DOI: 10.1016/j.renene.2016.05.024

# A simple hourly wind power simulation for the South-West region of Western Australia using MERRA data

**Dean Laslett[a]*, Chris Creagh[a], and Philip Jennings[a]**

*Murdoch University Western Australia*

(a) School of Engineering and IT, Murdoch University, South Street, Murdoch, WA 6150, Australia

*gaiaquark@gmail.com (Corresponding author)

## Abstract

A simple simulator capable of generating synthetic hourly values of wind power was developed for the South West region of Western Australia. The global Modern Era Retrospective Analysis for Research and Applications (MERRA) atmospheric database was used to calibrate the simulation with wind speeds 50m above ground level. Analysis of the MERRA data indicated that the normalised residual of hourly wind speed had a double exponential distribution. A translated square-root transformation function $y_n=(\sqrt{(1.96+y_e)}-1.4)/0.302$ was used to convert this to a normal-like distribution so that autoregressive (AR) time series analysis could be used. There was a significant dependency in this time series on the last three hours, so a third order AR model was used to generate hourly 50m wind speed residuals. The MERRA daily average 50m wind speed was found to have a Weibull-like distribution, so a square root conversion was used on the data to obtain a normal distribution. The time series for this distribution was found to have a significant dependency on the values for the last two days, so a second order AR model was also used in the simulation to generate synthetic time series values for the square root of the daily average wind speed. Seasonal, daily, diurnal, and hourly components were added to generate synthetic time series values of total 50m wind speed. To scale this wind speed to turbine hub height, a time varying wind shear factor model was created and calibrated using measured data at a coastal and an inland site. Standard wind turbine power curves were modified to produce an estimate of wind farm power output from the hub-height wind speed. Comparison with measured grid supervisory control and data acquisition (SCADA) data indicated that the simulation generated conservative power output values. The simulation was compared to two other models: a Weibull distribution model, and an AR model with normally distributed residuals. The statistical fit with the SCADA data was found to be closer than these two models. Spatial correlation using only the MERRA data was found to be higher than the SCADA data, indicating that there is still a further source of variability to be accounted for. Hence the simulation spatial correlation was calibrated to previously reported findings, which were similar to the SCADA data.

**Keywords:** Wind Power, Simulation, Western Australia, MERRA, ARMA, Exponential distribution.



## 1. Introduction

With the increasing focus on low emission power generation systems to mitigate global warming and the successful operation of several wind farms in the South West region of Western Australia (SWWA), it becomes worthwhile to consider the potential for expansion of wind power generation in this region. The SWWA is characterised by a Mediterranean climate [1], which is dominated by the eastward passage of high pressure sub tropical anti cyclonic cells. Mainly in winter, low pressure systems from the south cross the state every seven to ten days. Hence there are distinct differences in the seasonal wind speed variation at different places within SWWA. Frequently, there is a strong diurnal sea/land breeze along the coastline [2], more often in the summer months. This sea breeze can also penetrate as far inland as Kalgoorlie [3], which is about 350km from the nearest coast.

The wind speed at any site can be represented as the sum of several components operating at different temporal scales: seasonal, daily, diurnal, dependent and random. The seasonal component arises from the cyclical variation in the prevailing atmospheric systems as the earth orbits the sun. The daily component arises from the passage of weather systems across a region with typical durations from 2 to 8 days [4]. The diurnal component arises from the sea/land breeze system caused by temperature differences between the land and ocean. The dependent component arises because atmospheric phenomena can be persistent, resulting in a relationship between the wind speed at a particular time to the wind speed at previous times. Finally, most physical processes contain a random fluctuation component and wind speed is no different.

For a model to adequately represent the wind power generation potential at any one place in the SWWA, it is necessary to capture the variability at each temporal scale [5]. It will also be necessary to capture the spatial differences in these variabilities across the whole region of the SWWA. There have been several simple models that generate synthetic time series values of wind speed at one or more sites (eg. [6], [7]). These models attempt to mimic the observed statistical nature of the wind speed. There are also detailed models of wind speed at multiple sites or across a region that use meteorological physics, and tend to require much more computing power [8]. This study will focus on the development of a statistical model designed to operate across the SWWA region.

The two parameter Weibull distribution has been the most widely used simple statistical representation of overall wind speed behaviour [8]. The probability density function for this distribution is given by:

$$f(v)=\left(\frac{k}{\lambda}\right)\left(\frac{v}{\lambda}\right)^{k-1} e^{-\left(\frac{v}{\lambda}\right)^{k}} \quad v \geq 0$$

(1)

$$f(v)=0 \quad v < 0$$

Where v is the wind speed (m/s), f(v) is the probability density function, $k$ is the shape parameter, and $\lambda$ is the scale parameter. However, Carta et al. [9] also reviewed other probability density functions used to represent wind speed frequencies, and concluded that although the Weibull distribution has some advantages over other distributions, it cannot adequately represent many of the wind speed probability density functions that might be encountered in the real world. Gunturu and Schlosser [10] found that use of the Weibull distribution could lead to both over and under estimations of the wind power resource available.

Auto Regressive Moving Average (ARMA) models [11] have also been widely applied to the statistical representation and prediction of many kinds of time series data (for example [12], [13] and [14]) as well as wind speeds. ARMA models are a combination of Auto Regressive (AR) models, and moving average (MA) models, where the wind-speed value at time t is represented as the sum of a linear combination of wind speed values at previous times and the linear combination of a series of random values. Purely Auto Regressive models use only the random value at the present time:

$$y(t)=\sum_{k=1}^{p} \varphi_k \, y(t-k)+\rho \, r(t) \tag{2}$$

Where y(t) is the wind speed residual at time mark t, y(t-k) is the wind speed residual at timemark t-k, and r(t) is a series of uncorrelated white noise error values which is identically distributed with a normal frequency distribution, zero mean, and standard deviation of one. y(t) is multiplied by the wind speed standard deviation and then added to the mean wind speed to get a wind speed value. $\phi_k$ are the AR parameters, and $\sigma$ is the random noise parameter. The value of $\varrho$ is adjusted depending on the value of the AR parameters so that the standard deviation of y(t) remains at one. The AR order p is the maximum value of k with a non-zero value of $\phi_k$. This is commonly written as an AR(p) model. ARMA models can capture the temporal dependency inherent in wind speed time series, while using a simple Weibull distribution cannot. However, Papaefthymiou and Klockl [15] asserted that the frequency distribution (equivalent to the probability density function or PDF) of ARMA models rarely match the measured data, which can lead to under or over estimation of wind power.

Wind speed behaviour can also vary over several temporal scales, such as seasonal, daily, diurnal, and hourly. Seasonal variation is commonly modelled using one or more sinusoidal cycles (eg, [8] and [16]). Daily average wind speeds vary from the seasonal average and can have a skewed distribution [17]. Weibull, log-normal, modified normal and modified exponential distributions have been used to represent these distributions (eg [17], [18], [19]). Carlin and Haslett [20] proposed the use of a "squared normal" distribution to simply model Weibull-like distributions, based on Western Australian wind data. Daily wind speeds have also been found to have an autoregressive dependency (eg [21], [22], [23]).

A common way of modelling diurnal trends has been to calculate the average measured wind speed at every hour of the day for each month or season (eg [8]). Fixed cyclic functions have also been used (eg [19]). However these approaches don't explicitly catch the variation in peak daily wind speed magnitude and time that occurs throughout each month or season. ARMA models and high order AR models have also been developed that model diurnal variation (eg [6], [24]). Suomalainen et al. [23] concluded that these approaches were not sufficiently realistic and developed a model that identified day types defined by the time of day that the peak wind speed occurs, and defining a diurnal pattern for each day type.

After the seasonal, daily, and diurnal components of wind speed have been removed, what is left is the de-trended hourly wind speed. Similarly to the daily wind speed, the value at a particular time has a dependency on the values at previous times, and ARMA models have been commonly used to model this effect.

However, the above form of ARMA equation has been found to be generally suitable for use only if the time series data and the error values are normally distributed. If the data is not normally distributed, then the choice of distribution for the random error values needed to produce the same distribution as the data is not clear [25]. For example, Ward and Boland [16] found that de-trended wind speeds at sites in South Australia had a double exponential distribution (also called a Laplace distribution). But Damsleth and El-Shaarawi [26] found that even the simplest AR model (of order 1) would not necessarily generate a time series with a double exponential distribution, even if the random variable was given a double-exponential distribution. Lawrance and Lewis [25] suggested an alternate form of auto-regressive equation, but with impractical restrictions on the allowable values of the auto-regressive coefficients.

A possible solution is to convert the de-trended wind speed time series values into a normal distribution using a data transformation function. Mach et al. [27] tested a number of transformations on different types of data. If the data is found to have an exponential distribution, then the authors recommended a power transformation to convert the data to a normal distribution. Although a double exponential distribution is symmetric about the mean, unlike a standard exponential distribution, this might point the way to a suitable transformation function. If the data is found to have a Weibull-like distribution (such as daily wind speeds), then Mach et al. [27] recommended the use of a Box-Cox or power law transformation to convert to a normal distribution. Widger [28] used the square-root normal distribution to model wind speeds, suggesting that taking the square-root of the data (power law 1/2) may effectively convert a Weibull-like distributed wind speed time series into a normal-like distributed series. Carlin and Haslett [20] used a square-root transformation function on Western Australian wind data, and Brown et al. [29] used a square-root transformation function on data from the Pacific Northwest region of North America.

For an interactive hourly wind speed simulation, limiting the numerical complexity is important. However the simple Weibull model will be insufficient, as it does not account for persistence at this time scale. An ARMA model or some other model that accounts for persistence must be used. If a site has a significant diurnal component, then this must also be accounted for. This study found wind speed residuals that were not normally distributed. It was shown that using a model with normally distributed residuals led to significantly different, and less representative, statistical behaviour of the resulting wind power time series.

Several wind farms, each using a different wind turbine, are now often present on large scale electrical power grids. Hence once a representative time series of wind speeds at a reference height above ground has been generated, two further steps must be taken: scaling the wind speed to the hub height of a particular turbine, and then converting the scaled wind speed to an electrical power output.

The horizontal wind speed at different heights above ground often have different values. This effect is commonly called wind shear. There have been two equations commonly used to characterise the wind shear. The first is the logarithmic relationship:

$$\frac{v_2}{v_1}=\frac{\ln\left(\frac{h_2}{z_o}\right)}{\ln\left(\frac{h_1}{z_o}\right)} \tag{3}$$

Where $v_1$ is the wind speed at height $h_1$, $v_2$ is the wind speed at height $h_2$, and $z_o$ is the roughness length. This relationship is based on the physical aspects of atmospheric meteorological behaviour [30]. The simpler power law approximation is given by:

$$\frac{v_2}{v_1}\approx\left(\frac{h_2}{h_1}\right)^{\alpha} \tag{4}$$

Where $v_1$ is the wind speed at height $h_1$, $v_2$ is the wind speed at height $h_2$, and $\alpha$ is the wind shear exponent, often set to 1/7 [24]. The logarithmic relationship implies that the wind shear at any one site does not change with time, and is based on the assumption that the atmosphere is in a neutrally stable condition where vertical air movement is neither encouraged or resisted [10]. In wind power studies, this assumption is commonly justified by the idea that when horizontal wind speeds become high enough to start generating power, mixing will ensure the atmosphere becomes neutrally stable [31]. Thus inaccuracies due to the atmosphere being in a different state are more likely to occur at low wind speeds which will have little impact on the prediction of generated power. However, both Smith et al. [32] and Rareshide et al. [33] found that this is not always the case, especially in inland areas, and there might be significant diurnal and seasonal variation in wind shear factor at high wind speeds. Smith et al. [32] found that wind shear was generally higher at night and lower during the day, when it might even be negative.

The specific power per unit area, P ($W/m^2$), in wind flowing past a wind turbine is a function of air density $\sigma_{air}$ and the cube of wind-speed v:

$$P=\frac{1}{2}\sigma_{air}v^3 \quad W/m^2 \tag{5}$$

In practice, there is an upper limit to the fraction of this power that can be harvested, and a common approach to transforming wind speed into power output is to use a wind turbine power curve [8], which is a non-linear transformation function (Fig. 1). In part of the middle region of the curve, power output is proportional to the air density [10], hence an implicit assumption when using wind power curves is that the air density vertical profile at the turbine operational site is similar to the profile where the turbine was tested.

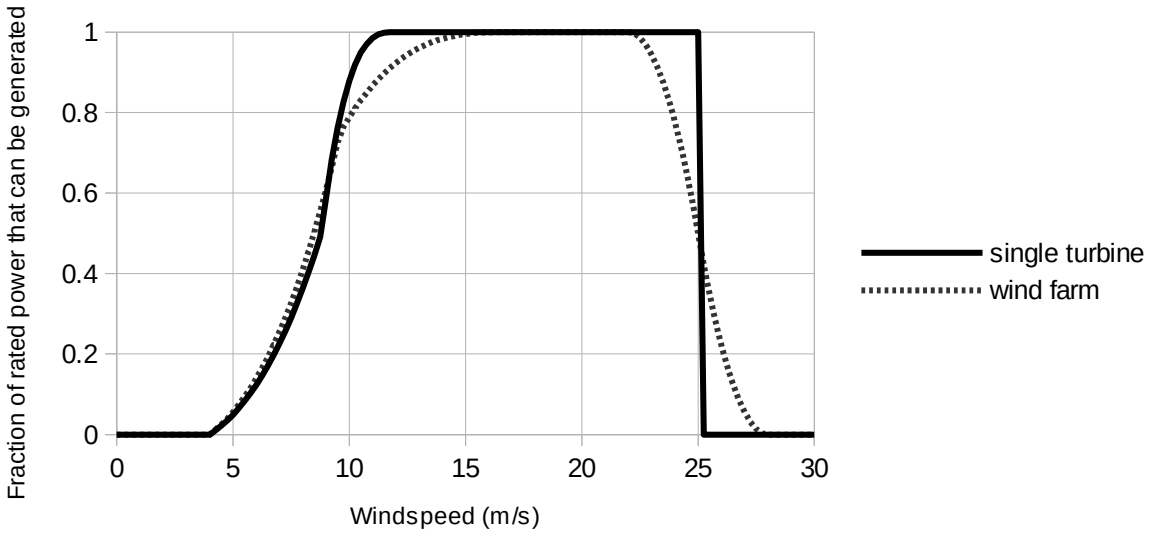


**Fig. 1.** Typical wind turbine and wind farm power curve.

Holttinen [34] reported that the wind power factor curve for an individual turbine must be modified if the wind power output from a whole wind farm, constructed using the same turbines, is required. This is probably due to variations in wind speed hitting different turbines within the wind farm. Generally Holttinen [34] used a gentler full power transition slope and shut down slope with a decreased shut-down wind speed (Fig. 1).

Because weather systems and hence wind patterns can extend over a wide area, wind farms sited close to each other are likely to have a significant correlation in wind speed and power output over time, but the correlation will decrease as the distance separating wind farms increases [35]. Kavasseri and Nagarajan [36] speculated that there would be less spatial correlation over shorter time scales because of local differences in topography and atmospheric behaviour, and more correlation over longer time scales due to global and regional weather system seasonal effects. Haslett and Raftery [24] examined sites in Ireland and found a decaying exponential relationship for the correlation between hourly wind speeds at two sites and the distance between sites, with the exception that sites very close together but not coincident can have a correlation significantly less than one. Carlin and Haslett [20] reported decreasing wind speed correlation with distance for sites in Western Australia.

A regional SWWA wind power simulation model should take this phenomenon into account. Correia and Ferreira de Jesus [37] developed a first order vector AR model with user specified spatial correlation between several sites, and Gibescu et al. [38] used a decaying exponential relationship to model the spatial correlation between wind speeds at different sites.

In this study, a simulation was developed for the purpose of modelling wind power generation for any site in the SWWA. Synthetic wind speeds were generated using square-root transformations of a normal distribution and AR models. Historical 50 metre wind speed data was used to calibrate the simulation at each temporal scale: seasonal, daily, diurnal, and hourly dependent, with a random fluctuation component

also added. Spatial correlation was introduced by creating distance weighted semi-dependencies in the random numbers used to generate the daily and diurnal components of wind speed.

A spatially and temporally dependent wind shear conversion factor model was developed so that wind speed at different turbine hub-heights could be estimated from these 50m wind speeds. Measured wind data at two sites was used to calibrate the wind shear conversion factor model. Finally, synthetic wind farm power output data was generated from the hub-height wind speed using modified wind turbine power curves. The simulation wind power output was compared to measured supervisory control and data acquisition (SCADA) wind farm power output data at 6 existing wind farm sites. Grasmere and Albany were considered to be separate, though adjacent, wind farms because different wind turbines are used at each site. The fit of the seasonal averages, and the daily, diurnal and hourly frequency distributions (equivalent to the probability density function or PDF) between the simulation and the SCADA data was compared, and it was shown that using a Weibull model, or even an auto regressive model with normally distributed residuals was not sufficient to represent the statistical behaviour of the measured wind power output.

## 2. Method

To obtain hourly wind speeds near the hub-heights commonly used in modern wind farms, the Modern-Era Retrospective Analysis for Research and Applications (MERRA) database [39] was accessed. A grid of hourly wind speeds at 50 metres above the surface of South Western Australia was obtained from this database. The grid contains 330 nodes (15 x 22) with a spacing of 2/3° in longitude (approximately 62.5 km), and 1/2° in latitude (approximately 56 km). The South West corner of the grid is at 112° east, 36° south, and the North East corner is at 126° east, 29° south. 106 of these nodes which are over land or close to the coastline were used to develop the simulation, with most concentrated near the coast-line (Fig. 2).

