Synthesis of wind time series for network adequacy assessment

G. S. Hawker, W. A. Bukhsh, S. Gill, K. R. W. Bell
Institute for Energy and Environment,
Department of Electronic and Electrical Engineering,
University of Strathclyde, Glasgow, United Kingdom.
{graeme.hawker, waqquas.bukhsh, simon.gill, keith.bell}@strath.ac.uk

***Abstract*—When representing the stochastic characteristics of wind generators within power system simulations, the spatial and temporal correlations of the wind resource must be correctly modelled to ensure that reserve and network capacity requirements are not underestimated. A methodology for capturing these correlations within a vector auto-regressive (VAR) model is presented, and applied to a large-scale reanalysis dataset of historical wind speed data for the British Isles. This is combined with a wind speed-to-power conversion model trained against historically metered data from wind farms on the Great Britain (GB) electricity system in order to derive a lightweight model for simulating injections of wind power across a transmission network. The model is demonstrated to adequately represent ramp rates, both at a site and network level, as well as the individual correlations between sites, while being suitable for network adequacy studies which may require the simulation of many years of operation.**

***Index Terms*—wind energy, wind power modeling, meteorology, autoregression.**

I. Introduction

The growth in wind generation in electricity markets has added additional complexity to power system modeling, with an increased background of non-synchronous generation and the requirement for adequate representation of the stochastic characteristics of wind generators. It is well-understood that wind is spatially correlated over large scales [1], and a failure to correctly model such correlations in power system simulations will lead to an over-smoothing of aggregate wind power and an underestimation of the required reserve capacity and inter-area flows on an electricity network [2].

Previous approaches to this problem fall into three broad categories: the direct use of historic wind farm generation data, e.g. [3]; direct use of wind data (either from historical or reanalysis data) through a deterministic wind speed-to-power conversion process, e.g. [4],[5],[6]; or the creation of a statistical model which seeks to characterise and reproduce the stochastic properties of the wind resource, e.g. [7], which analyses covariance in power output between sites; [8], which uses a Markov chain Monte Carlo (MCMC) approach; or [9], which uses copulas for their ability to model non-linear dependence structures as well as signals with non-Gaussian marginal distributions. [10] and [11] illustrate the smoothing effect of interconnection of multiple sites, and the demonstration of correlated output between locations in this manner points towards the need for models which can represent spatial correlations.

The use of historical or reanalysis meteorological data makes the assumption that historical conditions adequately represent all future operational conditions, which is dependent on the length of the dataset in use. In power system analyses which aim to identify edge cases, such as contingency scenarios for network adequacy studies, the tails of the distribution of possible operating conditions are of particular interest, and these may not be captured within existing data [12].

A statistical model, alternatively, aims to synthesise wind time series from a set of equations which correctly characterise the auto-correlative nature of the wind resource, and potentially allowing the generation of much larger amounts of time series data than may exist in historical datasets. This can also greatly reduce the computational burden, by allowing wind power data to be created as it is required by a wider simulation, rather than having to handle and store data in bulk. However, in such approaches it is key that the temporal and spatial correlations of the natural resource are correctly characterised, in order that the variability of wind injections across an electrical network adequately captures the variety of likely operating states.

The suitability of auto-regressive processes to modeling wind speeds, in order to preserve such correlations, has been previously demonstrated in e.g. [13] and extended to modeling of wind power in [14], [15], [16]. This paper aims to extend that work by selecting and parameterising a wind model suitable to a particular electricity system, and improving the modeling of actual power injections through the use of detailed metered wind generation data.

In this paper, a statistical wind model based on a Vector Auto-Regressive (VAR) process is presented and demonstrated for the electricity network of Great Britain (GB). In Section II, the statistical formulation is presented, followed by an application of the model to wide area historical reanalysis data for the British Isles (covering GB and relevant offshore

The authors kindly thank the EPSRC for providing funding under the SUPERGEN Energy Networks Grand Challenge Fund: 'Transformation of the Top and Tail of Energy Networks' (Project number EP/I031707/1, www.topandtail.org.uk).

development zones) in Section III. The conversion of wind speed to actual power injections into the transmission network, using historical recorded wind generation data, is covered in Section IV, and the results of synthesising power time series are compared to historical data in Section V. The paper ends with a discussion of the applicability of the presented model to network adequacy and similar power system analyses in Section VI.

