AIP Conference Proceedings

Wind farm power forecasting: New algorithms with simplified mathematical structure

Cite as: AIP Conference Proceedings **2191**, 020028 (2019); https://doi.org/10.1063/1.5138761
Published Online: 17 December 2019

Sebastian Brusca, Fabio Famoso, Rosario Lanzafame, Antonio Galvagno, Stefano Mauro, and Michele Messina

View Online

Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Vertical axis air turbine in oscillating water column systems
AIP Conference Proceedings **2191**, 020027 (2019); https://doi.org/10.1063/1.5138760

Biomass blend effect on energy production in a co-gasification-CHP system
AIP Conference Proceedings **2191**, 020082 (2019); https://doi.org/10.1063/1.5138815

Micro H-Darrieus wind turbines: CFD modeling and experimental validation
AIP Conference Proceedings **2191**, 020109 (2019); https://doi.org/10.1063/1.5138842

Meet the Next Generation of Quantum Analyzers
And Join the Launch Event on November 17th

Register now
Zurich Instruments

AIP Publishing

© 2019 Author(s).

Wind farm power forecasting: new algorithms with simplified mathematical structure

Sebastian Brusca[1], Fabio Famoso[2, a)], Rosario Lanzafame[2], Antonio Galvagno[1], Stefano Mauro[2], Michele Messina[2]

[1]*University of Messina, C.da Di Dio, Messina, 98166, Italy.*
[2]*Univeristiy of Catania, V.le A. Doria 6, Catania 95125, Italy.*
[a)]ffamoso@unict.it

Abstract. Nowadays wind power forecasting research is becoming more and more important in economic operation and safety management of power systems. Moreover, with a huge increase of data availability, data-driven models, especially for short-term forecast, represent a good compromise between precision and computational loads. This paper deals with a data-driven algorithm consisting of a combination of three main steps: gis-based data collection and pre-treatment, classification wind turbines step (Twostep cluster analysis) and short-term wind power forecast with artificial neural networks (ANN). The wind turbines were geo-referenced, the initial variables (single turbines power output, wind direction and speed) were pre-processed, a recursive optimized cluster analysis (CA) was performed to obtain a simplified mathematical structure for an artificial neural network. A windfarm (48 MW) placed in the eastern part of Sicily (Italy), was tested as case study. It consists of two different sites each one with 28 wind turbines at different rotor heights and different orography. The time series dataset consists of almost four years data (sampling time of 10 minutes). The consequent simplified mathematical structure leaded to perform good results in in a short-term wind power output forecast for both sites of the case study.

Kewords: gis, cluster analysis, short-term forecast, wind power.

INTRODUCTION

In last decades, the use of renewable energy systems and especially the wind power systems has considerably increased. In the last report of the World Wind Energy Agency (WWEA) it has been stated how the installed wind power capacity worldwide reached a total peak power of 597 GW with 50.1 GW added in 2018 [1]. The increase of the installed wind farms has encouraged the scientific community to better asses forecast methods for the wind energy production [2]. Traditional thermal generators, despite the fact are intermittent, are much more predictable than wind power generators. The primary energy source follows a deterministic process and the uncertainty grade in forecast models refers almost only about the random faults of the several system components. For these reasons these systems are usually described as “dispatchable” in contrast with wind generation systems described as “non-dispatchable”. Wind speed and direction are inherently variable and so it is difficult to evaluate wind power generation with traditional methods. The time horizon pays an important and crucial role in the model accuracy since forecast errors typically increase as the time range [3]. According to time range forecast methods may be divided in three main groups, each of them for different purposes. Short-term forecast methods usually range from 1 hours to 72 hours and are useful in power system planning for dispatch and for electricity trading where wind power or storage systems can be traded [4]. Medium-term methods range from 72 hours to one week and are usually performed to schedule plan or energy storage system maintenance [5]. Long-term methods range from weeks to months and are generally used to better evaluate wind farm overall energy production and investments payback [6]. Therefore, it appears evident how time horizon decision may affect the prediction models results. Current forecasting and prediction models may be also divided according to the model type itself. A first group refers to physical models. The physical models generally make use of global databases of meteorological measurements or atmospheric models, but they require large computational systems in order to achieve accurate results [7, 8].

