Estimating the Volatility of Wind Farm Output

John Boland,
Barbara Hardy Institute
School of Mathematics and Statistics
University of South Australia

23th September 2011

Acknowledgment

The work is supported by **Australian Research Council (ARC)** and **Australian Electricity Market Operators (AEMO)**

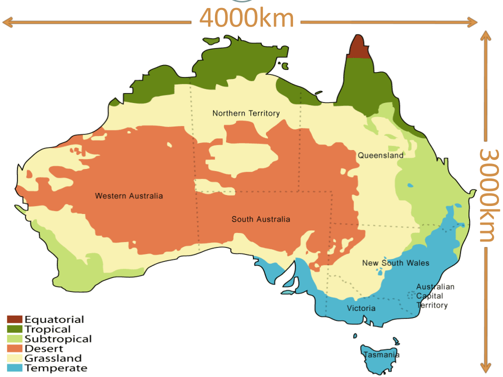
through the projects

- Discovery Grant - Strategic integration of renewable energy systems into the electricity grid.
- Linkage Grant - Unlocking the Grid: the future of the electricity distribution network.

UniSA

Context

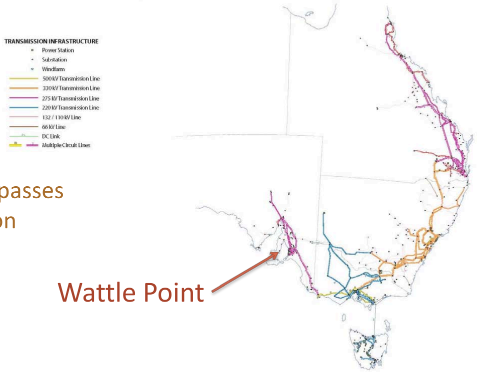
Australia



Context

The Australian National Electricity Market

Networks in Australia's National Electricity Market



This network encompasses 80% of the population

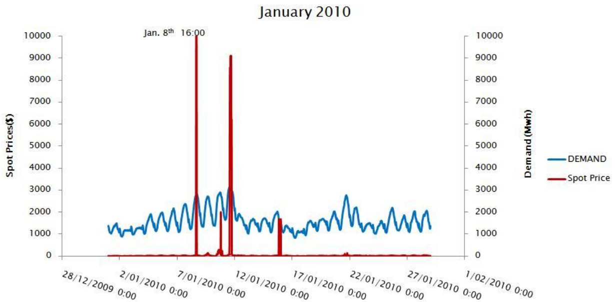
Wattle Point

Operation of the Electricity Market

- The market operates on a five minute basis and is cleared every half hour.
- Every five minutes, generators submit a bid stack detailing their offers of the volumes at each of ten price bands.
- The Australian Energy Market Operator (AEMO) runs a linear program to decide on how much energy to take from each generator.
- The marginal price for that five minute period is determined.
- At the end of the half hour, the spot price is set as the average of the five minute prices.
- All energy dispatched to the system during the half hour is at that spot price.

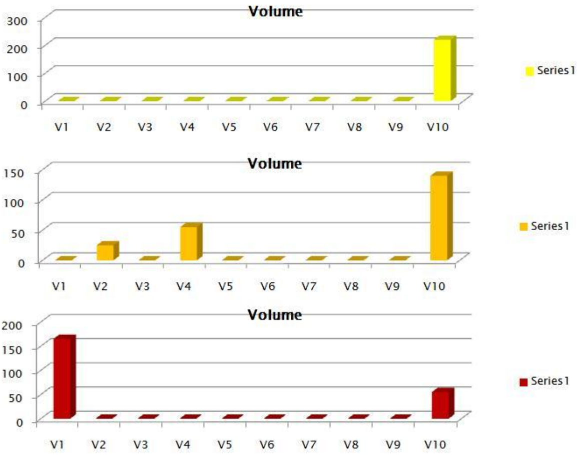
UniSA

Spot Price and Demand



UniSA

Bid Stacks



UniSA

Context

Wind energy in South Australia

UniSA

Motivation

- Reduce dependence on conventional centralised generation to a situation where more diversified, more volatile and less controllable generation sources contribute a significant percentage of the energy.
- Develop new techniques for modelling the volatility so that the output from these wind farms can be reliably estimated in order to enter fully into the competitive electricity market.
- Generation of synthetic data to be used as input into an optimisation model being constructed by others members of our team for the purpose of designing the future grid architecture to ensure the security of supply.

