# Analysis of the training metrics of ANNs and linear MCP models used for wind power density estimation at a candidate site

Sergio Velázquez [a]*, José A. Carta [a], José Mª Matías [b]

[a] *University of Las Palmas de Gran Canaria, Las Palmas, Canary Islands, Spain.*
[b] *University of Vigo, Vigo, Spain.*
**Corresponding author. tel.: +34 928 45 96 71, Fax: +34 928 457319, E-mail: svelazquez@diea.ulpgc.es.*

**Abstract:** In order to estimate the amount of electricity that can be produced by a potential wind farm it is important to know how the wind resource performs at the site where it is to be installed. Of fundamental importance in an analysis of the wind resource is the wind speed parameter. Understanding how this parameter behaves over periods of time that cover ten or more years (long-term) is vital for an accurate estimation that will span the working life of the wind installation. However, in most cases there is insufficient data available about the candidate site to enable a long-term study.

In this work, the long-term wind power density at a candidate site is estimated through the use of a Measure-Correlate-Predict (MCP) algorithm and an Artificial Neural Network model (ANN). To evaluate the accuracy of the estimations different metrics are used, with a comparison of the results obtained for each of them .

The mean hourly wind speeds and directions obtained from twenty-two weather stations located on different islands in the Canary Archipelago (Spain) are used for this study.

Among the conclusions that are reached is that the use of one or another metric (or combination of metrics) in the wind power density estimation process can lead to differing interpretations and/or conclusions. For this reason, it is important that the most appropriate metric (or set of metrics) is chosen at each moment for the study that is being carried out.

***Keywords:*** *Wind Power Density, Short-Term estimation, Long-Term estimation, Artificial Neural Networks, Measure Correlate Predict*

## 1. Introduction

The wind speed at any given site varies from one year to another [1]. For this reason, the long-term performance of the wind resource (10 years or more) needs to be known as accurately as possible to enable precise estimation of the power output of a wind farm over its working life [2-4]. In most cases there is insufficient data available about the candidate site to carry out long-term analyses. As a general rule, the information available about the wind resource at a candidate site only covers short periods of time (not more than one year).

In order to estimate the long-term wind speed at a candidate site, a number of authors have used long-term wind data obtained from reference stations in combination with estimation models. The traditional Measure Correlate Predict (MCP) algorithms [5-7] and methods which use Machine Learning [8-12] are the most commonly used techniques to generate the models in the estimation processes. The former generally use a single reference station to generate the model. With some exceptions, most of the MCP methods use linear regression algorithms to characterise the relationship between the wind speeds at the candidate and reference sites. Two different methods will be used in this paper. One employs the theory of Artificial Neural Networks (ANNs), and the other a traditional MCP linear regression algorithm.

The ANNs used in this paper were comprised of three layer networks with feedforward connections. More specifically, multilayer perceptron topologies (MLPs) were used [13]. A single hidden layer with 15 neurons was employed so as not to increase the training time. This architecture has demonstrated its ability to satisfactorily approximate any continuous transformation [13].

The models used to carry out the aforementioned estimations are trained, validated and tested using the available short-term (one year) wind data from reference and candidate weather stations. Using the model thereby obtained and the observed long-term reference station wind data, the candidate station long-term wind data can be estimated. In order to evaluate the performance of the models generated during the test stage of the study, different authors use a wide variety of metrics. Some of these metrics use the ratios between the mean observed and estimated values for different parameters of the wind resource [6,10], Eq. (1), while others [9,11,12] use point-to-point metrics such as, for example, the coefficient of correlation (CC) between the estimated and observed wind speeds, Eq. (2), and the Mean Absolute Percentage Error (MAPE), Eq. (3).

$$Ratio = \frac{\frac{1}{n}\sum_{i=1}^{n} E_i}{\frac{1}{n}\sum_{i=1}^{n} O_i} \tag{1}$$

$$CC = \frac{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)\left(E_i - \overline{E}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2\right] \times \left[\sum_{i=1}^{n}\left(E_i - \overline{E}\right)^2\right]}} \tag{2}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|O_i - E_i\right|}{O_i} \tag{3}$$

where $E_i$ are the estimated data; $O_i$ , are the observed or measured data and *n* is the number of data.

