forecasts.

The CDF of deterministic forecast is lower than the observed CDF at low PV generation values while is higher at high PV generation level. The greatest absolute difference between the forecast and the observed CDF is at PV generation near to zero. This means that in the early morning and in the late afternoon, the model provides a constant over-forecast. For this reason, the high KS of almost 0.1 is less representative than the second maximum of the absolute CDF difference around 0.5 that indicates an under-forecast around noon. The accuracy of the AE and the MLPNN models are very similar both in terms of RMSE and MAE and in term of CDF. The two models achieve a RMSE of 3.8% and 3.7% and a MAE of 2.5% and 2.45%, respectively. The CDF of the forecasts are very similar to the CDF of the observations with KS indexes of 0.02 and 0.017, respectively.

The daily K_{PV} index is the equivalent of the daily clear-sky index for PV generation [51], defined as the ratio between the measured daily PV generation and the theoretical PV production during clear-sky conditions, derived from a clear-sky irradiance model by the SAPM [31]. As the clear sky index, the daily K_{PV} can be used for meteorological classification: overcast days with $0 < \text{daily } K_{PV} \leq 0.4$, variable with $0.4 < \text{daily } K_{PV} \leq 0.8$ and clear sky days with daily $K_{PV} > 0.8$.

Dividing the daily K_{PV} into bins of width 0.1, we define the daily RMSE and MBE as the RMSE and MBE of each day averaged over all the days belonging to the same daily KPV bin. Figure 25 shows the daily RMSE and MBE obtained during overcast, variable and clear sky days. All the forecasts show the same trend of MBE, from an over-forecast during overcast days to an under-forecast during clear sky days. Nevertheless, coherently with the previous analysis, the KNN prediction is predominantly positive biased. The shape of the daily RMSE of the QRF and the deterministic predictions are similar, even if the latter obtains smaller errors. In the same way, also the KNN, AE and MLPNN forecasts show similar daily RMSE trends, with greater errors during overcast days and smaller errors during clear sky days. As expected, the RMSE

obtained by AE and MLPNN models is almost the same with the exception of the overcast days with $K_{PV} \leq 0.4$, where MLPNN performs better. Nevertheless, these days are only 30 on 366.

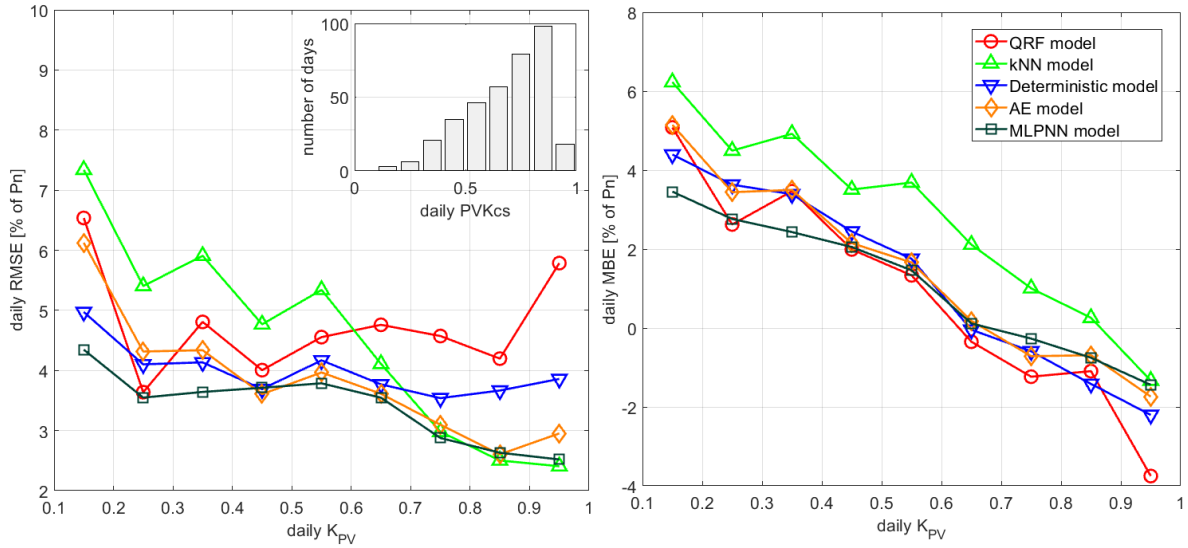


Figure 25: Daily RMSE (left) and MBE (right) of the forecast (i.e. the RMSE and MBE of each day averaged over all the days belonging to the same daily KPV bin).

Figure 26 summarizes the values of the Key Performance Indexes (KPIs). The improvement of the forecasts with respect to the persistence prediction ranges from 27.5% to 46.6%.

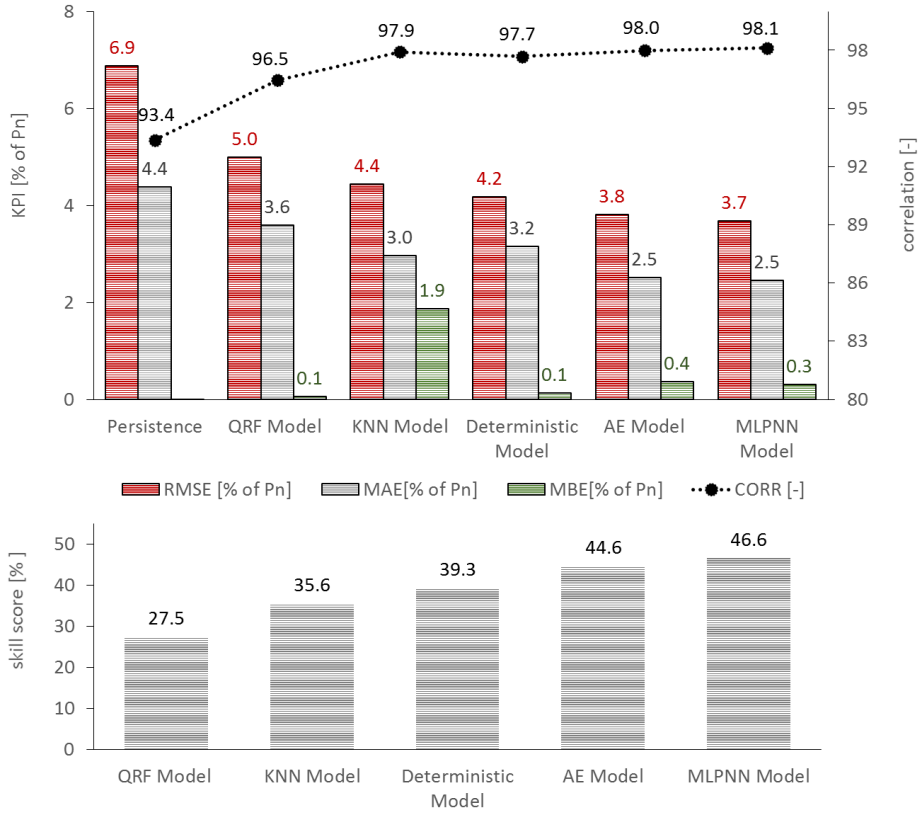


Figure 26: Main Key Performance Indexes (KPIs) of the forecasts.

The accuracy values of the outperforming models are coherent with the accuracy found in literature for other countries (Table 5).

Table 5: Comparison between the day-ahead forecast accuracy of the present models and some other approaches reported in literature.

Country	Regions size [km^2]	RMSE [% of Pn]	Skill score [%]
Italian outperforming forecast	300 x 10^3 (whole of Italy)	3.6	46.6
Germany [45]	214 x 10^3 - 104 x 10^3 (two regions)	4.1 - 4.3	48 – 52.8
Italian outperforming forecast	50 x 10^3 (average market zones)	5.5	31
NORD	119.5 x 10^3	4.1	35.8
CNORD	41 x 10^3	6.6	20.3
CSUD	41.5 x 10^3	5.8	33.7
SUD	49 x 10^3	5.6	34.7
SICI	26 x 10^3	5.1	27
SARD	24 x 10^3	5.5	31
Japan [47, 52]	32.4 x 10^3 – 72.6 x 10^3 (two regions)	6 - 7	50 - 60

5.4 Blending the Day-ahead Forecasts of the PV Generation All over Italy

We test the three methods: avg, SL and NN (Equations (11), (12) and (13) in section 4.7) to blend the day-ahead predictions obtained by the six models: MLPNN, AE, Deterministic, KNN, QRF and persistence, comparing the accuracy of the blending with the performance of the best prediction of the ensemble (MLPNN).