The contribution of this paper is in demonstrating the application of statistical wind and power models to large volumes of wind and power data covering a national electricity system, to both demonstrate the validity of reducing that data to a more concise form, as well as demonstrating the applicability of the approach to studies requiring the generation of a large number of steady-state or dynamic scenarios across a network with a high penetration of wind power. The use of a statistical model also allows the extrapolation of the methodology to project output of future sites, such as to model future scenarios for a network following increased growth in wind power.

II. A Vector Auto-Regressive (VAR) Wind Model

An n -variable vector autoregression of order p, VAR(p), is a system of n linear equations, with each equation describing the dynamics of one variable as a linear function of the previous p lags of every variable in the system, including its own p lags [17]. As a VAR model assumes stationarity of the underlying process, it is necessary to remove cyclically varying components of the wind speed. In [16], an inspection of seasonal and diurnal components is undertaken in the context of auto-regressive models, and that de-trending process is followed here.

The wind time series is modelled as including two deterministic components (Eq.1), representing the annual and diurnal trends respectively. For each measured wind node, harmonic analysis is used to derive an annual function (Eq.2) with Fourier terms Ω, 2Ω and 3Ω, with Ω representing the annual angular frequency. This is subtracted from the raw dataset, and a diurnal function (Eq.3) with Fourier terms ω and 2ω derived with ω representing the diurnal angular frequency is derived from the residuals. The remainder is termed the 'de-trended' value y_{dt}.

$$y_{dt}(t) = y(t) - y_a(t) - y_d(t) \tag{1}$$

$$\begin{aligned} y_a(t) =& a_0 + a_1 sin(\Omega t + b_1) + a_2 sin(2\Omega t + b_2) \\ &+ a_3 sin(3\Omega t + b_3) \end{aligned} \tag{2}$$

$$y_d(t) = c_0 + c_1 sin(\omega t + d_1) + c_2 sin(2\omega t + d_2) \tag{3}$$

Further partitioning of data prior to the application of Eq. 3 may be appropriate for locations subject to seasonal variance in diurnal trends - this is discussed further in section III.

The generalised vector auto-regressive (VAR) process of order p for the de-trended data at n nodes can be expressed by Eq. 4:

$$\begin{aligned} \vec{y}_{dt}(t) =& \vec{\Phi_1}\vec{y}_{dt}(t-1) + \vec{\Phi_2}\vec{y}_{dt}(t-2) + ... \\ &+ \vec{\Phi_p}\vec{y}_{dt}(t-p) + \vec{e}(t) \end{aligned} \tag{4}$$

where $\vec{y}_{dt}(t)$ is the vector of size n comprising the detrended values for all nodes at time t, and $\vec{\Phi_1}$,$\vec{\Phi_2}$...$\vec{\Phi_p}$ are the $n \times n$ matrices describing the inter-relationships between the nodes at time lags 1,2...p. $\vec{e}(t)$ is a Gaussian noise term.

Through the use of Eqs. 1 through 4, a model of time lag p capable of synthesising time series data for a wind field comprising n nodes, including initial values for $\vec{y}(t-1)$,$\vec{y}(t-2)$...$\vec{y}(t-p)$, can be represented by $pn^2+(p+12)n$ parameters.

Simulated wind data is synthesised iteratively, by deriving the next term of the detrended process via Eq. 4, and adding the annual, diurnal and noise terms at each node.

III. Derivation of a VAR Model for the British Isles

The NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA) [18] is a state of the art historical reanalysis of global atmospheric observations from the Goddard Earth Observing System (GEOS-5), covering 34 years of observations from weather stations, balloons, satellites, ships and aircraft, and resolved to a gridded global model covering all major meteorological parameters.

