Multifractal analysis of wind farm power output

L. McArthur [a], **S. Mackenzie** [b] **and J. Boland** [c]

[a]*School of Mathematical and Geospatial Sciences, RMIT University, GPO BOX 2476, Melbourne, 3000*
[b]*TTG Transport Technology, Sydney*
[c]*University of South Australia, Adelaide*
Email: lynne.mcarthur@rmit.edu.au

Abstract: Wind generated power is subject to fluctuations due to the intermittent nature of wind which causes unpredictable energy provision. Electrical power is not easily stored so the generation of power ideally would match demand. The amount of energy provided by wind, (and solar power) is determined by natural events which makes matching supply and demand very difficult. Wind speed data has been used in the past to assist with the prediction of power generation, but it is not a satisfactory predictor.

There have been numerous attempts to identify the characteristics of wind speed time series data and it has been found in previous studies that they exhibit long term dependence. Methods of analyzing wind speed and power output time series data have included traditional time series analyses such as ARMA and GARCH as well as the relatively new methods of detrended fluctuation analysis (DFA) and more recently multifractal detrended fluctuation analysis (MFDFA). This study extends these latter methods to examine the variation of output that is inherent in wind generated power.

Almost all chaotic systems can be characterized by a fractal dimension. The fractal dimension of an object is a measurement of a non-integer dimension which represents the degree of persistence, also known as scale or self similarity, embedded in the series.

The history of the Fractal Dimension Index (FDI) begins with the British dam builder and hydrologist H.E. Hurst, in 1951. He worked on the Nile River Dam Project searching for patterns in the Nile delta to solve problems related to the storage capacity of the dam reservoir.

Most hydrologists assumed that water inflow was a random process with no underlying order. After studying almost a millennium of Nile overflows Hurst found that large overflows tend to be followed by large overflows most of the time. There appeared to be cycles, but their lengths were non-periodic. Standard statistical analysis revealed no patterns between observations, but the concept of persistence was raised.

Hurst divided the Nile data into time segments and examined the variance of each segment in comparison with the number of total segments. The process is called re-scaled range analysis. Using this analysis Hurst showed that overflows tended to repeat, meaning that the natural overflows were partially predictable.

Mathematician Benoit Mandelbrot, used the Hurst exponent to experiment with time series found in nature. This led to his development of a method to measure the irregularity in natural objects. He named the measurement the Fractal Dimension Index. The development of fractal geometry in recent times has provided a means of characterizing a range of complex structures using the concept of the fractal dimension. The fractal dimension, which is not necessarily an integer, reflects the scale symmetry of random structures which include time series, and is measured by the Fractal Dimension Index (FDI). Scale symmetry refers to the self similarity patterns observed on different time scales, and a single scaling exponent characterizes the structure. The FDI determines the volatility, which is the unexpected event of extremity, of a given time series. This specialized indicator identifies the Fractal Dimension of the series by using re-scaled range analysis and an estimated Hurst exponent. This has been used by researchers to determine the persistence or anti-persistence of a series.

Previous studies have found that the wind farms in South Australia have some spatial correlation, but that it is not sufficient to be able to aggregate analyses into more than one farm. The farms are thus analysed individually, to assess the degree of persistence in the data for five minute time intervals. The results show that there is evidence of persistence in some of the wind farm time series data under study.

This paper describes the application of the MFDFA technique to a set of wind farms located in South Australia.

Keywords: *Multifractal detrended fluctuation analysis, wind farms, electricity generation, prediction*

1   INTRODUCTION

The objective of this study is to determine whether the power output from wind farms in South Australia illustrate any long-term correlation or self-similarity that hasn't been detected using traditional time series analyses. The application of multifractal detrended fluctuation analysis (MFDFA) to the data, will provide a method of determining if there is any evidence of self similarity or persistence in the series. The Hurst exponent is one of the measures that can provide this information and in turn will inform methods of forecasting. The data is then shuffled and the Hurst exponent is again estimated.

Data has been provided by AEMO, the Australian Energy Market Operator, and consists of 219 days of 5 minute power output data (approximately 63,000 data points) from five separate farms in South Australia. This paper describes the results of the analysis applied to farm 3, for both the raw data and the shuffled data.