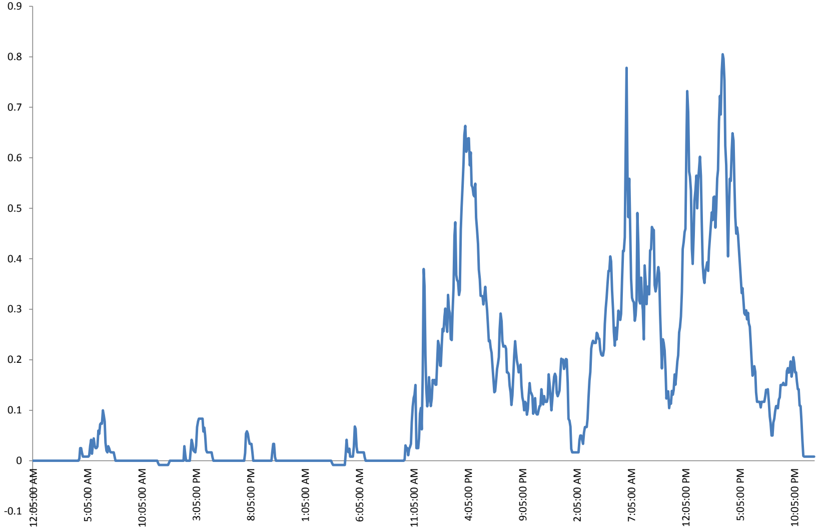
Modelling wind farm output

- We are focussing on two time scales of wind farm operation, 5 minute and half hour. These are the two most relevant time scales for the electricity market.
- We use both classical and modern time series analysis methods, the so-called modern being adapted from financial time series and dynamical systems.

Processing the data

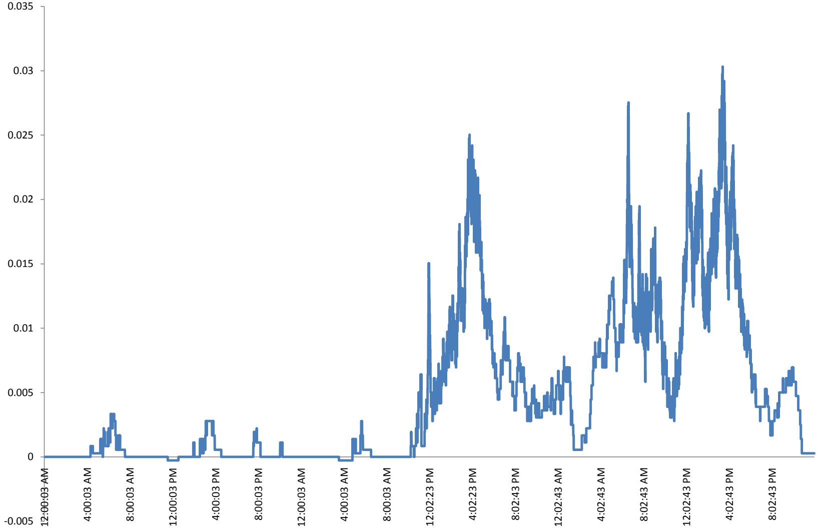
- 10 second data for 2 locations is used
- 5 and 30 minute data sets are formed by aggregating from 10 second data
- Processing of the data includes
  - Identifying and removing any seasonality
  - Forecasting the output level
  - Forecasting the volatility