The performance of the avg and NN blending methods depends on number of the forecasts included in the ensemble. On the contrary, the SL blending automatically selects the models to include in the ensemble. Indeed, during the training phase, the brute force optimization provides weight (W_i) different from zero only for the forecast models that contribute to the outperforming blend.

The avg blend outperforms the best forecast of the ensemble only if the difference between the RMSE of the best and worst forecast is lower than 1% of Pn (Figure 27). Same results were found in [42]. In this case, the highest accuracy of the avg blend method is reached averaging just the first two most accurate forecasts (obtained by the MLPNN and AE models). This method (avg 2 models blend) increases the skill score with respect to the persistence from the 46.6% of the MLPNN to almost 50%. Unlike the avg blend, the performance of the non-linear NN blending increases with the number of the forecasts included in the ensemble, even if the difference between the RMSE of the best and the worst prediction is greater than 3%. In this case, the NN blend that inputs all the 6 forecast trajectories (MLPNN, AE, Deterministic, KNN, QRF and persistence) is the outperforming blending method, enhancing the skill score from 46.6% to 51.3%. The smart linear blend does not obtain better performance than the avg 2 models blend. Afterward, we will explain this paradox.

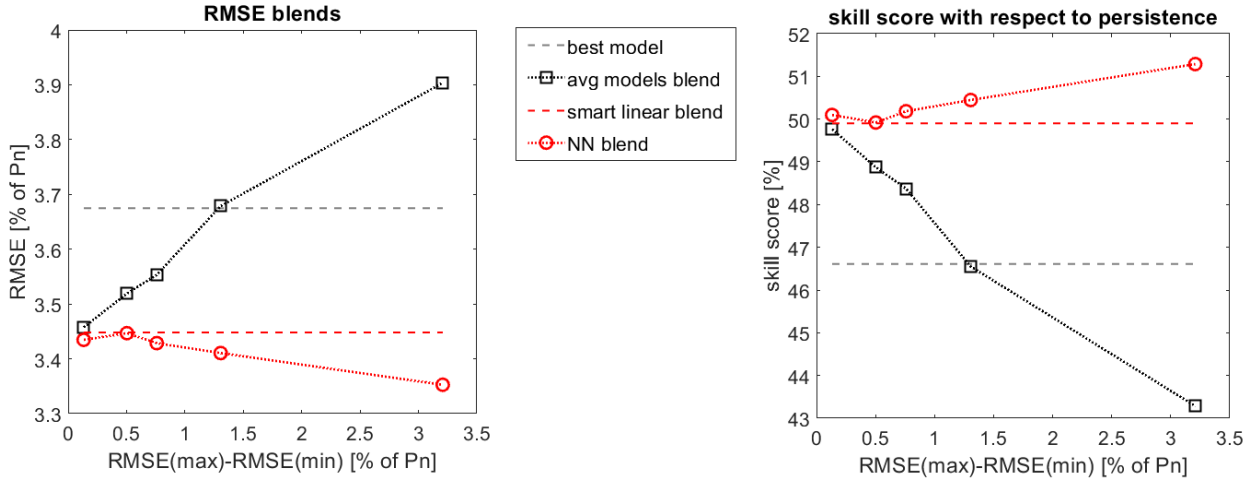


Figure 27: RMSE (left) and skill score (right) of the different blending methods versus the difference between the minimum and the maximum RMSE of the forecast trajectories included in the ensemble. Each points of the curves related to avg. models blend and NN blend correspond, from left to right, to an increasing number of forecast trajectories included in the blending in the order of increasing RMSE (exploiting RMSE values available in Figure 26): from an ensemble of two models (first point, MLPNN + AE) to an ensemble of six models (fifth point, MLPNN + AE + Deterministic + KNN + QRF + persistence).

The accuracy of the linear blending (such as the avg and SL methods) depends on the difference between the accuracy of each member of the ensemble and on the pair correlation between the forecast errors of all the predicted trajectories included in the ensemble (see Eq. (10). The lower the RMSE differences and the errors pair correlations, the higher is the accuracy of the blend. The extreme example is when we use two forecasts with no bias and same RMSE, but perfectly anti-correlated. In this case, the simple average will produce a perfect forecast.

The average is the simplest linear blending model and it should be the first model to test. Given N available forecasts, to identify which forecasts should be included in the ensemble to produce the outperforming average blending, all the possible ensemble including the best forecast should be tested and this could not be an easy task. Indeed, given N available forecasts there are $\left((N-1)+\sum_{k=3}^{N}\frac{(\mathbf{N-1})!}{(\boldsymbol{k-1})!(\boldsymbol{N-k}!)}\right)$ possible average, where k= 2, …, N is the # of predictions that should be included in the ensemble together with the outperforming one.

To reduce the number of attempts, we define the "average blending factor" (BF):

$$\boldsymbol{BF_i(N) = \left(\frac{RMSE_i^2}{RMSE_1^2} + 2 * \frac{RMSE_1 * RMSE_i}{RMSE_1^2} * \rho_{1i} + \sum_{j\neq i}^{2,N} \frac{RMSE_i * RMSE_j}{RMSE_1^2} * \rho_{ji}\right) \Big/ (N+1)} \qquad (23)$$

where N is number of available forecasts, $RMSE_1$ and $RMSE_i$ are the root mean square errors of the outperforming forecast and of i-th forecast while ρ_{ji} is the Pearson correlation between all the forecast errors.

This factor measures the average between the loss (with respect the best forecast) in blending a forecast with a lower accuracy (1 term: $RMSE_i^2/RMSE_1^2$) and the possible gains in blending the i-th forecast with low inter-correlated errors (N terms:

$$\boldsymbol{2 * \frac{RMSE_1 * RMSE_i}{RMSE_1^2} * \rho_{1i} + \sum_{j\neq i}^{2,N} \frac{RMSE_i * RMSE_j}{RMSE_1^2} * \rho_{ji}} \qquad (24)$$

The only forecasts that, for sure, will have a positive impact on the N average blending accuracy are the ones who have an Average Blending Factor lower than one and the lower are the $BF_i(N)$ the more accurate will be the simple average. Indeed, in case of the simple average, we can write Eq. (10) as in the following:

$$\boldsymbol{RMSE_{avg}(N) \cong \sqrt{\sum_{i=1,N}\sum_{j=1,N}\left(RMSE_i * RMSE_j * \rho_{ij}\right)/N^2} =} \qquad (25)$$

$$\boldsymbol{= RMSE_1\sqrt{\left[1 + (N^2-1)\left(\frac{\sum_{i=2}^{N} BF_i(N)}{N-1}\right)\right]/N^2} = RMSE_1\sqrt{[1 + (N^2-1)\langle BF_i(N)\rangle]/N^2}}$$

where N is the number of the available forecasts, $RMSE_1$is the RMSE of the outperforming forecast and BF_i is the Blending Factor of the forecasts (with i > 1). The RMSE of the average of N forecasts is proportional to root of the mean Blending Factor: $\langle BF_i(N)\rangle$.

It is worth noting that if all the blending factors of the N forecasts are equal to one, then $RMSE_{avg}(N) = RMSE_1$. In contrast, as long as the forecast errors are all positive correlated one to the other ($\rho_{ij} > 0$), the only forecasts that will have a positive impact on the N average accuracy are the $N1$ ones who have an Average Blending Factor lower than one (with $N1 \leq N$) and the lower is the $BF_i(N)$ the more accurate will be the average.

