Modelling the Volatility in Wind Farm Output

[1]Ward, Kathryn and [2]Boland, John

[1]School of Mathematics and Statistics, Division of Information Technology, Engineering and the Environment
University of South Australia, Mawson Lakes Campus, Mawson Lakes, SA 5095, Australia.
E-Mail: kathryn.gilbert@postgrads.unisa.edu.au
[2] E-Mail: john.boland@unisa.edu.au

Keywords: *Correlative Coherence, ARMA, Double Exponential Distribution, Fourier Series, Wind Farms*

EXTENDED ABSTRACT

The penetration of wind energy into the South Australian electricity market has been rapidly growing in the past decade. To be able to bid in the lucrative peak demand market, the wind farm industry needs to be able to predict, with some certainty, the amount of electricity that they can produce in the next trading cycle. We present here a methodology to model variability in the output from the various South Australian wind farms, both individually and combined, at the half hourly time interval.

This paper extends Boland's (2005) [1] paper in which he showed a methodology to determine the overall correlation, known as correlative cohesion [3], between wind farms. This correlative cohesion gives us an indication of how the wind farms work together so that the volatility effects of the individual turbine and farms are minimised.

We show a method to model the wind farm output volatility that shows the: estimation of the deterministic and random components of each farm's half-hourly time series, calculation of the overall correlative coherence between all of the farms and development of a null distribution for the correlative coherence between all the farms.

We begin by removing the cyclic component of the time series using Fourier analysis. We then remove the remaining deterministic structure using ARMA processes. We finally model the random component of the time series using a double exponential distribution (also known as the Laplace distribution) [2]. In order to smooth out the overall volatility of the wind farms, we need to determine how correlated they all are to each other. To begin with, we find the Spearman's correlation coefficient (non-parametric correlation) of each pair of wind farms. We build a symmetric matrix R using the pairs of Spearman correlations where $r_{ij} = Cor(X_i, X_j) = r_{ji}$ for $i \neq j$ and $r_{ij} = 1$ for $i = j$. Getz [3] presents a well defined measure of the diversity of the eigenvalues λ_i of R:

$$C(X^n) = 1 - \frac{1}{\ln(1/n)} \sum_{i=1}^{n} \left(\frac{\lambda_i}{n}\right) \ln\left(\frac{\lambda_i}{n}\right) \qquad (1)$$

This is a generalisation of the concept of coherence for periodic signals and gives us a measure of the degree at which the output from the farms vary in concert with each other [3].

If all off-diagonal correlations have the same value $r \in [0, 1]$ (i.e. the overall correlation between all of the farms) and the diagonal elements remain as 1, then the eigenvalues of the correlation matrix $R(r)$ are $\lambda_1 = 1 + (n-1)r$ and $\lambda_i = 1 - r$ for $i = 2, \ldots, n$ and Equation 1 reduces to

$$C_n(r) = \frac{(1+(n-1)r)\ln(1+(n-1)r)+(n-1)(1-r)\ln(1-r)}{n \ln n} \qquad (2)$$

The correlative coherence of any system of farms X^n is therefore the solution to the equation [3]

$$r = C_n^{-1}(C(X^n)) \qquad (3)$$

We use Newton's method to solve Equation 3 for r and obtain an overall correlative coherence of $r = 0.209662$ for the six wind farms. We also investigate an initial method for developing the null distribution of the correlative coherence.

A low r value is desired in the wind farm network because it is much easier to integrate the output into the grid if the farms are not acting in unison (which tends to produce large spikes in the output). Boland (2005) used wind data from the Bureau of Meteorology (BoM) sites, rather than the actual output from the wind farms, to estimate the correlative cohesion value for the farms . His analysis yielded a correlative coherence value of $r = 0.742$ which is much higher than the more recent analysis. The analysis using actual output gives a more optimistic view of the contribution that the wind farms produce together.

The correlative coherence value r can be used as an indication of the way in which a set of farms follow the same pattern. Adding a new farm into the analysis is a quite straightforward process. The r value can be recalculated and then analysed to see what the effect that the new farm has on the correlative cohesion among all of the farms. The r value can also be used in a similar manner to determine likely locations for new farms.

