

A new Measure-Correlate-Predict Wind Resource Prediction method

C. Skittides and W.-G. Früh¹

¹ Institute of Mechanical, Process and Energy Engineering,
School of Engineering and Physical Sciences
Heriot-Watt University
Riccarton, Edinburgh EH14 4AS (UK)

Phone number:+0044 131 4514374, Fax number:+0044 131 4513129, e-mail: cs261@hw.ac.uk, w.g.fruh@hw.ac.uk

Abstract. We present a new MCP method to estimate the wind resource at a target site based on a short set of concurrent measurements from that site and a reference site for which also a long historical wind speed record is available.

The method derives from time-series analysis methods used in nonlinear dynamical systems and uses Singular Systems Analysis to define the optimum correlation between the target and reference site, which is then used to build the model used for estimating the target site resource. The performance of this method is then applied to a set of Met.Office data from Scotland and benchmarked a basic linear-regression MCP.

Key words

Wind Energy, Wind Resource, Measure-Correlate-Predict, Singular-Systems Analysis.

1. Introduction

Given that modern wind farms have installed capacities of several hundred megawatts or more, even a small overestimate or an uncertainty in the predicted resource can result in a shortfall of income of several million pounds annually per wind farm. For this reason, wind resource assessment is an important part of siting and developing wind farms. In particular, correcting any short-term measurements at a site for inter-annual variability is required to estimate the long-term resource.

The principle behind the MCP methodology is to correlate short-term wind data of a target site, usually a site targeted for the development of a wind farm, with long-term wind data of a reference site, often a meteorological office site nearby, so that a relationship between them is established [1]. Carta et al. [2] recently published a review, which provides very detailed descriptions of the methods currently applied.

2. MCP method

Here we introduce first the most basic MCP method using linear regression, which is used as a benchmark to test the new methods against. After this, the principle of the new methods is introduced in section B.

A. Linear Regression MCP

In general, the following function mathematically describes MCP to estimate the wind speed at the target site, U_{target} :

$$U_{target} = f(U_{ref}, \theta_{ref}) \quad (1)$$

where U_{ref} is the wind speed at the reference site and θ_{ref} the wind direction at the reference site. The function f has a fixed form which is determined from the concurrent data and then applied to the historical record. The simplest example of this function is a linear function of the wind speed only and ignoring wind direction:

$$U_{target} = b + m U_{ref} \quad (2)$$

where b is a constant and m is the gradient, the values of which are determined from a linear regression of the available target site measurements from the measurement campaign against the reference data obtained at the same time from the reference site.

Once the coefficients, b and m , have been determined, eq.(2) can be applied to the existing historical data set from the reference site to calculate the corresponding wind speeds at the target site for that historical period.

Obviously, the fact that the wind direction does affect the local wind resource at the target site differently than at the reference site is acknowledged and many methods have been developed to refine the linear regression to take that into account [e.g., 2]. Other methods, for

example those based on correlating wind speed distributions, or using non-linear methods such as Artificial Neural Networks exist but, in the context of testing the new method against a basic benchmark, are not followed up in this paper.

B. Singular-Systems MCP

The majority of MCP methods are constrained by assuming one, or a few, pre-determined functions between either instantaneous wind speeds or wind speed distributions across the two sites. This new method is based on finding the functional form empirically and is inspired from time-series analysis methods in nonlinear dynamical systems [3,4].

The assumption is that the wind speed dynamics at the two sites together are governed by a common dynamical system for which the essential dynamics can be inferred from the measurements at both sites not at a single instance in time but from time series sections extending over a window of few hours. The ‘Singular Systems’ aspect then reformats that description obtained from the time series sections to isolate the most significant components to capture the common behaviour across the two sites. The optimisation uses the statistical tool known as Principal Component Analysis (PCA) or Empirical Orthogonal Function (EOF) analysis. This method was previously applied to wind speed measurements at a single site for forecasting the wind speed a few hours ahead [5,6]. To apply the PCA to the wind speed time series, it is advisable to express the wind speed and direction in wind vectors, $u = -U \sin\theta$ and $v = -U \cos\theta$, to avoid discontinuities across $0^\circ/360^\circ$.

By replacing the four measurements at time t of

$$(u_{ref}(t), v_{ref}(t), u_{target}(t), v_{target}(t))$$

by an extended set

$$y(t) = (u_{ref}(t, t-1, \dots, t-m), v_{ref}(t, \dots, t-m), u_{tar}(t, \dots, t-m), v_{tar}(t, \dots, t-m))$$

A time-delay matrix, Y , is created, where each row contains one of the extend sets. Applying a Singular Value Decomposition (SVD) to that time-delay matrix:

$$Y = P \Lambda S \quad (3)$$

Where P is the matrix of Principal Components, Λ the diagonal matrix of singular values, and S the matrix of singular vectors. Each column of S contains a singular vector describing the empirical pattern of behaviour across the four variables over the selected window. The corresponding diagonal entry in Λ contains the mean amplitude of that pattern contributing to the measurements, and the corresponding column in P contains the normalised amplitude of the pattern at each instance in time.

