European Association for the
Development of Renewable Energies, Environment
and Power Quality (EA4EPQ)

International Conference on Renewable Energies and Power Quality
(ICREPQ'12)
Santiago de Compostela (Spain), 28th to 30th March, 2012

Evaluation of simple wind power forecasting methods applied to a long-term wind record from Scotland

W.-G. Früh[1]

[1] School of Engineering and Physical Sciences
Heriot-Watt University
Riccarton, Edinburgh EH14 4AS, Scotland (UK)
Phone number:+0044 131 4514374, Fax number:+0044 131 4513129, e-mail: **W.G.Fruh@hw.ac.uk**

Abstract. We present an analysis of the ability to predict the power output from a nominal wind turbine or wind farm a few hours ahead using only locally available data – either the current and recent wind speed or power output. A third method combines the current state with knowledge of the long-term climatology.

The wind speed data were taken from a 46-year long record of hourly readings at a Scottish coastal site and converted to power output and electricity production using a generic wind turbine power curve. The wind speed data or the calculated power output at a given time were used to predict the output a few hours ahead, either using persistence, a linear model, or a model based on the mean daily cycle extracted from the long-term record.

Since many wind farm operators base their forecast on current wind speed or output measures alone, this analysis will provide some quantification of the quality of this approach, either to help them plan their operation or be able to put these simple methods in quantitative context of more complex methods.

Key words

Wind energy resource, wind farm output forecast.

1. Introduction

Wind power generation is one of the fastest growing industries in the developed world, with an installed capacity of 194 GW in 2010 through large wind farms, projected to grow by 15 - 23% per year [1] over the next 5 years. Considering that some of the wind farms now reach an installed capacity in the GW range, even a small relative change in output can amount to a significant variation of the supply to the network. One of the major concerns for wind farm operators in the electricity market is to forecast the expected available output from a wind farm a few hours ahead. In the UK, for example, contracts operate on a market with a day-ahead stage, a main 4-hour contracting, a gate-closure one hour ahead and a post-hoc imbalance settling [2]. As for the long-term prediction of the annual available output, the short-term forecasting should be as reliable as possible whilst being as simple as possible.

Recognising the importance of accurate and reliable forecasts has resulted in a substantial amount of research, which can be roughly classified as based in numerical weather prediction provided by the relevant meteorological office, Artificial Intelligence systems, or stochastic or statistical modelling.

The use of numerical weather prediction provides a prediction based in the actual physical processes determining the wind but the information tends to be at a spatial scale larger than wind farms and will take the local characteristics only at a very coarse scale into account; hence the challenge is to develop a reliable method to downscale the weather prediction to the site-specific wind forecast for the operational wind farm [3].

Artificial Intelligence approaches usually use artificial neural networks (ANN) where the processing of information in the networks is carried out through calculations which have been internally determined from a training period of available past data [4]. While they can be very powerful, it is very difficult to provide a rational explanation for the success or failure of a particular model. An alternative to building an empirical but deterministic model of the wind time series is based in dynamical systems theory where the underlying invariant dynamics are approximated by some basis function, such as radial basis functions [5], empirical orthogonal functions [6], or wavelets [7]. While their predictive response to new data can be better understood in terms of their most dominant basis functions, these methods are limited by assuming a relatively small number of constant basis functions or modes.

Stochastic and statistical modelling differs to ANN and NWP in the fundamental approach that the processes are not described by deterministic dynamics but can better be modelled as a process which depends both on the current and previous stage but also on a process which is apparent as a random fluctuation or force. The key examples for this approach are autoregressive models (AR) including autoregressive moving average (ARMA) [8] or regime switching models [9], where one approach implicitly assumes that the observed variability arises from a continuous random process within a single (but possibly very broad) regime whereas the other assumes that the variability is, at least partly, result of some random process switching the state of the system between a number of distinct regimes. The simplest method within this framework is the assumption that the dynamics are locally stationary, that is that the currently observed value persists until the next prediction point. This is followed by a simple linear prediction based on the current and a few previous observations as explained very clearly by Riahy and Abedi [10]. Riahy and Abedi demonstrate how the use of unfiltered input data can lead to large overshoot or undershoot predictions of the wind a few seconds ahead and suggest that the use of a filtered time series would be far more powerful. It has to be borne in mind, however, that they only demonstrate the improvement for predictions shorter than the filtering time scale and it is hence not surprising that they get near-perfect results.