2   THEORETICAL BACKGROUND

2.1   Multifractality

The behaviour of complex time series, such as in the financial sector, has been the topic of considerable study. Financial time series have been found to be generally non-linear, stochastic and chaotic, (Weron and Przybylowicz, 2000) and are characterized by Brownian motion, fractional Brownian motion, chaos and fractals, scaling behaviour and self organized criticality, (Weron and Przybylowicz, 2000).

An extension of the FDI analysis is a technique known as the detrended fluctuation analysis (DFA). This has been used to determine the fractal scaling properties of time series and, if they exist, the long-range correlations in noisy, non-stationary data (Kantelhardt et al., 2002). Fields of study include DNA sequences, cloud structure, geology and solid state physics, (Peng et al., 1994, and references therein). In addition, energy spot prices (Resta, 2004), wind speed data, wind generation and demand (Weron and Przybylowicz, 2000; Kavasseri and Nagarajan, 2004; Lei et al., 2007; Norouzzadeh et al., 2007; Erzgraber et al., 2008; Boland et al., 2009) and general nonstationary time series (Kantelhardt et al., 2002) have been analyzed using similar techniques.

Multifractality in time series can be the result of either temporal correlations or fat-tailed probability distributions Norouzzadeh et al. (2007). To measure the contribution from each source two procedures are applied: random shuffling of the data and phase randomization, (Norouzzadeh et al., 2007). Shuffling destroys temporal correlations while preserving distribution and phase randomization preserves the amplitudes of the Fourier transform, but randomizes the Fourier phases. These procedures show that a weaker multifractality will be obtained when determined from the shuffled data and phase randomized data when compared to that obtained from the original time series.

In this section a brief description of the methods adopted by previous researchers will be covered, including, (Peng et al., 1994; Weron and Przybylowicz, 2000; Kantelhardt et al., 2002; Resta, 2004; Norouzzadeh et al., 2007; Erzgraber et al., 2008; Kavasseri and Nagarajan, 2005; Koçak, 2008) and other researchers cited within.

2.2   Hurst Exponent

The Hurst exponent H, is estimated by the slope of the line produced by the log-log plot of the 'rescaled' time series against the size of the blocks, Figure 3. To rescale the series, the time series is divided into $m = 1, ..., n$ blocks of size s and the mean and standard deviation S_m is calculated for each block, Figure 2. The mean is removed from the original time series and the range R_m for each block is then calculated. The process is repeated for many values of s and the Hurst exponent is estimated using the slope of the log-log plot of s versus the mean of the rescaled subseries.

The values of the Hurst exponent, H, range between 0 and 1. A value of 0.5 indicates a true random walk (Brownian motion); a value of $0.5 < H < 1$ indicates long-term positive correlation. This implies that if there is an increase from time step t_{i-1} to t_i then there will probably be an increase from t_i to t_{i+1}. The same will be true of similar decreases. This is also termed 'persistence'. If $0 < H < 0.5$ the time series is displaying anti-persistent behaviour, and is sometimes referred to as 'mean reversion', indicating that the series is always reverting to the mean.

2.3 Detrended Fluctuation Analysis

This is an extension of the above method. Detrended Fluctuation Analysis (DFA) is a technique that provides a means of detecting self-similarity which may be embedded in non-stationary time series. It produces a self-similarity parameter, $h(q)$, which describes the scaling behaviour of the series. The parameter also approximates the Hurst exponent.

The method proceeds as follows: Let B_i be a time series of length k. The integrated time series $y(k)$ is constructed according to:

$$y(k) \quad = \quad \sum_{i=1}^{i=k} [B_i - B_{ave}] \tag{1}$$

where B_i is the ith data point and B_{ave} is the mean of the series.

This integration maps the original time series to a self-similar process. For a given block size n, the size of the fluctuation for the integrated and detrended time series is given by:

$$F(s) = \sqrt{\frac{1}{s} \sum_{i=1}^{i=s} [y(i) - y_s(i)]^2}. \tag{2}$$

This process is repeated over all time scales (block sizes) which then provides a relationship between $F(s)$ and the block size s.

2.4 Multifractal detrended fluctuation analysis

Multifractal detrended fluctuation analysis (MFDFA) is a generalization of DFA where the qth order moments are used rather than only the second moment.