74th ATI National Congress
AIP Conf. Proc. 2191, 020028-1–020028-9; https://doi.org/10.1063/1.5138761
Published by AIP Publishing. 978-0-7354-1938-4/$30.00

Computational fluid dynamics (CFD) is also used as an alternative method to the power law [9] or rotor performance [10]. In the statistical traditional methods, such as autoregressive moving mean models (ARMA) or autoregressive integrated moving mean models (ARIMA), a vast amount of data is analyzed and meteorological processes are not represented. They give good results in the estimation of mean monthly or even higher temporal scale wind speed [11]. Recently, the so-called learning approaches, or AI data-driven models, are considered a good compromise among the previous groups. They may be defined as "gray box" methods since they are able to explain physical phenomena such as wake effects [12, 13] or faults despite the fact they do not involve any physical model. Unfortunately, the latter methods are still not performing good results in near-real time given that the computation load increases as much as the time series and the number of observations [3]. The present work deals with an innovative method for short-term forecast wind power generation with a combination of GIS-based method, statistical analysis and neural network. The rest of this paper is organized as follows. Section 2 describes a wind farm located in the south of Italy treated as case of study. Section 3 illustrates the main steps of the developed model. Section 4 shows model results and Section 5 contains a brief discussion and conclusion.

CASE STUDY: A WIND FARM LOCATED IN THE SOUTH OF ITALY

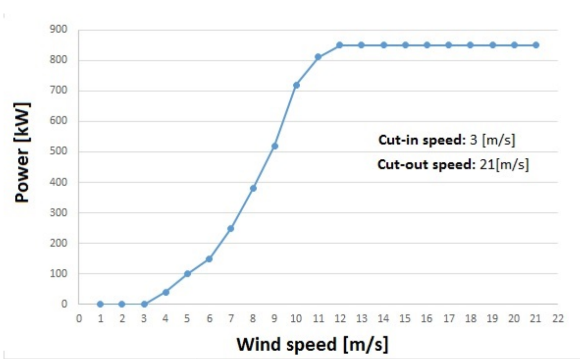
The proposed methodology illustrated in the next chapter has been tested for a real case study. A wind farm, as reported in Fig. 1, located in Sicily (south of Italy), was studied. This wind farm presents an overall P_{peak} of 48 MW consists of two main sites of 28 same wind turbines each one. The two sites are located in a mountain zone and present different quotes for each turbine, especially in site 1. Other main details are reported in Table 1.

FIGURE 1. The Wind farm position

TABLE 1. Main characteristics of the two sites.

	Site 1	**Site 2**
Position	37°58'32'' N, 15°07'00''E	37°56'47''N, 15°12'03''E
N° Turbines	28	28
Turbines quotes	732-1149 [m]	1038-1243 [m]
Turbine Height	52-58 [m]	52-58 [m]
N° blades	3	3
Turbine P_{peak}	850 [kW]	850 [kW]

The Turbine model presents a power curve with a very low Cut-in speed as reported in Fig. 2. The choice of this turbine model is probably due to the low wind velocities distribution of that geographical zone.

