20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

Probabilistic Forecasting of Wind Farm Output

M. Agrawal [a], **J. Huang** [a] **and J. Boland** [a]

[a] *Barbara Hardy Institute and School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, South Australia, 5095*
Email: manju.agrawal@unisa.edu.au

Abstract: We have previously developed a short time scale forecasting tool for solar radiation [Huang et al., 2013], and also a mechanism for estimating the conditional variance of wind farm output at particular time scales using data at a higher frequency, see [Agrawal et al., 2010, 2013]. The term conditional variance reflects the idea that the variance is changing with time (heteroscedastic) rather than being homogeneous in time (homoscedastic). In this paper, we will describe the application of the solar radiation forecasting tool (which is referred as CARDS model) to wind farm output to obtain forecasts of the level of output on two specific time scales, five minute and half hour. These are the time scales at which the Australian Electricity Market operates. Hence for efficient operation of the electricity grid, it is crucial to have knowledge of forecast of wind energy 5 minute ahead as well as half an hour ahead together with appropriate error bounds. This is exactly the aim of this paper which we achieve using the techniques developed in [Huang et al., 2013] and [Agrawal et al., 2013].

In more explicit terms, knowledge of $\{F_\tau\}_{\tau=t_0}^{t}$, the history of the wind energy output series up to time t allows us to forecast the level of output at time $t+1$, this we achieve using the forecasting tool developed in [Huang et al., 2013]. We then estimate the conditional variance at time t using the techniques developed in [Agrawal et al., 2013]. To facilitate this, we did have high frequency data available at the 10 second time scale. Once we obtained a time series of conditional standard deviation, $\{\sigma_\tau\}_{\tau=t_0}^{t}$, up to the current time step t, we reinvoke CARDS model to obtain a forecast of the conditional standard deviation at time $t+1$, that is, to get $\hat{\sigma}_{t+1}$. Upper and lower bounds of the forecasted wind farm output are thus constructed as $F_{t+1} \pm r\hat{\sigma}_{t+1}$ where r is a positive real number.

An interesting outcome is that 93.5% of the data coverage is contained in the interval $F_{t+1} \pm \hat{\sigma}_{t+1}$ for the 30 minute ahead forecast, while for 5 minute ahead forecast 94.2% of the data coverage is contained in the constructed interval $F_{t+1} \pm r\hat{\sigma}_{t+1}$ with $r = 0.65$. In other words, a lower rate of conditional standard deviation suffices to contain most of the observations at the 5 minute time scale.

This allows us to not only have a forecast of the output but to also put error bounds on that forecast. This type of information is crucial for efficient operation of the electricity grid. This is particularly true in South Australia where wind farms provided 26 % of the electricity generation in the financial year ending June 2012.

Keywords: *CARDS model, high frequency data, conditional variance, forecasting, wind energy*

1 INTRODUCTION

Simply obtaining the level of forecasting is often not adequate, specifically, when a time varying intermittency and the stakeholder's risk are involved. Therefore, we need to know error bounds of the forecast. This type of information is crucial for efficient operation of the electricity grid, in particular, in South Australia where wind farms provided 26 % of the electricity generation in the financial year ending June 2012.

We have previously developed a short time scale forecasting tool for solar radiation [Huang et al., 2013], and also a mechanism for estimating the conditional variance of wind farm output at particular time scales using data at a higher frequency time scale, see [Agrawal et al., 2013]. The term conditional variance reflects the idea that the variance is changing with time (heteroscedastic) rather than being homogeneous in time (homoscedastic). In this paper, we will describe an application of the solar radiation forecasting tool (which is referred as CARDS model) to wind farm output to obtain forecasts of the level of output on two specific time scales, five minute and half hour. These are the time scales at which the Australian Electricity Market operates. Hence for efficient operation of the electricity grid, it is crucial to have knowledge of forecast of wind energy 5 minute ahead as well as half an hour ahead together with appropriate error bounds. This is exactly the aim of this paper which we achieve using the techniques developed in [Huang et al., 2013] and [Agrawal et al., 2013].

