

An analog ensemble for short-term probabilistic solar power forecast



S. Alessandrini^{a,*}, L. Delle Monache^a, S. Sperati^b, G. Cervone^c

^a National Center for Atmospheric Research, Boulder, CO, USA

^b Ricerca Sistema Energetico, Milano, Italy

^c The Pennsylvania State University, University Park, PA, USA

HIGHLIGHTS

- A novel method for solar power probabilistic forecasting is proposed.
- The forecast accuracy does not depend on the nominal power.
- The impact of climatology on forecast accuracy is evaluated.

ARTICLE INFO

Article history:

Received 2 March 2015

Received in revised form 29 June 2015

Accepted 5 August 2015

Keywords:

Analog ensemble
Short-term solar power forecasting
Probabilistic predictions
Uncertainty quantification
Ensemble verification

ABSTRACT

The energy produced by photovoltaic farms has a variable nature depending on astronomical and meteorological factors. The former are the solar elevation and the solar azimuth, which are easily predictable without any uncertainty. The amount of liquid water met by the solar radiation within the troposphere is the main meteorological factor influencing the solar power production, as a fraction of short wave solar radiation is reflected by the water particles and cannot reach the earth surface. The total cloud cover is a meteorological variable often used to indicate the presence of liquid water in the troposphere and has a limited predictability, which is also reflected on the global horizontal irradiance and, as a consequence, on solar photovoltaic power prediction. This lack of predictability makes the solar energy integration into the grid challenging. A cost-effective utilization of solar energy over a grid strongly depends on the accuracy and reliability of the power forecasts available to the Transmission System Operators (TSOs). Furthermore, several countries have in place legislation requiring solar power producers to pay penalties proportional to the errors of day-ahead energy forecasts, which makes the accuracy of such predictions a determining factor for producers to reduce their economic losses. Probabilistic predictions can provide accurate deterministic forecasts along with a quantification of their uncertainty, as well as a reliable estimate of the probability to overcome a certain production threshold. In this paper we propose the application of an analog ensemble (AnEn) method to generate probabilistic solar power forecasts (SPF). The AnEn is based on an historical set of deterministic numerical weather prediction (NWP) model forecasts and observations of the solar power. For each forecast lead time and location, the ensemble prediction of solar power is constituted by a set of past production data. These measurements are those concurrent to past deterministic NWP forecasts for the same lead time and location, chosen based on their similarity to the current forecast and, in the current application, are represented by the one-hour average produced solar power.

The AnEn performance for SPF is compared to a quantile regression (QR) technique and a persistence ensemble (PeEn) over three solar farms in Italy spanning different climatic conditions. The QR is a state-of-the-science method for probabilistic predictions that, similarly to AnEn, is based on a historical data set. The PeEn is a persistence model for probabilistic predictions, where the most recent 20 power measurements available at the same lead-time are used to form an ensemble. The performance assessment has been carried out evaluating important attributes of a probabilistic system such as statistical

Abbreviations: T2M, air temperature at 2 m above ground; AnEn, analog ensemble; AZ, azimuth angle; BS, Brier score; BSS, Brier skill score; CRPS, continuous ranked probability score; CC, cloud cover; GHI, global horizontal irradiance; NP, nominal power; MAE, mean absolute error; MP, mean power; MRE, missing rate error; NN, neural network; NWP, numerical weather prediction; PeEn, persistence ensemble; PV, photovoltaic power; PDF, probability density function; QR, quantile regression; RAMS, regional atmospheric modeling system; RMSE, root mean squared error; EL, solar elevation; SPF, solar power forecasting; TSOs, transmission system operators; WPF, wind power forecast.

* Corresponding author at: National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307-3000, USA.

E-mail address: alessand@ucar.edu (S. Alessandrini).

<http://dx.doi.org/10.1016/j.apenergy.2015.08.011>

0306-2619/© 2015 Elsevier Ltd. All rights reserved.

consistency, reliability, resolution and skill. The AnEn performs as well as QR for common events, by providing predictions with similar reliability, resolution and sharpness, while it exhibits more skill for rare events and during hours with a low solar elevation.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Solar photovoltaic power (PV) generation has increased steadily in several countries in the last 10 years, becoming an important component of a sustainable solution of the energy problem. In Italy, for instance, thanks also to substantial government subsidies over the past 5 years, the annual generation by PV has reached 19,418 GWh in 2013, corresponding to 7% of the total Italian energy demand. The rate of growth is significant considering that in 2008 the generation from PV was just 200 GWh (0.1% of total) [1]. During the last 10 years wind power installed capacity has also followed the same trend in Italy, going from 664 MW to 8551 MW in 2013 [2], which has led to an annual generation of about 4.7% of the total Italian energy demand. As a consequence more than 10% of the total Italian energy production is now made of wind and solar resources that have a variable nature and a limited predictability. Astronomical and meteorological factors are the main causes of PV variability. The former are the solar elevation and the solar azimuth, which are easily predictable without any uncertainty. The amount of liquid water met by the solar radiation inside the troposphere is the main meteorological factor influencing the solar power production. In fact, a fraction of short wave solar radiation is reflected back outside the atmosphere or absorbed by the cloud water particles and cannot reach the earth surface. A similar process involves other kind of particles generally defined as aerosols that have different origins both natural (marine salt, soil erosion) and anthropogenic (combustion). The interaction between solar radiation and aerosols increases the amount of the diffused component (that still reaches the earth surface contributing to PV generation) but alters only marginally the reflected and absorbed components, which become significant only in the case of particularly high aerosol concentrations, such those found in the proximity of a volcanic eruption or during episodes of high urban pollution. The PV panel energy efficiency is also related to air temperature (the efficiency decreases at higher temperatures) [3]. Hence, the total cloud cover (CC), the global horizontal irradiance (GHI) and air temperature at 2 m above the ground (T2M) are the meteorological variables directly related to PV production and their limited predictability is also reflected on PV predictions. Aerosol interactions with solar radiation are generally taken into account by meteorological models, but the aerosol distribution is based on climatological estimates.

Electricity is distributed over a region via a grid, which is an interconnected network from suppliers to consumers. To maintain grid stability at an effective cost, it has now become crucial to be able to predict with accuracy the renewable energy production which is combined with other more predictable sources (e.g., coal, natural gas) to satisfy the energy demand [4,5]. The efforts of providing an accurate solar power prediction have now been mandated by recent legislation in Italy for utility companies, which have to pay penalties proportional to the forecast errors (defined as the difference between the day-ahead planned energy and the actual production). This has now put a greater focus on solar power forecasting (SPF). Even though the first methods for wind power forecasting (WPF) and SPF were both developed in early 1980s [6,7], the number of publications focused on WPF is much higher.

Predictions can be categorized into deterministic and probabilistic forecasts. A deterministic forecast consists of a single

predicted value of the variable for each prediction time, while probabilistic forecasting provides probability density functions (PDFs) from which probabilities of future outcomes can be estimated. Probabilistic forecasts also provide information about uncertainty in addition to the commonly provided single-valued (best-estimate) power prediction.

There are many examples of how probabilistic predictions can provide a higher value than deterministic ones. For instance, one example is estimating the optimal level of reserves that need to be allocated to compensate for wind and solar power variability and their limited predictability, as discussed in [8]. Another significant application is when renewable energy is traded in day-ahead electricity markets. In [9–11] it is shown that trading future wind energy production using probabilistic wind power predictions can lead to higher economic benefits than those obtained by using deterministic forecasts alone. Indeed, the maximum income for a producer is obtained by offering in the day-ahead market an amount of energy that can be different from the most expected one.

Two reviews of the status of forecasting GHI on different time scales for energy generation are reported in [12] and [13], while different forecasting techniques for PV power are evaluated and compared in [14], and a full review of SPF can be found in [15]. In this work we focus on short-range PV forecasting, i.e., 0–72 h ahead. An approach to forecasting on this time range can be based on applying statistical or machine learning techniques directly to historical time series of PV production data. Several applications of this kind can be found in the literatures [16–18]. In [19] it is shown that for forecasts up to 2 h ahead the most important input is the available observations of solar power, while for longer horizons numerical weather prediction (NWP) model output becomes crucial for a better accuracy. Even though there are several contributions on the topic of probabilistic WPF [20–23] for probabilistic PV forecasts only a few methods have been proposed in the literature. [24] and [25] propose probabilistic forecast systems for GHI only. In [24] the system is based on a stochastic differential equation framework together with NWP for modeling the uncertainty associated with the solar irradiance point forecast. In [25] the GHI forecast system is based on studying the correlation of uncertainty to local meteorological conditions describing synoptic-scale atmospheric flow. The direction and magnitude of geostrophic flow were used as an indicator of coastal cloud cover probability to produce regime-dependent forecast intervals.

In [26] PV energy probabilistic forecast for 1-h ahead are based on a Bayesian auto regressive time series model without using NWP. In [27] a normal distribution with zero mean and a standard deviation dependent on the solar zenith angle and the cloud situation is assumed to describe forecast errors of GHI. In this paper we propose the Analog Ensemble (AnEn) as a novel method for PV power forecasting over the 0–72 h lead time period. It has been originally proposed by [28] and [29] for deterministic and probabilistic meteorological forecasting, by [30] for WPF, and by [32] for wind resource assessment applications. The AnEn technique provides a set of likely PV predictions (i.e., an ensemble that is a Monte Carlo approximation of the PDF associated with future power production) using an historical dataset of observations and deterministic NWP. For each forecast lead time the ensemble

set of forecasts of PV is constituted by a set of its measurements from the past. These measurements are those concurrent to past NWP at the same lead time, chosen across the past runs most similar to the current forecast. In this paper AnEn performance is compared with a quantile regression (QR) technique. For QR, as for AnEn, only a single deterministic forecast is necessary along with an historical dataset of predictions and observations. The QR is a state-of-the-science method for probabilistic predictions based on historical data sets and has been widely used for probabilistic WPF [33,34,30]. Each desired quantile of the power forecast distribution is modeled separately defining different regression coefficients on past measured and observed PV values. Therefore, for each power quantile, a different prediction is issued based on the current deterministic meteorological forecast. Persistence ensemble (PeEn) is also used as a baseline reference. The PeEn takes the most recent 20 PV measurements at the same hour of the day to build the PDF of PV forecast. The performances of AnEn, QR and PeEn are compared at three solar PV farms in Italy with a 1-year data set of measurements, by evaluating important attributes of probabilistic predictions as resolution, reliability, spread-skill and statistical consistency. In terms of deterministic forecasting, the performance of AnEn and QR are also compared with those obtainable by applying a feed-forward Neural Network (NN) to produce a single-valued power forecast. In fact, computing the simple mean or median of an ensemble distribution allows to obtain a single-valued deterministic forecast. Also for this comparison PeEn is used as a baseline reference.