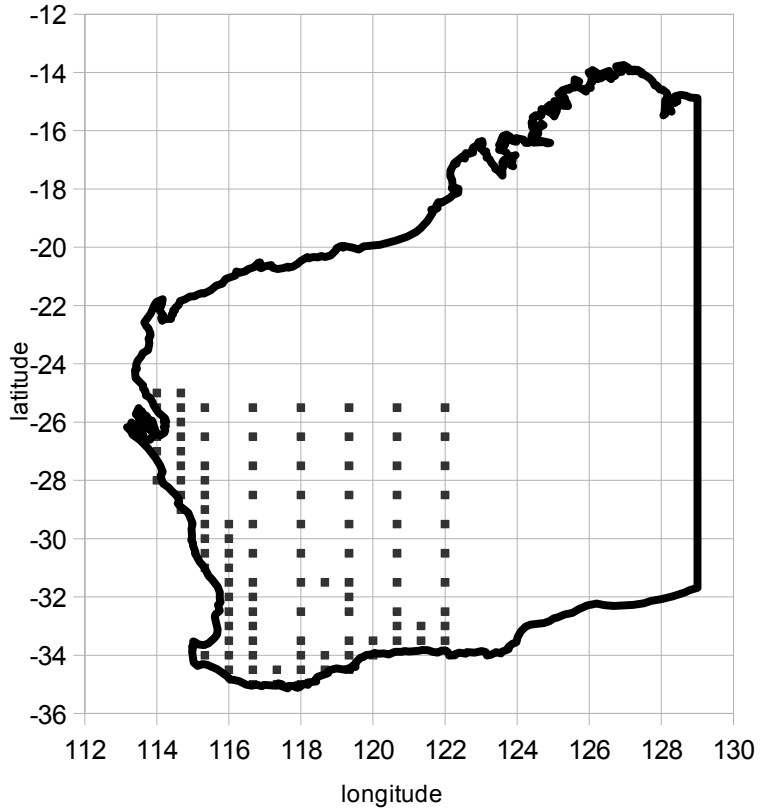


**Fig. 2.** MERRA nodes used to simulate wind speed in the South-West region of Western Australia.

The wind speeds were divided into 4 components for analysis: seasonal, daily, diurnal, and hourly. For all components, the distance from the coast of the wind farm was an important parameter. In a process similar to Laslett et al. [40], a shape map of the Western Australian Coastline was constructed from the GEODATA COAST 100K 2004 data package published by Geoscience Australia [41]. This data set is based on a 1:100,000 scale map sheet. The shape map consists of a vector map of the coastline and state border in longitude and latitude coordinates. It does not include any of the islands off the coast of Western Australia that are included in the data package. A global simplification algorithm [42] was used to simplify the map down to a 500 vertex coastline map.

The yearly average MERRA 50m wind speed varies across the SWWA with an average of 6.7 m/s and SD of 2.6 m/s. Two modes of seasonal variation were recognised in the monthly average MERRA wind speeds (For examples, see Fig. 3). The first mode has a maximum during the summer months and a minimum during winter, and is predominant at mid latitudes. The second mode is significant at southern latitudes. Both are attenuated at sites further inland.

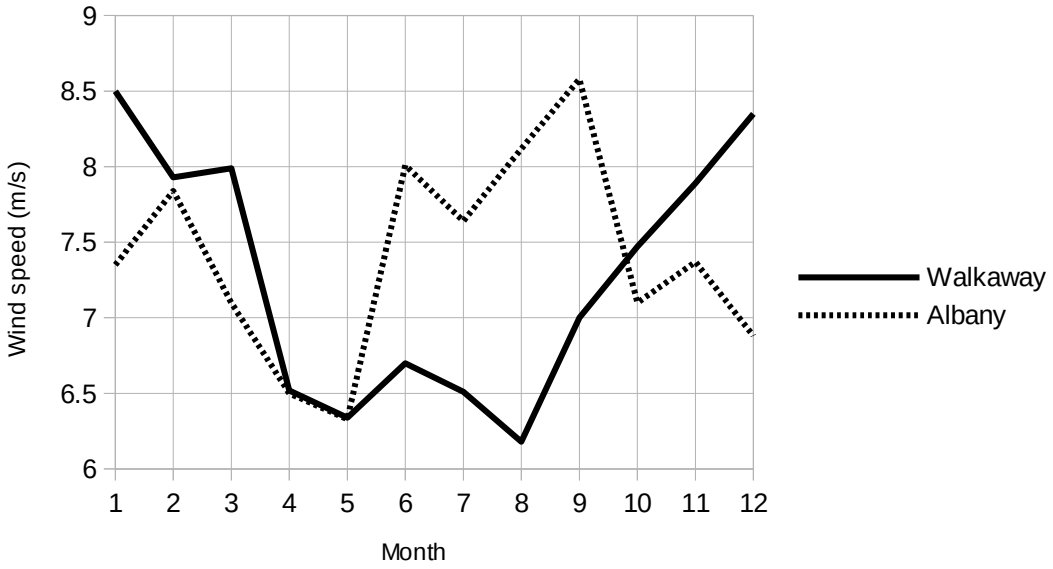


**Fig. 3.** Seasonal MERRA wind speeds near Walkaway and Albany wind farms.

To simulate seasonal wind speed at a particular site, the yearly average wind speed $V_{yav}$ was estimated to be the linear distance weighted average of the four yearly average wind speeds from the surrounding MERRA grid square. The seasonal wind speed at any particular day of the year was then calculated from $V_{yav}$ using a weighted combination of each seasonal mode:

$$V_{season} = V_{yav}\left(1 + kslat_1 V_{mode1}(DOY) + kslat_2 V_{mode2}(DOY)\right) \tag{6}$$

where DOY is the day of the year, $V_{season}$ is the seasonal wind speed at DOY, $kslat_1$ and $kslat_2$ are the weighting coefficients for each mode, and $V_{mode1}$ and $V_{mode2}$ are the magnitudes of each mode at DOY. $kslat_1$ and $kslat_2$ were found to have a dependency on latitude and distance from the coast. $V_{mode1}$ and $V_{mode2}$ were represented using piecewise linear functions. See Appendix A for the precise parameterisations.

The distribution of MERRA daily average 50m wind speeds was found to have a similar shape to a translated Weibull distribution (Fig. 4). Similarly to Carlin and Haslett [20], the square-root residual of the MERRA daily average wind speed was found to have a normal-like distribution (Fig. 5). The square-root residual was obtained by subtracting the mean of the square root wind speeds and then dividing by the standard deviation of the square root wind speeds. The standard deviation $\sigma_d$ was found to have both a spatial and seasonal dependency. See Appendix A for parameterisations.

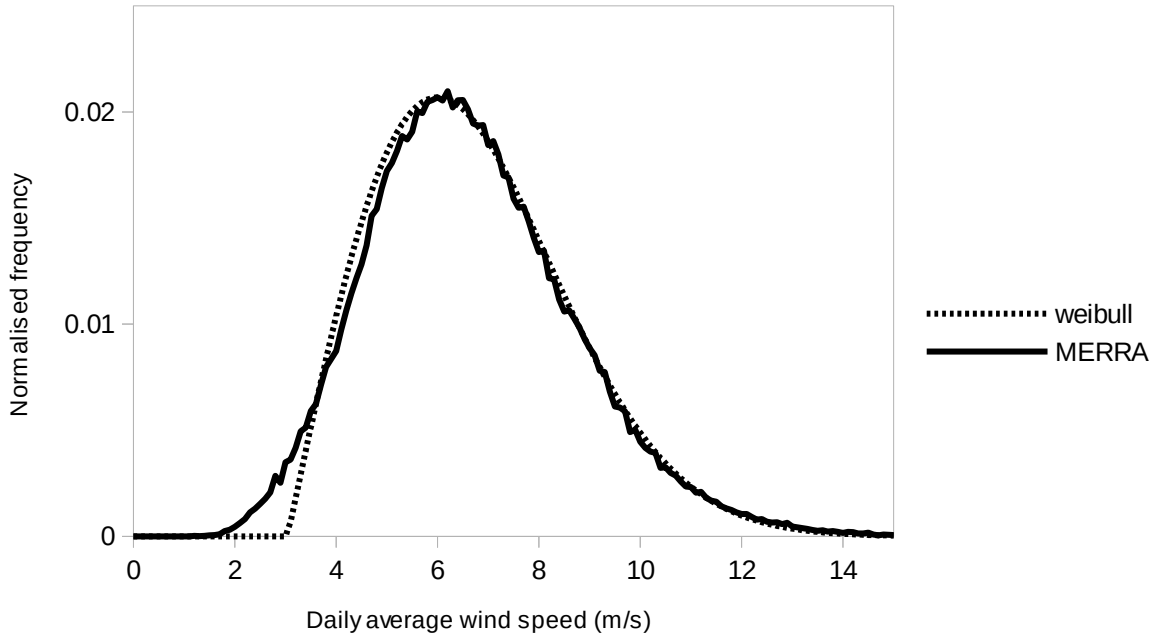


**Fig. 4.** Normalised frequency distribution of MERRA daily average 50m wind speed compared to a translated Weibull distribution. Horizontal axis bin width is 0.1 $ms^{-1}$, which is close to the value of 0.093 $ms^{-1}$ suggested by the Freedman-Diaconis rule [43].

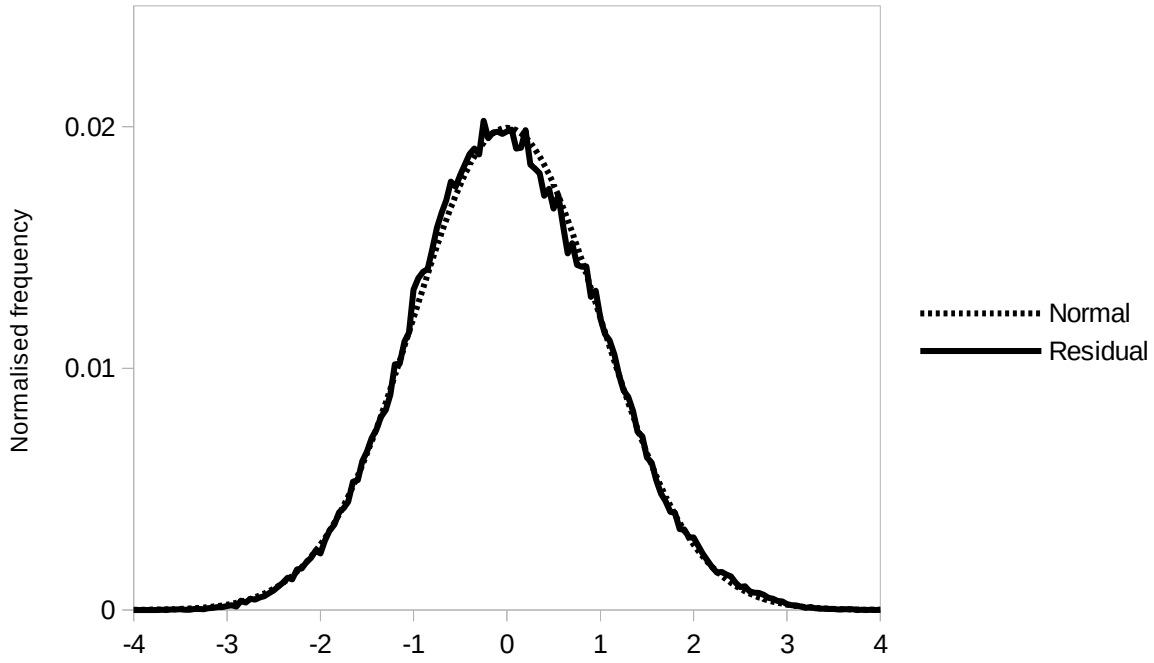


**Fig. 5.** Normalised frequency distribution of MERRA daily average 50m wind speed square-root residual compared to normal distribution. Horizontal axis bin width is 0.05, which is close to the value of 0.048 suggested by the Freedman-Diaconis rule [43].

Examination of the auto-regression and partial auto-regression coefficients of the square-root residuals at each MERRA node (for example Fig. 6 and Fig. 7) indicated a possible auto-regressive (AR) signature with dependency of order two. The dependency could also possibly be a second order moving average MA(2) or combined ARMA(1,1) model. For the residual at each node, the least squares method [44] and numerical maximum likelihood estimation was used to calculate the root mean square error (RMSE) for ARMA models with coefficients up to order (5,4) (for example Table 1). These indicated that the pure AR models gave sightly lower RMSE values. For increasing AR order, the RMSE initially decreased and then substantially levelled off after order 2. To confirm that an AR order of two was necessary and sufficient to capture most of the dependency within the time series, the Bayesian Information Criterion (BIC) [45] was used in the following form:

$$BIC = n\log_e\left(RMSE^2\right) + \left(p+q+1\right)\log_e n \tag{7}$$

Where n is the number of data points (1827), p is the AR order and q is the MA order. BIC was calculated and ranked in ascending order for each ARMA(p,q) model, $0 \le p \le 5$ and $0 \le q \le 4$, at each node (for example Table 2). The model with the lowest BIC was ARMA(2,0) or AR(2). The value of the AR(2) coefficients were found to be fairly consistent across all of the MERRA nodes, so the average coefficient values were used in the simulation.

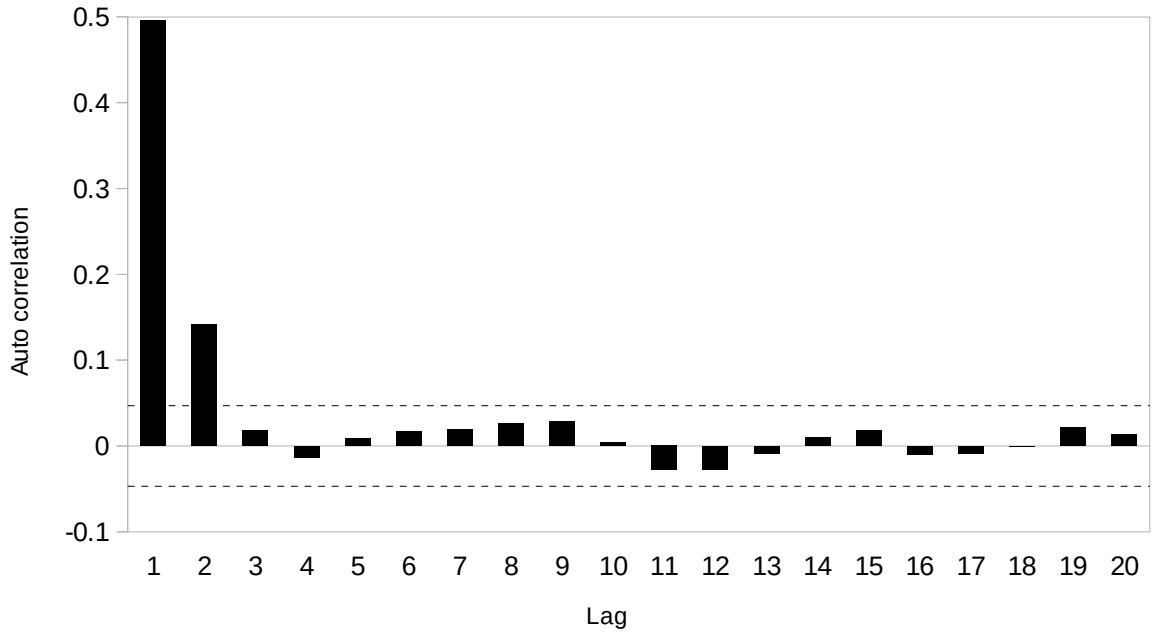


**Fig. 6.** Example auto-correlation of the MERRA daily average wind speed square-root residual at a single node. Dashed lines indicate 95% significance levels for a population value of zero.

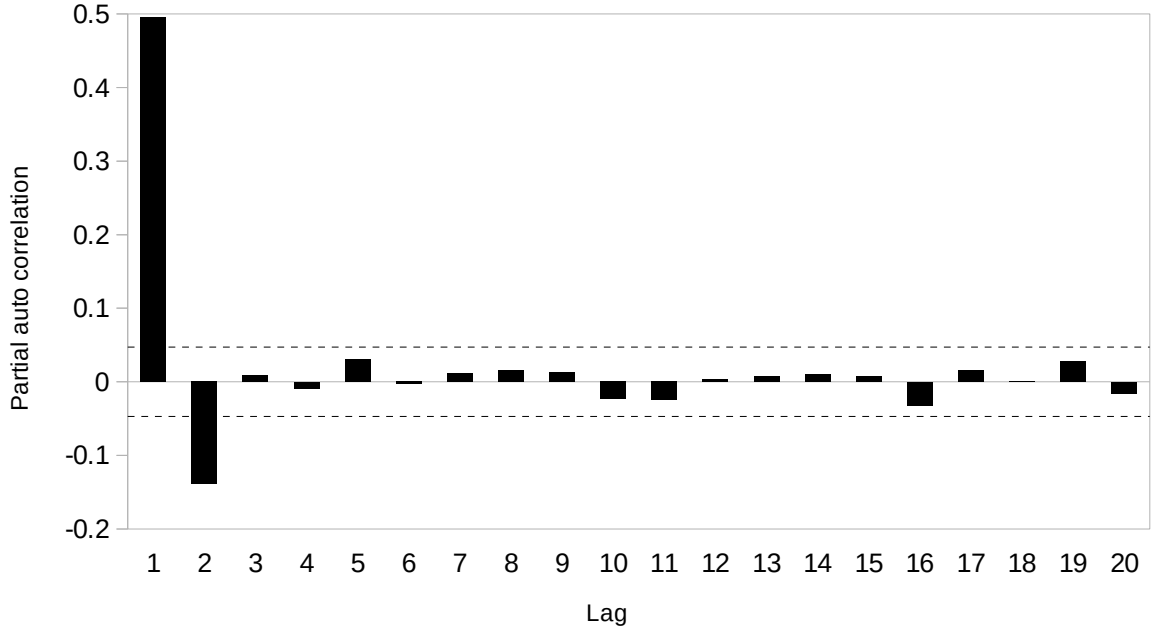


**Fig. 7.** Example partial auto-correlation of the MERRA daily average wind speed square-root residual at a single node. Dashed lines indicate 95% significance levels for a population value of zero.

