WIND ENERGY
Wind Energ., **1**, 29–34 (1998)

Research Article

A New Reference for Wind Power Forecasting

Torben Skov Nielsen, Alfred Joensen* and Henrik Madsen, Institute of Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark
Lars Landberg and Gregor Giebel, Department of Wind Energy and Atmospheric Physics, Risø National Laboratory, Building 125, DK-4000 Risø, Denmark

Key words: persistence; correlation; wind power; reference forecast model

In recent years some research towards developing forecasting models for wind power or energy has been carried out. In order to evaluate the prediction ability of these models, the forecasts are usually compared with those of the persistence forecast model. As shown in this article, however, it is not reasonable to use the persistence model when the forecast length is more than a few hours. Instead, a new statistical reference for predicting wind power, which basically is a weighting between the persistence and the mean of the power, is proposed. This reference forecast model is adequate for all forecast lengths and, like the persistence model, requires only measured time series as input. ©1998 John Wiley & Sons, Ltd.

Introduction

In this article we propose a new reference model which should be used instead of the persistence model (1) when short-term, say up to 48 h, forecasting models for wind power or energy are evaluated.

There are two types of wind power forecasting models: physical[1–3] and statistical.[4–6] Up to now the reference for these models and many other meteorological forecasting models has been the persistence model given by

$$p_{t+k} = p_t + \varepsilon_{t+k} \tag{1}$$

where t is a time index, k is the look-ahead time, p is e.g. wind power or energy and ε denotes the residual. The forecast $\hat{p}$ obtained using this model is

$$\hat{p}_{t+k} = p_t \tag{2}$$

which states that the expected value k time steps ahead is equal to the most recent value. In statistics this is called the persistence or naive predictor. In this article we shall refer to (1) as the persistence forecast model.

The model (2) is a simple description, but yet very powerful. This is because the atmosphere can be considered quasi-stationary, i.e. changing very slowly. A characteristic timescale in the atmosphere is f^{-1}, where f is the Coriolis parameter. Using $10^{-4}\,\mathrm{s}^{-1}$ for f gives that this time scale is approximately 3 h.[2]

To compare the forecasts with the observations, the root mean square error (RMS) or the mean square error (MSE) is usually used. The MSE for the persistence forecast model is given by

$$MSE_{\mathrm{p}} = \frac{1}{N-k}\sum_{t=1}^{N-k}(p_{t+k} - \hat{p}_{t+k})^2 = \frac{1}{N-k}\sum_{t=1}^{N-k}(p_{t+k} - p_t)^2 \tag{3}$$

*Correspondence to: A. Joensen, Institute of Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark.

CCC 1095–4244/98/010029–06$17·50

where N is the number of observations. The RMS is given by

$$RMS_{\mathrm{p}} = \sqrt{MSE_{\mathrm{p}}} \tag{4}$$

Owing to the quasi-stationarity of the atmosphere, p_{t+k} will be rather close to p_t when the time step k is less than a few hours, which means that the MSE will be small compared with the MSE for large k.

As k gets larger, $k \gg f^{-1}$, or say above 36 h, the flow in the atmosphere will no longer remain constant and the correlation between p_{t+k} and p_t will tend to zero. This means that the present flow provides no information about the future flow, and the model (1) which correlates the future flow to the present flow is no longer reasonable.

Instead, the mean of the flow could be used as a simple reference when the correlation is zero. In the Appendix it is shown that the MSE for the persistence is actually twice the MSE of the mean predictor when the correlation is zero.

It is thus quite obvious to suggest a new reference forecast model as a weighting between the persistence and the mean, where the weighting for different forecast lengths is determined by the correlation between p_t and p_{t+k}. In this article such a reference is proposed. Wind power is considered, but the proposed reference can be used for many other meteorological quantities, e.g. wind speed or energy.

New Reference Forecast Model

As outlined in the Introduction, the proposed reference forecast model is a weighting between the persistence and the mean, i.e. the k-step forecast is written as

$$\hat{p}_{t+k} = a_k p_t + (1 - a_k)\bar{p} \tag{5}$$

where p_t is the most recent measurement of the wind power and $\bar{p}$ is the estimated mean of the power given by

$$\bar{p} = \frac{1}{N}\sum_{t=1}^{N} p_t \tag{6}$$

When k is small, a_k should be approximately one and the reference thus corresponds to persistence, but when k is large and the correlation is zero, a_k should be zero and the forecast is simply the mean. It is thus reasonable to define a_k as the correlation coefficient between p_t and p_{t+k}:

$$a_k = \frac{\frac{1}{N}\sum_{t=1}^{N-k} \tilde{p}_t \tilde{p}_{t+k0}}{\frac{1}{N}\sum_{t=1}^{N-k} \tilde{p}_t^2} \tag{7}$$

where

$$\tilde{p}_t = p_t - \bar{p} \tag{8}$$

This actually corresponds to the value of a_k which minimizes the MSE for the new reference.

