DTU Library

Application of the measure-correlate-predict approach for wind resource assessment

Nielsen, M.; Landberg, L.; Mortensen, N.G.; Barthelmie, R.J.; Joensen, A.

Published in:
Wind energy for the new millennium. Proceedings

Publication date:
2001

Document Version
Early version, also known as pre-print

Application of the Measure-Correlate-Predict Approach for Wind Resource Assessment

Morten Nielsen[1], Lars Landberg[1], Niels G. Mortensen[1], Rebecca J. Barthelmie[1], Alfred Joensen[2]
[1]Risø National Laboratory, Wind Energy Department, DK-4000 Roskilde, Denmark
[2]WaveTech ApS, Store Kongensgade 81A, DK-1264 Copenhagen, Denmark
n.m.nielsen@risoe.dk, niels.g.mortensen@risoe.dk lars.landberg@risoe.dk, r.barthelmie@risoe.dk, aj@wavetech.dk

ABSTRACT: The correlation between wind measurements depends on the separation of the meteorological stations and the temporal resolution of the time series. This is shown by the band-limited correlation coefficient, calculated by a Fourier-based technique, and by examples of linear regression models for seven years of data from three Danish stations. Estimated long-term statistics based on regression models and a long record from the reference site becomes uncertain for data-overlap periods shorter than a year. The wind direction is harder to predict than the wind speed.
Keywords: Correlation; Averaging time; Meteorology; Resources;

1 INTRODUCTION

Accurate prediction of the wind resource is a key issue in the planning of new wind farms. If the investment is sufficiently large it may be beneficial to install a meteorological station to monitor the local wind conditions. The project will usually allow a limited measuring period only. This puts the wind engineer in a dilemma, since wind statistics are known to have significant variations - even on an annual basis. In this situation representative long-term statistics may have to be estimated by long-term measurements from a nearby reference station. A standard approach, known as Measure-Correlate-Predict (MCP) [1-4], is to establish a correlation between the two sets of measurements using data from the overlap period where observations from both stations are available. The next step is to use this stochastic model to estimate a long artificial time series, which both represent the local conditions and the general long-term variations. Exact match at any time is not required; the main issue in relation to wind resource assessment, is whether the statistics of the predicted time series are representative. In general, the success will depend on details of the stochastic model and the degree of correlation between wind conditions at the two sites.

Accurate wind-speed prediction is essential for the power prediction and in most cases the wind direction will also be quite important, e.g. when predicting production loss due to wake effects in wind turbine parks. The MCP model is usually established by linear regression for a number of wind-direction sectors at the reference site. In a study of data from the West Coast of UK [1] it was found that the inclusion of a deterministic diurnal model component led to insignificant improvement, although such an approach might be successful in other parts of the world. In cases where the observation heights of the two sites differ, a measure of atmospheric stability, such as the temperature gradient at the reference site, could improve the prediction [3]. This extension is not always advisable, e.g. when the new site is offshore and the reference site is onshore – surface temperatures differences often makes the atmospheric stabilities of the marine and onshore environments quite different [5]. Considerable directional veer may be observed in complex terrain, in particular when one of the sites is situated in a valley or on top of a ridge. This effect is included, either by an additional regression model for the wind veer [1-3], or by the matrix MCP method [4]. In the matrix MCP method, the data are sorted according to combinations of wind directions at both sites, usually in a 12×12 matrix, and predictions are based on wind-veer probabilities given a particular wind direction at the reference site.

The purpose of these model variants is to improve the correlation as far as possible. The optimal number of wind-direction sectors may be a compromise between the need for high directional resolution and the better statistical accuracy, which is obtained by a larger number of observations in wider sectors. It may also be relevant to remove low-wind situations from the data set, since wind direction is more unpredictable in weak winds.

The correlation between observations at two sites depends on their spatial separation. Variations in the atmospheric boundary layer and the effects of meso-scale systems, e.g. sea breezes, will decorrellate with increasing separation. Derrick [1] recommends that variations faster than the dirurnal cycle should be filtered out before correlation of observations with spatial separations on the order of 50 km. A moving average filter did not always solve the problem, since the wind near a front and low-pressure centre has significant spatial variation.