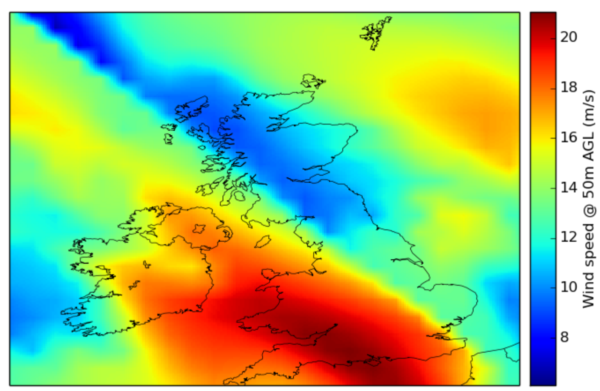

Figure 1. Snapshot of wind speeds at 50m AGL over the British Isles at midnight on the 1st of January 2015, interpolated from the MERRA dataset.

The MERRA dataset contains nodes corresponding to a gridded mesh with resolution $\frac{1}{2}$ degrees latitude by $\frac{2}{3}$ degrees longitude. In the latitude of the British Isles, this corresponds to a grid of approximately 55 by 44km. The dataset includes a set of 3 wind vectors averaged across each grid square at 2m, 10m and 50m above ground level, with a temporal resolution of 1 hour. A snapshot of this data for the British Isles is shown in Fig. 1.

For this study, only the 50m AGL values are used, as these are closest to the hub heights of modern wind generators. As a statistical wind-to-power conversion is used in the next section of this paper, extrapolation of wind speed to actual hub heights is not necessary - however, if this was required for a deterministic wind power model then the wind model presented

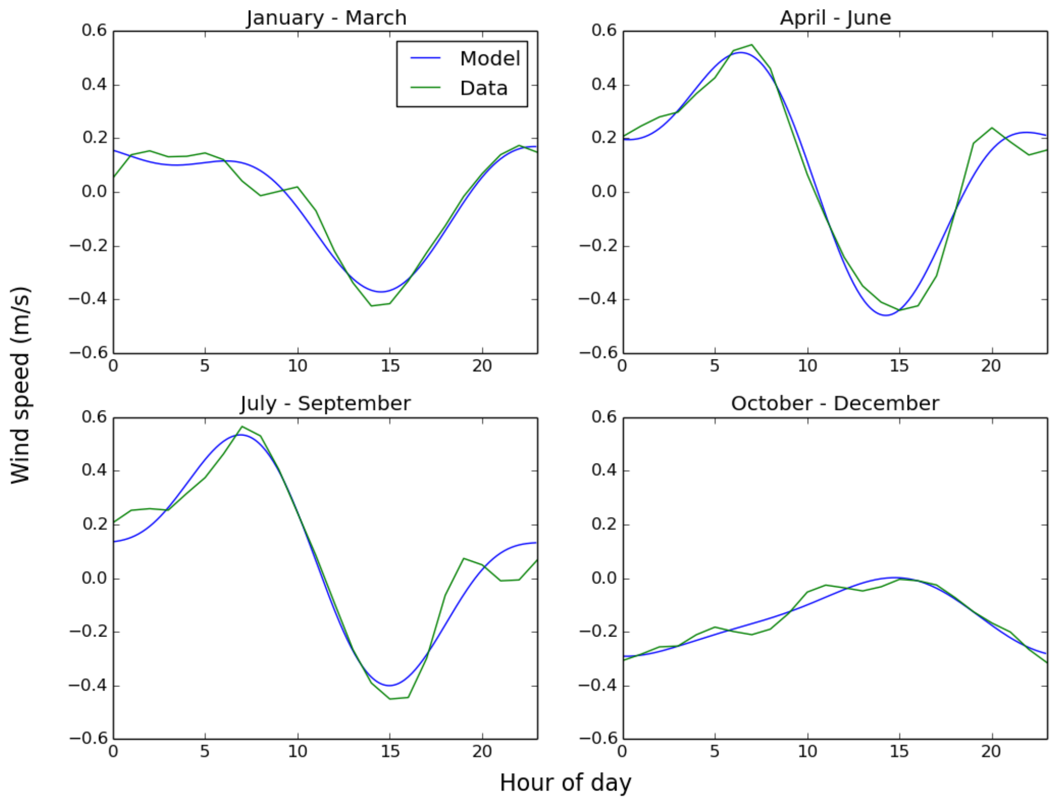

Figure 2. Fitted diurnal trends against hourly averages after annual detrending for node located at 55.5N,4.0W partitioned by season.

here could be repeated for the 3 height measurements and extrapolated at the site location following the methodology presented in [6].