The innovative aspect of the paper is to test a new method for probabilistic PV energy forecast in the 0–72-h lead time interval. To our knowledge there are no other contributions addressing this topic on the same forecast lead time range. Furthermore, for the first time an estimate on how different climatic conditions can influence the predictability of PV forecasts is drawn.

The paper is organized as follows. Section 2 describes the observational dataset on which the analysis is performed; Section 3 presents the prediction methods, while results are presented and discussed in Section 4; conclusions are drawn in Section 5.

2. Observational and meteorological forecast dataset

Data collected from three PV farms have been considered in this study. They are located in Milano (northern Italy) and Catania (Sicily, southern Italy), both with a nominal power (NP) of 5.21 kW, and in Calabria region (southern Italy; a non-disclosure agreement does not allow us to reveal the exact location) with a nominal power around 5 MW. The Milano PV farm is located in the east suburban/industrial area of Milano city. The Catania PV farm has the same set-up as that of Milano, in terms of type of photovoltaic panels and electronic components. It is located in the suburban area of Catania close to the Mount Etna volcano. Fig. 1 displays the plants' locations.

A high level of aerosol particle concentrations due to anthropogenic emissions, which, as previously explained, can sometimes affect the ground radiation, characterizes the Milano solar farm. Volcanic ash, released intermittently with a variable intensity by the nearby volcano Mount Etna, may influence the power production at the Catania solar farm, and the ash release, transport, and dispersion is difficult to predict with a real-time forecast system. The Calabria solar farm differs from the Milano and Catania ones, which can be considered similar in size to the common small rooftop installations. It is an industrial facility with a nominal power higher than the average found in Italy for PV plants. The climatic conditions of Calabria and Catania farms are similar, with an average number of sunny days per year larger than in Milano, where the weather is cloudier and the presence of fog during winter is common. A climatic analysis [35] based on meteorological

observations recorded close to the three solar farms indicates that the average fraction of days per year with an average cloud cover lower than 4/8 is around 45%, 60% and 66% at Milano, Calabria and Catania, respectively. This is reflected in the ratio between mean produced power and nominal power that is around 25% for Milano, and 30% and 35% for Calabria and Catania farms, respectively.

Hourly averaged power data are available for the periods January 2010 to December 2011 (Catania), July 2010 to December 2011 (Milano), and April 2011 to March 2013 (Calabria). Fig. 2 shows an example of time series of solar power for each one of the power plants. Time series are plotted for one month (i.e., July 2011).

The AnEn requires an historical set of forecast runs. To this aim we used the Regional Atmospheric Modeling System (RAMS) [36], which has been run for the entire period covered by PV measurements. The computational domain consists for each of the three solar farms of two nested grids with a resolution of $15 \times 15 \text{ km}^2$ and $5 \times 5 \text{ km}^2$. Every run starts at 00 UTC and is 72 h long. The boundary conditions used are the ECMWF deterministic forecast fields starting at 00 UTC (available every 6 h with 0.125° spatial resolution). The Harrington parameterization [37] is used as the radiation scheme in RAMS. Bulk microphysics parameterization is also activated in order to account for full moisture complexity. The forecast time series of GHI, CC, direct normal irradiance (DNI), and T2M have been computed at the solar farm positions through a bilinear interpolation from the finer grid. The two angles (azimuth, AZ and solar elevation, EL) defining the sun position are then added to the forecast set of meteorological variables used as analog predictors (Section 3.1). In Fig. 3 shows all pairwise scatter plots of solar power and its explanatory variables used in this application, which allow understand their inter dependency. The plots are produced from one month of training data (i.e., July 2010), for the Catania test case. Observed power is clearly strongly linearly dependent on forecasted GHI and solar elevation, while the relationship with the azimuth angle is not linear. T2M and CC show a less evident direct dependency and a low correlation with observed power.

3. Prediction methods

3.1. Analog Ensemble (AnEn)

For each forecast lead time t , the AnEn set of solar power forecasts is constituted by solar power observations from the past. These observations are those concurrent with the past forecast at the same lead time, chosen across the past runs most similar to the current forecast. The metric used to rank past forecasts' similarity to the current forecast is defined as follows [28,29]:

$$\|F_t, A_t\| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_{f_i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (F_{i,t-j} - A_{i,t+j})^2} \quad (1)$$

where F_t is the current forecast for the lead time t at a certain location; A_t is an analog forecast for the same lead time and location as F_t ; N_v and w_i are the number of physical variables and their weights, respectively, where i is a generic index referring to different variables; σ_{f_i} is the standard deviation of the time series of the past forecasts of a given variable at the same location; \tilde{t} is an integer equal to the half-width of the time window over which the metric is computed (e.g., if $\tilde{t} = 1 \text{ h}$, the distance will be computed over the three forecast lead times corresponding to hours $t-1$, t , and $t+1$), and $A_{i,t+j}$ and $F_{i,t-j}$ are the values of the analog and the forecast in the time window for a given variable. The goal is to find past forecasts of the meteorological variables (chosen among the ones with the highest correlation with the quantity to be predicted, solar

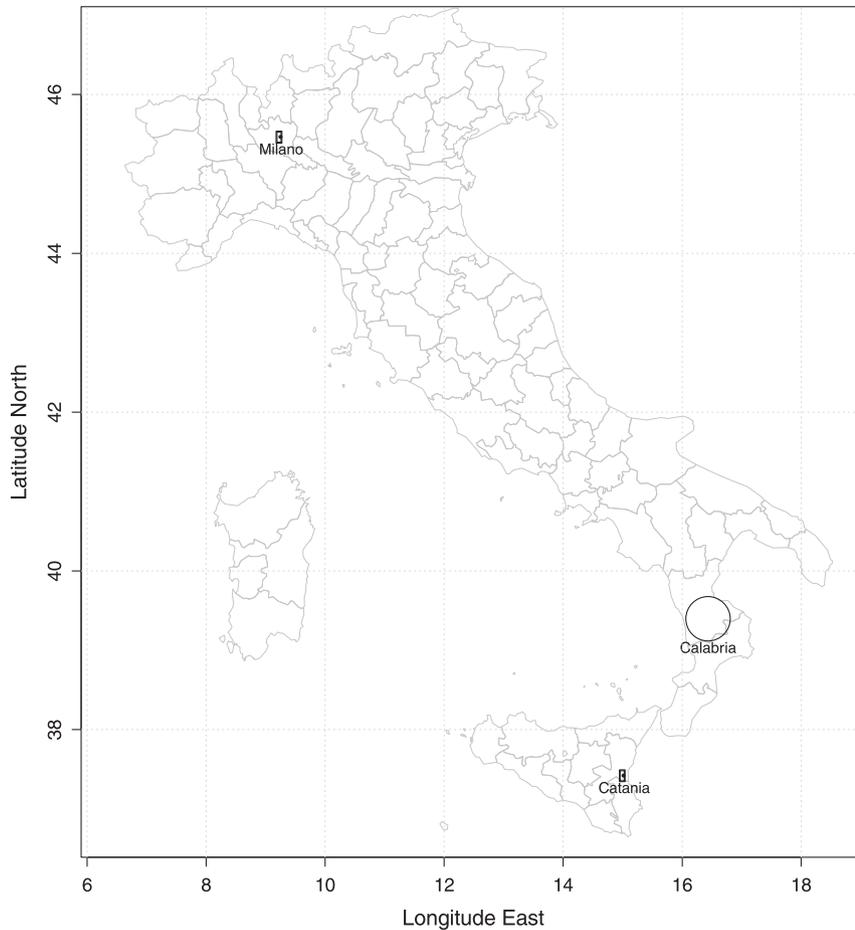


Fig. 1. Map of Italy with the positions of the solar farms. For Calabria only the approximate location is plotted.

power in this case) that were predicting similar values and temporal trend (i.e., which have a similar behavior as a function of time along the time interval defined by \tilde{t}) compared to the current forecast. The predictors used in this application are GHI, CC, T2M, AZ and EL with a half-width time window (\tilde{t}) equal to 1 h. The AZ and EL variables are added as predictors in order to define the sun's position and eventually take into account the presence of obstacle shadows (buildings or mountains, for example) influencing the power production at some particular time of the day.

The assumption is that if similar past forecasts are found, their errors will likely be similar to the errors of the current forecast, which can be inferred from theirs. The main steps of the algorithm can be summarized as follows:

- Retrieve a historical dataset of predictions (initialized at a given time) issued by a deterministic system, a meteorological model in this case, for the solar farm location (see Section 2 for a description of the NWP model).
- Retrieve an historical dataset of solar power observations at the locations of interest (see Section 2 for details).
- Choose the physical variables from the meteorological model to be used as predictors in Eq. (1) for the predictand variable (solar power in this application).
- For each lead time of the current forecast compute the distance (i.e., Eq. (1)) from every past forecast issued at the same lead time.
- For each lead time, rank all the past forecasts and select the n forecasts with the lowest distance.

- The concurrent n past measurements are the n members that constitute the current AnEn forecast for the lead time considered (in the current application 20 past observations are used).

The AnEn attempts to capture error dependent on a particular process of the atmospheric flow by assigning the observed errors from similar past situations, as described by the high-resolution deterministic model, to the current model forecast. The AnEn has potential advantages and disadvantages as compared to a NWP ensemble. One advantage of the AnEn may be to lower significantly the real-time computational expense of generating an ensemble, as AnEn only requires a single model forecast, as opposed to the multiple model runs of an NWP-based ensemble. Another advantage is that the forecast uncertainty is based solely upon past observations, thereby eliminating the need to simulate all sources of NWP forecast uncertainty via sophisticated and computationally intensive techniques, and, as will be shown, also avoiding the need for post-processing calibration. One disadvantage is the need of a “frozen” meteorological model in the training data set. Indeed, significant changes in the NWP forecast configuration may prevent the generation of skilful analogs.

