PO.ID 0111

AN APPLICATION OF PCA BASED APPROACH TO LARGE AREA WIND POWER FORECAST

F. Davò, S. Alessandrini, S. Sperati
RSE Spa (Ricerca Sistema Energetico, Milano, Italy)

Abstract

The purpose of this work is to predict the power produced by all the wind farms located over the entire area of Sicily island (one of the Italian market regions). The study has been conducted for a two-year period, considering hourly data of the aggregated wind power output of the island. For each day of the study and for the time horizons from 0 to 72 hours ahead, wind fields have been forecasted using a mesoscale meteorological model (RAMS) and ECMWF determinist forecast fields as boundary conditions. A Principal Component Analysis (PCA) has been applied on wind speed and wind direction data extracted at 50 m above ground on the model grid points inside the Sicily territory.

A Neural Network (NN) has been then used as post processing technique of the PCA output to obtain the final wind power forecast. The input of the NN, for every forecast lead time, are the final forecast data of wind speed and direction after the PCA application. The NN has been trained both with the PCA output and the power measurements on the first year of the analyzed period. For sake of comparison an alternative approach, applying NN directly on RAMS output (without PCA), has been adopted too. The study shows that the PCA introduction leads to better results in terms of RMSE, MAE and BIAS and to a lower computational time.

Site and wind data description

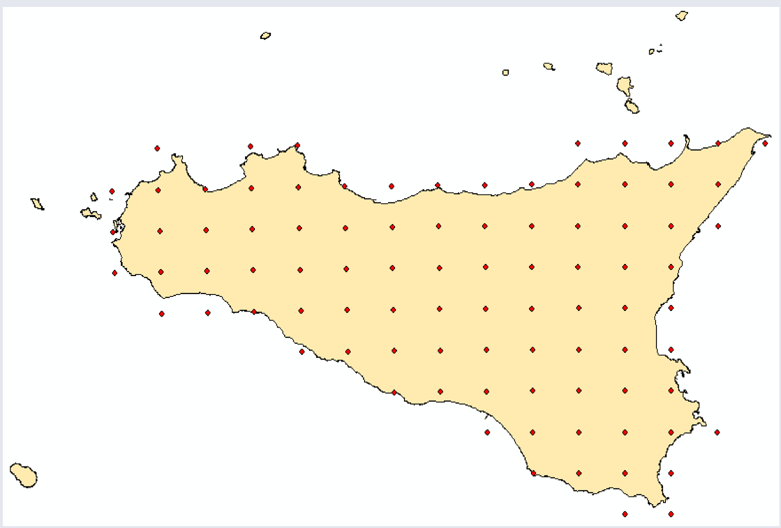

Figure 1. Points of the island considered for the power forecast.

The case study consists in all wind farms located in Sicily, whose total installed power (IP) is 1746 MW. Hourly measured power data were provided for the period January 2011-December 2012. Wind data have been forecasted using the model RAMS[1] (initialized with the boundary conditions of the ECMWF deterministic model) and then extracted at 80 m a.g.l., for 89 points inside the island, as shown in Figure 1 (the horizontal grid spacing is equal to 20 km).

Principal Component analysis

The Principal Component Analysis (PCA)[2,3] is a technique to reduce the dimensionality of a data set in which there is a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. Principal components are linear combinations of observed variables: the first component extracts from the data the maximum amount of information and the subsequent components optimize the remaining information under the constraint of non-correlation with the other components. The eigenvectors of the correlation matrix of input data determine the transformation matrix. In this context, the analysis has been performed only on wind speed, finding its principal components and using them also for direction. First, the correlation matrix and its eigenvalues and eigenvectors have been calculated. The criteria used for choosing the number of components are 3:

1. Kaiser rule: taking only those components that have an eigenvalue greater than one. In this application there are 6 eigenvalues greater than one.
2. Do a 'Scree Plot'[3]. The number of principal components is the number of components correspondent to the change of slope in the Scree plot (Figure 2).

