Post-processing Techniques and Principal Component Analysis for Regional Wind Power and Solar Irradiance Forecasting

Federica Davò^{*1,2}, Stefano Alessandrini³, Simone Sperati², Luca Delle Monache³, Davide Airolidi² and Maria T. Vespucci¹

[1] Università degli studi di Bergamo, Italy

[2] RSE, Ricerca Sistema Energetico, Milano, Italy

[3] NCAR, National Centre for Atmospheric Research, Boulder, CO, USA

*Correspondence to: F. Davò (federica.davo@unibg.it)

Abstract—This work explores a Principal Component Analysis (PCA) in combination with two post-processing techniques for the prediction of wind power produced over Sicily, and of solar irradiance produced over the Oklahoma Mesonet. For wind power, the study is conducted over a 2-year long period, with hourly data of the aggregated wind power output of the island. The 0-72 hour wind predictions are generated with the limited-area Regional Atmospheric Model System (RAMS), with boundary conditions provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic forecast. For solar irradiance, we consider daily data of the aggregated solar radiation energy output (based on the Kaggle competition dataset) over an 8-year long period. Numerical Weather Prediction data for the contest come from the National Oceanic & Atmospheric Administration – Earth System Research Laboratory (NOAA/ESRL) Global Ensemble Forecast System (GEFS) Reforecast Version 2. The PCA is applied to reduce the datasets dimension. A Neural Network (NN) and an Analog Ensemble (AnEn) post-processing are then applied on the PCA output to obtain the final forecasts. The study shows that combining PCA with these post-processing techniques leads to better results when compared to the implementation without the PCA reduction.

Index Terms—analog ensemble, neural network, principal component analysis, solar irradiance, wind power, forecasting

1. Introduction

In recent years, the development of wind and solar energy combined with their large-scale integration is creating a growing interest for predictions of the overall power production over large areas. In particular, Transmission System Operators (TSO) can benefit from such predictions to satisfy different purposes (e.g., overloads management or reserve estimation). The energy market operators need accurate predictions to reduce penalties, which are usually proportional to the unbalances, defined as the

difference between predicted and observed power. Individual electricity producers are also interested in accurate predictions. Some of them have now penalties when they produce less than forecasted or have to manage storage in batteries to ensure a smoother export to the grid as part of their contract. Predicting renewable energy at large scale means aggregating the forecasts of the individual power plants located in a geographical area of interest. Such areas can be defined by a TSO, or they can correspond to energy market regions.

The purpose of this study is to test a forecasting approach for large areas with two different test cases. First, the approach is tested on all the wind farms located over Sicily (one of the Italian energy market regions). The hourly data of wind speed are available over 2 years (2011 and 2012). The installed capacity (IC) during this period is 1746 MW. The second test includes 98 Oklahoma Mesonet solar monitoring stations. In this case, the period considered is from January 2000 to December 2007, and the dataset is composed of the daily data of the aggregated solar radiation energy output.

Large-scale wind power and solar irradiance forecasts have been the subject of several studies [1][2]. The most straightforward method to predict the power generated over an entire area is to sum the individual forecasts of each power plant. To do that, the locations and characteristics of each plant must be known. In this paper, an alternative method is proposed, which only requires historical time-series of power measurements produced by all the plants located within the area. In this study a Principal Component Analysis (PCA) technique is applied. PCA [3][4][5] is widely used in multivariate statistics and allows reducing the dataset size, retaining only the most relevant information. PCA is applied on the correlation matrix of the original forecasted variables over all the grid points inside the area, by computing its eigenvalues and eigenvectors. The eigenvalues allow establishing the optimal number of principal components, whereas the corresponding eigenvectors are used to transform the data, reducing the number of dimensions. A similar methodology was implemented to evaluate the performance of an analogous model on day-ahead forecasting of wind power production over large European regions in [6].

A Neural Network (NN) and an Analog Ensemble (AnEn) are applied on the PCA-transformed forecasts to obtain deterministic and probabilistic predictions of total wind power or total solar radiation energy. Probabilistic forecasting is useful to provide knowledge about uncertainty (e.g., a forecast of the PDF of power) in addition to a single-valued power prediction. For example, probabilistic predictions can be used by the day-ahead electricity markets for trading future energy production, obtaining higher results than those obtained by using deterministic forecasts alone, as shown in [7][8]. Indeed, by using probabilistic forecasting, it is possible to optimize the revenue for a producer in an economic model depending for example on the specific penalties for forecast errors valid in that market. The proposed method, AnEn [9][10], produces probabilistic power predictions from historical deterministic predictions and observations of the quantity to be predicted, as in [11][12][13][14][15]. It is possible to obtain a probabilistic power prediction also from Numerical Weather Prediction (NWP) ensembles [16][17][18]. The AnEn method was originally proposed in [19] for probabilistic meteorological forecasting. It

searches for the past forecasts most similar to the current one and keeps the correspondent past observations to form an ensemble. From any probabilistic prediction, it is possible to obtain a single-valued deterministic prediction by computing the mean or the median of the n analog members.

In literature [20][21] it is possible to find different methods for deterministic power forecasting, that consists in a single value for each time in the future for the variable to be predicted. Reviews of the state-of-the-art are reported in[22][23]. A deterministic forecast can be generated with an artificial NN [25][26][26]. Given one or more inputs, e.g., the principal components, the NN produces a response variable (which in the present application consists in wind power or solar irradiance) through neural interconnections. The NN is initialized with numeric weights and its neurons can be trained in order to make them able to learn how to associate to each input the correspondent output. In this context, a single-layer feed-forward NN with a back-propagation algorithm is used.

The paper is organized as follows: section 2 reports the observational and meteorological forecast data of Sicily and Oklahoma. The PCA technique is illustrated in section 3, within the post-processing methods AnEn and NN. Section 4 thoroughly describes the verification framework. Results of the application are given in section 5 and conclusions are illustrated in section 6.