**Table 1**
Example Root Mean Square Error (RMSE) for different orders of Auto Regressive Moving Average (ARMA) models of the MERRA daily average wind speed square root residual at a single MERRA node.

| | MA order |
|---|---|

| AR order | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 0.99993 | 0.87811 | 0.87811 | 0.87811 | 0.87811 |
| 1 | 0.87673 | 0.86891 | 0.86887 | 0.86887 | 0.86887 |
| 2 | 0.86863 | 0.86882 | 0.86870 | 0.86867 | 0.86864 |
| 3 | 0.86884 | 0.86887 | 0.86885 | 0.86884 | 0.86884 |
| 4 | 0.86829 | 0.86831 | 0.86830 | 0.86829 | 0.86829 |
| 5 | 0.86818 | 0.86819 | 0.86819 | 0.86819 | 0.86819 |

The Auto-Regressive (AR) order increases with each row downward, and the Moving Average (MA) order increases to the right.

**Table 2**
Example Bayesian Information Criterion (BIC) and ranking in ascending order for different orders of Auto Regressive Moving Average (ARMA) models of the MERRA daily average wind speed square root residual at a single MERRA node.

| | MA order | | | | |
|---|---|---|---|---|---|
| AR order | 0 | 1 | 2 | 3 | 4 |
| 0 | 7.25<br>30 | -459.93<br>20 | 452.42<br>24 | -444.91<br>27 | -437.40<br>29 |
| 1 | -465.69<br>15 | -490.90<br>2 | -483.59<br>5 | -476.07<br>8 | -468.57<br>14 |
| 2 | -492.09<br>1 | -483.78<br>3 | -476.79<br>7 | -469.39<br>12 | -462.01<br>18 |
| 3 | -483.69<br>4 | -476.06<br>9 | -468.61<br>13 | -461.17<br>19 | -453.67<br>23 |
| 4 | -478.51<br>6 | -470.89<br>11 | -463.46<br>17 | -455.96<br>22 | -448.46<br>26 |
| 5 | -471.46<br>10 | -463.90<br>16 | -456.38<br>21 | -448.86<br>25 | -441.35<br>28 |

The Auto-Regressive (AR) order increases with each row downward, and the Moving Average (MA) order increases to the right. The ARMA(2,0), or AR(2), model had the lowest BIC (rank 1).

To remove the seasonal and daily components, the 24 hour trend was found for each hourly MERRA wind speed value by calculating the average wind speed from 12 hours before to 11 hours after that hour. This trend was then subtracted to obtain a de-trended hourly MERRA wind speed dataset. Similarly to Skidmore and Tatarko [46], the simulation used a single sinusoid to represent the diurnal wind speed:

$$v_{diurnal} = t_{mag} \cos\left(2\pi \frac{(t - t_{peak})}{t_{period}}\right) \quad t_{peak} - 0.75t_{period} < t < t_{peak} + 0.25t_{period} \tag{8}$$

Where t is the time of day (hours), $t_{peak}$ is the time of day (hours) when the peak wind speed occurs within the de-trended dataset, $t_{period}$ is timespan between the beginning and the end of the sinusoid (hours), and $t_{mag}$ is the magnitude of the sinusoid ($ms^{-1}$). For each day, the difference between maximum and minimum wind speed values (in the de-trended MERRA dataset), and the hour when these occurred was used to formulate the magnitude, period, and peak hour of the sinusoid for the simulation. The average peak hour was found to occur later as distance from the coast increased (Fig. 8). This indicated that the peak sea-breeze front travels inland initially at about 33 $kmhr^{-1}$ (9.17 $ms^{-1}$), which is consistent with the average offshore land-breeze propagation speed of 32.4 ± 14.4 $kmhr^{-1}$ (9 ± 4 $ms^{-1}$) reported by Gille et al. [47].

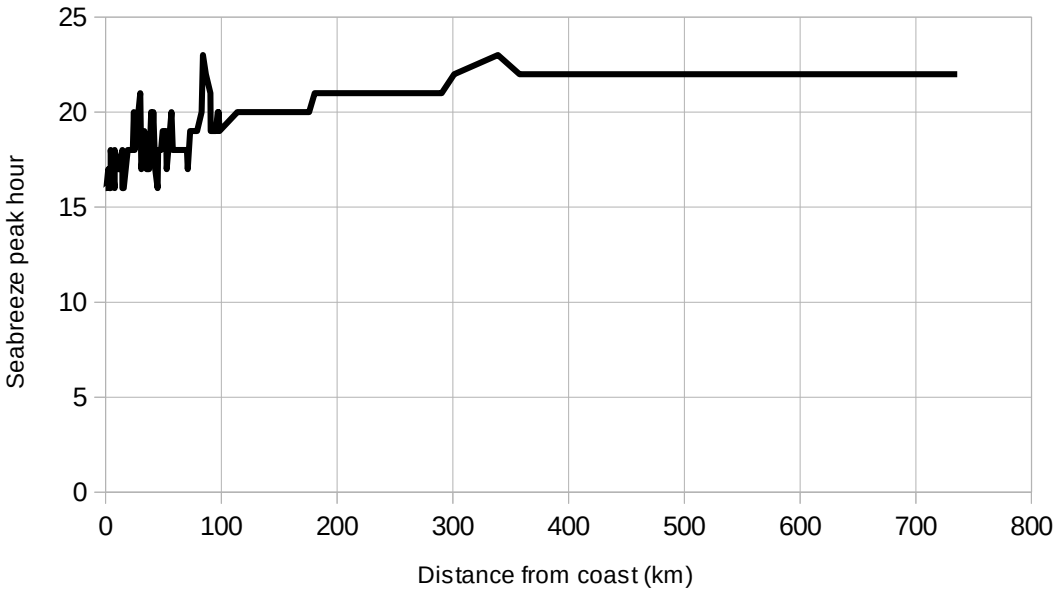


**Fig. 8.** Variation in summer sea-breeze peak hour with distance from the coast-line.

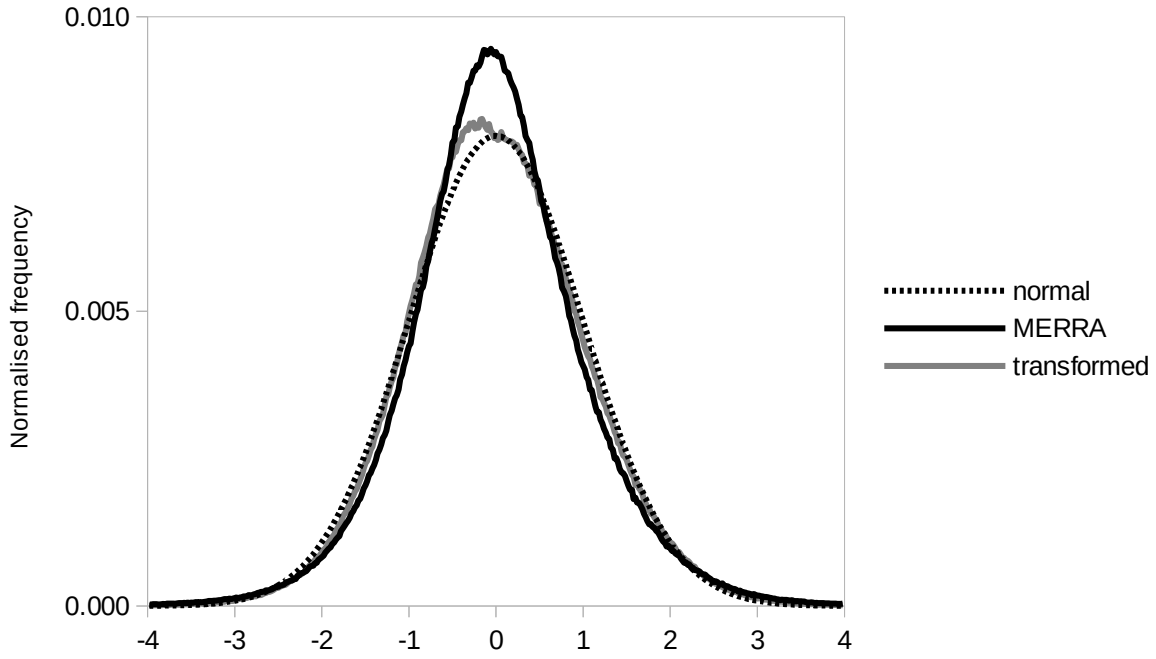


**Fig. 9.** Normalised frequency distribution of MERRA 50m wind speed residual and transformed residual compared to normal distribution. Horizontal axis bin width is 0.02, which is close to the value of 0.016 suggested by the Freedman-Diaconis rule [43].

The sea-breeze magnitude, period, and peak hour variabilities were also found to have a seasonal dependence as well as dependence on distance from the coast-line. The diurnal component was subtracted from the de-trended hourly wind speed to obtain the hourly MERRA wind speed residual y. y was then normalised by subtracting the overall mean and dividing by the overall standard deviation, σ. σ was found to have a spatial dependence. Hill et al. [8] found the de-trended wind speed distribution for sites in the UK to follow a normal distribution. However in this study the normalised residual y was found to roughly follow a double exponential distribution with a slight skew, rather than a normal distribution (Fig. 9). Ward and Boland [16] also found a double exponential distribution for wind data in South Australia.

As pointed out by Lawrance and Lewis [25] and Damsleth and El-Shaarawi [26], there is a tendency for autoregressive equations to produce time series with normal distributions, even if the distribution of the random term is not normal. Hence a data transformation function was required to convert the distribution of y to a normal-like distribution. Because a double exponential distribution is symmetric about the mean, a simple symmetric form of square-root conversion was used on the data:

$$y_n(t)=\frac{\left(1.4-\sqrt{1.96-y(t)}\right)}{0.302} \quad y(t)<0$$

$$y_n(t)=\frac{\left(\sqrt{y(t)+1.96}-1.4\right)}{0.302} \quad y(t)\geq 0 \qquad (9)$$

Where $y_n(t)$ is the transformed hourly MERRA wind speed residual. The transformed distribution is more normal-like (Fig. 9). The auto-correlation coefficients and partial auto-correlation coefficients of $y_n(t)$ for each node indicated that there was an auto-regressive (AR) dependency of order three within the $y_n(t)$ time series (For example Fig. 10 and Fig. 11). However there remained possibly significant low levels of dependence at lags greater than three. In a similar procedure to the daily average wind speed square root residuals, The least squares method and numerical maximum likelihood estimation was used to calculate the root mean square error (RMSE) for ARMA models with coefficients up to order (5,4) for each residual (for example Table 3). These indicated that the pure AR models gave sightly lower RMSE values. For increasing AR order, the RMSE initially decreased and then substantially levelled off after order three. To confirm which AR order was necessary and sufficient to capture most of the dependency within the time series, the Bayesian Information Criterion (BIC) was calculated and ranked in ascending order for each ARMA(p,q) model, $0 \leq p \leq 5$ and $0 \leq q \leq 4$, at each node (for example Table 4). For hourly data, n = 43824. The results were not definitive as models with the lowest BIC were a mixture of ARMA(3,0) and ARMA(4,0). ARMA(3,0), or AR(3), models were chosen as the difference in RMSE values between the two models was small (typically < 1%). The AR(3) coefficient values were found to have a spatial dependency. See Appendix A for the full parameterisation of the spatial dependencies.

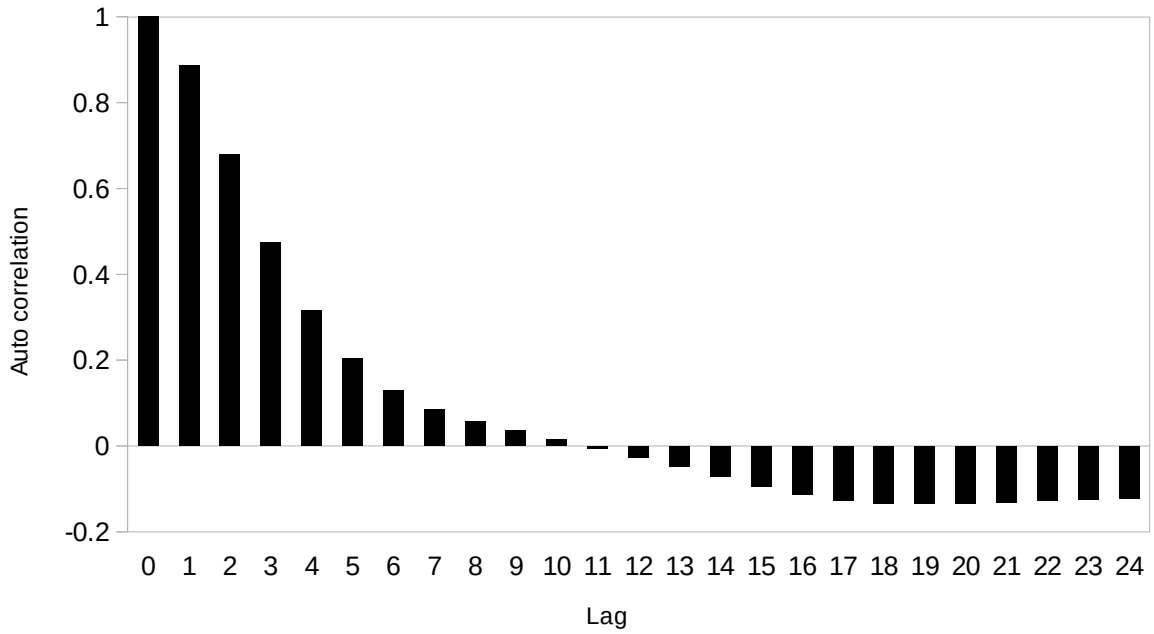


**Fig. 10.** Example auto-correlation of the MERRA hourly wind speed transformed residual at a single node. 95% significance level for a population value of zero is ~0.01.

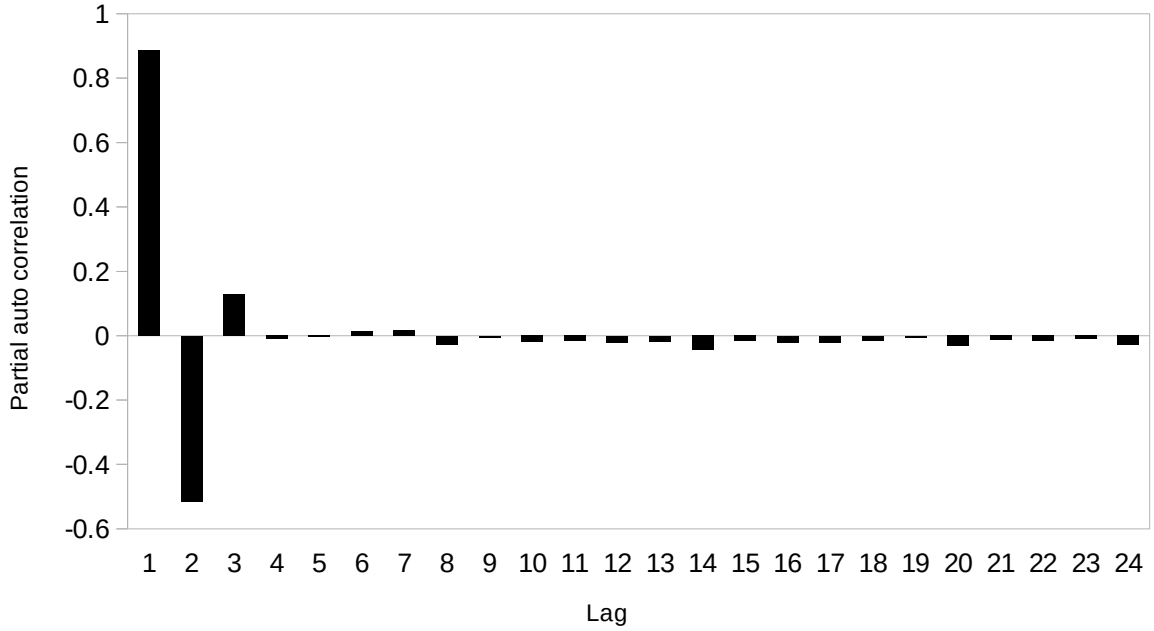


**Fig. 11.** Example partial auto-correlation of the MERRA hourly wind speed transformed residual at a single node. 95% significance level for a population value of zero is ~0.01.