Theoretically, the pair-correlation between the forecast errors (ρ_{ij}) could be also negative but this is very unlikely since the physics laws underlying the numerical weather prediction models are the same.

Furthermore, it is possible to iterate the process computing again $BF_i(N1)$ (with i > 1) and verify if all the $N1$ selected forecasts still have a blending factor $BF_i(N1)$ lower than one. Otherwise it should be selected only the $N2$ with $BF_i(N1) < 1$ (with $N2 < N1$) , and so on.

Finally, once $BF_i(N2) < 1$ (for $i = 1, \dots, N2$) between the $N2$ forecasts there is certainly an average that outperforms the "best" forecast and all the other possible average blendings, at least the average between the "best" forecast and the forecast with the lowest blending factor ($BF_2(N2) = min\{BF_i(N2)\}_{N2}$).

Therefore, the outperforming average blending should be chosen among the ensemble of only the $N2 \leq N$ forecasts with blending factor lower than one and it should be adopted as reference blending model.

In our case, only the AE model provides a $BF_2(6)$ less than one (Figure 28a), and for this reason the simple average of the MLPNN and AE predictions provides the outperforming average, while the accuracy of average models blending decreases when the other forecasts are included in the ensemble (Figure 27).

Moreover, Figure 28(c) also shows the forecast accuracy of the average models blend of a part of the possible forecast ensembles including the best prediction (MLPNN): ensemble from 2 to maximum 6 members. For instance, the elements of the first row are the RMSE obtained by the blend of the ensembles that include two, three, four, five and six forecasts. It can be observed that the exclusion of the AE forecasts (with $BF_2(6) < 1$) from the ensemble will never lead to a better blending performance even if the error correlation of the other members is lower (Figure 28(b)).

Using the blending factor approach we are able to immediately find the reference prediction without testing all the 31 possible average combinations.

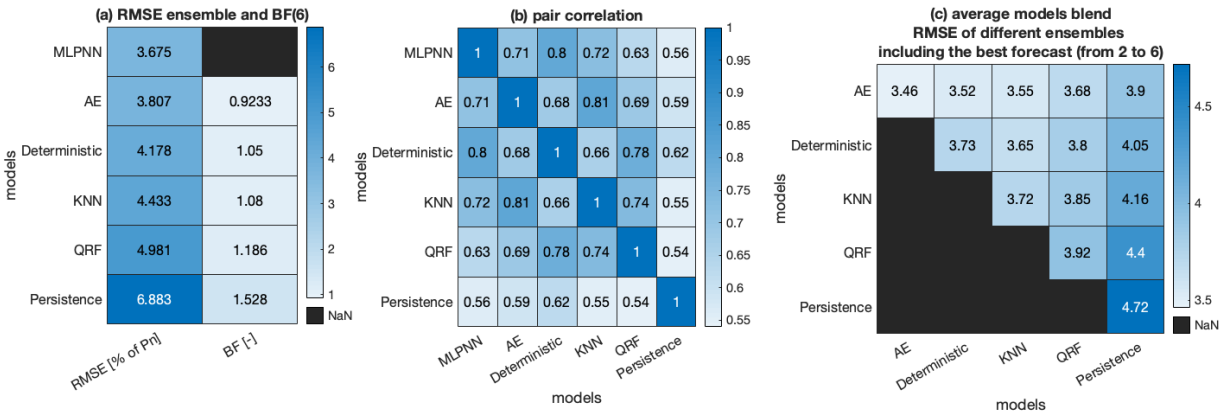


Figure 28: RMSE and Blending Factor of each forecast of the ensemble, (a); matrix of the Pearson cross correlation between the forecasts errors of each ensemble member, (b); RMSE of the average blend of all the possible forecast ensembles that include the best model (MLPNN), (c).

Nevertheless, by setting suitable weights and/or suitable forecast partitions the predictions with blending factor greater than one can also effectively contribute to improve the blend accuracy.

Figure 29 shows the weight coefficients (W_i) for each $K_{PV}^{\text{best for}}$ intervals (forecast partitions) of the smart linear method obtained by the brute force optimization procedure.

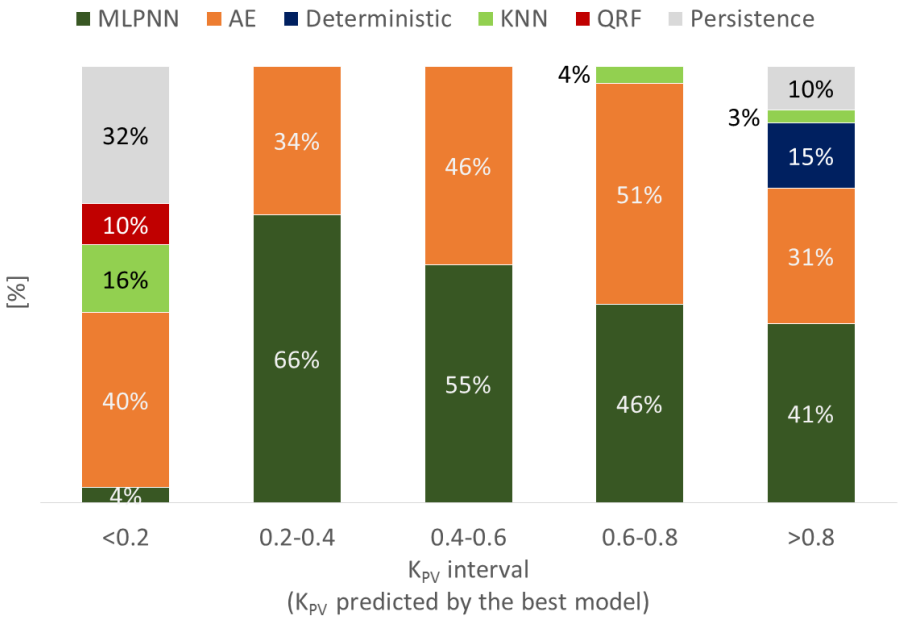


Figure 29: Weight coefficients (W_i) for each $K_{PV}^{best\ for}$ intervals of the smart linear method obtained by the brute force optimization procedure.

The SL method uses a complex combination of the forecast models only during very overcast ($K_{PV}^{best\ for} < 0.2$) or very clear sky ($K_{PV}^{best\ for} > 0.8$) predicted conditions. It is interesting to note that under these weather conditions the persistence model provides a remarkable contribution. Indeed, during consecutive very overcast or very clear sky days, persistence model can provide an accurate forecast. Nevertheless, the majority of the PV generation occurs at overcast ($0.2 < K_{PV}^{best\ for} < 0.4$) or variable/fine ($0.4 < K_{PV}^{best\ for} < 0.8$) weather conditions. During these conditions the brute force optimization mainly selects the first two models of the ensemble (MLPNN and AE) with an average weight close to 50% and 50% (exactly the same of the avg 2 models blend). Furthermore, the brute force optimization of SL model that provides the best linear combination of all forecast trajectories of the training set is affected by under or over fitting problems. Thus, the SL blending method, when applied on the test set, could obtain also worse performance than the simple average (that does not need any training). Indeed, to reduce these problems, different $K_{PV}^{best\ for}$ interval partitions have been tested: few intervals produce under-fitting while too many intervals generate overfitting. For these reasons, the SL blend obtains almost the same accuracy of the average of the first two outperforming forecasts.