At present, it is far from clear which of these approaches is optimal but recent work has presented a systematic comparison of a number of techniques against measurements from a wind farm in Southern Italy [11] and has found that different methods have different strength, and the 'optimum model' depends very much on the desired application, such as the prediction horizon. While that research is progressing, many wind farm operators still base their operational forecasts on the assumption that the wind four hours ahead will be '*the same as now*' or '*persistence'*.

In this paper, we evaluate the predictive reliability of this persistence-based approach against two other simple forecasting techniques. One of these is a prediction based on a linear regression of the most recent wind speed data, while the other uses the current wind speed and predicts according to a daily and seasonal cycle obtained from the long-term wind climatology at the site.

These three forecasting methods were applied to a 46-year long wind speed record from a site on the West coast of Scotland, which provided hourly mean wind readings from an anemometer 10 m above ground. The wind speed readings were then used to calculate the corresponding power output using a generic wind turbine power curve, The analysis quantifies the mean error as well as the likelihood of predicting the wind power to within a specified error margin for a given time step ahead. While the first is a standard measure, it is felt that the second may be more useful for operational purposes, where the operator is less interested in the mean error of their operation but more how good their chances are to operate within a given operational limit.

2. Wind data and turbine power output

A. Wind data

The wind data were provided by the UK Meteorological Office through their British Atmospheric Data Centre (BADC) [13] from one of their land surface weather stations which have anemometers at 10 m above ground. The data presented here are from Machrihanish, a coastal site near Campbelltown on the western coast of the Kintyre peninsula. The hourly mean wind speeds are stored to the nearest knot (1 kn = 0.5144 m/s). For the analysis, the hourly wind speeds were converted to m/s, and the uncertainty in each measurement was assumed to be ± 0.257 m/s.

Comparison with other sites across Scotland showed that this site was a good representation for the entire region, and it had a long continuous wind record from 1969 until present. We applied the analysis also to extrapolated heights of typical large turbines. While the power output and capacity factors increased significantly quantitatively, the qualitative result and the relative errors were identical to those from the direct readings at 10 m above ground. For this reason, we present here the results from the direct measurements rather than extrapolations which involve further assumptions on the wind shear profile.

B. Turbine power output

The wind speed is converted to a nominal power output through a generic turbine performance curve representative for many modern large wind turbines, with a cut-in wind speed of 4 m/s, a rated wind speed of 12 m/s, and a cut-out wind speed of 25 m/s, as shown in Figure 1. At the rated wind speed, the power output, P(u), reaches the rated power which is here taken as unity, $P_R = 1$.

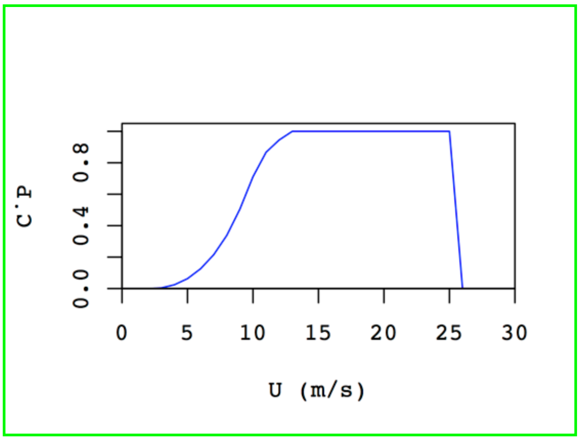

Fig. 1. Power curve of the generic wind turbine.