2 METHODOLOGY

To obtain the expected value of the wind energy output, that is the forecast of wind energy at time $t+1$ from knowledge of the history of output up to time t, we use the method developed in [Huang et al., 2013]. This model was developed for forecasting solar radiation, and is called the CARDS model[1]. Fundamentally, the CARDS model is based on a combination of an autoregressive (AR) model and a dynamical system model, observing a certain set of rules, which is further enhanced using a predictor-corrector kind of component, called the 'fixed component'. While autoregressive models are widely known in the time series literature, the dynamical system model considered in the method is comparatively less well known, although, it has existed in the literature since as early as 1960's. The model which originated from biophysics (known in the literature as the FitzHugh-Nagumo system, [FitzHugh, 1961]), performed very well for power market modelling, see [Lucheroni, 2007], [Lucheroni, 2009]. The same version as presented in [Lucheroni, 2009] was adapted in [Huang et al., 2013] in the construction process of the CARDS model. For a detailed description and algorithm of the CARDS model, the reader is referred to [Huang et al., 2013].

Let $\{F_t\}$ denote the time series of wind energy output from a windfarm at a time scale of interest (5 minute or half hourly). Given the level of wind energy output up to time t, using the CARDS model, suppose we obtained forecasted energy F_{t+1} at time t.

Our next aim is to construct error bounds of this forecast. To construct the error bounds one has to know the variance at the time step of the forecast. When using time series models for the forecast, an estimate of the variance at any lead is straightforward if the series is stationary in the variance. However, with wind farm output, one observes stochastically changing variance similar to that exhibited by financial market variables, which is referred to as conditional volatility (e.g., see [Poon and Granger, 2003]). In [Agrawal et al., 2013], the authors developed a method of estimating time varying conditional variance using the higher frequency data. We briefly outline the method here.

Suppose the high frequency data time series $\{X_\tau\}$ can be modelled as an autoregressive model, say an $AR(p)$ model: $X_\tau = \alpha_1 X_{\tau-1} + \alpha_2 X_{\tau-2} + \ldots + \alpha_p X_{\tau-p} + Z_\tau$, where $\alpha_1, \alpha_2, \ldots \alpha_p$ are the parameters, and $\{Z_\tau\} \sim WN(0, \sigma^2)$. Let the time scale of interest consist of N_0 high frequency data points. For example, if high frequency data is available at the time scale of 10 seconds and the time scale of interest is 5 minute, then $N_0 = 30$ because there are 30 *ten seconds* in a 5 minute duration. Let $\{Y_t\}$ be the time series at the time scale of interest, so that the two time series are related in the following manner:

$$Y_t \quad = \quad X_t + X_{t-\frac{1}{N_0}} + X_{t-\frac{2}{N_0}} + \ldots + X_{t-\frac{N_0-1}{N_0}} \tag{1}$$

It is proved in [Agrawal et al., 2013], under a reasonable assumption that the residuals $\{Z_\tau\}'s$ remain i.i.d. (independent and identically distributed) within the each time duration $(t-1, t]$, that the variance, $\sigma^2(Y_t)$, at the time step t is given by the following expression:

[1] The acronym CARDS stands for Coupled AutoRegressive and Dynamical System.

$$
\begin{aligned}
\sigma^2(Y_t) \quad = \quad & \left(\sum_{n=0}^{N_0-1} \left(\sum_{i=0}^{n} \psi_i \right)^2 \right) \sigma^2(Z_t) + \\
& \left(\sum_{n=N_0}^{2N_0-1} \left(\sum_{i=0}^{n} \psi_i - \sum_{i=0}^{n-N_0} \psi_i \right)^2 \right) \sigma^2(Z_{t-1}) + \\
& \left(\sum_{n=2N_0}^{3N_0-1} \left(\sum_{i=0}^{n} \psi_i - \sum_{i=0}^{n-N_0} \psi_i \right)^2 \right) \sigma^2(Z_{t-2}) + \ldots, \qquad (2)
\end{aligned}
$$

