A Data Science Approach to Forecast Electricity Demand in NSW

Student1 (z123456),

Student2 (zID),

Student3 (zID).

25/07/2020

Table of Contents

# 1 Introduction

This R Markdown template can be used for the ZZSC9020 course report. You can incorporate R chunks and Python chunks that will be run on the fly. You can incorporate commands.

TBW

# 2 Literature Review

As stated in drivers of electricity consumption and demand forecasts can be split into two different types: structural drivers such as population, economic growth, electricity price, technology adoption, energy efficiency etc which can be estimated based on past trends and expert judgement, and random drivers such as weather, customer behavior etc, which can be modeled as probability distributions. There are many factors that drive customers to make similar choices regarding electricity consumption at the same time - work and school schedules, traffic and social norms around mealtimes, weekdays, public holidays, weekends, due to weather the use of heating and cooling appliances, and many other societal factors, such as whether the beach is pleasant, or the occurrence of retail promotions.

NSW is largely self-sufficient in relation to electricity supply according to , meeting most of its local demand through state generation. The remaining electricity is purchased from other states (particularly Victoria and Queensland) through the National Electricity Market. Electricity imports enable NSW to manage its supply at the lowest cost to consumers. In 2019–20, renewable energy sources provided around 19% of the state’s total electricity generation, which is more than four times that provided in 2008–09. This does not include energy supplied by solar hot water heating, which provided an estimated supply of around 4.4 PJ in 2019–20, equivalent to 1,222 gigawatt-hours (GWh).

Load forecasting is usually concerned with the prediction of hourly, daily, weekly, and annual values of the system demand and peak demand. Such forecasts are sometimes categorized as short-term, medium-term and long-term forecasts, depending on the time horizon. In terms of forecasting outputs, load forecasts can also be categorized as point forecasts (i.e., forecasts of the mean or median of the future demand distribution), and density forecasts (providing estimates of the full probability distributions of the possible future values of the demand).

In the text, see Figure .

As the graph shows the temperature in Australia is rising and forecast show that it will continue to rise. According to the climate of NSW is changing, with 6 of the 10 warmest years on record occurring in the past 10 years. The warmest year on record for both average temperature and maximum temperature in NSW was 2019, when average temperature was 1.2°C above the 1990–2009 average. Across NSW, average temperatures will continue to increase throughout this century. By 2090, average temperature is projected to rise by around 1.3°C under a low emissions scenario and around 4.0°C under a high emissions scenario. So temperature is key factor while determining the electricity demand. The usage of heating in colder nights and cooling system during the hot days directly impacts the demand of electricity.

In STLF, both univariate and multivariate models have been discussed in the literature [[**7**](https://www.mdpi.com/2571-5577/6/6/100#B7-asi-06-00100),[**11**](https://www.mdpi.com/2571-5577/6/6/100#B11-asi-06-00100),[**12**](https://www.mdpi.com/2571-5577/6/6/100#B12-asi-06-00100)]. Univariate models use only historical demand data to make future predictions, while multivariate models consider other variables such as atmospheric variations and calendars along with historical demand data. Taylor et al. [[**11**](https://www.mdpi.com/2571-5577/6/6/100#B11-asi-06-00100),[**12**](https://www.mdpi.com/2571-5577/6/6/100#B12-asi-06-00100)] developed both univariate and multivariate models, and stated that the univariate models had good prediction capability. In univariate time series models, the historical electricity demand data are arranged with correlated past lags to capture the demand patterns even when the data are limited [[**13**](https://www.mdpi.com/2571-5577/6/6/100#B13-asi-06-00100)]. McCulloch et al. [[**6**](https://www.mdpi.com/2571-5577/6/6/100#B6-asi-06-00100)] obtained improved accuracy by including the temperature as a variable, recognizing that weather conditions play a crucial role in forecasting performance.

Short-term load forecasting is an essential instrument in power system planning, operation and control. Many operating decisions are based on load forecasts, such as dispatch scheduling of generating capacity, reliability analysis, and maintenance planning for the generators. Overestimation of electricity demand will cause a conservative operation, which leads to the start-up of too many units or excessive energy purchase, thereby supplying an unnecessary level of reserve.

In addition to the basic components of electricity demand forecasting, several external factors play a significant role in refining predictions and improving accuracy.

Temperature is a crucial factor, as it directly influences electricity consumption patterns. Extreme temperatures, whether very hot or cold, can lead to higher demand for heating or cooling, respectively. Forecasts often need to account for temperature variations to predict demand more accurately, especially during seasonal extremes or unusual weather patterns.