For a turbine with unit rated power, the capacity factor, C_C, of the turbine in an environment with a normalised wind distribution, Φ_u, is identical to its mean power output and can be calculated by the convolution integral

$$C_C = \int \Phi_u \, P(u)\, du \qquad (1)$$

This approach has a number of implicit assumptions, most importantly that the turbine responds instantly to changes in wind speed or direction.

C. Data analysis

The analysis of the data was carried out using the statistical package R [14]. For the presentation of data distributions, extensive use of box plots was made, in which the distribution of a quantity around its median is shown in terms of a box and whiskers. For data with no clear outliers the box and whiskers show the range of the observations in their quartiles; the first quartile is represented by the lower whisker, the second by the part of the box below the median line, the third quartile by the part of the box above the median line, and the final quartile by the upper whisker. However, if there are outliers, then they are shown separately, and the whiskers only cover the data which are defined as within the expected range of the distribution. Throughout this analysis, the standard setting for box plots was used which defines the maximum range of the first quartile as 1.5 times the range of the second quartile and similarly the maximum range of the 4th quartile as 1.5 times the range of the 3rd quartile. In box plots where a dark circle is shown within the box, this circle represents the arithmetic mean of the data.

D. The predictors

The forecasting of the power output was carried out by using the information available at the time of the forecasting to forecast the power output during an hour a specified number of hours, H_p, ahead.
The three predictors used are referred to as

1. *Persistence*
   This predictor simply assumes that the power output H_p hours ahead would be equal to the current output.

2. *Slope*
   This predictor attempts to predict the power not only from the current measurement but also from the trend over the last few hours, where a weighted average of the slope between the current measurement and that one, two, and three hours ago is averaged with a weighting of 1/2, 1/3, and 1/6, respectively, to smooth out strong random fluctuations.

3. *Daily cycle*
   This predictor uses the current measurement and the mean daily cycle. The power output is calculated as the sum of the current output plus the difference between the mean climatological value at the hour to be predicted and the climatological value at the current hour.

The prediction from these three predictors was then compared to the actual output at the predicted time. From this prediction error, two complementary measures of predictive ability were calculated, the mean prediction error and the likelihood of predicting within a given error margin, where the error margin is specified in units of the capacity factor.

3. Results

A. Daily and seasonal coherence

As a precursor to the forecasting, the autocorrelation of the wind speed or power output time series was computed as this will give an indication if there is any correlation between a current wind or power measurement and some measurement in the past or future. The autocorrelation function for the wind speed against the time lag, shown in Fig. 2, highlights a rapid decrease in correlation as the time difference between the measurements is increased but there is evidence of a weak daily cycle and a slightly stronger seasonal cycle, as well as a feature at a time scale of 22 days, possibly indicative of the typical time scale of mid-latitude synoptic weather systems. The presence of daily and annual variability suggests that an analysis of the resource at an annual and daily cycle could provide useful information for seasonal forecasting of the electricity production from a wind farm or for operational forecasting of the output.

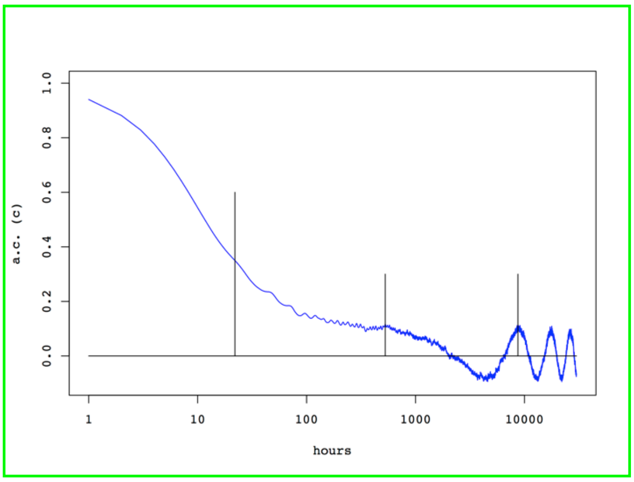

Fig. 2. Autocorrelation curve for the hourly wind speed data.