## 2. Meteorological data used

The meteorological data used in this paper correspond to the mean hourly wind speed and directions of twenty-two weather stations located in six of the seven islands that make up the Canary Archipelago (Spain).
The data series used were provided by the State Meteorological Agency (Spanish initials: AEMET) of the Ministry of the Environment and Rural and Marine Environs of the Spanish Government and by the Canary Islands Technological Institute (Spanish initials: ITC).
The available wind data are as follows:

- Six (6) weather stations with 10 years of available wind data (1999-2008)
- Five (5) weather stations with data available for the year 2002.
- Eleven (11) weather stations with data available for the year 2006.

## 3. Methodology

The methodology employed in the analysis undertaken in this paper consisted of: the generation of models for short-term (one year) estimation and the generation of models for long-term (ten years or more) estimation. The first type of model generation used data from all twenty-two weather stations, while the second (long-term) type used only those stations for which meteorological data was available for a ten year period (six weather stations).

As many models (cases) were generated for the short-term study as possible combinations, taking as the starting point one station as reference station and another as candidate station. In this way, 55 combinations were generated for the year 2002 and 136 for 2006, making a total of 191 different cases. On the same basis, and using the six weather stations for which ten years worth of data were available, a total number of 15 cases were generated for the long-term study.

Following is an explanation of the different baseline scenarios used in the study.

Scenario A): Estimation of the short-term wind data using ANNs.

The models are generated from the known wind data for the reference and candidate stations (one year). The data available for the reference and candidate stations is randomly divided for use in the training, validation and test stages in respective proportions of 60%, 20% and 20%.

Different networks or estimation models are generated using the data from the training and validation stages. Using these models and the test data for the reference station (which is not used in the generation of the model), data estimation is performed for the candidate station. The estimated data is then compared with the observed data to generate the different metrics used in the analysis.

The wind speed and direction of the reference station (input weather station) are used as input parameters of the neural network. The candidate station (target weather station) wind speed is used as the output parameter (Fig. 1).

Scenario B): Short-term estimation using an MCP method

A simple linear regression between the wind speed of the reference and candidate stations is used for the MCP method. Wind direction is considered in the generation of the models, with the parameters of the model calculated for twelve direction bins of 30º. A simple validation is carried out to evaluate the quality of the models. That is, 20% of the data is reserved as a test subset, which is not used in the construction of the model. The remaining 80% is used for training. The data is randomly allocated to these two groups.

Scenario C): Long-term estimation using ANNs

Only the six weather stations for which data is available for a ten year period are used in the generation of the different networks or models in the long-term estimation study. One of the available years (in this case 2008) is used for generation of the network. All the data for the year is randomly allocated to two sub-groups (80% to training and 20% to validation). As in the case of Scenario A), the different models or networks are generated using this information. The candidate station long-term data is estimated using these models and the data corresponding to the other years of the reference station. By comparing the estimated data and the observed long-term data at the candidate station, the different metrics to be used in the analysis of the results are calculated.

Scenario D): Long-term estimation using an MCP method

The same stations are used as in Scenario C), as well as the same reference year. The models are generated in the same way as in Scenario B), except that in this case 100% of the year's data is used in the construction of the models. Once the parameters of the model have been calculated, the long-term candidate station data is estimated using the data from the remaining years of the reference station.

Matlab software (the MathWorks, Inc) was used to implement the different scenarios.

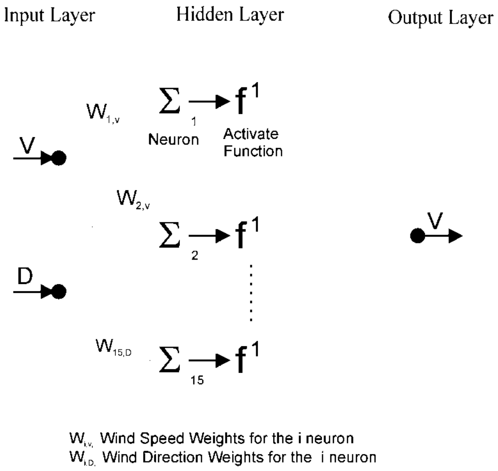


*Fig. 1. ANN Schematic diagram with wind speed (V) and wind direction (D) of a reference weather station as input signals, and wind speed (V) of a candidate (target) station as output signal.*

## 4. Analysis of Results.

Figure 2 shows the results obtained for the mean wind speed ratio in the case of Scenarios A) and B). This comparison is performed by representing on the x-axis the existing coefficients of correlation R, Eq. (4), between the wind speeds measured at the reference and candidate stations.