2. Observational and meteorological data description

Hourly wind speed forecasts are generated over the area of Sicily using the Regional Atmospheric Modeling System (RAMS)[27]limited-area meteorological model. Initial and boundary conditions are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic model[28].

The gridded wind field is obtained by interpolating data from the closest vertical model levels at 80 m a.g.l. which is the average hub height in Sicily. The whole computational domain is made of 900 grid points with 20 km horizontal grid increments. The RAMS simulation covers the period January 2011-December 2012, for which the hourly data of aggregated wind power produced in Sicily are made available by Terna, the Italian Transmission Operator (TSO) [29]. The average installed capacity (IC) during this period is equal to 1746 MW. We have applied a correction factor to account for an increasing IC during the overall period.

The first year of the dataset (2011) is used as a training period, whereas the second year (2012) is kept as an independent verification period. The training period is split into two sub-periods, of six months each; the first is used to choose the optimal number of principal components. We test this choice over the second part of the training dataset. Each forecast run is initialized

at 12 UTC and covers the period from 0 to 72 hours ahead. We use a Geographic Information System (GIS) [30] to select from the whole domain only the 89 grid points within the territory of Sicily, as shown in Fig. 1.

Fig. 1. Grid points by the RAMS model computational domain over Sicily (Italy), used for wind power predictions.

The solar dataset was provided for a Kaggle competition [31]. Input NWP data for the contest are retrieved from the National Oceanic & Atmospheric Administration – Earth System Research Laboratory (NOAA/ESRL) Global Ensemble Forecast System (GEFS) Reforecast Version 2 [32]. Data include 11 ensemble members with three-hour time resolution from 12 to 24 hours ahead. Each forecast run is initialized at 00 UTC. Locations of the Mesonet sites relative to the GEFS data ($1^\circ \times 1^\circ$ of spatial resolution) are shown in Fig. 2. The total daily incoming solar radiation energy per square meter (referred as energy density, $J \cdot m^{-2}$) measured at 98 Oklahoma Mesonet sites is available since 1-January 1994. Solar radiation energy was measured by a pyranometer at each Mesonet site every five minutes and integrated from sunrise to 23:55 UTC of each day. For this work, only data of the period January 2000-December 2007 are considered. The maximum observed radiation energy density (MED) is equal to $22 \times 10^8 J \cdot m^{-2}$ in winter, $29 \times 10^8 J \cdot m^{-2}$ in spring and summer, and $21 \times 10^8 J \cdot m^{-2}$ in autumn. In this study we use only forecasts of the downward short-wave irradiative average flux at the surface cumulated on a daily basis ($J \cdot m^{-2}$) to be consistent with the observations. These data are provided on 9×16 points grid covering Oklahoma, for a total of 144 points. Homogeneity across the different seasons is obtained by normalizing solar data on a daily basis. The normalization has been carried out by subtracting the mean and dividing by the standard deviation computed over the eight values corresponding to the same Julian day of the eight years in the dataset. This is performed for both forecasted and measured data. The dataset is split into two parts of four years each: a training period (2000-2003) and a test period (2004-2007). The training period is split into two sub-periods of two years each with the same purpose as for the wind case.

Fig. 2. NOAA/ESRL Global Ensemble Forecast System (GEFS) Reforecast Version 2 (big grey points) and Mesonet stations (small black points, see text for additional details) used for solar irradiance prediction.

3. Prediction methods

3.1 Principal Component Analysis (PCA)

PCA is a variable reduction procedure useful in case of datasets with a large number of correlated information. Principal components are linear combinations of observed variables, ordered according to a criterion of relevant information which is expressed by their variance: the first component extracts the maximum amount of information from the data and the other ones optimize the leftover information under the constraint of non-correlation with the other components. The primary purpose of this technique is to reduce the number of variables. This is achieved through a linear transformation of the m observed variables that are projected by the eigenvectors of the correlation matrix on a Cartesian system. The eigenvectors are sorted by descending order with respect to the amount of information; therefore, the variable with greater variance is projected on the first axis, the second on the second axis, and so on. To reduce the dataset dimension, the first $p \ll m$ components are taken.

To perform a PCA on a dataset, it is necessary to normalize the data by their mean and standard deviation in order to have more homogeneous distributions. Moreover, PCA constructs orthogonal, mutually uncorrelated, linear combinations that explain as much common variation as possible. PCA can be performed on both the covariance and the correlation matrix. Scaling the data matrix such that all variables have zero mean and unit variance (i.e., normalizing them) makes the two approaches identical. Let X be a $n \times m$ matrix of the normalized data, where n is the number of total forecast events used for the study, and m is the

number of different grid points where the forecasts are available. In these applications $m = 89$ in the case of Sicily and $m = 144$ for Oklahoma.

The second step is to compute the eigenvalues and eigenvectors of the correlation matrix. Let Γ denote the eigenvectors matrix. It is now possible to transform the original variables into new variables without information redundancy (uncorrelated) as follows:

$$X_{pca} = \Gamma * X \quad (1)$$

After this operation, the original data are transformed in new uncorrelated data, without reducing the dimensionality. In order to reduce the data to a p -dimensional space, the original matrix X is multiplied by only the first p eigenvectors (called *loadings*). Each column of the resulting matrix, X_{pca} is a principal component, also known as *score*.

In the literature three methods [3] have been introduced to choose the dimension p :

- 1) Cumulative percentage of total variation: only those components representing the 80%-90% of the total variance are retained, i.e.,

$$\frac{\lambda_1 + \dots + \lambda_p}{\lambda_1 + \dots + \lambda_m} \approx 80\% - 90\% \quad (2)$$

where λ_i is the i -th eigenvalues of the correlation matrix; the numerator represents the variance of the first p principal components, whereas the denominator represents the variance of all the main components.

- 2) The so-called *Kaiser rule* (proposed by Kaiser [33]) takes only those components having an eigenvalue greater than one.
- 3) Doing a ‘Scree Plot’ where the number of principal components corresponds to the change of slope in the Scree Plot.