An analysis of the wind speeds by month across all years, shown in Figure 3 shows that there is a moderate but clear seasonal cycle with the mean winds larger during the winter months (November to February) and lower in the summer months (June to August). This cycle is much more pronounced in the occurrence of strong winds, where the wind only goes rarely above 10 m/s in the summer (at 10 m above ground) but that the wind was above 10 m/s for 25% of the time in winter. In contrast to this, the number or extent of calmer periods with wind speeds less than 4 m/s, varies only slightly across the seasons.

Considering that the strongest variability is where the performance curve of turbines is very sensitive to changes, or where they approach their rated power, it is instructive to compare the wind speed records with their equivalent capacity factors. Figure 4, which shows the statistics of the hourly capacity factor, demonstrates this sensitivity of the power output against wind speed fluctuations as the seasonal cycle is amplified compared to the wind speed cycle.

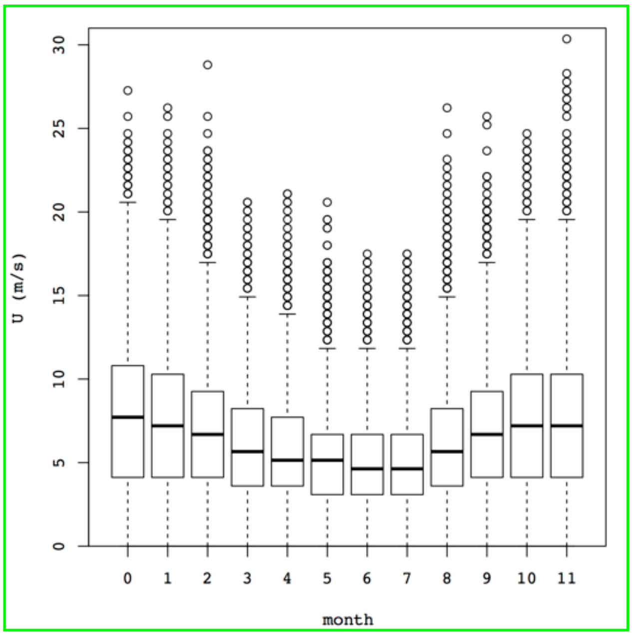

Fig. 3. Distribution of hourly wind speeds for different months in the year.

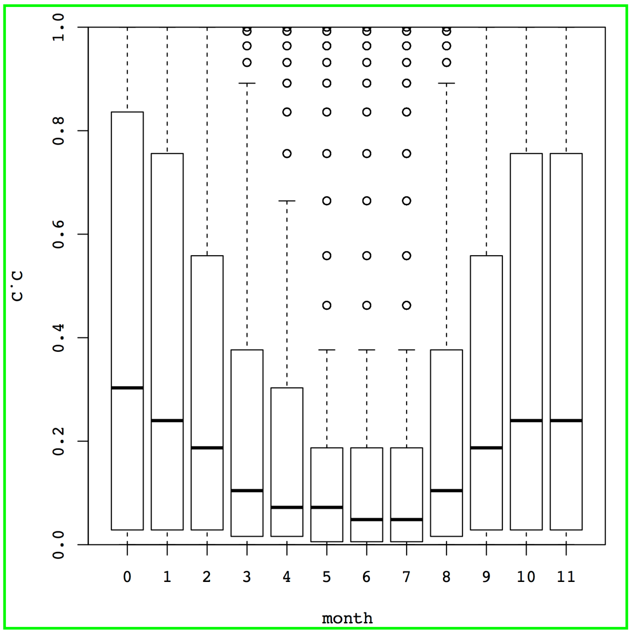

Fig. 4. Distribution of the hourly capacity factor.