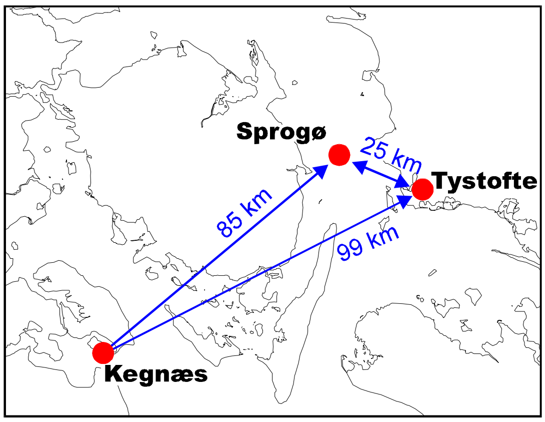

Figure 1: Meteorological stations.

Table 1: Meteorological stations.

	Tystofte	Sprogø	Kegnæs
Location	Inland	Island	Peninsula
Dir height	36m	70m	10m
Speed height	39m	70m	23m
Data coverage	98.5%	99.7%	95.7%

2 MEASUREMENT

The time series selected for numerical tests in this paper were measured during the period 1991-1997 at three stations in the inner Danish waters, see Figure 1. One of the stations is situated on a small island (Sprogø) representative for the offshore wind climate in the Great Belt. The second stations is situated somewhat inland 25 km away from Sprogø (Tystofte), and the third one is on a peninsula near an inlet 85 km away from Sprogø (Kegnæs). The masts had similar equipment, however with variable measurement heights, as read in table 1. The data coverage is >95% at all sites.

3 CORRELATION

3.1 The significance of time resolution

Before the MCP analysis it is relevant to consider the possible need for time averaging of the data. The ultimate goal is to find reliable predictions of short-time statistics applicable for the wind resource assessment (~10 min), but unless the stations are closely situated, variations on short time scales are unlikely to correlate. Ayotte et al. [6] analysed data in rolling-hill terrain in Southeast Australia and suggested a useful analysis of the temporal and spatial correlation problem.

Consider first the discrete Fourier transformation of each of time series

$$U(f) = \sum_{n=0}^{N-1} u_n e^{2\pi i f n/N}, \quad f = 0 \ldots N/2 \tag{1}$$

where the wind vector is modelled as a complex number, $u = x + iy$, in which the real part and imaginary parts are the east-west and north-south wind components. Typical meteorological time series miss some data points, and in general, we must be prepared to calculate the sum in Eqn. 1 directly instead of using the efficient FFT algorithm. Figure 2 shows normalised cumulated power spectral for the selected data set. These distributions are very similar for the three sites and 90-92% of the energy is seen to relate to periods longer than one day.

Ayotte et al. [6] defined the band-limited correlation coefficient

$$R(F) = \frac{\sum_{f=1}^{F} U_1(f) U_2^*(f)}{\sqrt{\sum_{f=1}^{F} U_1(f) U_1^*(f) \cdot \sum_{f=1}^{F} U_2(f) U_2^*(f)}} \tag{2}$$

where * denotes complex conjugation. *R(F)* is seen to be a measure of the correlation of low-pass filtered time series, which disregard variations beyond the cut-off frequency *F*. Note that the sum starts at index 1 rather

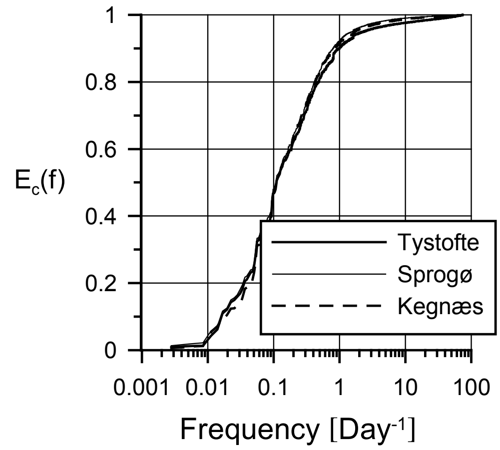

Figure 2: Normalised cumulative spectral energy.