**Table 3**
Example Root Mean Square Error (RMSE) between different Auto Regressive Moving Average (ARMA) models and the MERRA hourly wind speed transformed residual at a single node.

| | MA order | | | | |
|---|---|---|---|---|---|
| AR order | 0 | 1 | 2 | 3 | 4 |
| 0 | 0.99501 | 0.59737 | 0.59737 | 0.59737 | 0.59737 |
| 1 | 0.45858 | 0.39830 | 0.39830 | 0.39830 | 0.39830 |
| 2 | 0.39308 | 0.39035 | 0.39018 | 0.39010 | 0.39010 |
| 3 | 0.38983 | 0.38982 | 0.38982 | 0.38982 | 0.38982 |
| 4 | 0.38981 | 0.38982 | 0.38987 | 0.38984 | 0.38983 |
| 5 | 0.38981 | 0.38983 | 0.38985 | 0.38983 | 0.38986 |

The Auto-Regressive (AR) order increases with each row downward, and the Moving Average (MA) order increases to the right.

**Table 4**
Example Bayesian Information Criterion (BIC) and ranking in ascending order for different orders of Auto Regressive Moving Average (ARMA) models of the MERRA hourly wind speed transformed residual at a single node.

| | MA order | | | | |
|---|---|---|---|---|---|
| AR order | 0 | 1 | 2 | 3 | 4 |
| 0 | -428.19 | -45136.44 | -45125.75 | -45115.06 | -45104.38 |
| | 30 | 26 | 27 | 28 | 29 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | -68311.50<br>25 | -80651.62<br>21 | -80640.93<br>22 | -80630.25<br>23 | -80619.34<br>24 |
| 2 | -81807.56<br>20 | -82409.56<br>19 | -82436.04<br>17 | -82442.40<br>16 | -82435.33<br>18 |
| 3 | -82524.81<br>1 | -82517.95<br>3 | -82507.11<br>5 | -82496.83<br>7 | -82485.26<br>10 |
| 4 | -82518.49<br>2 | -82505.93<br>6 | -82485.36<br>9 | -82481.62<br>11 | -82472.63<br>13 |
| 5 | -82507.70<br>4 | -82494.41<br>8 | -82477.41<br>12 | -82471.13<br>14 | -82453.59<br>15 |

The Auto-Regressive (AR) order increases with each row downward, and the Moving Average (MA) order increases to the right. In this case, the ARMA(3,0), or AR(3), model had the lowest BIC (rank 1).

It was now possible to start generating synthetic hourly wind speed values. Firstly, the synthetic normally distributed hourly residual $y_{ns}(t)$ was generated using a standard AR(3) equation:

$$y_{ns}(t)=\varphi_1 y_{ns}(t-1)+\varphi_2 y_{ns}(t-2)+\varphi_3 y_{ns}(t-3)+\rho r(t) \tag{10}$$

Where r(t) is a normally distributed random variable and ϱ is set so that the standard deviation of $y_{ns}(t)$ is one. $\phi_1$, $\phi_2$ and $\phi_3$ are the AR(3) coefficients (see Appendix A). The initial values of $y_{ns}(t-1)$, $y_{ns}(t-2)$ and $y_{ns}(t-3)$ were set to standard normally distributed random values. The computational benefit of using the square-root data transformation in equation (9) is that generation of synthetic wind speed residuals requires a simple reverse transformation involving a calculation of the square:

$$\begin{aligned} y_s(t)&=1.96-(1.4-0.302\,y_{ns}(t))^2 \quad y_{ns}(t)<0 \\ y_s(t)&=(1.4+0.302\,y_{ns}(t))^2-1.96 \quad y_{ns}(t)\geq 0 \end{aligned} \tag{11}$$

Where $y_s(t)$ is the synthetic hourly wind speed residual for wind farm w. The distribution of $y_s(t)$ was similar to the MERRA wind speed residual distribution, but without the slight skew (Fig. 12).

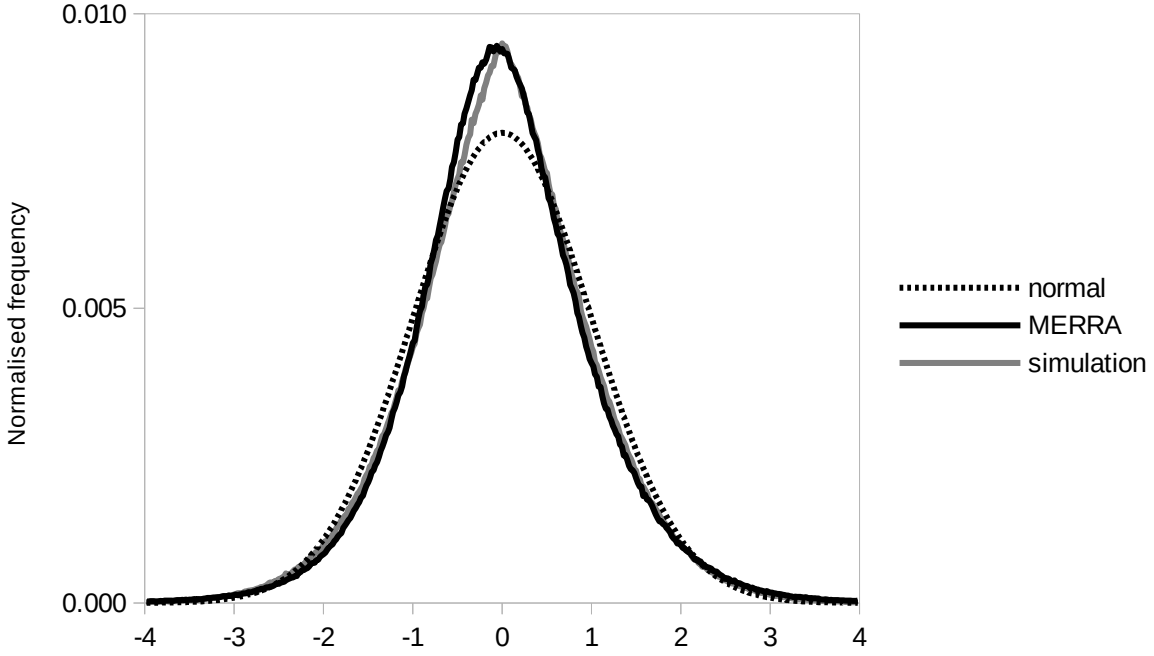


**Fig. 12.** Normalised frequency distribution of MERRA and simulation hourly 50m wind-speed residual. Horizontal axis bin width is 0.02, which is close to the value of 0.016 suggested by the Freedman-Diaconis rule [43].

The synthetic average daily wind speed residual $y_{ds}(t)$ was generated using a standard AR(2) equation:

$$y_{ds}(t)=\varphi_{d1}\,y_{ds}(t-1)+\varphi_{d2}\,y_{ds}(t-2)+\rho_d\,r_d(t) \tag{12}$$

Where $\varrho_d$ was set to a value such that the standard deviation of $y_{ds}(t)$ is 1. $r_d(t)$ is a standard normally distributed random value. The initial values of $y_{ds}(t-1)$, and $y_{ds}(t-2)$ were set to standard normally distributed random values. Synthetic daily average wind speeds $v_{ds}(t)$ were generated by squaring $y_{ds}$ and using $V_{season}$ as the average:

$$v_{ds}(t)=\left(\sqrt{V_{season}}+\sigma_d\,y_{ds}(t)\right)^2-\sigma_d^2 \tag{13}$$

The $\sigma_d^2$ term is present to make the mean of $v_{ds}(t)$ be $V_{season}$. The hourly synthetic wind speed $v_s$ could now be assembled as the sum of the daily average component, the diurnal component, and the hourly dependent component:

$$v_s(t)=v_{ds}(t)+v_{diurnal}(t)+\sigma y_s(t) \tag{14}$$

See Appendix A for the full wind speed simulation algorithm. $v_s$ is the wind speed 50 metres above the ground, but as the hub height of most modern wind turbines is higher than 50 metres, $v_s$ must be scaled to the hub-height wind speed $v_{hh}(t)$. The simpler power law estimation for wind shear (equation (4)) was used because no extra information about surface friction is required. MERRA data was only available for one height, so measured data at different heights from two sites, one coastal and one inland, was used. This data indicated that the wind shear is more pronounced in inland areas, and varies with hour of the day, with wind shear exponent $\alpha$ being usually larger at night (Fig. 13). There was also a seasonal variation superimposed on this, with $\alpha$ being even greater at night during the winter months.

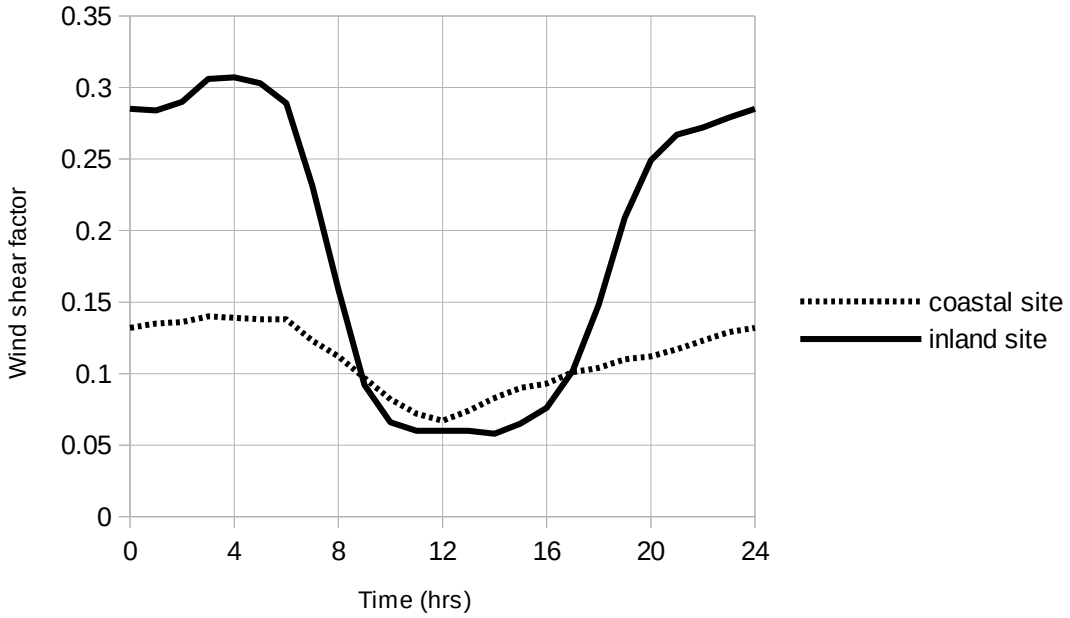


**Fig. 13**. Change in average wind shear factor with time of day for a coastal site and an inland site.

Hourly wind farm power output was estimated from the hub height wind speed $v_{hh}$ using wind turbine power curves modified according to the findings of Holttinen [34] (Fig. 1). The parameters for wind turbines used in the SWWA are given in Table A1 of Appendix A. Since the SWWA has a generally low elevation, it was assumed that there was no significant difference in air density between the sites used to measure the turbine technical specifications, and the actual air density encountered by the turbines used in Western Australia.

The model presented in this study, called here the 'transformed residual' model, was run for a period of 5 years, from 2009 to 2013. The simulation was started by calculating the spatial and seasonal parameters for a chosen day of the year, then generating power output values hour-by-hour. Daily and seasonal parameters were recalculated at the beginning of each day. The simulated wind farm power output was compared to actual wind farm power output data for the SWIS grid, measured using the Supervisory Control And Data Acquisition (SCADA) system. Six of the largest wind farms connected to the SWIS grid were chosen for comparison. All of these wind farms have capacities greater than 10MW.

The parameters used to generate the model synthetic seasonal wind speed $V_{season}$ have a dependence on latitude and distance from the coast (see equation (A7)), so there is already an implicit correlation in the seasonal wind speed between two nearby sites. Correlation between the daily and diurnal components of

wind speed for different wind sites was introduced using a matrix of distance weighted combinations of the random numbers used to generate these values. The weighting factor between two sites was given an inverse relationship to distance apart (equations (A11) to (A15)), so that distant sites would be less correlated than nearer sites. The hourly autoregressive and random components of wind speed at each site were assumed to be uncorrelated.

The MERRA 50m wind speeds were also scaled to hub-height and converted to wind power values. It was found that the distance correlation between these MERRA wind power values was greater than the SCADA data distance correlation (Fig. 14), suggesting that there is an extra source of spatial variability other than the wind speed. Therefore the model was instead calibrated to the distance correlation values reported in Carlin and Haslett [20] for wind measurements at several sites in Western Australia, which correspond more closely to the SCADA correlation.

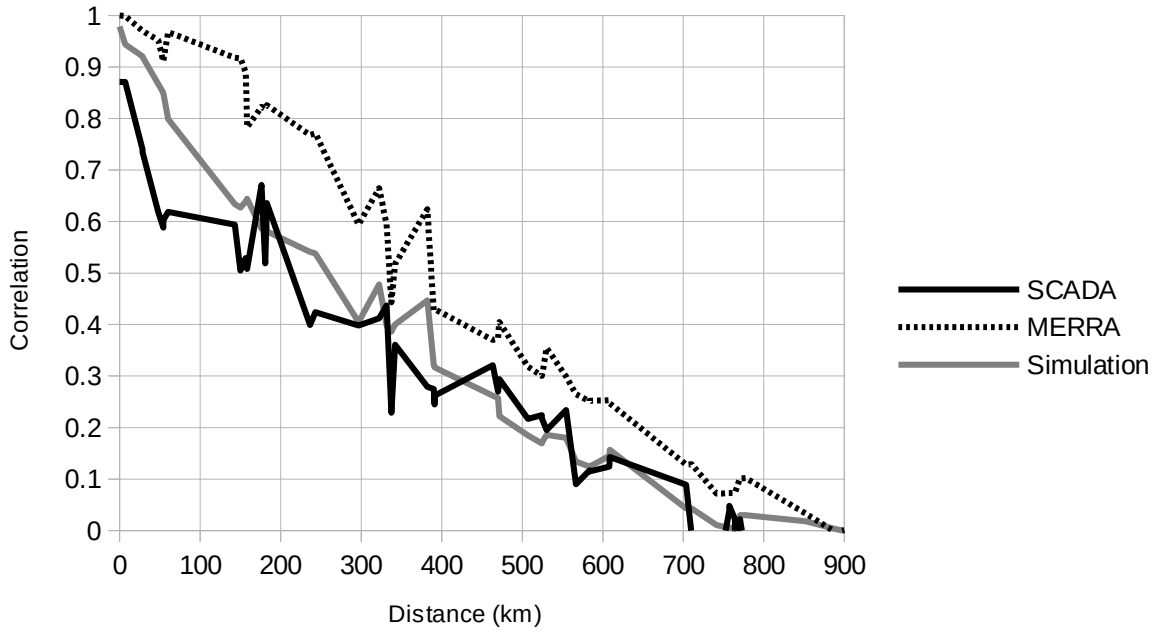


**Fig. 14**. Average correlation with inter wind farm distance for wind farm power output for the SWWA.

To assess the simulation, the results from two other models were also compared. The first model, called here the 'Weibull' model, used the Weibull distribution to generate hourly time series wind speed data with no dependency on previous values of wind speed. The seasonal wind speed $V_{season}$ was used to calculate the scale parameter $\lambda$, and the shape parameter $k$ was estimated using the maximum likelihood method from the hourly wind speed data. The second model was the same as the transformed residual model, except that normally distributed residuals were used, with no data transformation. This model was called the 'normal residual' model. The average of 10 simulation runs of all three models were compared to the measured SCADA wind power output data. Two statistical measures used to compare the models with the measured data were the Root Mean Square Error (RMSE), and the Mean Bias Error (MBE). The RMSE is a measure of the magnitude of the difference between individual data points in each data set. The sign of the difference is ignored. The MBE is a measure of the average difference between individual data points in each data set, or whether the model generated data set is biased higher or lower compared to the measured data on the whole. Here the sign of the difference is not ignored. Generally, a model with a lower RMSE than another fits

the data more closely. In this study a model with a negative MBE might be considered more favourably than a model with a similar but positive MBE because under predicting wind power generation on the whole is more desirable than over predicting. Representing RMSE and MBE as a percentage gives an idea of how significant the error is compared to the average value of the measured data. See Appendix B for definitions of these measures.

## 3. Results

The three simulation generated overall average Capacity Factor (CF) values for the six largest wind farms connected to the SWIS (Albany and Grasmere are considered separate wind farms) were generally comparable to the measured SCADA values (Table 5), with differences less than 12%, except for Walkaway wind farm, where the three models underestimated the yearly average CF by 9-16%. The normal residual model slightly overestimated the yearly average CF. The Weibull and transformed residual models underestimated the overall average CF. These two models generated similar CF values for 4 out of the 6 wind farms. For Grasmere and Albany wind farms, the transformed residual model generated slightly lower values than the Weibull model. However the magnitude of the differences between these two models and the SCADA overall CF were similar, indicating that the models were similarly close.