Since the first and the third method depend on an arbitrary choice of the eigenvalues (i.e., there can be several choice of the number of eigenvalues), the second method is used in this work.

In this study, the PCA is applied to wind speed gridded forecast data for the wind case and to the predicted gridded downward short-wave irradiative average flux at the surface for the solar case.

Starting from the 89 dimensions corresponding to the number of forecast grid points considered in Sicily, the PCA shows that only the first 6 principal components have eigenvalues greater than one; a sensitivity analysis on the number of components over the second part of the training dataset confirms this choice. For the solar case, the optimal number of principal components is 17.

3.2 Analog Ensemble (AnEn)

The AnEn method builds an ensemble forecast whose members are made by past observations appropriately selected across a historical dataset. of deterministic meteorological forecasts. For these applications, the historical data-sets of forecasts are made

by gridded data, the output of meteorological models. For the wind case, wind speed data from 89 points of RAMS simulation are used (see Section 2 for more details). For the solar case, solar irradiance data from 144 points are provided by GEFS simulations.

Specifically, given a historical dataset of predictions and corresponding observations, for each lead-time of the current forecast, distances from every past forecast issued at the same lead-time are computed. All these distances are sorted from the lowest and the first n correspondent past forecasts are chosen. The concurrent n past measurements (wind power or solar radiation density in this work) are the n members that constitute the current analog ensemble forecast for the lead-time considered. It is important to underline that the AnEn, by selecting past power measurements which become the current forecast, naturally takes care of transforming meteorological predictions into power (or energy) predictions.

The metric used to assess past forecasts similarity to the current forecasts is defined as follows [10][19][34]:

$$\|F_t A_{t'}\| = \sum_{j=-\tau}^{j=\tau} \sqrt{\sum_{i=1}^p \lambda_i (F_{i,t+j} - A_{i,t'+j})^2} \quad (3)$$

where F_t is the current forecast at time t at a certain location, $A_{t'}$ is an analog forecast at time t' before F_t was issued and at the same location, λ_i is the i -th eigenvalue coming from the PCA, p is the dimension of the subspace in which the data are projected, τ is an integer equal to half width of the time window over which the metric is computed, 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. In this context, the forecast F and A are the arrays resulting from the PCA analysis. To assign different weights to each principal component, the eigenvalues λ_i are introduced in (3). Equation 3 differs from the metrics already used in [10] and [9]. In fact, in those works the forecasts F and A were not transformed by the PCA and represented arrays of different meteorological variables (predictors) extracted in one location. Also, λ_i were referred as the weights assigned to the different p meteorological predictors. In this work we just use one meteorological predictor (wind speed or solar irradiance) and p represent the number of principal components selected. When the AnEn is applied without the PCA dimension reduction (see section 3.4), p is the number of grid points (89 for the wind case, 144 for the solar case) and λ_i is kept constant equal to 1.

The AnEn provides probability density functions (PDF) from which the likelihood of future events can be estimated. The PDF of the power or energy density forecast is estimated using the set of the n members from the analog ensemble forecast. On the other hand, computing the mean or the median of the n analog members allows obtaining a deterministic forecast as well.

3.3 Neural Network (NN)

A NN links the input variables and one or more output variables through an interconnected group of nodes called neurons. The back-propagation algorithm used in this work is based on the programming language R [35], and allows minimizing an error function depending on the different weights associated to the different interconnections. Each neuron assigns a weight to the input signals, and through an activation function, it produces an output term. The learning procedure of each neuron is performed through different iterative cycles, each making up a training phase for the network. At each iteration, the input signals in are multiplied by an appropriate value w representing the signal weight, usually within the range $(-1,1)$. This operation produces the resultant $net_i = \sum_j w_{ij} in_j$. The term $net_i - \theta_j$, where θ_j is an arbitrary value, forms the activation function argument and generates the output. The NN input is usually normalized to values between 0 and 1, in order to adapt the activation function to the weight values.

The strength of the learning phase is evaluated by the difference between the forecast output signal and the measured value; this difference represents the error. It is not possible to define *a priori* the optimal number of neurons, but in literature it is possible to find different rules [36][37].

The advantage of using an NN is that different types of data and variables, e.g., lead-time, wind speed and wind direction, can be combined. Also, being the training carried out using observed power data, it allows the NN to perform the conversion into power after getting meteorological variables as input. An independent iterative training is performed starting from different random weights: the NN that provides the minimum Root Mean Square Error (RMSE, see section 4.1), compared to the observations over a subset of the training period, is selected and applied over the test period. In this work we have applied the NN using between 6 and 8 neuron depending on the different number of input data determined by each approach (see section 3.4). A sensitivity analysis (not shown) carried out on the training period has indicated the best NN configuration for each application.

3.4 Reference Methods

To better understand the PCA contribution to the overall forecast system's performance, we introduce a simpler dimension reduction technique as a baseline reference; we apply the AnEn and NN using the mean, the maximum, the minimum values and the standard deviation (MMMS) of the gridded forecast data as input. Also, we apply the AnEn and NN using the forecasts data from all the grid points as predictors. To summarize, we compare six approaches: NN after PCA (PCA+NN), AnEn after PCA (PCA+AnEn), NN after MMMS (MMMS+NN), AnEn after MMMS (MMMS+AnEn), and finally NN (NN) and AnEn (AnEn) without any dimension reduction.

We also want to understand the contribution of NN and AnEn as post-processing techniques to convert meteorological data into power data minimizing the BIAS. To do that, we compare the NN and AnEn after the PCA to a less sophisticated post processing method, i.e. a linear regression (PCA+LR).

In Table 1 the computational time of PCA+NN, NN, PCA+AnEn and AnEn for both wind and solar case are compared. The work has been implemented in R [35][35] and run on a single processor Intel® Core® i7-4510U, CPU 2.00 GHz.

Table 1. Computational time for four different methods in both wind and solar case.