3 days of wind energy output at 5 minute intervals



UniSA

3 days of wind energy output at 10 second intervals

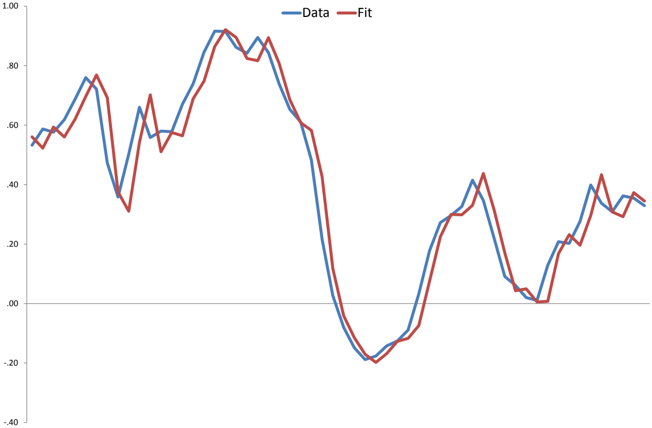


UniSA

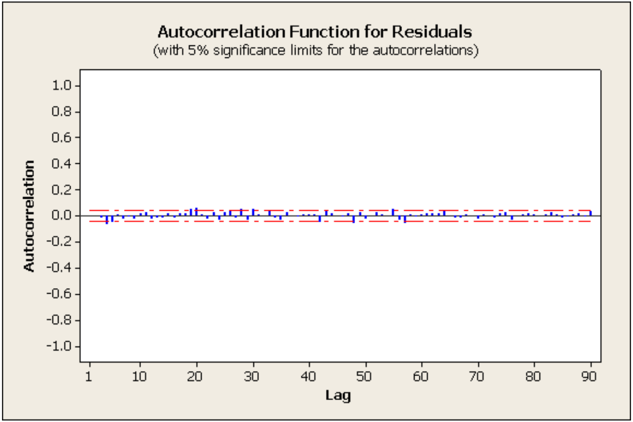
Time series modelling of 5 minute data

- Identification of underlying process relies on investigation of autocorrelation function (ACF) and partial autocorrelation function (PACF).
- The sample ACF is a measure of the linear relationship between time series observations separated by some time period, denoted the lag k.
- If X_t is correlated with X_{t-1}, and X_{t-1} is correlated with X_{t-2}, and $\ldots$, X_{t-k+1} is correlated with X_{t-k}, it will seem like X_t is correlated with X_{t-k}.
- The sample PACF sorts out this interaction arising through a transitive action.

Data and model fit

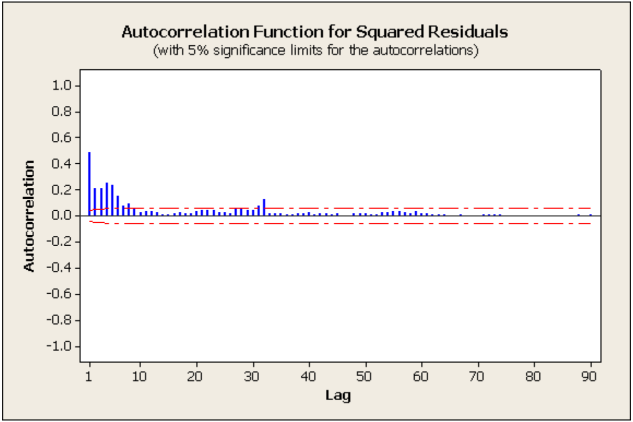


ACF of Residuals



UniSA

ACF of Squared Residuals



ARCH and GARCH

- The noise is uncorrelated but dependent. This phenomenon is prevalent in financial markets - it called volatility clustering. Periods of high volatility are followed by periods of low volatility.
- Engle developed the autoregressive conditional heteroscedastic (ARCH) model to cater for this. The figure above indicates that the model will have to have a long lag AR structure.
- For this lack of parsimony and other reasons, Bollerslev developed the generalised ARCH or GARCH model, where we replace the long lag AR model with a short lag ARMA model.
- Often, an ARMA(1,1) for the residuals squared is sufficient and the GARCH model is derived from that
$\sigma_t^2 = 0.006 + 0.122a_{t-1}^2 + 0.821\sigma_{t-1}^2$

UniSA

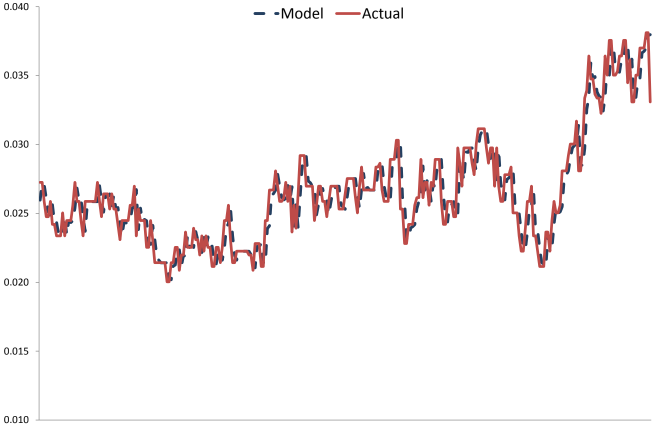
Hidden Markov Model

- We also applied an HMM to modelling the variance.
- It gave more physically interesting results in that for every farm, only two states were required.