For the non-linear blending methods, the accuracy of the forecast included in the ensemble is less important. Indeed the performance of the NN blend increases with the number of the ensemble members regardless of the RMSE of each prediction. Nevertheless, a low pair correlation between the forecast errors of each ensemble member becomes much more important. Indeed, for any data driven method (ANN, SVR, QRF, etc.) the use of high correlated inputs can completely threaten the performance since the input features do not bring relevant information, while the model complexity increases dramatically with the inputs number. Thus the performance of a non-linear blend based on any machine learning technique that inputs highly correlated forecasts can be easily lower than the accuracy of the best member of the ensemble. Low correlated PV generation forecast errors can be generated only by the use of completely different prediction models, i.e. different NWP models, as well as different up-scaling methods and/or different machine learning forecast techniques. Perez et

al. [39, 40, 41] developed a smart linear method to blend the GHI forecasts of 4 different NWP and the resulting forecast was identified by EPRI as the most accurate among thirteen operational models, after an independent evaluation in two climatically distinct US regions. The blend of the day-ahead forecast reduces the MAE of best forecast member by 15% (relative reduction). Similar results were found by Huertas-Tato et al. [53] that blended four different GHI predictions in four cities (in Spain and Portugal). They achieved by non-linear blending (SVM) an improvement of best prediction MAE ranging from 12% to 14% depending on the city and a RMSE improvement of 11%-16% (MAE/RMSE relative reduction). Pierro et al. [42], combining six PV power forecasts of a single plant generation obtained by 6 different data-driven models and two different NWP, obtained an RMSE improvement of 6% just with the simple average blending method. They also showed that averaging only the forecast obtained through the same NWP the improvement will be reduced to 3%. Gigoni et al. [43] combined five different day-ahead forecasts of 32 PV plants generation in Italy, obtained by five data-driven models. In this case, they used a smart-linear blending model which, in average on the 32 PV power generations, obtained a MAE and RMSE improvement of 1.9% and 0.6%. This small improvement probably depends on the use of the same NWP as models input, so that the five forecasts were all strongly correlated one to the other.

Also in this work, all the forecast results from the use of the same NWP but very different up-scaling approaches have been adopted. The pair correlation of the forecast errors ranged from 0.81 to 0.54 (Figure 28(b)), so that the NN blend could reduce the MAE and the RMSE of the best forecast by 8.6% and 7.1%, values comparable with the ones of [41, 53, 42].

Figure 30 shows scatter plots and the Cumulative Density Function (CDF) of the best forecast model and of the blending forecast methods. The avg. 2 model and the smart linear blend methods improve the RMSE of the best forecast member of the ensemble by almost 6% (as in [8]). The NN 6 models blend is almost unbiased and obtains 7% of improvement, as mentioned above. Furthermore, this method provides a CDF of the PV generation more similar to the one observed with a KS index of 0.012 with respect to 0.017 of the best forecast.

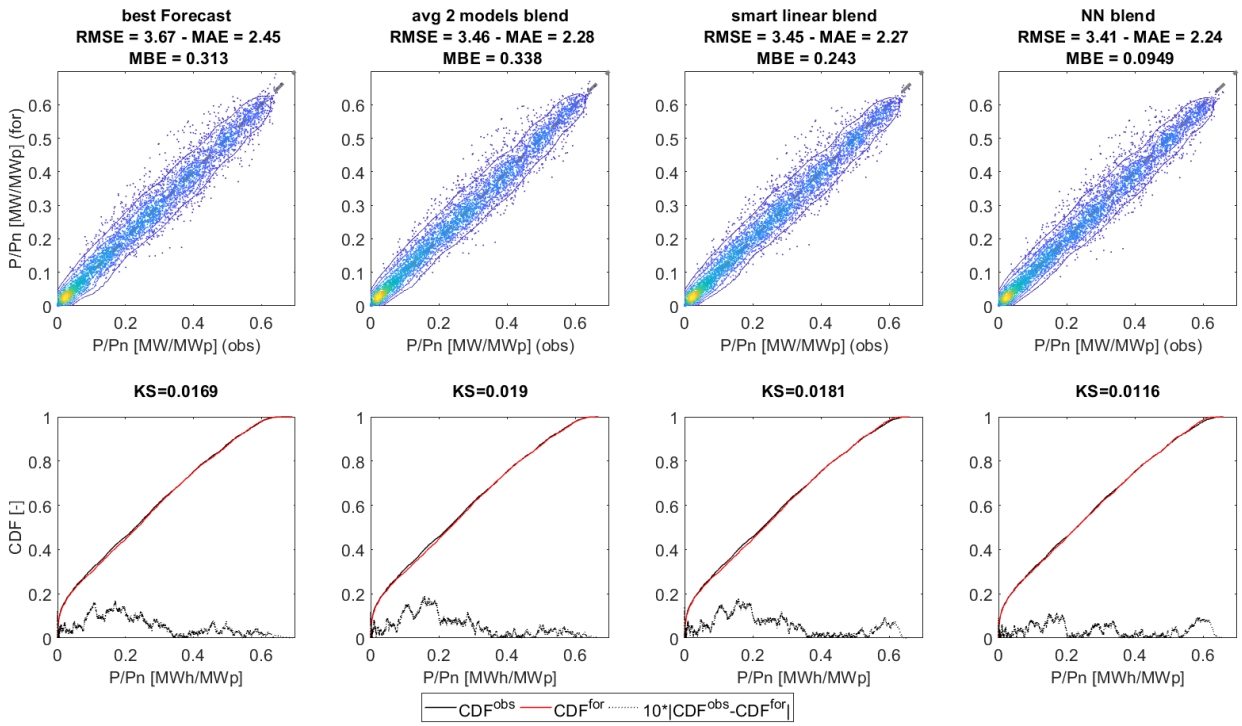


Figure 30: Scatter plots (top) and Cumulative Density Function (CDF, bottom) of the best forecast model and of the blending forecast methods.

The NN blend method reduces the daily RMSE and MBE during overcast days and provides a RMSE lower than the best forecast model during variable and clear sky days. On the contrary, the best forecast is outperforming only during perfect clear sky days (Figure 31).

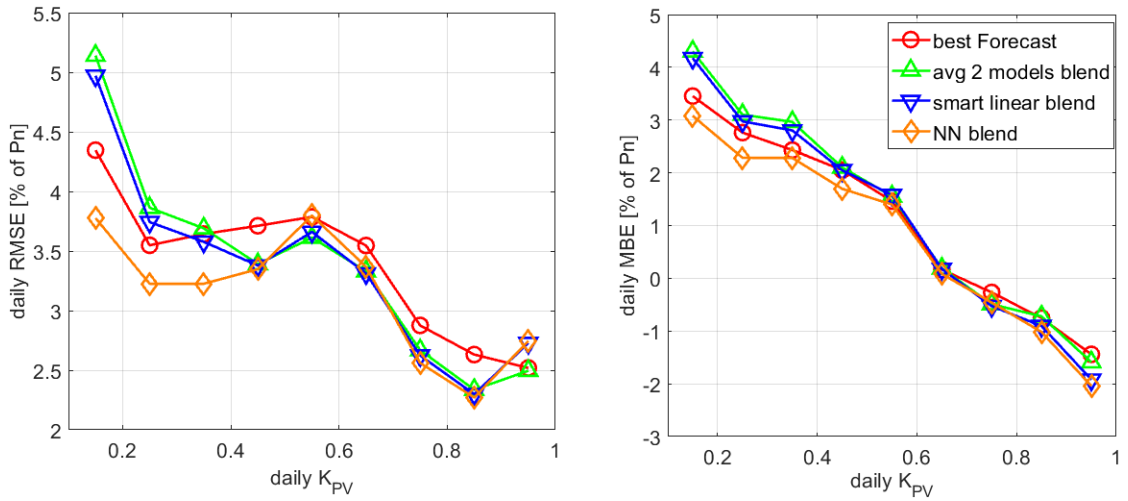


Figure 31: Daily RMSE (left) and MBE (right) of the best forecast model and of the blending forecast methods.

It is worth noticing, from Figure 31, that the performance of the 2avg and SL models are basically the same other than during the very clear sky days in which SL obtains a lower bias but an even higher RMSE. Nevertheless, 2avg model is the simplest possible blending while

SL is a remarkably complex model obtained by a brute force optimization of 30 weight coefficients.

This theoretical and experimental study leads to an important general result: to assess the quality of a blending method is not enough to compare the obtained accuracy with the one achieved by the best forecast of the ensemble, but it is also essential to benchmark the performance with the simplest blending model i.e. the average. In particular, to select the outperforming reference model, it should be used the outperforming average chosen among all the forecasts that has a blending factor lower than one . Any blending method should outperform this reference model, otherwise the use of a more complex approach cannot be justified.