	PCA+NN	NN	PCA+AnEn	AnEn
Wind	3' 41''	6h 39' 44''	1h 50' 6''	28h 4' 35''
Solar	1' 57''	18' 7''	20' 31''	3h 8' 52''

The improvement computational time reduction lead by PCA application is evident. When applied with the AnEn the reduction is of about 90% for both the wind and the solar case and when applied with the NN the reduction is of about 98% for the wind case and of about 88% for the solar case.

4. Forecast evaluation

4.1 Deterministic Forecast Evaluation

RMSE, Mean Absolute Error (MAE), bias and Pearson correlation indices are used to assess the performances of the different prediction systems. Bias is defined as the difference between the arithmetic mean of observed and forecasted values. It is obtained by the formula

$$bias = \frac{1}{N} \sum_{i=1}^N (f_i - o_i) \quad (4)$$

where f_i represents the forecasted data and o_i the measured one. The bias evaluates the systematic error and its optimal value is 0; a positive (negative) value indicates a tendency to overestimate (underestimate) the observations, respectively.

MAE is given by the formula

$$MAE = \frac{1}{N} \sum_{i=1}^N |f_i - o_i| \quad (5)$$

i.e., it is the average absolute error between the forecasts and the observations. Lower MAE values indicate greater forecast accuracy.

RMSE is the squared root of the residual variance: it is given by the formula

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2} \quad (6)$$

i.e., it is the quadratic average of the errors and it indicates how much the forecasted data are far from the observed data.

Similarly to MAE, a lower RMSE indicates greater accuracy.

The Pearson correlation index is a coefficient that assesses the linearity between the covariance of two random variables and the product of their standard deviation, based on the formula

$$cor = \frac{cov(o_i, f_i)}{\sqrt{sd(o_i)sd(f_i)}} \quad (7)$$

The range of this coefficient is [-1,1]. It evaluates how the forecast and the measurement temporal trend are similar. A correlation equal to 1 indicates a perfect forecast.

Bias, MAE and RMSE are normalized with respect to IC and MED.

4.2 Probabilistic forecast evaluation

The quality of probabilistic predictions can be evaluated by assessing several key attributes, some of which are described below.

4.2.1 Reliability and sharpness diagram

A reliability diagram allows visualizing the similarity between forecasts and observations in a probabilistic framework. The class intervals of forecast probabilities are plotted against the observed frequencies [38][39][40]. A probabilistic system is reliable if, for any given level of probability, the rate of occurrence of an event (i.e., its observed relative frequency) is equal to its forecasted probability. In other words, if an event is predicted with a certain probability, such event should occur in the same percentage of cases. A perfect probabilistic system would correspond to a reliability diagram with a 1:1 diagonal line.

Together with reliability it is also useful to show the sharpness of a probabilistic system, i.e., the ability to issue forecasts with both low and high probabilities [41]. By plotting the relative frequency (i.e., the number of cases) in each probability class interval, it is possible to assess how sharp the system is. Forecast systems with low sharpness exhibit a flat frequency line. For reliable forecasts sharpness leads to better resolution (i.e., predictive skill). Reliability diagrams, for each probabilistic system,

are produced using a threshold equal to 50% of IC in the case of wind forecast and equal to 50% of seasonal MED in the case of solar forecasts, i.e., for each season the threshold is equal to 50% of maximum energy density of that season. Therefore four different thresholds are used for solar forecasts.

4.2.2 Relative Operating Characteristic Skill Score

The relative operating characteristic (ROC) skill score (ROCSS) index is based on the ROC curve [42], which is particularly useful in decision-making. The decision to undertake actions about the occurrence of a particular event can depend on the forecasted probability of the same event exceeding a certain threshold. The ROC curve is produced by plotting the false alarms divided by the total number of non-occurred event against the hit rate (i.e., the correct forecasts divided by the total occurrences of the event). A lower curvature corresponds to a higher number of false rates, showing a lower ability of the forecast system to discriminate. The ROC (as well as the ROCSS) thus depends upon resolution and not upon reliability. ROCSS translates the ROC score into a standard skill score estimating the resolution of probabilistic predictions, for which a value equal to 1 denotes a perfect forecast and a ROCSS lower than 0 indicates a system that performs worse than climatological forecasts.

4.2.3 Continuous Ranked Probability Score

The continuous ranked probability score (CRPS) compares a full probability distribution with the observations, when they are both represented as cumulative distribution functions (CDFs) [43]. It is given by the formula

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{+\infty} \left(F_i^f(x) - F_i^o(x) \right)^2 dx \quad (8)$$

where $F_i^f(x)$ is the CDF of the probabilistic forecast for the i -th values and $F_i^o(x)$ is the CDF of the observations. The CRPS is a proper score and it is expressed in the same unit as the observed variable. A lower value of CRPS indicates better performance. The CRPS generalizes the mean absolute error [44] and therefore it provides a direct way of comparing various deterministic and probabilistic forecasts by using a single metric.

4.2.4 Rank histogram

Rank histogram is a diagnostic tool to evaluate the calibration and the consistency of an ensemble distribution, indicating how the ensemble spread represents the observation uncertainty [45]. The ensemble members are ranked to delineate ranges of the

predicted variable; in a perfect ensemble the probability of occurrence of the observation within each range is equal. A calibrated system should have a flat distribution, in which the verification analysis ranks approximately with the same frequency in each interval. However, in operational situations it is more frequently to observe different shapes of the distribution. When the distribution takes a U-form, it indicates an over-confident prediction: this means that most of the observations fall outside the extremes of the probability distribution, the ensemble members are not well calibrated and the spread is not sufficient to include all measurements. Conversely, an inverted U-shape indicates an under-confident estimate, in which the spread of the forecasts is excessively large: even in this case the members do not appear properly calibrated, since the observations fall mostly close to the center and not close enough to the extremes.

In the rank histograms, the horizontal line represents the frequency for a perfectly uniform distribution. Vertical confidence bars can be calculated with a quantile function for a binomial distribution, to show a range in which a deviation from the perfect distribution is still consistent with reliability. The bars delimit the 90% confident interval. 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 underdispersive (overdispersive) ensemble.