The following is a simple case to illustrate how the AnEn algorithm works, based on using one meteorological predictor (GHI) and a time window equal to 0, i.e., $N_p = 1$ and $\tilde{t} = 0$ in Eq. (1). Let the current 12-h ahead forecasted GHI at a given location be equal to 300 W/m^2 and assume that the past set of GHI forecasts is made by only five forecasts for the same lead time. They are equal to: 200 W/m^2 , 300 W/m^2 , 350 W/m^2 , 420 W/m^2 and 500 W/m^2 . From

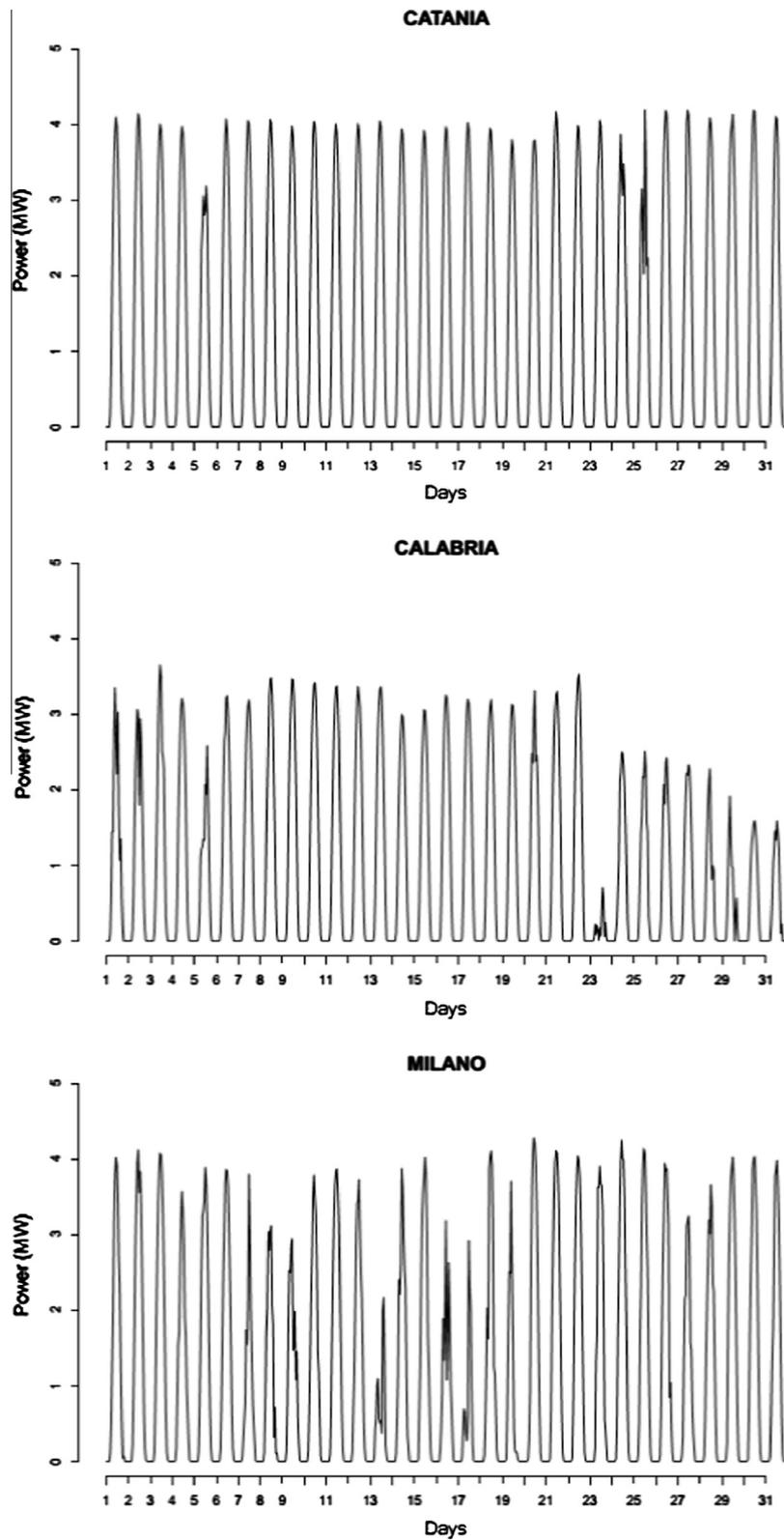


Fig. 2. Time series of solar power at the test locations for July 2011.

the past data set, the five PV concurrent observations are: 300 kW, 400 kW, 390 kW, 450 kW and 600 kW, respectively. Considering that the standard deviation of the set of GHI forecast is equal to approximately 114.3 W/m^2 , it is possible to compute the five

correspondent distances between the current and the past GHI forecast by means of Eq. (1). Those normalized distances are, respectively: 87.4, 0, 21.8, 125.9, and 349.7. To generate a 3-member ensemble power forecast (in real cases usually more than

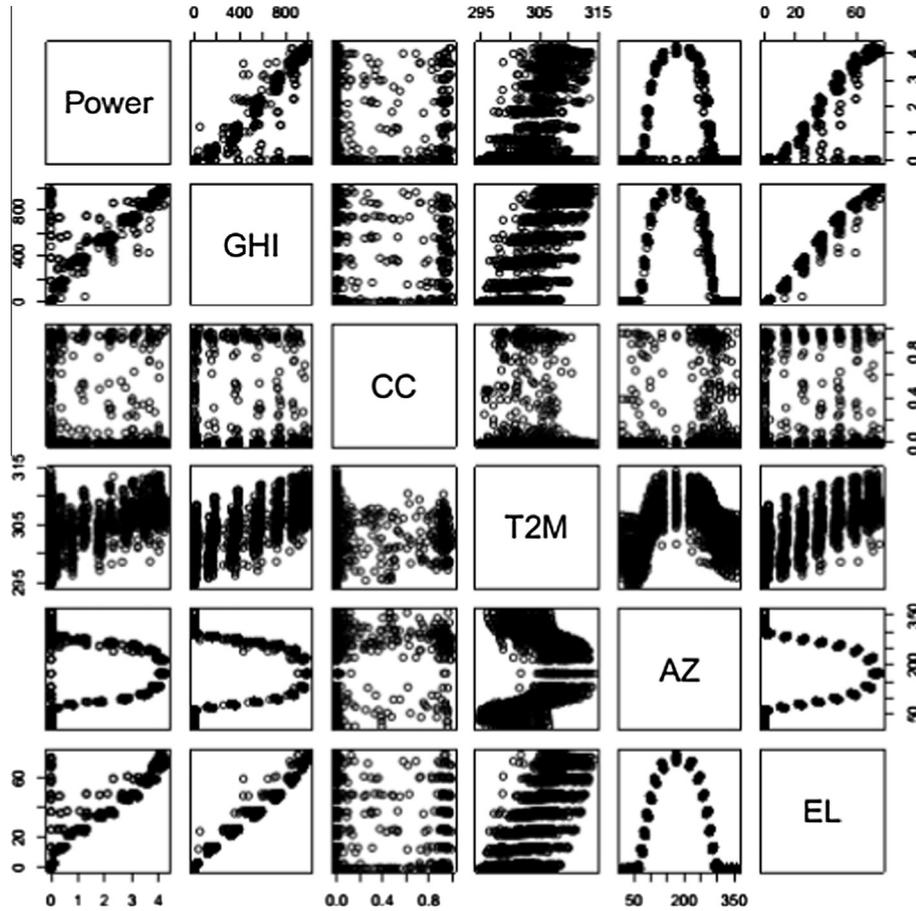


Fig. 3. Pairwise scatter plots of the explanatory variables (Catania, July 2010).

10 members are used), the members will be chosen from the past power data whose concurrent forecasts result to have the lower three distances. In this example they are 300 kW, 400 kW and 390 kW, which now represent the current 12-h ahead ensemble solar power forecast.

Similarly to what has been done in [30,31], an optimization method has been applied to define the set of weights (w_i) used in Eq. (1). For all the solar farms the data sets are split into two parts. The first 365 days (183 days for Milano) are considered as a training period while the evaluation of the AnEn model, shown in the following sections, is carried out on the remaining portion of the data sets (verification period). The set of optimal weights is defined by choosing the combination that minimizes the continuous ranked probability score (CRPS, see Section 4.3) over the last 60 days of the training periods (from day 306 to day 365 for Calabria and Catania, from day 124 to day 183 for Milano). As we are using five predictors (GHI, CC, T2M, AZ and EL), five corresponding weights can be set. All the possible combinations defined with the constraint $\sum_{i=1}^5 w_i = 1$, where $w_i \in [0 - 0.2, \dots, 1]$, are considered for the AnEn computation over the last 60 days of the training period. The weight values for the five predictors (GHI, T2M, CC, EL and AZ) leading to the minimum CRPS following this procedure are: 0.3, 0., 0.5, 0.1, and 0.1 for Catania, 0.4, 0.5, 0.1, and 0.0 for Milano, and 0.3, 0.1, 0.5, 0.0, and 0.1 for Calabria, respectively. Those values are kept constant over all the verification periods and are used for the results shown hereafter. An example of AnEn forecast is illustrated in Fig. 4 (top panel, see comments in Section 3.3).

3.2. Quantile Regression (QR)

A probabilistic power forecast system based on QR [33,34] is included in this analysis as a reference method. Considering a random variable Y , $Q(\tau)$ is defined as the value for which the probability of obtaining values of Y below $Q(\tau)$ is τ . In QR, $Q(\tau)$, with $0 < \tau < 1$, is expressed as a linear combination of known regressors and unknown coefficients, exactly as the mean of the random variable Y is modeled in (multiple) linear regression. Thus the τ -quantile is modeled as:

$$Q(\tau) = \beta_0(\tau) + \beta_1(\tau)x_1 + \dots + \beta_p(\tau)x_p \quad (2)$$

where x_p are the p known regressors, also called explanatory variables, and β_p are unknown coefficients, depending on τ , to be determined from N observations.

The sample β_p coefficients for the τ quantile can be found by minimizing the cost function:

$$\sum_{i=1}^N w_i \cdot \rho_\tau [y_i - (\beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p})] \quad (3)$$

where w_i is a vector of weights (between 0 and 1) given at the observations y_i of the random variable Y , and ρ_τ is the check function defined as:

$$\rho_\tau(e) = \begin{cases} \tau e, & e > 0 \\ (\tau - 1)e, & e \leq 0 \end{cases} \quad (4)$$