Five years of data spanning the whole years 2010 to 2014 inclusive was retrieved from the MERRA dataset for the 552 nodes covering the British Isles as depicted in Fig. 1. For each node, the best-fit parameters for Eq. 2 were derived, and the best-fit parameters for Eq. 3 derived from the residuals. As the diurnal trends in the British Isles appear to vary significantly between seasons [16], the annual residuals were first partitioned into 4 3-month datasets in order to derive separate diurnal trends for each period - these are shown for an example node in Fig. 2. The summed annual and diurnal trends for one year are illustrated against the raw values for a sample node in Fig. 3. Fig. 4 illustrates the normality of the data once after detrending, which is required as the VAR process assumes normality.

The matrices for a VAR(3) model were then fitted to the remaining residuals across all nodes, using the Python Statsmodels package[1]. 10 years of data was simulated from the fitted model, using the final value of the training dataset as initial values, and added to the annual and diurnal trends at each node to produce the simulated wind time series for each node.

[1] http://statsmodels.sourceforge.net/

IV. Conversion of Wind Speed to Power

The wind speed to power conversion model for a wind generator at a specified location consists of two steps: firstly, the weighted average u_g of the wind speeds u_i at the nearest k weather nodes is calculated from a set of weightings $w_1, w_2, ..., w_k$, as in Eq. 5.

$$u_g(t) = \frac{\sum_{i=1}^{k} w_i u_i}{\sum_{i=1}^{k} w_i} \tag{5}$$

Secondly, the site power output value p_g is linearly interpolated by applying a discrete power curve function consisting of wind speed/power pairs $u_{pc}(m), p_{pc}(m)$, as in Eq. 6.

$$p_g(t) = p_{pc}(m) + \frac{u_g(t) - u_{pc}(m)}{u_{pc}(m+1) - u_{pc}(m)}(p_{pc}(m+1) - p_{pc}(m)) \tag{6}$$

where:

$$u_{pc}(m) <= u_g(t) < u_{pc}(m+1)$$

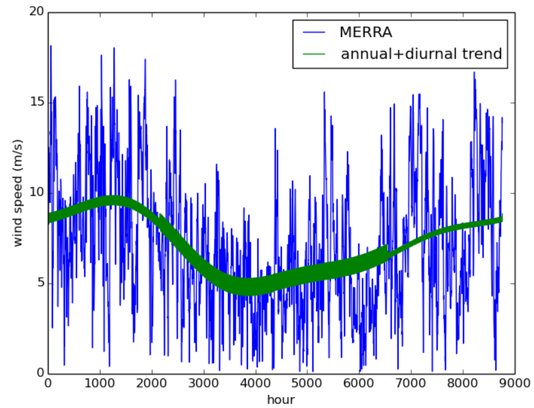

Figure 3. Raw MERRA wind data from 2014 for node located at 55.5N,4.0W with summed fitted annual and diurnal trends.

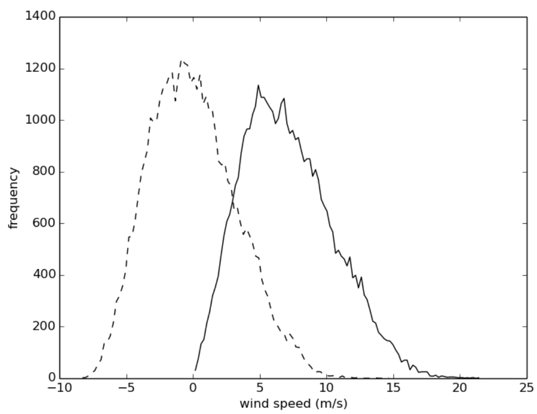

Figure 4. Distribution of wind speeds for node located at 55.5N,4.0W before (solid line) and after (dashed line) detrending.

In order to derive the node weightings and power curve speed/power pairs, metered energy data (on a half-hourly resolution) for wind generators participating in the GB Balancing Mechanism was retrieved for the period 1st January 2013 to 31st March 2015. This includes all transmission-connected wind generators, as well as larger embedded wind generators. The geographical distribution of the sites for which data was available throughout this period is shown in Fig. 5. Six further sites were excluded from the dataset due to being subject to significant export constraints during this period, leaving 46 generators totalling 4869MW of installed capacity. The data was aggregated by hour to match the resolution of the MERRA wind model.

