

A Data Science Approach to Short-Term Electricity Demand Forecasting

Louis Long (z5162359)

Mark Budrewicz (z5353932)

Hao Chiou (z5131909)

Charlet Jeyapaul (z5375906)

07/10/2023

Contents

1	Abstract	2
2	Motivation	2
3	Introduction	2
4	Literature review	3
5	Material and Methods	5
5.1	Software	5
5.2	Description of the Data	5
5.3	Pre-processing Steps	5
5.4	Data Cleaning	7
5.5	Assumptions	7
5.6	Modelling Methods	7
6	Exploratory Data Analysis	7
7	Analysis and Results	7
8	Discussion	7
9	Conclusion	7
	References	8
	Appendix	9

1 Abstract

2 Motivation

Electricity demand forecasting is a topic that is within the interest of government and market bodies, with the consensus of accurate forecasts of energy demand driving profitability. Electricity in Australia is generated primarily using oil, coal, and gas as fuel (Bhattacharya, Inekwe, and Sadorsky 2020). Also, with oil prices increasing, this implies that the cost to generate energy will also increase (Kpodar and Liu 2022). Therefore, profit maximization is achieved when energy production is minimized to be near actual energy demand. Thus, if electricity demand can be modelled accurately, government and market bodies can then utilize such information to tailor policies as well as optimize pricing. A recent phenomenon in the electricity grid has been the shift towards renewable energy sources such as solar photovoltaic and wind generation. While these technologies are indeed cleaner they bring with them risks (Agency, n.d.) due to their intermittency and inability to be dispatchable as compared to electricity produced from fossil fuels. As of 2021, renewables accounted for 23% of total electricity generation in the state of NSW (Department of Climate Change and Water, n.d.). As it is non dispatchable, renewable electricity generation acts to lower the energy demand of the grid, so much so, that the Australian Energy Market Operation (AEMO) has already acknowledged that the increasing uptake in rooftop solar will shift maximum load demand to later in the day at sunset as temperatures remain high but roof top solar output falls to nil (Operator, n.d.c). The increase in renewable electricity generation will see a parallel requirement for flexible electricity generation such as gas, pumped hydro and grid scale batteries with quick ramp up ability to cope with the variability in supply (Energy, n.d.). And with that, a more robust shorter term electricity demand forecasting process will be required to ensure that the quick start alternative electricity generators have sufficient warning to ramp up when required.

3 Introduction

The aim of this paper is to investigate different modelling techniques to generate a rolling 60 min forward demand forecast for the state NSW electricity grid that could be used by flexible quick start generators (gas, pumped hydro, and grid scale batteries) to help minimise operational costs and give a competitive edge in pricing which will benefit its profitability and flow down to lower costs for the end users of the electricity grid.

Quick start electricity generators are essential to the electricity generation mix as they add flexibility to the grid to cope with highly variable demand (Energy, n.d.). This flexibility ensures a sufficient balance of supply and demand as these quick start generators can supply with a shorter ramp up period. While the ramp up period is shorter than conventional base load generators, it is still important for an electricity generator to be able to estimate how much electricity is required by the grid within a short forecast window so that it can strategically plan generation output to minimise its startup and shut down costs and maximise profit. Times of high load requirement in the grid inevitably lead to higher spot prices, so it is in the interests of an electricity supplier to be able to accurately forecast and capitalise on such situations, inline with its operational constraints (startup and shutdown times, maintenance requirements). While an electricity demand forecast is already generated by AEMO (Operator, n.d.b) for participants in the grid, this analysis aims to investigate a forecasting model that will outperform AEMO's demand forecast, allowing a supplier to take advantage of potential misalignment in AEMO's forecast to actual load demand.

AEMO produces a short term forecast model in 30 minute intervals (the 'Half-hourly demand model') (Operator, n.d.b), that uses features such as a categorical value for each half hour block of the day and month of the year, dummy variables for weekends and public holidays and transformations of half hourly temperature inclusive of interaction terms. This model uses a Least Absolute Shrinkage and Selection Operator (LASSO) regularisation algorithm which is a regression analysis method that automatically performs feature selection by including a penalty term to the loss function which acts to shrink the non significant variable coefficients to zero.