$$R = \frac{\sum_{i=1}^{n}\left(Vr_i - \overline{Vr}\right)\left(Vc_i - \overline{Vc}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(Vr_i - \overline{Vr}\right)^2\right] \times \left[\sum_{i=1}^{n}\left(Vc_i - \overline{Vc}\right)^2\right]}} \tag{4}$$

where $Vr_i$ and $Vc_i$ are the measured wind speeds at the reference and candidate weather stations, respectively. $\overline{Vr}$ and $\overline{Vc}$ are the mean wind speeds at the reference and candidate weather stations.

Based on Figure 2, and for the cases studied, the following conclusions were reached: a) the ratio between the estimated and observed mean wind speed is independent of the existing coefficient of correlation, R, between the mean wind speeds at the reference and candidate stations. This becomes more noticeable for coefficients of correlation greater than 0.4. b) for cases where the coefficient of correlation is less than 0.4, the dispersion in the results obtained for the different cases analysed is relatively high (in the range between 0.82 and 0.99). For coefficients of correlation higher than 0.4, the results are concentrated principally between the values 0.95 and 1.01.

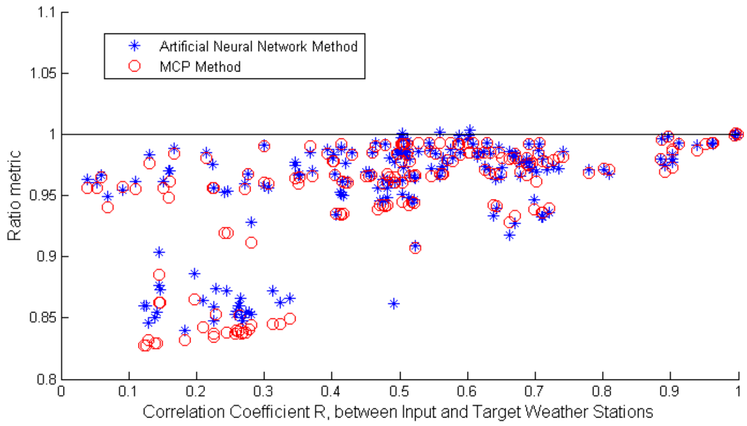


*Fig. 2. Variability of the ratio of the mean wind speed with the correlation coefficient between reference (input) and candidate (target) weather station wind speed. Scenarios A) and B).*

In the same analysis, but using the Mean Absolute Relative Error (MARE) point-to-point calculation metric for the wind speed, S. Velázquez et al. [11] found that this was dependent on the coefficient of correlation, R, between the wind speeds at the reference and candidate stations.

Figure 3 shows the results for the wind power density ratio, for the different cases studied in Scenarios A) and B). Unlike the previous results, obtained for the mean wind speed ratio, the mean wind power density ratio does depend on R, Eq. (4).

The wind power density $P_i$, or power per unit of area perpendicular to the direction from which the wind is blowing, is given by Eq. (5). Where $\rho_i$ is expressed in kg m$^{-3}$ and $V_i$ is expressed in m s$^{-1}$, $P_i$ is obtained in W m$^{-2}$. $P_i$, which depends on the air density $\rho_i$ and on the wind speed $v_i$, is the basic unit for measuring the power contained in the wind.

$$P_i = \frac{1}{2}\rho_i V_i^3 \tag{5}$$

If the results obtained in Fig. 2 and Fig. 3 are compared it can be observed in many cases that, though a good result is obtained for the mean wind speed ratio (values close to 1), the same cannot be said for the mean wind power density ratio. This can be seen, for example, in the results in the range of coefficients of correlation between 0.4 and 0.8. In the case of the mean wind speed ratio, these values are generally between 0.95 and 1.01, while for the mean wind power density ratio the results are between 0.5 and 0.8. It can be deduced from this analysis that a good result in the mean wind speed ratio is not always equivalent to a good result in the estimation of the wind power density.

The coefficient of correlation, CC, between the estimated and observed wind speeds at the candidate site, Eq. (2), depends on the existing coefficient of correlation, R, between the wind speeds at the reference and candidate stations, Eq. (4) [11]. So, the higher CC is, the closer to 1 will be the mean wind power density ratio.