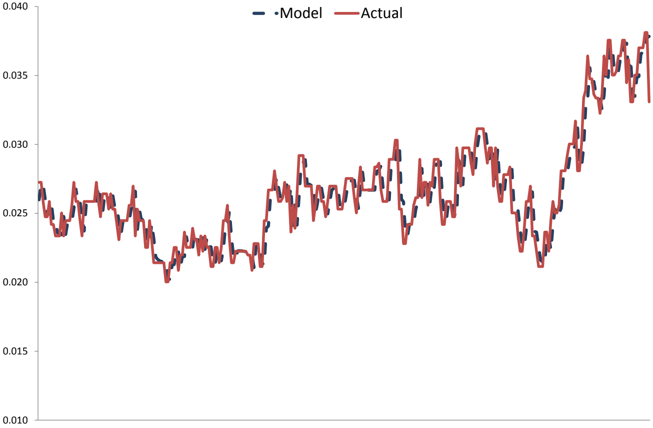
Alternate Formulation

- It is crucial to obtain accurate estimates of the volatility and prediction of the wind farms' output so that the wind energy can enter fully and reliably into the competitive electricity market.
- Aim is to estimate the volatility at 5 minute time scale.
- We do have high frequency data available (at every 10 second)
- The 10 sec data follow an AR(p) process [indeed, an $AR(8)$ process, but if we go for a simpler model, it could be taken as an $AR(3)$].

$AR(3)$ - fit of 10 sec data



$AR(8)$ - fit of 10 sec data



Volatility (unobservable)

- We develop a method to estimate volatility when high frequency data follow an $AR(p)$ process.
- Many researchers have made use of high frequency data to estimate the volatility. Their approach involved computation of covariance etc.
- Our approach is different, as we use model of high frequency data to estimate the volatility.
- We will describe how to use ten second wind farm output to estimate the volatility on a five minute time scale.

10 sec data (X_t), an $AR(3)$ process

- 10 sec data (X_t) is an $AR(3)$ process:

  $$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} + Z_t$$

  Or equivalently,

  $$\phi(B) X_t = Z_t$$

  where $\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \alpha_3 B^3$, and B denotes the backshift operator, that is, $BX_t = X_{t-1}$.

- As $\phi(B)$ is invertible, the process is equivalent to an infinite moving average process.

  $$X_t = \psi(B) Z_t$$

  where $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots$

- Thus, in simple terms, we got

  $$X_t = \psi_0 Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \psi_3 Z_{t-3} + \dots,$$

UniSA

10 sec data (X_t), an $AR(3)$ process

- It can be shown that

$$\psi_j = \alpha_1 \psi_{j-1} + \alpha_2 \psi_{j-2} + \alpha_3 \psi_{j-3} \quad (1)$$

with $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$.

- We develop an expression for ψ_j, $j > 0$ in subsequent slides.

UniSA

Summing variables that follow an AR(3) process to get a 5 minute variable

Let $\{X_t\}$ denote the time series of wind energy output at intervals of every 10 seconds, and let $\{Y_t\}$ denote the time series of aggregated energy output at every 5 minute. The five minute process as a sum of thirty 'ten second observations' can be expressed as

$$Y_t \quad = \quad X_t + X_{t-\frac{1}{30}} + X_{t-\frac{2}{30}} + \ldots + X_{t-\frac{29}{30}} \tag{2}$$

It is understood throughout that $X_{t-\frac{i}{30}}$ represents the wind energy output at the i^{th} 10 second prior to time t, so that $t-1$ remains the consistent notation for five minutes to t.