4.2.5 Binned Spread/Skill Diagram

A binned-spread/skill diagram assesses the relationship between the ensemble spread and the RMSE of the ensemble mean over small intervals (i.e., bins) of spread, instead of considering its overall average; it assesses the ability of a probabilistic prediction to quantify its uncertainty [46][47][48]. Each bin has the same number of data points, which results in bins of different width. If the observations and the ensemble members are samples from the same distribution, it can be shown that the ensemble spread and the RMSE of the ensemble mean should be equal. This requirement becomes a necessary condition if the ensemble members and the ensemble mean errors do not have a Gaussian distribution. In this case, a good level of consistency is reached if the ensemble spread and the RMSE match at all values, resulting in a trend lying on the plot 1:1 diagonal. In this case the ensemble system is able to forecast its own error.

5. Results

5.1 Wind power results

The different methods are first compared in a deterministic framework. Fig. 3 shows examples of the times series for PCA+NN, PCA+AnEn probabilistic forecasts, and the measured wind power. The PCA+AnEn quantile ranges are consistent with the wind power observations variability and, as expected, they increase for larger lead times.

Fig. 3. PCA+NN, PCA+AnEn and measured wind power times series for 2 January 2012 for the wind case. The shadings correspond to the 25–75 (lighter) and 10–90 (darker) quantiles.

Fig. 4 shows RMSE (a), MAE (b), BIAS (c) (all normalized by IC) and correlation (d). Only lead-times from 36 to 59 hours ahead (i.e., the day-ahead) are shown (considering that lead time 0 corresponds to 12 UTC). All the forecast methods's performances deteriorate for lead times between 48-59 h.

Fig. 4. RMSE (a), MAE (b), BIAS (c) and correlation (d) of PCA+NN, PCA+AnEn, MMMS+NN, MMMS+ AnEn, NN, AnEn and PCA+LR as a function of forecast lead time for 36-59h ahead for the wind case. The arrows to the left of each plot represent the bootstrap confidence intervals of each index obtained grouping data over the different lead times together.

This is a typical pattern caused by the presence of more variability in wind speed (hence less predictable conditions) due to stronger turbulent convection in the afternoon hours (i.e., lead times 48-59h). Using the PCA-reduction procedure instead of all the 89 grid points improve the results, in particular when the NN is applied. This is due to the high number of input that prevent the NN from working properly. In fact, with an increased number of degrees of freedom, it's more difficult to reach a

convergence during the NN training. The AnEn can work with a high number of predictors (the results of PCA+AnEn, MMMS+AnEn and AnEn are comparable), but the advantage of using a PCA relies in the significant computational time reduction (Table 1). When comparing the three different post-processing methods (NN, AnEn and LR), the simple LR performs worse than the NN and the AnEn. For the remaining part of the paper we have compared only the NN and AnEn methods applied on the PCA-reduced data (i.e. PCA+NN, PCA+AnEn, MMMS+NN and MMMS+AnEn).. Using PCA instead of MMMS allows achieving better performances. PCA+NN shows generally slightly better performance in terms of RMSE, BIAS and correlation, whereas PCA+AnEn shows better performance in terms of MAE. NN performs slightly better than AnEn in terms of RMSE and correlation but, as shown by the confidence intervals in proximity of the left vertical axis (computed with all the lead times pulled together), the difference is not statistically significant. AnEn shows lower BIAS than NN.

Regarding the probabilistic evaluation, Fig. 5 reports the reliability diagrams calculated using a threshold equal to 50% of IC corresponding to an observed frequency around 40% (see horizontal dashed line). The use of the PCA leads to a slightly more reliable forecast, in fact its forecast probabilities appear more similar to the observed frequencies in the range between 0.4 and 0.8. Regarding sharpness, both methods show similar results (i.e., the majority of cases occurs in the 0-20% and 90-100% range). The vertical bars in the diagrams are consistency intervals showing the range of empirical probabilities that could be observed even for perfect probabilistic forecast, due to sampling effects. They are derived from binomial distribution with the same probability as the corresponding class [49]. Bars are calculated using the 5% and 95% quantiles of this distribution. The horizontal dashed line is the climatological frequency (i.e., the mean of the observed frequencies).

Fig.5. Reliability (orange/green line) and sharpness (grey squares) diagrams for probabilistic forecasts of IC greater than 0.5, for PCA + AnEn (a) and MMMS + AnEn (b) for the wind case. The vertical bars indicate the potential range for perfectly reliable forecasts. The dashed horizontal line is the climatological frequency of the event.

Fig.6. ROCSS (a) as a function of forecast lead time (event forecasts greater than 0.5 IC) and CRPS (b) as a function of forecast lead time for the wind case. The orange line represents PCA + AnEn whereas the green one MMMS + AnEn. The arrows to the left of each plot represent the bootstrap confidence intervals of ROCSS and CRPS obtained by grouping data over the different lead-times together.

Fig. 6 (a) shows ROCSS values as a function of forecast lead time. The indices are computed using a threshold corresponding to 50% of IC. PCA+AnEn and MMMS+AnEn show a similar level of performance with values around 0.9 and always greater than 0.8 except for lead time 51-h. PCA+AnEn reaches higher values than MMMS+AnEn, in particular for lead time range 1-20. For all other lead times the two approaches are comparable.

In Fig. 6 (b) CRPS as a function of forecast lead-time is plotted. PCA+AnEn shows lower CRPS values for all of the lead times, indicating better performance. A daily cycle is also clear, with lower performances in early afternoon hours, when lower ROCSS and higher CRPS values are observed.

Fig. 7 shows the rank histogram produced for PCA+AnEn (a) and MMMS+AnEn (b). The horizontal line in the diagrams represents the frequency of a perfectly uniform distribution. The vertical bars delimit the 90% confidence interval. Both methods appear slightly overdispersive with very similar rank histograms, with small MRE values and a good statistical consistency.

Fig. 7. Rank histogram, PCA + AnEn data (a), and MMMS + AnEn (b) for the wind case.

