



A vector auto-regressive model for onshore and offshore wind synthesis incorporating meteorological model information

D. Hill¹, K. R. W. Bell¹, D. McMillan¹, and D. Infield²

¹Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK

²Centre for Doctoral Training in Wind Energy Systems, University of Strathclyde, Glasgow, UK

Correspondence to: D. Hill (david.c.hill@strath.ac.uk)

Received: 15 January 2014 – Revised: 27 March 2014 – Accepted: 11 April 2014 – Published: 14 May 2014

Abstract. The growth of wind power production in the electricity portfolio is striving to meet ambitious targets set, for example by the EU, to reduce greenhouse gas emissions by 20 % by 2020. Huge investments are now being made in new offshore wind farms around UK coastal waters that will have a major impact on the GB electrical supply. Representations of the UK wind field in syntheses which capture the inherent structure and correlations between different locations including offshore sites are required. Here, Vector Auto-Regressive (VAR) models are presented and extended in a novel way to incorporate offshore time series from a pan-European meteorological model called COSMO, with onshore wind speeds from the MIDAS dataset provided by the British Atmospheric Data Centre. Forecasting ability onshore is shown to be improved with the inclusion of the offshore sites with improvements of up to 25 % in RMS error at 6 h ahead. In addition, the VAR model is used to synthesise time series of wind at each offshore site, which are then used to estimate wind farm capacity factors at the sites in question. These are then compared with estimates of capacity factors derived from the work of Hawkins et al. (2011). A good degree of agreement is established indicating that this synthesis tool should be useful in power system impact studies.

1 Introduction

The potential of wind power to contribute to demand for electric power needs to be fully assessed if effects of climate change on food production, water resources, insurance costs and so on are to be avoided. In order to facilitate such assessments, a Vector Auto-Regressive (VAR) model-based synthesis has been developed previously by the authors using just onshore data (Hill et al., 2012). These syntheses contain information about the annual and diurnal trends as well as the stochastic component. Many studies have addressed the potential impacts of wind power on a power system (e.g. Strbac, 2002; Gross et al., 2006; Cox, 2009). Different approaches have been taken to the representation of wind speed structure – numerical weather prediction models (e.g. Aigner and Gjengedal, 2010), an approach incorporating reanalysis and mast datasets (Kubik et al., 2013) or purely statis-

tical approaches. A literature review of the latter appears in Hill et al. (2012).

The spatial correlation of wind speeds (and hence wind power) over scales over several hundred km is well known (Hill et al., 2012). This was discussed by Oswald et al. (2008) who looked at an extreme event and evaluated quantitatively by Sinden (2007) and Gibescu et al. (2006). Failure to model such correlations will result in excessive aggregate smoothing of wind power output, and an underestimate of the need for reserve and unrealistic inter-area power flows important in network constraint cost analysis.

A de-trending process is briefly described in the Methodology. The Results section depicts forecasting skill and also shows model to model comparisons of capacity factors with work carried out by Hawkins et al. (2011), which describes the meteorological model used in their work and incorporating a calibration process.

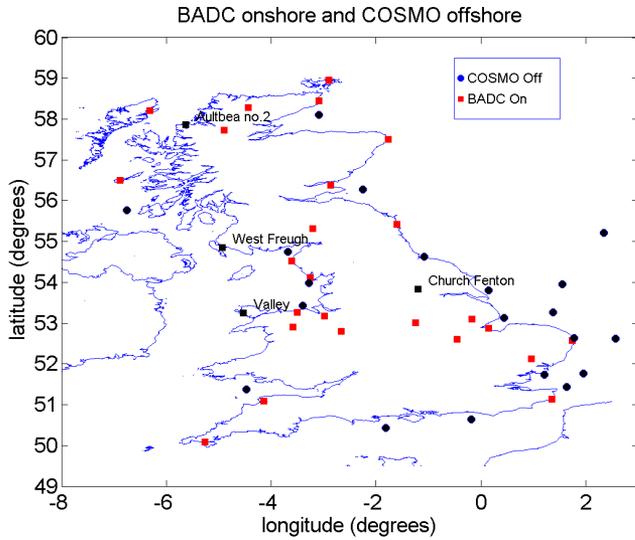


Figure 1. Map to show locations of sites onshore and offshore where time series input to VAR model, results of labelled sites appear in Fig. 3.