In our case, the outperforming reference model is the 2avg that obtains a SS with respect to the RMSE of the best forecast of 5.7% (Table 6). The smart linear blending SL achieves a SS with respect to the best forecast of 6%, improving the performance of the reference model of 0.3% only (relative value). The NN non-linear blending gets a skill scores of 7.1% and 1.4% with respect to the best forecast and the reference blending, respectively. Therefore only the NN method could be an acceptable alternative to the simple average blending.

Table 6: Summary of model performances.

Methods	MBE [% of Pn]	MAE [% of Pn]	RMSE [% of Pn]	SS (best for) [%]	SS (2avg) [%]
best forecast	0.313	2.45	3.67	0.0	-6.1
2avg (reference)	0.338	2.28	3.46	5.7	0.0
SL	0.243	2.27	3.45	6.0	0.3
NN	0.095	2.24	3.41	7.1	1.4

It is worth noting that none of the aforementioned works [39, 40, 41, 53, 43] have used a simple reference blending model.

6 RESULTS OF BENCHMARK FOR UTRECHT (THE NETHERLANDS)

6.1 Performance Metrics

The error metrics used to evaluate the results are given by equations (16), (17), (18) and (19) as defined in section 5.1. The only difference here is that the P_n is constant during the studied period, and equal to the combined installed capacity of the 150 PV systems.

6.2 Utrecht PV Production

The PV-systems have a combined installed capacity of 301 kWp [26]. The average annual energy yield of these systems in the period 2014-2016 is 1'039 kWh/kWp (see Table 7). In this same period, the yield of individual systems is found to differ between 799 and 1’271 kWh/kWp per year.

Table 7: Average annual power production of 150 PV-systems / unit of installed capacity.

Year	Yield [kWh/kWp]
2014	1'042
2015	1'073
2016	1'003
Average	1'039

Figure 32a shows the monthly energy yield for all systems combined during the period studied. The figure indicates variations both in the PV power yield per month and interannual variations. On an annual basis most power is produced in the months April to August. The difference in the PV yield from one month or year to another can be explained by fluctuations in the global horizontal irradiance, which is highly dependent on the cloud cover. For example, the limited production in June 2016 compared to the PV yield during the same month in 2014 and 2015 is due to a lower observed irradiance in that month [54].

In addition, Figure 32b shows the distribution of the hourly PV generation in Utrecht for daylight hours only. The maximum PV power generation observed in an hour is 0.93 kW/kWp, whereas in 90% of time production is found to be below 0.70 kW/kWp.

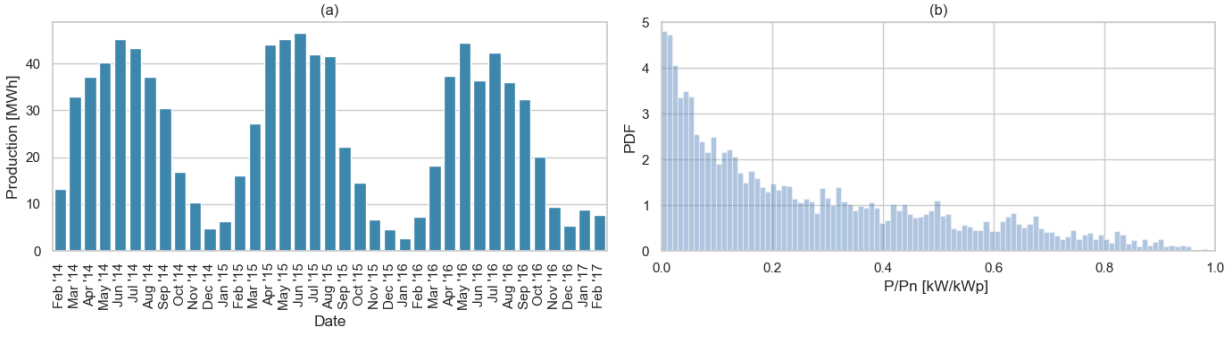


Figure 32: monthly PV power yield for all 150 PV systems (a); Distribution of the PV power generation (b).

6.3 Benchmark Results of Day-ahead Forecasts of PV Generation for 150 PV-systems in Utrecht

The regional forecasts generated consider day-ahead, hourly PV power forecasts for the combined production of 150 PV-systems in Utrecht. Figure 33 shows a scatterplot holding the measurements and predictions for all forecasting models. The evaluation of the forecasting results in terms of the RMSE, MAE, MBE and Skill Score can be found in Figure 34 and Figure 35. It is important to note that the results only consider diurnal hours for the period February 2016 until February 2017.

From Figure 34 it can be found that the results obtained for the K-SVM, RF, GB and FNN model are very similar. Moreover, based on the MAE and RMSE these models prove to be very competitive. From these models, RF is found to have the best performance when considering the MAE, with an error of 6.5%. Furthermore, when assessing the RMSE, K-SVM model is found to outperform all other models with an error of 9.9%. In particular, a lower MAE score and a higher RMSE value indicates that the RF model generates more outliers compared to K-SVM. From these 4 models, the poorest results are observed for the FNN model. Finally, Figure 34 shows that all four models significantly outperform the classical persistence model.

Figure 35 presents the model performance results in terms of the Skill Score. Since the Skill Score is directly dependent on the RMSE (see equation (19)), the K-SVM model also achieves the best results in terms of the Skill Score. Moreover, compared to the RMSE and MAE, the observed difference between the obtained Skill Scores per model is more significant. RF and GB achieve more or less similar results and, similar to RMSE, the poorest performance is found for the FNN model.

From Figure 34 it can also be observed that all models, except RF, have a positive bias. This indicates that these models overestimate the PV power production in the validation year (February 2016 until January 2017). This deviation could be explained by inter-annual variations in the PV power production. Moreover, the total power production is found to be approximately 6% lower in the validation year compared to the average of the model training years [26]. The stronger bias of the K-SVM and FNN models indicate that these models are more affected by these inter annual deviations.

As the order of best performing models differ per error metric considered, the preferred model depends on its application and therefore the interest of the user.

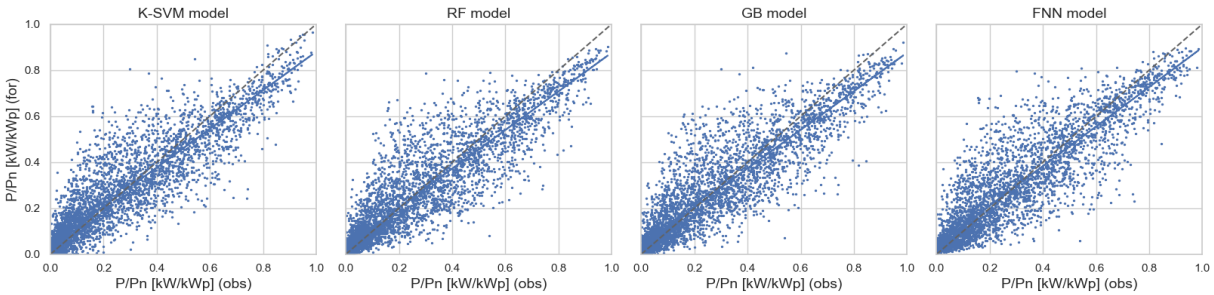


Figure 33: scatterplot of the forecasts.