where $\sigma^2(Z_{t-k})$ is the variance of $Z_{t-k}, \ldots Z_{t-k-\frac{N_0-1}{N_0}}$. One does not need to know each individual ψ_i but the partial sums $\sum_{i=0}^{n} \psi_i$, which appear in equation (2) as a building block, can be expressed entirely in terms of the $AR(p)$ parameters as

$$
\sum_{k=0}^{n} \psi_k = \sum_{k=0}^{n} \sum_{(n_1,\ldots,n_p)\in A_k} \frac{(n_1+n_2+\ldots+n_p)!}{n_1!\,n_2!\ldots n_p!} \alpha_1^{n_1} \alpha_2^{n_2} \ldots \alpha_p^{n_p} \qquad (3)
$$

where $n_1, n_2, \ldots, n_p$ are non-negative integers and $A_k = \{(n_1, n_2, \ldots n_p) \mid n_1 + 2n_2 + \ldots + pn_p = k\}$ for each $k = 0, 1, 2, \ldots n$. See [Agrawal et al., 2013] for more details.

We emphasise that it is a method of nowcasting (not forecasting). Using this method, we construct a time series of conditional standard deviations $\{\sigma_i\}_{i=i_0}^{t}$ up to the current time step t. We then reinvoke the methods developed in [Huang et al., 2013], namely the CARDS model, to get $\hat{\sigma}_{t+1}$, the forecast for the conditional standard deviation at next time step $t+1$. Upper and lower bounds of the forecasted wind farm output are thus constructed as $F_{t+1} \pm r\hat{\sigma}_{t+1}$ where r is a positive real number.

3 Results

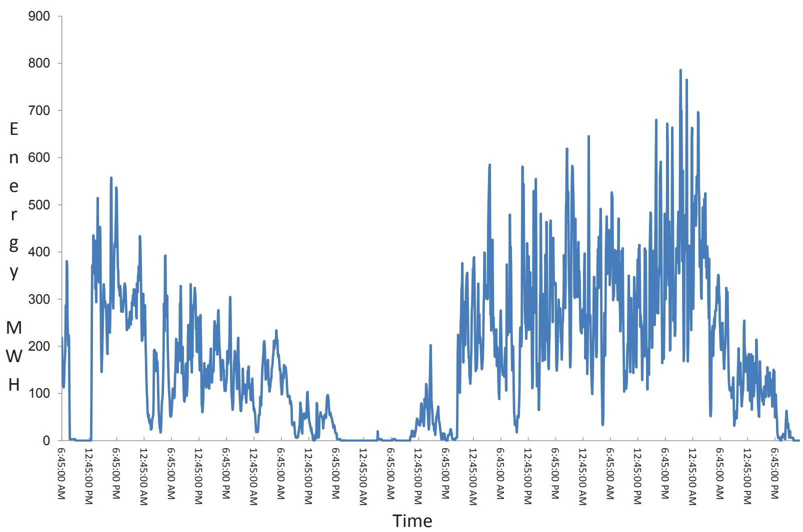

Figure 1: Wind energy output from a South Australian windfarm at 5 minute time scale.

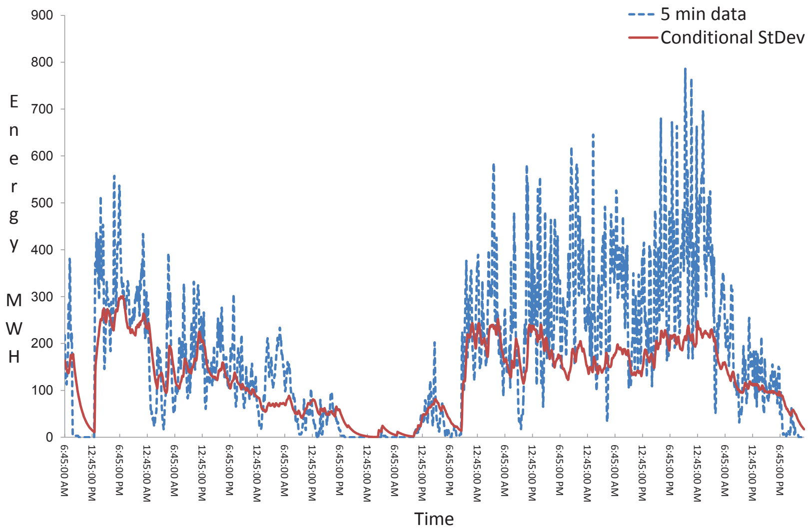

Figure 2: Estimated conditional standard deviation at 5 minute time scale.

