METMA V International Workshop on Spatio-Temporal Modelling
Santiago de Compostela
30 Jun - 2 Jul 2010

Reading the Message of the High Frequency Wind Data. A Short-Term Wind Speed Forecast with Periodic Components and Long Range Dependence

G. Beblo[1] and W. Schmid[2]

[1] *Department of Statistics, European University Viadrina PO Box 1786, 15207 Frankfurt (Oder), Germany. Beblo@euv-frankfurt-o.de*
[2] *Department of Statistics, European University Viadrina PO Box 1786, 15207 Frankfurt (Oder), Germany. Schmid@euv-frankfurt-o.de*

Abstract. *Accurate short-term wind forecasts are crucial for the whole energy industry that is currently moving toward clean and renewable energy sources like, e.g., the wind energy. These forecasts can be used to optimize the performance of wind farms. They are also very helpful for planning purposes that aim at smoothing the overall electricity supply. In this work, we present a short-term wind speed forecast using 10-minute high frequency data collected in Manschnow, Germany. To model the wind speed we use the information about its daily and monthly patterns extracted from the data and we apply an ARFIMA (p,d,q) process for modelling its persistent nature. As a result we obtain a model that beats the standard persistence benchmark at one hour time horizon.*

Keywords. *High frequency data; Long range dependence; Time series regression; Wind speed; Environmental statistics.*

1 Introduction

Wind power could be a perfect solution to the problem of world's increasing energy needs. It is pollution-free, renewable and inexpensive in generation, once the wind farm has been installed. However, wind fluctuations resulting in varying electricity generation make that aim difficult to reach. Therefore accurate and reliable short-term wind forecasts are so important to level the capacity of electric utilities and to create secure energy supplies. In fact, it is possible to predict the wind power directly, however, it will always strongly depend on the wind turbines used in a certain wind farm (their specifications, capacity, size etc., see, e.g., Genton and Hering (2009)). Because the wind power output can be directly calculated from the wind speed it seems to be more natural to forecast the wind speed. Numerical Weather

Prediction models fulfill that task very well for forecasting periods exceeding 6 hours (cf. Giebel et al. (2003)). For shorter time horizons statistical methods outperform them significantly (e.g. autoregressive time series methods). Such short-term wind forecasts are frequently used to optimize the performance of the wind farms. They provide a valuable piece of information to wind energy investors on the stock exchange as well.

Our aim is to forecast the wind speed by taking into account daily and monthly factors as well as the wind persistence. To capture the daily factor we make use of the results of spectral analysis. Peaks in the spectral density graph give a hint on periodic daily component. Over the years we find also some interesting pattern in the average wind speed during single months. These monthly influences are incorporated into our model by a locally constant function which is monthly dependent. The parameters for both daily and monthly factors are estimated using the method of least squares. After subtracting both of them from the wind time series we obtain residuals that are highly correlated. Because of the long-memory feature of the residuals we model them by the means of an ARFIMA (Autoregressive Fractionally Integrated Moving Average) process.

2 Data and Exploratory Analysis

The high-frequency data used in this study were collected at the weather station in Manschnow, an area in Eastern Germany ($52°32'50''N$ $4°32'49''E$). They consist of approximately 157100 10-minute observations from January 2005 to December 2007. Because the data were not entirely complete, 0.36% had to be estimated. This was done by using a linear interpolation. By sorting the data with respect to certain criteria wind features like diurnal or seasonal pattern become obvious.

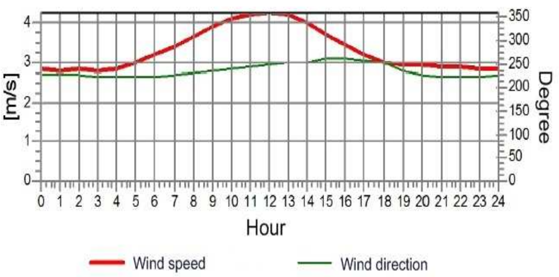

Figure 1: Hourly averages

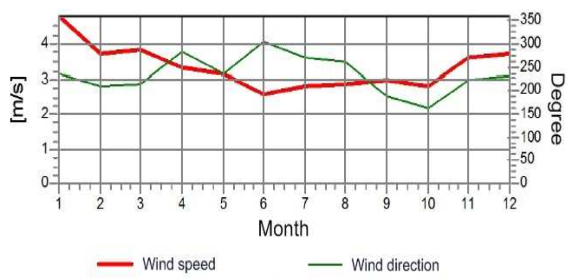

Figure 2: Monthly averages

Figure 1 presents the average hourly wind speed over 2005 to 2007. There is a clear regular pattern within a single day; the wind speed peaks between 11 AM and 12 AM. We investigated this pattern using spectral analysis that confirmed our daily factor assumption. We found out that peaks in the periodogram of our data occurred always at a frequency of 0.043. This value corresponds to the period P of 144 observations which is equal to 24 hours (144×10 minutes $= 1440$ minutes $= 24$ hours). Figure 2 shows the average monthly wind speed and wind direction over the same years. It indicates that the wind blows much stronger during the cool season (October-March) than during the warmer one. But even within the seasons there are visible differences between the single months. Using the SAS program we estimated the values of the monthly parameters and we proved their significance for the model by rejecting the null hypothesis that they are equal to 0 at the significance level of 99.99%. Finally, we decided to include them into the model as indicator functions for each month apart from December in order to make the model uniquely identifiable.