**Table 5**
SCADA and simulated overall average capacity factor (CF) for six wind farms within the SWWA using three simulation models: the Weibull model, the normal residual model, and the transformed residual model.

| Name | Capacity (MW) | Distance from Coast (km) | SCADA CF | Weibull model CF | Normal residual model CF | Transformed residual CF |
|---|---|---|---|---|---|---|
| Grasmere | 13.8 | 0.67 | 0.33 | 0.33 | 0.36 | 0.32 |
| Albany | 21.6 | 0.67 | 0.32 | 0.33 | 0.36 | 0.32 |
| Mumbida | 55 | 14.6 | 0.39 | 0.38 | 0.41 | 0.38 |
| Emu Downs | 79.2 | 23.6 | 0.35 | 0.33 | 0.36 | 0.33 |
| Walkaway | 89.1 | 15.8 | 0.43 | 0.36 | 0.39 | 0.36 |
| Collgar | 206 | 255 | 0.37 | 0.34 | 0.36 | 0.34 |

Model values are the average of 10 simulation runs.

The errors between the simulation and SCADA yearly average capacity factors (Table 6) indicated that the Weibull model was slightly closer to the measured SCADA data. The greatest RMSE error for the Weibull and transformed residual model occurred at the Walkaway wind farm, and for the normal residual model, the greatest error was at the Albany windfarm. These results indicated that actual power generation at the Walkaway wind farm is significantly greater than predicted by all the models, which are based on MERRA data. Local effects may be increasing wind speeds at this site. Comparing the SCADA and simulation monthly average capacity factors, the Weibull model achieved a lower RMSE than the other two models. However, the normalised frequency distribution of average daily capacity factors (Fig. 15) indicated that the Weibull model generated a significantly different distribution to the measured SCADA data, with CF values concentrated on intermediate values between 0.2 and 0.5. Although less pronounced, the normal residual model also generated a distribution more concentrated on intermediate values of CF. The transformed residual model generated a distribution closest to the SCADA distribution, with a slight skew towards lower

CF values, reflecting the slight conservative bias of this model. These discrepancies were reflected in the error values (Table 6). The normal and transformed residual models achieved much lower RMSE values than the Weibull model. The transformed residual model achieved the lowest RMSE values overall, although the value for Mumbida wind farm was comparable to the normal residual model, and the value for Walkaway was significantly higher. The Weibull and normal residual models also generated a significantly different diurnal peak hour distribution (Fig. 16). Peak hour is the hour of the day when CF (and hence wind power output) is at a maximum. The transformed residual model achieved lower RMSE values than the Weibull or normal residual models, except for Emu Downs wind farm, where the values were comparable. The hourly CF normalised frequency distribution (Fig. 17) of the Weibull model fitted the SCADA distribution slightly better than the other two models. The transformed residual model again exhibited a slight skew towards lower CF values.

**Table 6**
Errors in the yearly and monthly average capacity factor (CF) estimation, daily average CF frequency distribution, diurnal peak hour distribution and hourly average CF frequency distribution of three simulation models compared to measured SCADA wind power data.

| Time scale | Measure | Weibull model | | Normal residual model | | Transformed residual model | |
|---|---|---|---|---|---|---|---|
| Yearly | RMSE(%) | Grasmere | 1.74 | Grasmere | 10.4 | Grasmere | 5.3 |
| | | Albany | 9.77 | Albany | 19.2 | Albany | 9.6 |
| | | Mumbida | 4.59 | Mumbida | 3.9 | Mumbida | 5.4 |
| | | Emu Downs | 7.71 | Emu Downs | 8.4 | Emu Downs | 9.7 |
| | | Walkaway | 15.11 | Walkaway | 9.6 | Walkaway | 16.1 |
| | | Collgar | 9.0 | Collgar | 5.6 | Collgar | 10.4 |
| | | **Average** | **8.0** | **Average** | **9.5** | **Average** | **9.4** |
| | MBE(%) | Grasmere | 1.2 | Grasmere | 9.8 | Grasmere | -1.1 |
| | | Albany | 8.6 | Albany | 18.2 | Albany | 6.1 |
| | | Mumbida | -4.6 | Mumbida | 3.9 | Mumbida | -5.4 |
| | | Emu Downs | -4.1 | Emu Downs | 4.0 | Emu Downs | -4.2 |
| | | Walkaway | -14.5 | Walkaway | -7.6 | Walkaway | -15.2 |
| | | Collgar | -8.1 | Collgar | -1.3 | Collgar | -8.2 |
| | | **Average** | **-3.6** | **Average** | **4.5** | **Average** | **-4.6** |
| Monthly | RMSE(%) | Grasmere | 15.4 | Grasmere | 25.5 | Grasmere | 30.0 |
| | | Albany | 20.4 | Albany | 30.4 | Albany | 30.3 |
| | | Mumbida | 12.5 | Mumbida | 31.1 | Mumbida | 26.8 |
| | | Emu Downs | 16.9 | Emu Downs | 22.1 | Emu Downs | 23.4 |
| | | Walkaway | 18.9 | Walkaway | 21.2 | Walkaway | 24.0 |
| | | Collgar | 17.3 | Collgar | 22.1 | Collgar | 24.1 |
| | | **Average** | **16.9** | **Average** | **25.4** | **Average** | **26.4** |
| | MBE(%) | Grasmere | 1.3 | Grasmere | 9.5 | Grasmere | -2.5 |
| | | Albany | 8.1 | Albany | 16.6 | Albany | 5.0 |
| | | Mumbida | 1.6 | Mumbida | 12.2 | Mumbida | 5.7 |
| | | Emu Downs | -4.1 | Emu Downs | 4.0 | Emu Downs | -4.0 |
| | | Walkaway | -14.4 | Walkaway | -8.0 | Walkaway | -14.8 |
| | | Collgar | -5.2 | Collgar | 1.5 | Collgar | -4.1 |
| | | **Average** | **-2.1** | **Average** | **6.0** | **Average** | **-2.4** |
| Daily frequency distribution | RMSE(%) | Grasmere | 143 | Grasmere | 32.5 | Grasmere | 25.2 |
| | | Albany | 147 | Albany | 40.7 | Albany | 17.0 |
| | | Mumbida | 122 | Mumbida | 35.1 | Mumbida | 35.9 |
| | | Emu Downs | 126 | Emu Downs | 25.5 | Emu Downs | 23.5 |
| | | Walkaway | 129 | Walkaway | 28.8 | Walkaway | 37.8 |
| | | Collgar | 136 | Collgar | 25.1 | Collgar | 22.3 |
| | | **Average** | **134** | **Average** | **31.3** | **Average** | **26.9** |
| Diurnal peak hour distribution | RMSE(%) | Grasmere | 47.1 | Grasmere | 38.0 | Grasmere | 25.9 |
| | | Albany | 47.0 | Albany | 39.9 | Albany | 18.5 |
| | | Mumbida | 58.6 | Mumbida | 55.3 | Mumbida | 43.8 |
| | | Emu Downs | 49.5 | Emu Downs | 48.5 | Emu Downs | 50.6 |
| | | Walkaway | 58.0 | Walkaway | 52.0 | Walkaway | 42.0 |
| | | Collgar | 107 | Collgar | 77.1 | Collgar | 20.5 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Average** | **61.2** | **Average** | **51.8** | **Average** | **33.5** |
| Hourly frequency distribution | RMSE(%) | Grasmere | 27.2 | Grasmere | 79.0 | Grasmere | 18.4 |
| | | Albany | 15.6 | Albany | 65.1 | Albany | 18.9 |
| | | Mumbida | 19.7 | Mumbida | 43.2 | Mumbida | 30.8 |
| | | Emu Downs | 48.9 | Emu Downs | 26.6 | Emu Downs | 56.3 |
| | | Walkaway | 27.8 | Walkaway | 48.1 | Walkaway | 36.6 |
| | | Collgar | 35.2 | Collgar | 78.6 | Collgar | 27.2 |
| | | **Average** | **29.0** | **Average** | **56.8** | **Average** | **31.3** |

RMSE and MBE values are given as a percentage of the average SCADA capacity factor or average SCADA CF frequency. Values for individual wind farms are the average of 10 simulation runs.

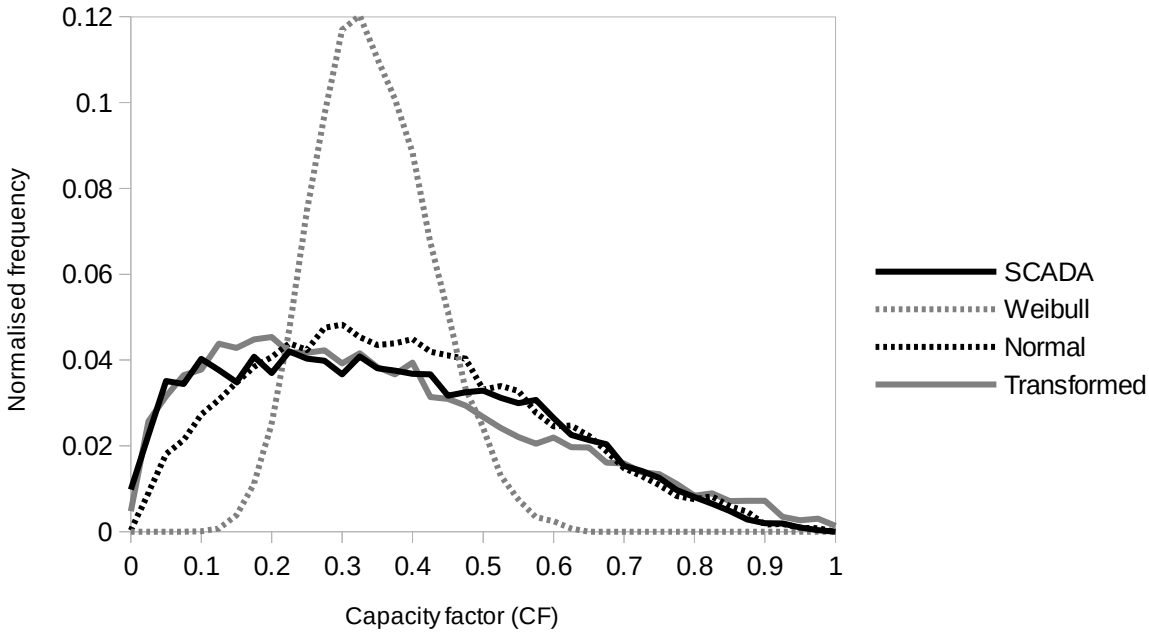


**Fig. 15.** Normalised frequency distribution of average daily capacity factors. Three models (Weibull, normal residual, and transformed residual) are compared to measured SCADA wind power output data. Horizontal axis bin width is 0.025, which is close to the value of 0.032, suggested by the Freedman-Diaconis rule [43].

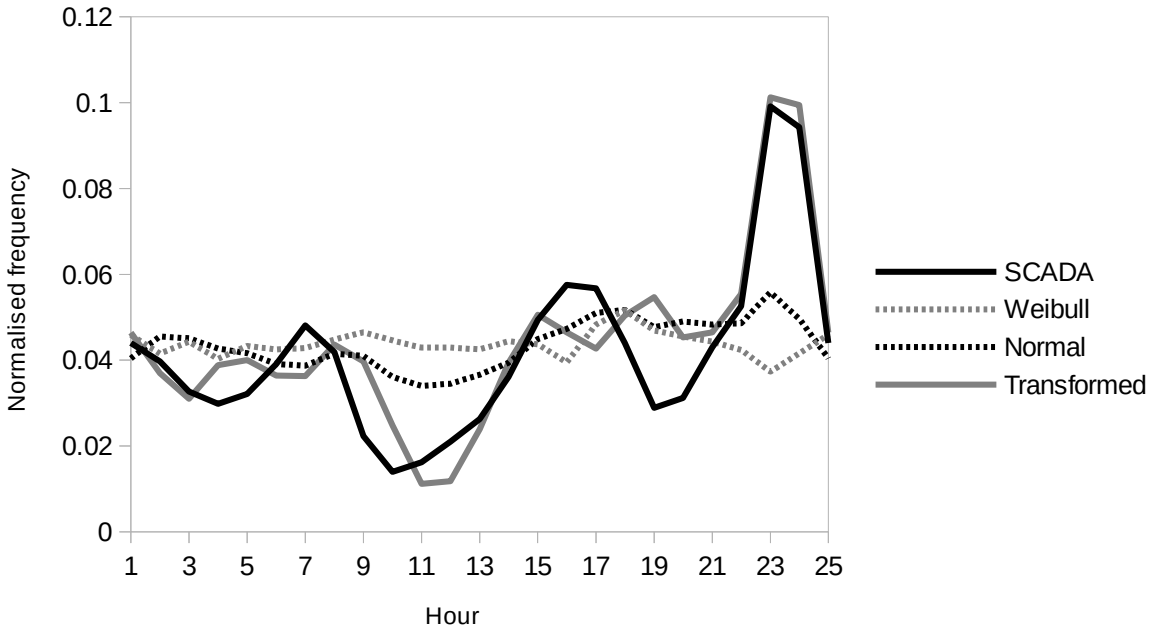


**Fig. 16.** Diurnal capacity factor (CF) peak hour distribution. Peak hour is the hour of the day when CF (and hence wind power output) is at a maximum. Three models (Weibull, normal residual, and transformed residual) are compared to measured SCADA wind power output data. The 25$^{th}$ hour is the same data point as the 1$^{st}$ hour and is provided for continuity.

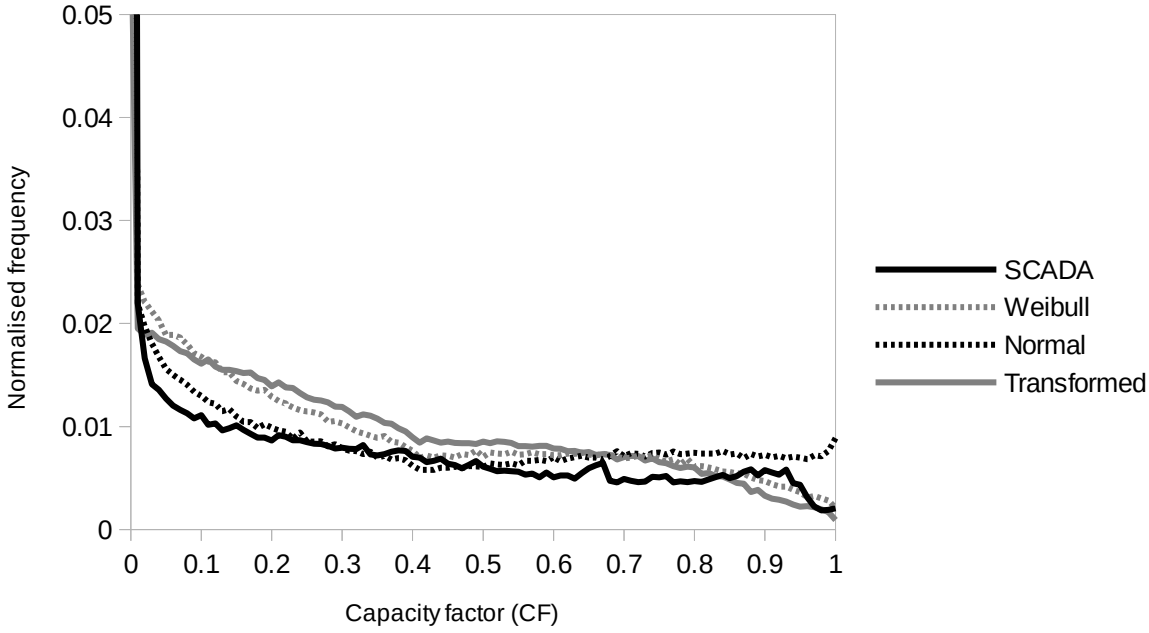


**Fig. 17.** Normalised frequency distribution of average hourly capacity factors. Three models (Weibull, normal residual, and transformed residual) are compared to measured SCADA wind power output data. Horizontal axis bin width is 0.01, which is close to the value of 0.016, suggested by the Freedman-Diaconis rule [43].

## 4. Discussion and Conclusions

Even though the Weibull model generated yearly capacity factors slightly closer to the measured SCADA data than the transformed residual model, both the Weibull and normal residual models generated significantly different daily average capacity factor and diurnal peak hour distributions to the measured data, and hence would generate unrealistic statistical behaviour if they were used to simulate existing or hypothetical wind power systems in the SWWA. The transformed residual model generated daily average capacity factors and diurnal peak hours with a much closer distribution to the measured data and demonstrated the necessity of characterising the wind speed residual properly, and not blindly assuming that it has a normal distribution.

Wind power is a distributed resource that is increasing in use world-wide. Therefore simulating the operation of large scale electrical grids with significant levels of wind power is becoming more important. To do this it is necessary to build regional scale wind power simulations that can account for spatial, seasonal and hour-by-hour variation. The results from this study indicated that using MERRA data as the basis of such a wide area simulation is a viable method. The MERRA dataset is freely available and covers the whole surface of the world, including many regions that would have no access to suitable hub-height wind speed data.

The wind farm capacity factors were found to have a greater distance correlation when estimated from MERRA data, than when calculated directly from the measured SCADA data. Hence there may be more variability in air density, wind shear factor, wind farm wide wind speed variability, or another factor than accounted for here. It is important to confirm this finding in future studies over other regions, as wind variability can have a significant effect on the operation of a large scale electrical grid. Simulations based on MERRA data can be built for any site or region in the world, but they must incorporate a means for correctly setting the distance correlation between wind farm sites.

