Climate Change & Its Impact On Victorian Energy Demand

ADS1002

Adam Choong | Jack Dong | Junwei(Alex) Huang | Daniel Naude | Bob Wu

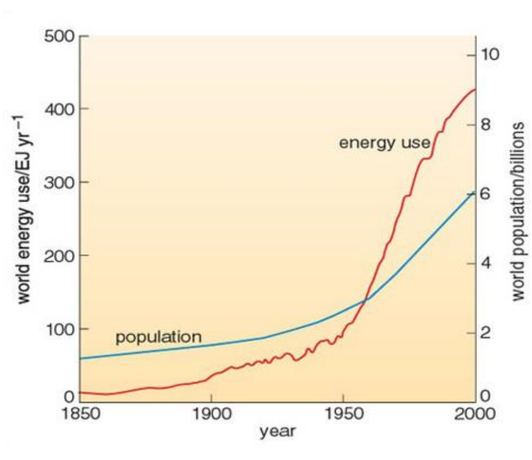
General statement of contribution

All members listed were active in the completion of this project. Bob was chiefly responsible for the background research components of this project. This included the construction of a detailed introduction into the project's topic, the effect of climate change on energy demand, and an insightful summary of the conclusions made about the results from the data analysis and modelling. Jack was responsible for performing the linear modelling component of the project. The intention of this section was to provide sufficient evidence to reject the possibility of using simple linear regression on the provided data set. Alex was responsible for conducting exploratory data analysis to determine the appropriateness of the project's research question and to provide evidence which justified the level of validity that the project's results had. Adam was responsible for exploring the use of the random forest regression algorithm on the provided data, with the intention of predicting the effects of rudimentary climate change on forecasted energy demand. Similarly, Daniel was responsible for exploring the use of support vector machines in the prediction of energy demand. Daniel also oversaw the analysis and application components of the results. Adam served as chief editor of the project report. Alex served as chief editor of the PowerPoint and written the presentation.

Introduction:

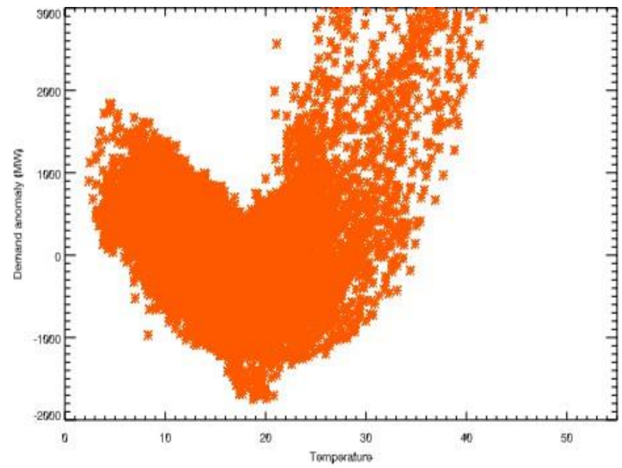
Power, defined as the amount of energy transferred or converted per unit of time, is also the very thing that keeps modern society functioning. Governmental institutions track the consumption habits of millions of Australians and deliver energy (measured in MW) across the country to those in need of it. In Australia, the main source of energy generation and its subsequent conversion into power is electricity. According to the National Electricity Market, the government supplies around 200-terawatt hours of electricity annually or around 80% of Australia's total electricity consumption (NEM, u.d.), demonstrating how dependent Australians are on electricity.

Figure 1: Energy use mapped against world population



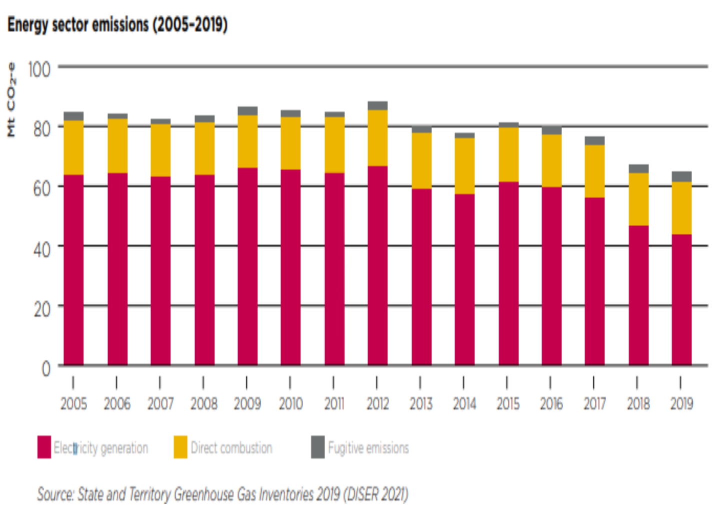
The trend describing the increase of electricity demand was accelerated during the 1920's and 1930's as more households gained access to electricity more readily. For instance, only 6% of households in Britain had electricity in 1919, but this figure had grown to two thirds of households two decades later as described by Science Museum (Science Museum, 2020). Consequently, not only were more and more people gaining access to electricity as a means of power generation, but newly invented and popularised electric appliances and systems that gradually became commonplace within households and institutions, also caused each individual resident to consume more energy per capita as time progressed, thus increasing demand exponentially.