2 Methodology

A key innovation in the approach taken here is a de-trending process to represent the cyclical varying components of wind speed and then a stochastic VAR model is applied to the de-trended component. The MIDAS set of onshore weather data produced by the British Atmospheric Data Centre (UK Meteorological Office, 2012) and representing observations at 10 m above ground level was used as the initial source of onshore wind speeds. Modelled sites were chosen onshore that have > 90 % coverage for the period January 1996 to December 2005, taken as the trend formation period onshore. These sites also had > 90 % coverage from January 2006 to December 2007, for use in the VAR modelling process. Observations from multiple offshore locations could not be obtained. Instead, offshore data for the latter period was obtained from the Consortium for Small-scale Modelling (COSMO, 2010), meteorological model (Doms and Schättler, 2002), which does not suffer from data gaps and represent wind speeds at 10 m above sea level. See Fig. 1 for site locations, whose offshore locations are the same as those used in the Scottish Electricity Dispatch Model (SEDM) which made use of the weather model developed by Hawkins et al. (2011).

Some form of de-trending of the data is necessary to ensure a reasonable degree of stationarity in the de-trended series, an assumption of the VAR modelling. In Hill et al. (2012) the wind speed data were first de-trended site by site by the suitable application of harmonic analysis to whatever period of data existed for those sites. A concurrent period common to all sites was then taken for which to calculate the de-trended series. However, in previous work, for some sites, a longer period than that was used to calculate the

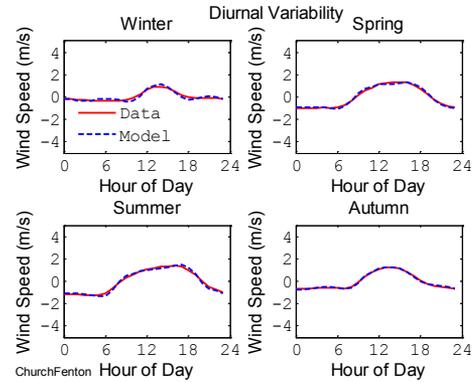


Figure 2. Diurnal trends at Church Fenton, particularly pronounced in Spring and Summer.

trends. In the current work, the trends for all the onshore sites are calculated based on data from the same period, i.e. January 1996 to December 2005, while the trends for all the offshore sites are based on data from January 2006 to December 2006, referred to henceforth simply as the year 2006. These sets of site-by-site trends are subtracted from observations at each site in 2006 to form the training set for the VAR model.

The harmonic analysis described in Hill et al. (2012) remains the same in this work, with an annual component modelled with Fourier terms Ω , 2Ω and 3Ω (where Ω is the annual angular frequency) and a seasonally varying diurnal component with ω and 2ω terms (where ω is the 24 h angular frequency) – see Fig. 2 for an example of the diurnal trends.

These deterministic trends, represented by y_a (annual) and y_{ds} (seasonally varying diurnal), are then subtracted from the data, y , as in Eq. (1), to obtain the de-trended series y_{dt} , before commencing the process of Vector Auto-Regression.

$$y_{dt}(t) = y(t) - y_a(t) - y_{ds}(t) \quad (1)$$

The VAR model is represented by Eq. (2) where Φ_1 and Φ_2 are 50×50 matrices (known simply as the VAR coefficients) reflecting the influences between the 50 sites under consideration at lags $t-1$ and $t-2$ h respectively. There is no simple physical interpretation of Φ_1 and Φ_2 but an equation in the multivariate case can be derived which is similar, but not identical, to the Yule-Walker equations of the univariate case, which in turn relates the covariances of the observations to the auto-regressive coefficients (see for example Sect. 3 in Lütkepohl, 2005).