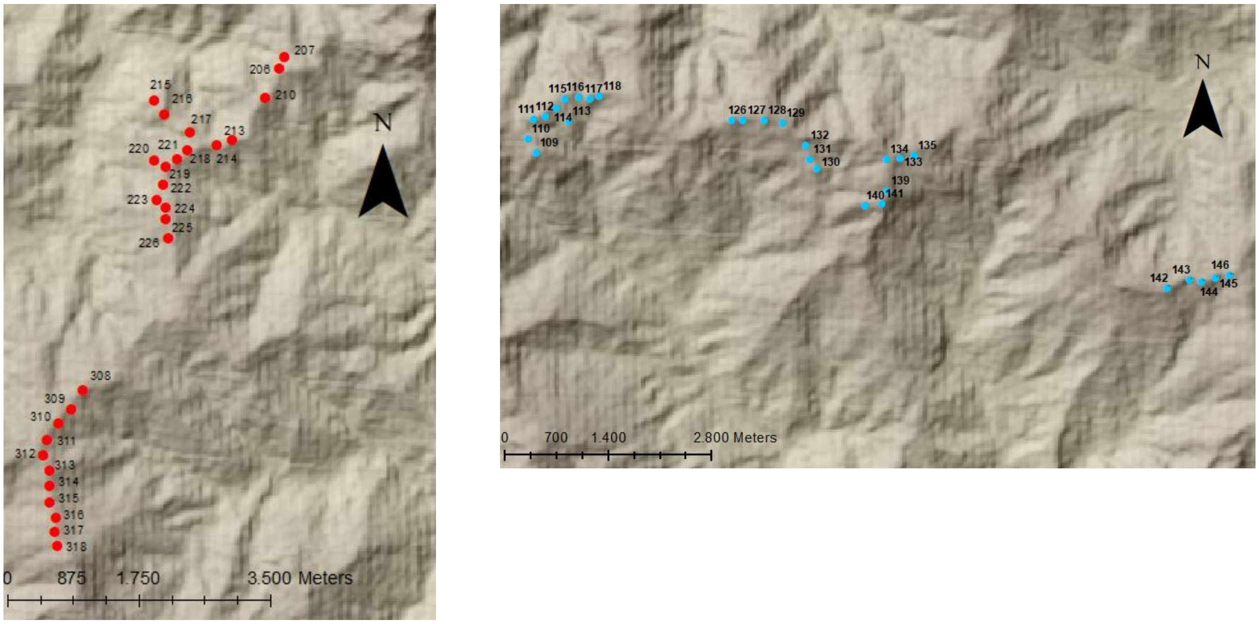


(a)

(b)

FIGURE 2. (a) The power curve of the turbine model for an air density of 1.225 kg/m^3. (b) A picture of some turbines in site 2.

Despite the sites are closed each other, they have been treated separately because of their different geographical distribution turbines. Fig. 3 shows the different distribution of each site. Site 1 presents two geographical clusters distribution orientated in the north direction, while site 2 presents three geographical clusters oriented to East.



(a) (b)

FIGURE 3. The location of the 28 turbines of site 1 (a) and site 2 (b).

The time series consists of data from 2015 to 2017. For each turbine (56 turbines) there were collected: Power peak [kW], Wind speed [m/s], Wind direction [°] with a time step of 10 minutes. In this work 2017 was deeply analyzed and the model was applied with a time range of a month in the two sites. For the lack of space in this paper the results refer to only one month (February 2017). In the following Table 2 the synthetic values are presented.

TABLE 2. Main data for site 1 and site 2 in February 2017.

Turbine	Medium P_{peak} [kW]	Max P_{peak} [kW]	Dev.std P_{peak} [kW]	Medium Wind speed[m/s]	Max Wind speed[m/s]	Dev.std Wind speed[m/s]
			Site 1			
210	163.75	850	199.61	5.67	23.71	3.27
220	75.89	828.73	146.39	4.07	16.21	2.38
310	351.51	843.11	194.18	7.71	21.45	3.48
316	144.21	848.55	198.83	5.85	20.35	3.53
			Site 2			
112	170.69	849.81	206.23	5.57	27.76	3.30
128	216.21	849.99	229.90	6.51	25.17	3.59
139	163.88	848.45	204.53	5.21	33.81	3.47
144	111.42	849.41	165.92	4.85	8.86	2.54

Wind direction frequencies are presented in Fig. 4. Despite the geographical turbines positions, this wind directions distribution reveal some possible wake effects (and the consequent power losses effect) occurred in the proximity of the turbine 316 for site 1 and in turbine 112 for site 2 .