Two filters were applied to the data in order to ensure, as far as possible, the dataset captured normal generator operation:

1) A curtailment filter was applied to remove all hourly datapoints where sites were subject to external constraints by the GB System Operator, using reported Bid-Offer Acceptance instructions from the published Balancing Mechanism data[2].

[2]www.bmreports.com

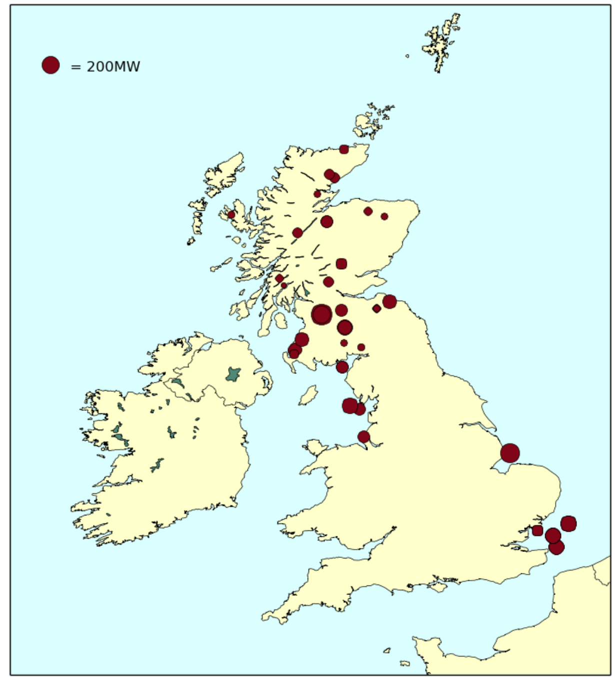

Figure 5. Location and relative capacities of the 46 wind generators used in the power simulation for Great Britain. Some smaller sites are obscured by larger generators in their vicinity.

2) A coarse availability filter was applied to remove hourly datapoints where it appeared that the site was unable to generate at non-nominal levels. This was achieved by performing an initial run of the power-curve fitting described below, and then filtering all points where the total output of the site was below 20% of the expected value, and the site was expected to be producing more than 10% of its rated power.

The dataset was split into training data covering 1st January 2013 to 30th June 2014, and validation data covering 1st July 2014 to 31st March 2015.

The power curve and wind weightings were derived by least-squares fitting against the filtered training data using the Levenberg-Marquardt algorithm. The 4 nearest weather nodes were used for each site, corresponding to the bounding points of the gridded model. Initial node weightings were set to being equal $w_i = 0.25 \ \forall i \in [1, 4]$, and the initial wind/power pairs are for a current pitch-regulated wind generator, aggregated across multiple turbines, as published in [19]. Power values outside of the domain of this power curve are assumed to be zero.

An example of a power curve generated by this process for one wind generator is given in fig. 6, illustrating the datapoints which were removed either due to being periods of external curtailment, or through the coarse availability filter described above.

In order to synthesise power time series for each generator, the derived node weightings and power curve values were applied on a per-site basis to the wind time series generated from the VAR(3) model.