The detrending described above can also be performed using polynomials of order ν. In this case, we obtain

$$F(\nu, s) \quad = \quad \sqrt{\frac{1}{s} \sum_{i=1}^{i=s} \{y[N - (\nu - n_s)s + i] - y_\nu(i)\}^2} \tag{3}$$

$$\Rightarrow F^2(\nu, s) \quad = \quad \frac{1}{s} \sum_{i=1}^{i=s} \{y[N - (\nu - n_s)s + i] - y_\nu(i)\}^2 \tag{4}$$

where equation (4) represents the corresponding variance (Kantelhardt et al., 2002; Norouzzadeh et al., 2007).

The qth order fluctuation function is calculated from averaging over all segments:

$$F_q(s) \quad = \quad \left[\frac{1}{2n_s} \sum_{i=1}^{i=2n_s} [F^2(\nu, s)]^{q/2} \right]^{1/q} \tag{5}$$

where q can take any real value except zero.

Again, the process is repeated over different time scales s, and again the log-log plots of $F_q(s)$ versus s for each q reveal the scaling behaviour. For each q, a different line will be produced, and the slope of the line gives an approximation for the Hurst exponent, $H = h(q)$. If the series contains multifractal characteristics, then $h(q)$ will vary with q. If the original series, $B(i)$ is power-law correlated, indicating self-similarity, then the fluctuation function, equation (5), will vary according to:

$$F_q(s) \sim s^{h(q)} \tag{6}$$

where the exponent $h(q)$ describes the scaling behaviour of the qth order fluctuation function (Norouzzadeh et al., 2007).

Positive and negative values of $h(q)$ describe the scaling behaviour of segments with large and small fluctuations respectively. The exponent $h(q)$ is known as the generalized Hurst exponent. For stationary time series

$h(2)$ is identical to the Hurst exponent and for multifractal series $h(q)$ varies with q, which is considered to be characteristic of multifractal processes.

For positive values of q the exponent $h(q)$ describes the scaling behaviour of segments with large fluctuations, while for negative values of q the exponent refers to the small fluctuation behaviour (Norouzzadeh et al., 2007). A multifractal time series has the property that $h(q)$ varies with q.

2.5 Shuffling the series

Multifractality can be distinguished by two different characteristics in the time series (Kantelhardt et al., 2002): one is due to a broad probability density function and the other is due to long-range correlations. The way to identify the two types is to 'shuffle' the series or randomly order it. If the process of shuffling does not alter the outcome of the multifractal analysis, then it is of the first type. If the shuffling changes the value of the Hurst exponent, then the series is of the second type, since the shuffling has changed the presence of correlations. The wind farm data is shuffled for each farm and the exponents are recalculated. In this second case, the estimated Hurst exponent would be close to 0.5, indicating a random series.

3 CASE STUDY: SOUTH AUSTRALIA

The objective of this study is to determine whether the power output from five wind farms in South Australia illustrate any long-term correlation or self-similarity that hasn't been detected using traditional time series analyses. The application of multifractal detrended fluctuation analysis (MFDFA) to the data, will provide a method of determining if there is any evidence of self similarity or persistence in the series. The Hurst exponent is one of the measures that can provide this information and in turn will inform methods of forecasting. The data is then shuffled and the Hurst exponent is again estimated.

Shuffling the data removes any time dependence and so should show no evidence of correlation.

Data has been provided by AEMO, the Australian Energy Market Operator, and consists of 219 days of 5 minute power output data (approximately 63,000 data points) from separate farms in South Australia. This report concentrates on the output from one farm, Farm 3.

4 RESULTS

This paper details the results for farm 3. The original time series consisted of 63,121 data points of 5 minute sampling over approximately 7 months. Figure 1 illustrates the original time series and Figure 2 shows the first 500 data points divided into $n = 10$ blocks of size $s = 50$.

The Hurst exponent for the time series for farm 3 has been estimated using the MFDFA technique described above. The time scale is five minutes. The algorithm was simulated using MATLAB and six values of q were selected; $q \in \{-6, -2, -1, 1, 2, 6\}$. The results indicate that there is some self-similarity in the data. Figure 3 illustrates the fluctuation function for 6 values of q and 4 shows the plot of $h(q)$ versus q. Table 1 contains the values of $h(q)$ for each q.

