22nd International Conference on Renewable Energies and Power Quality (ICREPQ'24)
Bilbao (Spain), 26th to 28th June 2024
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.22, September 2024

Evaluation of MCP Correlation Algorithms Applied to Wind Data Series

A. Moreira[1], T. Rocha[2], J. Mendonça[3], R. Pilão[4] and P. Pinto[1]

[1] MEGAJOULE S.A.
Rua do Divino Salvador de Moreira, 255; 4470-105 Maia (Portugal)

[2]INESC TEC, Centre for Power and Energy Systems
Rua Dr. Roberto Frias, 4200-465 Porto (Portugal)

[3] SIIS, Instituto Politécnico do Porto,
R. Dr. Roberto Frias, 4200-465 Porto (Portugal)

[4] CIETI, Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino de Almeida, 341; 4200-072 Porto (Portugal)

Abstract. This work aimed to develop methodologies for analysing statistical correlations among wind data series using various Measure-Correlate-Predict (MCP) methods, with the goal of selecting the most suitable method for extrapolating long-term data with minimal associated uncertainty. It was analysed the minimum time required for a wind measurement campaign when applying this methodology. Fifteen local wind measurement stations were selected. The long-term wind data reanalysis series that exhibited the strongest correlation with the measured wind data at each station was then chosen. Multiple tests were conducted with different simultaneous periods between the measured data series and the long-term series. Fifteen correlation algorithms were tested for each concurrent period. The performance of each model was evaluated using the RMSE (Root Mean Square Error) and MBE (Mean Bias Error) associated with each MCP. Analysis of the errors identified measurement periods with the lowest associated error ranging from 1 to 5 years and a single-factor ANOVA analysis was conducted. Finally, t-significance tests were performed. The study concluded that the Neural Network was the most effective MCP method. Additionally, it was determined that the minimum number of years required for a local measurement campaign should be between 2 and 3 years.

Key words. MCP, correlation algorithms, local measurement station, wind regime, reanalysis data series

1. Introduction

From the standpoint of wind energy, the variability of wind resources stands out as a significant characteristic. Wind conditions exhibit considerable variation both spatially and temporally [1]. This variability presents a particular challenge when assessing the long-term feasibility of wind projects, which typically have a lifespan of around 20 years. Wind data collected over short periods is inadequate for accurately representing average conditions over the project's lifespan due to year-to-year fluctuations in wind speed [2].

To tackle this challenge, researchers have proposed the Measure-Correlate-Predict (MCP) methodology to evaluate the Long-Term characteristics of the wind resource. This methodology (Figure 1) involves utilizing concurrent wind data from a relatively short-term local measurement campaign and correlating it with simultaneous data from a reference long-term/historical data source. The correlation established between the two datasets allow for extrapolation of the short-term measurement data into a representative long-term time series of wind data [3].

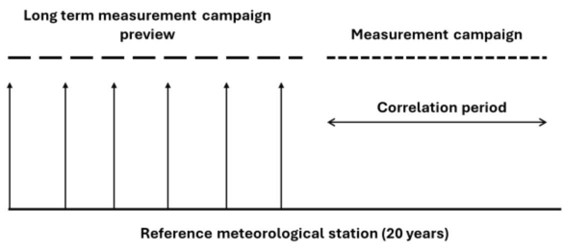

Fig. 1. Example of MCP methodology.

The MCP correlation procedure can be implemented using various methods [4]:

A. regression method where each regression method may disaggregate the data into different sectors (e.g., 36 or 360), seasons (e.g., all year, 2 or 4 seasons), and diurnal/nocturnal periods - a linear regression is then fitted to each subset of the original data [5].

B. matrix method that describes the connection between two wind climates in a matrix form. It expresses the wind speed and direction at a particular site as a matrix of speed-ups (variations in wind speed) and deflections (variations in wind direction) relative to the reference site. [6].

C. artificial neural networks (ANNs) are another powerful tool commonly used. ANNs are computational models inspired by the structure and function of the human brain's neural networks. In the case of wind data analysis, ANNs can be trained to learn complex patterns and relationships between input variables (such as wind speed, direction, temperature, etc.) and output variables (such as long-term wind predictions). [7].