Applying this SVD to the delay matrix created from the set of concurrent data from the measurement campaign at the target site and the data available from the reference site determines the values for Λ and S . The standard SVD implementations in analysis packages such as MATLAB

or R return these matrices ordered in descending order of Λ . That means that the first singular vector contributes most to the variance of the measurements. By truncation the Λ and S matrices to r columns only, the system is truncated to retain only the r strongest pattern, effectively removing small-scale localised fluctuations but maximising the coherent behaviour across the two sites. These truncated Λ_r and S_r can then be used to calculate new principal components, P_n from a delay matrix consisting of new measurements, $P_n = Y_n S_r^T \Lambda_r^{-1}$, where ‘ T ’ indicates the transpose.

The new aspect of the method developed here stems from the fact that in MCP, the historical information is only from the reference site and we are seeking the corresponding information for the target site. In practical terms, this means that we can only fill the first set of columns of Y_n , say Y_h and that we can only use the first half rows of S_r , say S_h , and the first half of Λ_r , Λ_h :

$$P_n = Y_n S_r^T \Lambda_r^{-1} \quad (4)$$

However, as each row in S or S_r or S_h contains information about all four sites, applying eq.(4) to the historical data only activates the corresponding target components, and we can estimate the target values as

$$Y_e = P_p \Lambda_h S_h = Y_n S_r^T \Lambda_r^{-1} \Lambda_h S_h \quad (4)$$

From this estimated time-delay matrix, the target velocities can then be extracted as the appropriate column vectors.

3. The data source and MCP application

A. Data Source

A typical (concurrent) data measurement period used is a year or more [7]. Our data set contains twelve years from 1999 to 2010 of wind speed and direction measurements from eight Met. Office stations across Scotland through the British Atmospheric Data Centre [8], listed in Table I, of which seven cover a fairly narrow latitude range but extending from the western island of Islay to Edinburgh at the east coast, and another site from the far north-west on the Isle of Harris and Lewis. All data were recorded with anemometers at 10 m above ground at an hourly interval with the data stored in *knots*, rounded to the nearest ± 1 kn but converted to m/s for this analysis.

Table I. – Anemometer locations

No	Location	Latitude	Longitude
1	Stornoway	58.2138	-6.31772
2	Port Ellen	55.6813	-6.24866
3	Machrihanish	55.4408	-5.69571
4	Prestwick	55.5153	-4.58343
5	Glasg airport	55.9068	-4.53122
6	Salsburgh	55.8615	-3.87409
7	Edinb airport	55.9284	-3.34294
8	Edinburgh south	55.9228	-3.18750

A. MCP settings

The analysis covered all possible pairs of stations, treating each both, as target and reference site. Two different lengths of training periods of concurrent data were used, a single calendar year and two consecutive calendar, and then using in turn each possible year or 2-year period. In the analysis presented here, the window length to create the delay matrix covered either a 24h or a 48 period, and truncation level ranging from 3 to 12 retained singular vectors were explored, leading to 896 different permutations explored in the analysis.

Based on the window length to create the delay matrix and the number of retained singular values and vectors, the resource at each target site was predicted for the entire twelve-year period.

C. Analysis method

To quantify the degree to which the wind resource is estimated correctly, we chose to calculate the error in the wind speed distribution obtained from the MCP method. To do this, first the distribution from the predicted resource was determined as the histogram of predicted wind speeds in bins of 1 m/s, then the absolute difference between the likelihood of a particular wind speed bin to that actually observed in the full record was calculated and finally, the errors for each wind speed bin were added to calculate a total absolute error.

From the total absolute error, a performance index was defined as the ratio of the total error from that particular PCA-MCP prediction over that from the linear regression. A performance index of 1 implies that both methods perform equally well, but a performance index < 1 implies that the PCA-MCP is better than linear regression.

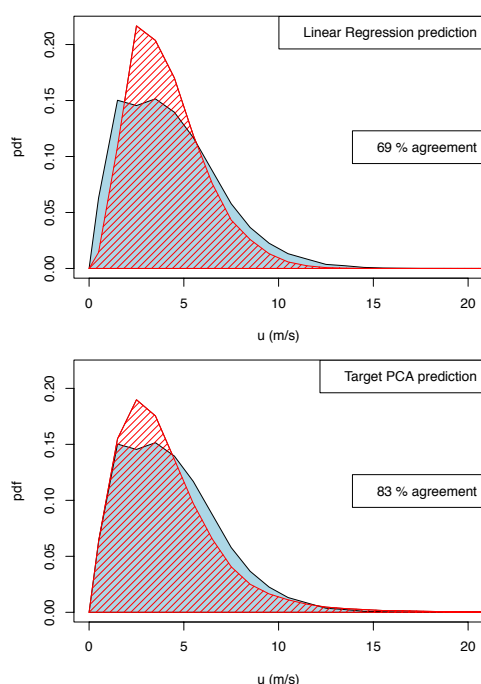


Figure 1 – Actual (blue shaded) and predicted (red cross-hatched) wind speed distributions at the target site using linear regression (upper frame) and the PCA-MCP (lower frame).