1 INTRODUCTION

The penetration of wind energy into the South Australian electricity market has been rapidly growing in the past decade. Wind farms at six sites across South Australia already produce 388 MW of electricity with a further 341 MW under construction. Additional wind farms are also proposed which significantly increase the potential capacity for electricity to be fed into the grid.

To be able to bid in the lucrative peak demand market, the wind farm industry needs to be able to predict, with some certainty, the amount of electricity that they can produce in the next trading cycle. We present here a methodology to model variability in the output from the various South Australian wind farms, both individually and combined, at the half hourly time interval.

This paper extends Boland's (2005) [1] paper in which he showed an application of a methodology, first described by Getz (2003)[3], to determine the overall correlation between wind farms. This gives us an indication of how the wind farms work together so that the volatility effects of the individual turbine and farms can be minimised.

We will show a method to model the wind farm output volatility that contains three parts:

- estimating the deterministic and random parts of each farm's half hourly time series
- calculating the overall correlative coherence between all of the farms
- developing a null distribution for the correlative coherence between all the farms

2 DATA

In this analysis we have used 365 days of wind farm output data in half hourly intervals from the following sites: Canunda, Starfish Hill, Lake Bonney, Wattle Point, Mt. Millar and Cathedral Rocks. This data is given in MWh and our thanks are extended to the Electricity Supply Industry Planning Council of South Australia for providing this data.

3 METHODS

3.1 Distributional Attributes

Initially we want to model the distributional attributes of the output variability. This will give us an idea of the characteristics of the volatility inherent in the system behaves. To do this, we first begin by removing the deterministic component so that the underlying random structure can be estimated.

The first step in this process is to look for cycles within the data that may occur throughout the year (365 days). Using Fourier series analysis we can determine significant cycles within the data. Figure

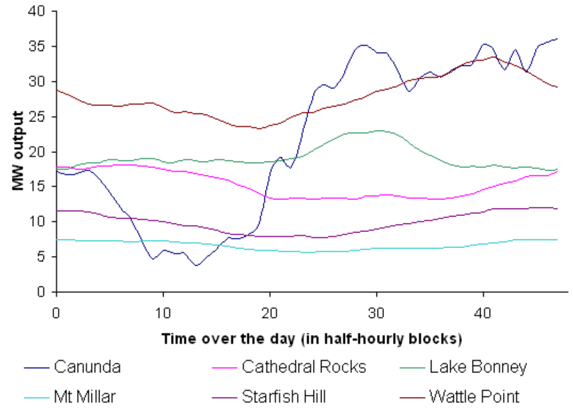

Figure 1. Average Daily Wind Farm Output

1 shows that there is a distinct drop in output followed by a significant peak for each of the sites. So we have found that there is a diurnal pattern in each set of data. The Fourier series for each wind farm's diurnal cycle is given by

$$\begin{aligned} x_t &= y_t + r_t \\ y_t &= \beta_0 + \beta_1 \cos\left(\tfrac{2\pi t}{48}\right) + \beta_2 \sin\left(\tfrac{2\pi t}{48}\right). \end{aligned} \tag{4}$$

We estimate the coefficients (β_0, β_1 and β_2) of the cycles using regression. We then remove the cycles from the data by obtaining the residuals ϵ_t from the regression process.

After we have removed the cycles from the data, we need to find if there are persistent effects evident. To do this we examine the autocorrelation and partial autocorrelation functions to see if there are any lags present in each set of data. This can be done in most statistical analysis packages and from these two functions, we get a good estimate of how many lags (if any) are within the data. We can model the dependence on previous time steps using Auto-regressive Moving-Average (ARMA) modelling.

An ARMA(p, q) model follows the structure shown in Equation 5 [5]. Here, y_t is the observed value of the series at time t, a_t are the residuals and no restrictions are placed on their distribution, x_{it} is the observed value of explanatory variable i at time t, ϕ_i is the coefficient of the auto-regressive component and θ_i is the coefficient of the moving-average component.