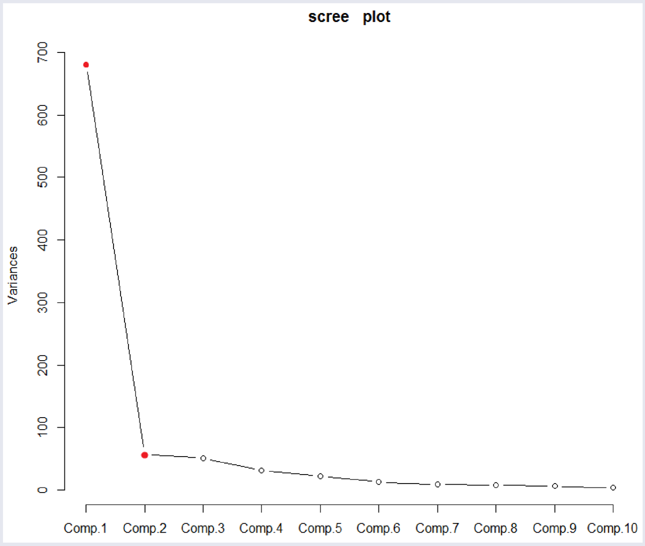

Figure 2. Scree Plot of the first ten components

3. Taking only those components that represent the 80%-90% of the total variance. The command "princomp()" of the statistic software R[4] returns Figure 3 (here only the first 20 components are shown).

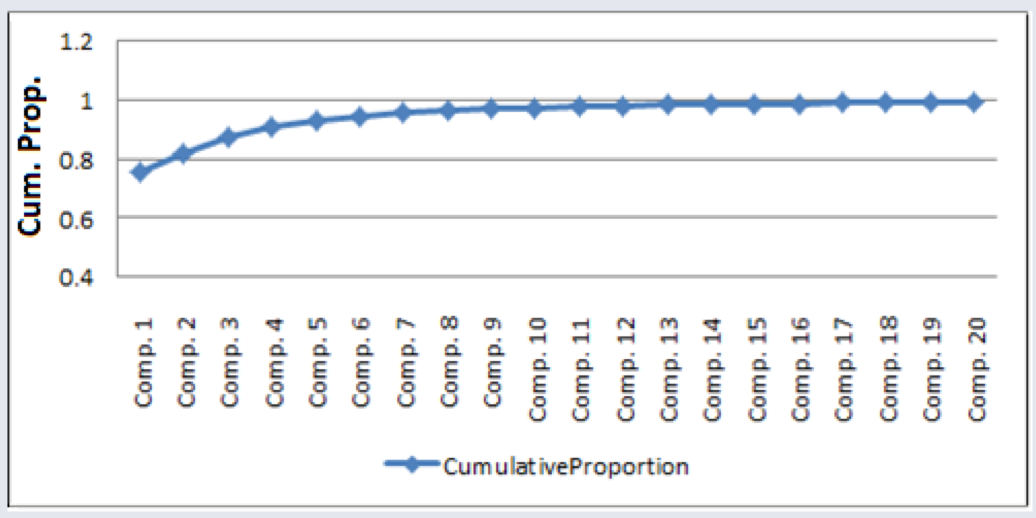

Figure 3. Cumulative proportion of the first twenty components.

After the analysis, it has been assessed that the optimal number N of main components should be between 3 and 6. It is now possible to obtain the transformed (final) data of wind speed and direction from the original forecast data of wind speed and direction as follows:

$$Final\ Data = (Eigenvectors)^T * (Initial\ Data)^T$$

Forecast evaluation

A Model Output Statistics (MOS) technique based on the use of a feed-forward NN[5] has been applied, using wind speed and direction found after the PCA as input and measured power as output. The NN has been trained on the year 2011 data and then applied on the year 2012 (test period) to test the use of different number of components. The results have been compared with an approach that applies NN directly on the RAMS output without doing the PCA (No PCA). Statistical indexes have been used as performance evaluation. The best performance is obtained with 4 principal components (4 P.C.). The use of PCA plus NN outperforms the forecast configuration with NN alone. After this first test, others NN configurations have been tested. A recursive approach (multiple NN training as new parts of the test period are added to the training data set) with 10 neurons results as the best one (4 P.C rec NN), as shown in the following figures.