Examples

In this section, measured wind power is used to calculate the correlation, and the *RMS* for the new reference is compared with the *RMS* for the mean and persistence.

Correlation

Measurements of half-hourly mean values of wind power from a wind farm located in Hollandsbjerg, Denmark have been used to calculate an estimate of the correlation as a function of the forecast length. Two data sets are considered, namely measurements from a summer and a winter period. Each data set contains 4380 measurements. The estimated correlation as a function of the forecast length from the summer period is shown in Figure 1 and from the winter period in Figure 2.

From both figures it is seen that the correlation seems to decrease exponentially as a function of the forecast length. Therefore the figures also show the values of the function

$$f(k) = \phi^k \tag{9}$$

where the values used for ϕ are the estimated correlation coefficients for $k = 1$.

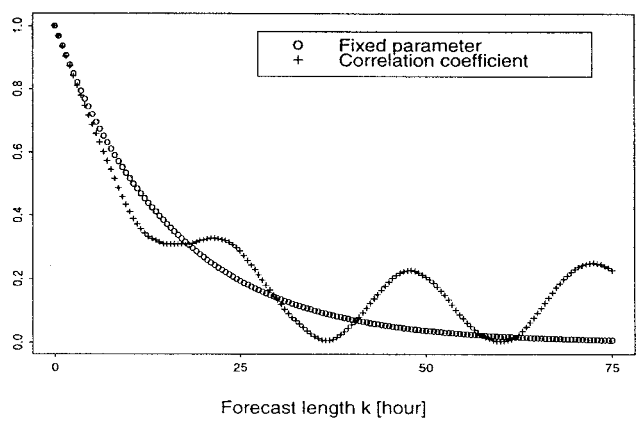

Figure 1. Estimated correlation as a function of the forecast length for 4380 half-hourly mean values of observed wind power in a summer period, and values of the fixed parameter function

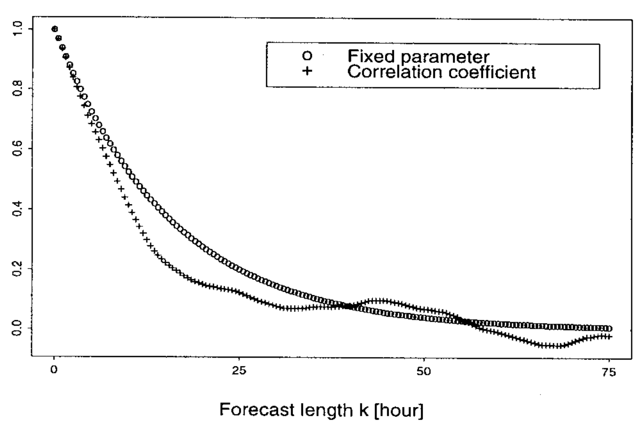

Figure 2. Estimated correlation as a function of the forecast length for 4380 half-hourly mean values of observed wind power in a winter period, and values of the fixed parameter function

The correlation for the half-hourly forecast ($k = 1$) is 0·968 for both periods and the agreement between $f(k)$ and the correlation is good for both periods as long as the forecast length is small. However, for the summer period the correlation is seen to be highly periodic owing to the diurnal variation in the wind speed, because in latitudes such as Denmark's this diurnal variation is most significant during the summer period.

Thus the correlation is not independent of the location of the wind farm or the time of year. Therefore it is not possible to use a simple expression like (9) or to assume global values for the correlation. It is thus recommended that the correlation is calculated for each forecast length using (7) and (8) and that the correlation which is calculated using measurements from a given location should not be used for any other locations.

Performance

In this subsection the measurements from Hollandsbjerg are used to show how the RMS of the forecast error depends on the forecast length. One year of half-hourly mean values of the power are used and the RMS is calculated using the new reference, the persistence and the mean of the power. The result is shown in Figure 3.

The figure clearly demonstrates the need for a new reference forecast model, since the RMS for the persistence model for large horizons is larger than the RMS obtained using the mean value as a forecast. For small forecast lengths, $k \leqslant f^{-1} \approx 3\,\text{h}$, the RMS for the new reference is almost identical to the RMS for the persistence forecast model; and for larger horizons, say $k > 24\,\text{h}$, the RMS for the new reference approximates the RMS of the mean. For the intermediate horizons it is clearly seen that the new reference combines the forecasts from the persistence and the mean in such a way that the RMS is significantly below the RMS of these last two approaches.

Summary

In this article we have proposed a new reference forecast model for predictions related to wind speed and power. This reference should be used instead of the commonly used persistence forecast model, which is shown not to be reasonable for forecast lengths above a certain limit. The algorithm for calculating predictions from the new reference model is summarized below.