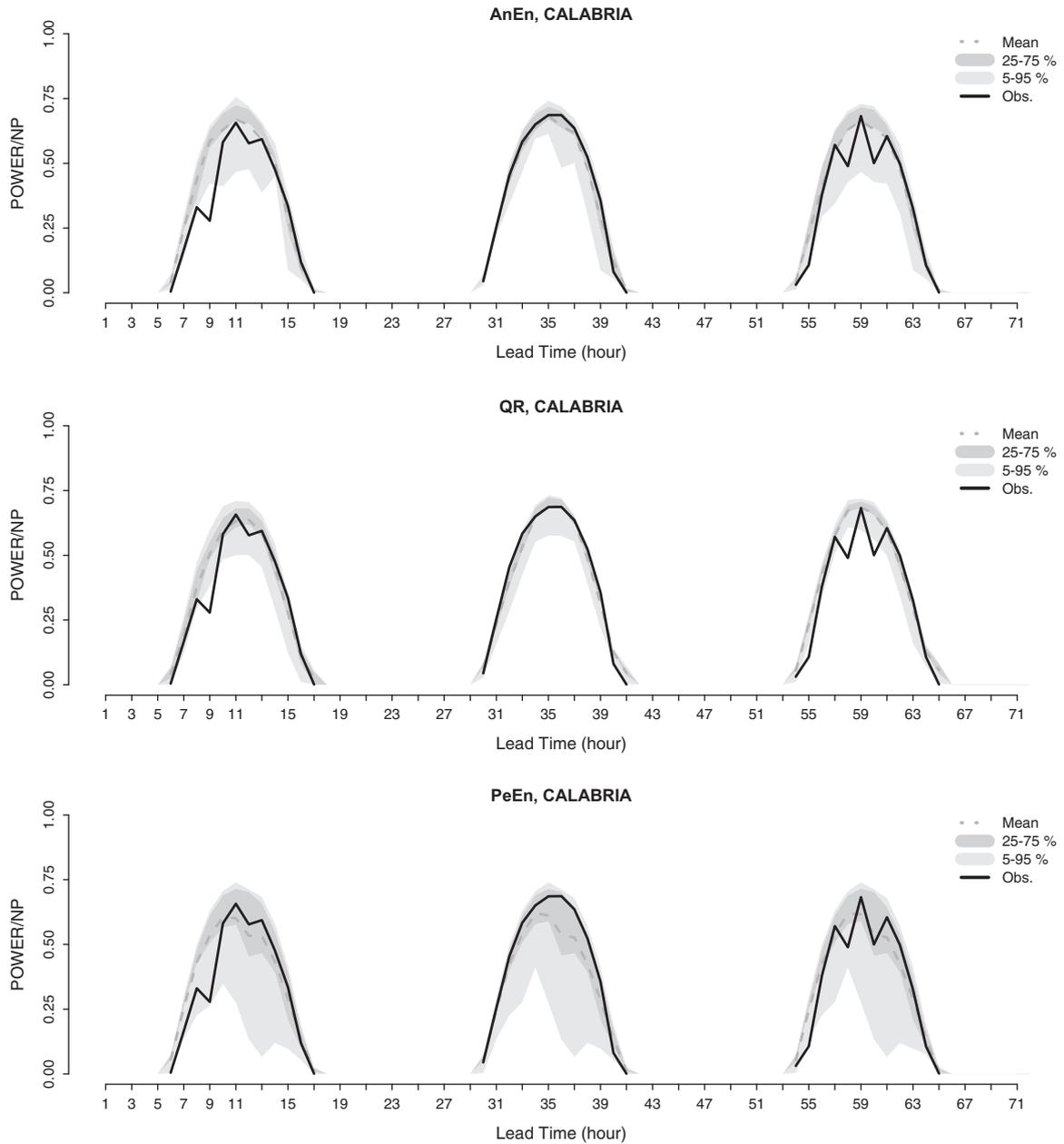


Fig. 4. Example of (top) analog ensemble (AnEn), (middle) quantile regression (QR), and (bottom) persistence ensemble (PeEn) forecast PDFs for 27 September 2011. The grey shadings correspond to the 25–75 (darker) and 5–95 (lighter) quantiles. The black and dashed lines represent the power observation and ensemble mean, respectively.

The coefficients β_p can be estimated with linear programming techniques. In this case, as in [34], the R add-on package *quantreg* has been used.

Additive models [38] are also used to account for the non-linearity between solar power and its explanatory variables. Such models can be expressed as follows:

$$y = a + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \varepsilon \quad (5)$$

where the constant a and the functions $f_j(x_j)$ are estimated from data. It is important to underline that the estimates are non-unique (e.g., it is possible to add a constant to one function and subtracting it from another), unless restrictions are imposed on the functions. In [38] the authors impose on each function restrictions of having zero mean over the data. In that case, it is possible to approximate each function using linear combination of known basis functions of the corresponding input variable as follows:

$$f_j(x_j) = \sum_{k=1}^{n_j} b_{jk}(x_j) \theta_{jk} \quad (6)$$

where $b_j(x_j)$ are the basis functions and θ_j are the unknown coefficients. In order to obtain unique estimates, restrictions are then imposed on Eq. (6) and the resulting basis functions are derived. Imposing $f_i(0) = 0$, then $\theta_{j1} = -\sum_{k=2}^{n_j} \theta_{jk} b_{jk}(0) / b_{j1}(0)$. Using θ_{j1} in Eq. (6) results in the following expression:

$$f_j(x_j) = \sum_{k=2}^{n_j} \left(b_{jk}(x_j) - \frac{b_{jk}(0)}{b_{j1}(0)} b_{j1}(x_j) \right) \theta_{jk} \quad (7)$$

where the $n_j - 1$ new basis functions are defined by the term in front of the coefficients θ_{jk} . The result is then a linear regression model in which the estimates of a and θ_{jk} are unique.

Examining Eq. (2) it can be seen that the model can be generalized to quantile regression by modeling $Q(\tau)$ as follows:

$$Q(\tau) = a(\tau) + \sum_{j=1}^p f_j(x_j; \tau) = a(\tau) + \sum_{j=1}^p \sum_{k=1}^{n_k} b_{jk}(x_j) \theta_{jk}(\tau) \quad (8)$$

In Eq. (8), the basis functions are constructed under appropriate restrictions as described above.

In [34] QR is used to forecast errors of WPF but, as mentioned by the authors, this can lead to unrealistic quantiles (i.e., a lower quantile crossing a higher one). As suggested in [34] a possible way to avoid such unrealistic quantiles is to forecast the power production instead of the forecast error, which is what is performed in this study and already implemented successfully for WPF in [30]. For all the variables a cubic spline basis with 10 degrees of freedom is constructed, using no intercept; this implies that the functions are restricted to be zero at the lower boundary knot. The boundary knots are placed at the limits of the data, while the inner knots are placed at equally distant quantiles of the cumulative function of the individual explanatory variables. In this way the knots result relatively closer where the observations of the variable are denser. In the prediction phase it is then important to use the same knots defined during the training phase, where both the observations and the explanatory variables are available. Since none of the bases allow for a free intercept, this is handled by an intercept in the model. The forecast lead time is introduced as an explanatory variable because the meteorological model accuracy depends on it. Similarly to what has been done for AnEn, an optimization process is performed in order to choose the best set of regressors among the available variables, i.e., GHI, T2M, CC, EL, AZ and the forecast lead time (LT), keeping PV measurements as realizations y_i of the random variable Y . All the possible combinations are used to compute the QR forecast over the last 60 days of the training period and for each of them the CRPS is computed. The lowest CRPS value resulted from the following set of regressors: GHI, CC, EL, LT for Catania, GHI, T2M, CC, EL, AZ, LT for Milano and GHI, CC, LT for Calabria. Those set of variables are kept unchanged over the verification periods and are used for the results shown hereafter.

A similar optimization is performed on the same periods in order to define the weights vector w_i . The function $w_i = e^{-i/\tau_s}$ (where i indicates the time distance expressed as number of hours between the current forecast and the past forecast in the training data set) is introduced in order to let the weights assigned to the past observations decrease in time with a “forgetting factor” equal to τ_s . Several attempts are made, with τ_s changing between 400 h, and $+\infty$ (i.e., when all the weights w_i are set equal to 1); the choice of $\tau_s = 800$ h resulted in the minimum CRPS for all three solar farms.

Similarly to AnEn, the solar power forecast PDF is estimated with 20 quantile intervals from 0.0 to 1.0 with step 0.05. The same training and test periods previously described have also been used. The QR coefficients are updated using new available power observations every 24 h during the verification period, and then applied on the following 0–72 h predictions. An example of QR forecast is illustrated in Fig. 4 (middle panel, see comments in next section).

3.3. Persistence ensemble (PeEn)

The PeEn forecast is used as a baseline prediction to be compared with AnEn and QR forecasts. In this work the PeEn ensemble for each of the 72 h ahead is made of the most recent available 20 measured PV values at the same hour. This set of values can eventually be ranked to define a set of power quantile intervals. The PeEn ensemble forecast can be accurate if clear sky conditions persist for several consecutive days. Indeed, at the same hour of the

day similar levels of power are expected as the sun position is quite similar.

An example of PeEn forecast is illustrated in Fig. 4 (bottom panel). The PeEn forecast provides very similar PDFs for the three days ahead, while both AnEn (top panel) and QR (middle panel) PDFs change depending on the underlying meteorological conditions. This results in quantile ranges wider for PeEn with respect to AnEn and QR, which can be excessive in the clear sky conditions observed in day 2, when a lower prediction uncertainty is expected. Indeed, AnEn PDF becomes sharper on the second forecast day, while is able to account for a wider uncertainty on the first and the third day when cloudier situations are observed. Also, QR is sharper than AnEn on the first and the third day when the power variability introduced by cloudy conditions is higher.

3.4. Neural Network (NN)

A forecasting system based on NN is included to compare forecast performance from a deterministic point of view. The system is based on artificial, feed-forward NN with single hidden layer and is applied using a dedicated package (“nnet”) of R [39]. This NN has a simple structure that connects some inputs variable with a response variable, passing by one hidden layer of neurons, which process the information. The authors have applied this type of NN in [20,30].

The NN is configured using forecast lead time, GHI, CC and T2M as input variables and measured power as the output variable. The NN initial configuration is defined by a set of weights, which are randomly initialized. For this reason, multiple trials are performed, training the NN starting from a different set of random weights. The training is performed on the same period previously defined for the other models. The last 60 days of the training periods are not used in the training phase but to select the best NN. For each independent trial the root-mean-square-error (RMSE) between the output of the NN and the measured power is computed. The NN providing the lowest RMSE is then selected to compute the forecasts for test period. An optimization is also made to choose the optimal set of input variables. In all the test cases, the optimal set is made of GHI, CC, T2M, EL and forecast lead time. Furthermore, the choice of 4 neurons in the hidden layer provides the best results for each test case.

4. Results

4.1. Statistical consistency

Consistency indicates whether the members of an ensemble system are statistically indistinguishable from the observations. If an ensemble system is statistically consistent, an observation ranked among the corresponding ordered ensemble members is equally likely to take any rank in the range of the whole PDF [40]. A rank histogram can be used to assess whether the observations are equally distributed among the forecasted PDF. A perfectly uniform rank histogram would be flat, i.e., with uniform rank probability of $1/(n+1)$ [41] where n is the number of ensemble members.