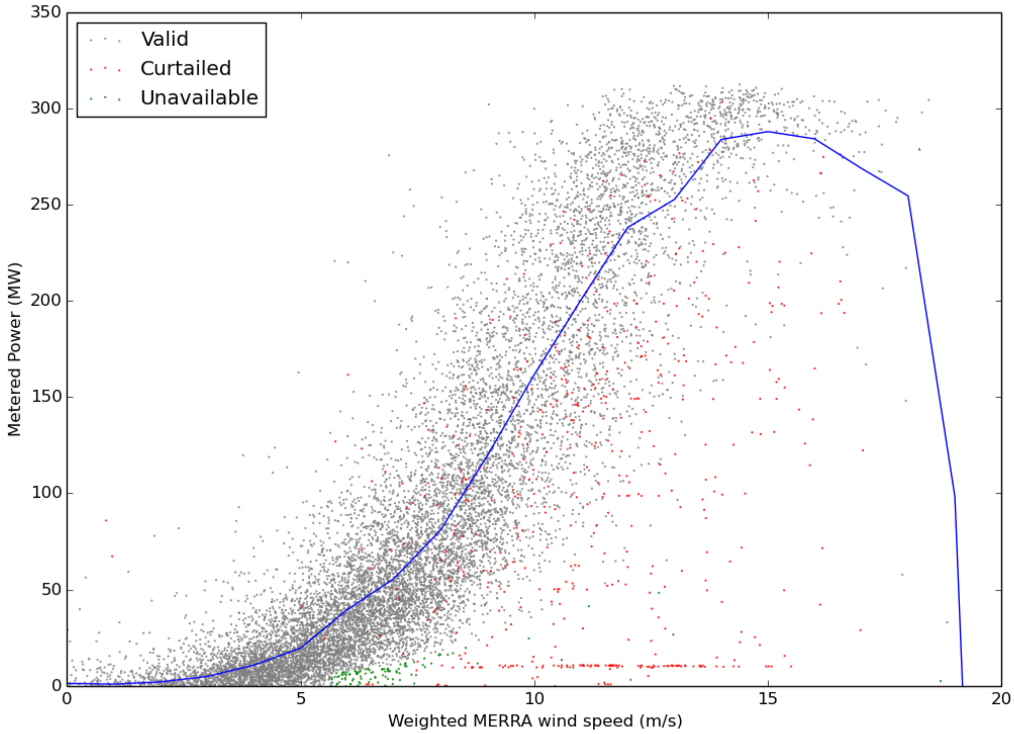

Figure 6. Derived power curve for Whitelee 1 Wind Farm, showing filtered data points.

V. Validation

The validation dataset specified in the previous section was used to determine the accuracy of the derived power curves to the metered wind generator output. The average R^2 value across all 46 wind generators was 0.839, with a range of 0.770 to 0.899. A large proportion of the error appears to be generated around the rated power of the site, where a small change in wind speed can have a large effect on the power output, from turbines shutting down due to high wind speeds. Other sources of error will include time-variant properties such as site operational availability (other than that captured by the data filter), local terrain effects in combination with wind direction, and local turbulence conditions. An interesting outcome was that the average r-squared value improved by 0.02 if the wind weightings used in Eq. 5 were not constrained to positive values - this may reflect the temporal lag in weather fronts moving across neighbouring MERRA nodes.

The outputs of the VAR simulation must be validated according to two criteria: the extent to which spatial correlations between wind locations are preserved, and the extent to which rates of change of power are reproduced. It is not possible to directly compare the VAR synthesis to measured data and extract e.g. a Mean Absolute Error (such as for a traditional forecasting methodology), as only the starting vector of wind speeds $\vec{y}(t_0)$ is derived from physical data. Hence the distributions of linear and spatial correlations are compared to that of the original data instead.

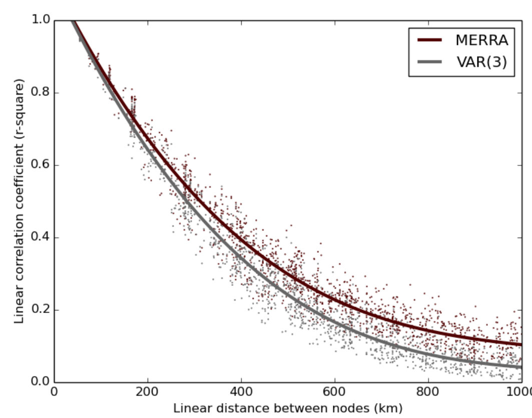

Figure 7. Comparison of linear correlation between randomly sampled node pairs from the original MERRA wind data and the VAR(3) simulated wind data.