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \theta_i \xi_{t-i} + \xi_t \tag{5}$$

where $\xi_t \sim WN(0, \sigma_\xi^2)$ We find that in nearly all of the farms there is an ARMA(1,1) process acting on our deseasoned residuals. We remove the ARMA(1,1) component and calculate the remaining residuals ξ_t. Our data now has the structure shown in Equation 6

$$x_t = \text{Seasonal component} + \text{ARMA component} + \xi_t \tag{6}$$

After the deterministic structure of the data is removed, we estimate the distributional attributes of the random process ξ_t. The residuals for Canunda wind farm are shown in Figure 2.The distribution of the remaining residuals for each farm were found to be symmetric and leptokurtic and thus, not normal. These characteristics suggest that each set (one for

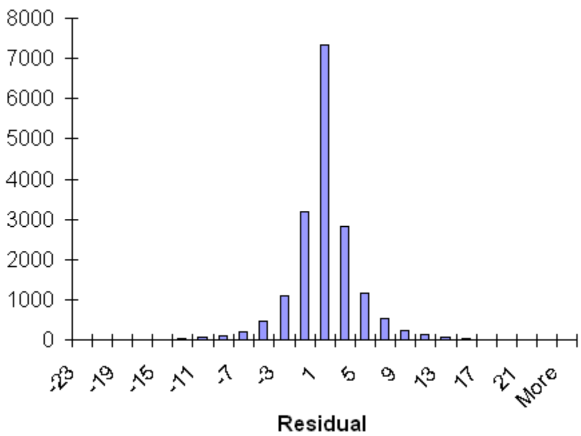

Figure 2. Histogram of the Canunda residuals after seasons and ARMA process is removed.

each farm) of residuals follow a double exponential distribution whose probability density function (pdf) is given by [2]:

$$f(x) = \frac{\exp - \left|\frac{x-\mu}{\beta}\right|}{2\beta}.$$

The parameters for each double exponential distribution were estimated using maximum likelihood estimation where,

$$\begin{aligned} \hat{\mu} &= \mu^* \\ \hat{\beta} &= \frac{\sum_1^n (x_i - \hat{\mu})}{n}. \end{aligned}$$

Here μ^* is the median of the data. Figure 3 shows the goodness-of-fit of the double exponential function for the Canunda wind farm residuals against the original residuals. Both the actual and estimated data can be seen to peak sharply at the same point and then drop away. From this we can see that the peak and spread of the original residuals are captured well by the double exponential distribution.

We thus have a model for the overall output of each wind farm which is given by Equation 7 where $\xi_t \sim$ Double Exponential (μ, β).

$$\begin{aligned} x_t = \beta_0 + \beta_1 \cos\left(\tfrac{2\pi t}{48}\right) + \beta_2 \sin\left(\tfrac{2\pi t}{48}\right) \\ + \phi_0 + \phi_1 r_{t-1} + \theta \xi_{t-1} + \xi_t \end{aligned} \tag{7}$$

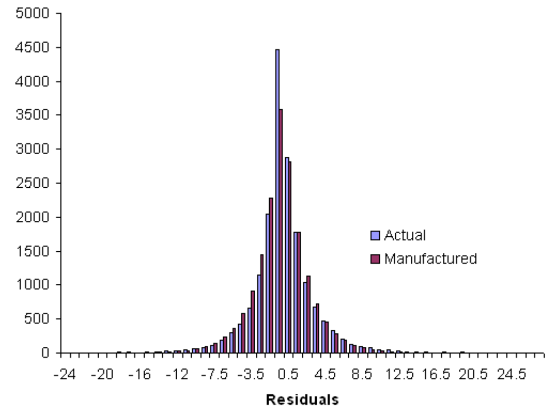

Figure 3. The original residuals for Canunda wind farm and their fitted values

3.2 Correlative Coherence Analysis

In order to smooth out the overall volatility of the wind farms, we need to determine how correlated they all are to each other. We use a method described by Getz [3] and applied by Boland [1] to determine the correlative coherence between the wind farms.

To begin with, we find the Spearman's Correlation coefficient (non-parametric correlation) of each pair of wind farms (so there are 15 distinct comparisons for the six farms). We build a symmetric matrix R using these correlations where $r_{ij} = Cor(X_i, X_j) = r_{ji}$ for $i \neq j$ and $r_{ij} = 1$ for $i = j$.