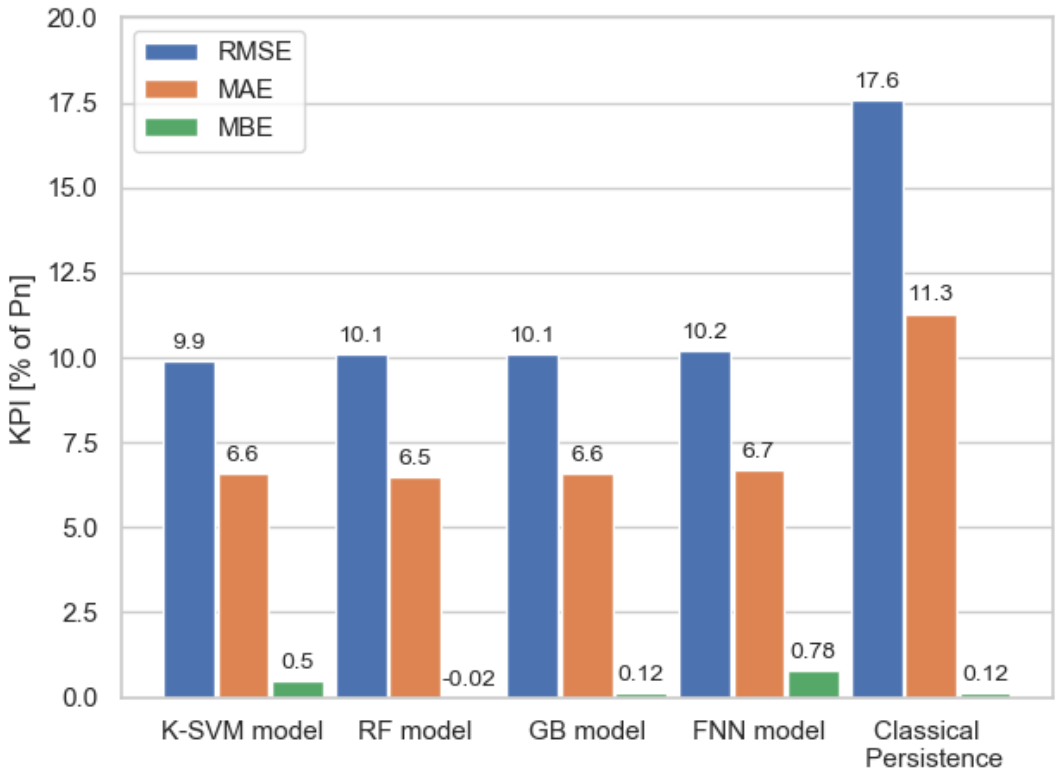


Figure 34: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Bias Error (MBE) of the K-SVM, RF, GB, FNN and Persistence forecasting models. All values are given in percentages.

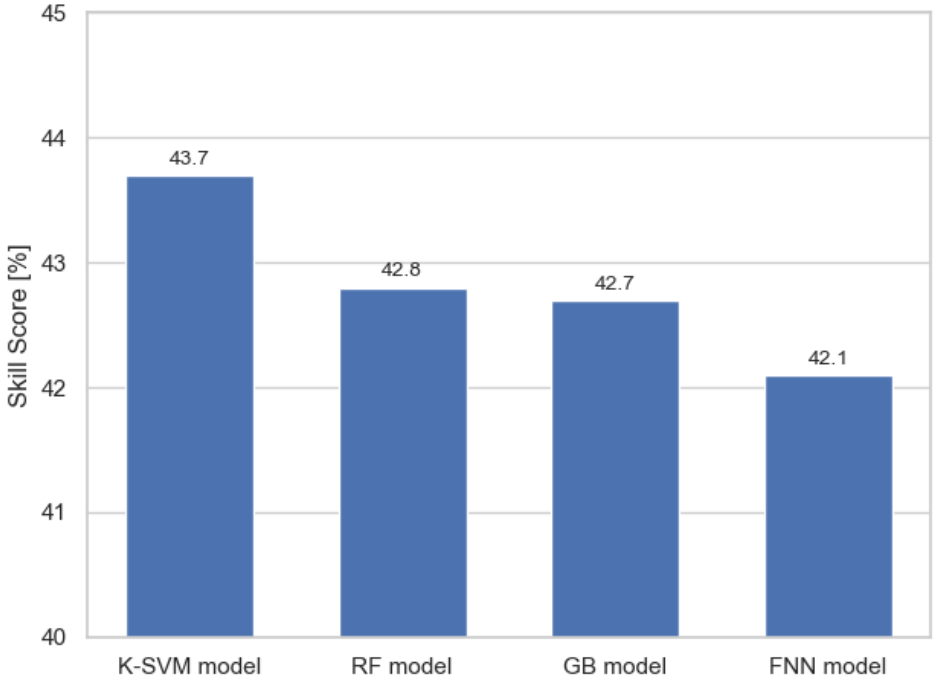


Figure 35: the Skill Score of the K-SVM, RF, GB and FNN forecasting models relative to Classical Persistence.

7 CONCLUSION

In this work, we compare the accuracy of several up-scaling methods for regional PV power forecast from 1 up to 3 days-ahead using three different case studies characterized by different size of the area of interest and located in two different countries:

- Forecast of the aggregated PV power output of 150 PV systems in the province of Utrecht (Netherlands) within a controlled area of 1.4 $x10^3$ km^2;
- Forecast of the regional PV generation of six Italian macro zones[6] (market zones) with controlled areas that range from $24x10^3$ to $119x10^3$ km^2;
- Forecast of the regional PV generation of the whole of Italy with a controlled area of 300 $x10^3$ km^2.

All the tested up-scaling methods directly predict the regional PV power generation, i.e. consider the PV power output of the whole PV fleet as if it had been produced by a single "virtual" solar power plant.

The examined models gain different performances depending on the following: input and power measurements pre-processing, numerical weather prediction (NWP) used as forecast model inputs, aggregation methods adopted for input dimensionality reduction, PV forecast models, post-processing correction, forecast horizon and size of the area under examination.

Table 8 summarizes the main features of all the benchmark methods, for each case study and reports the obtained ranges of the day-ahead forecast accuracy.

For all the case studies, the tested approaches obtain a wide range of accuracy. The greater the differences between the up-scaling steps of each method, the larger is the range of accuracy that can be obtained. In the regional power forecast for the small area of the Utrecht province, the examined methods differ only in the data-driven forecast models using all the IFS forecast by ECMWF as a basis and the same post-processing and pre-processing steps . The skill score ranges from 42.1% to 43.7%: this marks a relative performance increase of almost 4% for the best approach with respect to the worst.

In the PV power forecast of the Italian market zones, the skill score ranges from 16% to 30.5%, with almost 49% of relative improvement. In this case, the two tested methods differ in the PV forecasting models and in the post-processing technique.

With regard to the forecast of the PV generation for the whole of Italy, we obtained skill scores from 27.5% to 46.7%, hence almost 43% of difference. Also in this case, there are many differences between the forecasting procedures, even if the same NWP data (WRF with IFS – ECMWF boundary conditions) were used as methods inputs: input aggregation methods, forecasting models and output post-processing.

These results indicate that, in regional PV power forecast, there is a wide margin of performance improvement achievable. All the forecasting steps should be carefully tuned to get the highest accuracy.

[6] Nowadays, the Italian TSO has divided Italy in these six macro regions since between these zones there are bottlenecks of the National grid. In each of these regions, the Energy Markets take place. For this reason, the aggregated PV power generation of each market zone is currently predicted by the Italian TSO.

We also show that the accuracy does not remarkably change passing from one day to three days-ahead forecast horizon. Indeed, considering the outperforming up-scaling method, the average RMSE of the day-ahead forecast of the market zones PV generation is 5.5% of Pn, while the accuracy of the three days-ahead forecast is 5.6% of Pn. This is because the accuracy of the NWP used as models input does changes slowly in this horizon range. This result is confirmed by literature (for instance in [55]).

It is well known, that the performance of the regional PV power forecast improves with the size of the forecast footprint because of the “smoothing effect”. We study this effect on Italy proving that the forecast error (RMSE) decreases with the root of controlled area following an exponential or hyperbolic trend (similar results could be found in [35, 37]). As a result, the RMSE of the forecast of the Italian PV generation can be reduced by 34% passing from the prediction of the market zones solar generation to the prediction of the national solar production (test period 2017). Indeed, we found that mean RMSE of the market zones PV power forecast is 5.5% of Pn, while the RMSE of the Italian PV power generation is 3.6%.