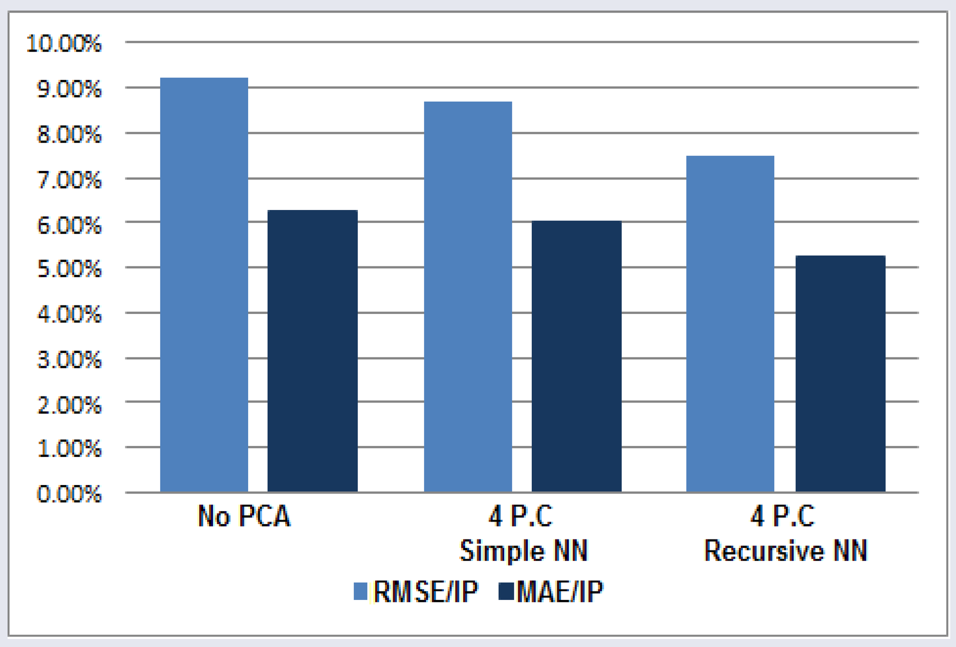

Figure 4. Histogram of the statistical indexes RMSE and MAE normalized by IP for the 3 forecast configurations.

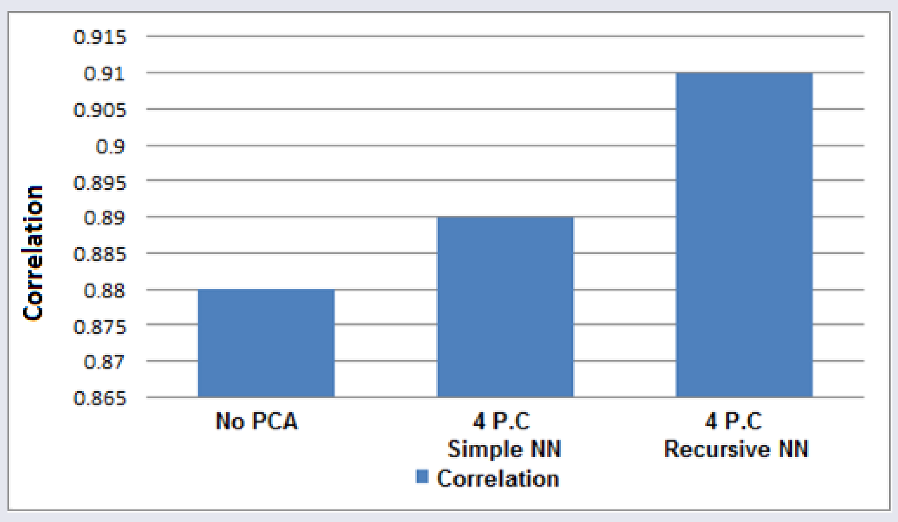

Figure 5. Histogram of the statistical index correlation for the 3 forecast configurations.