$$y_{dt,t} = \Phi_1 y_{dt,t-1} + \Phi_2 y_{dt,t-2} + e_t \quad (2)$$

The vector $y_{dt,t}$ is a column vector containing the de-trended series, one site per row and e_t is a Gaussian (normally distributed) noise term. Equation (2) is first applied to the training period of year 2006 to define the VAR model by finding the VAR coefficients Φ_1 and Φ_2 which minimise the sum of

squares of the residuals, e_t , e_t^T in a least squares approach. The year 2007 is reserved for assessing the forecasting accuracy and the determination of the standard deviation of the noise term. The VAR coefficient matrices are used to drive a synthesis of wind speeds by utilising Eq. (2) as a recursive equation using observations at each hour $t-1$ and $t-2$ with the output at time t given simply as Eq. (2) with no noise term. This output is compared with observed value at time t to determine the one step ahead forward prediction error. The process is stepped forward in time and repeated so that a succession of error values is found enabling its standard deviation σ to be calculated. Along with a mean of 0, this is used in a Normally distributed ($N(0, \sigma)$) noise term e_t , thus completing the VAR model.

The number of VAR matrix terms in the model was chosen so that this one-step ahead error was minimised. The addition of the 2nd term provides a small improvement of 2.4%, whilst a possible $t-3$ term did not provide further improvements.

Application of the model to synthesise a wind speed time series starts from some initial condition at times $y_{dt,0}$ and $y_{dt,1}$ together with the one-step ahead forward prediction error. Finally, the value of the trend for the synthesised hour is added to the output of the VAR model.

The aim of this modelling process is to allow the underlying local meteorological characteristics to be captured so that the developed model is able to describe all sites and their interactions, in particular the differences between actual conditions, hour by hour, and the underlying trends. The model also provides a prediction (forecast) of the next hour, given observations in the current and previous hours. These observations depend on the time of day and day of the year and are sensitive to year to year variations. That is, a particular year might be “windier” or “less windy” than the year or years on which the trends were calculated; the de-trended data which is then used to determine the parameters of the auto-regressive model would, as a consequence, not necessarily reflect the current year and the differences would be handled as best it can by the VAR modelling. This highlights the particular challenge of accurately modelling inter-year variability. Particular sites may have observations in some years that differ markedly from the “current” year to which the model is applied for modelling or forecasting purposes.

It should also be noted that diurnal trends are not significant far offshore due to the absence of thermal effects, and that coastal and inland sites may also have distinct patterns associated with them.

The synthesized wind speed series can then be converted to hub height (of 80 m) using a power law (Wieringa, 1993) and then passed through a wind power curve such as that used in the TradeWind report (2009), to produce a wind power series. These power series are obtained at all the modelled offshore sites (chosen to represent the main existing or proposed wind farm locations) and then calculations of capacity factors are made (The capacity factor is the ratio of energy produced

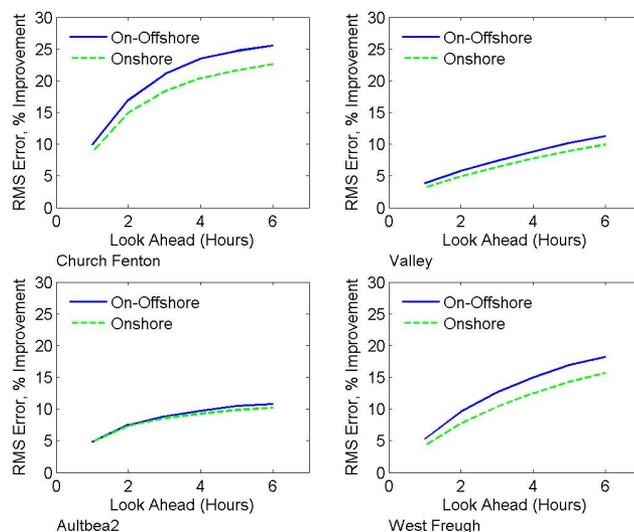


Figure 3. Forecasting skill of VAR model with improvements compared with persistence forecasting, at 4 sample sites for purely onshore model and mixed onshore/offshore model.