The vertical bars represent the confidence intervals [42,43] and are calculated with a quantile function for a binomial distribution. The confidence intervals delimit the 5–95% quantile interval of the binomial distribution. They assess the confidence in the estimated distance from the perfect forecast (the one lying along the horizontal bar), which varies based on the available samples in each bin. E.g., with a limited number of samples a deviation from the perfect forecast could just occur by chance, while with an infinite number samples would reduce the confidence interval length to 0. The

Missing Rate Error (MRE), which is the fraction of observations lower (higher) than the lowest (highest) ranked prediction above or below the expected missing rate of $1/(n+1)$, is also shown. A larger positive (negative) MRE reveals a more under dispersive (over dispersive) ensemble. Fig. 5 shows the rank histograms for the different forecast techniques and solar farms. Only the forecast lead times when the hourly average solar elevation is greater than 0 are considered. The AnEn shows the overall best level of statistical consistency with the lowest values of MRE. There is a slight tendency of underestimating the frequency of the highest bins at Milano and Catania. Considering that for AnEn the 20-member ensemble is made of past measurements (selected as described in Section 3.1), this means that the observed PV is more likely to be lower than the highest ensemble members, which could possibly indicate that the produced power decreases with time. This decrease in power could potentially be explained by a lower frequency maintenance of the PV panels, which would result in a performance degradation. The latter could be a

plausible explanation considering that, differently from Calabria, both Milano and Catania plants are not industrial facilities aiming to optimize the energy production, and they are likely not as well maintained. If the maintenance schedule were known, it would be possible to correct the past observations taking into account for the performance degradation, and likely further improve the AnEn performance. The PeEn and QR are both under dispersive. The PeEn is made of the most recent 20 observations that are not able, in general, to capture the observed variability. For QR the under dispersive behavior could be attributed to the optimization process described in Section 3.2, which sets the forgetting factor to a relatively low value (i.e., 800 h) for the optimal weights. While this choice leads to the best performances in term of CRPS, on the other hand it reduces the average inter-quantile distance (i.e., the ensemble spread). In other tests (not shown here) with higher values of τ_s we have obtained better-calibrated ensembles by the QR model but with a higher (worse) CRPS.

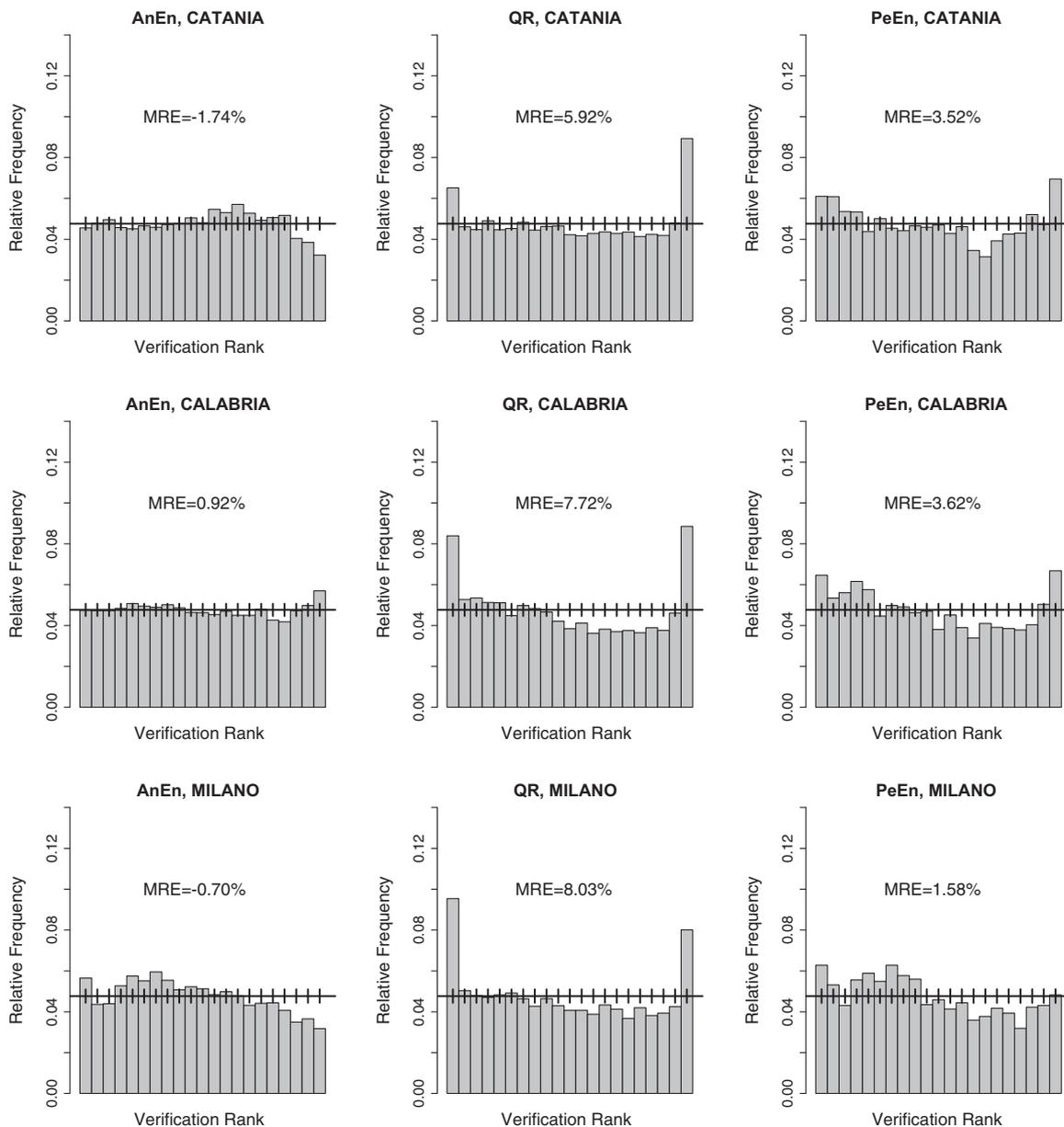


Fig. 5. Rank histogram for Analog Ensemble (AnEn), Quantile Regression (QR) and Persistence Ensemble (PeEn) computed over all the lead times with a positive hourly average solar elevation. Each row refers to a different solar plant (Catania, top, Calabria, middle, and Milano, bottom).

4.2. Spread-skill consistency

The ability of a probabilistic prediction to quantify its uncertainty can be assessed by compiling binned-spread/skill diagrams. In a spread/skill diagram the ensemble spread is compared to the RMSE of the ensemble mean over small class intervals (i.e., bins) of spread, instead of considering its overall average [44–46]. A good correlation in the spread/skill diagram is an indication that an ensemble system is able to forecast its own error [46]. If the observations are indistinguishable from the predictions, i.e., the observations and the ensemble members are samples from the same distribution, it can be shown that the ensemble standard deviation (i.e., the ensemble spread) and the RMSE of the ensemble mean should be equal. However, if the ensemble members and ensemble mean errors do not have a Gaussian distribution, the spread-RMSE matching requirement is only a necessary condition for spread-skill consistency. The latter is satisfied if the two indices match at all values (i.e., the resulting trend lies on the plot's 1:1 diagonal). Binned-spread/skill diagrams for each prediction system are reported in Fig. 6. Each bin has the same number of data points, which results in bins of different width. The QR and PeEn behavior is consistent with what is seen in the rank histograms. There is an under estimation of the spread that is often lower than the RMSE (i.e., the methods are always under dispersive). The PeEn shows the largest range of forecast errors (RMSE). The AnEn has a similar trend to PeEn and QR for Milano, even if the spread is slightly over estimated for the bins corresponding to the highest RMSE. For Catania and Calabria the spread/skill relationships are generally better for AnEn, considering the confidence intervals overlap or lie very close to the diagonal.

4.3. Continuous ranked probability score

In order to assess the quality of the two ensemble systems without being limited by a specific threshold value, as is the case with reliability diagrams, Brier score, or relative operating characteristic (ROC) skill score, the continuous ranked probability score (CRPS) is computed. The CRPS is the equivalent of the Brier score integrated over all possible threshold values. In other words, it compares a full probabilistic distribution with the observations, where both are represented as cumulative distribution functions (CDF) [47]. The more the PDF originating the CDF is sharp and

centred on each observation, the lower the CRPS is. It is given by the formula

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\infty} (F_i^f(x) - F_i^o(x))^2 dx \quad (9)$$

where $F_i^f(x)$ is the CDF of the probabilistic forecast and $F_i^o(x)$ is the CDF of the observation for the i th ensemble prediction/observation pair, and N is the number of available pairs. In the literature it is shown that the CRPS reduces to the mean absolute error (MAE) for a deterministic forecast [48]. A lower value of the CRPS indicates better performance, with 0 being a perfect score. The index is expressed in the same unit as the forecasted variable.

In Fig. 7 the CRPS as a function of forecast lead time is plotted for the three solar farms and the three probabilistic prediction methods, normalized by both nominal power (left axis) and mean measured power (right axis). Next to the left vertical axes the CRPS (and its confidence interval) computed with all available lead times (considering only the hours with a positive average solar elevation), is shown. These average CRPS values for QR and AnEn are consistently lower than those of PeEn. When looking at CRPS computed independently for each lead time, in the early morning and late afternoon when the solar elevation is low, QR is at times worse than PeEn and AnEn (see for instance lead time 29 for Catania and lead time 65 for Calabria). This can be explained by the lower correlation between the forecasts and the observations when the solar elevation is low. This lower correlation affects the QR performance more than AnEn, because in QR the regression coefficients are computed considering all the data in the training data set, while in AnEn only a few cases are selected. At Milano QR shows the lowest total average CRPS and also the best performance during the peak production hours. At Catania and Milano there is no statistically significant difference between QR and AnEn as the total average CRPS bars overlap. To analyze the relative performances of the different forecast systems at the different plants, the CRPS can be also normalized with the mean measured power (MP) to account for different climatic conditions (i.e., different MP/NP ratio). For Catania and Calabria QR and AnEn show an average CRPS/MP of about 15%, i.e., a similar level of accuracy in similar climatic conditions. At Milano, QR and AnEn are close to 20% and 21%, respectively, i.e., they exhibit a worse performance in a climate with more cloudy days, which is expected given the reduced solar power predictability.

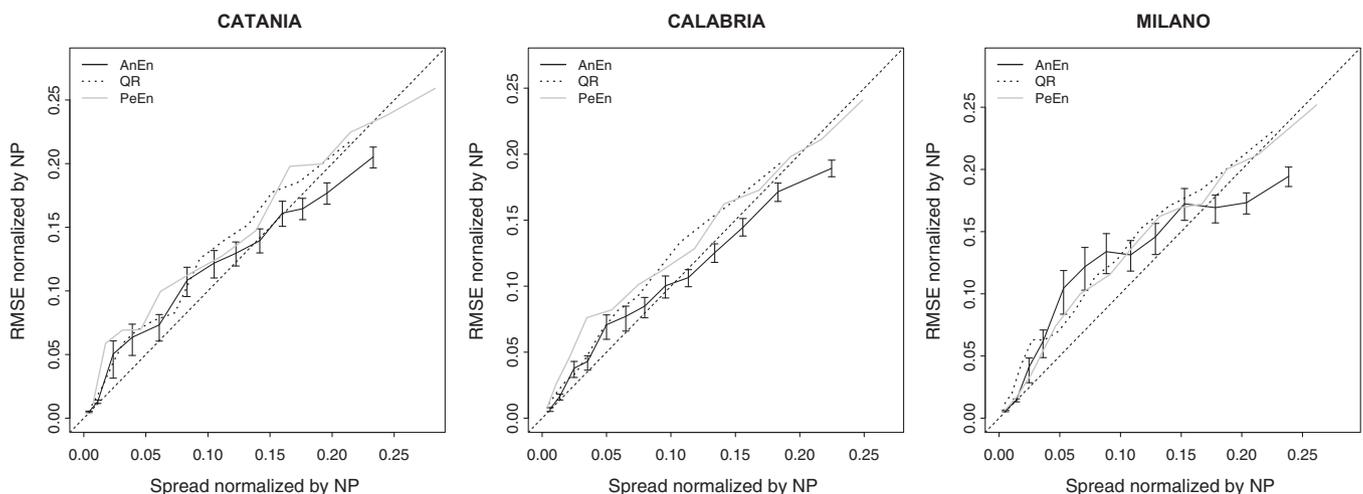


Fig. 6. Binned Spread/skill plots for the three solar farms and three different forecast methods. Both root mean square error (RMSE) and spread values are normalized by nominal power (NP); the 5–95% bootstrap confidence intervals are plotted for AnEn only to reduce clutter.

