AMSI VACATION RESEARCH SCHOLARSHIPS 2018-2019

A Markov-Chain-Based Investigation into Renewable Energy Storage in South Australia

Scott Carnie-Bronca

Supervised by Dr Giang Nguyen

The University of Adelaide

Vacation Research Scholarships are funded jointly by the Department of Education and Training and the Australian Mathematical Sciences Institute.

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Australian Government
Department of Education and Training

Abstract

Energy storage is an important aspect of maintaining a reliable power grid sourced by renewable power generation. We aim to investigate the quantity of renewable energy generation and storage required to support the South Australian power grid.

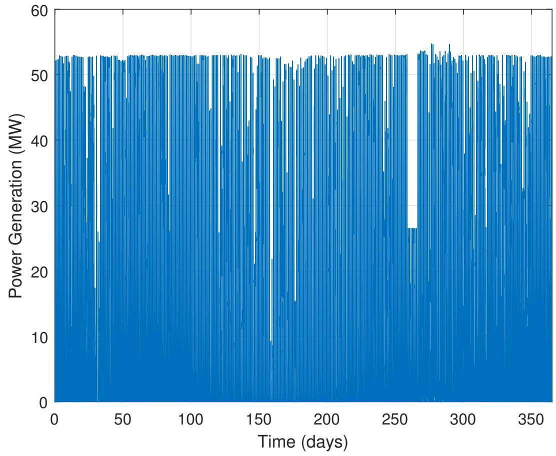
1 Introduction

Renewable energy sources are highly desirable for future power generation due to growing concerns regarding excessive pollution from fossil fuel-based power generation. However, they often suffer from a lack of reliability, for example, unpredictable long-term weather patterns can severely limit the production of photovoltaic cells. Due to this, energy storage is an important aspect of maintaining a reliable power grid sourced by renewable power generation. Australia, and in particular South Australia, is well-suited to maintaining a renewable power system due to the abundance of solar radiation [2, 3].

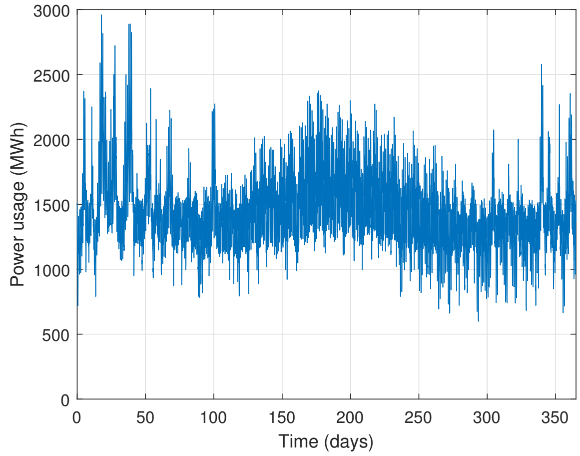
Several large-scale electricity storage technologies exist and are in current use globally, including mechanical storage such as flywheels and pumped hydroelectric power storage, and chemical methods such as batteries and hydrogen storage, with a majority of South Australian storage in battery systems. This includes the Tesla battery installed in 2017 near Hornsdale, SA, with a capacity of 129MWh [1]. However, each type of storage system has several limitations, for example, the high expense of battery storage, and the environmental concerns regarding pumped hydroelectric power storage [2]. In this project, solar/photovoltaic power generation and lithium-ion battery storage are considered.

Section 2.1 describes the collection of data on both generation from the Broken Hill solar farm, located in New South Wales near the South Australian border, and South Australian electricity usage. A brief analysis of the data is also presented. Filtering of the data is performed using Fourier analysis in Section 2.2 to separate the long-term deterministic components of the usage and generation from the random elements. The short-term fluctuations are then modelled using time series models in Section 2.3, including separate models for nighttime, defined as the period of time between 11pm and 4am each day, and whether the day is a weekday or a weekend. In Section 2.4, we construct a Markov Chain model from the time series models, to simulate numerous combinations of battery storage capacity and amounts of generation, as discussed in Section 2.5.

I, Scott Carnie-Bronca, declare that this report titled "A Markov-Chain-Based Investigation into Renewable Energy Storage in South Australia", and the work presented in it are my own.



(a) Electricity generation from the Broken Hill solar farm, with a maximum generation capacity of 52MW, over 2018.



(b) South Australian electricity usage over 2018.

Figure 1: Plot of electricity usage and generation data.

2 Method

2.1 Data Collection

In order to develop and apply the model, we obtained data from the Australian Energy Market Operator (AEMO) for 2018, for both energy generation across the National Energy Market and electricity usage for South Australia [6, 7]. The generation data was then filtered to select the Broken Hill solar farm, with a maximum capacity of 52MW, for model development. This solar farm was chosen as a full data set for a South Australian solar farm was unavailable, and Broken Hill exhibits similar climate conditions to outback South Australia. As the generation was measured in MW (megawatts) every 5 minutes, it was converted to MWh (megawatt-hours) over 30 minute intervals, the same units and time interval as the usage data, by averaging the generation over six 5-minute intervals and dividing by 2. Plots of the generation and usage are provided in Figure 1.

As shown in Figure 1a, there is a short period of time at about day 265, corresponding to late September, with a lower generation amount. This is likely due to maintenance, and so we removed it from the data set.