The simulation in this instance was conservative. Although measured data was used to calibrate the wind shear factor, similar simulations could be developed for regions with no available measured data. Setting the wind shear exponent to zero would generate even more conservative synthetic power output data, but still usable because the 50m height above ground level is within the range of most modern wind turbine hub heights.

The frequency distribution of the AR-based simulation capacity factor was similar to the measured data capacity factor, due to the use of the novel data transformation functions, which also have applicability to other sites that have an exponential wind speed residual distribution. Matching the measured and simulated wind power frequency distribution is important for detailed estimation of wind power potential.

Similarly to Ward et al. [16], it was found that wind sites closer to the coast tended to have diurnal output peaks in the afternoon, earlier than those wind sites further inland which peak later in the evening. This implies that a mix of coastal and inland wind farm sites is beneficial for avoiding large peaks and lulls in wind power generation and maintaining a supply of wind power that is consistent with the peak in electricity demand.

## Acknowledgements

Thanks to the National Aeronautics and Space Administration (NASA) for providing open access to the MERRA database, and thanks to Murdoch University for supporting the candidature that allowed this simulation to be developed.

## References

[1] Tan HJ. Directional Waves in the Nearshore Coastal Region of Perth, Western Australia. University of Western Australia, 2004.

[2] Pattiaratchi C, Hegge B, Gould J, and Eliot I. Impact of sea-breeze activity on nearshore and foreshore processes in south western Australia. Continental Shelf Research 1997;17:1539–60.

[3] Clarke RH. Sea-breezes and waves: the 'Kalgoorlie sea-breeze' and the 'Goondiwindi breeze'. Australian Meteorological Magazine 1989;37:99-107.

[4] Burton T, Sharpe D, Jenkins N, Bossanyi E. Wind Energy Handbook. 2nd ed. 2001. Chichester, England: John Wiley & Sons, Ltd.

[5] Suomalainen K, Silva CA, Ferrão P, Connors S. Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning. Energy 2012;37:41–50. doi:10.1016/j.energy.2011.08.001.

[6] Billinton R, Chen H, Ghajar R. Time-series models for reliability evaluation of power systems including wind energy. Microelectronics Reliability 1996;36(9):1253-1261.

[7] Karki R, Hu P, Billinton R. A Simplified Wind Power Generation Model for Reliability Evaluation. IEEE Transactions on Energy Conversion 2006;21:533–40. doi:10.1109/TEC.2006.874233.

[8] Hill DC, McMillan D, Bell KRW, Infield D. Application of Auto-Regressive Models to U.K. Wind Speed Data for Power System Impact Studies. IEEE Transactions on Sustainable Energy 2012;3:134–41. doi:10.1109/TSTE.2011.2163324.

[9] Carta JA, Ramírez P, Velázquez S. A review of wind speed probability distributions used in wind energy analysis. Renewable and Sustainable Energy Reviews 2009;13:933–55. doi:10.1016/j.rser.2008.05.005.

[10] Gunturu UB, Schlosser CA. Characterization of wind power resource in the United States. Atmospheric Chemistry and Physics 2012;12:9687–702. doi:10.5194/acp-12-9687-2012.

[11] Box, GEP, Jenkins GM. Time Series Analysis, Forecasting and Control, revised edition. 1976. San Francisco: Holden-Day.

[12] T. Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence. 24 (2011) 164–181. doi:10.1016/j.engappai.2010.09.007.

[13] S.S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium (NAPS), 2010, IEEE, 2010: pp. 1–8. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5619586 (accessed November 26, 2015).

[14] Valipour M. Parameters Estimate of Autoregressive Moving Average and Autoregressive Integrated Moving Average Models and Compare Their Ability for Inflow Forecasting. Journal of Mathematics and Statistics 2012; 8 (3):330-338.

[15] Papaefthymiou G, Klockl B. MCMC for Wind Power Simulation. IEEE Transactions on Energy Conversion 2008;23:234–40. doi:10.1109/TEC.2007.914174.

[16] Ward K, Boland J. Modelling the volatility in wind farm output. Modelling and Simulation Society of Australia and New Zealand, 2007.

[17] Klink K. Trends and interannual variability of wind speed distributions in Minnesota. Journal of Climate 2002;15:3311–7.

[18] Bogardi I, Matyasovszky I, Estimating daily wind speed under climate change. Solar Energy 1996.;57(3):239-248.

[19] Donatelli M, Bellocchi G, Habyarimana E, Confalonieri R, Micale F. An extensible model library for generating wind speed data. Computers and Electronics in Agriculture 2009;69:165–70. doi:10.1016/j.compag.2009.07.022.

[20] Carlin J, Haslett J. The probability distribution of wind power from a dispersed array of wind turbine generators. Journal of Applied Meteorology 1982;21:303-313.

[21] Mohandes MA, Rehman S, Halawani TO. A neural networks approach for wind speed prediction. Renewable Energy 1998;13:345–54. doi:10.1016/S0960-1481(98)00001-9.

[22] Tol RSJ. Autoregressive conditional heteroscedasticity in daily wind speed measurements. Theoretical and Applied Climatology 1997;56:113–22.

[23] Suomalainen K, Silva CA, Ferrão P, Connors S. Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning. Energy 2012;37:41–50. doi:10.1016/j.energy.2011.08.001.

[24] Haslett J, Raftery AE. Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. Applied Statistics 1989;38:1. doi:10.2307/2347679.

[25] Lawrance AJ, Lewis PAW. The Exponential Autoregressive-Moving Average EARMA (p,q) Process. Journal of the Royal Statistical Society Series B (Methodological) 1980;42:150–61.

[26] Damsleth E, El-Shaarawi AH. ARMA Models with Double-Exponentially Distributed Noise. Journal of the Royal Statistical Society Series B (Methodological) 1989;51:61–9.

[27] Mach P, Thuring J, Samal D. Transformation of Data for Statistical Processing. 29th International Spring Seminar on Electronics

Technology, 2006. ISSE '06, 2006, p. 278–82. doi:10.1109/ISSE.2006.365112.

[28] Widger WK. Estimations of Wind Speed Frequency Distributions Using Only the Monthly Average and Fastest Mile Data. J Appl Meteor 1977;16:244–7. doi:10.1175/1520-0450(1977)016<0244:EOWSFD>2.0.CO;2.

[29] Brown BG, Katz RW, Murphy AH. Time Series Models to Simulate and Forecast Wind Speed and Wind Power. J Climate Appl Meteor 1984;23:1184–95. doi:10.1175/1520-0450(1984)023<1184:TSMTSA>2.0.CO;2.

[30] Kubik M, Coker PJ, Hunt C. Using meteorological wind data to estimate turbine generation output: a sensitivity analysis. World Renewable Energy Congress, Sweden, 2011, p. 4074–81.

[31] Coppin P, Ayotte K, Steggel N. Wind resource assessment in Australia: a planners guide. CSIRO Wind Energy Research Unit; 2003.

[32] Smith K, Randall G, Malcolm D, Kelley N, Smith B. Evaluation of wind shear patterns at midwest wind energy facilities. American Wind Energy Association (AWEA) WINDPOWER 2002 Conference, 2002.

[33] Rareshide E, Tindal A, Johnson C, Graves A, Simpson E, Bleeg J, et al. Effects of complex wind regimes on turbine performance. Scientific Proceedings. American Wind Energy Association WINDPOWER Conference, 2009.

[34] Holttinen H. Hourly wind power variations in the Nordic countries. Wind Energy 2005;8:173–95.

[35] Sinden G. Characteristics of the UK wind resource: Long-term patterns and relationship to electricity demand. Energy Policy 2007;35:112–27. doi:10.1016/j.enpol.2005.10.003.

[36] Kavasseri RG, Nagarajan R. Evidence of Crossover Phenomena in Wind-Speed Data. IEEE Transactions on Circuits and Systems I: Regular Papers 2004;51:2255–62. doi:10.1109/TCSI.2004.836846.

[37] Correia PF, Ferreira de Jesus JM. Simulation of correlated wind speed and power variates in wind parks. Electric Power Systems Research 2010;80:592–8. doi:10.1016/j.epsr.2009.10.031.

[38] Gibescu M, Ummels BC, Kling WL. Statistical wind speed interpolation for simulating aggregated wind energy production under system studies. Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on, IEEE; 2006, p. 1–7.

[39] Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, et al. MERRA: NASA's modern-era retrospective analysis for research and applications. Journal of Climate 2011;24:3624–48.

[40] Laslett D, Creagh C, Jennings P. A method for generating synthetic hourly solar radiation data for any location in the south west

of Western Australia, in a world wide web page. Renewable Energy 2014;68:87–102. doi:10.1016/j.renene.2014.01.015.

[41] Geoscience Australia. GEODATA COAST 100K 2004 dataset ANZCW0703006621. Available from: http://www.ga.gov.au/meta/ANZCW0703006621.html; 2012.

[42] Visvalingam M, Whyatt JD. Line generalisation by repeated elimination of points. Cartographic Journal, The 1993;30:46–51.

[43] D. Freedman, P. Diaconis, On the histogram as a density estimator: L 2 theory, Probability Theory and Related Fields. 57 (1981) 453–476.

[44] Zheng WX. Study of a least-squares-based algorithm for autoregressive signals subject to white noise. Mathematical Problems in Engineering 2003;2003:93–101. doi:10.1155/S1024123X03210012.

[45] G. Schwarz, Estimating the Dimension of a Model, Ann. Statist. 6 (1978) 461–464. doi:10.1214/aos/1176344136.

[46] Skidmore E, Tatarko I. Stochastic wind simulation for erosion modeling. SMR 1990;705:16.

[47] Gille ST, Llewellyn Smith SG, Statom NM. Global observations of the land breeze. Geophysical Research Letters 2005;32(5):1-4.

## Appendix A. Detailed Wind Simulation Model

The following algorithm for generating synthetic hourly wind farm output power at any location within the SWWA was used. This algorithm is split into 4 sections depending on how often computation is required. Latitude and longitude values are in degrees, but all sine and cosine terms assume the argument is in radians, and the $\cos^{-1}$ term produces a value in radians. The simulation can be started on any day of the year, and at any hour of the day, by setting the day of the year variable DOY and hour of the day variable hr to the desired values. DOY can range from 1 to 365 (or 1 to 366 for modelling leap years), and hr from 0 to 23. The initial values of daily and hourly residuals $yd_{w,0}$, $yd_{w,1}$, $yns_{w,0}$, $yns_{w,1}$ and $yns_{w,2}$ are set to standard normally distributed random values. For a simulation of $N_{wf}$ wind farms, to enable distance correlation to be set, at each hour the wind power for each wind farm w = 1,$N_{wf}$ is calculated together.

The first section must be computed once before the simulation begins:

(1.1) Calculate wind shear seasonal coefficients. For month m, m = 1 to 12, calculate:

$$cs_m = \cos\left(\frac{\pi}{6}(m-6)\right) \tag{A1}$$

(1.2) Set daily auto regression coefficients

$$\varphi_{d1}=0.523237$$
$$\varphi_{d2}=-0.160552 \quad \text{(A2)}$$
$$\rho_d=0.88102$$

The second section must be computed once per wind farm w, w = 1,$N_{wf}$, before the simulation begins:

(2.1) From the latitude and longitude of the location of wind farm w ($lon_w$, $lat_w$), use Euclidean geometry and the coastline shape map coordinate data to calculate the distance from the coast $cdist_w$ (in km). The coastline shape map consists of 500 vertices in longitude and latitude coordinates ($lon_i$,$lat_i$), i = 1 to 500. The first vertex is the start of the coastline and is where the coast crosses the Northern Territory border. The coastline is approximated by a set of line segments, each defined by a pair of adjacent vertices ($lon_i$,$lat_i$) ($lon_{i+1}$,$lat_{i+1}$), i = 1 to 499. First calculate the horizontal and vertical position coordinates for wind farm w.

$$x_w=111.195(lon_w-129)\cos\left(\frac{\pi}{180}lat_w\right) \quad \text{km}$$
$$y_w=111.195\,lat_w \quad \text{km} \quad \text{(A3)}$$

Set $d_{min} = 10^8$. For each line segment i = 1 to 499, calculate

$$x_i=111.195\left(\frac{(lon_i+lon_{i+1})}{2}-129\right)\cos\left(\frac{\pi}{180}\frac{(lat_i+lat_{i+1})}{2}\right) \quad \text{km}$$
$$y_i=111.195\frac{(lat_i+lat_{i+1})}{2} \quad \text{km} \quad \text{(A4)}$$
$$d=(x_w-x_i)^2+(y_w-y_i)^2 \quad \text{km}^2$$
$$if\,(d<d_{min})\,d_{min}=d$$

Calculate

$$cdist_w = \sqrt{(d_{min})} \text{ km} \tag{A5}$$

(2.2) If the grid of MERRA yearly average wind speeds have locations [latmerra$_i$,lonmerra$_j$], where latmerra$_i$ is the latitude (degrees) and lonmerra$_j$ is the longitude (degrees) of node i,j, i = 1 to $N_{lat}$, j = 1 to $N_{lon}$, then let the matrix of MERRA year average wind speeds be denoted by Vyav$_{ij}$. For wind farm w, if lat$_w$ is the latitude, and lon$_w$ is the longitude, then find latmerra$_{in}$, the latitude of the nearest MERRA node with latitude greater than lat$_w$, and find lonmerra$_{jn}$, the longitude of the nearest MERRA node with longitude greater than lon$_w$, then calculate the average yearly wind speed vyav$_w$:

$$\begin{aligned}
fi_{lat} &= \frac{(lat_w - latmerra_{\text{in}})}{(latmerra_{\text{in}} - latmerra_{\text{in-1}})} \\
fi_{lon} &= \frac{(lon_w - lonmerra_{\text{jn}})}{(lonmerra_{\text{jn}} - lonmerra_{\text{jn-1}})} \\
vi_1 &= (1 - fi_{lat})\, vyav_{\text{in-1,jn-1}} + fi_{lat}\, vyav_{\text{in,jn-1}} \\
vi_2 &= (1 - fi_{lat})\, vyav_{\text{in-1,jn}} + fi_{lat}\, vyav_{\text{in,jn}} \\
vyav_w &= (1 - fi_{lon})\, vi_1 + fi_{lon}\, vi_2
\end{aligned} \tag{A6}$$

(2.3) For each wind farm w, calculate the seasonal variation coefficients, the diurnal and wind shear distance coefficients, and the diurnal latitude and peak time coefficients:

$$k0_w = \frac{0.0395956\,(lat_w + 34)}{1 + 0.00794402\,(lat_w + 34)^3} + \frac{lat_w + 34}{2\,(50 + cdist_w)}$$

$$k1_w = \frac{0.0804696}{(1 + 0.741463\,(lat_w + 35.1))\,(1 + \frac{cdist_w}{200})}$$

$$fc_w = \frac{1}{1 + (\frac{cdist_w}{75})^2}$$

$$fdist_w = \frac{(100 + cdist_w)}{(200 + cdist_w)}$$

$$fshear_w = \frac{cdist_w}{(50 + cdist_w)}$$

$$awsf_{w,m} = 0.005\,(1 - cs_m)\,(1 - fshear_w) \quad \text{for month m = 1,12}$$

$$bwsf_w = 0.056 + 0.0625\,fshear_w$$

$$cwsf_w = 0.01 + 0.1\,fshear_w \qquad \text{(A7)}$$

$$dwsf_w = 0.005 + 0.08125\,fshear_w$$

$$flat_w = \frac{1}{36 + lat_w}$$

$$asb_w = 16.5 + \frac{32.5\,cdist_w}{(700 + 3\,cdist_w)}$$

$$bsb_w = 1.75 - \frac{6.25\,cdist_w}{(125 + 5\,cdist_w)}$$

$$af_w = 0.72 + 0.21\,fdist_w$$

$$bf_w = 0.394\,fdist_w - 0.08\,flat_w - 0.572$$

$$cf_w = 1.75 + 0.13\,fdist_w - 0.5\,flat_w$$

$$df_w = 0.261 - 1.16\,fdist_w$$

(2.4) For each wind farm w, calculate the hourly and daily auto regression coefficients:

$$farlat = 0.05 + 0.4\frac{(36 + lat_w)}{(37 + lat_w)}$$

$$fardist = \frac{1}{1 + \frac{cdist_w}{100}}$$

$$\varphi_{w,1} = farlat(1.28 + 0.17\, fardist)$$

$$\varphi_{w,2} = farlat(-0.55 - 0.27\, fardist)$$

$$\varphi_{w,3} = farlat(0.095 + 0.07\, fardist) \qquad \text{(A8)}$$

$$\rho_w = 0.45 - \frac{0.051}{(1 + \frac{cdist_w}{50})}$$

$$\sigma_w = (1 - \frac{0.15}{(1 + 0.01\, cdist_w)})(1 - \frac{0.15}{(36 + lat_w)})$$

$$\sigma_{dbw} = 0.43(0.91 + \frac{0.09}{(1 + 0.01\, cdist_w)})(0.67 + \frac{1.32}{(39 + lat_w)})$$

(2.5) For each wind farm w, and month m, m = 1 to 12, calculate dawn and dusk times. Let mdoy = [ 15,44,75,105,136,166,197,228,258,289,319,350 ]

$$DOY = mdoy[m]$$

$$b = 0.017453(DOY - 81)$$

$$eot = \frac{(9.87\sin(2b) - 7.53\cos(b) - 1.5\sin(b))}{60}$$

$$ds = -0.40928\sin(0.0172142(284 + DOY))$$

$$noon = 12.275 - eot - \frac{(lon_w - 115.87)}{15}$$

$$lr = 0.017453\,lat_w$$

$$hsd = 3.8197186\cos^{-1}\left(\frac{-\sin(lr)\sin(ds)}{\cos(lr)\cos(ds)}\right) \quad \text{(A9)}$$

$$if\,(hsd < 0)\,hsd = -hsd$$

$$dawn = noon - hsd$$

$$dusk = noon + hsd$$

$$justafterdawn_{w,m} = dawn + 2$$

$$afterdawn_{w,m} = dawn + 4$$

$$beforedusk_{w,m} = dusk - 1$$

$$night_{w,m} = dusk + 3$$

$$wsfbase_{w,m} = 0.11 + 0.0625\,fshear_w + 0.02\,cs_m$$

(2.6) For each wind farm w, that uses wind turbine type wt, modify reference velocities for the individual turbine to represent overall wind farm operation. If another wind farm uses the same type of turbine, this step does not have to be repeated.

$$
\begin{aligned}
vc_{wt} &= vco_{wt} - 0.5 \\
vr_{wt} &= vro_{wt} + 5 \\
vf_{wt} &= 0.4\, vc_{wt} + 0.6\, vro_{wt} \\
vs_{wt} &= vso_{wt} - 3 \\
c_{wt} &= \frac{0.5}{\left(vr_{wt} - vf_{wt}\right)^3} \\
b_{wt} &= \frac{0.5}{\left(vf_{wt}^3 - vc_{wt}^3\right)} \\
a_{wt} &= -b_{wt}\, vc_{wt}
\end{aligned}
\tag{A10}
$$

Where parameters with suffix o refer to the original individual wind turbine parameters. $vco_{wt}$ is the cut-in wind speed, $vro_{wt}$ is the reference wind speed (the speed at which the turbine reaches full power output), and $vso_{wt}$ is the shut-down wind speed. These parameters can be obtained from the wind turbine technical specifications. Values for some turbines which are used in the simulation are given below in Table A1. $a_{wt}$, $b_{wt}$, and $c_{wt}$ are constants associated with the partial power section of the turbine power curve.