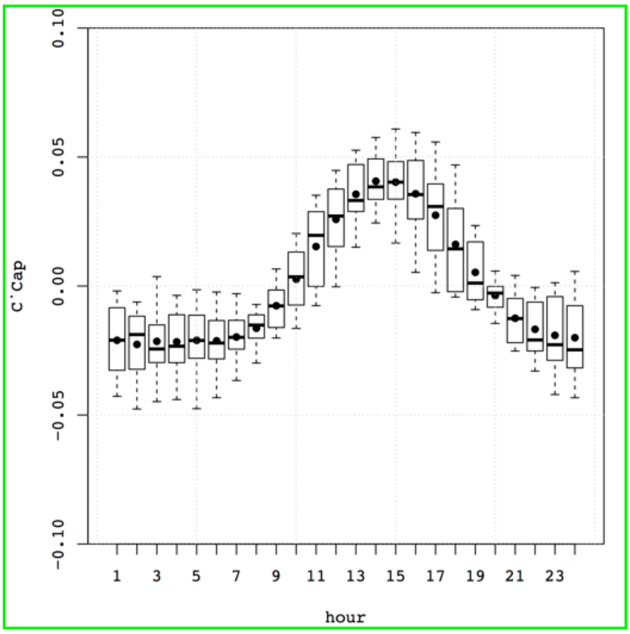

Fig. 5. Deviation of the hourly capacity factor against the monthly mean.

Since the autocorrelation function suggested the presence of a weak daily cycle as well as the seasonal cycle, Figure 5 presents by how much the monthly mean of the difference between the hourly output and its overall mean deviates from that mean. This shows that the available capacity factor is below the monthly mean by about 0.2 at night and in the morning but above the mean in the afternoon, reaching a peak of, on average, 0.4 above the monthly mean.

B. Mean prediction error

Figure 6 shows the mean prediction error for the power output, separated into over-prediction (positive) and under-prediction (negative). While all predictors show a similar mean error for predicting a single hour ahead, it is immediately obvious that a linear prediction based on the observed slope leads to a rapidly increasing mean error. The other two methods follow qualitatively and quantitatively similar patterns. It appears that the mean error is lower for using persistence when one has over-predicted the output but that the daily cycle reduces the mean error when one has under-predicted the output.

C. Likelihood of good prediction

Figures 7 and 8 show that the ability of all predictors decreases as the prediction step increases. The difference between the two figures is that for Figure 7, the prediction is based on the power output data and the daily cycle of the power output as shown in Figure 5, whereas the predictive quality in Figure 8 refers to prediction of the wind speed which is converted to power output after the prediction step. In both cases the slope predictor is always the worst, though much more so if the predictions are based on the power output rather than the wind.

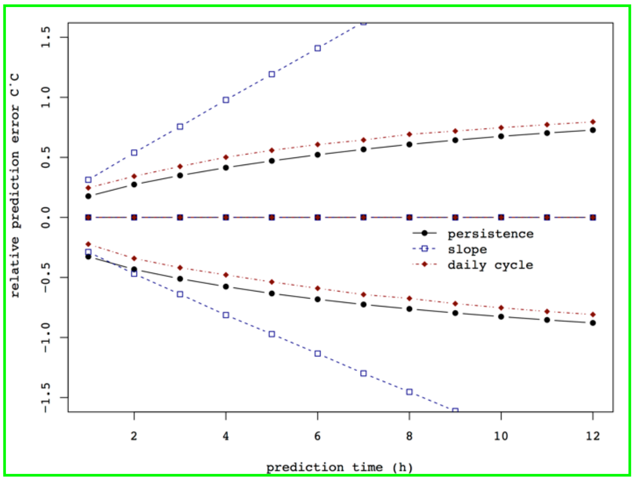

Fig. 6. Mean prediction error for the three predictors.