by a wind farm in a given period to the total amount of energy it would have generated had it been generating at full power). Comparisons are then made with the results achieved in the development of a prototype Scottish Electricity Dispatch Model (SEDM) for analysis of the GB power market and which used a large set of data derived from a meteorological model without any statistical model, based on (Hawkins et al., 2011), referred to below as “the Edinburgh model”.

3 Results

The work here utilises a training window for model estimation and a forecasting window for its subsequent validation (it does not use “in-sample” data). As stated above, the model was trained on January 2006 to December 2006 and then the year 2007 was used to evaluate the model’s predictive ability. Figure 3 shows the improvements gained over persistence forecasting (where the wind speed at time “now” is held fixed for all look aheads) in the RMS errors at 4 candidate sites. The results from the purely onshore model are seen to be inferior to the mixed onshore/offshore model at nearly all sites by a few per cent, except around the boundaries of the model, such as those at station Aultbea2 and Valley. For sites like these it is hard to see the level of improvement as the lines in Fig. 3 are so close, so further calculations have been made as described below. Improvements over persistence are easier to identify in Fig. 3 and can be seen to be up to 25% and more at 6 h ahead, such as those at Church Fenton, although at sites such as Valley and Aultbea2, located around the edges of the model’s geographical limits, percentages are down to 10% at 6 h ahead.

Further calculations to explore the benefits of extending the data used to include offshore sites (albeit in the form of meteorological model output) have been undertaken for all sites for look ahead periods from 1 to 6 h. We concentrate here on the results for just 3 and 6 h ahead forecasts. The improvement (in percentage terms) achieved by the addition of offshore met. model data to the onshore data for the 6 h ahead forecasts, compared with the purely onshore analysis, varied across the sites studied, ranging from -1.8% (i.e. a slight degradation) for the Western Isles, up to a 22.1% improvement for Walney Island, with an overall mean forecasting improvement across all sites of 10.4% . Interestingly, at 3 h ahead, improvements rose by up to 59.6% (for Tiree) with an overall mean improvement of 16.3% . The main reason for the better performance of the onshore/offshore model would appear to be simply due to the presence of additional relevant (i.e. correlated) data, that happens in this case to cover a wider region. Additional onshore data of good quality would be expected to also improve the modelling in a similar manner. However if the offshore data was not of good quality, or there was no significant relationship between the offshore and onshore data, then no improvement would be anticipated. Further research is required to explore the strength of the relationship of any new data that might be added to an existing model, and the improvements (or otherwise) in model performance. It would be attractive then to derive mathematical constraints based on correlation that would determine when an additional measurement site would improve the performance of an existing model.

How well the VAR model presented here compares with other “state-of-the-art” techniques is difficult to say as model results are highly data dependent. A proper comparison would require an exchange of data sets, or the implementation in full of the methods used by others. This is beyond the scope of the present research. See the reviews by Giebel et al. (2011) and Ma et al. (2009) for more information on other modelling approaches.

Figure 4 shows the two sets of wind turbine capacity factors calculated for the Edinburgh model and the VAR model at all the offshore sites used in the VAR model and, whilst a few significant discrepancies do exist, the overall pattern of results is similar, even though the two models were based on completely independent data. It should be noted that the Edinburgh model was based on longer term information covering nearly 10 years (from April 2001 to December 2010). The capacity factors for the VAR model synthesis used here are based on just two years, 2006 and 2007; this period was seen from a GL Garrad Hassan paper (Hodgetts, 2011) to be fairly “typical” in that it was close to the long term mean.