Our models seek to build upon the AEMO 'Half-hourly demand model', to test if other meteorological data such as solar exposure can capture the impact of household solar photovoltaic generation in the grid, and whether rainfall has an impact on indoor electricity consumption. Our investigation seeks to compare the

performance various time series and machine learning models such as LASSO, Autoregressive integrated moving average (ARIMA), Support Vector Regression (SVR) and Random Forrest Regression with the aim of improving the short term electricity demand forecast with a slight increase in computational cost. Each of the models will then be further refined via training with simulated data. The final model will be selected after comparison of various generated models and features, specifically through consideration of feature selection, accuracy, and interpretability and performance relative to AEMO’s forecast.

4 Literature review

Initial review of the literature on electricity demand forecasting pointed to a well understood relationship between temperature and electricity demand that has already been incorporated by the AEMO’s forecasting procedures (Operator, n.d.b). This relationship has been widely explored with a specific focus on higher frequency time intervals. (McCulloch and Ignatieva 2020) showed a strong relationship when modelling Australian 5-minute electricity demand with intraday temperature using a Generalised Additive Model. This same research also indicated that sensitivity of demand to temperature was heavily dependent on the time of day with higher demand sensitivity during hours of high human activity. In (G. Zhang and Guo 2020), their support vector machine (SVM) model found other meteorological features such as wind speed, relative humidity, precipitation and air pressure impacted electricity demand. Interestingly the behaviour of these variables on demand was non linear, which the authors hypothesised could be due to their impact on relative temperature (ie high wind can act to help cool on warmer days, but make it feel colder on cooler days). (Vu, Muttaqi, and Agalgaonkar 2015) regression model for monthly electricity demand for the State of NSW showed that demand was “predominantly dependent on temperature, humidity and the number of rainy days”. While the study was done on monthly electricity demand data, it points to a link in meteorological features impacting NSW electricity demand which may also transfer to more granular time periods. Other literature included additional features and characteristics of the data such as seasonality and cyclicity (from hourly to monthly) in the electricity demand time series. As described in (Operator, n.d.b), AEMO’s half hourly model includes categorical values for half hourly block of the day and month of the year, and include dummy variables for non work days. In (Clements, Hurn, and Li 2016) study on Queensland electricity demand, seasonality of demand including hour of the day, day of the week were identified as important features that needed to be factored into the modelling. One approach by (Jiang, Li, Liu, et al. 2020) was to remove the seasonality as a preprocessing step through a Fast Fourier Transform method which converted the electricity demand from time series domain to frequency domain. In this model, the seasonality was removed so as not to affect the SVM algorithm. Other approaches such as (Alonso, Nogales, and Ruiz 2020), chose to let their Long Short-Term Memory (LSTM) recurrent neural network to capture the seasonal idiosyncrasy naturally. The daily seasonality of hourly blocks was identified in the data exploration, so an LSTM with 24 hour time steps was used to capture the seasonality. As our paper sets out to investigate multiple forecasting techniques, the seasonality will need to be incorporated through different methods. As the complexity of electricity demand has evolved so has the research into techniques for its forecasting. Earlier models explored Box and Jenkins time series modelling (Hagan and Behr 1987) or what is commonly known today as ARIMA. They found that a seasonal ARIMA model was well suited to electricity load forecasting where the forecast model was a linear combination of previous days electricity load with 24 hour (daily) and 168 hour (weekly) autoregressive lag included in their model. It was noted that including a transfer function to model the non linear relationship between temperature and electricity load should improve the forecasting error. A similar autoregressive approach in [VuD.H.2017Sedf] using NSW daily electricity demand data looked at grouping days of the week with similar electricity demand profiles together (weekends) in the modelling to improve the robustness of the forecast and to vary its coefficients with time to better capture the relationship to more recent data points.