This work aims to develop methodologies for analyzing statistical correlations among wind data series using various MCP methods and select the most appropriate method for extrapolating long-term data with the least associated uncertainty. Furthermore, the study intends to investigate how the concurrent period used to build the correlation can affect the performance indicators of MCP methods.

2. Methodology

Initially, 15 wind measurement stations were selected, considering the widest possible extension of the measurement campaign period, and ensuring wind speed and direction data coverage above 90 %. Nine measurement stations were chosen in Portugal, two in Romania, two in Poland, and two in South Africa, as shown in Table I. This multinational selection aimed to broaden the scope of analysis through a broad database covering different terrain complexities, characterized by the Ruggedness Terrain Index (RIX), orography, roughness, and local wind regime characteristics.

Table I. Characteristics of the 15 measuring stations.

Station	Location	Orographie	RIX (%)	Roughness	A (m/s)	k	V_{med} (m/s)	Heights (m)	Years
A1	Shout Africa	Little Complex	2.1	Low vegetation	7.2	1.71	6.4	40/60	5
A2		Medium Complex	4.7	Forest	8.1	2.74	7.8	40/60	5
PT1	Portugal	Little Complex	1.5	Low vegetation	7.4	1.74	6.8	40/60	10
PT2		Flat	0.8	Villages,wooded. rural space	9.1	2.86	7.9	40/60/80	5
PT3		Medium Complex	4.1	Agrícultural fields	8.7	2.87	7.5	40/60/80	5
PT4		Complex	7.8	Vegetation	6.5	2.21	8.0	40/60	7
PT5		Complex	7.3	Vegetation	6.0	2.04	5.4	40/60	8
PT6		Complex	6.8	Forest	7.4	2.11	6.6	40/60	7
PT7		Complex	7.6	Agrícultural fields	6.2	1.80	5.6	30/60	7
PT8		Flat	1.1	Vegetation	5.6	2.28	5.1	30/60	3
PT9		Medium Complex	3.8	Forest	9.0	2.20	7.9	10/40/60	2
PO1	Poland	Flat	0.6	Vegetation	8.3	2.60	7.5	30/60	5
PO2		Medium Complex	3.1	Agrícultural fields	7.5	2.18	6.9	30/60	5
R1	Romania	Flat	1.9	Forest	6.5	2.06	6.1	20/40/60	6
R2		Medium Complex	4.0	Agrícultural fields	8.1	2.74	7.8	20/40/60	6

K is the shape parameter and A is the scale parameter of the Weibull distribution that characterizes the local wind regime.

For each measuring station, the reanalysis series closest to the measuring location were selected:

- **ERA5**: 4 sets at 100 m above ground level (a.g.l) and 4 sets at 10 m a.g.l;
- **MERRA2**: 4 series at 50 m a.g.l..

The reanalysis series for calculating MCP at each station were selected based on the best correlation obtained, measured by the coefficient of determination R^2, between the long-term series and the measured data series, considering the total concurrent period of measured data. Several tests were conducted with different periods between the measured data series and the long-term series. For each measurement mast, tests were performed using periods ranging from 6 months to a maximum of 10 years, according to the available total measured data period.

The correlation algorithms were tested for each measurement period, including 13 regressions, 1 matrix, and 1 artificial neural network.

Subsequently, the quality of the correlation obtained was evaluated by calculating errors such as RMSE and MBE between the wind speed data measured and predicted by the different MCP algorithms tested, on an hourly basis.

For the years with the lowest associated errors, single-factor *ANOVA* [8] analyses were conducted to understand the significance of the measurement year, considering as null hypothesis that there are no discrepancies among the errors of each model, while alternative hypothesis suggested that at least two models possess different error means. Additionally, t-significance tests were performed for the selected years to identify MCP algorithms with significantly different means. A significance level of 5 % was adopted for the analysis.

3. Results and Discussion

For each of the 15 measurement stations, ERA5 reanalysis data has been selected, following the defined criterion based on the value of the coefficient of determination R^2, obtained for the correlation between the long-term series and the measured data series, considering the concurrent data period.