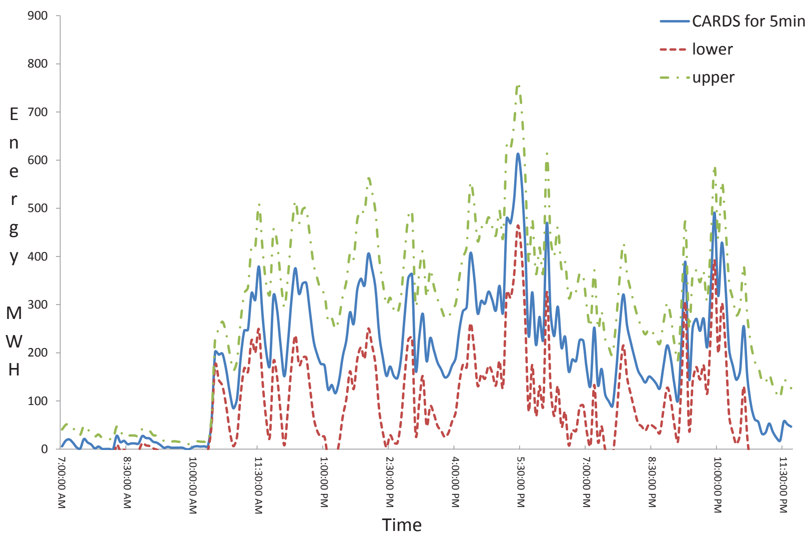

Figure 3: Upper and Lower bounds of 5 minute ahead wind energy forecast.

3.1 Results at 5 minute time scale

Figure 1 shows a snapshot of energy output from a wind farm in South Australia at 5 minute time scale. The two main features of the wind farm output are evident from the Figure 1, a highly volatile nature and intermittency. The conditional standard deviation time series estimated using the higher frequency data (which was available to us at 10 second time scale), is shown in Figure 2. The volatility clustering is clearly present in the wind energy data, that is, the periods of low volatility are followed by the periods of high volatility. The CARDS model is reinvoked on the conditional standard deviation time series to obtain a 5 minute ahead forecast of conditional standard deviation (that is, to get $\hat{\sigma}_{t+1}$). However, as it turned out, at the 5 minute time scale, an $ARMA(2,2)$ model performs as well as does a CARDS model for the conditional standard deviation time series. We rely on the use of an $ARMA(2,2)$ at this stage, for the sake of parsimony.

The 5 minute ahead forecast obtained using the CARDS model, together with upper and lower bounds of the forecast is displayed in Figure 3. It is an interesting finding that 94.2% of the data coverage is contained in the constructed interval $F_{t+1} \pm r\hat{\sigma}_{t+1}$ with $r = 0.65$ for the 5 minute ahead forecast. The average thickness of the forecasted band (that is, average of the difference of upper and lower bound of the forecast) is found to be 173.5, although it should be noted that the bounds are tighter when the level of wind energy output is low.

3.2 Results at 30 minute time scale

The conditional standard deviation time series estimated using the higher frequency data (at 10 second time scale) from the same windfarm, is shown in Figure 4. The CARDS model was invoked on the conditional standard deviation time series to obtain a 30 minute ahead forecast of conditional standard deviation. However, it turned out that at the 30 minute time scale, an $AR(3)$ model performs as well as does a CARDS model for the conditional standard deviation time series. We go with an $ARMA(2,2)$ at this stage, for the sake of parsimony.