As previously discussed, AEMO’s half hourly model is a linear regression technique that uses the LASSO regularisation algorithm to produce a forecast which aims to ‘describe the relationship between underlying demand and key explanatory’ (Operator, n.d.b). In this respect it is more of an explanatory model in contrast to the predictive ARIMA methods above that have no explanatory power. Multiple regression techniques have provided a simple, yet powerful approach to forecast short term electricity demand. In (Charlton and Singleton 2014) a study was performed on 20 different locations, with 24 linear regression models generated

to forecast electricity demand for each hour of the day at each location, which included variables such as temperature (and polynomial transformation), dummy variables for season and weekday/weekend with good performance of 0.70 R². In the Australian context, (S. Fan and Hyndman 2012) regression study on Victorian and South Australian electricity demand, produced a separate model (48) for each half-hourly period of the day, and included variables such as temperature (current and lagged, 24hr min and max), previous demand observations (24hr min, max and average) and calendar variables to forecast current electricity demand. A stepwise variable selection was carried out from a wider set of variables with a process of elimination and analysis of the effect on mean absolute percentage error (MAPE) used as the selection criterion. In (“Understanding Intraday Electricity Markets: Variable Selection and Very Short-Term Price Forecasting Using LASSO” 2019), a linear regression forecasting model was used to predict 30min electricity pricing utilising the LASSO algorithm for its variable selection. This allowed an almost ‘unlimited’ number of variables to be used in the forecasting. In keeping with the approach by AEMO, we look to also implement a LASSO regularisation algorithm but our approach will investigate additional transformations and meteorological variables. While there was originally more focus on regression modelling of short term electricity demand, the research has shifted to techniques that can model the complex nonlinear relationship of electricity demand to temperature and seasonality. One of the most recent papers in the literature, (Bashari and Rahimi-Kian 2020), forecasts 24 hour ahead electricity load for the city of Toronto by combining a LSTM network using historical electrical demand with a deep Feedforward Neural Network (FNN) using forecasted meteorological inputs (temperature). The combination of the two artificial neural networks outperformed each individual network as measured by Root Mean Square Error (RMSE). Another approach to model the non-linear relationships is forecasting with a Support Vector Regression (SVR) which uses a kernel to convert non-linear data to a high-dimensional space and then separates the data with a hyperplane. (Jiang, Li, Lu, et al. 2020) utilised SVM algorithms with 30min time series electricity demand data to generate a forecast for both NSW and Singapore electricity demand, performing better than their benchmarking models. In (Li et al. 2021), a least squares support vector machine was used to model half hourly electricity demand for New South Wales, Queensland and South Australia which as a method is less computationally expensive than the standard SVM technique. For this model, the data was split between work days and weekends and a lagged time series was used with a radial basis function (RBF) kernel to model each half hour forecast. The authors did not include temperature, which we seek to utilise for our SVM analysis. Finally, Random Forest regression models have also been used as a machine learning technique for short term electricity demand forecasting. They are a type of ensemble model that utilises classification and regression trees to apply bagging to generate a series of decision/regression trees that each vote on a result. One of the benefits of random forest models is they provide a feature importance metric that can be utilised for feature selection in high dimensions of data. In (Huang, Lu, and Xu 2016), particular attention was paid to removing outliers as they could have a ‘significant effect on the importance of the features’. The paper’s analysis found that for 1 hour forward electricity demand forecast, the random forest model’s most important features were the previous 24 hours of actual demand, whether it was a workday, and the day of the week. A more recent paper (Dudek 2022) applied the random forest method to forecasting short term electricity demand but adding additional features for daily and weekly seasonality. Again, only calendar features were used in the modelling for both papers, with no temperature data utilised. While there has been the use of random forest techniques for short term electricity demand forecasting in the Australian context (G.-F. Fan et al. 2021), they have been used within combined hybrid machine learning models, so we believe our research into applying solely random forest regression model for short term forecasting is novel, which will also seek to use meteorological data. Our paper seeks to build upon the success in the literature above by applying similar machine learning methods (with additional meteorological data) to generate a comparison of the techniques to identify the best performing model specific to the NSW grid. In (Z. Wang et al. 2021), it is apparent that the best model approach and subset of features are heavily dependent on characteristics of that grid, so while these machine learning methods above have been heavily researched, we seek to understand which one of them can outperform the AEMO demand forecast.