R is a correlation matrix and its eigenvalues λ_i, $i = 1 \ldots, n$, have special properties, in that $0 \leq \lambda_i \leq n$ and $\sum_{i=1}^n \lambda_i = n$ which we can rewrite as $0 \leq \lambda_i/n \leq 1$ and $\sum_{i=1}^n \lambda_i/n = 1$. Given these properties, Getz [3] presents a well defined measure of the diversity of the eigenvalues λ_i of R. This measure is a generalisation of the concept of coherence for periodic signals [3] and gives us a measure of the degree at which the output from the farms vary in concert with each other.

$$C(X^n) = 1 - \frac{1}{\ln(1/n)} \sum_{i=1}^{n} \left(\frac{\lambda_i}{n}\right) \ln\left(\frac{\lambda_i}{n}\right) \tag{8}$$

If all the off-diagonal correlations have the same value $r \in [0, 1]$ (i.e. the overall correlation between all of the farms) and the diagonal elements remain as 1, then the eigenvalues of the correlation matrix $R(r)$ are $\lambda_1 = 1 + (n-1)r$ and $\lambda_i = 1 - r$ for $i = 2, \ldots, n$ and Equation 8 reduces to

$$C_n(r) = \frac{(1+(n-1)r)\ln(1+(n-1)r)+(n-1)(1-r)\ln(1-r)}{n \ln n} \quad (9)$$

Therefore, if $r = 0$ then $C(X^n) = 0$ and if $r = 1$ then $C(X^n) = 1$. The correlative coherence of any system of farms X^n is therefore the solution to the equation

$$r = C_n^{-1}(C(X^n)) \quad (10)$$

We use Newton's method to solve Equation 10 for r and obtain an overall correlative coherence of $r = 0.209662$ for the six wind farms.

We can also investigate r values to see how the addition of a new wind farm changes the correlative coherence of the system. We can determine if the addition of the farm has significantly increased or decreased the r value.

We next needed to determine which values of r corresponded to various combinations of correlated and uncorrelated farms. To begin with, we only knew that an r value of 0 meant that the farms were completely uncorrelated and that an r value of 1 meant that the farms were completely correlated.

We developed data sets in which some or all of the farms were highly correlated. We then used equations 8,9 and 10 to determine the resulting r values for various combinations. Table 1 shows the r values for farms with different correlation combinations. Here 3,1,1,1 (for example) means 3 highly correlated farms, and 3 uncorrelated (from the first 3 farms, and each other) farms. We can see, that for a set of farms where five were highly correlated and one was highly uncorrelated from all of the others, we get an r value of .717. For a set of farms in which two farms are highly correlated and the rest of the farms are highly uncorrelated, we get an r value of .274, and if we have a set of farms that are all highly uncorrelated we get an r value of .043. We can see that our actual r value of .209662 is quite low, in comparison to these, and there is no evidence of a strong overall link between the farms.

Table 1. r values for different correlated combinations of six wind farms.

Correlation configuration	r value
6	.95
5,1	.830041
4, 2	.765551
4,1,1	.684299
3, 3	.74424
3, 2, 1	.620647
3, 1,1,1	.515792
2, 2, 2	.57879
2, 2 , 1, 1	.465469
2, 1, 1, 1,1	.321067
1, 1, 1, 1, 1, 1	.049996

3.3 Null distribution of r

We have shown that for different pair-wise correlations among a set of six farms, we can obtain very different values of r. We have also shown that the r value for the actual wind farm data is 0.209662. We would now like to estimate the null (i.e. that there is no correlative cohesion between the farms) distribution for r. In this paper we will show an initial attempt at developing this null distribution.

We start by generating a sequence of Uniform data between 0 and 1. We transform this data into a doubly exponential sequence by applying the inverse double exponential function [4](shown in Equation 11)

$$a = \begin{cases} \mu + \beta \ln(2F(a)), & \text{if } F(a) \leq 0.5 \\ \mu - \beta \ln(2(1 - F(a))), & \text{otherwise} \end{cases} \quad (11)$$

to each point in the sequence. Here μ and β are the estimates for one of the farms found using the method in Section 3.1. We repeat this generation of data a number of times (in our case we used 20) for each farm. We now have a large set of synthetic sequences of data, which are not correlated, but have the same distributional attributes as our original farms.