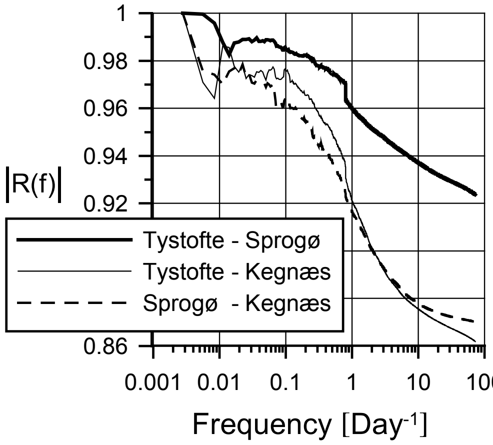

Figure 3: Modulus of the band-limited correlation coefficients.

than 0, which in time domain corresponds to subtraction of the average value. Figure 3 shows that the modulus of the band-limited correlation coefficient is close to unity for low frequencies and declines with increasing frequency. For a given time resolution the correlation improves when the distance between the stations is short. This result is in agreement with fig. 7 of the original article [6], where the one-hour band-limited correlation coefficient varied from $R(F) = 0.83$ for 104-km separation over $R(F) = 0.94$ for 22-km separation up to $R(F) = 0.99$ for 1.4-km separation.

The time series correlation coefficient is an upper limit of the correlation within the reach of regression models. For large spatial separation between prediction and reference site, there will be more successful regression model for data with longer averaging time. The disadvantage of the increased averaging time is, however, that a smaller fraction of the spectral energy is resolved, as shown by figure 2.

The authors [6] proceeded by the calculation of the phase of the cross-correlation of a given Fourier mode.

$$\varphi = \mathrm{Arg}\left(U_1(f) U_2^*(f)\right) \tag{3}$$

An example of this measure is seen in figure 4, which shows cross-correlation phases for Sprogø and Tystofte.

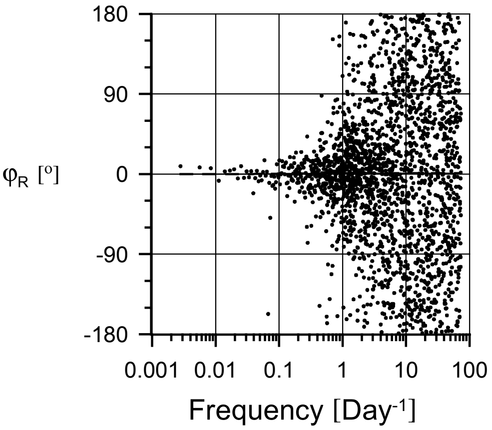

Figure 4: Cross-correlation phases for Sprogø and Tystofte. For clarity, some points have randomly been removed from the high-frequency end of the diagram.

Table 2: Standard variation of the cross-correlation phases shown in figure 4.

Frequency [Day^{-1}]	0.2-0.6	0.6-2	2-6	2-20
Variation, σ_φ	33°	52°	76°	98°

As expected, the cross-correlation phases are small for low frequencies and scattered for high frequencies, see standard deviation in Table 2. Though some outliers may be associated with Fourier modes of insignificant amplitude, Fig. 4 does indicate that correlation of short-term wind direction may be difficult.