Table 1. The estimated Hurst exponent, $h(q)$, for different q values; original data, OR, and shuffled data, SH.

q	-6	-2	-1	1	2	6
Original Series	0.870	0.757	0.728	0.675	0.651	0.563
Shuffled series	0.440	0.431	0.431	0.436	0.440	0.459

As can be seen from the Figures and Tables, there is evidence to suggest that the time series contains multifractality of the type characterized by correlation. These correlations were not apparent from traditional time series analysis conducted previously by Magnano and Boland (2007); Boland et al. (2009); Ward et al. (2009), but our analysis has indicated that there is a degree of self-similarity as evidenced by the values of the estimated Hurst correlation and the exponents of multifractality have been found.

The values obtained by the analysis of the shuffled data indicates that the random data shows no self-similarity or correlation, and this result confirms that there could be some reasonable expectation of using techniques that can capture the correlations in the original series.
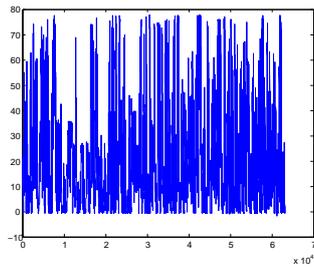
Figure 1. The original time series of power output for farm 3.
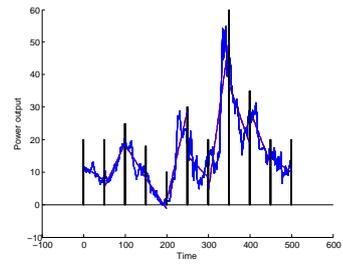
Figure 2. The division of the time series into $n = 10$ blocks of size $s = 50$ and the respective trend line in each block.
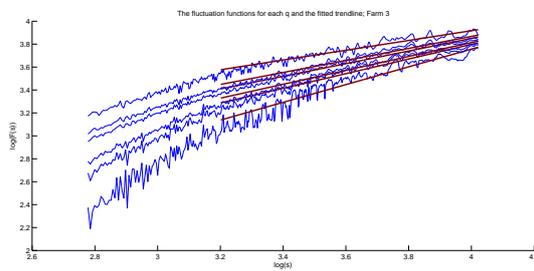
Figure 3. The fluctuation function for farm 3.
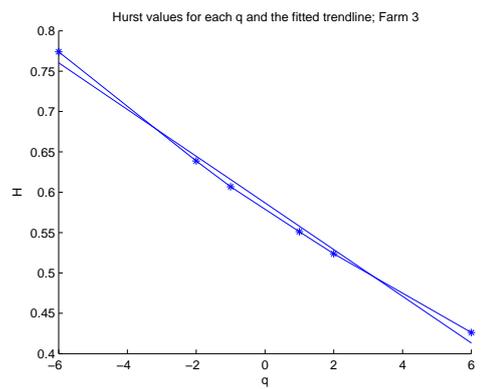
Figure 4. The plot of $h(q)$ versus q for farm 3.

5 DISCUSSION

As illustrated in this paper, the presence of multifractality in a time series can be indicated by determining the value of the Hurst exponent (Hurst, 1951).

The Hurst exponent was estimated for the wind farm output data for time scales ranging from ten seconds to one day, and $AR(p)$ models estimated for the data on these scales as well. This will show the correspondence between the Hurst exponent, a measure of persistence, and the AR(p) models that explicitly show the level of persistence. Table 2 gives the Hurst exponent for the different time scales for the Adelaide wind farm.

When $AR(p)$ models were estimated for the six time scales, similar results were obtained, but this statement requires explanation. The principal stationarity condition for an $AR(p)$ process,

$$X_t = \alpha_1 X_{(t-1)} + \alpha_2 X_{(t-2)} + \ldots + \alpha_p X_{(t-p)} + Z_t \tag{7}$$

is that $\alpha_1 + \alpha_2 + \ldots + \alpha_p < 1$. Note that the closer to unity this sum is, the greater the persistence, and physically this means that if the value at time t deviates from the long term mean of zero, it will take a longer time to revert to the mean. For example, for an $AR(1)$ process, if $\alpha_1 = 0.99$, when the process goes to one side of zero, it will persist there much longer than for an $AR(1)$ process where $\alpha_1 = 0.25$. The results in Table 2 for the sum of the α_is in each case shows the concordance with the Hurst exponent results. The only

Table 2. Estimated Hurst exponents for various time scales at one wind farm.