Holidays and weekdays also affect electricity usage. On holidays, electricity consumption patterns can differ significantly from regular weekdays due to changes in work routines and social activities. For instance, public holidays might see a decrease in commercial and industrial electricity use, while residential usage could increase due to family gatherings and home activities. Understanding these variations helps in adjusting forecasts to better reflect the actual demand patterns during such periods.

Overall, incorporating temperature data and considering holiday effects are essential for creating more accurate and reliable electricity demand forecasts, which in turn support better decision-making for power system management and operational planning.

According to (www.soe.epa.nsw.gov.au, 2021) electricity demand from the NSW grid is projected to experience a slight decline over the next five years before rising once more.

The main reason for the decline in energy consumption is the lower consumption by the NSW industrial sector, particularly the manufacturing industry, since 2012–13. The closure of the Kurnell and Clyde petroleum refineries in October 2014 was the largest single cause of reduced consumption by the manufacturing industry (Nsw.gov.au. (2018)). Overall energy consumption has remained stable over the last three years with little change in sectoral demands. There was a noticeable decline in industrial energy consumption following the closure of the Kurnell and Clyde petroleum refineries in October 2014, but few major changes since then (www.soe.epa.nsw.gov.au. (n.d.)). This initial decrease is anticipated to be driven by population growth being counterbalanced by enhancements in the energy efficiency of appliances and machinery. Furthermore, the growing adoption of rooftop solar panels and battery storage systems is expected to further reduce residential demand on the electricity grid. However, beyond the five-year mark, consumption is forecasted to increase as electric vehicle charging and broader electrification begin to significantly impact electricity demand. According to NSW Climate and Energy Action. (n.d.)  the share of solar and wind in NSW’s energy mix has more than DOUBLED from 5% in 2015 to 12% in 2019.

In (Forecasting Approach - Electricity Demand Forecasting Methodology Electricity Statement of Opportunities Forecast, 2023) AEMO uses a monthly regression model based on five years (60 months) of historical data. The choice of five years strikes a balance between ensuring that the model considers only relatively recent consumption trends and behaviours, while being long enough to capture seasonality and contain enough observations to be statistically meaningful. At this stage, structural shocks which affect the data series (such as COVID-19) can also be captured using dummy variables, where applicable.

In the Fan and Hyndman in the research paper (Fan and Hyndman, 2012) proposed a semi-parametric additive models are proposed to estimate the relationships between demand and the driver variables. Specifically, the inputs for these models are calendar

variables, lagged actual demand observations and historical and forecast temperature traces for one or more sites in the target power system. The proposed methodology has been used to forecast the half-hourly electricity demand for up to seven days ahead for power systems in the Australian National Electricity Market (NEM).

# 3 Material and Methods

## 3.1 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!

## 3.2 Description of the Data

How are the data stored? What are the sizes of the data files? How many files? etc.

## 3.3 Pre-processing Steps

What did you have to do to transform the data so that they become useable?

## 3.4 Data Cleaning

How did you deal with missing data? etc.

## 3.5 Assumptions

What assumptions are you making on the data?

## 3.6 Modelling Methods

TBW

# 4 Exploratory Data Analysis

This is where you explore your data using histograms, scatterplots, boxplots, numerical summaries, etc.

import numpy as np  
np.random.seed(1)  
np.random.normal(0.0, 1.0, size=10)

## array([ 1.62434536, -0.61175641, -0.52817175, -1.07296862, 0.86540763,  
## -2.3015387 , 1.74481176, -0.7612069 , 0.3190391 , -0.24937038])

# 5 Analysis and Results

## 5.1 XGBoost

Having a very simple model is always good so that you can benchmark any result you would obtain with a more elaborate model.

For example, one can use the linear regression model

where it is assumed that the ’s are i.i.d. .

## 5.2 Facebook Prophet

TBW

# 6 Discussion

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

# 7 Conclusion and Further Issues

What are the main conclusions? What are your recommendations for the “client”? What further analysis could be done in the future?

A figure:

In the text, see Figure .

# References

# Appendix

## **Codes**

Add you codes here.

## **Tables**

If you have tables, you can add them here.

Use <https://www.tablesgenerator.com/markdown_tables> to create very simple markdown tables, otherwise use .

| Tables | Are | Cool |
| --- | --- | --- |
| col 1 is | left-aligned | $1200 |
| col 2 is | centered | $12 |
| col 3 is | right-aligned | $1 |