5 Material and Methods

5.1 Software

R and Python of course are great software for Data Science. Sometimes, you might want to use **bash** utilities such as **awk** or **sed**.

Of course, to ensure reproducibility, you should use something like **Git** and **RMarkdown** (or a **Jupyter Notebook**). Do **not** use **Word**!

5.2 Description of the Data

The data sets used in the analysis contained historical data from the state of NSW inclusive of the period 01-01-2010 to 31-01-2021 comprising historical electricity demand time series and meteorological observations over the period. Each data set has been stored in our public Github (project-group-L, n.d.). The data includes the following extracts;

- . Total Electricity Demand (totaldemand_nsw.csv 42.9mb) Actual total electricity demand for the state of NSW in 5 minute increments measured in megawatts (MW). This data was sourced from the Energy Market Management System database of AEMO (Operator, n.d.a) made available by UNSW for the purposes of this research project.

- . Total Forecasted Electricity Demand (forecastdemand_nsw.csv.zip.partaa 94.4mb, forecastdemand_nsw.csv.zip.partab 21mb) Also known as the ‘Half Hourly forecast’, it is the forecasted electricity demand measured in MW for the state of NSW for a series of future ‘half hourly’ time series increments generated by AEMO per the ‘Forecasting Approach Electricity Demand Forecasting Methodology’ paper (Operator, n.d.b). This AEMO forecast data will be used as a benchmark to compare against the other forecasting models that will be investigated. This data was sourced from the Energy Market Management System database of AEMO (Operator, n.d.a) made available by UNSW for the purposes of this research project. The 1 hour forward accuracy of the forecast vs the actuals as measured by error rate was calculated as ###%.

- . Air Temperature (temperature_nsw.csv 8mb) Also known as ‘dry bulb air temperature’ (Meteorology, n.d.c), air temperature in Celsius was taken from periodic (approximately 30min) observations from Bankstown Airport weather station and is used as a proxy to represent statewide temperature in NSW. It has been included in our analysis to include the well-known impact that temperature has to electricity demand (McCulloch and Ignatieva 2020). This data was sourced from the Australian Data Archive for Meteorology (Meteorology, n.d.b), made available by UNSW for the purposes of this research project.

- . Solar Exposure (solar_nsw.csv 416.4kb) Solar exposure is the total amount of solar energy falling on a horizontal surface, measured in megajoules per square metre (MJ/m²) (Meteorology, n.d.b), with ‘values highest in clear sky conditions during the summer, and lowest during winter or very cloudy days’. Solar exposure was included in this analysis to analyse the impact of solar photo voltaic electricity generation which acts as a net reduction to total electricity demand. The data used is based on daily solar observations derived from satellite data for the coordinates of Bankstown Airport weather station which will be a proxy for statewide solar exposure for NSW. This data was sourced from the Australian Data Archive for Meteorology (Meteorology, n.d.b).

- . Rainfall (rainfall_nsw.csv 744.5kb) Rainfall in millimetres (mm) was taken from observations from Bankstown Airport weather station, with the data the representing daily amount of rainfall (Meteorology, n.d.b). It will be used as a proxy for rainfall for the state of NSW. It has been included in the analysis to study the potential indoor/outdoor impact that rainfall may have on electricity demand. This data was sourced from the Australian Data Archive for Meteorology (Meteorology, n.d.b).

5.3 Pre-processing Steps

R package In this project, data will be pre-processed via R. Data transformation will be completed via the tidyverse package (Wickham et al. 2019). Whereas date time related transformations will be completed via

utilising the Lubridate package (Grolemund and Wickham 2011).

Data cleaning and preparation All data files used in this project will be in csv format. Data files will be read into R and processed to become fit for purpose via having the correct format and sufficient integrity. Data will be processed to ensure consistency, as well as having appropriate treatment for null values.

Prior transformation takes place, all data files will be read into R, converted into dataframes, and confirmed that the data is relevant to NSW only (via relevant region columns).

Date-matrix data frames with intervals one, five, thirty minutes, as well as one day will be created. The purpose of the date matrices would be to serve as base tables for the datasets to be left joined to, to ensure consistency in time gaps of every measure. The reason why left join will be used is to ensure date without measure will still be retained, where the null value will be displayed as measure value, and will be treated post join with date matrix data frames.