**Table A.1**
Wind turbine power curve parameters.

| Turbine | Capacity (MW) | vco (m/s) | vro (m/s) | vso (m/s) | hub-height (m) |
|---|---|---|---|---|---|
| ENERCON-E40 | 0.6 | 2.5 | 12 | 28 | 46 |
| ENERCON-E48 | 0.8 | 2.5 | 14 | 28 | 50 |
| ENERCON-E66 | 1.8 | 2.5 | 15 | 28 | 65 |
| ENERCON-E70 | 2.3 | 2.5 | 15 | 28 | 64 |
| ENERCON-E126 | 7.5 | 2.5 | 17 | 28 | 135 |
| VESTAS-V82 | 1.65 | 3.5 | 12.5 | 20 | 78 |
| VESTAS-V90 | 1.856 | 4 | 12 | 25 | 80 |
| VESTAS-V112 | 3 | 3 | 12 | 25 | 119 |
| GE 2.5-100 | 2.5 | 3 | 12.5 | 25 | 75 |
| REPOWER 3.4M104 | 3.4 | 3.5 | 13.5 | 25 | 78 |
| ENERCON-E53 | 0.8 | 2.5 | 14 | 28 | 73 |
| ENERCON-E40 | 0.5 | 2.5 | 12 | 25 | 44.2 |

(2.7) For each wind farm w, calculate distance correlation weighting coefficients $f_{dc}$. Estimate $dist_{w1w2}$, the distance between wind farm $w_1$ and wind farm $w_2$ (km), and $fdc_{w1w2}$, the distance weighting factor between wind farm $w_1$ and wind farm $w_2$ for every possible pair of wind farms $w_1$ and $w_2$. For wind farms $w_1 = 1$ to $N_{wf} - 1$, and for wind farms $w_2 = w_1+1$ to $N_{wf}$:

$$dx = 111.195\left(\left(lon_{w1} - 129\right)\cos\left(\frac{\pi}{180} lat_{w1}\right) - \left(lon_{w2} - 129\right)\cos\left(\frac{\pi}{180} lat_{w2}\right)\right)$$

$$dy = 111.195\left(lat_{w1} - lat_{w2}\right)$$

$$dist_{w1w2} = \sqrt{dx^2 + dy^2}$$

$$d_1 = \frac{dist_{w1w2}}{60}$$

$$d_2 = \left(\frac{dist_{w1w2}}{500}\right)^8 \qquad \text{(A11)}$$

$$fdc_{w1,w2} = \frac{1}{\left(1 + d_1\right)\left(1 + d_2\right)}$$

$$fdc_{w2,w1} = fdc_{w1,w2}$$

$$fdc_{w1,w1} = 1$$

For wind farms w = 1 to $N_{wf}$

$$fdctot = \sum_{w_1 = 1, N_{wf}} fdc_{w,w1} \qquad \text{(A12)}$$

$$fdc_{w,w1} = \frac{fdc_{w,w1}}{fdctot} \quad \text{for } w_1 = 1 \text{ to } N_{wf} \qquad \text{(A13)}$$

$$fdctotsq = \frac{1}{\sum_{w_1 = 1, N_{wf}} fdc_{w,w1}^2} \quad \text{for } w_1 = 1 \text{ to } N_{wf} \qquad \text{(A14)}$$

$$fdc_{w,w1} = fdc_{w,w1}\, fdctotsq \quad \text{for } w_1 = 1 \text{ to } N_{wf} \qquad \text{(A15)}$$

$$fdc_{w,N_{wf}+1} = fdctotsq \qquad \text{(A16)}$$

The third section must be calculated at the beginning of each day (hr = 0), and at the beginning of the simulation if hr does not start at 0.

(3.1) Calculate the seasonal variation coefficient fseason, first seasonal mode maximum amplitude coastal and inland components f0c and f0i, and the second seasonal mode maximum amplitude $f_1$ (all same for every wind farm). Let:

dom[13] = {0,31,59,90,120,151,181,212,243,273,304,334,366},

dmm[14] = {-15,15,44,75,105,136,166,197,228,258,289,319,350,380},

fsm[13] = { -1,-1,-1,-0.5,0,0.5,1,1,1,0.5,0,-0.5,-1 },

f0mi[14] = { 0.75,1.8,0.75,0.6,-1.0,-0.4,-0.5,-1.7,-1.7,0.4,0.5,1.4,0.75,1.8 },

f0mc[14] = { 1.2,1.5,0.5,0.7,-0.6,-0.7,-0.6,-0.7,-1.25,-0.3,0,0.6,1.2,1.5 }, and

f1m[14] = { -0.3,0,0.3,-0.25,-0.6,-0.75,0.5,0.12,0.25,1,-0.25,0.1,-0.3,0 },

If not at beginning of simulation, then increment day of year DOY by 1. From DOY, find the month m (m = 1..12) such that dom[m-1] < DOY <= dom[m], and the mid-month number mm (mm = 1..13) such that dmm[mm-1] < DOY <= dmm[mm].

$$fseason = 1 + fsm[m]$$

$$f_{mm} = \frac{(doy - dmm[mm-1])}{(dmm[mm] - dmm[mm-1])}$$

$$f0c = f0mc[mm-1] + f_{mm}(f0mc[mm] - f0mc[mm-1]) \quad \text{(A17)}$$

$$f0i = f0mi[mm-1] + f_{mm}(f0mi[mm] - f0mi[mm-1])$$

$$f1 = f1m[mm-1] + f_{mm}(f1m[mm] - f1m[mm-1])$$

(3.2) For each wind farm w, calculate the first seasonal mode maximum amplitude:

$$f0_w = fc_w f0c + (1 - fc_w) f0i \quad \text{(A18)}$$

(3.3) For each wind farm w, calculate the seasonal wind speed:

$$Vseason_w = Vyav_w(1 + k0_w f0_w + k1_w f1) \quad \text{(A19)}$$

At the beginning of the simulation only, set $vdav_w$ to $Vseason_w$.

(3.4) For each wind farm w, calculate the sea breeze peak time:

$$tsb_w = asb_w + bsb_w\, fseason$$

if ($tsb_w < 0$) $tsb_w = 0$

if ($tsb_w > 23$) $tsb_w = 23$ (A20)

(3.5) For each wind farm w, generate four normally distributed random numbers, $r_{w,i}$:

$$r_{w,i} = \frac{\left(z^{0.135} - (1-z)^{0.135}\right)}{0.1975}$$ i = 1 to 4, z uniformly distributed on (0,1) (A21)

(3.6) For each wind farm w, distance weight the random variables $rdc_{w,i}$, i = 1 to 4:

$$rdc_{w,i} = \sum_{q=1,N_{wf}} fdc_{w,q}\, r_{q,i}$$ (A22)

(3.7) For each wind farm w, calculate the average daily velocity. The previous daily autoregressive residual calculations are moved one day back and today's daily residual $yd_{w,0}$ is calculated.

$$yd_{w,2} = yd_{w,1}$$
$$yd_{w,1} = yd_{w,0}$$
$$yd_{w,0} = \varphi_{d1}\, yd_{w,1} + \varphi_{d2}\, yd_{w,2} + \rho_d\, rdc_{w,1}$$ (A23)

Move previous daily average wind speed one day back, and apply reverse square root transformation to $yd_{w,0}$.

$$vdavl_w = vdav_w$$

$$\sigma_{dw} = (1 + 0.225\, fseason)\, \sigma_{dbw}$$

$$vdav_w = \left(\sqrt{Vseason_w} + \sigma_{dw}\ yd_{w,0}\right)^2 - \sigma_{dw}{}^2 \qquad \text{(A24)}$$

$$if\,(vdav_w > 16.5)\, vdav_w = 16.5$$

$$if\,(vdav_w < 1)\, vdav_w = 1$$

(3.8) For each wind farm w, calculate the diurnal component coefficients. The previous day's calculations are moved one day back (a late peaking diurnal component from the previous day may still remain active into the present day).

$$tpeak_{w,0} = tpeak_{w,1} - 24$$

$$tperiod_{w,0} = tperiod_{w,1}$$

$$tfreq_{w,0} = tfreq_{w,1}$$

$$tstart_{w,0} = tstart_{w,1} - 24 \qquad \text{(A25)}$$

$$tstop_{w,0} = tstop_{w,1} - 24$$

$$tmag_{w,0} = tmag_{w,1}$$

(3.9) Calculate today's diurnal component peak time of day, $tpeak_{w,1}$

$$f_{pk} = af_w + bf_w fseason$$

$$f_{pk2} = cf_w + df_w fseason$$

$$tpeak_{w,1} = 15 - 0.5\left(f_{pk2} - rdc_{w,2}\right) \quad \text{rdc}_{w,2} < -\text{f}_{pk2}$$

$$tpeak_{w,1} = 7.5 + 0.5\, fseason - 3\left(f_{pk} - rdc_{w,2}\right) \quad -\text{f}_{pk2} < \text{rdc}_{w,2} < -\text{f}_{pk}$$

if $-\text{f}_{pk} < \text{rdc}_{w,2} < \text{f}_{pk}$

{ $tsb_w = asb_w + bsb_w fseason$

if ($\text{tsb}_w < 0$) $\text{tsb}_w = 0$

if ($\text{tsb}_w > 23$) $\text{tsb}_w = 23$

$tpeak_{w,1} = tsb_w + 3\, rdc_{w,2}$ } (A26)

$$tpeak_{w,1} = 7.5 + 0.5\, fseason + 3\left(rdc_{w,2} - f_{pk}\right) \quad \text{f}_{pk} < \text{rdc}_{w,2} < \text{f}_{pk2}$$

$$tpeak_{w,1} = 15 + 0.5\left(rdc_{w,2} - f_{pk2}\right) \quad \text{rdc}_{w,2} > \text{f}_{pk2}$$

while $\text{tpeak}_{w,1} < 0$ add 24 hours to $\text{tpeak}_{w,1}$

while $\text{tpeak}_{w,1} >= 36$ subtract 24 hours from $\text{tpeak}_{w,1}$

(3.10) Calculate this days diurnal component period $\text{tperiod}_{w,1}$

$$tperiod_{w,1} = 24 + 2\, rdc_{w,3} \quad \text{tpeak}_{w,1} < 6$$

$$tperiod_{w,1} = 16 - fseason + \left(3 - 0.75\, fseason\right) rdc_{w,3} \quad 6 \le \text{tpeak}_{w,1}$$

(A27)

if $\text{tperiod}_{w,1} < 6$ $\text{tperiod}_{w,1} = 6$

if $\text{tperiod}_{w,1} > 36$ $\text{tperiod}_{w,1} = 36$

(3.11) Calculate this days diurnal component magnitude $\text{tmag}_{w,1}$

$$dt = tpeak_{w,1} - 8.5$$

$$dv = vdav_w - 5 + 0.25\, fseason$$

$$am = -0.825 - 0.66\, fseason$$

$$bm = 0.1485 + 0.033\, fseason$$

$$cm = 3.2959 - 0.21327\, fseason - 0.7755 / \left(1 + 0.5\, dt^2\right)$$

$$dm = 0.275 - 0.1155\, fseason + 0.11\, vdav_w \qquad \text{(A28)}$$

$$tmag_{w,1} = \left(1 - 0.15\, flat_w \left(2 - fseason\right)\right)\left(am + bm\ vdav_w + cm / \left(1 + 0.15\ dv^2\right)\right) + dm\ rdc_{w,4}$$

if ($tmag_{w,1} < 0$) $\quad tmag_{w,1} = 0$

if ($tmag_{w,1} > vdav_w$) $\quad tmag_{w,1} = vdav_w$

if ($tmag_{w,1} > 7$) $\quad tmag_{w,1} = 7$

(3.12) Calculate $tstart_{w,1}$, $tstop_{w,1}$, and $tfreq_{w,1}$. Make sure today's diurnal component doesn't start before midnight (hr = 0).

$$if\left(tpeak_{w,1} < 12\right) \quad \{ \quad tmag_{w,1} = -tmag_{w,1} \quad , \quad tpeak_{w,1} = tpeak_{w,1} + \frac{tperiod_{w,1}}{2} \quad \} \qquad \text{(A29)}$$

$$tstart_{w,1} = tpeak_{w,1} - \frac{3\, tperiod_{w,1}}{4} \qquad \text{(A30)}$$

$$if\left(tstart_{w,1} < 0\right) \quad \{ \quad tperiod_{w,1} = 1.333\, tpeak_{w,1} \quad , \quad tstart_{w,1} = 0 \quad \}$$

$$tstop_{w,1} = tpeak_{w,1} + \frac{tperiod_{w,1}}{4} \qquad \text{(A31)}$$

$$tfreq_{w,1} = \frac{2\pi}{tperiod_{w,1}}$$

The fourth section must be calculated each hour for each wind farm:

(4.1) for each wind farm w, calculate total hourly wind speed, where hr is the hour of the day. First generate a normally distributed random number, then move the previous hourly autoregressive calculations one hour back and calculate this hour's synthetic normally distributed residual $yns_{w,0}$:

$$r_h = \frac{\left(z^{0.135} - (1-z)^{0.135}\right)}{0.1975} \quad \text{z uniformly distributed on (0,1)}$$

$$\begin{aligned} yns_{w,3} &= yns_{w,2} \\ yns_{w,2} &= yns_{w,1} \\ yns_{w,1} &= yns_{w,0} \\ yns_{w,0} &= \varphi_{w,1}\, yns_{w,1} + \varphi_{w,2}\, yns_{w,2} + \varphi_{w,3}\, yns_{w,3} + \rho_w r_h \end{aligned} \tag{A32}$$

Apply reverse data transformation to calculate this hour's synthetic wind speed residual $ys_w$:

$$\begin{aligned} ys_w &= 1.96 - \left(1.4 - 0.302\, yns_{w,0}\right)^2 \quad yns_{w,0} < 0 \\ ys_w &= \left(1.4 + 0.302\, yns_{w,0}\right)^2 - 1.96 \quad yns_{w,0} \geq 0 \end{aligned} \tag{A33}$$

Calculate diurnal wind speed:

$$\begin{aligned} vdiurnal_w &= 0 \\ vdiurnal_w &= vdiurnal_w + tmag_{w,0} \cos\left(tfreq_{w,0}\left(hr - tpeak_{w,0}\right)\right) \quad tstart_{w,0} < hr < tstop_{w,0} \\ vdiurnal_w &= vdiurnal_w + tmag_{w,1} \cos\left(tfreq_{w,1}\left(hr - tpeak_{w,1}\right)\right) \quad tstart_{w,1} < hr < tstop_{w,1} \end{aligned} \tag{A34}$$

Calculate daily average wind speed trend:

$$vhm_w = vdavl_w + \frac{hr\left(vdav_w - vdavl_w\right)}{24} \tag{A35}$$

Calculate total hourly wind speed:

$$v_w = vhm_w + vdiurnal_w + \sigma_w ys_w$$
$$if\,(v_w < 0)\, v_w = 0 \qquad \text{(A36)}$$

(4.2) for each wind farm w, calculate wind shear factor and hub height wind speed. m is the number of the current month.

$$if\,(hr < justafterdawn_{w,m})\; hr = hr + 24$$
$$if\,(hr < beforedusk_{w,m})$$
{ $if\,(hr < afterdawn_{w,m})$
$$wsf = wsfbase_{w,m} + 0.5(hr - justafterdawn_{w,m})(awsf_{w,m}(v_w - 5) - bwsf_w)$$
else
$$wsf = wsfbase_{w,m} + awsf_{w,m}(v_w - 5) - bwsf_w \quad \}$$
else
{ $if\,(hr < night_{w,m})$
$$wsf = wsfbase_{w,m} + 0.25\,(hr - beforedusk_{w,m})(cwsf_w + dwsf_w(8 - v_w)) \qquad \text{(A37)}$$
else
$$wsf = wsfbase_{w,m} + cwsf_w + dwsf_w(8 - v_w) \quad \}$$
$$if\,(wsf < 0)\, wsf = 0$$
$$if\,(wsf > 0.7)\, wsf = 0.7$$

$$vhh_w = v_w \left(\frac{hh_{wt}}{50}\right)^{wsf}$$

Where $hh_{wt}$ is the hub height of turbine wt used by the wind farm, and $vhh_w$ is the hub-height wind speed of wind farm w.