2.2 Long-Term Trend Removal

Both the generation and usage show strong yearly, seasonal and daily cycles. To remove these from the stochastic components, a form of Fourier filtering was applied. Let the generation data set be $\{y(t)\}$, for $t = 1, 2, ..., 17520$. Then, $y(t)$ can be written as

$$y(t) = \mu + \sum_{j=1}^{\frac{n}{2}-1} \left(a_j \cos\left(\frac{2\pi j}{n} t\right) + b_j \sin\left(\frac{2\pi j}{n} t\right)\right) + a_{\frac{n}{2}} \cos(\pi t),$$

where n is the number of data points, in this case 17520, μ is the mean of the data set, and a_j, b_j are estimated by

$$\widehat{a}_j = \frac{2}{n} \sum_{t=1}^{n} y(t) \cos\left(\frac{2\pi j}{n} t\right),$$
$$\widehat{b}_j = \frac{2}{n} \sum_{t=1}^{n} y(t) \sin\left(\frac{2\pi j}{n} t\right).$$

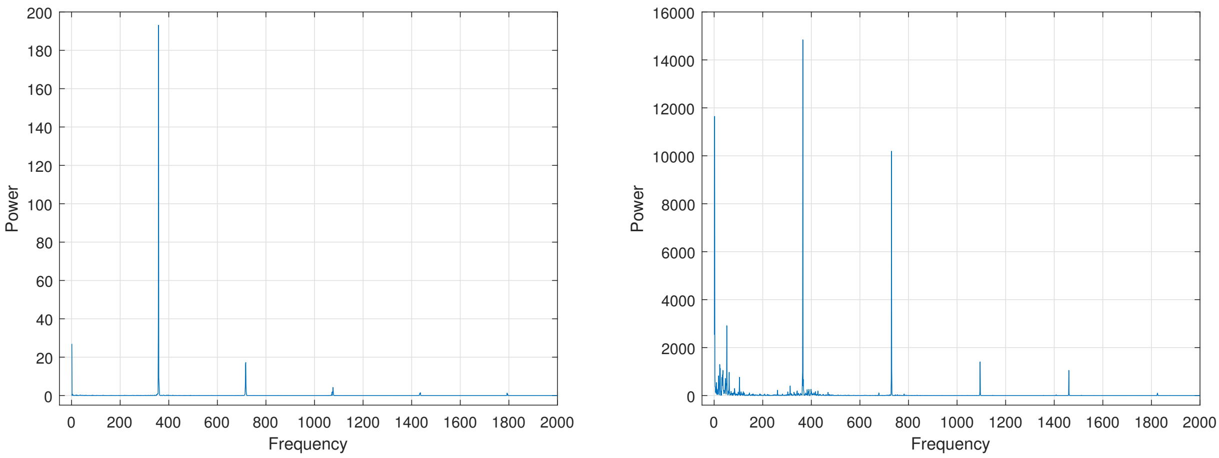
These a_j and b_j terms then represent the strengths of cycles with frequencies corresponding to j; for example, a_1 represents the strength of the cosine component of a cycle with the same period as the data set, i.e. one year, and b_{365} corresponds to the magnitude of the sine component of the daily cycle. A power spectrum can be constructed from the coefficients by plotting the power, $p_j = (\widehat{a}_j^2 + \widehat{b}_j^2)/2$, against j.

As shown in Figure 2a, the spectrum has peaks at 1, 365 and integer multiples of 365, for both generation and usage. This indicates that strong yearly and daily cycles exist in the data. In the usage power spectrum, in Figure 2b, there is also a peak at 52, corresponding to a weekly cycle. To separate the trends, we created a new set of the coefficients, containing the values for $j = 1, 365, ...$, and reconstructed the long-term trend from these coefficients, with each set of cycles being displayed in Figure 3. The short-term fluctuations are then the difference between the real data set and the long-term trends.

2.3 Time Series Models

To model the short-term fluctuations, we fit time series models. In particular, we used AutoRegressive (AR(1)) models to model the generation and usage, as they are relatively simple models, and it is expected that the generation and usage at one time would depend on the value at a previous time. Denote the short-term fluctuations by Z_t. Then, an AR(1) model takes the form

$$Z_t = \alpha Z_{t-1} + \delta \varepsilon,$$



(a) Power spectrum for solar generation, including peaks at 1 and multiples of 365, corresponding to strong yearly and daily cycles.

(b) Power spectrum for electricity usage, including peaks at 1, 52, and multiples of 365, corresponding to yearly, weekly and daily cycles.

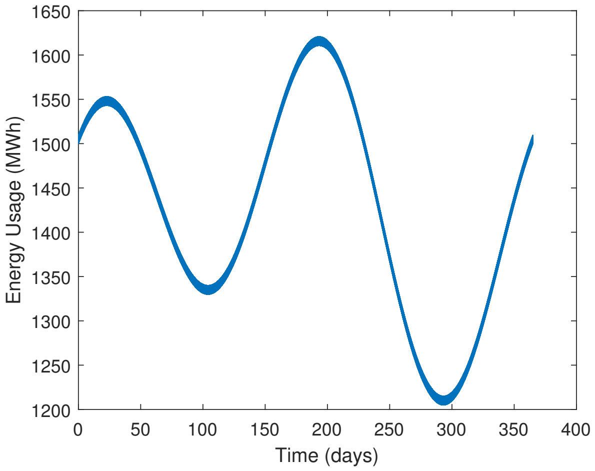
Figure 2: Plots of power spectra of generation and usage, constructed from the data.

where $\varepsilon \sim N(0,1)$ and α and δ are constants. We estimated the parameters using maximum likelihood estimation through MATLAB's econometrics toolbox. The estimated models for generation and usage are as follows:

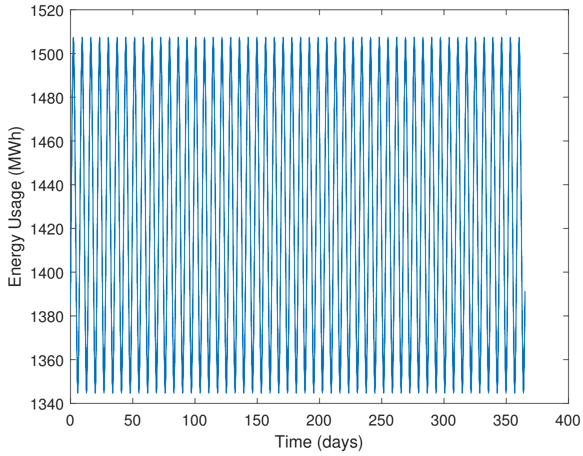
$$\text{Usage: } X_t = \begin{cases} 0.96858X_{t-1} + 3733.9\varepsilon_1 \text{ (weekday)}, \\ 0.96266X_{t-1} + 3852.5\varepsilon_1 \text{ (weekend)}, \end{cases}$$

$$\text{Generation: } Y_t = 0.81245Y_{t-1} + 178.36\varepsilon_2.$$

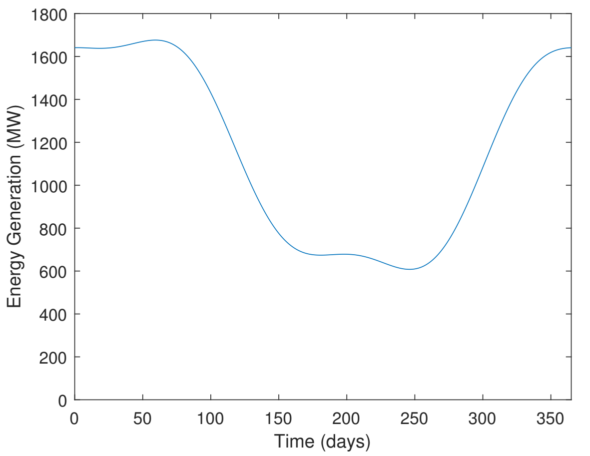
where $\varepsilon_1, \varepsilon_2$ are i.i.d $N(0,1)$. Using these models, simulations were performed for the short-term fluctuations. Figure 4 shows simulations of the short-term fluctuations compared those present in the data, and simulations of the usage and generation across the year compared to the data including the long-term trends. As shown in the simulated fluctuation plots for usage and generation, the time series models (in blue) accurately capture the spread and general distributions of the data (in orange), however in the case of the usage do not fully capture the sharp peaks during summer. These peaks may be captured by modifying the model, for example by adding a Markov Switching regime [10]. The full simulated paths for usage also closely approximate the data. However, the simulated generation appears to follow the data less accurately, due to the bounding observed in the data. This may be able to be improved by transforming the data set, fitting the models then transforming back. Alternatively, any values in the time series that are more extreme than the maximum value of the



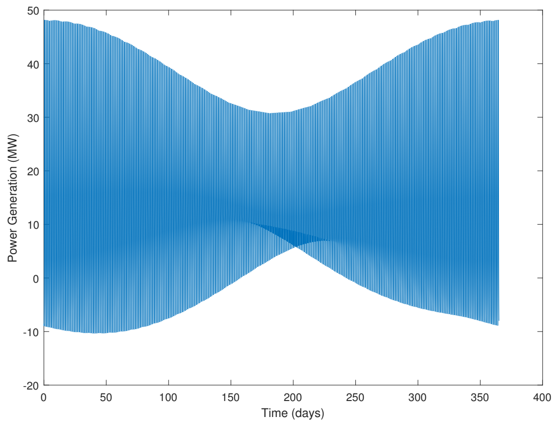
(a) Yearly and half-yearly cycles for usage.



(b) Weekly cycles for usage.



(c) Yearly and half-yearly cycles for generation.



(d) Daily cycles for generation.

Figure 3: Plots of long-term trends detected in the generation and usage data. As shown, both generation and usage values are much higher during summer, however usage also peaks in winter.

short-term fluctuations in the data could be ignored to introduce bounding in the simulations. Neither of these approaches were used, as it would over-complicate the simulations.

The difference of the models, $Y_t - X_t$ (the short-term power transfer), follows an AR MA(2,1) model, as the difference of two AR(1) models is AR MA(2,1) [4]. An AR MA(2,1) model $\{T_t\}$is of the form

$$T_t = \alpha T_{t-1} + \beta T_{t-2} + \delta\varepsilon_t + \gamma(\delta\varepsilon_{t-1})$$

where α, β, δ and γ are parameters, and $\varepsilon_t \sim N(0, 1)$ i.i.d. In this case, if X_t is the usage and Y_t is the generation, the model can be calculated from the parameters listed above:

$$\begin{aligned} X_t &= \alpha X_{t-1} + \delta_1\varepsilon_{1,t}, \\ Y_t &= \beta Y_{t-1} + \delta_1\varepsilon_{2,t}, \\ \therefore T_t = Y_t - X_t &= (\beta - \alpha)T_{t-1} - \alpha\beta T_{t-2} + \left(\frac{\alpha^2\delta_2 + \beta^2\delta_1}{\delta_1 + \delta_2}\right)(\delta_1 + \delta_2)\varepsilon_{t-1} + (\delta_1 + \delta_2)\varepsilon_t. \text{ [4]} \end{aligned}$$

We evaluated these coefficients for both weekday and weekend models to be

$$\begin{aligned} T_{t,weekday} &= -0.1561T_{t-1} - 0.7869T_{t-2} + 2630\varepsilon_{t-1} + 3912\varepsilon_t, \\ T_{t,weekend} &= -0.1502T_{t-1} - 0.7821T_{t-2} + 2708\varepsilon_{t-1} + 4031\varepsilon_t, \end{aligned}$$

and used these models to construct a Markov chain model to simulate the short-term fluctuations.