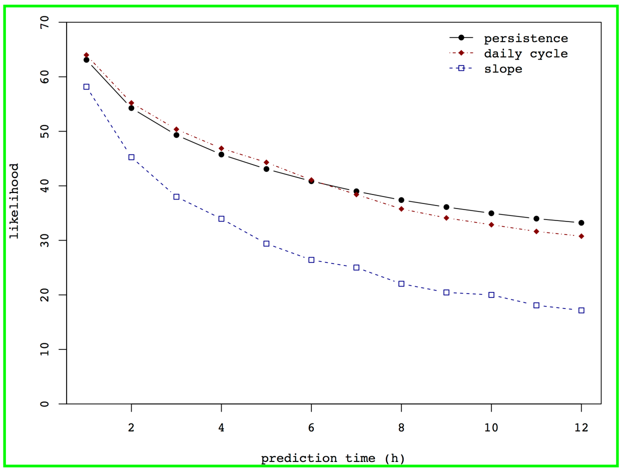

Fig. 7. Likelihood of predicting the power output to within ±0.1 for the three predictors from the recent power output information.

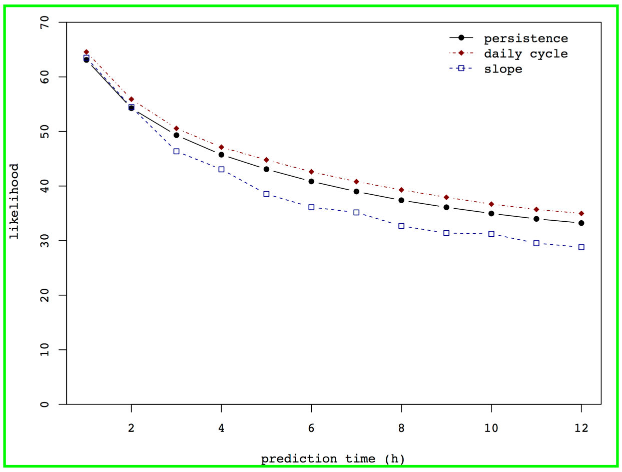

Fig. 8. Likelihood of predicting the power output to within ±0.1 for the three predictors from the recent wind speed information.

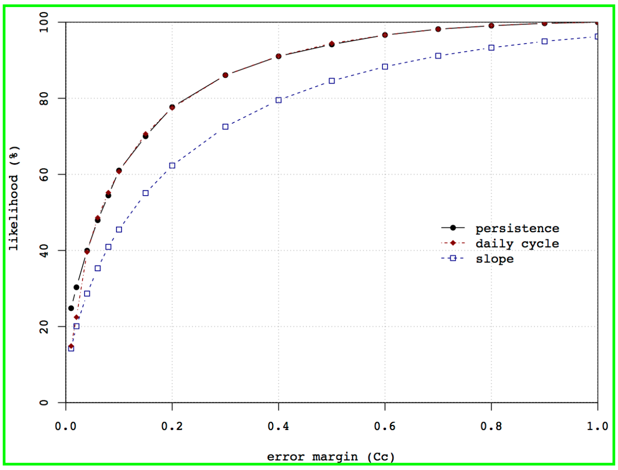

Fig. 9. Likelihood of predicting the power output 4 hours ahead against the error for the three predictors.

Even though the daily cycle for the power output or capacity factor appears more clearly than the corresponding wind speed daily cycle, the predictor based on the wind speed daily cycle outperforms the persistence predictor for all prediction time steps, whereas the output-based prediction shows a cross-over between the daily cycle and persistence at a prediction horizon of 6 hours. For both methods, the predictive power to predict the power output to within ±0.1 drops from about 65% for one hour ahead to around 30% for 12 hours ahead.