We now take a combination of the synthetic data (one sequence for each farm) and calculate the pairwise Spearman's correlation coefficient for each distinct combination. Equation 12 shows an example of these correlations that we have built into an R matrix.

$$R = \begin{pmatrix} 1 & -0.10 & 0.33 & 0.04 & -0.04 & -0.01 \\ -0.10 & 1 & 0.01 & 0.19 & -0.01 & 0.06 \\ 0.33 & 0.01 & 1 & -0.01 & 0.05 & -0.04 \\ 0.04 & 0.19 & -0.01 & 1 & 0.02 & -0.03 \\ -0.04 & -0.01 & 0.05 & 0.02 & 1 & 0.10 \\ -0.01 & 0.06 & -0.04 & -0.03 & 0.10 & 1 \end{pmatrix} \quad (12)$$

We next apply the method described in Section 3.2 to calculate the correlative cohesion r for this combination of data. By repeating this process of calculating the correlative cohesion with different combinations of the manufactured data, we can build a distribution for the null r value. Figure 4 shows the histogram of 5000 r values calculated in this manner. We can see that the r value of 0.209662 for our actual farm data is a long way outside the 99% region, which shows that the wind farms do work in concert with each other to some extent.

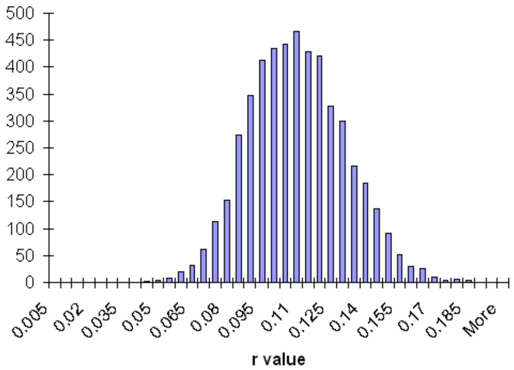

Figure 4. Histogram of 5000 r values calculated from the manufactured data

4 DISCUSSION

In performing this analysis, many interesting features of the data have come to light. The MWh output for each farm shows a diurnal cycle which indicates that the wind speed drops down through the day time and then picks up again in the early evening and night. The peak domestic demand for electricity is at around 6 PM (depending on the month) and the introduction of extra wind-powered electricity into the grid would be advantageous at this time of day.

Although each farm displayed this diurnal cycle, with generally a peak in the afternoon or evening, each farm's average peak was at different times. We believe that this is caused by the local effects at the farm. Wind farms that are close to the coast generally have an afternoon peak in output, whereas those further inland tend to have their peak in the late evening. These local effects are very good for the network stability overall. Having different peak times for each of the farms means that it is rare for all of the farms to peak at once (and overload the physical grid) or to all drop out at once (which leads to other sources of power having to be used).

It is interesting to note that the while the Lake Bonney and Canunda farms are physically very close together, they have quite different average days (see Figure 1). Both of the farms peak at about 2 PM but Lake Bonney's MWh output varies between around 17 MWh and 23 MWh, while Canunda varies between 4 and 36 MWh (see Figure **??**). Lake Bonney is thus the more stable wind farm and is easier to integrate into the grid. This difference is caused by the number of turbines at each farm. Lake Bonney has 46 turbines while Canunda only has 23 turbines. The extra turbines at the Lake Bonney site smooth out the variational effects and give a quite narrow band of possible output. The Canunda farm does not have as many turbines, so the variations in wind speed effect it's output to a greater extent.

The random component of each farm was found to follow a double exponential distribution. Mostly, the output varies by some mean value. Sometimes, however, we get strong peaks or drop outs. This is represented by the high centre peak and thick tails that are present in the double exponential distribution so we do regularly get values that are quite a distance from the centre. This is quite an intuitive distribution for the random part of the wind farm output.