Figure 2: Demand anomaly mapped against temperature



Going into the future, to ensure that there is always an adequate supply of power to match demand, it is necessary to forecast what the peak demand of power is likely to be during each hour and each day of the week. Furthermore, temperature is a key variable observed in this project because it influences peak demand and can therefore be strongly considered as the main indicator that dictates overall demand. For example, as days become increasingly hot, cooling needs to be used for longer periods of time and to a greater effect, especially in Australia where temperatures are exceedingly hot during summer as seen in figure 2, therefore increasing demand. Similarly, during colder periods, heating is required to a great extent, thereby increasing demand as well. These changes would result in a dramatic shift in the shape of graphs like in Figure 2.

Figure 3: The trend of Victorian energy sector emissions



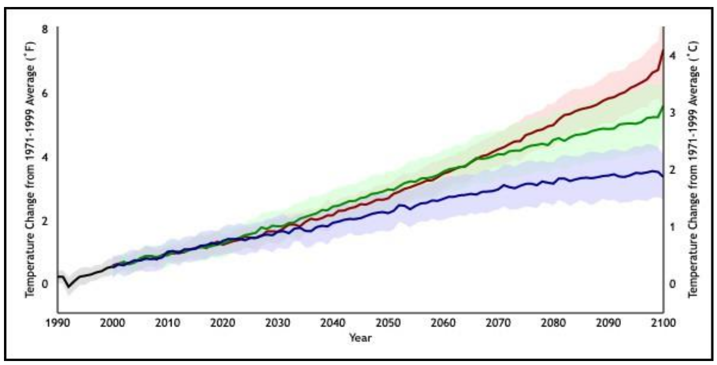
In response to soaring energy demand across the state, the Victorian government has published an energy sector emissions reduction pledge detailing programs to control emissions. One such pledge is the Solar Homes Program launched in 2018 by the Victorian State Government (Victoria State Government, 2021), which involves installing 135,000 rebated solar panels, batteries and hot water systems, generating 790 megawatts of clean power for 800,000 Victorian households. Another is the Victorian Energy Upgrade program (VEU), where $1.3 billion is invested towards the installation of 9 million energy-saving items like energy management information systems, smart thermostats, ceiling insulation and more to make businesses and households more energy and cost efficient.

Altogether, these government plans helped households and businesses to save $110 and $3,700 respectively per year and resulted in a 11 million tonne reduction in greenhouse gas emissions. As a result, it is observed in figure 4 published by the Victorian State Government (Victoria State Government, 2021) that since 2012 they have successfully reduced carbon emissions from 86 million tonnes to 63 million tonnes in 2019, and this reduction is forecasted to decrease even further in the upcoming years as Victoria becomes more energy efficient.

Regardless, temperatures would continue to increase due to global warming in the coming years, which would affect the local energy demand. Hence for this project, the following question will be inquired into:

"Through plotting local temperature against energy demand in MW, how accurate of a prediction can be made with regards to forecasting the expected energy demand in a chosen year assuming a 2 degree increase in global temperature due to global warming?"

Figure 4: Predicted emission pathways according to climate.gov



To achieve this, weather station data from Olympic Park and Melbourne's Central Business District was obtained and utilised to construct various models in an attempt to simulate rudimentary climate change. Through this process, the goal is to extrapolate a value that would reflect predicted net increases in energy demand, and finally evaluate the significance of such value to answer the research question. The 2-degree value was selected because it followed the most realistic emission pathway where renewable and non-renewable sources of energy generation were balanced according to Herring (2012), who displayed results from numerous climate model simulations as shown in figure 4.

Data

Preprocessing and data manipulation

The raw data came from 1143 .csv files containing half-hourly observations from NSW, Queensland, Victoria, South Australia, and Tasmania. The features of the data include the time and date of the observations, the level of energy demand at the time (measured in Megawatts), the price of energy in dollars, and the air temperature in ˚C. A wide range of meteorological information was also included in the files, such as humidity, wind speeds, and air pressure observations, though this information was removed from consideration as it was not relevant to the project. Air temperature rather than apparent temperature was used because data on apparent temperature was not provided despite its greater appropriateness to the project. It was possible to calculate apparent temperature using all the data provided but due to a lack of time and knowledge, apparent temperature was not calculated. A python script was constructed to match each temperature observation to the energy demand that occurred at the same time, and a single dataframe was produced per state. In order to narrow the analysis to the most relevant information. Only the data from Victoria over the period from 1st January 2015 – 1st January 2020 was considered, as this was the most relevant to the project's current situation.