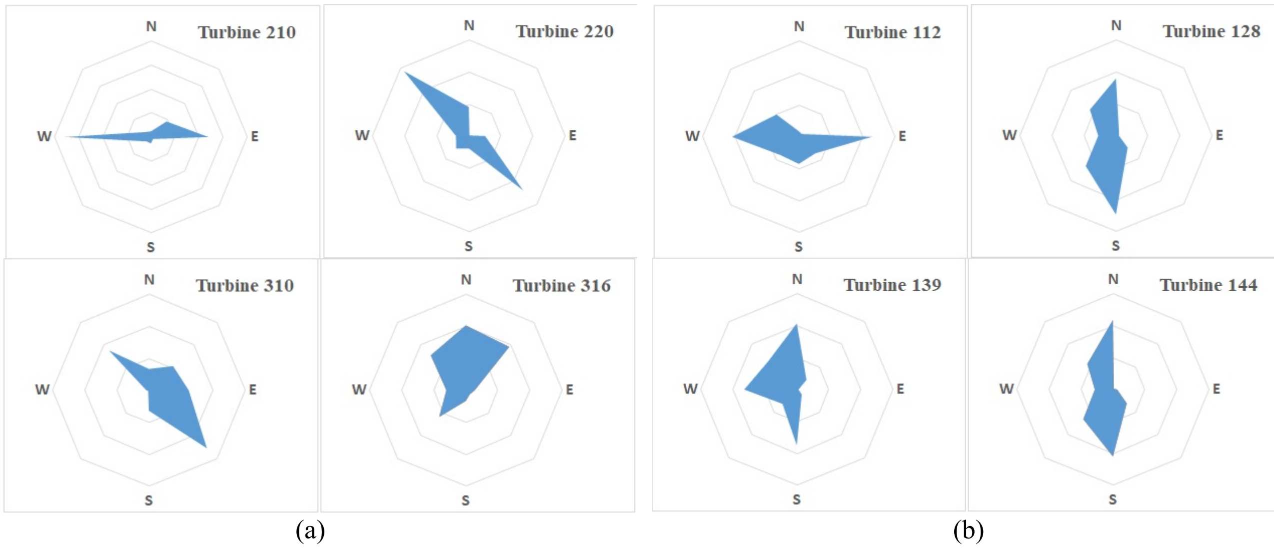


FIGURE 4. The distribution of the Wind directions in February 2017 for site 1 (a) and site 2 (b).

METHODS AND APPLICATION

In order to perform a short-term forecast overall power production in the two different sites, it was implemented a model consisting of three main steps as illustrated in Fig. 5.

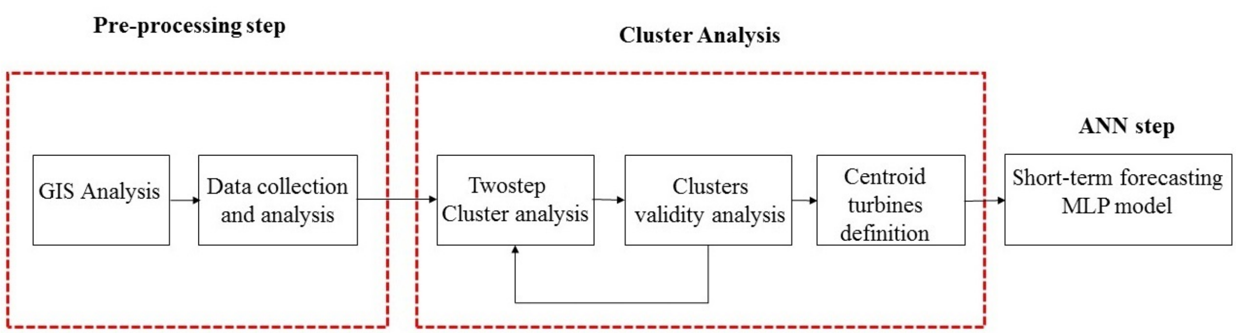


FIGURE 5. The main scheme of the implemented model.

In the following sub-chapters the three main steps are explained.

Pre-processing step

The first step consisted of data collection from all turbines (28 for each site), data pre-treatment and a preliminary data analysis. First of all, the wind turbines were georeferenced in a Gis-based system. In this way, all turbines were geographically clustered. The integration of spatial coordinates (x,y,z) to the presented methodology gave the opportunity to better assess the cluster analysis data results. A next step involved data collection from time series samplers. For each month, with a time step of 10 minutes, there were collected P_{peak} [kW], Wind speed [m/s] and Wind direction [°] for each turbine. In this paper it is presented just February 2017 (time series of 4032 samplers). Once collected, data was pretreated in order to handle missing values and wrong values (such as negative P_{peak}). A Python script was implemented in order to replace missing values with a moving mean calculation (by considering similar subsets), anyway missing values were about 1% of the total subset. For the considered variables (P_{peak}, Wind speed, Wind direction) there were calculates some preliminary statistical indices such as: Mean, Maximum, Minimum, median, dev. std. (σ). In Cluster Analysis there were used 6 variables for each turbine: mean P_{peak}, σ (P_{peak}),mean Wind speed, σ (Wind speed), mean Wind direction, σ (Wind direction) for each turbine. Since the high difference of scales variables, it was crucial for Cluster analysis to standardize them. It made easier to compare scores, even if those scores were measured on different scales. The main criteria is the follows:

$$Z = \frac{X-m}{\sigma} \tag{1}$$

Where Z is the standardize value, X is the variable non-standardize, m is the mean and σ is the standard deviation.

Cluster Analysis

Once obtained a matrix of 28 (turbines) x 6 (std.variables) elements, it was performed the Cluster Analysis part. It consisted of three sub-steps:

- TwoStep Cluster algorithm
- Clusters validity analysis
- Centroid turbine definition

TwoStep Cluster algorithm

The two-steps Cluster Analysis is an exploratory algorithm designed to reveal natural groupings (or clusters) within a dataset that would otherwise not be apparent. Clustering techniques have been widely used in many applications involving environmental researches [14, 15]. Clustering methods are divided in two main groups: hierarchical and partitioning algorithms. The first methods usually form clusters successively, on the basis of a pre-selection clusters established at the beginning. The second ones split data in clusters at the same time according to some criteria. The TwoStep Cluster algorithm is a hierarchical clustering method. It involves the main two steps:

- Pre-clustering;
- Clustering.