Figure 9 shows the complementary picture to Figure 8 in that the prediction horizon and using the wind speed is now fixed at four hours and the specified error margin is varied from δC_C = ±0.01 to ±0.1. As expected, all predictors get more predictions right within the given error margin as that margin is relaxed. While the persistence and slope predictors follow similar curves, with the persistence substantially outperforming the slope predictor, the daily cycle predictor shows an unexpected behaviour in that it shadows the slope predictor for tight error margins of ±0.05 or less but approaches the persistence predictor for higher error margins.

4. Conclusion

The analysis presented in section 3 has demonstrated that of three simple predictors, the most basic of all, namely persistence, and one based on the daily cycle lead to the smallest prediction error and the highest likelihood of predicting the power output to within a given error margin. Even though the daily cycle itself appeared to be more pronounced in the power output than in the wind speed directly, the prediction of the power output achieved better results if the wind speed was predicted, and the predicted output calculated from this predicted wind, rather than using the power output directly for prediction.

This analysis was applied to six other sites within Scotland with qualitatively identical findings even though some sites were in sheltered inland locations with different mean wind characteristics and different strengths of their daily and seasonal cycles.

A clear recommendation to operators, at least those operating within similar climate conditions as Scotland, is to either continue with their current basic approach or to engage with research in much more powerful techniques. Incremental sophistication of the forecasting methods appears to lead to a deterioration rather than improvement of the predictive ability. However, the ability to predict the correct output within a 10% error only a third of the time suggests that there is a clear need to develop more reliable forecasting methods as the number and size of wind farms becomes larger.

Acknowledgement

The author wishes to thank the UK Meteorological Office for supplying the wind speed data.

References

[1] GWEC, Global wind statistics 2010, Global Wind Energy Council (2011), pp.
[2] National Grid. Chapter 10 Market Overview in NETS Seven Year Statement, 2010 http://www.nationalgrid.com/uk/Electricity/SYS/
[3] R.J. Davy, M.J. Woods, C.J. Russell and P.A. Coppin, “Statistical downscaling of wind variability from meteorological fields” in Boundary Layer Meteorology 2010, Vol. 135, 161 – 175.
[4] G. Li and J. Shi, “On comparing three artificial neural networks for wind speed forecasting” in Applied Energy 2010, Vol. 87, 2313 – 2320.
[5] L.A. Smith, “Local optimal prediction: exploiting strangeness and the variation of sensitivity to initial conditions” in Philosophical Transactions of the Royal Society of London (A) 1994, Vol. 348, 371 – 381.
[6] M.R. Allen and L.A. Smith, “Optimal filtering in Singular Systems Analysis”, Physics Letters A 2997, Vol. 234, 419 – 428.
[7] X. An, D. Jian, C. Liu and M Zhao, “Wind farm power prediction based on wavelet decomposition and chaotic time series”, in Expert Systems with Applications 2011, Vol. 38, pp. 11280 – 11285.
[8] E. Erdem and J. Shi, “ARMA based approaches for forecasting the tuple of wind speed and direction” in Applied Energy 2011, Vol. 88, 1405 – 1414.
[9] C. Gallege, P. Pinson, H. Madsen, A. Costa and A Cuerva, “Influence of local wind speed and direction on wind power dynamics – Applications to offshore very short-term forecasting”, in Applied Energy 2011, Vol. 88, pp. 4087 – 4096.
[10] G.H. Riahy and M. Abedi, “Short term wind speed forecasting for wind turbine applications using linear prediction method” in Renewable Energy 2008, Vol. 33, 35 – 41.
[11] M.G. de Giorgi, A. Ficarella and M. Tarantino, “Error analysis of short term wind power prediction models” in Applied Energy 2011, Vol. 88, 1298 – 1311.
[12] UK Meteorological Office. MIDAS Land Surface Stations data (1853-current) through the British Atmospheric Data Centre: http://badc.nerc.ac.uk
[13] K. Hornik, The R FAQ, ISBN 3-900051-08-9, 2011, http://CRAN.R-project.org and J. Adler, R in a Nutshell. O’Reilly, 2010.