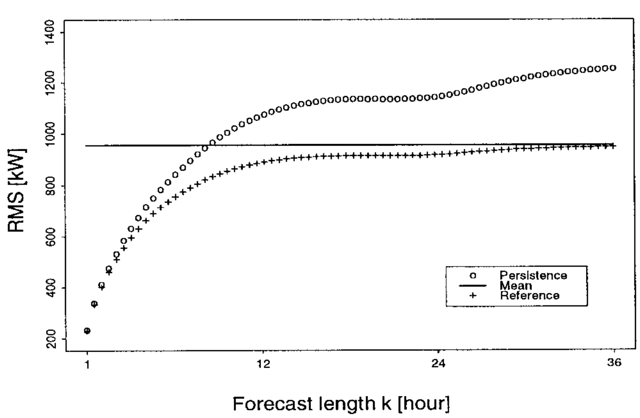

Figure 3. RMS for the three simple forecast models—the mean, the persistence and the new reference—calculated using one year of half-hourly mean values of measured wind power

- Calculate the mean $\bar{p}$ using (6).
- For each forecast length k:
  —calculate the correlation coefficient a_k using (7);
  —calculate the predictions $\hat{p}_{t+k}$ from the reference forecast model using (5).

The main difference between this algorithm and the persistence forecast model is that the correlation coefficient has to be calculated for each forecast length. If the correlation were the same all over the world, i.e. not depending on the location of a wind farm, the above algorithm could be simplified by omitting the calculation of the correlation coefficient. In this case the correlation coefficients could be given in a table, which could be considered globally valid. However, the results in the previous section indicate that this is not the case.

The new reference forecast model is still almost as simple as the persistence forecast model, since it only requires time series of measured wind power as input. It is clearly demonstrated that if the forecast length k is larger than $f^{-1} \approx 3\,\text{h}$, then the new reference should be used.

Appendix: Mean Square Error (MSE)

Here it is shown that the *MSE* for the persistence forecast model is twice the *MSE* if the mean is used as a forecast model when the flow can be considered uncorrelated.

The *MSE* given by (3) can be rewritten as

$$MSE_{\mathrm{p}} = \frac{1}{N-k}\left(2\sum_{i=k+1}^{N-k} p_i^2 + \sum_{i=1}^{k} p_i^2 + \sum_{i=N-k+1}^{N} p_i^2 - 2\sum_{i=1}^{N-k} p_i p_{i+k}\right) \tag{10}$$

As the number of observations $N \to \infty$ and $k \ll N$, it is seen that the second and third sums in (10) become negligible and hence

$$MSE_{\mathrm{p}} \approx \frac{2}{N-k}\left(\sum_{i=k+1}^{N-k} p_i^2 - \sum_{i=1}^{N-k} p_i p_{i+k}\right)$$

Using that the mean of two multiplied uncorrelated random variables X and Y is given by $E(XY) = E(X)E(Y)$, the *MSE* for large k can be rewritten as

$$MSE'_{\mathrm{p}} \approx \frac{2}{N-k}\left[\sum_{i=k+1}^{N-k} p_i^2 - \frac{1}{N-k}\left(\sum_{i=1}^{N-k} p_i\right)^2\right]$$

If instead the mean of the flow is used as a forecast model, i.e.

$$\hat{p}_{t+k} = \bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i, \quad 1 \leqslant t \leqslant N$$

we see that the *MSE* for this model is

$$MSE_{\mathrm{m}} = \frac{1}{N}\sum_{i=1}^{N}\left(p_i - \frac{1}{N}\sum_{j=1}^{N} p_j\right)^2 = \frac{1}{N}\left[\sum_{i=1}^{N} p_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} p_i\right)^2\right] \approx \frac{1}{2} MSE'_{\mathrm{p}}$$

which means that the *MSE* for the persistence model will be twice the *MSE* of the mean model for large k, where p_{t+k} and p_t are uncorrelated.

References

1. L. Landberg, 'Short-term prediction of power production from wind farms', *J. Wind Engng. Ind. Aerodyn.*, submitted (1998).
2. L. Landberg, *Short-term Prediction of Local Wind Conditions*, Risø-R-702(EN), Risø National Laboratory, Roskilde, 1994.
3. L. Landberg and S. J. Watson, 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorol.*, **70**, 171–195 (1994).
4. A. Joensen, *Models and Methods for Predicting Wind Power*, Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, 1997 (in Danish).
5. T. S. Nielsen and H. Madsen, *Using Meteorological Forecasts in On-line Predictions of Wind Power*, Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, 1996.
6. H. Madsen, *Models and Methods for Predicting Wind Power*, Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, 1996.