2.4 Markov Chain Model

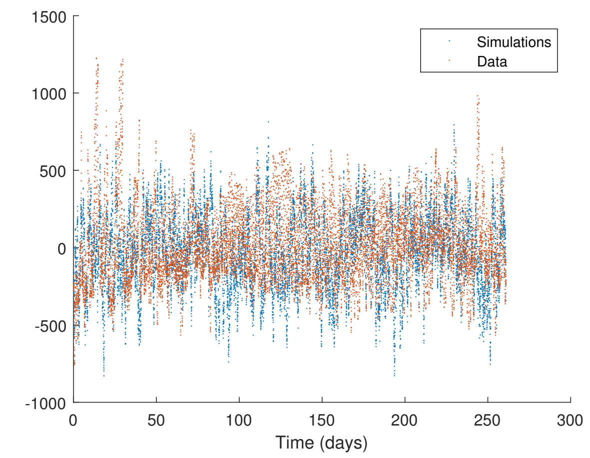
A Markov chain is a statistical model used to model changes between discrete states in a system. This is done using matrix algebra, with a transition matrix, usually denoted P, containing the probabilities of transitioning between one state to another at each time step. For a Markov chain with N states from 1 to N, P takes the form

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,N} \\ \vdots & \vdots & \vdots & & \vdots \\ p_{N,1} & p_{N,2} & p_{N,3} & \cdots & p_{N,N} \end{bmatrix},$$

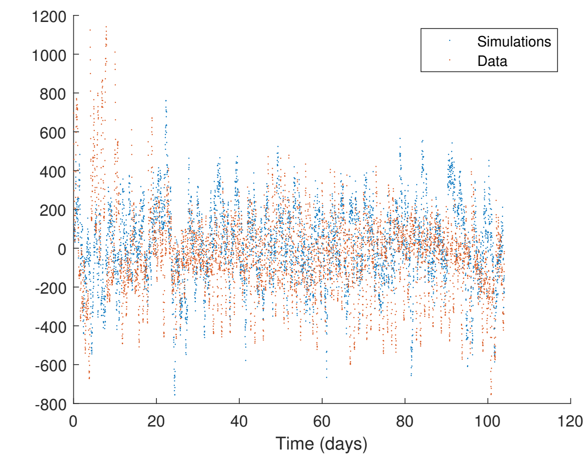
where $p_{i,j}$ denotes the probability of transitioning from state i to state j at each time step.

A property of Markov chains is the Markov property, which means that the probabilities of being in any state at one time step depend only on the state of the process in the previous time step, and not on any previous time steps, i.e.

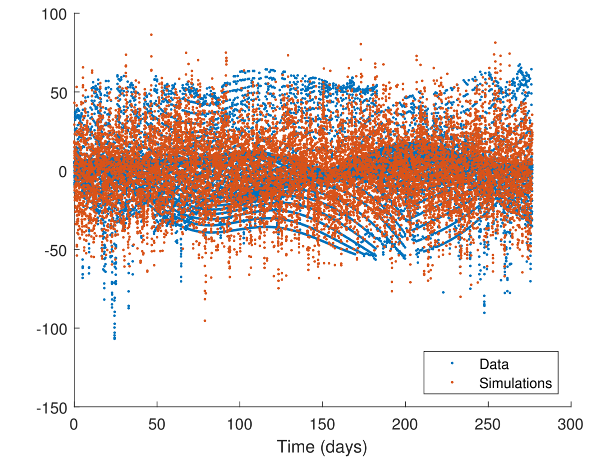
$$P(X_t = j|X_{t-1} = i_1, X_{t-2} = i_2, ...) = P(X_t = j|X_{t-1} = i_1) = p_{i,j}.$$



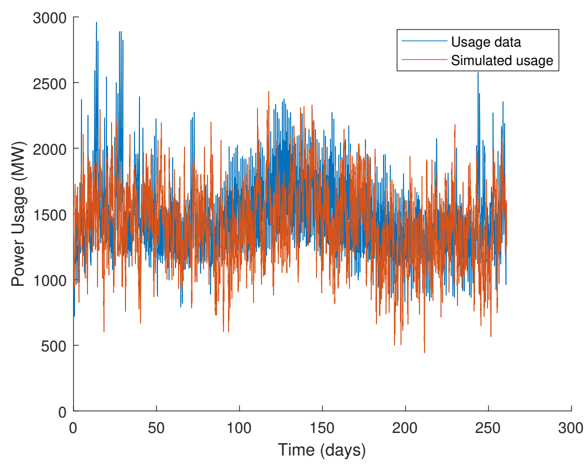
(a) Simulated short-term fluctuations for weekday usage.



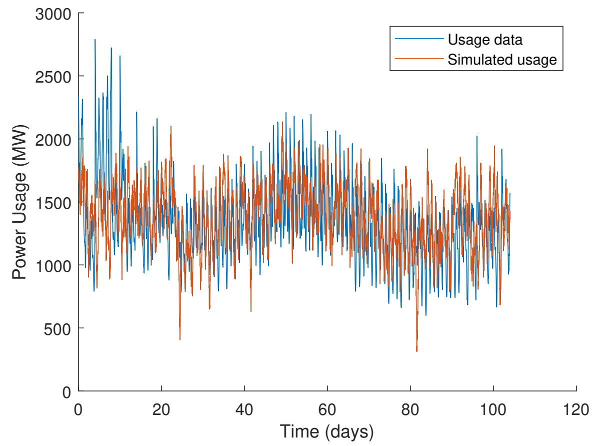
(b) Simulated short-term fluctuations for weekend usage.