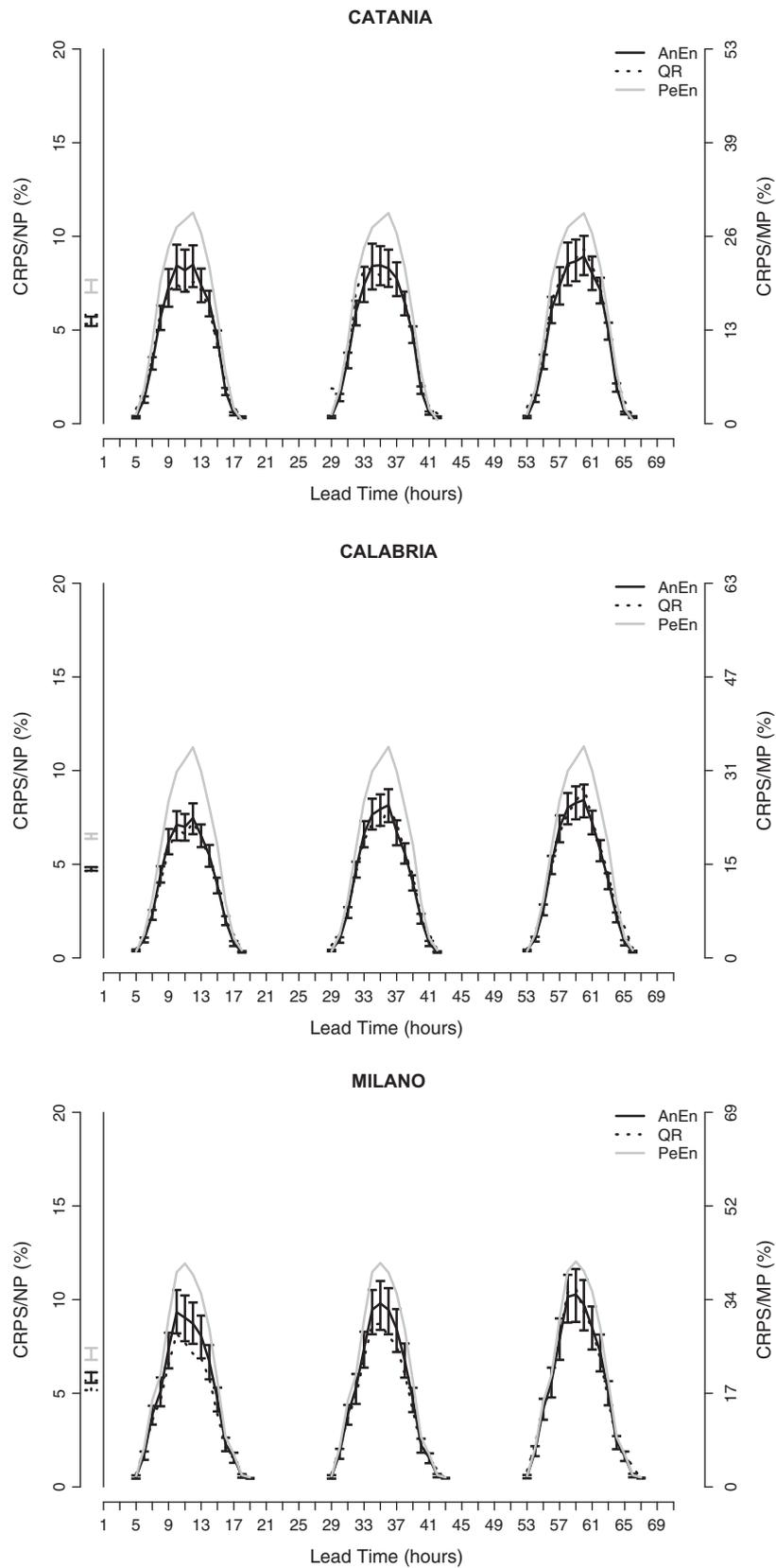


Fig. 7. Continuous ranked probability score (CRPS) normalized by nominal power (NP, left axis) and mean measured power (MP, right axis) as a function of forecast lead time; 5–95% bootstrap confidence intervals are plotted for AnEn only to reduce clutter. The vertical bars next to the left vertical axis indicate CRPS considering all the lead times together (excluding the hours with a negative average solar elevation).

4.4. Brier skill score (reliability and resolution)

Reliability is a measure of the similarity between the forecasts and the observations in a probabilistic system. It is not sufficient to characterize the quality of a probabilistic forecast, though it is an essential attribute of probabilistic predictions conditional to resolution [49].

A probabilistic forecasting system would be perfectly reliable if, statistically, the forecast probabilities were equal to the observed frequencies. In other words, taking into consideration all the cases in which the occurrence of an event is predicted with a certain probability bin, such an event should occur with a frequency equal to the mid-point of the probability bin [50]. Resolution measures the forecasts' ability to predict when an event occurs or not [51]. Probability forecasts with perfect resolution forecast 100% on occasions when the event occurs and forecast 0% when the event does not occur.

An assessment of reliability and resolution can be performed calculating the Brier score (BS) components [51,52]. The Brier score is equivalent to RMSE for a deterministic forecast and it measures the difference between the predicted probability of an event and its occurrence with the formula:

$$BS = \frac{1}{N} \sum_{i=1}^N (p_n - o_n)^2 \quad (10)$$

where p is the forecasted probability of a categorical event, o_n is the categorical observation (1 if the event occurs, 0 if it does not occur), and N is the total number of (p_n, o_n) pairs. As with the RMSE, a lower value of the Brier score implies a better performance, with 0 being a perfect score. The Brier score can be decomposed in the reliability, resolution, and uncertainty components. The uncertainty part is related to the variance of the observations and is not an attribute of the forecasting system. The highest

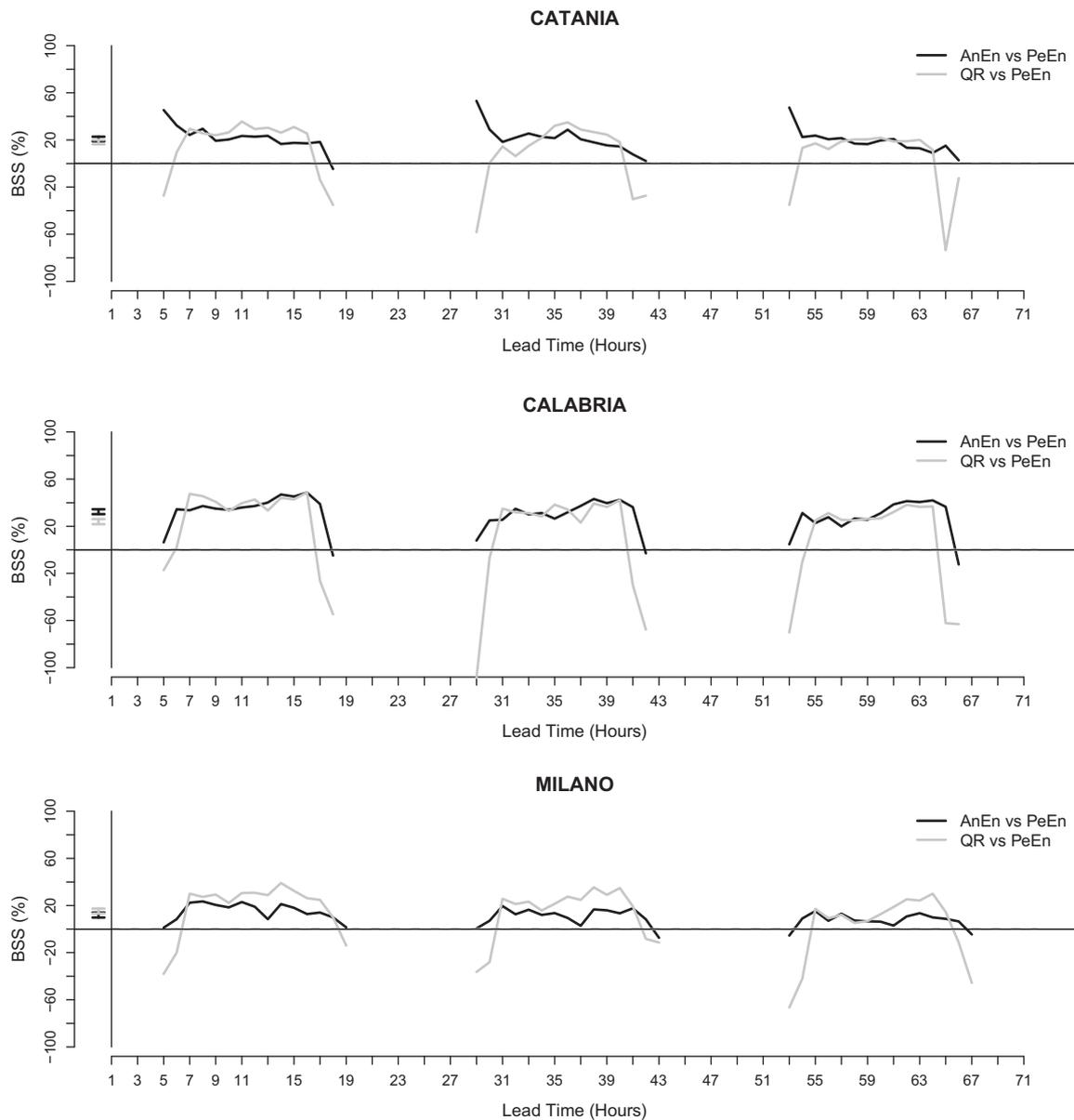


Fig. 8. Brier skill score (BSS) of analog ensemble (AnEn) and quantile regression (QR) (with persistence ensemble, PeEn, as reference) as a function of forecast lead time for the three different solar power plants. Next to the left axis is shown the BSS 5–95% bootstrap confidence interval computed considering all the lead times together and only the hours with a positive average solar elevation. The event considered is solar power forecasts greater than the mean power at each lead time.

uncertainty occurs when the climatological frequency of a given event is equal to 0.5.