(4.3) Calculate wind farm power output. For each wind farm w, using wind turbine wt:

$$if\left(vhh_w \le vc_{wt}\right) CF = 0$$

else

$$\{ \quad if\left(vhh_w \ge vr_{wt}\right)$$

$$\{ \quad if\left(vhh_w > vs_{wt}\right)$$

$$\{ \quad vd = vhh_w - vs_{wt}$$

$$if\left(vd \ge 6\right) CF = 0$$

else

$$if\left(vd \ge 3\right)$$

$$CF = \frac{\left(vd-6\right)^2}{18}$$

else

$$CF = 1 - \frac{vd^2}{18} \quad \}$$

else $CF = 1 \quad \}$

else

if $\left(vhh_w \le vf_{wt}\right) CF = a_{wt} + b_{wt}\, vhh_w^3$

else

$$\{ \quad vd = vr_{wt} - vhh_w$$

$$CF = 1 - c_{wt}\, vd^3 \quad \}\}$$

(A38)

$$power_w = CF * capacity_w$$

where $capacity_w$ is the full power capacity of wind farm w.

(4.4) Advance the hour of the day variable hr by 1. If hr ≥ 24, advance the day of year variable DOY by 1 and set hr = 0. If DOY > 365 (or DOY > 366 if a leap year is being modelled), set DOY = 1.

## Appendix B. Formulation of statistical measures

The root mean square error (RMSE) is calculated using:

$$\mathrm{RMSE}=\frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(h_m-h_d\right)^2}$$

where $h_d$ is the measured data value, $h_m$ is the model generated synthetic data value, and N is the number of data points. RMSE can be represented as a percentage value by dividing by the mean value of $h_d$ and multiplying by 100. The mean bias error (MBE) is calculated using:

$$\mathrm{MBE}=\frac{1}{N}\sum_{1}^{N}\left(h_m-h_d\right)$$

MBE can be represented as a percentage value by dividing by the mean value of $h_d$ and multiplying by 100.The RMSE is a measure of the magnitude of the difference between individual data points in each data set. The sign of the difference is ignored. The MBE is a measure of the average difference between individual data points in each data set, or whether the model generated data set is biased higher or lower compared to the measured data on the whole. Here the sign of the difference is not ignored. Generally, a model with a lower RMSE than another fits the data more closely. In this study a model with a negative MBE might be considered more favourably than a model with a similar but positive MBE because under predicting wind power generation on the whole is more desirable than over predicting. Representing RMSE and MBE as a percentage gives an idea of how significant the error is compared to the average value of the measured data.

# Appendix C. Nomenclature

| | | | |
|---|---|---|---|
| $\alpha$ | wind shear exponent | $kslat_2$ | weighting coefficient for seasonal mode 2 |
| $\lambda$ | weibull scale parameter | $lat_i$ | latitude of coast-line map vertex i (degrees) |
| $\sigma_{air}$ | air density ($Kg/m^3$) | $lon_i$ | longitude of coast-line map vertex i (degrees) |
| $\varrho$ | generic hourly wind speed residual random noise component standard deviation | $lat_w$ | latitude of wind farm w (degrees) |
| $\varrho_d$ | synthetic daily average wind speed residual random noise component standard deviation | $lon_w$ | longitude of wind farm w (degrees) |
| $\varrho_w$ | hourly wind speed residual random noise component standard deviation for wind farm w | $latmerra_i$ | latitude of MERRA grid nodes i,j, j = 1 to $N_{lon}$ (deg) |
| $\sigma$ | generic hourly wind speed standard deviation coefficient | $latmerra_{in}$ | latitude of nearest MERRA node with latitude greater than $lat_w$ (deg) |
| $\sigma_{dbw}$ | base daily wind speed standard deviation coefficient for wind farm w ($m^{1/2}s^{-1/2}$) | $lonmerra_j$ | longitude of MERRA grid nodes i,j, i= 1 to $N_{lat}$(deg) |
| $\sigma_{dw}$ | seasonally adjusted daily wind speed standard deviation coefficient for wind farm w ($m^{1/2}s^{-1/2}$) | $lonmerra_{jn}$ | longitude of the nearest MERRA node with longitude greater than $lon_w$ (deg) |
| $\sigma_w$ | hourly wind speed standard deviation coefficient for wind farm w ($ms^{-1}$) | lr | latitude of wind farm w in radians (radians) |
| $\phi_{d1}$ | synthetic daily wind speed residual first order auto-regression coefficient | MA | moving average model |
| $\phi_{d2}$ | synthetic daily wind speed residual second order auto-regression coefficient | MAPE | mean absolute percent error |
| $\phi_k$ | generic k-th order auto-regression parameter | MBE | mean bias error |
| $\phi_{w,1}$ | synthetic hourly wind speed residual first order auto-regression coefficient for wind farm w | mdoy | mid-month day of year array |
| $\phi_{w,2}$ | synthetic hourly wind speed residual second order auto-regression coefficient for wind farm w | $night_{w,m}$ | time of day at 3 hours after dusk for wind farm w during month m (hr) |
| $\phi_{w,3}$ | synthetic hourly wind speed residual third order auto-regression coefficient for wind farm w | $N_{lat}$ | number of horizontal MERRA grid lines (22) |
| $af_w$ | diurnal peak time coefficient for wind farm w | $N_{lon}$ | number of vertical MERRA grid lines (15) |
| $asb_w$ | diurnal peak time coefficient for wind farm w | noon | local time at which solar altitude is maximum (hr) |
| $a_{wt}$ | wind farm wide power curve parameter for turbine type wt ($m^{-2}s^2$) | $N_{wf}$ | number of wind farms |
| $afterdawn_{w,m}$ | time of day at 4 hours after dawn for wind farm w during month m (hrs) | p | autoregressive model order |

| | | | |
|---|---|---|---|
| am | diurnal wind magnitude constant component coefficient | P | specific power per unit area ($W/m^2$) |
| AR | auto regressive model | $power_w$ | power output for wind farm w (MW) |
| ARMA | auto regressive moving average model | q | moving average model order |
| $awsf_{w,m}$ | wind shear factor coefficient for wind farm w | r | generic normally distributed random variable |
| b | day of year angle for equation of time (radians) | $r_d$ | generic normally distributed random variable for daily synthetic wind speed residual generation |
| $b_{wt}$ | wind farm wide power curve parameter for turbine type wt ($m^{-3}s^3$) | $r_{w,i}$ | normally distributed random variables for wind farm w, i = 1 to 4 |
| $beforedusk_{w,m}$ | time of day at 1 hour before dusk for wind farm w during month m (hrs) | $rdc_{w,i}$ | distance weighted combination random variables for wind farm w, i = 1 to 4 |
| $bf_w$ | diurnal peak time coefficient for wind farm w | $r_h$ | standard normally distributed random variable for hourly synthetic wind speed residual generation |
| BIC | Bayesian Information Criterion | RMSE | root mean square eror |
| bm | diurnal wind magnitude daily wind speed component coefficient | t | time of day (hr) |
| $bsb_w$ | diurnal peak time coefficient for wind farm w | $tfreq_{w,0}$ | frequency of yesterday's diurnal wind sinusoid for wind farm w (radians per hr) |
| $bwsf_w$ | wind shear factor coefficient for wind farm w | $tfreq_{w,1}$ | frequency of today's diurnal wind sinusoid for wind farm w (radians per hr) |
| $c_{wt}$ | wind farm wide power curve parameter for turbine type wt ($m^{-3}s^3$) | $t_{mag}$ | generic magnitude of diurnal wind sinusoid ($ms^{-1}$) |
| $capacity_w$ | full power capacity for wind farm w (MW) | $tmag_{w,0}$ | magnitude of yesterday's diurnal wind sinusoid for wind farm w ($ms^{-1}$) |
| $cdist_w$ | distance inland of wind farm w from the nearest part of the coastline (km) | $tmag_{w,1}$ | magnitude of today's diurnal wind sinusoid for wind farm w ($ms^{-1}$) |
| CF | capacity factor | $t_{peak}$ | generic time of day when peak wind speed occurs in de-trended dataset (hr) |
| $cf_w$ | diurnal peak time coefficient for wind farm w | $tpeak_{w,0}$ | peak hour of yesterday's diurnal wind sinusoid for wind farm w (hr) |
| cm | diurnal wind magnitude peak component coefficient | $tpeak_{w,1}$ | peak hour of today's diurnal wind sinusoid for wind farm w (hr) |
| $cs_m$ | wind shear seasonal coefficient for month m | $t_{period}$ | generic timespan of diurnal wind sinusoid (hr) |
| $cwsf_w$ | wind shear factor coefficient for wind farm w | $tperiod_{w,0}$ | period of yesterday's diurnal wind sinusoid for wind farm w (hr) |
| d | square of distance between wind farm w and coast-line segment i ($km^2$) | $tperiod_{w,1}$ | period of today's diurnal wind sinusoid for wind farm w (hr) |
| $d_1$ | distance factor 1 between two wind farms (km) | $tsb_w$ | sea breeze peak time of day base value for wind farm w (hr) |

| | | | |
|---|---|---|---|
| $d_2$ | distance factor 2 between two wind farms ($km^8$) | $tstart_{w,0}$ | start time of day of yesterday's diurnal wind sinusoid for wind farm w (hr) |
| $d_{min}$ | square of distance inland from the nearest part of the coastline ($km^2$) | $tstart_{w,1}$ | start time of day of today's diurnal wind sinusoid for wind farm w (hr) |
| dawn | sunrise local time (hr) | $tstop_{w,0}$ | stop time of day of yesterday's diurnal wind sinusoid for wind farm w (hr) |
| $df_w$ | diurnal peak time coefficient for wind farm w | $tstop_{w,1}$ | stop time of day of today's diurnal wind sinusoid for wind farm w (hr) |
| $dist_{w1w2}$ | distance estimate between wind farm w1 and wind farm w2 (km) | v | wind speed (m/s) |
| dm | diurnal wind magnitude variation coefficient | $v_1$,$v_2$ | wind speed at heights $h_1$,$h_2$ (m/s) |
| dmm | mid-month cumulative day of year array | $vc_{w,t}$ | wind farm wide cut-in wind speed for turbine type wt ($ms^{-1}$) |
| dom | end of month cumulative day of year array | $vco_{w,t}$ | single turbine cut-in wind speed for turbine type wt ($ms^{-1}$) |
| DOY | day of year | $Vdav_w$ | Daily average wind speed base value ($ms^{-1}$) |
| ds | solar declination angle (radians) | $Vdavl_w$ | Daily average wind speed base value from yesterday ($ms^{-1}$) |
| dt | diurnal wind magnitude peak hour factor | $V_{diurnal}$ | generic diurnal wind speed ($ms^{-1}$) |
| dusk | sunset local time (hr) | $Vdiurnal_w$ | diurnal wind speed for wind farm w ($ms^{-1}$) |
| dx | horizontal distance estimate between two wind farms (km) | $v_{ds(t)}$ | synthetic daily average wind speed ($ms^{-1}$) |
| dy | vertical distance estimate between two wind farms (km) | $vf_{w,t}$ | wind farm wide power curve level-off threshold wind speed for turbine type wt ($ms^{-1}$) |
| dv | diurnal wind magnitude daily wind speed factor | $vhh_w$ | hub height wind speed for wind farm w ($ms^{-1}$) |
| $dwsf_w$ | wind shear factor coefficient for wind farm w | $vhm_w$ | daily average wind speed trend ($ms^{-1}$) |
| eot | equation of time | $vi_1$ | interpolated wind speed 1 ($ms^{-1}$) |
| f | probability density function (also frequency distribution) for wind as a function of wind speed | $vi_2$ | interpolated wind speed 2 ($ms^{-1}$) |
| f0c | coastal first seasonal mode maximum amplitude | $V_{mode1}$ | magnitude of seasonal mode 1 |
| f0mc | coastal first seasonal mode maximum amplitude by month array | $V_{modee2}$ | magnitude of seasonal mode 2 |
| f0w | first seasonal mode maximum amplitude for wind farm w | $vr_{w,t}$ | wind farm wide reference wind speed for turbine type wt ($ms^{-1}$) |
| f0i | inland first seasonal mode maximum amplitude | $vro_{w,t}$ | single turbine reference wind speed for turbine type wt ($ms^{-1}$) |
| f0mi | inland first seasonal mode maximum amplitude by month array | $vs_{w,t}$ | wind farm wide shut down wind speed for turbine type wt ($ms^{-1}$) |
| f1 | second seasonal mode maximum amplitude | $v_s(t)$ | synthetic hourly wind speed ($ms^{-1}$) |

| | | | |
|---|---|---|---|
| f1m | second seasonal mode maximum amplitude by month array | $v_{season}$ | generic seasonal wind speed ($ms^{-1}$) |
| fardist | AR parameter distance factor | $Vseason_w$ | seasonal wind speed for wind farm w ($ms^{-1}$) |
| farlat | AR parameter latitude factor | $vso_{w,t}$ | single turbine shut down wind speed for turbine type wt ($ms^{-1}$) |
| $fdc_{w1,w2}$ | distance correlation factor between wind farm w1 and wind farm w2 | $v_{yav}$ | yearly average wind speed ($ms^{-1}$) |
| fdctot | sum of raw distance correlation factors between wind farm w and all other wind farms | $Vyav_{ij}$ | yearly average wind speed at MERRA node i,j i = 1 to $N_{lat}$, j = 1 to $N_{lon}$ ($ms^{-1}$) |
| fdctotsq | reciprocal of the sum of squares of the fractional distance correlation factors between wind farm w and all other wind farms | $Vyav_w$ | estimate of yearly average wind speed at wind farm w ($ms^{-1}$) |
| $fdist_w$ | diurnal distance coefficient | wsf | wind shear factor |
| $fi_{lat}$ | latitude interpolation factor | $wsfbase_{w,m}$ | base wind shear factor for wind farm w during month m |
| $fi_{lon}$ | longitude interpolation factor | $x_i$ | horizontal position coordinate for coast-line map line segment i (km) |
| $flat_w$ | diurnal latitude coefficient | $yd_{w,0}$ | daily average wind speed residual for wind farm w |
| fmm | seasonal mode maximum amplitude linear interpolation coefficient | $yd_{w,1}$ | daily average wind speed residual for wind farm w from yesterday |
| $f_{pk}$ | diurnal wind sinusoid peak hour distribution factor 1 | $yd_{w,2}$ | daily average wind speed residual for wind farm w from two days ago |
| $f_{pk2}$ | diurnal wind sinusoid peak hour distribution factor 2 | $y_i$ | vertical position coordinate for coast-line map line segment i (km) |
| fseason | seasonal coefficient | $x_w$ | horizontal position coordinate for wind farm w (km) |
| $fshear_w$ | wind shear distance coefficient | $y_w$ | vertical position coordinate for wind farm w (km) |
| fsm | seasonal coefficient by month array | y | hourly wind speed residual |
| ft | wind shear factor time coefficient | $y_{ds}$ | synthetic daily average wind speed residual |
| $h_1$ | height above ground (m) | $y_n$ | transformed hourly MERRA wind speed residual |
| $h_2$ | height above ground (m) | $y_{ns}$ | generic synthetic normally distributed hourly wind speed residual |
| $hh_{wt}$ | hub height for wind turbine wt (m) | $yns_{w,0}$ | synthetic normally distributed hourly wind speed residual for wind farm w for present hour |
| hr | hour of day | $yns_{w,1}$ | synthetic normally distributed hourly wind speed residual for wind farm w for previous hour |
| hsd | time between sunrise and noon, or noon and sunset (hrs) | $yns_{w,2}$ | synthetic normally distributed hourly wind speed residual for wind farm w for two hours before present |

| | | | |
|---|---|---|---|
| $justafterdawn_{w,m}$ | time of day of 2 hours after dawn for wind farm w during month m (hrs) | $y_s$ | generic synthetic hourly wind speed residual |
| k | weibull shape parameter | $ys_w$ | present hour synthetic wind speed residual for wind farm w |
| $k0_w$ | seasonal variation mode 0 coefficient | z | random variable uniformly distributed between 0 and 1 |
| $k1_w$ | seasonal variation mode 1 coefficient | $z_o$ | roughness length (m) |
| $kslat_l$ | weighting coefficient for seasonal mode 1 | | |