We perform an additional assessment of statistical consistency by compiling binned-spread/skill diagrams as in Fig. 8. The two different approaches show a good distribution of their values along the diagonal, even if they are both slightly overdispersive. The two approaches are both adequately able to represent forecast uncertainty, since their spread reflects the deterministic error variance.

Fig. 8. Binned ensemble spread versus standard deviation of the ensemble mean error for the probabilistic forecasts: PCA + AnEn (orange line) and MMMS + AnEn (green line) for the wind case. The bars delimit the 90% bootstrap confidence interval.

5.2 Solar irradiance results

Fig. 9 shows time series of forecasted solar irradiance for PCA+NN and PCA+AnEn and the observed values. The forecasted values look in good agreement with the observations.

Fig. 9. PCA+NN, PCA+AnEn and measured solar irradiance times series for April 2004 for the solar case. The shadings correspond to the 25–75 (lighter) and 10–90 (darker) quantiles.

As shown in Fig. 10, for deterministic predictions, a reduction procedure (PCA or MMMS) leads to better results than the use of the entire dataset, in particular for the NN. The AnEn also performs worse when applied to all the 144 grid points exhibiting worse BIAS values. Concerning the post-processing methods, the NN and the AnEn perform better than the LR. As for the wind case, in the remaining part of the paper we have tested the PCA+NN and the PCA+AnEn only. The PCA approach leads to better results than MMMS in terms of RMSE (a), MAE (b), BIAS (c) (normalized by the maximum energy density) and correlation (d). The trends are similar for all four cases, showing good results during winter and autumn seasons and lower performances during spring and summer. Indeed, prediction errors are proportional to the solar radiation energy that is lower during winter[34]. NN leads to slightly better results than AnEn, both coupled to PCA and MMMS. RMSE and MAE follow the same trend, showing higher performance for NN, in particular for PCA+NN. BIAS shows a tendency to underestimate the forecasts in spring summer and autumn. Looking at the bootstrap intervals (in proximity of the left vertical axis computed with all the lead times pulled together), the improvement in correlation on PCA+NN and PCA+AnEn is closer to be statistically significant than for other metrics.

Fig.10. RMSE (a), MAE (b), BIAS (c) and correlation (d) of PCA+NN, PCA+AnEn, MMMS+NN, MMMS+ AnEn, NN, AnEn, PCA+LR as a function of the seasons for the solar case. The arrows to the left of each plot represent the total value of each index computed on the entire dataset with the bootstrap confidence interval

In Fig. 11 PCA+AnEn (a) and MMMS+AnEn (b) are compared by plotting reliability diagrams, where the threshold is set to a value equal to 50% of MED. The two models show the same level of reliability and sharpness. The horizontal dashed line is the climatological frequency around 55%.

Fig. 11. Reliability (orange/green line) and sharpness (grey squares) diagrams (event forecast greater than 0.5 MED), for PCA + AnEn (a) and MMMS + AnEn (b) for the solar case. The vertical bars indicate the potential range for perfectly reliable forecasts. The dashed horizontal line is the climatological frequency of the event.

Fig. 12 shows ROCSS (a) and CRPS (b) diagrams, where the corresponding values are reported as a function of the season. In the ROCSS diagram the indices are computed for a threshold corresponding to 50% of MED. PCA+AnEn shows a level of performance with values always greater than 0.9, whereas MMMS+AnEn shows lower values, with a minimum around 0.7 in summer. The bootstrap intervals indicate that the differences are statistically significant only during summer. This means that PCA+AnEn performs statistically significantly better than MMMS+AnEn only during summer. PCA+AnEn model shows lower CRPS values than MMMS+AnEn during all the seasons. The rank histograms and binned spread-skill diagrams (not shown) for both PCA+AnEn and MMMS+AnEn show an optimal and similar level of statistical consistency.

Fig. 12. ROCSS (a) as a function of seasons for the probabilistic forecasts of forecasted energy density greater than 0.5 MP and CRPS (b) a function of seasons for the solar case. The orange line represents PCA + AnEn whereas the green one MMMS + AnEn. The arrows to the left of each plot represent the total value of each index computed on the entire dataset with the bootstrap confidence interval.

6. Conclusion

We test different approaches to predict the hourly/daily wind power or solar irradiance produced by several plants spread over a vast geographic region on two separate datasets: Sicily (South of Italy) for wind power and Oklahoma for solar irradiance.

These approaches are based on using gridded forecast data from Numerical Weather Prediction (NWP) models over the geographic region of interest after a Principal Component Analysis (PCA). We use the PCA to reduce the size of the forecast gridded data. After the PCA, starting from the dimensionally reduced datasets, we apply the Analog Ensemble (AnEn) technique and a Neural Network (NN) algorithm to generate deterministic and probabilistic predictions of the regional power output,

In both the solar (Oklahoma) and wind power (Sicily) applications, the PCA-based reduction procedure leads to better results than those obtained by applying the NN and AnEn directly over the entire gridded prediction data. The improvements rely both on the forecast accuracy (lower Root Mean Squared Error, RMSE) and on consistent computational time reduction (between 88% and 98% depending on the application). Also, AnEn and NN both outperform a simpler bias reduction method as a linear regression.

We also compared the PCA with a less sophisticated dimension reduction procedure, as taking the minimum the maximum and the mean values of the gridded forecast data (MMMS)

For Sicily, PCA leads to better results than MMMS, particularly when coupled with NN. RMSE, MAE, BIAS, and correlation show that PCA performs better than MMMS, which leads to a lower performance possibly caused by a loss of information. Indeed, the selected principal components contain a large portion of the available information (described by the variance) of the original data. Thus, they are a valid proxy of the entire dataset.