In the first step the algorithm scans the dataset samplers one by one and decides whether the current sampler has can be added to the previously formed clusters or it will be the centroid of a new cluster based on a distance criterion.

In the presented model Euclidian distance was chosen as distance criterion. Considering two generic turbines (T1 and T2) Euclidian distance is defined as follows:

$$Ed(\mathrm{T}1, T2) = \sqrt{\sum_{i=1}^{n}(T1_i - T2_i)^2} \tag{2}$$

Where T = (T_1, T_2,., T_n) with n the number of variables described before (mean P_{peak}, mean Wind speed etc.). The following step considers the sub-clusters already formed and group them in the desired number of clusters. The process starts with defining an initial cluster for each sub-cluster. Then, all clusters are compared and the pair of clusters with the smallest distance between them is merged into one cluster. The model was performed according to a prior decision about the number of clusters. It was decided to test it from 2 clusters (minimum) up to 8 clusters (maximum) and classify all turbines according to these 7 possibilities.

Clusters validity analysis and Centroid turbine definition

Since the model was tested to group turbines in different clusters sizes (from 2 clusters to 8 clusters) it was crucial to choose the optimal clusters number. There are many clustering validity indices to assess the validity of clustering. In this paper, the Sihoulette (S) coefficient was used. The S coefficient in d-dimensions is shortly defined as follows:

$$S_i = \frac{b_i - a_i}{\max\left[a_i, b_i\right]} \tag{3}$$

Where a_i is the average distance between i and all other data points in the same clusters, b_i is the smallest average distance of i to all points in any other cluster of which i is not a member. The average S_i over all points of a cluster is a measure of how tightly grouped all the points in the cluster are. Thus the average S_i over all data of the entire dataset is a measure of how appropriately the data have been clustered. S varies in the range [-1,1]. $S \geq 0{,}5$ was considered as criterion for the correct number of clusters.

With the appropriate number of cluster it was chosen a most representative turbine for each cluster. The chosen turbines was that turbine whose Euclidean distance to the centroid of the cluster was the minimum.

ANN Step

In this paper, a Multilayer perceptron (MLP) neural network was used to conduct a short-term overall wind power forecast (about 70 hours) in advance. This common neural network has been widely used in similar researchers [4,16]. MLP can give a reasonable solution for such a highly nonlinear problem as wind power forecast. It uses a feedforward architecture. It was modeled with three layers of nodes: input layer, hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. As activation function it was used the sigmoid function which is probably the most historically common one. MLP uses a supervised learning technique called backpropagation for training. The Data subset was divided in two categories: the 70% was used for the learning phase and the 30% as test phase. MLP received as input the wind speeds and wind directions and gave as output the overall P_{peak} of each site. It was tested with the best turbines of the selected clusters and it was modeled for other cases to make a comparison of results. The results were discussed according to the overall relative error (E%) with real power and the sum of squares Error (SSE).

RESULTS

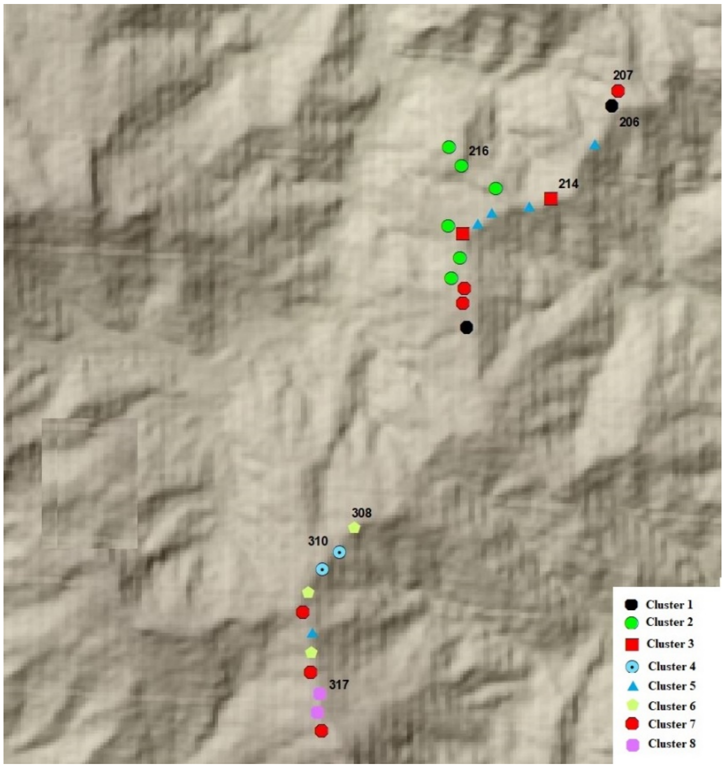
The results of the cluster analysis accordingly the recursive validity test by S coefficient are presented in Table 3. The site 1, due to its high variability of all samplers (all variables presented a high σ), showed a lower S coefficients and just the case of 8 Clusters fulfilled the criterion (S>0.5). Site 2 presented a lower variability samplers and the criterion was respected with seven clusters.