Fig. 7 shows the spatial correlations between randomly sampled wind node pairs in the original MERRA dataset compared to the outputs of the VAR(3) model. This demonstrates a close reproduction of the extent of spatial correlation observed in the reanalysis data. While the VAR(3) model shows a slightly

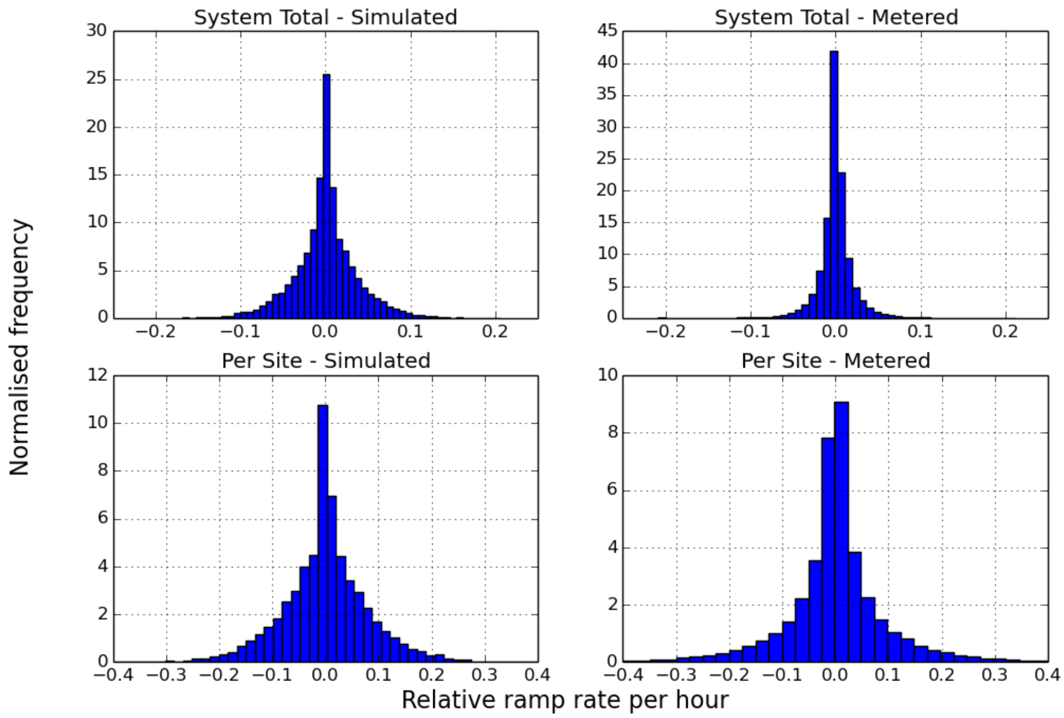

Figure 8. Ramp rate distributions as a proportion of total rated power.

lower spatial correlation for node pairs located at significant distances, the correlation values have a lower variance around the best-fit line. One possible explanation for this difference is that large-scale wind patterns are often driven by complex weather systems which may lead to increased correlation over significant distances during particular events - as the VAR model is a statistical model rather than a meteorological one, it is incapable of reproducing these events.

Fig. 8 shows the ramp rates at each wind generator, and the total ramp rates for all generators in the system, derived as a proportion of rated power (either for the individual site or the total capacity of 4869MW). This shows that the distributions of ramp rates for the simulated data are very similar to that of the raw metered dataset. The simulated distributions show a slightly higher variance (i.e. display proportionately greater ramp rates) than in the original metered data.

VI. Conclusions

The statistical wind power model presented in this paper demonstrates a means of representing wind injections into a power network using a relatively low number of parameters with few computational requirements compared to direct use of large-scale numerical weather models. Similarly, the iterative nature of the process of synthesising time series data means that it can be used for 'just-in-time' simulation without the need for storage of large volumes of data. This makes it particularly suitable for power system studies concerned with issues such as network and reserve adequacy, which may require the simulation of many years of operation in order to analyse all possible operational scenarios.

The statistically-derived wind-to-power conversion model provides a good fit against metered data, but use of, and validation against, data from extant wind generation remains difficult in terms of quantifying the impact of local conditions both in terms of wind flow/turbulence and operational availability of turbines. While theoretical wind-to-power models provide some ability to quantify these factors in power system simulations, validation of such models against national-scale datasets still requires these issues to be addressed.