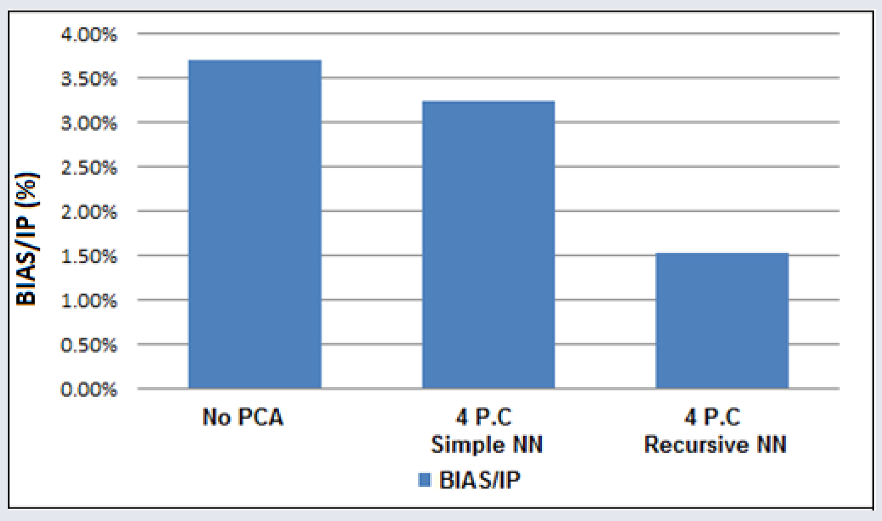

Figure 6. Histogram of the statistical index BIAS normalized by IP for the 3 forecast configurations.

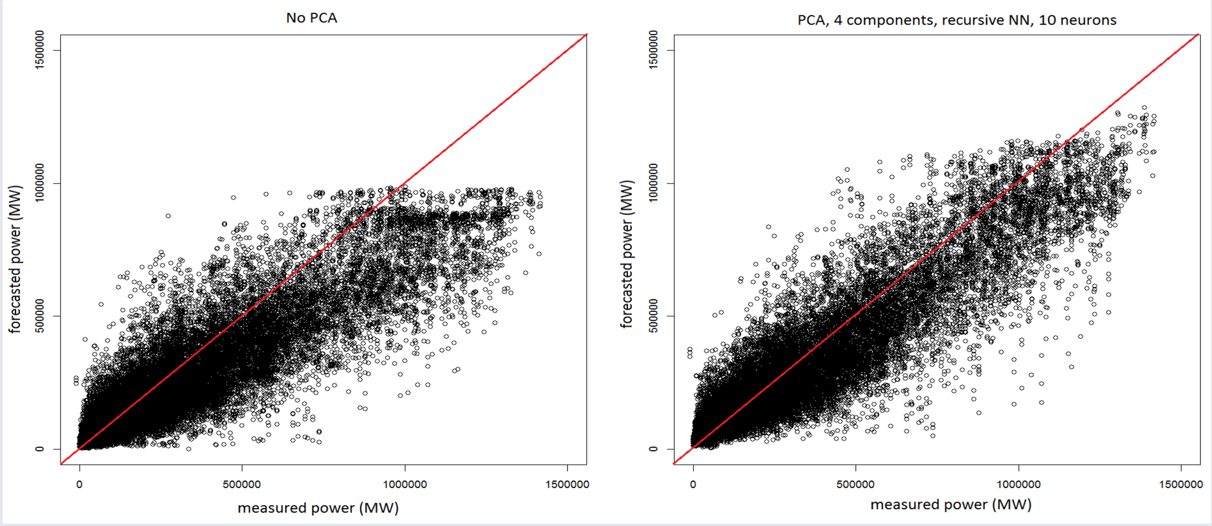

Figure 7. Scatter plot of measured series VS forecasted series.