Summing variables that follow an AR(3) process to get a 5 minute variable

$$
\begin{aligned}
Y_t \quad = \quad & \psi_0 Z_t + (\psi_0 + \psi_1) Z_{t-\frac{1}{30}} + (\psi_0 + \psi_1 + \psi_2) Z_{t-\frac{2}{30}} \\
& + (\psi_0 + \psi_1 + \psi_2 + \psi_3) Z_{t-\frac{3}{30}} + \ldots \\
& + (\psi_0 + \psi_1 + \ldots + \psi_{29}) Z_{t-\frac{29}{30}} \\
& + (\psi_1 + \psi_2 + \psi_3 + \ldots + \psi_{30}) \color{red}{Z_{t-1}} \\
& + (\psi_2 + \psi_3 + \psi_4 + \ldots + \psi_{31}) Z_{t-\frac{31}{30}} + \ldots \\
& + (\psi_{29} + \psi_{31} + \ldots + \psi_{58}) Z_{t-\frac{59}{30}} \\
& + (\psi_{30} + \psi_{32} + \ldots + \psi_{59}) \color{red}{Z_{t-2}} + \ldots \qquad (3)
\end{aligned}
$$

Note that in (3), up to 30^{th} term the coefficients have different form than those after 30^{th} term.

$\sigma^2(Y_t)$ in terms of ψ_i's

We will assume that within each 5 minute interval, the Z_t's are i.i.d. with zero mean. The variance of Y_t is thus

$$
\begin{aligned}
\sigma^2(Y_t) \quad = \quad & [\psi_0^2 + (\psi_0 + \psi_1)^2 + (\psi_0 + \psi_1 + \psi_2)^2 \\
& +(\psi_0 + \psi_1 + \psi_2 + \psi_3)^2 + \ldots \\
& +(\psi_0 + \psi_1 + \psi_2 + \ldots + \psi_{29})^2]\sigma^2(Z_t) \\
& +[(\psi_1 + \psi_2 + \psi_3 + \ldots + \psi_{30})^2 \\
& +(\psi_2 + \psi_3 + \psi_4 + \ldots + \psi_{31})^2 + \ldots \\
& +(\psi_{30} + \psi_{31} + \ldots + \psi_{59})^2]\sigma^2(Z_{t-1}) + \ldots
\end{aligned} \tag{4}
$$

$\sigma^2(Y_t)$ in terms of ψ_i's

$$\begin{aligned}
\sigma^2(Y_t) \quad = \quad & \left(\sum_{n=0}^{29} \left(\sum_{i=0}^{n} \psi_i \right)^2 \right) \sigma^2(Z_t) + \\
& \left(\sum_{n=30}^{59} \left(\sum_{i=0}^{n} \psi_i - \sum_{i=0}^{n-30} \psi_i \right)^2 \right) \sigma^2(Z_{t-1}) + \\
& \left(\sum_{n=60}^{89} \left(\sum_{i=0}^{n} \psi_i - \sum_{i=0}^{n-30} \psi_i \right)^2 \right) \sigma^2(Z_{t-2}) + \dots
\end{aligned} \tag{5}$$

UniSA

The components $\sum_{i=0}^{n} \psi_i$

The basic components in the expression for $\sigma^2(Y_t)$ are $\sum_{i=0}^{n} \psi_i$.

We are able to prove that

$$\sum_{i=0}^{n} \psi_i \;=\; \sum_{k=0}^{n} \sum_{(n_1,n_2,n_3)\in A} \frac{(n_1+n_2+n_3)!}{n_1!\,n_2!\,n_3!} \alpha_1^{n_1}\alpha_2^{n_2}\alpha_3^{n_3}$$

where the triplets (n_1, n_2, n_3) come from the set

$A = \{(n_1, n_2, n_3) \,|\, n_1 + n_2 + n_3 \leq k \;\&\; n_1 + 2n_2 + 3n_3 = k\}$

UniSA

Results: For any $AR(p)$

- Equation(5) is valid for the situation when the high frequency time series (X_t) follow an $AR(p)$ process. But REMEMBER, ψ_i's would be described differently for different p.
- It can be proved that

$$\psi_j = \alpha_1\psi_{j-1} + \alpha_2\psi_{j-2} + \ldots + \alpha_p\psi_{j-p} \tag{6}$$

with $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$.