3.2 Regression models

The wind at the new site is usually related to the reference-site wind by a linear regression model.

$$s_p = a_s + b_s s_r \tag{4}$$

Here s_r is the reference-site wind speed and the regression coefficients (a_s, b_s) depend on the reference-site wind direction θ_r. The simplest way to manage the directional dependence is to divide the data according to the wind direction at the reference station, and construct a linear regression model for each class. Similarly, the model for wind veer is expressed by $\Delta\theta_p = a_\theta + b_\theta \Delta\theta_r$, where $(\Delta\theta_p, \Delta\theta_r)$ are deviations from the mid-angle of the sector. The performance of these regression models could be evaluated by the product-moment correlation coefficient or Pearsons's r [7] defined by

$$r = \frac{N\sum \xi_r \xi_p - \sum \xi_\mathrm{r} \sum \xi_\mathrm{p}}{\sqrt{\left[N\sum \xi_r^2 - \left(\sum \xi_\mathrm{r}\right)^2\right] \cdot \left[N\sum \xi_p^2 - \left(\sum \xi_\mathrm{p}\right)^2\right]}} \tag{5}$$

where ξ could mean either speed s or direction $\Delta\theta$.

As an alternative, one may fit a vector model of the form $u_p = a + B\, u_r$, i.e. a linear vector relation of the form

$$\begin{bmatrix} x_\mathrm{p} \\ y_\mathrm{p} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_\mathrm{r} \\ y_\mathrm{r} \end{bmatrix} \tag{6}$$

where $(x_\mathrm{P}, y_\mathrm{P})$ and $(x_\mathrm{r}, y_\mathrm{r})$ are velocities in orthogonal co-ordinates. The least-square fit of the vector-based model is:

$$\begin{bmatrix} a_1 & a_2 \\ b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \tag{7}$$

$$\begin{bmatrix} N & \sum x_\mathrm{r} & \sum y_\mathrm{r} \\ \sum x_\mathrm{r} & \sum (x_\mathrm{r})^2 & \sum y_\mathrm{r} y_\mathrm{r} \\ \sum y_\mathrm{r} & \sum x_\mathrm{r} y_\mathrm{r} & \sum (y_\mathrm{r})^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum x_\mathrm{p} & \sum y_\mathrm{p} \\ \sum x_\mathrm{r} u_\mathrm{p} & \sum x_\mathrm{r} v_\mathrm{p} \\ \sum y_\mathrm{r} u_\mathrm{p} & \sum y_\mathrm{r} v_\mathrm{p} \end{bmatrix}$$

where each sum contains N elements. Similar to Eqn. 2, the correlation coefficient for the 2-D model is defined by

$$r = \frac{N\sum u_r u_p^* - \sum u_\mathrm{r} \sum u_p^*}{\left\{\sqrt{\left[N\sum u_r u_r^* - \sum u_\mathrm{r} \sum u_r^*\right]} \cdot \sqrt{\left[N\sum u_p u_p^* - \sum u_\mathrm{p} \sum u_p^*\right]}\right\}} \tag{8}$$

with $u = x + iy$. Formally, this regression coefficient is in itself a complex number, but the imaginary part is usually small. Exceptions from this rule would imply a systematic time delay between the events at the two stations.

3.3 Model performance

Fig. 5 shows the performance of the models when correlating the winds at Sprogø and Tystofte, i.e. offshore and onshore with 25-km distance. The data are from the period 1991-1997 and divided according to the wind direction sector at Tystofte. The analysis is repeated for time series with different averaging times and correlation is seen to improve for longer periods. The simple speed correlation model has a better performance than the vector model, especially for the shorter averaging time. The correlation of the wind directions is surprisingly poor. In applications where an accurate direction model is of high importance, one would have to use long averaging times and/or fewer wind sectors.