A brief investigation of possible regional variations in the agreement in capacity factor estimates from the two methods was made. It was tentatively concluded that the South Eastern offshore sites on the whole showed better agreement than those on the West of the country, although the sites of Lynn and Inner Dowsing, and Dudgeon showed poor agreement

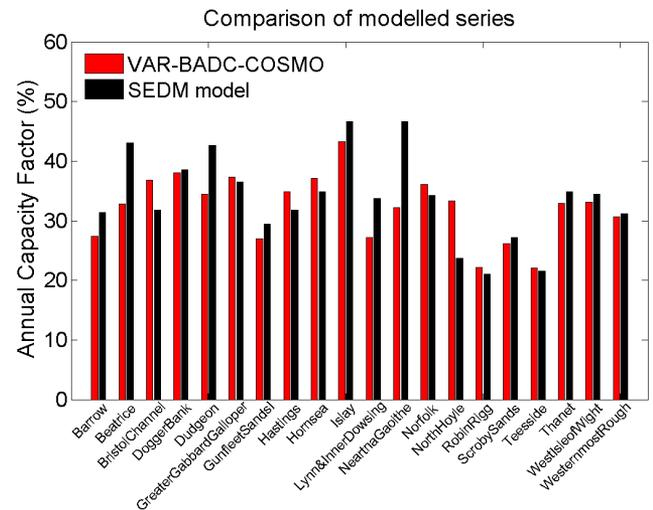


Figure 4. Comparison of capacity factors derived from VAR model and Edinburgh model.

with differences up around 24% . It is thought that these results may be due to the greater concentration of input data to the VAR model in the regions showing better agreement.

A judgment of which of a number of candidate models is most “accurate” depends on the purpose to which the models are being put. For example, the priority might be to adequately reproduce monthly average wind speeds, or alternatively the hour-by-hour variability, and a given model might not achieve both of these equally well in comparison with other models. Moreover, such an evaluation depends on the availability of actual observations to which the models can be compared. Although many models are based on wind speed, to be of value in the energy sector, they must in the end quantify wind power, and this depends on an accurate wind speed to power conversion which, in itself, is far from trivial. For example, in practice it depends on wind direction, the size of a wind farm, the terrain in which it is located and the design of wind turbines at the farm and their spacing (mainly due to wake effects). Furthermore, the validation of a wind power time series depends on hour-by-hour observations of wind power. In Britain, only monthly total energy outputs are publicly available which limits the extent of the validation. The Edinburgh model was compared with these monthly outputs and here, a comparison is made with the results from that model.

4 Conclusions

A method capable of providing both wind forecasts and a synthesis of wind speeds (and wind power) across the GB power network system, useful for planning and operation purposes, has been presented. Further improvements to published work have been made, including a more rigorous approach to separate training from application. In addition, the

VAR model has been extended to include offshore meteorological model data. The performance of this extended VAR model forecasting has been assessed, and improvements over persistence by as much as 25 % have been demonstrated. Up to 3.7 % of this improvement can be attributed to the addition of the offshore model data at 6 h ahead.

Capacity factors at offshore wind farm sites (both existing and planned) have been compared with an approach based purely on a meteorological model and, with a few exceptions, it has been shown that the two approaches are in good agreement at most sites. Future work will involve the development of interpolation techniques (e.g. Kriging) to facilitate the estimation of capacity factors at onshore wind farm sites away from meteorological measurement sites, with which it will be possible to validate against measured wind power data from the Renewables Obligation Certificate register.

Acknowledgements. The authors kindly thank the EPSRC for providing funding under the “Top and Tail” grant (Project number EP/I031707/1, <http://www.topandtail.org.uk/>). The provision of data from the BADC and COSMO is also gratefully acknowledged and a special note of thanks goes to Tobias Aigner for his help in unpacking the COSMO model data.