The components $\sum_{i=0}^{n} \psi_i$: For any $AR(p)$

For $AR(p)$ case, we prove that

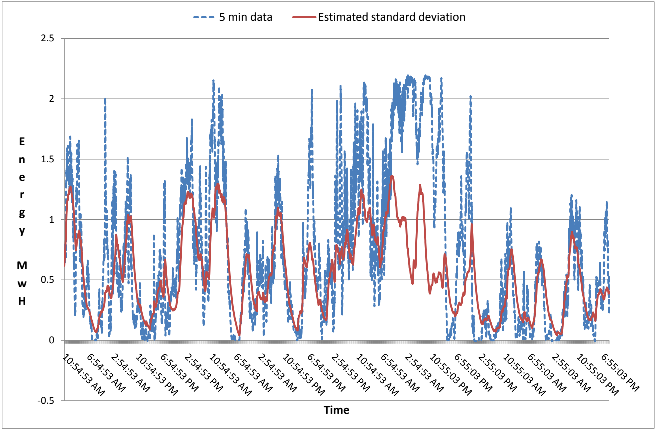
$$\sum_{i=0}^{n} \psi_i = \sum_{k=0}^{n} \sum_{(n_1,n_2,\ldots n_p)\in A} \frac{(n_1+n_2+\ldots+n_p)!}{n_1!\,n_2!\,\ldots\,n_p!} \alpha_1^{n_1}\alpha_2^{n_2}\ldots\alpha_p^{n_p}$$

where $A = \{(n_1, n_2, \ldots, n_p) \,|\, n_1 + n_2 + \ldots + n_p \leq k$ &
$n_1 + 2n_2 + \ldots + pn_p = k\}$

How do we apply these results?

- The prime assumption is that over a five minute interval, the noise Z_t in the ten second model is sufficiently close to being i.i.d.
- With this assumption, we use the 30 ten second Z_t values (the residuals after fitting the model) to calculate the variances $\sigma^2(Z_t)$, $\sigma^2(Z_{t-1})$ ... to use in the calculations in Eqn (5).
- We calculate a separate variance for each five minute interval.

Volatility Estimates



Further Work

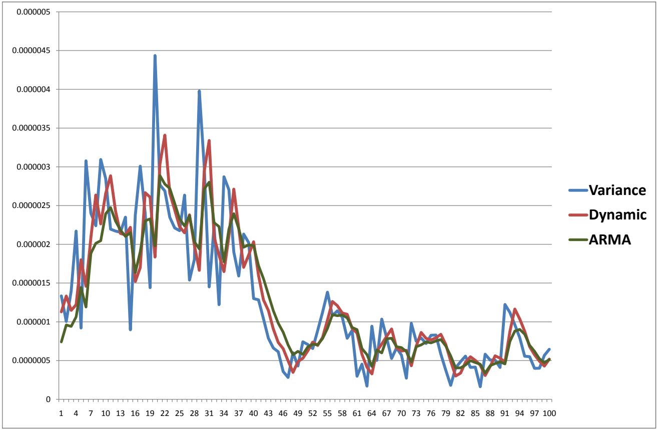
- Similar arguments will work if high frequency data follow an $ARMA(p, q)$ process. However, the expressions for the $\sigma^2(Y_t)$ (Eqn (5)) will no longer be the same.
- It would be good to explore the distribution of the aggregated variable.
- The theoretical results are as well applicable to similar other situations, e.g., in financial time series.
- We are presently comparing the results from this formulation with more empirically based studies in the literature.
- Modelling the realised volatility series. We use a resonating model from a paper entitled **A resonating model for the power market and its calibration** - Lucheroni.

Resonating Model

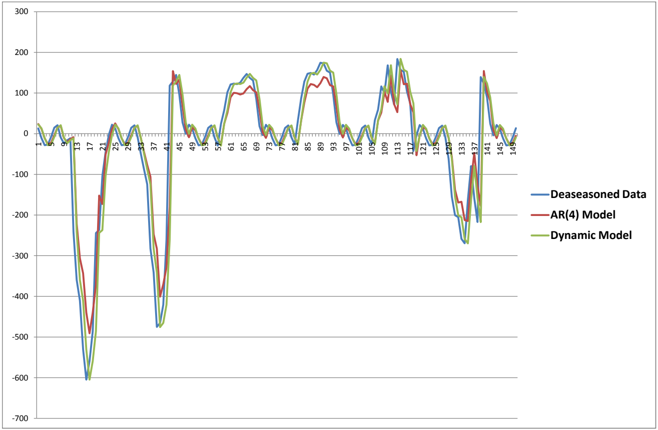
$$
\begin{aligned}
x_{i+1} &= x_i + z_i \delta t \\
z_{i+1} &= z_i + [\kappa(z_i + x_i) - \lambda(3x_i^2 z_i + x_i^3) \\
&\quad - \epsilon z_i - \gamma x_i - b + f(t_i)]\frac{\Delta t}{\epsilon} \\
&\quad - \sigma(d)\sqrt{\Delta t}\psi_i
\end{aligned}
$$