TABLE 3. Sihoulette results for site 1 and site 2.

	2 Clusters	3 Clusters	4 Clusters	5 Clusters	6 Clusters	7 Clusters	8 Clusters
Site 1	0.423	0.445	0.467	0.472	0.485	0.495	**0.511**
Site 2	0.447	0.465	0.476	0.479	0.491	**0.506**	0.512

The two cases (8 Clusters for site 1 and 7 Clusters for site 2) are reported in Fig. 6 and 7, Table 4 and 5. In Site 1, most of the clusters (2, 4, 5, 7, 8) despite the fact they were defined with a numerical method, followed a geographical clustering distribution. It reveals how this "black box" method was able to follow physical phenomena such as power loss due to wake effects, wind variation due to quotes and terrain roughness.

TABLE 4. The clusters results for site 1.

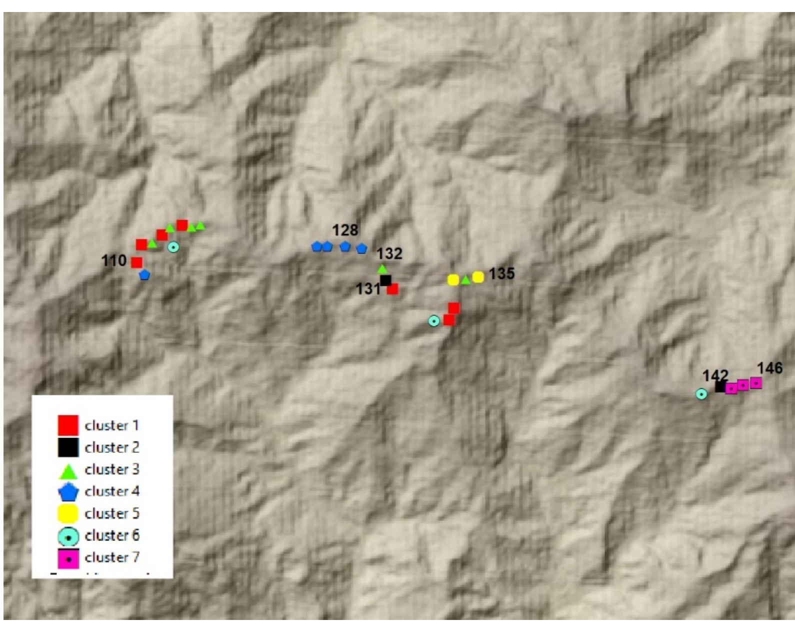


Clusters	Turbines	Centroid Turbine
Cluster 1	206, 226	206
Cluster 2	216, 217, 220, 222, 223	216
Cluster 3	214, 221	214
Cluster 4	309, 310	309
Cluster 5	210, 213, 318, 219, 313	213
Cluster 6	308, 311, 314,	308
Cluster 7	207, 224, 225, 312, 315,318	207
Cluster 8	316, 317	317

FIGURE 6. The geographical clusters distribution for site 1.

In the case of Site 2, the same considerations may be applied to Clusters 1, 3, 4, 5 and 7.