Having a reference forecast (e.g., PeEn) it is then possible to calculate the Brier skill score (BSS), which is given by the formula:

$$BSS = 1 - \frac{BS}{BS_{ref}} \tag{11}$$

where BS is the Brier score of the system to be evaluated, and BS_{ref} is the BS of the reference forecast. The BSS measures the ability of the

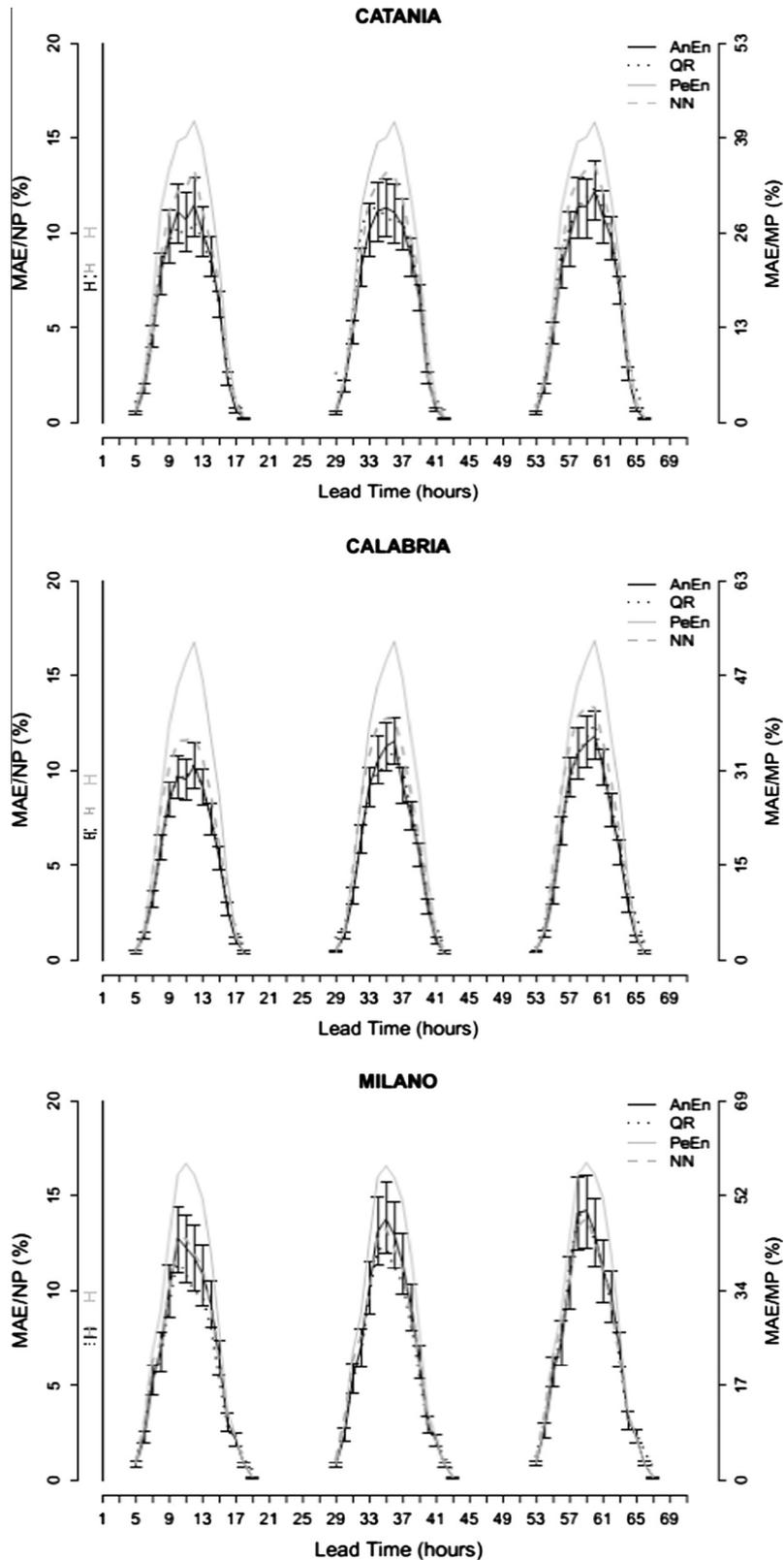


Fig. 9. Percentage mean absolute error normalized by nominal power (MAE/NP%, left y axis) and by mean power (MAE/MP%, right y axis) of analog ensemble (AnEn), quantile regression (QR), Neural Network (NN) and persistence ensemble (PeEn) as function of forecast lead time. Bootstrap 5–95% confidence intervals are plotted for AnEn only to reduce clutter. Next to the left y-axis the metric confidence interval is computed by pulling all the lead times together for the hours with a positive average solar elevation.

model to issue a better probabilistic forecast than the reference. Positive (negative) values of *BSS* indicate better (worse) performances than the reference system.

Fig. 8 shows the *BSS* values of AnEn and QR (setting PeEn as reference) as a function of forecast lead time and for the three solar power plants. Similarly to Fig. 5, next to the left vertical axis are shown the *BSS* 5–95% bootstrap confidence intervals computed considering all the lead times together and only the hours with a positive average solar elevation. In Fig. 8 the event considered is PV greater than the mean power. Given the diurnal variation of the mean power, different thresholds are taken at each lead time. In Fig. 8, AnEn and QR show similar performances (except at Milano, see discussion below) and are better than PeEn, with values of *BSS* between 20% and 40%, except for lead times corresponding to a low solar elevation (early morning and late afternoon). At those lead times QR becomes worse than PeEn. The reason for that can be addressed to the lack of correlation between the observations and the regressors as already discussed in Section 4.3. At Milano and Catania the total *BSS* of AnEn and QR are not statistically different and are close to 20% and 15%, respectively. At Calabria AnEn is statistically significantly better than QR with values close to 35% vs. 25%.

4.5. Evaluation of deterministic predictions

In this section a quantitative assessment of the deterministic prediction quality (also referred to as spot power forecasts) is presented, given that they are still used by several end users for a range of applications. A straightforward way to obtain a deterministic prediction from a probabilistic one is to take, for every issued forecasted PDF, its mean or median.

A common verification framework, as suggested for WPF by [53], should be used to evaluate a deterministic forecast. The mean absolute error is used to evaluate the three different deterministic forecast, which can be expressed as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |o_i - f_i| \quad (12)$$

where o_i is the i th observed value and f_i the i th forecasted value. MAE allows measuring the average error magnitude in the forecasts. As a selection criteria of forecast–observation pairs in Eq. (8) we kept only the hours with a positive average solar elevation that are relevant for PV forecasting. The median of every issued forecast PDF is used as the f_i because it provides lower MAE than the mean value of the PDF [54].

The RMSE is alternative common verification index for deterministic predictions, which completely defines the error distribution under the assumption that errors are unbiased and follows a normal distribution. However, RMSE is a quadratic score index, that gives higher weights to larger forecast errors. The MAE, being a linear score, gives on average the same weight. Considering that the penalties paid by the producers are usually proportional to the imbalances (i.e., the differences between forecasted and produced power) the MAE choice should be preferable to evaluate the PV forecast performances. In Fig. 9 MAE values are normalized by NP (left vertical axis) and reported as a percentage for every lead time separately and over all the lead times together (as in Fig. 7). The MAE is also normalized by MP (right vertical axis) during the test period. The AnEn and QR show a total MAE/NP statistically significantly lower than PeEn with values close to 6–7% for all three sites. The AnEn is slightly better than QR at Catania but slightly worse at Milano, while they have similar total MAE/NP at Calabria. The QR achieves better accuracy than AnEn at Milano during the central hours of the first two forecast days, while again QR performance degrades at early morning and late afternoon. NN

shows slightly worse results than both AnEn and QR at Catania and Calabria, exhibiting higher errors during the central hours of each forecast day, while at Milano it performs similarly to AnEn. The different climatology of the three solar farms affects the performances of QR and AnEn particularly in terms of MAE/MP. In fact, at Catania and Calabria, MAE/MP total average values are under 20% while at Milano they are around 25%. Therefore, the impact of more cloudy weather and a polluted environment in terms of predictability in the 1–72 h range can be quantified for this data set to be 7–10% of MAE/MP. A higher impact on this index may occur when dealing with locations with weather more cloudy and higher power variability than Milano.

5. Conclusions

This study evaluates the analog ensemble (AnEn) method, which was originally designed for meteorological probabilistic forecasting and already tested for wind power forecasting, for solar power probabilistic and deterministic forecasting, and for wind resource assessment applications. The test is carried out using measured hourly power data from three solar farms located in Italy. The AnEn best configuration is found through an iterative optimization process over a period of 60 days of the training data set carried out independently for each solar farm. A 3-h time window and 20 analogs are also chosen as parameters of the AnEn configuration. The AnEn performance is compared to other methods such as quantile regression (QR) and persistence ensemble (PeEn). A comparison with a Neural Network (NN) method is also performed to assess forecast performance from a deterministic point of view. These methods have the advantage of not requiring an irradiance-to-power conversion function. In fact, only the meteorological forecasted variables global horizontal irradiance (GHI), cloud cover (CC), 2 m temperature (T2M), solar azimuth (AZ), solar elevation (EL) and photovoltaic power (PV) measurements are used for this application. Using an irradiance-to-power conversion function (that would be specific for each solar farm) would require the additional availability of GHI observations time series, to post-process the GHI by a correction procedure before applying the conversion function. An optimal choice of predictors of the QR weights function is also found on the same 60-day periods used for the AnEn optimization. NN is also optimized by choosing its best set of initial parameters on the same 60-day period as above, before being applied over the test period. The PeEn is a 20-member ensemble including the most recent 20 observations available at the same forecast lead time.

The AnEn shows in general the best level of statistical consistency compared to QR and PeEn, which are both under dispersive, as shown by the rank histograms. The PeEn being made of the last 20 observations is not able to capture the whole observations variability, while QR results are under-dispersive because of the relatively short forgetting factor chosen in the weighting function to optimize continuous ranked probability score (CRPS).