3 Modelling High Frequency Wind Speed Data

The above considerations give rise for introducing the model

$$\begin{aligned} W_t &= \beta_0 + \beta_1 cos(2\pi t/P) + \beta_2 sin(2\pi t/P) \\ &\quad + \beta_3 I_{\{Jan\}}(t) + ... + \beta_{13} I_{\{Nov\}}(t) + \varepsilon_t . \end{aligned} \tag{1}$$

where W_t is the wind speed at time t, P is a period equal to 144, $I_A(t)$ denotes the indicator function of A, and $\{\varepsilon_t\}$ represents an ARFIMA(p,d,q) process. This means that

$$\phi(B)\nabla^d \varepsilon_t = \theta(B) a_t .$$

Here B stands for the backward shift operator defined by $B\varepsilon_t = \varepsilon_{t-1}$. The fractional difference operator ∇^d is given by

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k .$$

$\{a_t\}$ is a white noise process of uncorrelated variables with mean 0. It is assumed that $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ and $\theta(z) = 1 - \sum_{j=1}^{q} \theta_j z^j$ have no zeros on or inside the unit circle and no zeros in common.

We model the residuals using the ARFIMA process because of their strong positive autocorrelation confirmed by the Durbin-Watson statistics. ARFIMA processes have been intensively discussed in, e.g., Beran (1994), Taniguchi and Kakizawa (2000), and Robinson (2003).

The above model is a regression model with 13 regressors and long-memory stationary errors. The parameters β_0 to β_{13} were estimated using the ordinary least squares estimator (LSE) method. The autoregressive and moving-average order of the process (p and q) were set equal to 3 according to the results based on the minimum information criterion. The parameter d as well as the three AR and the three MA coefficients are estimated for each forecast separately using the ARFIMA package in the Ox program of J.A. Doornik. It has to be noted that the estimator of d always takes values between 0 and 0.5. The estimation procedure was performed given the previous month observations which complies with the results of Gneiting et al. (2006). They observed that training periods between 30 and 50 days perform best for AR forecasts. Moreover, all regressors of the model are non-stochastic. Such processes have been treated among others by Yajima [10] and Dahlhaus [3]. Using the matrix notation the process (1) can be written as

$$W_t = \mathbf{z}_t' \beta + \varepsilon_t$$

with $\mathbf{z}_t = (1, cos(2\pi t/P), sin(2\pi t/P), I_{\{Jan\}}(t), ..., I_{\{Nov\}}(t))'$ and $\beta = (\beta_0, ..., \beta_{13})'$. Various estimators of the parameter vector β have been considered in literature. The most obvious approach to use is the LSE $\hat{\beta}_{LSE} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}$ where $\mathbf{Z}' = (\mathbf{z}_1, ..., \mathbf{z}_n)$ and $\mathbf{W} = (W_1, ..., W_n)'$. Yajima [10] analyzed the asymptotic behavior of this estimator for a polynomial regression. Assuming that $0 < d < 1/2$ he showed that under certain conditions the least squares estimator is strongly consistent and consistent but with a lower order than for short-range dependent disturbances. He calculated the asymptotic covariance matrix of the LSE as well. Moreover, he studied the asymptotic behavior of the best linear unbiased estimator (BLUE) of β. It is possible to extend his results to more general regression models if the so-called Grenander's conditions are fulfilled (see, e.g., Taniguchi and Kakizawa [9]).

4 Forecast and Evaluation

To arrive at the final forecast we first estimate the ARFIMA parameters that we use as an input for the residual prediction. This step is performed by using a *forecast* option in the ARFIMA package of the Ox program. Finally, we add the daily and monthly components. Theoretically, it could happen that the forecast result is negative, especially when there is a strong declining trend in the daily component and the ARFIMA residual forecast. In such a case the forecast result should be set to 0. However, this has never happened in our study.

A typical benchmark for evaluating the short-term forecasts of the wind speed is the no-change assumption (persistence). In general, the shorter the forecasting period, the more competitive the persistence forecast turns out to be. It is trivial, but it works well due to the fact that the changes in the atmosphere occur usually within days (at least in Europe). To compare the persistence and our point forecasts we used the root mean squared error (RMSE). Our model beats the persistence forecast relatively best at one hour time horizon $+/-$ 10 minutes. In such a case the RMSE for a persistence model equals 7.24 while for our model it is 6.55. This gives a 9.42% improvement over the persistence assumption. We observed this is mostly due to a very good ARFIMA performance when the wind speed changes rapidly.

References

[1] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall. New York.

[2] Brown, B. G. Katz, R. W. and Murphy, A. H. (1984). Time series Models to Simulate and Forecast Wind Speed and Wind Power. *Journal of Climate and Applied Meteorology* **23**, 1184–1195.

[3] Dahlhaus, R. (1995). Efficient location and regression estimation for long range dependent regression models. *Annals of Statistics* **23**, 1029–1047.

[4] Fuller, W. A. (1996). *Introduction to Statistical Time Series*. Wiley. New York, Chichester, Brisbane, Toronto, Singapure.

[5] Genton, M. G. and Hering A. S. (2009). Powering Up With Space-Time Wind Forecasting. *Journal of the American Statisticial Association*, in press.

[6] Giebel, G., Brownsword, R., Kariniotakis G. (2003). *The State-Of-The-Art in Short-Term Prediction of Wind Power. A Literature Overview*. Risø National Laboratory.

[7] Gneiting, T. Larson, K. Westrick, K. Genton, M. G. and Aldrich, E. (2006). Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: The Regime-Switching Space-Time Method. *Journal of the American Statistical Association*, Vol. 101, **475**.

[8] Robinson, P. M. (2003). *Time Series with Long Memory*. Oxford University Press. Oxford, New York.

[9] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer. New York, Berlin, Heidelberg.

[10] Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Annals of Statistics* **19**, 158–177.