The stochastic nature of the wind in different seasons of the year makes it important to determine the minimum period necessary for its correct characterization. To find this minimum period the correlation algorithms were tested, using the presented methodology, for each measuring stations using different simultaneous periods between the measured data series and the long-term series predicted.

The analysis of results from the different study stations concerning the variation of RMSE and MBE errors with respect to the correlation time applied in the MCP algorithm indicated a reduction in errors over time. This reduction was particularly evident in the case of RMSE.

Figures 2 to 4 show, for the 15 stations under analysis, the evolution of the RMSE and MBE values as a function of measurement time for the matrix, neural network and 4 Seasons 360s DN methods. These three methods were chosen because they present the lowest error values as well as the same tendency for values to vary over time. The results highlight both the occurrence of the lowest error values and the consistent trend in their variation over time.

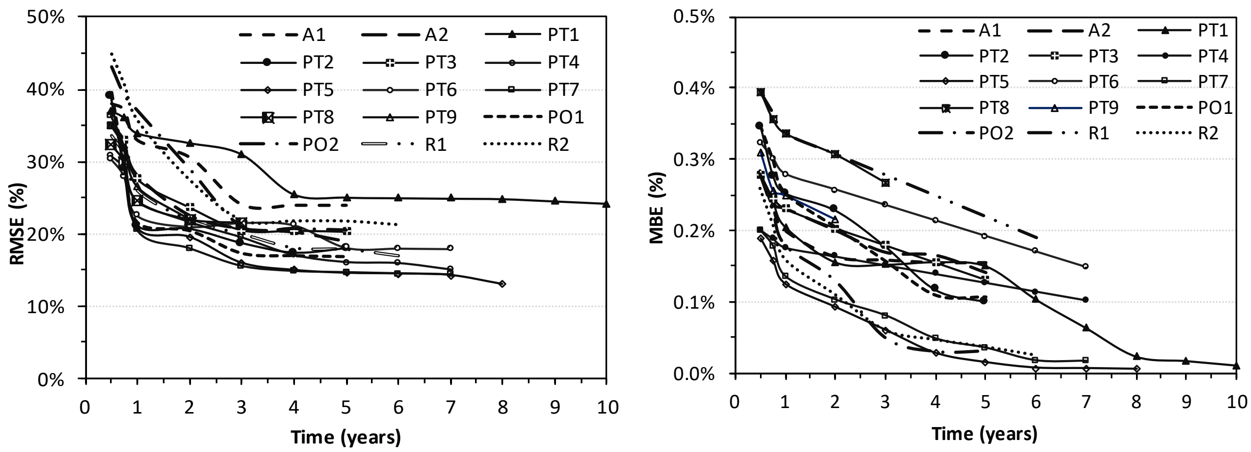

Fig. 2 RMSE and MBE as a function of time for the matrix method.