Data transformation In this project, the five primary data sources utilised will be total energy demand, historical forecast, historical temperature, daily rainfall, and daily solar exposure. The timeframe of the data of interest will be from 2010-01-01 to 2022-08-01. Data transformation will be explained separately by each dataset individually as per below:

Total Demand The total energy demand dataset contains total energy demand in megawatts measures. There are no unreasonable high or low measures identified, via looking at maximum and minimum total demand recorded, thus no filtering is required. Measures are mostly taken with five minute intervals, therefore a five minute interval date-matrix data frame will serve as the base table for left join.

Forecast Demand The forecast demand dataset contains forecasts made by AEMO. Forecasts generated by AEMO are nearly always on a half-hourly basis. Within each unique timestamp, AEMO will make multiple forecasts with the new forecast taking account of the previous, where forecasts are sequences via PERIODID (indexed from 1 ~ n, 1 being the first forecast, and n-th forecast being the final). Within the forecast demand dataset, all forecast values are within a reasonable range of having no outlier, therefore no filtering is required. In this project only the first and final forecast will be used, via transformation of the tables from long to wide format, removing the PERIODID and forecast measure columns, whilst adding two new columns for initial and final forecast. The transformed table will then be left joined onto a date-matrix data frame of thirty minutes.

The temperature dataset contains temperatures measured degrees celsius with sporadic measure timestamps, usually between zero to thirty minutes, as well as larger gaps of up to three days. One thing to note with the temperature data would be that there are outlier temperatures recorded, for example -9999 degrees celsius, whilst historically the lowest temperature ever recorded in NSW is -23 degrees celsius (Meteorology, n.d.d). Therefore, the temperature data will filter out columns where temperatures are below -23 degrees celsius. The filtered temperature table will then be joined via left join to date-matrix with one minute intervals, due to sporadic timestamps mentioned above.

Daily Rainfall and Daily Solar Exposure The rainfall dataset contains NSW daily rainfall in millimetres, while the solar dataset contains NSW daily solar exposure measured in megajoules per square metre. For these two tables, they will be joined via left join to date matrix table of one day interval.

Joining of datasets All data frames will then be joined into a single table in preparation for the modelling step. The cleaned dataset of total energy demand dataset will be used as base for join for other tables. Temperature and forecast dataset will be used via datetime field. Solar exposure and rainfall datasets will be joined via relevant days, ignoring the hour minute numbers.

Aggregation Given the objective of this project is to create a model capable of forecasting hourly demand, the combined data will be aggregated to hourly level. Given such, the total and forecast energy demand will be aggregated by summing the total/forecasted energy per hour. Temperature on the other hand will be averaged. Rainfall and solar will be untouched given the measures were at daily level.

Additional fields derived from existing data:

Year, Month, Day, Hour fields will be extracted from date time field

Seasons will be derived from month field

DuringDay from Hour, where 8 am ~ 8 pm will return 1 otherwise 0

NSW Public Holiday via utilising the tsibble package in R (E. Wang, Cook, and Hyndman 2020)

Extreme temperature of day and night fields will be derived via guideline provided by Australian government (Meteorology, n.d.a)

Extreme Hot Day = over 40 degrees during day

Hot Day = over 35 degrees during day

Extreme Hot Night= over 25 degrees during day

Hot Night= over 20 degrees during day

Extreme Cold day= under 10 degrees during day

Cold day = under 15 degrees during day

Extreme Cold Night= under 0 degrees during day

Cold Night= under 5 degrees during day

5.4 Data Cleaning

To clean the data to ensure there are no null measurements for any timeframe, for any field, methods such as linear interpolation and last observation carried forward will be utilised (Y. Zhang and Thorburn 2022). Initially, simple imputation was considered due to simplicity, however such a method is also prone to bias therefore will not be applied (Emmanuel et al. 2021). Another idea that was populating missing temperature via regression based interpolation methods (Xu et al. 2013). However, given the temperature data covers only one region, as well as missing timegaps being relatively small, linear interpolation will be sufficient for this case due to simplicity. Hence overall, missing total demand, temperature, daily rainfall, and daily solar exposure will be filled via linear interpolation, and missing AEMO forecast will be populated via last observation carried forward to prevent skewing from original forecast.