Finally, we investigate the accuracy improvement that can be reached by blending the different predictions of the Italian PV generation. Thus, instead of simply selecting the outperforming forecast, we studied how the different predictions can be combined to increase the accuracy of the Italian PV power forecast. We studied both linear and non-linear combinations. We proved that the simplest linear blending method, i.e. the forecasts average, has an accuracy strongly dependent on the RMSE of the forecasts included in the blending ensemble. Conversely, Smart linear blending methods can suitably combine any forecast regardless of their RMSE. Also, Non-linear model performance is not affected by the RMSE of the forecasts to be blended but depends on the pair-correlation between the forecast errors. Non-linear blending models that combine forecasts with highly correlated errors can easily obtain an accuracy lower than the one of the outperforming forecast of the blending ensemble.

We further showed that, in order to assess the quality of a blending method, it is not enough to compare its accuracy with the one of the best forecast of the ensemble, but also the outperforming simple average must be used as reference. To identify this reference model, we defined a new index, called “average blending factor” (BF) proving that the outperforming average blending should be chosen among all the possible averages of all the forecasts that have a blending factor lower than one. Any blending method should outperform both the best forecast and this reference blending.

In our case, the reference model was the average of only two forecasts over six available (with BF equal to 1 and 0.92). It improves the RMSE of the best forecast of 5.7%. The smart linear blending model (SL) improves the accuracy of the best forecast of just 6% while the non-linear model (NN) of 7.1%. Therefore, the model SL, in spite of its much higher complexity, does not achieve a significant performance improvement compared to the simple reference model. In contrast, NN is the outperforming method.

We conclude that the accuracy of regional forecasts can be further improved by a blending model that appropriately combines the output of different up-scaling approaches, provided that these approaches are based on quite different procedures, so that their forecasting errors have a low correlation with each other.

Table 8: Summary of the process of the benchmark up-scaling methods and accuracy results (day-ahead forecast horizon).

FORE-CASTING STEPS / ACCURACY CASE STUDIES	(1) DATA PRE-PROCESSING	(2) NWP	(3) INPUTS AGGREGATION	(4) FORECAST MODELS	(5) POST-PROCESSING	SIZE OF THE AREA [10^3 km^2]	RANGE OF RMSE [% of Pn]	RANGE OF SKILL SCORE [%]
PROVINCE OF UTRECHT (test period from February 2016 to February 2017)	SPATIAL AND TEMPORAL INTERPOLATION OF NWP TIME AGGREGATION AND EXCLUSION OF OUTLIERS FROM POWER MEASUREMENTS	IFS OF ECMWF	-	SUPPORT VECTOR MACHINE	REMOVE OUTLIERS FOR POWER PREDICTIONS	1.4	From 9.9 to 10.2	From 42.1 to 43.7
				RANDOM FOREST				
				GRADIENT BOOSTING				
				FEED FORWARD NEURAL NETWORK (MLP)				
MARKET ZONES OF ITALY (test period 2017)	EXCLUSION OF OUTLIERS FROM POWER MEASUREMENTS	MESO-SCALE MODEL (RAMS) + MOS	AGGREGATION BY PROVINCES	ANALOG ENSEMBLE	-	50 (zones average)	From 5.5 to 6.6 (zones average)	From 30.5 to 16 (zones average)
				KNN + QUANTILE REGRESSION FORESTS	POWER OUTPUT RESCALING			
ITALY (test period 2016)	-	MESO-SCALE MODEL (WRF) + MOS	PRINCIPAL COMPONENT ANALYSIS	QUANTILE REGRESSION FORESTS	-	300	From 5.0 to 3.7	From 46.6 to 27.5
	-		AGGREGATION BY PROVINCES	KNN + QUANTILE REGRESSION FORESTS	POWER OUTPUT RESCALING			
	RETRIEVE EQUIVALENT POA		AGGREGATION BY ITALY	DETER-MINISTIC (SAPM)	PER-FORMANCE RESCALING			
	-		AGGREGATION BY PROVINCES	ANALOG ENSEMBLE	-			
	RETRIEVE EQUIVALENT POA		AGGREGATION BY MARKET ZONES	FEED FORWARD NEURAL NETWORK ENSEMBLE (MLP)	PER-FORMANCE RESCALING			

LIST OF FIGURES

LIST OF TABLES

REFERENCES

[1] I. R. E. A. (IRENA), "Global Energy Transformation - A roadmap to 2050," 2019 Edition.

[2] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada and C. Cornaro, "Data-driven up-scaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data," *Solar Energy,* pp. 1026-1038, 2017.

[3] U. Das, K. Tey and M. Seydmahmoudia, "Forecasting of photovoltaic power generation and model optimization: a review," *Renewable and Sustainable Energy Reviews,* vol. 81, pp. 912-928, 2018.

[4] R. Ulbricht, U. Fischer, W. Lehner and H. Donker, "First steps towards a systematical optimized strategy for solar energy supply forecasting," in *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, Prague, 2013.

[5] R. Perez and M. Perez, "A Fundamental Look At Supply Side Energy Reserves For The Planet," *International Energy Agency, Solar Heating and Cooling,* 2015.

[6] R. Perez, T. E. Hoff, J. Dise, D. Chalmers and S. Kivalov, "Mitigating short-term PV output variability," in *Proceedings of the 28th European Photovoltaic Solar Energy Conference and Exhibition*, Paris, France, 2013.

[7] P. Denholm, R. Margolis and J. Milford, "Production Cost Modeling for High Levels of Photovoltaics Penetration," NREL, 2008.

[8] R. Perez, M. David, T. Hoff, M. Jamaly, S. Kivalov, J. Kleissl, P. Lauret and M. Perez, "Spatial and Temporal Variability of Solar Energy," *Foundational Treds in Renewable Energy,* pp. 1-44, 2016.

[9] J. A. Duffie and W. A. Beckman, Solar Engineering of Thermal Processes, Wiley, 2013.

[10] J. Widén and J. Munkhammar, Solar Radiation Theory, Uppsala: Uppsala University, 2019.

[11] CAISO, 04 11 2018. [Online]. Available: www.caiso.com.

[12] J. Kleissl, Solar Energy Forecasting and Resource Assessment, Boston: Academic Press, 2013.

[13] D. W. van der Meer, J. Widén and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renewable and Sustainable Energy Reviews,* pp. 1484-1512, 2018.

[14] M. Diagne, M. David, P. Lauret, J. Boland and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews,* pp. 65-76, 2013.

[15] M. Pierro, M. De Felice, E. Maggioni, A. Perotto, F. Spada, D. Moser and C. Cornaro, "Deterministic and Stochastic Approaches for Day-Ahead Solar Power Forecasting," *Journal of Solar Energy Engineering* , p. 021010, 2016.

[16] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann and C. Kurz, "Regional PV power prediction for improved grid integration," *Progress in Photovoltaics,* pp. 757-771, 2010.

[17] S. Killinger, B. Müller, B. Wille-Haussmann and 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.

[18] S. Killinger, P. Guthke, A. Semmig, B. Müller, B. Wille-Haussmann and F. Fichtner, "Up-scaling PV Power Considering Module Orientations," *IEEE Journal of Photovoltaics,* pp. 941-944, 2017.

[19] M. Zamo, O. Mestre, P. Arbogast and O. Pannecoucke, "A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production part 1: Deterministic forecast of hourly production," *Solar Energy,* pp. 792-803, 2014.

[20] "www.terna.it/it-it/sistemaelettrico/transparencyreport/generation/expostdataontheactualgeneration.aspx," [Online].

[21] "Sunrise," [Online]. Available: http://sunrise.rse-web.it/.

[22] ECMWF, "ECMWF (European Centre for Medium-range Weather Forecasts)," 2019. [Online]. Available: https://www.ecmwf.int/en/forecasts/datasets/archive-datasets.