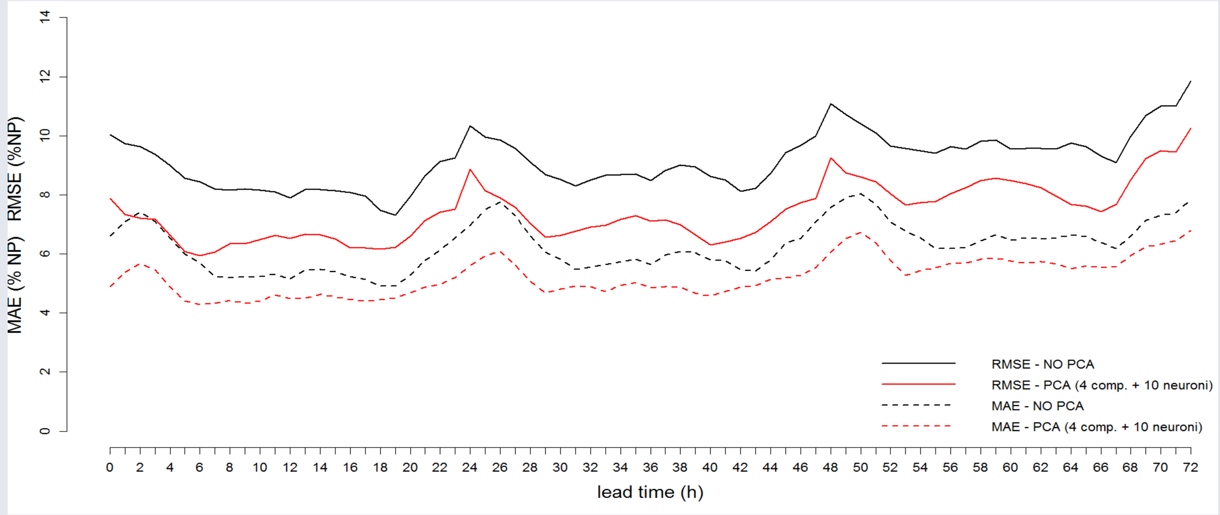

Figure 8. MAE and RMSE normalized by IP as a function of forecast lead time.

Conclusions

The use of Principal Component Analysis (PCA) together with Neural Network (NN) has demonstrated to be a useful way to forecast the total power produced by all wind farms on a large area (Sicily) with a total installed power of 1746 MW (IP). The PCA allows the dimension reduction of large data set (as those coming from a meteorological model grid) used as input to the NN. The main benefits from the PCA application are both a reduction of computational time and a simplification of the NN structure with less neurons. In terms of performance indexes the PCA+NN configuration outperforms the NN alone one.

References

[1] Pielke R. A., Cotton W. R., Walko R. L. et al., *A Comprehensive Meteorological Modeling System - RAMS*, Meteorology and Atmospheric Physics, 49, 69.
[2] L. I Smith. *A tutorial in Principal Components Analysis*, 2002.
[3] S.M. Holland. *Principal Components Analysis (PCA).* Department of Geology, University of Georgia, Athens, GA 30602-2501, May 2008.
[4] R Development Core Team R, *A language and environment for statistical computing*. http://www.R-project.org, ISBN 3-900051-07-0, 2006.
[5] S. Alessandrini, S. Sperati, P. Pinson. *The use of different ensemble forecasting systems for wind power prediction on a real case in the South of Italy*. EWEA 2012, Copenhagen.

Acknowledgements

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 March 8, 2006.