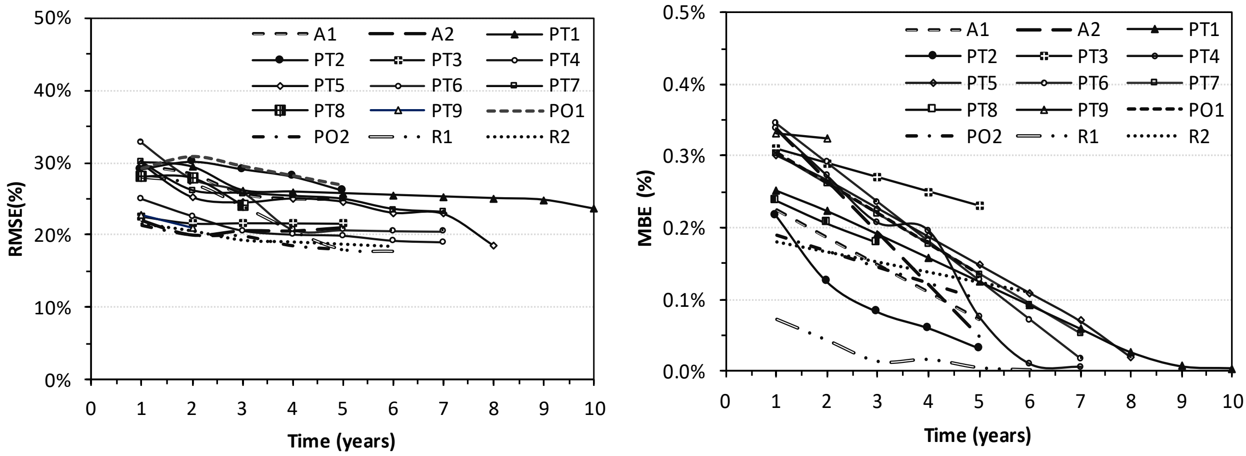

Fig. 3. RMSE and MBE as a function of time for the 4 Seasons 360s DN method.

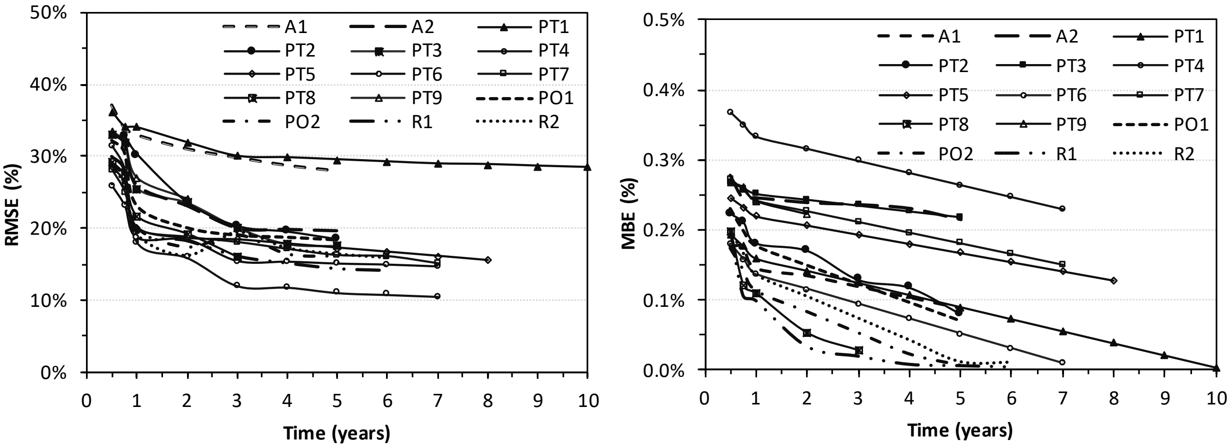

Fig. 4. RMSE and MBE as a function of time for artificial neural network method.

RMSE evaluates the dispersion between the measured data and the data predicted by the MCP algorithm used. For wind data measured over periods shorter than one year, the neural network and matrix method showed RMSE errors ranging between 20 % and 40 %, and 25 % to 45 %, respectively (Fig. 2, 4). The 4 Seasons 360s DN model does not consider trials shorter than a year due to seasonal division constraints. Generally, for all three methods, there was an average decrease of approximately 10 % in errors from one to five years of data collection across all stations. However~~,~~ both the matrix and neural network models exhibited a significant drop in RMSE values within the first three years of data. Subsequently, after five years of measurement, RMSE errors somewhat stabilized.

Regarding the duration of measured data required for MCP, the results indicated that the minimum period of measured data should lie between 2 and 3 years. Further extending the measurement campaign beyond 3 years may not offer significant improvements in cost-effectiveness.

Regarding the MBE, all three algorithms yielded positive error values, indicating a tendency to overestimate the models, suggesting that, on average, the predictions are higher than the actual measurements (Figures 2-4). Typically, a sharp decrease in MBE is observed from six months to one year of measurement. Subsequently, after the first year, MBE decreases more gradually and follows an approximately linear trend with increasing measurement time.

Additionally, among the stations analysed, the neural network method exhibited the highest MBE value of 0.37 % at station PT4. For the matrix method, the highest error value was 0.39 % recorded at station R01, while the 4 Seasons 360s DN model showed the highest MBE value of 0.36 % at station PT6, considering one year of measured data.