[23] B. Elsinga and W. G. J. H. M. v. Sark, "Short-term peer-to-peer solar forecasting in a network of photovoltaic systems," *Appl. Energy,* vol. 206, no. June, p. 1464–1483, 2017.

[24] H. Akima, "A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures," *J. ACM,* vol. 17, no. 4, p. 589–602, 1970.

[25] P. I. a. R. Perez, "A new airmass independent formulation for the linke turbidity coefficient," *Sol. Energy,* vol. 73, no. 3, pp. 151-157, 2002.

[26] L. Visser, T. Alskaif and W. Van Sark, "Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions," *IEEE 46th Photovoltaic Specialists Conference (PVSC),* pp. 1-6, 2019.

[27] E. Maxwell, "A Quasi-Physical Model for Converting Hourly Global Horizontal to Direct Normal Insolation," *1987.*

[28] B. Liu and R. C. Jordan, "Daily insolation on surfaces tilted towards equator," *ASHRAE,* vol. 10, p. 526–541, 1961.

[29] D. King, J. Kratochvil and W. Boyson, "Photovoltaic Array Performance Model," Sandia National Laboratories, Albuquerque, NM, USA, 2004.

[30] M. Pierro, G. Belluardo, P. Ingenhoven, C. Cornaro and D. Moser, "Inferring the Performance Ratio of PV systems distributed in an region: a real-case study in South Tyrol," in *44th IEEE Photovoltaic Specialists Conference (PVSC)*, Washington, DC, USA, 2017.

[31] N. Engerer, "KPV: A Clear-Sky Index for Photovoltaics," *Solar Energy,* vol. 105, pp. 679-693, 2014.

[32] C. Cornaro, M. Pierro and F. Bucci, "Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast," *Solar Energy,* vol. 111, pp. 297-312, 2015.

[33] S. Alessandrini, L. Delle Monache, S. Sperati and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied Energy,* vol. 157, pp. 95-110, 2015.

[34] M. Moschella, M. Tucci, E. Crisostomi and A. Betti, "A Machine Learning Model for Long-Term Power Generation Forecasting at Bidding Zone Level," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Bucharest, Romania, 2019.

[35] T. Hastie, R. Tibsharani and J. Friedman, The Elements of Statistical Learning, Springer, 2009.

[36] N. Meinshausen, "Quantile Regression Forests," *Journal of Machine Learning Research,* vol. 7, pp. 983-999,

2006.

[37] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001.

[38] N. Meinshausen, "quantregForest: Quantile Regression Forests," [Online]. Available: https://CRAN.R-project.org/package=quantregForest.

[39] R. Perez, J. Schlemmer, K. Hemker, S. . Kivalov, A. Kankiewicz and Gueymard, "Satellite-to-Irradiance Modeling – A New Version of the SUNY Model," in *42nd IEEE PV Specialists Conference*, 2015.

[40] R. Perez, J. Schlemmer, K. Hemker, S. . Kivalov, A. Kankiewicz and J. Dise, "Solar Energy Forecast Validation for Extended Areas & Economic Impact of Forecast Accuracy," in *43th IEEE PV Specialists Conference*, 2016.

[41] R. Perez, J. Schlemmer, S. Kivalov, J. Dise, P. Keelin, M. Grammatico, T. Hoff and A. Tuhoy, "A New Version of the SUNY Solar Forecast Model: A Scalable Approach to Site-Specific Model Training.," in *IEEE PV Specialists Conference WCPEC-7*, Waikoloa, HI., 2018.

[42] M. Pierro, M. De Felice, E. Maggioni, A. Perotto, F. Spada, D. Moser and C. Cornaro, "Multi-Model Ensemble for day ahead prediction of photovoltaic power generation," *Solar Energy,* vol. 134 , p. 132–146, 2016a.

[43] L. Gigoni, A. Betti, E. Crisostomi, A. Franco, M. B. Tucci and D. Mucci, "Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants," *EEE Transactions of Sustainable Energy,* vol. 9, no. 2, pp. 831 - 842, 2018.

[44] T. Hoff and R. Perez, "Modeling PV fleet output variability," *Solar Energy,* vol. 86 , pp. 2177-2189, 2012 .

[45] E. Lorenz , T. Scheidsteger, J. Hurk, D. Heinemann and C. Kurz, "Regional PV power prediction for improved grid integration," *Progress in Photovoltaics: Research and Applications,* vol. 19 , p. 757–771, 2011.

[46] E. Lorenz, . J. Hurka, D. Heinemann and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," *IEEE ,* vol. 2 (1), pp. 2-10, 2009.

[47] J. G. Fonseca, T. Oozeki, H. Ohtake, . K. Ichi Shimose, T. Takashima and K. Ogimoto, "Regional forecasts and smoothing effect of photovoltaic power generation in Japan: An approach with principal component analysis," *Renewable Energy,* vol. 68, no. 403-413, 2014.

[48] M. David, T. Hoff, M. Jamaly, S. Kivalov, J. Kleissl, P. Lauret, M. Perez and R. Perez, "Spatial and Temporal Variability of Solar Energy," *Foundations and Trends R in Renewable Energy,* vol. 1 (1), pp. 1-44, 2016.

[49] "Characterization of the spatio-temporal variations and ramp rates of solar radiation and PV," IEA PVPS Task 14 Report, [Online]. Available: http://www.iea-pvps.org/index.php?id=336.

[50] B. Espinar, L. Ramirez, A. Drews, H. G. Beyer, L. F. Zarzalejo, J. Polo and L. Martin, "Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations," *Solar Energy,* vol. 83, no. 1, pp. 118-125, 2009.

[51] N. Engerer and M. F., " KPV: A Clear-Sky Index for Photovoltaics,," *Solar Energy ,* vol. 105, pp. 679-693, 2014.

[52] J. G. Fonseca, T. Oozeki, H. Ohtake, T. Takashima and O. Kazuhiko, "On the use of maximum likelihood and input data similarity to obtain prediction intervals for forecasts of photovoltaic power generation," *Journal of Electrical Engineering and Technology,* vol. 10(3), pp. 1342-1348, 2015.

[53] J. Huertas-Tato, R. Aler, I. M.Galvan, F. Rodríguez-Benítez, C. Arbizu-Barrena and D. Pozo-Vázquez, "A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning," *Solar Energy,* vol. 195, pp. 685-696 , 2020.

[54] KNMI, "Uurgegevens van het weer in Nederland," [Online]. Available: https://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi.

[55] L. Bengtsson, U. Andrae, T. Aspelien, Y. Batrak and et al, "The HARMONIE-AROME Model Configuration in the ALADIN-HIRLAM NWP System," *Monthly Weather Review,* vol. 145, no. 5, pp. 1919-1935, 2017.

[56] M. Mller, M. Homleid, K. I. Ivarsson, A. Morten and et al, "Arome-metcoop: A nordic convective-scale operational weather prediction model," *Weather and Forecasting,* vol. 32, no. 2, pp. 609-627, 2017.

[57] J.-J. Morcrette, H. W. Barker, J. N. S. Cole, M. J. Iacono and R. Pincus, "Impact of a New Radiation Package, McRad, in the ECMWF Integrated Forecasting System," *Monthly Weather Review,* vol. 136, no. 12, pp. 4773-4798, 2008.

[58] E. J. Mlawer, S. J. Taubman, P. D. Brown, M. J. Iacono and S. A. Clough, "Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave," *Journal of Geophysical Research: Atmospheres,* vol. 102, no. D14, pp. 16663-16682, 1997.

[59] B. Elsinga and W. G. J. H. M. van Sark, "Short-term peer-to-peer solar forecasting in a network of photovoltaic systems," *Appl. Energy,* vol. 206, no. June, p. 1464–1483, 2017.

[60] e. a. Lorenz, "BENCHMARKING OF DIFFERENT APPROACHES TO FORECAST SOLAR IRRADIANCE," in *24th European Photovoltaic Solar Energy Conference*, 21-25 September 2009, Hamburg, Germany.

ISBN 978-3-906042-88-6