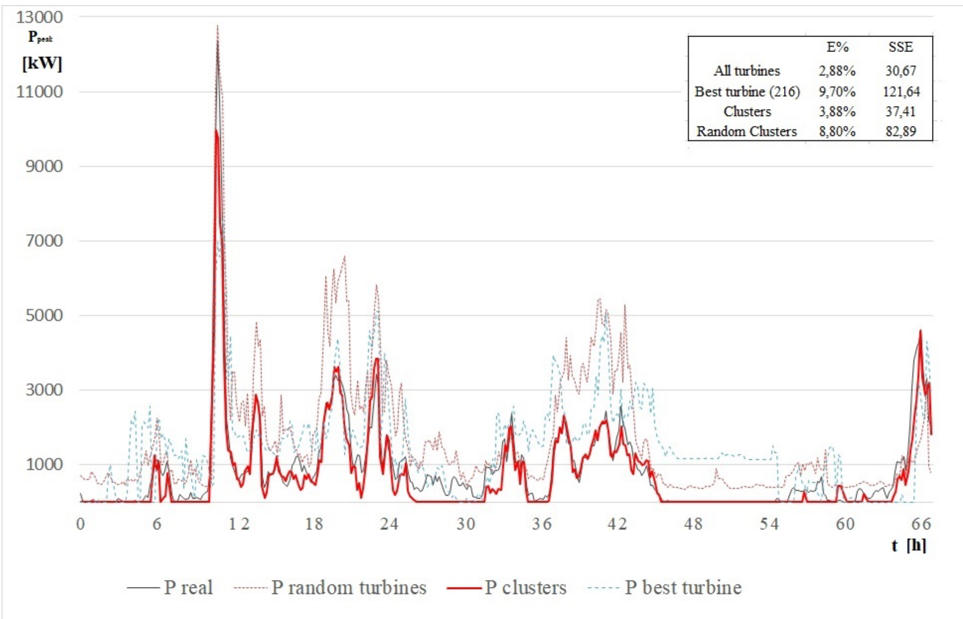
TABLE 5. The clusters results for site 2.



Clusters	Turbines	Centroid Turbine
Cluster 1	110, 111, 114, 116, 130, 141	110
Cluster 2	131, 143	131
Cluster 3	112, 115, 117, 118, 132, 133	132
Cluster 4	109, 126, 127, 128, 129	128
Cluster 5	135, 139	135
Cluster 6	113, 134, 140, 142	142
Cluster 7	144, 145, 146	146

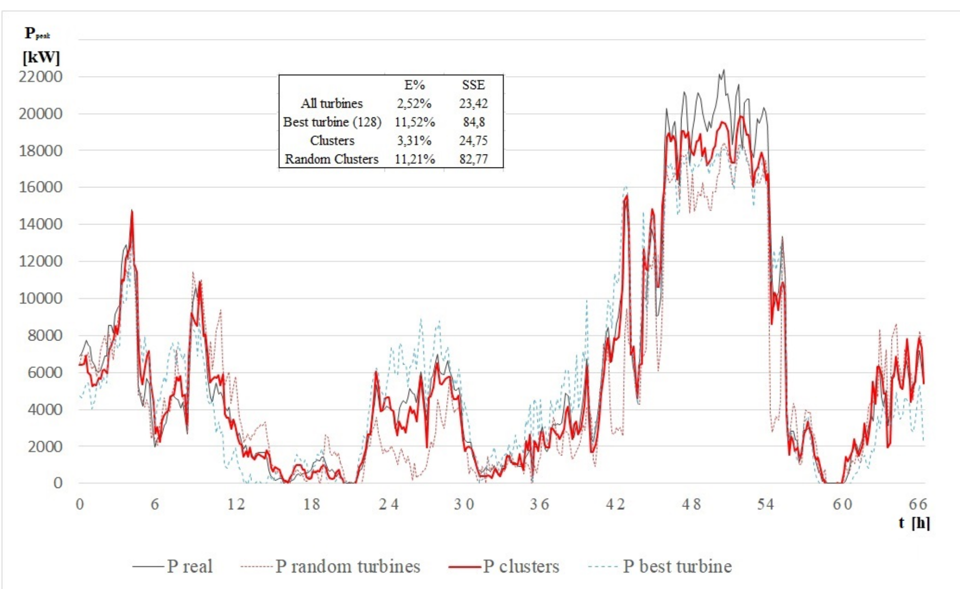
FIGURE 7. The clusters distribution for site 2.

The short-term power forecast results are showed in Fig. 8 for site 1, and Fig. 9 for site 2. The overall power output was calculated in different scenario configuration. *P clusters* refers to the output wind power of the proposed model. *P real* represents the real overall wind power. *P random turbines* represents the output overall wind power forecast using an equal number of turbines of the clusters results (8 turbines for site 1 and 7 turbines for site 2) but chosen randomly. *P best turbine* represents the overall output wind power forecast using just one turbine and the results refers to the best turbine output. It was also calculated the forecast wind power using all turbines, but it was presented just E% and SSE. Both in site 1 and in site 2, the proposed methodology reveals good performances. The Clustering method gave good results for both sites. The relative error was almost three time less than randomly chosen turbines criterion. Moreover, the errors were, in both cases, almost the same in comparison with the application of MLP for all turbines.