5.5 Assumptions

What assumptions are you making on the data?

5.6 Modelling Methods

6 Exploratory Data Analysis

7 Analysis and Results

8 Discussion

Put the results you got in the previous chapter in perspective with respect to the problem studied.

9 Conclusion

—

References

- Agency, Australian Renewable Energy. n.d. “THE GENERATOR OPERATIONS SERIES: Report Two: Ramp Rates for Solar and Wind Generators on the NEM.” <https://arena.gov.au/assets/2021/08/the-generator-operations-series-report-two.pdf>.
- Alonso, Andrés M., Francisco J. Nogales, and Carlos Ruiz. 2020. “A Single Scalable LSTM Model for Short-Term Forecasting of Massive Electricity Time Series.” *Energies (Basel)* 13 (20): 5328.
- Bashari, Masoud, and Ashkan Rahimi-Kian. 2020. “Forecasting Electric Load by Aggregating Meteorological and History-Based Deep Learning Modules.” In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. <https://doi.org/10.1109/PESGM41954.2020.9282124>.
- Bhattacharya, Mita, John N. Inekwe, and Perry Sadorsky. 2020. “Convergence of Energy Productivity in Australian States and Territories: Determinants and Forecasts.” *Energy Economics* 85: 104538.
- Charlton, Nathaniel, and Colin Singleton. 2014. “A Refined Parametric Model for Short Term Load Forecasting.” *International Journal of Forecasting* 30 (2): 364–68.
- Clements, A. E., A. S. Hurn, and Z. Li. 2016. “Forecasting Day-Ahead Electricity Load Using a Multiple Equation Time Series Approach.” *European Journal of Operational Research* 251 (2): 522–30.
- Department of Climate Change, the Environment, Energy, and Water. n.d. “Australian Electricity Generation - Fuel Mix Calendar Year 2021.” <https://www.energy.gov.au/data/australian-electricity-generation-fuel-mix-calendar-year-2021>.
- Dudek, Grzegorz. 2022. “A Comprehensive Study of Random Forest for Short-Term Load Forecasting.” *Energies (Basel)* 15 (20): 7547.
- Emmanuel, Tlameo, Thabiso Maupong, Dimane Mpoeleng, Thabo Semong, Banyatsang Mphago, and Oteng Tabona. 2021. “A Survey on Missing Data in Machine Learning.” *Journal of Big Data* 8 (1): 140–40.
- Energy, AGL. n.d. “Spotlight on: Flexible Generation.” <https://www.agl.com.au/thehub/articles/2020/05/spotlight-on-flexible-generation>.
- Fan, Guo-Feng, Meng Yu, Song-Qiao Dong, Yi-Hsuan Yeh, and Wei-Chiang Hong. 2021. “Forecasting Short-Term Electricity Load Using Hybrid Support Vector Regression with Grey Catastrophe and Random Forest Modeling.” *Utilities Policy* 73: 101294.
- Fan, Shu, and R. J. Hyndman. 2012. “Short-Term Load Forecasting Based on a Semi-Parametric Additive Model.” *IEEE Transactions on Power Systems* 27 (1): 134–41.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with Lubridate.” *Journal of Statistical Software* 40 (3).
- Hagan, Martin T., and Suzanne M. Behr. 1987. “The Time Series Approach to Short Term Load Forecasting.” *IEEE Transactions on Power Systems* 2 (3): 785–91.
- Huang, Nantian, Guobo Lu, and Dianguo Xu. 2016. “A Permutation Importance-Based Feature Selection Method for Short-Term Electricity Load Forecasting Using Random Forest.” *Energies (Basel)* 9 (10): 767–67.
- Jiang, Ping, Ranran Li, Ningning Liu, and Yuyang Gao. 2020. “A Novel Composite Electricity Demand Forecasting Framework by Data Processing and Optimized Support Vector Machine.” *Applied Energy* 260: 114243.
- Jiang, Ping, Ranran Li, Haiyan Lu, and Xiaobo Zhang. 2020. “Modeling of Electricity Demand Forecast for Power System.” *Neural Computing & Applications* 32 (11): 6857–75.
- Kpodar, Kangni, and Boya Liu. 2022. “The Distributional Implications of the Impact of Fuel Price Increases on Inflation.” *Energy Economics* 108: 105909.
- Li, Ranran, Xueli Chen, Tomas Balezentis, Dalia Streimikiene, and Zhiyong Niu. 2021. “Multi-Step Least Squares Support Vector Machine Modeling Approach for Forecasting Short-Term Electricity Demand with Application.” *Neural Computing & Applications* 33 (1): 301–20.
- McCulloch, James, and Katja Ignatieva. 2020. “Intra-Day Electricity Demand and Temperature.” *The Energy Journal (Cambridge, Mass.)* 41 (3): 161.
- Meteorology, Bureau of. n.d.a. “About the Climate Extremes Analyses.” <http://www.bom.gov.au/climate/change/about/extremes.shtml>.
- . n.d.b. “Climate Data Online.” <http://www.bom.gov.au/climate/data/>.
- . n.d.c. “Observation of Air Temperature.” <http://www.bom.gov.au/climate/cdo/about/airtemp-measure.shtml>.
- . n.d.d. “Rainfall and Temperature Records.” <http://www.bom.gov.au/climate/change/about/extremes>