Fig. 6 shows the correlation between data from Sprogø and Tystofte (25 km) and Sprogø and Kegnæs (85 km), for time series with 3-hour averaging time. As expected, the correlation is best over the short distance. The values is in good agreement with the values in Fig. 3, which for the frequency of 12 day^{-1} are 0.94 for the 25-km distance and 0.88 for the 85-km distance

4 PREDICTION

In practice the regression is established by a relatively short data overlap period and the long-term series is predicted from a long time series at the reference station. The significance of a long data

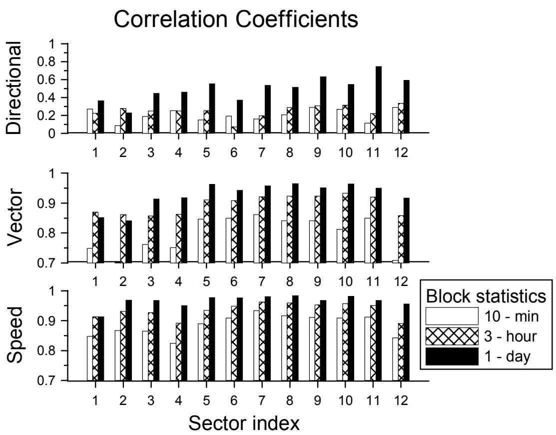

Figure 5: Correlation coefficients for regression models predicting winds at Sprogø by the wind at Tystofte. The analysis is repeated for time series with variable averaging times.

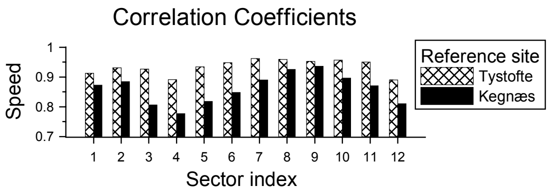

Figure 6: Correlation coefficients for regression models predicting winds at Sprogø by the wind at Tystofte (25km) or Kegnæs (85km).

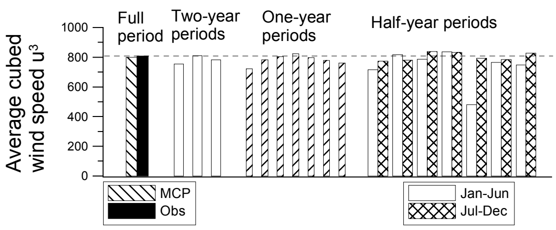

Figure 7: Time averages of the third power of wind speeds at Sprogø. The solid bar represents the observed full-period average value. The rest of the bars represent predictions by the MCP method using data from Tystofte 25km and variable data overlap.

overlap has been tested by truncated time series of variable duration [1,4] and the general recommendation is to use at least one year. Obviously, the reference period should contain a fair amount of data in every wind sector of the reference site. The one-year rule might also be the conclusion from the results shown in Fig. 7. The MCP estimates based on half-year periods may be in fortuitous good agreement with those of the full record, but there is a risk of false predictions. The MCP model reproduces the mean wind speed, but with a correlation coefficients less than unity, the higher order moments are predicted with a negative bias, e.g. the full-period MCP prediction ($\overline{u^3} = 800\ \mathrm{m^3/s^3}$) is a little less than the value based on corresponding observations ($\overline{u^3} = 808\ \mathrm{m^3/s^3}$).

ACKNOWLEDGEMENTS

Various colleagues at Risø maintained the measuring stations mentioned in this work. Gunnar Jensen scrutinised and organised the data.

REFERENCES

[1] A. Derrick, In: EWEC 93 proceedings, H.S. Stephens & Ass. UK, Eds.: Garrad, Palz and Scheller (Eds.), (1993), 681-685.

[2] J. C. Woods and S. J. Watson, J. Wind Eng. Ind. Aerodyn. (1997) **66**, 85-94

[3] A. Joensen, L. Landberg, and H. Madsen, In: EWEC 99 proceedings, James and James Science Publishers, UK, Eds.: E.L. Petersen, P. H. Jensen, K. Rave, P. Helm and H. Ehmann (1999) 1157-1160.

[4] J. R. Salmon and J. L. Walmsley, Wind Eng. Ind. Aerodyn. (1999) **79**, 233-268

[5] R.J. Barthelmie, Meteorol. Appl., (1999) **6**, 39-48.

[6] K. W. Ayotte, R. J. Davy, P. A Coppin, Boundary-Layer Meteorol. (2001) **98**, 275-295

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes in C, Second edition, Cambridge University Press, UK (1992).