	E%	SSE
All turbines	2,88%	30,67
Best turbine (216)	9,70%	121,64
Clusters	3,88%	37,41
Random Clusters	8,80%	82,89

FIGURE 8. The real and forecast power for site 1 for about 70 hours.



	E%	SSE
All turbines	2,52%	23,42
Best turbine (128)	11,52%	84,8
Clusters	3,31%	24,75
Random Clusters	11,21%	82,77

FIGURE 9. The real and forecast power for site 2 for about 70 hours.

CONCLUSIONS

This paper deals with a new methodology application in short-term wind power forecast applications. This methodology consisted in the application of three main steps with a combination of Gis-based techniques, classifying statistical methods and ANN methods. The main goals reached can be summarized as follows:

- The proposed classified method (Twostep clustering) was effective in generating strong wind turbines clusters for both sites treated as case study.
- This numerical classification followed a natural geographical distribution of the wind turbines.
- The results of the model improved considerably the accuracy of the short-term wind power forecast. The application of this lower number of turbines was almost the same of the application of all turbines. This latter case usually suffer a heavy computation load not really adapted for near-real time applications such this one.

For future work it is under investigation the integration of other variables such as: slopes, terrain roughness, quotes, directions of all turbines. All these gis-based variables should lead to better performances for the model itself.

REFERENCES

1. "The World Wind Energy Report 2018". World Wind Energy Agency (WWEA), (2019).
2. J. Yan, Y. Liu, S. Han, Y. Wang and S. Feng. "Reviews on uncertainty analysis of wind power forecasting" in *Renewable and Sustainable Energy Reviews*, **52**, 1322-1330 (2015).
3. B. Hodge and M. Milligan. "Wind power forecasting error distributions over multiple timescales" in *2011 IEEE Power and Energy Society General Meeting*, edited by IEEE, (2011).
4. G.W. Chang, H.J. Lu, Y.R. Chang and Y.D. Lee. "An improved neural network-based approach for short-term wind speed and power forecast" in *Renewable Energy,* **105**, 301-311, (2017).
5. J. Wang, S. Qin, Q. Zhou and H. Jiang. "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China" in *Renewable Energy,* **76**, 91-101, (2015).
6. N. Kirchner-Bossi, R. Garcia-Herrera, L. Prieto and R.M. Trigo. "A long-term perspective of wind power output variability" in *International Journal of Climatology,* **35**, 2635-2646, (2015).
7. Landberg, L.A. "Mathematical look at physical power prediction model" in *Wind Energy,* 1-23, (1998).
8. Landberg L., Myllerup L., Rathmann O., Lundtang Petersen E., Hoffmann Jørgensen B., Badger J. and Gylling Mortensen N. "Wind resource estimation – an overview" in *Wind Energy,* **6(3)**, 261-271, (2003).
9. D. Elia. "Wind turbine control in computational fluid dynamics with OpenFOAM" in *Wind Engineering*, **41(4)**, (2017).
10. S. Mauro, S. Brusca, R. Lanzafame, M. Messina. "CFD modeling of a ducted Savonius wind turbine for the evaluation of the blockage effects on rotor performance" in *Renewable Energy,* **141**, 28-39, (2019).
11. R. G. Kavasseri, K. Seetharaman. "Day-ahead wind speed forecasting using f-ARIMA models" in *Renewable Energy,* **34(5)**, 1388-1393, (2009).
12. S. Brusca, F. Famoso, A. Galvagno, R. Lanzafame, S. Mauro and M. Messina. "Wind Turbine Wake Mathematical Models Validation by Means of Wind Field Data" in *International Journal of Appplied Engineering Research,* **12(24)**, 16068-16076, (2017).
13. S. Brusca, R. Lanzafame, F. Famoso, A. Galvagno, M. Messina, S. Mauro and M. Prestipino. "On the Wind Turbine Wake Mathematical Modelling" in *Energy Procedia*, **148**, 202-209, (2018).
14. R. Lanzafame, P.F. Scandura, F. Famoso and P. Monforte. "NO_2 concentration analysis in urban area of Catania" in *Energy Procedia*, **45**, 671-680, (2014).
15. A. Astel, S. Tsakovski, P. Barbieri and V. Simeonov. "Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets" in *Water Research,* **41(19)**, 4566-4578, (2007).
16. M. Carolin Mabel, E. Fernandez. "Analysis of wind power gneration and prediction using ANN: A case study" in *Renewable Energy,* **33(5)**, 986-992, (2008).