To investigate the impact of measurement duration ranging from 1 to 5 years, a one-way analysis of variance (ANOVA) was conducted for each year. This aimed to compare the average errors of the 15 MCP models employed and ascertain if the differences are statistically significant.

Both RMSE and MBE tests showed highly statistically significant results (p-value << 0.05). This means that the null hypothesis can be rejected and that all error means are equal. In other words, there are statistically significant differences between the models' error performances.

Figures 5 and 6, the Boxplots for 1 and 5 years of measured data are depicted considering RMSE values, respectively.

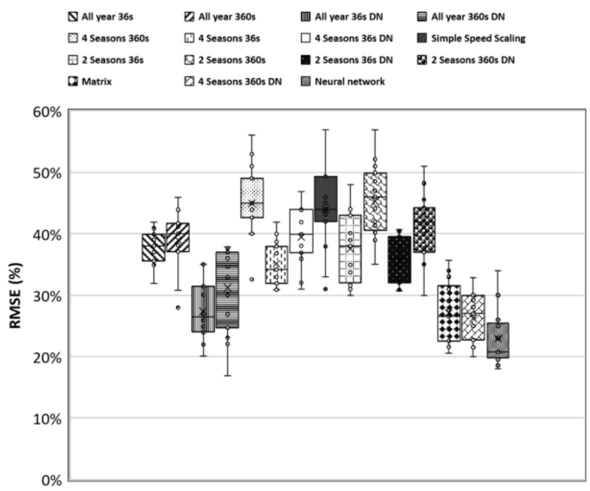

Fig. 5. Boxplots for 1 year of measured data relative to RMSE.

Most MCP models displayed outliers-data points that diverge from the general trend. These outliers may persist due to subjective decisions regarding their elimination.

Additionally, distinct average values were observed among algorithms. The neural network model, 4 Seasons 360s DN, and the matrix exhibited the lowest RMSE values across the five years of data considered.

Finally, to ascertain which algorithms exhibit statistically significant differences, t-tests for the difference of means were conducted between the neural network models, 4 Seasons 360s DN, and the matrix for each of the five years analysed (Table II and III). Thus, for every year, three t-tests were performed to compare: Matrix versus 4 Seasons 360s DN, Matrix versus Neural Network, and 4 Seasons 360s DN versus Neural Network.

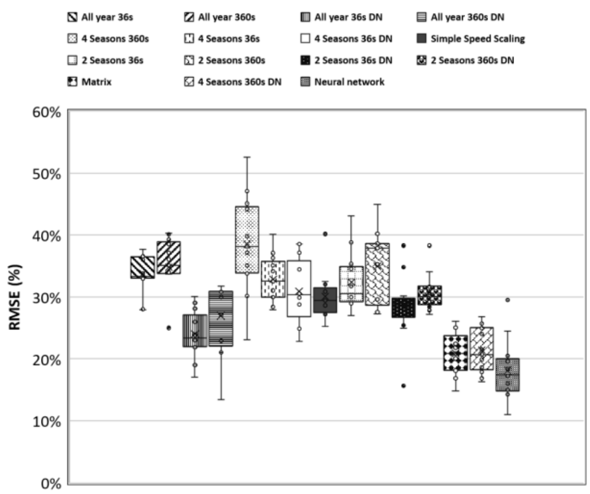

Fig. 6. Boxplots for 5 years of measured data relative to RMSE.

Tables II and III provide the stat_t and p-values resulting from the one-sided t-tests of significance. These tests were conducted for the three combinations between the models, considering each year of measured data.

Table II - RMSE *stat_t* and *p_value* values resulting from the left-sided t tests of significance for the three combinations between MCP models.

Years	Matriz vs *4 Seasons 360s DN*	Matriz vs Neural Network	*4 Seasons 360s DN* vs Neural Network
	stat_t/valor_p	*stat_t/valor_p*	*stat_t/valor_p*
1	-0.0555/0.4780	**2.0657/0.02448**	**2.5436/0.0086**
2	-1.3282/0.0978	**1.7130/0.04930**	**3.0167/0.0028**
3	**-2.4752/0.0104**	1.1091/0.1392	**3.5452/0.0008**
4	**-2.7288/0.0061**	0.7711/0.2244	**3.0526/0.0029**
5	**-2.0739/0.0055**	0.7343/0.2353	**2.9767/0.0034**

Table III - MBE *stat_t* and *p_value* values resulting from the left-sided t tests of significance for the three combinations between MCP models.