The correlative coherence value r can be used as an indication of the way in which a set of farms follow the same pattern. A low r value is desired in the wind farm network because it is much easier to integrate the output into the grid if the farms are not acting in unison (which tends to produce large spikes in the output). A high r value indicates that the set are highly correlated and so follow the same sort of pattern. If one farm has a surge in output, then the others are also likely to. A high r value has a negative impact on the network stability as the farms will all peak (or drop out) at a roughly similar time. This can cause trouble with the supply and distribution of electricity through the grid. A low r value, however, can lead to quite good stability within the grid, which is highly desirable.

Boland (2005) used the method described in Section 3.2 but with wind data from the Bureau of Meteorology (BoM) sites, rather than the actual output from the wind farms. At the time, most of the wind farms were in the early stages of construction and the output data was unavailable. Boland found a correlative coherence value of $r = 0.742$ for his data which is much higher than the $r = 0.209662$ found for the wind farm output in this paper. The local land features plus the smoothing effects of the multiplicity of turbines means that the analysis using actual output gives a more optimistic view of the contribution that the wind farms produce together.

We show in Section 3.3 a first attempt at a methodology to determine the null value of the correlative cohesion of a set of farms. From Equation 12 we can see that two farms in particular (Farm 1 and Farm 3) are highly correlated. We developed the methodology using each individual farms characteristics to build new 'uncorrelated' data. Unfortunately, due to the proximity of the Canunda and Lake Bonney (unsurprisingly Farm 1 and Farm 3) sites to each other, the random structure in their time series is quite similar. This causes the high correlation and so our results don't show the null distribution that we would like. In our future work we will attempt to overcome this by merging the two sites together.

Adding a new farm into the analysis is a quite straightforward process. The r value can be recalculated and then analysed to see what the effect that the new farm has on the correlative cohesion between all of the farms. This is a very useful tool which we can use in two ways: to predict the effect of a proposed wind farm and to determine the specifications for where a new wind farm should be located.

In the first case, we know where a new wind farm has been constructed. Before it begins operation, we can analyse the effect it will have on the grid stability by calculating the r value in a similar manner to the method used by Boland in his 2005 paper. In this paper, Bureau of Meteorology (BoM) recordings of wind speed were used to estimate the correlative coherence between wind farms. This method of takeing BoM measures can be used to estimate the output for the new farm, and hence calculate the new correlative cohesion r value for the new wind farm.

In the second case, we would choose a desired r value and determine the best conditions for achieving it, given the wind farms that we already have. This could involve specifying the average MW output needed, along with the variability desired. This translates into determining the location (coastal or inland) along with an estimate of the number of turbines needed in the new farm.

5 CONCLUSION

Volatility within wind farm output is currently preventing wind farm operators to compete in peak electricity demand bidding. We have presented here a method to model the volatility within individual wind farms. We have also shown an approach that can be used to determine to what extent that a number of wind farms are correlated overall. The correlative coherence can be used when considering where to build new wind farms and what effect they will have on the existing wind farms.

An extension for this work is to repeat the analysis with five minute data instead of the half-hourly. Electricity is dispatched in five minute intervals so this work will be of particular interest to the wind farm operators. We will work on forecasting the wind farm output at various time intervals with high accuracy so that the wind farms have a chance to compete in the peak demand market.

The connector between the South Australian and Victorian grids means that the interaction between the two is very important. We will investigate the correlation between South Australian and Victorian wind farms to see what sort of effect they have together, on the SA grid. We will also redevelop the null distribution methodology with the farms that show high correlation combined.

6 REFERENCES

Boland, J.W. (2005) *Windfarm Output Variability in South Australia*, Proceedings of the 14th International Conference on Applied Simulation and Modelling, Benalmadena, Spain, June 15-17.

Evans, M., N. Hastings, et al. (2000) *Statistical Distributions*. New York, John Wiley & Sons, Inc., pp 117 - 120

Getz, W.M. (2003) "Correlative coherence analysis: variation from intrinsic and extrinsic sources in competing population", *Theoretical Population Biology*, **64**, pp.89-99

Magnano, L. & J.W. Boland (2006) *Generation of synthetic sequences of electricity demand: Application in South Australis*. Adelaide, The University of South Australia

Tsay, R.S. (2005) *Analysis of Financial Time Series*. Hoboken, N.J., John Wiley & Sons, Inc., pp 56-60