For probabilistic wind power predictions over Sicily, the MMMS and PCA are almost similar, even though MMMS shows a tendency to overestimate the forecasts and to be less reliable than PCA. The benefit of using a PCA-based approach is more evident for the Oklahoma dataset. When NN is applied after PCA, it provides the most accurate deterministic forecast, particularly during summer. PCA+AnEn is more reliable than MMMS+AnEn when used to generate probabilistic solar power predictions. Also, lower values of CRPS and higher values of relative operating characteristic skill score (ROCSS) indicate a higher quality of the PCA-based probabilistic forecasts.

Acknowledgments

This work has been financed by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE S.p.A. and the Ministry of Economic Development-General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency in compliance with the Decree of 8 March 2006. This work has been partly funded by the U.S. Department of Energy's SunShot Initiative, under award DE-EE0006016.

References

- [1] R. Girard, K. Laquaine, G. Karioniotakis, "Assessment of wind power predictability as a decision factor in the investment phase of wind farms", *Appl. Energy* [Online], vol. 101, pp. 609-617, 2013. Available: <http://www.sciencedirect.com/science/article/pii/S0306261912005090>
- [2] P.-J. Trombe, P. Pinson, H. Madsen, "A general probabilistic forecasting framework for offshore wind power fluctuations", *Energies* [Online], vol. 5, pp. 621-657, 2012. Available: <http://www.mdpi.com/1996-1073/5/3/621>
- [3] I. T. Jolliffe, "Principal Component Analysis", Springer, New York, 1986.
- [4] S. Wold, "Principal component analysis", *Chem. and Intell. Lab. Systems* [Online], vol. 2, pp. 37-52, 1987. Available: http://imedeia.uib-sic.es/master/cambioglobal/Modulo_V_cod101615/Theory/lit_support/pca_wold.pdf

- [5] H. Abdi, L. J. Williams, "Principal component analysis", *WIREs Comput. Stat.* [Online], vol. 2, pp. 433-459, aug. 2010.
Available: <http://onlinelibrary.wiley.com/doi/10.1002/wics.101/abstract>
- [6] M. L. Martin, F. Valero, A. Pascual, J. Sanz, L. Frias, "Analysis of wind power production by means of an analog model", *Atmos. Research* [Online], vol. 143, pp. 238-249, 2014. Available:
<http://www.sciencedirect.com/science/article/pii/S0169809514001069>
- [7] M. S. Roulston, D. T. Kaplan, J. Hardenberg, L. A. Smith, "Using medium-range weather forecasts to improve the value of wind power production", *Renew. energy* [Online], vol. 28, pp. 585-602, 2002. Available:
http://people.maths.ox.ac.uk/lenny/wind2003_roulston.pdf
- [8] S. Alessandrini, F. Davò, S. Sperati, M. Benini, L. Delle Monache, "Comparison of the economic impact of different wind power forecast system for producers", *Adv. Sci. Res.* [Online], vol. 11, pp. 49-53, 2014. Available:
<http://dx.doi.org/10.5194/asr-11-49-2014>.
- [9] S. Alessandrini, L. Delle Monache, S. Sperati, J. N. Nissen, "A novel application of an analog ensemble for short-term wind power forecasting", *Renew. Energy* [Online], vol. 76, pp. 768-781, 2015. Available:
<http://www.sciencedirect.com/science/article/pii/S0960148114007915>
- [10] L. Delle Monache, T. Nipen, Y. Liu, G. Roux, R. Stull, "Kalman filter and analog schemes to post-process numerical weather predictions", *Mon. Wea. Rev.* [Online], vol. 139, pp. 3554-3570, 2011. Available:
<http://journals.ametsoc.org/doi/full/10.1175/2011MWR3653.1>
- [11] J. B. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts", *Wind Energy* [Online], vol. 9(1-2), pp. 3-11, 2006. Available: <http://onlinelibrary.wiley.com/doi/10.1002/we.182/full>
- [12] J. K. Møller, H. A. Nielsen, H. Madsen, "Time-adaptive quantile regression", *Comp. Stat. and Data Anal.* [Online], vol. 52, pp. 1292-1303, 2006. Available: <http://www.sciencedirect.com/science/article/pii/S0167947307002502>
- [13] H. A. Nielsen, H. Madsen, T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts", *Wind Energy* [Online], vol. 9(1-2), pp. 95-108, 2006. Available:
<http://onlinelibrary.wiley.com/doi/10.1002/we.180/full>
- [14] J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba et al., "Wind power forecasting uncertainty an unit commitment", *Appl. Energy* [Online], vol. 88, pp. 1014-1023, 2011. Available:
<http://www.sciencedirect.com/science/article/pii/S0306261911002339>
- [15] P. Pinson, "Estimation of the uncertainty in wind power forecasting", Ph.D. dissertation, Ecole des Mines de Paris, 2006.