UniSA

Comparison of Model versus a GARCH Model



Use of the Resonating Model with Deseasoned Hourly Solar Data

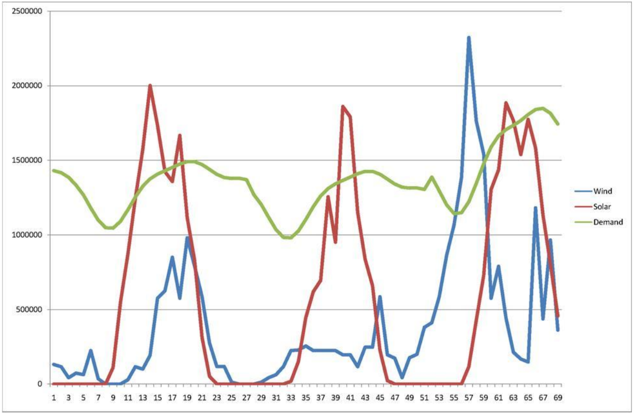


Augmentation of the Electricity Grid

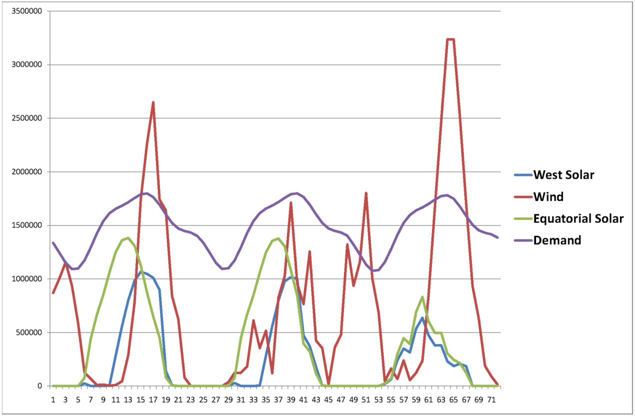
$$
(P)\begin{cases}
\textit{minimize } \{\sum_{(i,j)\in\Omega} c_{ij} n_{ij}\} & \\
s.t. & \\
Sf + g = d & \text{Kirchoff I,power balance} \\
f_{ij} - \gamma_{ij}(n_{ij}^0 + n_{ij})(\theta_i - \theta_j) = 0 & \text{Kirchoff II,voltage balance} \\
|f_{ij}| \leq (n_{ij}^0 + n_{ij})\bar{f}_{ij} & \text{line limits in link}(i,j) \\
0 \leq g \leq \bar{g} & \text{generation limit at every node} \\
0 \leq n_{ij} \leq \bar{n}_{ij} & \text{link expansion constraint} \\
n_{ij} & \text{integer variables} \\
f_{ij}, \theta_j & \text{free continuous variables} \\
(i,j) \in \Omega & \text{set of all possible links}
\end{cases} \tag{7}
$$

UniSA

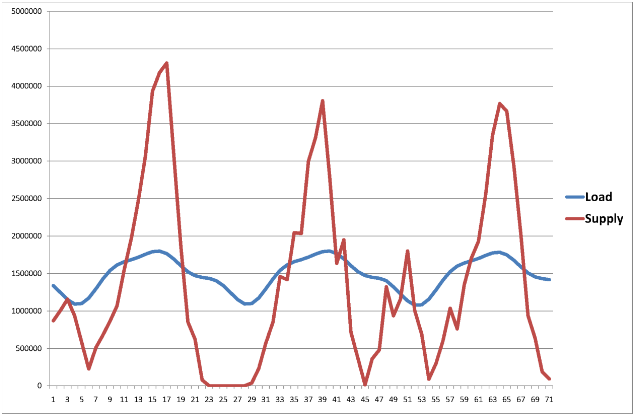
Storage - Renewable Supply and Demand



Storage - Mix your solar



Storage - Add the Supply



UniSA