The 30 minute ahead forecast obtained using the CARDS model, together with upper and lower bounds of the forecast is displayed in Figure 5. It should be noted that 93.5% of the data coverage is contained in the interval $F_{t+1} \pm \hat{\sigma}_{t+1}$ for the 30 minute ahead forecast. The average thickness of the forecasted band (that is, average of the difference of upper and lower bound of the forecast) is 1514.9, although, the bounds are tighter when the level of wind energy output is low.

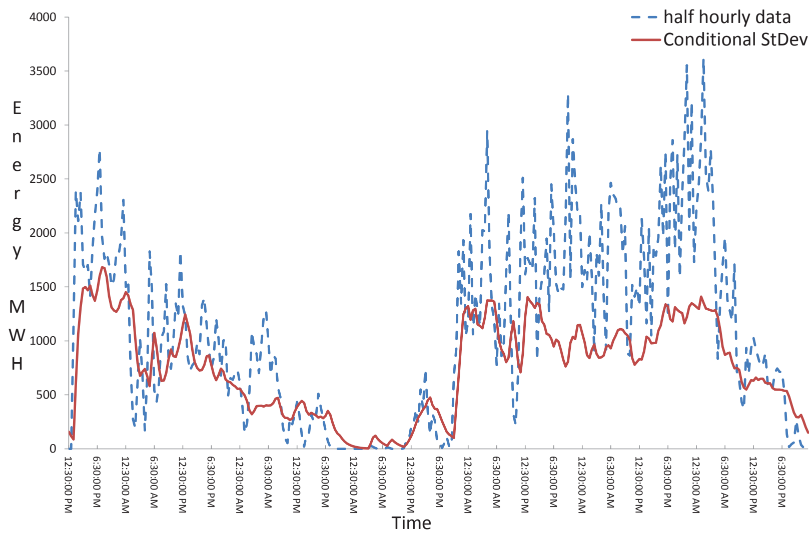

Figure 4: Estimated conditional standard deviation at 30 minute time scale.

4 Conclusions

It is shown that the the CARDS model which was originally created for solar radiation forecasting in [Huang et al., 2013], performs perfectly well for the wind energy forecasting at both the time scales of interest. The

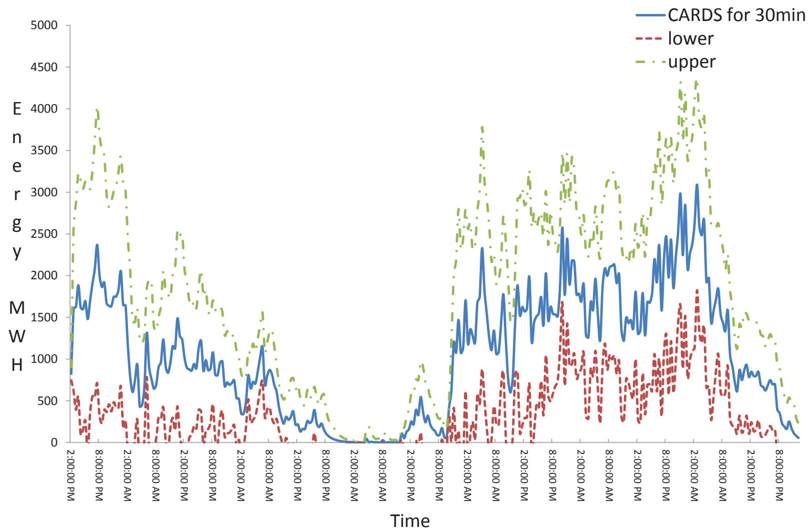

Figure 5: Upper and Lower bounds of 30 minute ahead wind energy forecast.

conditional standard deviation time series at the 5 minute time scale constructed from the high frequency data turn out to be an $AR(2,2)$ process (as a parsimonious model). Consequently, this $AR(2,2)$ model is exploited to forecast the conditional standard deviation 5 minute ahead, which is then utilised to obtain prediction intervals for the 5 minute ahead forecast of wind energy. The prediction intervals are narrower at times when the volatility is low.