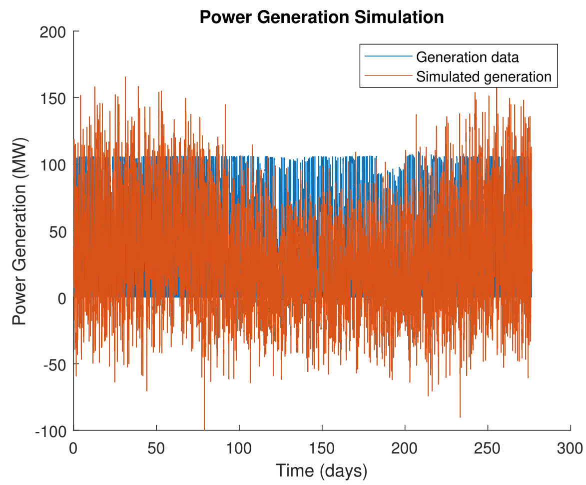
(c) Simulated short-term fluctuations for solar generation.



(d) Simulation of weekday power usage.



(e) Simulation of weekend power usage.



(f) Simulation of solar generation.

Figure 4: Simulations of generation and usage, compared to the data. The simulations for electricity usage follow the general trend and spread of the data; however, they do not show the sharp peaks during summer. For generation, the simulations also follow the trend and spread of the data but lack the bounding shown in the data.

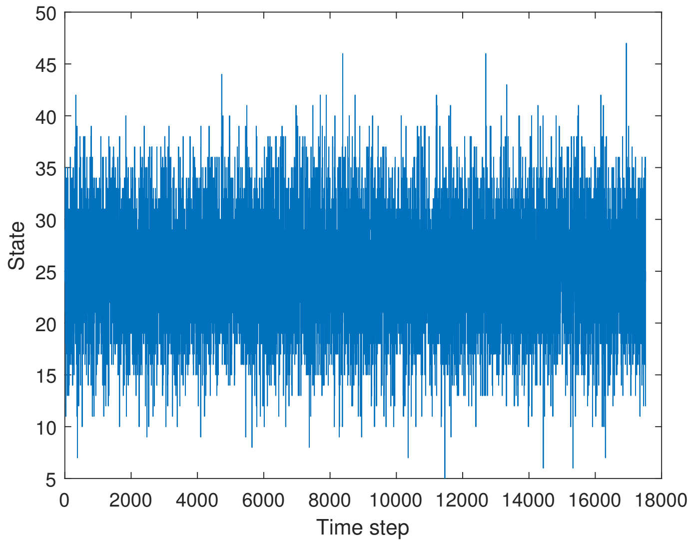


Figure 5: Simulated sample path of the Markov chain. The path mainly fluctuates about state 25, corresponding to no net transfer, with a slight bias towards positive transfer into a battery.

Due to this property, a Markov chain is useful in computations as it provides an efficient method to simulate a path of a random process, where each time step depends only on the previous time step.

In this project, each state of the Markov chain corresponds to a pair of previous states in the power transfer, as the distribution of T_t at each time step depends on both T_{t-1} and T_{t-2}. In order to model the power transfer accurately, 51 states, from 0 to 50, were constructed for T_t, resulting in $51^2 = 2601$ states in the Markov chain, corresponding to each pair of states. The Markov chain used to model the power transfer was also time-inhomogeneous; at each time step the transition probabilities $p_{i,j}$ depend on the value of ε_{t-1}, and the time of the week (whether it is a weekend or weekday, and if it is day or night).

In the simulations, at each time t we calculated the transition probabilities from the time series models, with the value ε_{t-1} and the time as inputs along with a range of values for both T_{t-1} and T_{t-2} corresponding to each state. We selected the column of the transition matrix according to the previous state, then an output state was randomly selected according to the probability distribution in the column. We ran the simulations for 17,520 time steps, corresponding to a period of one year, with initial values $T_{t-1} = T_{t-2} = \varepsilon_{t-1} = 0$. A sample path is shown in Figure 5. As shown, the path mainly fluctuates about state 25, corresponding to no net transfer, with a slight bias towards the higher states which correspond to a small amount of charge being gained by the battery.

2.5 Battery Charge Simulations

We repeated the process described in Sections 2.2-2.4 numerous times to investigate the amount of generation required. Seven amounts of generation were considered, from 70 to 120 solar farms the size

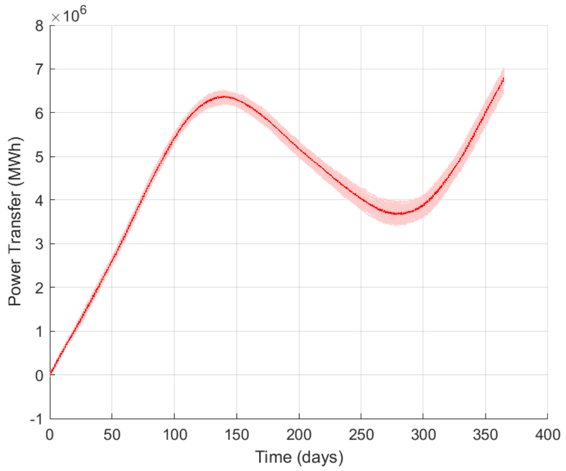


Figure 6: Plot of 100 simulated paths (light red) of power transfer and mean power transfer over time (red); 85 solar farms, showing the typical spread of the simulations.