The QR and AnEn exhibit similar results in terms of CRPS and mean absolute error (MAE) when computed overall lead times. The AnEn is slightly better at Catania and slightly worse at Milano than QR, and both methods are better than PeEn with statistically significant differences. Looking at specific lead times, QR tends to lose accuracy when the solar elevation is low (i.e., early morning and late afternoon) when the regressors (i.e., the meteorological model forecasts) and the observed variables (i.e., power) are less correlated. The different climatology at the three solar farms affects the performances of QR and AnEn particularly in term of MAE normalized by mean power (MAE/MP). At Catania and Calabria MAE/MP total average values are under 20% while at Milano they are around 25%. There is not a significant impact on

predictability with respect to the nominal power. In fact, the Calabria nominal power (NP) is around one thousand times greater than Catania but the MAE/MP of AnEn and QR results are similar, at about 18–19% on both plants. NN shows slightly worse results than AnEn and QR during the central hours of each forecast day at both Catania and Calabria, while at Milano its performance is similar to AnEn. This points out the applicability and, in general, the appreciable performance of probabilistic forecast systems when used to deliver also a single-valued, short-term forecast, which appears comparable with that of a common deterministic forecasting system based on NN. The Brier skill score (BSS) is then used to test AnEn and QR skill against PeEn on predicting two events: PV power greater than the mean power and lower than 25% of the mean power, with the mean power computed separately for each lead time. The correct prediction of the second event is particularly important for transmission system operators (TSO) that, to maintain grid stability, must find other sources of power production to compensate for the lower solar power availability. Both QR and AnEn have a better BSS than PeEn during the central hours of the day, but they (in particular QR) lose skill in the early morning and late afternoon. The AnEn has a statistically significant higher total BSS than QR at Catania and Calabria plants, while at Milano they are statistically equivalent.

It is important to note that the AnEn algorithm described in this paper requires minimal computational resources. To process a single three day-ahead AnEn forecast, a fraction of a second is sufficient on a common personal computer CPU. The most substantial computational effort is carried out to obtain the deterministic meteorological forecast used to generate the time series input to AnEn.

The impact of different climatologies on AnEn performance is also assessed. However, future work should expand the tests presented here to other locations spanning a wider range of climatological conditions, to assess the generality of the results reported in this study.

Acknowledgements

This material is based upon work supported by the Department of Energy, under Award Number DOE E0006016. This work has been partly funded by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE (formerly known as ERSE) and the Italian Ministry of Economic Development – General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency stipulated on July 29, 2009 in compliance with the Decree of March 19, 2009. The authors acknowledge Umberto De Angelis and Giuseppe Tommaselli of Enel S.p.A. for providing meteorological and power data used in this study, Sue Ellen Haupt and Jared Lee of NCAR for useful comments on the manuscript, and Constantin Junk for useful suggestions on the AnEn optimization procedure. The COST Action “WIRE” is also acknowledged for supporting two short-term scientific missions of Stefano Alessandrini and Simone Sperati to NCAR, which have contributed to these results.

References

- [1] Early data on 2013 electricity demand: 317 billion kWh of demand, –3.4% compared to 2012 Terna (company) press release. <<http://www.terna.it/LinkClick.aspx?fileticket=CjQzJQkNXhM%3d&tabid=901&mid=154>>.
- [2] EWEA. Wind in power, European statistics; 2014. <http://www.ewea.org/fileadmin/files/library/publications/statistics/EWEA_Annual_Statistics_2013.pdf>.
- [3] Skoplaki E, Palyvos JA. On the temperature dependence of photovoltaic module electrical performance: a review of efficiency/power correlations. *Sol Energy* 2009;83(5):614–24.
- [4] Mahoney WP, Parks K, Wiener G, Liu Y, Myers WL, Sun J, et al. A wind power forecasting system to optimize grid integration. *IEEE Trans Sustain Energy* 2012;3(4):670–82.
- [5] Marquis M, Wilczak J, Ahlstrom M, Sharp J, Stern A, Smith JC, et al. Forecasting the wind to reach significant penetration levels of wind energy. *Bull Am Meteorol Soc* 2011;92(9):1159–71.
- [6] Brown BG, Katz RW, Murphy AH. Time series models to simulate and forecast wind speed and wind power. *J Clim Appl Meteorol* 1984;23(8):1184–95.
- [7] Jensenius J, Cotton GF. The development and testing of automated solar energy forecasts based on the model output statistics (MOS) technique. In: 1st Workshop on terrestrial solar resource forecasting and on use of satellites for terrestrial solar resource assessment. Washington (DC); 1981.
- [8] Doherty R, O'Malley M. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Trans Power Syst* 2005;20(2):587–95.
- [9] Roulston MS, Kaplan DT, Hardenberg J, Smith LA. Using medium-range weather forecasts to improve the value of wind energy production. *Renew Energy* 2003;28(4):585–602.
- [10] Pinson P, Madsen H. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy* 2009;12(2):137–55.
- [11] Alessandrini S, Davò F, Sperati S, Benini M, Delle Monache L. Comparison of the economic impact of different wind power forecast systems for producers. *Adv Sci Res* 2014;11(1):49–53.
- [12] Diagne HM, Lauret P, David M. Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids. In: WREF 2012–world renewable energy forum; 2012.
- [13] Pedro HT, Coimbra CF. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Sol Energy* 2012;86(7):2017–28.
- [14] Heinemann D, Lorenz E, Girodo M. Forecasting of solar radiation. *Solar energy resource management for electricity generation from local level to global scale*. New York: Nova Science Publishers; 2006.
- [15] Pelland S, Remund J, Kleissl J, Oozeki T, De Brabandere K. Photovoltaic and solar forecasting: state of the art. IEA PVPS, Task 2013;14.
- [16] Chen JL, Liu HB, Wu W, Xie DT. Estimation of monthly solar radiation from measured temperatures using support vector machines—a case study. *Renew Energy* 2011;36(1):413–20.
- [17] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Sol Energy* 2010;84(5):807–21.
- [18] Amrouche B, Le Pivert X. Artificial neural network based daily local forecasting for global solar radiation. *Appl Energy* 2014;130:333–41.
- [19] Bacher P, Madsen H, Nielsen HA. Online short-term solar power forecasting. *Sol Energy* 2009;83(10):1772–83.
- [20] Alessandrini S, Sperati S, Pinson P. A comparison between the ECMWF and COSMO ensemble prediction systems applied to short-term wind power forecasting on real data. *Appl Energy* 2013;107:271–80.
- [21] Wang J, Botterud A, Bessa R, Keko H, Carvalho L, Issicaba D, et al. Wind power forecasting uncertainty and unit commitment. *Appl Energy* 2011;88(11):4014–23.
- [22] Kou P, Gao F, Guan X. Sparse online warped Gaussian process for wind power probabilistic forecasting. *Appl Energy* 2013;108:410–28.
- [23] Pinson P, Nielsen HA, Madsen H, Kariniotakis G. Skill forecasting from ensemble predictions of wind power. *Appl Energy* 2009;86(7):1326–34.
- [24] Iversen EB, Morales JM, Møller JK, Madsen H. Probabilistic forecasts of solar irradiance using stochastic differential equations. *Environmetrics* 2014;25(3):152–64.
- [25] Mathiesen P, Brown JM, Kleissl J. Geostrophic wind dependent probabilistic irradiance forecasts for coastal California. *IEEE Trans Sustain Energy* 2013;4(2):510–8.
- [26] Bracale A, Caramia P, Carpinelli G, Di Fazio AR, Ferruzzi G. A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control. *Energies* 2013;6(2):733–47.
- [27] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE J Select Top Appl Earth Observat Remote Sens* 2009;2(1):2–10.
- [28] Delle Monache L, Eckel FA, Rife DL, Nagarajan B, Searight K. Probabilistic weather prediction with an analog ensemble. *Mon Weather Rev* 2013;141(10):3498–516.
- [29] Delle Monache L, Nipen T, Liu Y, Roux G, Stull R. Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon Weather Rev* 2011;139(11):3554–70.
- [30] Alessandrini S, Delle Monache L, Sperati S, Nissen JN. A novel application of an analog ensemble for short-term wind power forecasting. *Renew Energy* 2015;76:768–81.
- [31] Junk C, Delle Monache L, Alessandrini S, Cervone G, von Bremen L. Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Meteorologische Z* 2014.
- [32] Vanvyve E, Delle Monache L, Monaghan AJ, Pinto JO. Wind resource estimates with an analog ensemble approach. *Renew Energy* 2015;74:761–73.
- [33] Bremnes JB. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy* 2006;9(1–2):3–11.
- [34] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy* 2006;9(1–2):95–108.
- [35] Servizio meteorologico dell'Aeronautica Militare, sezione atlante climatico. <<http://clima.meteoam.it/atlanteClimatico.php>>.

- [36] Pielke RA, Cotton WR, Walko REA, Tremback CJ, Lyons WA, Grasso LD. A comprehensive meteorological modeling system—RAMS. *Meteorol Atmos Phys* 1992;49(1–4):69–91.
- [37] Harrington JY. The effects of radiative and microphysical processes on simulated warm and transition season Arctic stratus (Doctoral dissertation, Colorado State University); 1997.
- [38] Hastie TJ, Tibshirani RJ. *Generalized additive models*. London: Chapman and Hall; 1990.
- [39] R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <<http://www.R-project.org/>>.
- [40] Anderson JL. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J Clim* 1996;9(7):1518–30.
- [41] Hamill TM. Interpretation of rank histograms for verifying ensemble forecasts. *Mon Weather Rev* 2001;129(3):550–60.
- [42] Bröcker J, Smith LA. Increasing the reliability of reliability diagrams. *Weather Forecast* 2007;22(3):651–61.
- [43] Pinson P, McSharry P, Madsen H. Reliability diagrams for non-parametric density forecasts of continuous variables: accounting for serial correlation. *Q J R Meteorol Soc* 2010;136(646):77–90.
- [44] Van den Dool HM. A new look at weather forecasting through analogues. *Mon Weather Rev* 1989;117(10):2230–47.
- [45] Wang X, Bishop CH. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J Atmos Sci* 2003;60(9):1140–58.
- [46] Hopson TM. Assessing the ensemble spread-error relationship. *Mon Weather Rev* 2014;142(3):1125–42.
- [47] Carney M, Cunningham P. Evaluating density forecasting models. Trinity College Dublin, Department of Computer Science; 2006.
- [48] Hersbach H. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast* 2000;15(5):559–70.
- [49] Gneiting T, Raftery AE, Balabdaoui F, Westveld A. Verifying probabilistic forecasts: calibration and sharpness. In: Preprints, 17th Conf. on probability and statistics in the atmospheric sciences, vol. 2. Seattle, WA: Amer Meteor Soc; January 2004.
- [50] Toth Z, Zhu Y, Hagedorn R. Ensemble forecasts and their verification.
- [51] Murphy AH. A new vector partition of the probability score. *J Appl Meteorol* 1973;12(4):595–600.
- [52] Stephenson DB, Coelho CA, Jolliffe IT. Two extra components in the Brier score decomposition. *Weather Forecast* 2008;23(4):752–7.
- [53] Madsen H, Pinson P, Kariniotakis G, Nielsen HA, Nielsen TS. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Eng* 2005;29(6):475–89.
- [54] Gneiting T. Making and evaluating point forecasts. *J Am Stat Assoc* 2011;106(494):746–62.