4. Results

A. Wind resource predictions

Figure 1 shows the example of predicting the wind speed at Edinburgh Airport from the wind speed at Salsburgh, around 30 km to the West based on a training period of 2009-2010. In both figures, the predictions are the red cross-hatched distributions, overlaid over the actual distribution (blue shaded region). It is clear that the basic MCP using linear regression substantially overestimates the wind speed bins between 3 and 6 m/s and underestimates the lower wind speeds as well those between 6 and 14 m/s. While the PCA-MCP also overestimates the moderate wind speeds, it does so to a much lesser degree and only underestimates the wind speed range from 5 to 10 m/s. This substantial improvement is reflected in the degree of overlap with the actual distribution of 83% compared to an overlap of only 69% for the linear regression. Expressing the overlap as the absolute error between the distributions, they translate to an error of $e_{LR} = 1 - 0.69 = 0.31$ for the linear regression and $e_{PCA} = 1 - 0.83 = 0.17$ for the PCA-MCP, giving a performance index of $PI = e_{PCA}/e_{LR} = 0.55$. That means that the prediction error has been reduced to 55% of that made by using linear regression

B. Performance test

Exploring the PCA-MCP and comparing against the linear regression for all possible pairs of stations from Table I, and using the set of window lengths and truncation levels listed in §3.A for the PCA-MCP, leads to an overall pattern of the performance index as shown in Figure 2. This shows that the prediction error through the PCA-MCP is in most of the 896 cases between 10 and 50% of that from the linear regression. It is extremely rare that PCA-MCP is no better, or even slightly worse, than linear regression.

Initial thoughts that the calculation of the PCA-MCP prediction of the reference site itself might give a measure of the expected prediction error at the target site appears not to hold, and current further work is aimed at determining a rational method to identify the best window length and truncation.

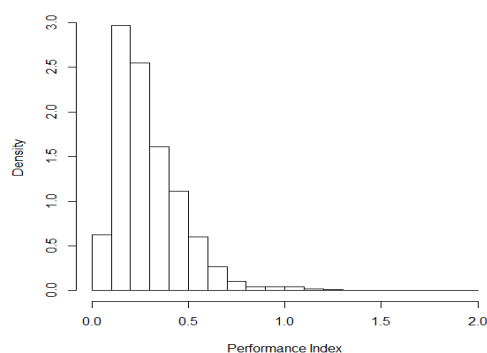


Figure 2 – Histogram of Performance indices, PI , achieved by PCA-MCP against linear regression for all pairs of stations and PCA-MCP settings explored.

4. Conclusion

In this paper, we developed a new wind resource assessment method, taking the concept of the traditional measure-correlate-predict (MCP) and developing it using a time-series analysis technique from Dynamical Systems (Chaos) Theory. In that approach, the formalism determines the shape and coefficients of the best relationship between the target and a reference site by treating the measurements as representative of the joint dynamical system, rather than one as input and the other as output.

The work has progressed so far to demonstrate on a variety of station pairs, some several hundred kilometres apart, that this method is almost always superior to the basic standard MCP using linear regression.

The next stage of the work is to subject it to a systematic analysis to identify if it is possible to judge the quality of the prediction at the target site from the information available to the analyst. The PCA-MCP method using the long-term reference data not only predicts the wind speeds at the target site but also returns a new estimate of the reference site data. The initial idea that this reference estimate might give clues about the quality of the target prediction could not be substantiated. Other possible quantities to test in the next stage of development are the estimates returned from applying the truncated PCA-MCP predictor to the data from the reference site for the training period and thus predicting the target wind speed for the training period.

Acknowledgement

We are grateful for the support of this work through SgurrEnergy Ltd. and the Energy Technology Partnership of the Scottish universities and Scottish Government. The authors would like to thank the UK Meteorological office for providing access to the MIDAS record wind data through the British Atmospheric Data Centre.

References

- [1] A. Dinler, "A new low-correlation MCP (measure-correlate-predict) method for wind energy forecasting", in *Energy*, vol. 63, pp. 152-160, 2013.
- [2] J. A. Carta, S. Velázquez, and P. Cabrera, "A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site" in *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 362-400, 2013.
- [3] D. S. Broomhead, R. Jones, J. P. King, and E. R. Pike, *Singular Systems Analysis with application to dynamical systems*: Adam Hilger, 1987.
- [4] P. L. Read, "Phase portrait reconstruction using multivariate singular systems analysis", in *Physica D*, 69:353-365, 1993.
- [5] C. Skittides and W. Früh, "Wind speed forecasting using singular systems analysis", in *International Conference on Renewable Energies and Power Quality (ICREPQ'13)*, volume 11. *Renewable Energy and Power Quality Journal*, 2013.
- [6] C. Skittides and W.-G. Früh, "Wind forecasting using principal component analysis", in *Renewable Energy*, 69:365 - 374, 2014.
- [7] M. A. Lackner, A. L. Rogers, and J. F. Manwell, "Uncertainty analysis in MCP-Based wind resource assessment and energy production estimation", in *Journal of Solar Energy Engineering-Transactions of the ASME*, vol. 130, 2008.
- [8] British Atmospheric Data Centre (BADC). Available at: www.badc.ac.uk