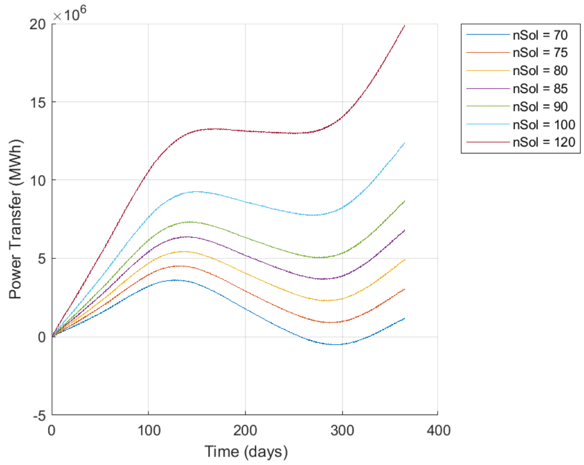


Figure 7: Comparison of power transfer for different amounts of generation. In all cases a negative trend during winter is observed, with the slope decreasing in magnitude as the number of solar farms increases.

of the Broken Hill solar farm, by scaling the data by a constant factor. In each case, 100 simulations were performed so that probabilities could be calculated, using the Phoenix supercomputer at the University of Adelaide. This number of simulations for each case allowed for estimations of the probability of the grid having a power deficiency, however 100 is likely not enough to accurately calculate the probabilities. More simulations were not performed due to time constraints and computational limits. The simulated paths, and the mean at each time step, for the case of 85 solar farms were plotted in Figure 6, with plots for the other cases provided in Appendix A. Figure 7 provides a comparison on the net transfer by plotting the mean of the simulated paths, at each time step, for each number of solar farms considered.

For all cases considered, there is a negative trend during winter. This is to be expected, as both the generation is generally lower during the winter period, and the electricity usage is higher. As more solar farms are added, for example in the 120 solar farm case, the slope gets shallower, due to the winter generation being non-zero.

In order to adapt the transfer simulations into battery charge simulations, we applied real-world effects and limitations. Firstly, we added the maximum capacity of the battery to the model by checking, at each time step, if the cumulative charge would exceed the maximum capacity (or obtain a value below 0). If this occurred, the charge was assumed to be equal to the maximum capacity (or 0), with the remainder charge being ignored, as wasted energy (or an energy deficit). We also considered battery self-discharge in the model. Estimates for the self-discharge for a lithium-ion battery range

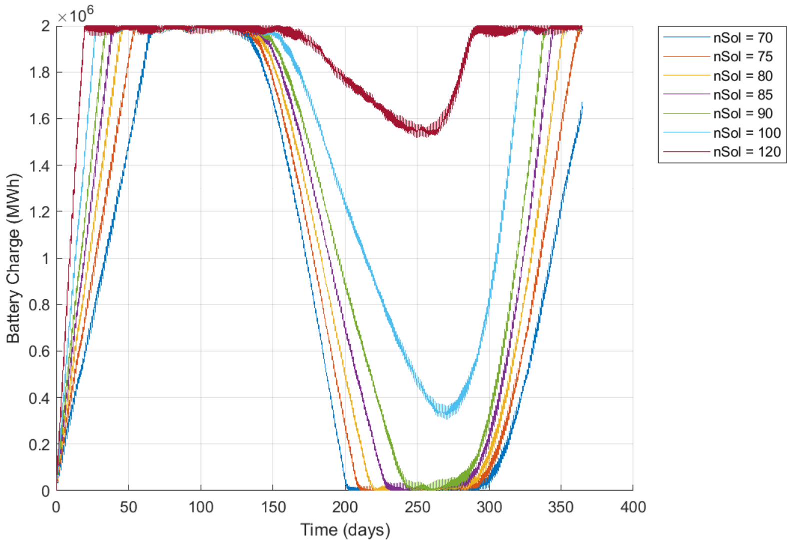


Figure 8: Plot of simulated battery charge over time, for 2,000,000MWh of battery storage. In all cases, the battery reaches its maximum capacity during summer, and in all but two cases the battery has no charge by the end of winter.

between 2-4% of the current charge per month [5], so at each time step 0.00208% of the current charge was removed, which corresponds to 3% per month. Long-term battery degradation effects, which take place over several years, were ignored in the model. Figure 8 shows the battery charge over time for each number of solar farms, with a battery capacity of 2,000,000 MWh.

As shown by the paths reaching zero towards the end of the winter period, there is a severe power deficit in all but two of the cases, which correspond to the highest amounts of solar generation. However, these cases are also not optimal, as the excess electricity generated while the path is at the maximum capacity is wasted. To find the optimal amount of battery storage and solar generation, we constructed a surface plot of the availability, here defined as the probability of the battery charge either being empty or at the maximum state, against the number of solar farms and the battery storage, and is shown in Figure 9. The vertical axis, corresponding to the availability, is logarithmic, so a value close to zero corresponds to a high probability that the battery charge is neither full, and hence wasting energy, or empty, which could lead to an energy deficit. From the plot, it is clear that the availability of the system generally increases as the storage capacity increases. For the case of 70 solar farms, the availability is limited even with large amounts of storage, due to the electricity deficit discussed earlier. The availability is also lower for the case of 120 solar farms, as the battery charge stays at the maximum capacity for a large proportion of the time. According to the plot, the optimal amount of solar generation and storage appears to be about 80-85 solar farms connected to a battery

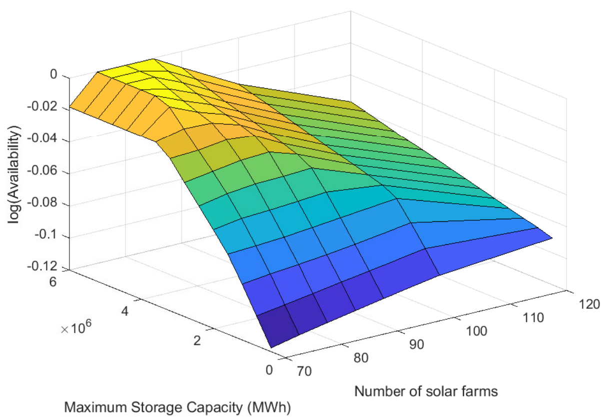


Figure 9: Surface plot of availability, against the number of solar farms and battery capacity, with a peak at approximately 80 solar farms, with a storage capacity of 5,200,000MWh.