In the case of half hourly time scale the forecasting it done using CARDS model, and the conditional standard deviation series is constructed using the higher frequency data at 10 second time scale. The conditional standard deviation series so constructed turns out to be an $AR(3)$ model (as a parsimonious model) which is then used to forecast conditional standard deviation 30 minute ahead. Perhaps, the conditional standard deviation series being an $AR(3)$ process at the half hourly time scale can be perceived as the persistence level of volatility is higher at half hourly scale as compared to that at 5 minute time scale.

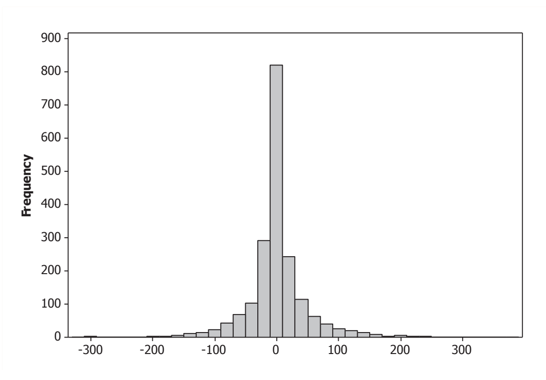

Figure 6: Distribution of the residuals from the CARDS model at 5 minute time scale.

Remarkably, almost 94% of the data coverage is contained in the constructed interval $F_{t+1} \pm r\hat{\sigma}_{t+1}$ with $r = 0.65$ for the 5 minute ahead forecast, while with the half hourly timescale the same percentage of the observed data falls within the constructed interval $F_{t+1} \pm r\hat{\sigma}_{t+1}$ with $r = 1$. Arguably, a lower rate of conditional standard deviation suffices to contain most of the observations at the 5 minute time scale. This can be explained through the distribution of the errors from the CARDS model, see Figure 6 and Figure 7. As the

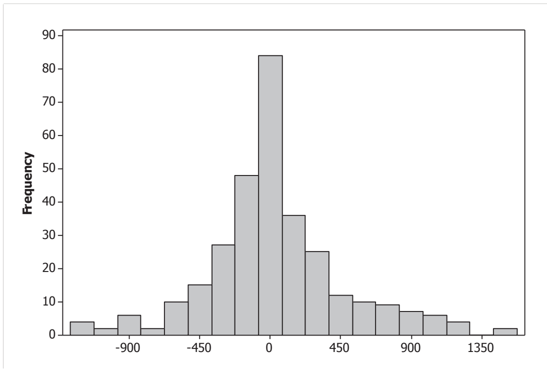

Figure 7: Distribution of the residuals from the CARDS model at 30 minute time scale.

peak is sharper and more concentrated around the center for the 5 minute time scale in comparison to that for the half hourly case, a lower rate of conditional standard deviation suffices to provide the same percentage of coverage.

Acknowledgments

Authors would like to acknowledge ASI (Australian Solar Institute) grant number SKS:PGM:26-6328791 to undertake Project 2-A004 "*Forecasting and Characterising Grid Connected Solar Energy and Developing Synergies with Wind"*.

References

Agrawal, M., J. Boland, and B. Ridley (2010). Volatility of wind energy using high frequency data. In M. Hamza (Ed.), Proceedings of IASTED International Conference: Modelling, Identification and Control (AsiaMIC 2010), November 24-26, 2010, Phuket, Thailand*, pp. 1–3.*

Agrawal, M., J. Boland, and B. Ridley (2013). Analysis of wind farm output: Estimation of volatility using high-frequency data. Environmental Modeling and Assessment 18*, 481–492.*

FitzHugh, R. (1961). Impulses and physiological states in models of nervemembrane. Biophys Journal 1*, 445–466.*

Huang, J., M. Korolkiewicz, M. Agrawal, and J. Boland (2013). Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (cards) model. Solar Energy 87*, 136–149.*

Lucheroni, C. (2007). Resonating models for the electric power market. Physical Review E 76*(5), 9835.*

Lucheroni, C. (2009). A resonating model for the power market and its calibration, EEM 2009 6th International conference on the European, May 2009.

Poon, S. and C. W. J. Granger (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature 41*(2), 478–539.*