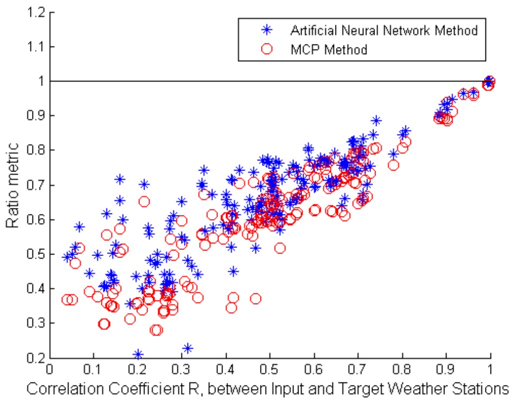


*Fig. 3. Variability of the ratio of the mean wind power density with the correlation coefficient, R, between reference (input) and candidate (target) weather station wind speed. Scenarios A) and B).*

Figure 4 show the results obtained in the 15 cases analysed in Scenario C) for the mean wind speed ratio metric and the mean wind power density ratio metric. Also shown, for each case, is the coefficient of correlation, R.

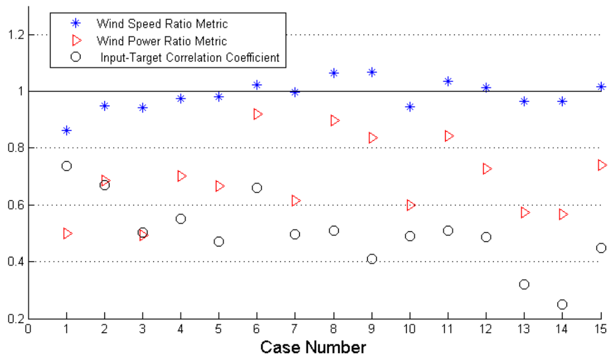


*Fig. 4. Results for the ratio of the mean wind speed and ratio of the mean wind power density (long-term). Scenario C).*

The basic conclusions obtained from the results for Scenario C) are the same as for Scenarios A) and B). The values, for example, in cases 7, 12 and 15, of 0.9959, 1.0132, and 1.0165, respectively, for the mean wind speed ratio are close to the target value of 1. However, the results in the same cases for the mean wind power density ratio are, respectively, 0.6141, 0.7272 and 0.7392, some way off the target value of 1.

If a point-to-point calculation metric like the Mean Absolute Percentage Error (MAPE) is used for the same Scenarios C) and D), then the results shown in Figures 5 and 6 are obtained.

If Figures 4 and 5 are compared, cases such as case 1 are observed which, while displaying the worst result for the mean wind speed ratio, has the third best result for the MAPE-based analysis. Meanwhile, case 7 gives a mean wind speed ratio of 0.9959 (the best of the results for this metric), which is practically equal to the target value, but has one of the worst results for the MAPE metric of wind speed, with values of 55.90% and 56.41% with respect to the

observed value depending on whether the estimation is conducted using ANNs or MCP methods, respectively. Identical conclusions are obtained if the ratio of the mean wind power density, Fig. 4, is compared with the MAPE metric of wind power density, Fig. 6.

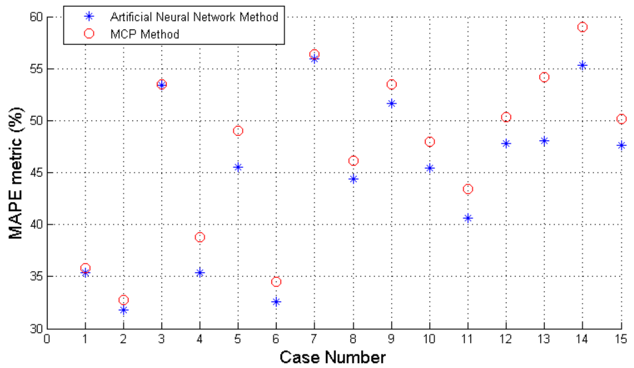


*Fig. 5. Results for the MAPE metric of the wind speed. Scenarios C) and D).*

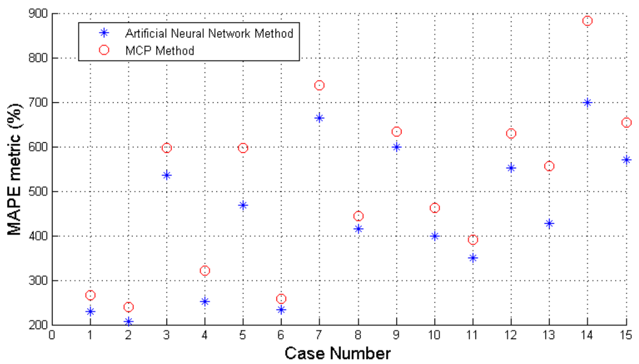


*Fig. 6. Results for the MAPE metric of wind power density. Scenarios C) and D).*