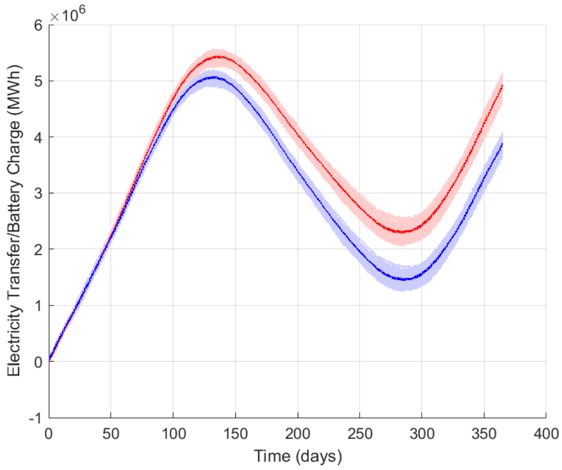


Figure 10: Plot of power transfer (red) and battery charge (blue), for 80 solar farms with 5,200,000MWh storage.

with a capacity of approximately 5,200,000 MWh, to minimise both power deficits and wasted energy. Figure 10 shows the battery charge over time for this case, including the lack of the bounding that is present in Figure 8.

3 Conclusion

According to the model developed in Section 2, the optimal amount of solar energy generation and storage required is 80 solar farms of the approximate size of the Broken Hill solar farm, or about 4160MW of reported solar generation, connected to a battery with a capacity of at least 5,200 GWh. This amount of storage is much larger than what is currently installed in Australia, with the Tesla battery at Hornsdale, the largest battery storage system in Australia, having a capacity of 129MWh, or 0.0025% of the required storage. However, this storage amount could be reduced significantly by introducing alternative generation methods, such as wind and geothermal generation, that have relatively steady energy production throughout the year. This would have the effect of both reducing the amount of solar generation required, and reducing the negative slope present in the power transfer, hence reducing the amount of storage needed.

The model could be improved in several ways, to provide a more accurate analysis of the required generation and storage. Firstly, generation from other sources, such as wind, geothermal and hydro-electric generators, could be modelled and included. This was not done in the model above due to a lack of data, or in the case of wind generation difficulties in filtering the data. Another improvement for the model is the introduction of a more accurate battery model, including additional effects such as a maximum charge/discharge rate, long-term degradation, and the effects of environmental conditions such as temperature on the effective battery capacity. Factors such as the cost of different energy generation methods and electricity storage could also be applied to the model to optimise the amounts of generation and storage with respect to both availability and cost.

Acknowledgements

I would like to thank Fergus Willsmore for his assistance in filtering the data, and Angus Lewis for explaining how to use Phoenix. This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

References

[1] Harmsen, N 2018, "How Elon Musk's big Tesla battery is changing Australia's power landscape". Available at: https://www.abc.net.au/news/2018-04-06/tesla-battery-outperforms-coal-and-gas/9625726

[2] Moore, J & Shabani, B 2016, "A Critical Study of Stationary Energy Storage Policies in Australia in an International Context: The Role of Hydrogen and Battery Technologies", *Energies 2016*, vol. 9, article 674. Available at: https://www.mdpi.com/1996-1073/9/9/674

[3] Australian Energy Resources Assessment 2018, available at: https://aera.ga.gov.au/#!/solar

[4] Zhu, W 2018, "Lecture Notes: Sum of Independent ARMA Processes", Stony Brook University, New York, USA. Available at: http://www.ams.sunysb.edu/žhu/ams586/SUM.pdf

[5] Woodbank Communications 2005, "Battery Performance Characteristics". Available at: https://www.mpoweruk.com/performance.htm

[6] Australian Energy Market Operator 2018, "Next Day Actual Gen". Available at: http://www.nemweb.com.au/REPORTS/ARCHIVE/Next_Day_Actual_Gen/

[7] Australian Energy Market Operator 2018, "Operational Demand/Actual Daily". Available at: http://nemweb.com.au/Reports/ARCHIVE/Operational_Demand/ACTUAL_DAILY/

[8] Australian Energy Market Operator 2018, "AEMO Annual Report 2018". Available at: https://www.aemo.com.au/-/media/Files/About_AEMO/Annual-Report/AEMO-Annual-Report-2018.pdf

[9] Climate Council 2018, "Fully Charged: Renewables and Storage Powering Australia". Available at: https://www.climatecouncil.org.au/wp-content/uploads/2018/03/Fully-Charged-Renewables-and-Storage-Powering-Australia.pdf

[10] Cai J 1994, "A Markov Model of Switching-Regime ARCH", *Journal of Business Economic Statistics* vol. 12, no. 3, pp. 309-316