Years	Matriz vs *4 Seasons 360s DN*	Matriz vs Neural network	*4 Seasons 360s DN* vs Neural network
	stat_t/valor_p	*stat_t/valor_p*	*stat_t/valor_p*
1	-1.4776/0.0757	1.4060/0.0858	**2.7377/0.0055**
2	-1.2422/0.1126	0.9236/0.1821	**2.0374/0.0259**
3	-0.8572/0.1999	0.4675/0.3222	1.2792/0.1065
4	-0.9942/0.1654	-0.3219/0.3752	0.5333/0.2996
5	-0.0125/0.4950	-0.2034/0.4204	-0.1995/0.4218

According to Table II, there is a statistically significant difference (p-value below 5 %) between Matrix and Neural Network for up to 2 years of data. Similarly, there is a statistically significant difference (p-value below 5 %) between 4 Seasons 360s DN and Neural Network for up to 5 years of data. Therefore, we rejected the null hypothesis, indicating significant differences between these models. Additionally, in both cases, the Neural Network model exhibited the lowest average RMSE value.

However, the results presented in Table III showed statistically significant differences between 4 Seasons 360s DN and Neural Network, with a positive *stat_t* value, for up to 2 years of data. This suggests that the Neural Network model had the lowest mean MBE for this period.

In summary, the t-tests suggest that the Neural Network algorithm is the most suitable model for MCP based on the given data.

4. Conclusions

A study was developed with the aim of selecting the most appropriate method for extrapolating long-term data and investigate how the concurrent period used to build the correlation between wind data series using different MCP methods can affect its performance indicators. Among the 15 models tested, matrix algorithms, neural networks, and the 4 Seasons 360s DN method exhibited the lowest error values. The analysis of errors revealed that the Neural Network was the most suitable MCP method due to its lowest RMSE and MBE values.
Regarding the period of concurrent data to be used to perform an MCP, it was concluded that the minimum period of measured data should be between 2 and 3 years.

Acknowledgment

This work was supported by multi-year financing of FCT- *Fundação para a Ciência e Tecnologia* (grant UIDB/04730/2020).

References

[1] J. V. P. Miguel, "Influência da duração da campanha de medição anemométrica na avaliação de recursos eólicos com base na aplicação de métodos de MCP," Instituto da energia e do ambiente, São Paulo (2016), pp 21-26.
[2] J. T. Jervell, "Estudo da influência das características do vento no desempenho de aerogeradores", Tese de Mestrado em Engenharia Mecânica, Porto (2008). Porto, pp 8-18.
[3] R. M. T. Gomes, "Análise de Incerteza em Estimativa de Produção de um Parque eólico", Tese de Mestrado em Engenharia Eletrotécnica e de Computadores, Porto (2018). pp 5-15.
[4] H. L. S. d. Reis, "Estudo sobre a aplicação do método MCP em dados de longo período para a estimativa de recurso eólico e solar", Brasília (2013). pp 5-20.
[5] L. Landberg, L. Myllerup, O. Rathmann,, E. L. Petersen, B. H. Jørgensen, J. Badger, N. G. Mortensen, "Wind resource estimation—an overview", Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology (2003). Vol. 6 (3), pp. 261-271.
[6] J. C. Woods, S. J. Watson, (1997). "A new matrix method of predicting long-term wind roses with MCP", *Journal of Wind Engineering and Industrial Aerodynamics* (1997). Vol. *66*(2), pp. 85-94.
[7] S. A. Kalogirou, "Artificial neural networks in renewable energy systems applications: a review",. Renewable and sustainable energy reviews (2001). Vol. 5(4), pp. 373-401.
[8] J. Hair, W. Black, B. Babin, R. Anderson, (2019). "Multivariate Data Analysis", Prentice Hall, (2019).