It is noted that in this study, and others which use similar data sources, large-scale wind datasets have (by necessity of computational burden) a low temporal resolution, and a power system planner may be interested in ramp rates and correlations on a sub-hourly resolution. However, spectral analysis of wind speeds shows a significant spectral gap at a frequency of around 1 hour [20], suggesting that it may be suitable to utilise a seperate statistical model - likely with a high proportion of noise - to interpolate between hourly values derived from a model such as has been presented.

References

[1] J. O. Oswald, M. Raine, and H. Ashraf-Ball, "Will British weather provide reliable electricity?" *Energy Policy*, vol. 36, pp3212-3225, 2008.

[2] S. Hawkins, D. Eager and G. P. Harrison, "Characterising the reliability of production from future British offshore wind fleets." in: *IET Renewable Power Generation Conference*, September 2011.

[3] ESB, "Impact of wind power generation in Ireland on the operation of conventional plant and the economic implications." Technical Report, February 2004.

[4] D. J. Cannon, D. J. Brayshaw, J. Methven, P. J. Croker and D. Lenaghan, "Using reanalysis data to quantify extreme wind power generation statistics: A 33-year case study in Great Britain." *Renew. Energy*, vol. 75, pp767–778, 2015.

[5] M. L. Kubik, D. J. Brayshaw, P. J. Croker and J. F. Barlow, "Exploring the role of reanalysis data in simulating regional wind generation variability over Northern Ireland." *Renew. Energy*, vol. 57, pp558–561, 2013.

[6] I. Staffell and R. Green, "How does wind farm performance decline with age?" *Renewa. Energy*, vol. 66, pp775–786, 2014.

[7] M. Gibescu, B. C. Ummels and W. L. Kling, "Statistical wind speed interpolation for simulating aggregated wind energy production under system studies," in: *Proc. 9th int. Conf. Probabilistic Methods Applies to Power Systems*, Stockholm, Sweden, June 1-7, 2006.

[8] G. Papaefthymiou and G. Klöckl, "MCMC for wind power simulation." *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 234 240, 2008.

[9] G. Papaefthymiou and D. Kurowicka, "Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis." *IEEE Transactions on Power Systems* vol. 24, no. 1, pp. 4049, 2009

[10] C. L. Archer, and M. Z. Jacobson, "Supplying Baseload Power and Reducing Transmission Requirements by Interconnecting Wind Farms." *Journal of Applied Meteorology and Climatology*, vol. 46, no. 11, pp. 1701-1717, Nov. 2007.

[11] W. Kempton, F. M. Pimenta, D. E. Veron, and B. A. Colle, Electric power from offshore wind via synoptic-scale interconnection. *Proceedings of National Academy of Sciences USA*, vol. 107, no. 16, pp. 72407245, April 2010.

[12] S. Zachary, C. Dent, D. J. Brayshaw, "Challenges in quantifying wind generation's contribution to securing peak demand." In: *IEEE Power and Energy Society General Meeting*, Detroit, MI, 2011.

[13] D. A. Smith and K. C. Mehta, "Investigation of stationary and nonstationary wind data using classical Box Jenkins models." *J. Wind Eng. Ind. Aerodynamics*, vol. 44, pp319–328, 1993.

[14] R. Billinton, H. Chen and R. Ghajar, "Time-series models for reliability evaluation of power systems including wind energy." *Microelectron. Reliability*, vol. 36, no. 9, pp1253-1261, 1996.

[15] M. J. Duran, D. Cros and J. Riquelme, "Short-term wind power forecast based on ARX models." *J. Energy Eng.*, vol. 133, no. 3, pp172–180, 2007.

[16] D. Hill, D. McMillan, K. R. W. Bell and D. Infield, "Application of Auto-Regressive Models to UK Wind Speed Data for Power System Impact Studies." *IEEE Trans. Sus. Energy*, vol. 3, no. 1, pp134–141, 2012.

[17] H. Lutkepohl, "New introduction to multiple time series analysis." Springer-Verlag, New York, 2005.

[18] M. M. Rienecker et al, "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications." *J. Climate*, vol. 24, pp3624–3648, July 2011.

[19] Garrad Hassan and Partners Ltd, "TradeWind Work Package 2.6 Report: Equivalent Wind Power Curves" Technical Report, July 2008

[20] I. Van der Hoven, "Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour." *J. Meteor.*, vol. 14, pp160-164, 1957.