- .shtml.
- Operator, Australian Energy Market. n.d.a. “Energy Market Management System (EMMS).” [https://visualisations.aemo.com.au/aemo/web-help/Content/EMMScommon/MMS.htm?tocpath=Energy%20Market%20Management%20System%20\(EMMS\)%7C_____0](https://visualisations.aemo.com.au/aemo/web-help/Content/EMMScommon/MMS.htm?tocpath=Energy%20Market%20Management%20System%20(EMMS)%7C_____0).
- . n.d.b. “Forecasting Approach -Electricity Demand Forecasting Methodology.” https://aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/nem_esoo/2023/forecasting-approach_electricity-demand-forecasting-methodology_final.pdf?la=en.
- . n.d.c. “Maximum and Minimum Demand.” <https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-planning-data/nem-electricity-demand-forecasts/2017-electricity-forecasting-insights/summary-forecasts/maximum-and-minimum-demand>.
- project-group-L. n.d. “Project-Group-l.” <https://github.com/llog-unsw/project-group-L>.
- “Understanding Intraday Electricity Markets: Variable Selection and Very Short-Term Price Forecasting Using LASSO.” 2019. *International Journal of Forecasting* 35 (4): 1533–47.
- Vu, D. H., K. M. Muttaqi, and A. P. Agalgaonkar. 2015. “A Variance Inflation Factor and Backward Elimination Based Robust Regression Model for Forecasting Monthly Electricity Demand Using Climatic Variables.” *Applied Energy* 140: 385–94.
- Wang, Earo, Dianne Cook, and Rob J. Hyndman. 2020. “A New Tidy Data Structure to Support Exploration and Modeling of Temporal Data.” *Journal of Computational and Graphical Statistics* 29 (3): 466–78.
- Wang, Zhe, Tianzhen Hong, Han Li, and Mary Ann Piette. 2021. “Predicting City-Scale Daily Electricity Consumption Using Data-Driven Models.” *Advances in Applied Energy* 2 (C): 100025.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686.
- Xu, Cheng-Dong, Jin-Feng Wang, Mao-Gui Hu, and Qing-Xiang Li. 2013. “Interpolation of Missing Temperature Data at Meteorological Stations Using p-BSHADE.” *Journal of Climate* 26 (19): 7452–63.
- Zhang, Guoqiang, and Jifeng Guo. 2020. “A Novel Method for Hourly Electricity Demand Forecasting.” *IEEE Transactions on Power Systems* 35 (2): 1351–63.
- Zhang, Yifan, and Peter J. Thorburn. 2022. “Handling Missing Data in Near Real-Time Environmental Monitoring: A System and a Review of Selected Methods.” *Future Generation Computer Systems* 128: 63–72.

Appendix