Appendix A Power Transfer Plots

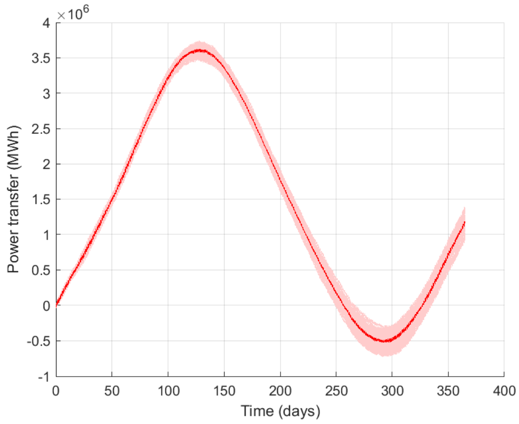


Figure 11: Power transfer simulations for 70 solar farms

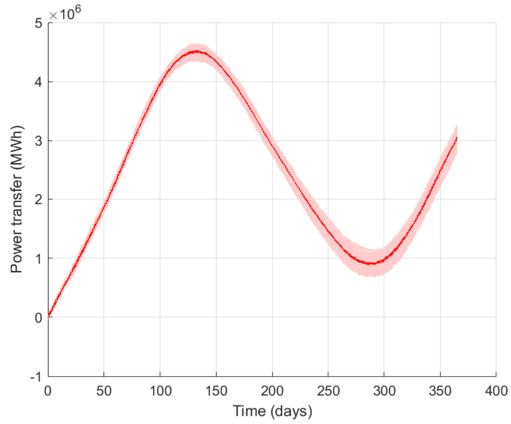


Figure 12: Power transfer simulations for 75 solar farms

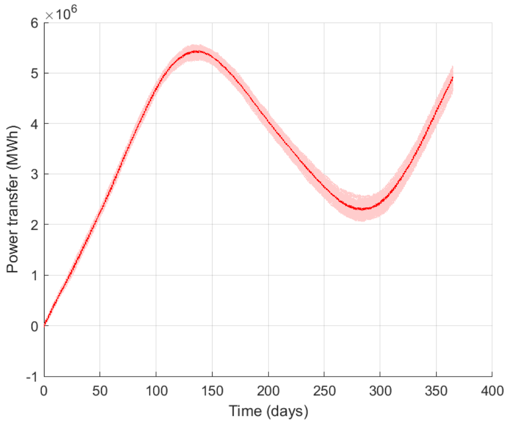


Figure 13: Power transfer simulations for 80 solar farms

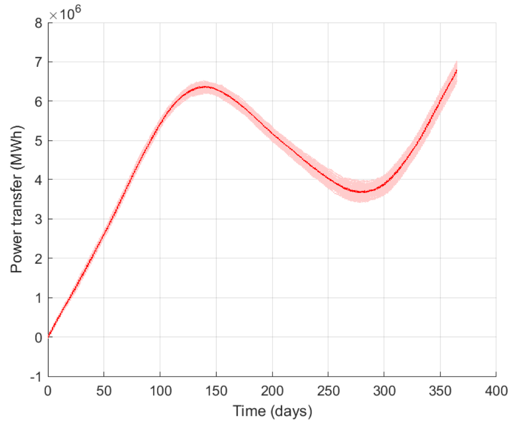


Figure 14: Power transfer simulations for 85 solar farms

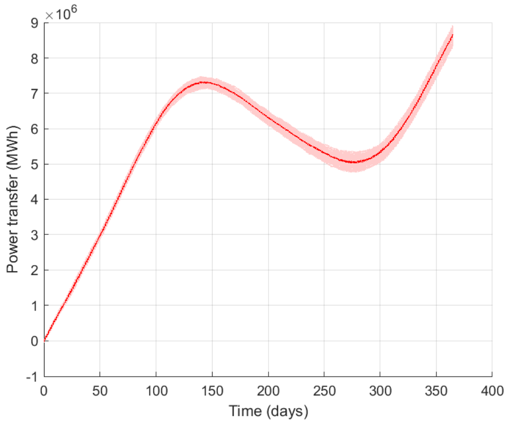


Figure 15: Power transfer simulations for 90 solar farms

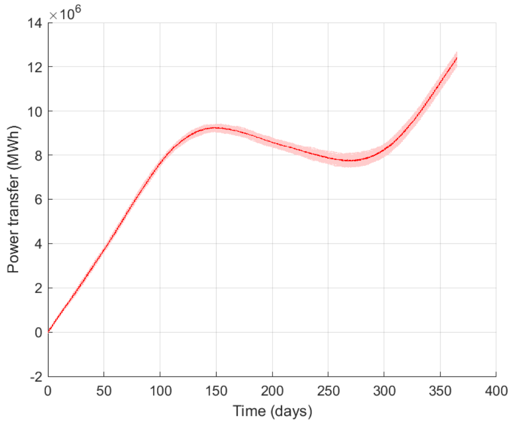


Figure 16: Power transfer simulations for 100 solar farms

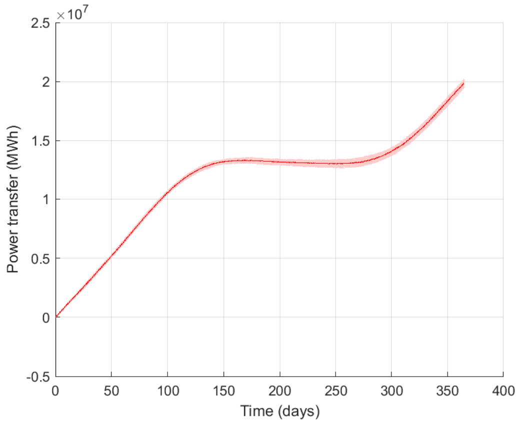


Figure 17: Power transfer simulations for 120 solar farms