It can also be concluded from Figures 4 to 6 that the errors calculated when using point-to-point calculation metrics are much higher than when using metrics such as the ratios between the mean values of the entire data series.

## 5. Conclusions

The most important conclusions that have been reached in this paper are as follows:

a) The results of the metric of the ratio between mean estimated and observed wind speeds are independent of the coefficient of correlation between the reference and candidate wind speeds, Eq. (4). This is not the case with the metric of the ratio of the mean wind power density which is dependent on this coefficient of correlation.

b) In the estimation of short and/or long-term wind power density, the ratios between the mean values of the observed and estimated parameters, Eq. (1) are, on their own, not good indicators for decision-taking in analyses of the estimation of the wind resource, since values close to the target value of 1 in the ratio of the mean wind speed are not always equivalent to good results in the estimation of the wind power density, Figs. 4-6.

If the above metrics are considered for use in analysis of the performance of the estimation models, additional metrics should also be used such as the coefficient of correlation, CC, between the estimated and observed values of the wind speed. This metric takes into account the combined performance over time of the estimated and observed values.

c) The interpretations that can be made when using metrics such as the ratio between the mean values of parameters and point-to-point calculation metrics such as the MAPE, are very different. Case 7, for example (Fig. 4), with a mean wind speed ratio close to 1, gives a relative error (MAPE), with respect to the wind speed, higher than 55%, Fig. 5.

d) When the objective is the estimation of the wind power density for a subsequent point-to-point calculation (as is, for example, an hourly calculation), metrics like the MAPE and the CC are considered to be better indicators when it comes to analysing the accuracy of the models.

**References**

[1] TR. Hiester, WT. Pennell, The siting handbook for large wind energy systems. 1st ed. New York: WindBook; 1981.

[2] GW. Koeppl, Putnam's power from the wind. Second ed. New York: Van Nostrand Reinhold Company; 1982.

[3] CG. Justus, K. Mani, AS. Mikhail, Interannual and month-to-month variations of wind speed. Journal of Applied Meteorology 18, 1979, 913–932.

[4] CI. Aspliden, DL. Elliott, LL. Wendell, Resources assessment method, siting, and performance evaluation. In: Guzzi R, Justus CG. Physical climatology for solar and wind energy, New Jersey: World Scientific; 1988, pp 321-76.

[5] R. García-Rojo, Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach, Wind Engineering 28, 2004; pp 213-236.

[6] AL. Rogers, JW. Rogers, JF. Manwell, Comparison of the performance of four Measure-Correlate-Predict algorithms. Journal of Wind Engineering and Industrial Aerodynamics 93, 2005, pp 243-264.

[7] JC. Woods, SJ. Watson, A new matrix method of predicting long-term wind roses with MCP, Journal of Wind Engineering and Industrial Aerodynamics 66, 1997, pp 85-94.

[8] JA. Carta, S. Velázquez, JM. Matías, Use of Bayesian Networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site, Energy Conversion and Management 52, 2011, pp 1137-1149.

[9] A. Oztopal, Artificial Neural Network approach to spatial estimation of wind velocity, Energy Conversion and Management 47, 2006, 395-406.

[10] P. Lopez , R. Velo, F. Maseda, Effect of direction on wind speed estimation in complex terrain using neural networks, Renewable Energy 33, 2008;33, pp 2266-2272.

[11] S. Velázquez, JA. Carta, JM. Matías, Influence of the input layer signals of ANNs on wind power estimation for a target site: A case study, Renewable and Sustainable Energy Reviews, 15, 2011, pp 1556-1566

[12] DA. Fadare, The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria, Applied Energy, 87, 2010, pp 934-942.

[13] JC. Principe, NR. Euliano, WC. Lefebvre, Neural and Adaptive Systems. Fundamentals Through Simulations, first ed. New York: John Wiley & Sons, Inc.; 2000.