Time scale	Estimated H	$\sum_{i=1}^{p} \alpha_i$
Ten Second	0.966044	0.9991
Five Minute	0.912679	0.9964
Half Hour	0.822751	0.9875
Hour	0.799138	0.9798
Half Day	0.689918	0.8653
Day	0.659201	0.8507

difference one can see in the two sets of results is that there is a general decrease in the Hurst exponent from the ten second to hourly time scale and then a significant drop, while with the sum of the α_i s, they only drop off slowly until they move from the hourly to half day scale. So, there is the same significant drop from the hourly to half day, but there is a difference at shorter scales.

In summary, there is a consistency in the results of the two approaches to typifying the persistence of the wind farm output at the various time scales. In other words, the Hurst exponent is a valuable tool for identifying the level of persistence in a time series.

6 CONCLUSIONS

The Figures and Tables illustrate the results of the MFDFA analysis for the original and shuffled time series of the wind farm power output. It is clear that the time series is not completely random, and that there is some degree of persistence in the data. That is, there is evidence from this analysis that the wind farm power output has long-term correlation, and that this can be used to better predict future power output, based upon previous power output.

This preliminary study has indicated that persistence exists in the wind power time series at the five minute scale, and decreases as the time scale increases. This is consistent with both the small sample taken for testing and by the evidence of other researchers.

This project has established that further research will contribute to the body of knowledge in the field of both wind power generation prediction. Future research will include analysis of all wind farms in Australia and incorporation of any persistence in the series into predictive models.

REFERENCES

Boland, J., B. Ridley, and M. Agrawal (2009). Modelling the conditional variance of wind farm output utilising realised volatility. In *Proceedings of the ISES Solar World Congress 2009: Renewable Energy Shaping Our Future, Johannesburg, South Africa, 11 - 14 October 2009*, pp. 25–33.

Erzgraber, H., F. Strozzi, J.-M. Zaldivar, H. Touchette, E. Gutierrez, and D. K. Arrowsmith (2008). Time series analysis and long range correlations of Nordic spot electricity market data. *Physica A: Statistical Mechanics and its Applications 387*(26), 6567–6574.

Hurst, H. E. (1951). Long-term storge of reservoirs: an experimental study. . *Transactions of the American society of civil engineers 116*, 770–799.

Kantelhardt, J. W., S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley (2002). Multifractal detrended fluctuation analysis of nonstationalry time series. *Physica A 316*, 87–114.

Kavasseri, R. G. and R. Nagarajan (2004, November). Evidence of crossover phenomena in wind speed data. [Online: Accessed March 17, 2010] URL: http://arxiv.org/abs/physics/0411194,.

Kavasseri, R. G. and R. Nagarajan (2005). A multifractal description of wind speed records. *Chaos, Solitons & Fractals 24*(1), 165–173.

Koçak, K. (2008). Practial ways of eveluating wind speed persistence. *Energy 33*(1), 65–70.

Lei, D., W. Lijei, H. Shi, G. Shuang, and L. Xiaozhong (2007). Prediction of Wind Power Generation based on Chaotic Phase Spce Reconstruction Models. In *Proceediings of the International Conference on Power Electronic Drive Systems, Bangkok, Thailand, November 27-30*, pp. 744–748.

Magnano, L. and J. Boland (2007). Generation of synthetic sequences of electricity demand: Application in south australia. *Energy 32*(11), 2230–2243.

Norouzzadeh, P., W. Dullaert, and B. Rahmani (2007). Anti-correlation and multifractal features of Spain electricity spot market. *Physica A 380*, 333–342.

Peng, C., S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A. L. Goldberger (1994). Mosaic organization of DNA nucleotides. *Physical Review E 49*(2), 1685–1689.

Resta, M. (2004). Multifractal analysis of Power Markets. Some empirical evidence. EconWPA, [Online: Accessed March 17, 2010]. URL: http://ideas.repec.org/p/wpa/wuwpem/0410002.html.

Ward, K., M. Korolkiewicz, and B. J. (2009). Modelling residual wind farm variability using HMMs. In R. Anderssen, R. Braddock, and L. Newham (Eds.), *18th World IMACS MODSIM Congress, Cairns, Australia 13-17 July, 2009*, pp. 4467–4473.

Weron, R. and B. Przybylowicz (2000). Hurst analysis of electricity price dynamics. *Physica A 283*, 462–468.