- [16] S. Alessandrini, S. Sperati, P. Pinson, "A comparison between the ECMWF and COSMO Ensemble Prediction System applied to short-term wind power forecasting on real data", *Appl. Energy* [Online], vol. 107, pp. 271-280, 2013. Available: <http://www.sciencedirect.com/science/article/pii/S0306261913001499>
- [17] S. Alessandrini, P. Pinson, R. Hagedorn, G. Decimi, S. Sperati, "An application of ensemble/multi model approach for wind power production forecasting", *Adv. Sci. Res.* [Online], vol. 6, pp. 35-37, 2011. Available: <http://www.adv-sci-res.net/6/35/2011/asr-6-35-2011.html>
- [18] M. Lange, U. Focken, "Physical approach to short-term wind power forecast", Springer, ISBN 3-540-25662-8, 2005.
- [19] L. Delle Monache, A. Eckel, D. Rife, K. Searight, "Probabilistic weather predictions with an analog ensemble". *Mon. Wea. Rev* [Online], vol. 141, pp. 3498–3516, 2013. Available: <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00281.1>
- [20] F. Cassola, M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output", *Appl. Energy* [Online], vol. 99, pp. 154–166, 2012. Available: <http://www.sciencedirect.com/science/article/pii/S0306261912002747>
- [21] P. Pinson, H. A. Nielsen, H. Madsen, T. S. Nielsen, "Local linear regression with adaptive orthogonal fitting for the wind power application", *Statistics and Computing* [Online], vol. 18, pp. 59-71, 2008. Available: <http://link.springer.com/article/10.1007%2Fs11222-007-9038-7>
- [22] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, C. Draxl, "The state-of-the-art in short-term prediction of wind power: a literature overview", *ANEMOS plus*, 2011.
- [23] S. Sperati, S. Alessandrini, P. Pinson, G. Kariniotakis, The "Weather Intelligence for Renewable Energies" Benchmarking Exercises on Short-Term Forecasting n Wind and Solar Power Generation, *Energies*, 8(9), pp. 9594-9619, 2015.
- [24] U. E. Cali, B. Lange, J. Dobschinski, M. Kurt, C. Moehrlen, B. Ernst, "Artificial neural network based wind power forecasting using a multi-model approach", 7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms, Madrid, 2008.
- [25] B. D. Ripley, "Pattern recognition and neural networks" Cambridge, 1996.
- [26] L. Von Bremen, "Combination of deterministic and probabilistic meteorological models to enhance wind farm forecast", *J.Phys. Conf. Ser.* [Online], 75, 2007. Available: <http://iopscience.iop.org/article/10.1088/1742-6596/75/1/012050>
- [27] R. A. Pielke, W. R. Cotton, R. L. Walko et al., "A comprehensive meteorological modeling system – RAMS", *Met. and Atm. Physics* [Online], vol. 49, pp. 69-91, 1992.
- [28] European Centre for Medium-Range Weather Forecasts, ECMWF, <http://www.ecmwf.int/>.
- [29] Terna, <http://www.terna.it/default.aspx?tabid=1778>
- [30] QGIS, <http://www2.qgis.org/it/site/about/index.html>.

- [31] Kaggle competition, <https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest/data>
- [32] T. Hamill, G. Bates, J. Whitaker, D. Murray, M. Fiorino, T. Galarneau, "A description of the 2nd generation NOAA global ensemble reforecasts data set", *NOAA Earth System Research Lab*, Physical Sciences Division Boulder, Colorado, USA, 2015.
- [33] H. F. Kaiser, "The application of electronic computers to factor analysis", *Educ. Psychol. Meas.* [Online], vol. 20, pp. 141-151, 1960. Available: <http://epm.sagepub.com/content/20/1/141.extract>
- [34] S. Alessandrini, L. Delle Monache, S. Sperati, G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast", *Appl. Energy* [Online] vol. 157, pp. 95-110, Nov. 2015. Available: <http://www.sciencedirect.com/science/article/pii/S0306261915009368>
- [35] "R, A language and environment for statistical computing", *R Found. for Statis. Computing*, Vienna, Austria ISBN:3-900051-07-0, <http://www.R-project.org>.
- [36] Z. Boger, H. Guterman, "Knowledge extraction from artificial neural network models", *IEEE Systems, Man, and Cybernetics Conference* [Online], Orlando, FL, USA, 1997. Available: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=633051&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D633051
- [37] A. Blum, "Neural networks in C++", NY, Wiley, 1992.
- [38] I. T. Jolliffe, D. B. Stephenson, "Forecast verification: a practitioner's guide in atmospheric science", Wiley and Sons, Chichester, 2003.
- [39] D. S. Wilks, "Statistical methods in the atmospheric sciences", 2nd Ed. Academic Press, 2006.
- [40] P. Pinson, P. McSharry, H. Madsen, "Reliability diagrams for nonparametric density forecasts of continuous variables: accounting for serial correlation", *Quart. Journ. Royal, Meteor. Soc.* [Online], vol. 136, pp. 77-90, jan 2010. Available: <http://onlinelibrary.wiley.com/doi/10.1002/qj.559/abstract>
- [41] T. Gneiting, A. E. Raftery, F. Balabdaoui, A. Westveld, "Verifying probabilistic forecasts: calibration and sharpness", *Proc. Workshop on Ensemble 670 Weather Forecasting in the Short to Medium Range*, Val-Morin, QC, Canada, 2003.
- [42] I. B. Mason, "A model for the assessment of weather forecasts", *Aust. Meteorol. Mag.*, vol. 30, pp. 291-303, 1982.
- [43] E. P. Gritmit, T. Gneiting, V. Berrocal, N. A. Johnson, "The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification", Technical report 493, Department of statistics, university of Washington.
- [44] M. Carney, P. Cunningham, "Evaluating density forecasting models", Technical report, Trinity College Dublin, Depart. of comp. Scien., Dublin, 2012

- [45] T. Hamill, "Interpretation of rank histograms for verifying ensemble forecasts", *Mon. Wea. Rev.* [Online] vol. 129, pp. 550-560, 2001. Available:
[http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(2001\)129%3C0550%3AIORHFV%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(2001)129%3C0550%3AIORHFV%3E2.0.CO%3B2)
- [46] H. M. Van Den Dool, "A new look at weather forecasting through analogues", *Mon. Wea. Rev.* [Online] vol. 117, pp 2230-47, 1989. Available:
[http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(1989\)117%3C2230%3AANLAWF%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1989)117%3C2230%3AANLAWF%3E2.0.CO%3B2)
- [47] X. Wang, C. H. Bishop, "A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes", *J.Atmos. Sci.* [Online], vol. 60(9), pp. 1140–58, 2003. Available:
[http://journals.ametsoc.org/doi/abs/10.1175/15200469\(2003\)060%3C1140%3AACOB AE%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/15200469(2003)060%3C1140%3AACOB AE%3E2.0.CO%3B2)
- [48] T. M. Hopson, "Assessing the ensemble spread-error relationship", *Mon. Wea. Rev.* [Online], vol. 142(3), pp. 1125–42, 2014. Available: <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00111.1>
- [49] J. Bröcker, L. A. Smith, "Increasing the reliability of reliability diagrams", *Wea. Forec.* [Online], vol. 22, pp 651-661, 2007. Available: <http://journals.ametsoc.org/doi/pdf/10.1175/WAF993.1>