Edited by: D. Brayshaw

Reviewed by: two anonymous referees

References

- Aigner, T. and Gjengedal, T.: Detailed wind power production in northern Europe, Renewable Energy Conference, 27 June–2 July, Yokohama, Japan, 2010.
- COSMO (Consortium for Small-scale Modelling): available at: www.cosmo-model.org (last access: January 2013), 2010.
- Cox, J.: Impact of intermittency: how wind variability could change the shape of the British and Irish electricity markets, Pöyry Energy (Oxford) Ltd, Oxford, 2009.
- Doms, G. and Schättler, U.: A description of the nonhydrostatic regional model, LM, Part 1: Dynamics and Numerics, Deutscher Wetterdienst, 2002.
- Gibescu, M., Ummels, B. C., and Kling, W. L.: Statistical wind speed interpolation for simulating aggregated wind energy production under system studies, in: Proc. 9th Int. Conf. Probabilistic Methods Applied to Power Systems KTH, Stockholm, Sweden, June, 1–7, 2006.
- Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., and Draxl, C.: The state-of-the-art in short-term prediction of wind power – A literature review, 2nd Edn., Technical Report, Technical University of Denmark, 2011.
- Gross, R., Heptonstall, P., Anderson, D., Green, T., Leach, M., and Skea, J.: The costs and impacts of intermittency: an assessment of the evidence on the costs and impacts of intermittent generation on the British electricity network, UK Energy Research Centre, available at: <http://www.ukerc.ac.uk/support/tiki-index.php?page=Intermittency>, March, 2006.
- Hawkins, S., Eager, D., and Harrison, G. P.: Characterising the reliability of production from future British offshore wind fleets, IET Renewable Power Generation Conference, September, 2011.
- Hill, D., McMillan, D., Bell, K. R. W., and Infield, D.: Application of auto-regressive models to UK wind speed data for power system impact studies, IEEE Trans. Sustain. Energ., 3, 134–141, January, 2012.
- Hodgetts, B.: A statistical review of recent wind speed trends in the UK, RenewableUK conference, Manchester, available at: http://www.gl-garradhassan.com/assets/downloads/A_Statistical_Review_of_Recent_Wind_Speed_Trends_in_the_UK.pdf, October, 2011.
- Kubik, M. L., Brayshaw, D. J., Coker, P. J., and Barlow, J. F.: Exploring the role of reanalysis data in simulating regional wind generation variability over Northern Ireland”, Renew. Energ., 57, 558–561, 2013.
- Lütkepohl, H.: New introduction to multiple time series analysis, Springer-Verlag, New York, 2005.
- Ma, L., Luan, S., Jiang, C., Liu, H., and Zhang, Y.: A review on the forecasting of wind speed and generated power”, Renew. Sustain. Energ. Rev., 13, 915–920, 2009.
- Oswald, J. O., Raine, M., and Ashraf-Ball, H.: Will British weather provide reliable electricity?,” Energ. Pol., 36, 3212–3225, June, 2008.
- Sinden, G.: Characteristics of the uk wind resource: Long term patterns and relationship to electricity demand, Energ. Pol., 35, 112–127, January, 2007.
- Strbac, G.: Quantifying the system costs of additional renewables in 2020, report to UK DTI, available at: <http://www.dti.gov.uk/files/file21352.pdf>, 2002.
- TradeWind Project Final Report: available at: http://www.trade-wind.eu/fileadmin/documents/publications/Final_Report.pdf, May 2009.
- UK Meteorological Office: Met. Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853–current), NCAS British Atmospheric Data Centre, available at: http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent__ukmo-midas, 2012.
- Wieringa, J.: Representative roughness parameters for homogeneous terrain, Bound. Layer Meteorol., 63, 323–363, 1993.