After compiling the data into one dataframe and dropping irrelevant features, the process to impute missing values was initiated. Only the air temperature variable had any missing values, and there were very few overall. The missing values were recorded as five empty spaces, which were first replaced with NaN values. A k-Nearest-Neighbours imputation was attempted, but was not useful with the limited feature set provided. Instead, the missing values that had neighbouring values (30 minutes before and 30 minutes after) were replaced with the mean of those neighbouring temperatures. Any missing values with no neighbouring temperatures were dropped entirely, though this only represented about 0.1% of all data. The resulting dataset had over 90,000 Victorian observations between 1/1/2015 and 1/1/2020, each containing the time, temperature, energy demand, and energy price.

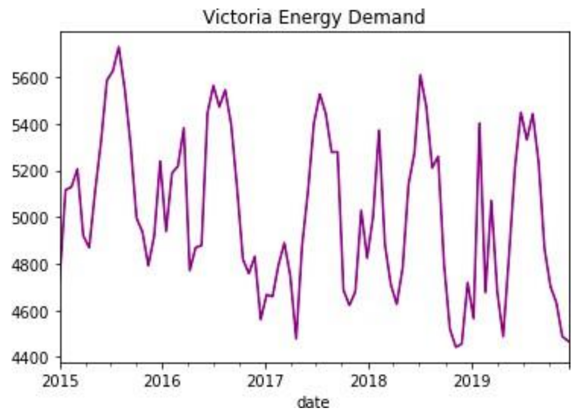
Figure 5: A glimpse of the dataframe structure.

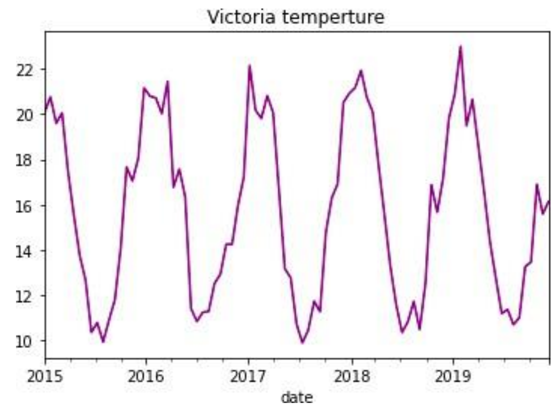
date	DEMAND	RRP	TEMP
2015-01-01 00:00:00	4833.57	20.39	16.7
2015-01-01 00:30:00	4594.19	20.10	16.6
2015-01-01 01:00:00	4395.72	20.20	16.5
2015-01-01 01:30:00	4225.04	15.20	15.7
2015-01-01 02:00:00	4051.82	14.20	15.9

Exploratory Data Analysis

In this project, the data used was specifically from the Victorian data set and analysis was performed over the last 5 years. The Victorian data, of all states, was chosen because of the abundance of research shown on Victoria's energy market and the project members' preference to conduct analysis on a location that had a familiar landscape with regards to climate and energy usage. Furthermore, only one explanatory variable was used during the analysis due to external recommendations. To demonstrate the reliability or lack thereof of the results produced from this study, a correlation coefficient was calculated between the explanatory variable (temperature measured in ˚C) and the response variable (demand measured in megawatts). Moreover, the relationship between temperature and demand was analysed on an annual and seasonal scale to suit the research question's purposes. The correlation coefficient calculated was relatively low (0.174), implying either that these variables had weak association or that there was no fixed relationship between them. In support of the case that there was no strictly fixed relation between these variables, graphical evidence was provided to show that demand is determined by the extremity of temperature.

Figures 6 and 7: A demonstration of the inverse relationship between temperature and demand

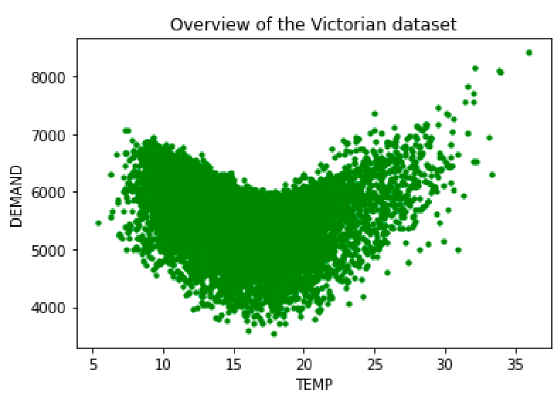




In figures 6 and 7, it can be seen that when demand is high, temperatures are either at the relatively extreme upper range (relatively hot) or relatively extreme lower range (relatively cold). The inference with regards to the patterns of temperature and demand was made based on the assumption that during the winter months (midyear period), temperatures were generally colder than those of the summer months (end of year and beginning of year periods).

Another issue with the data provided was the lack of observations, particularly towards the extreme upper range of temperatures. This was becoming a growing concern because of the inability to remove what would typically be considered outliers that greatly affected each model's accuracy. The reasons why justification to remove outliers was lacking were that there was a sufficient volume of data in the extreme upper range and contextually, the study aimed to analyse the relationship between temperature and demand at the aforementioned region, thus adding statistical significance to these data points. The succeeding figure demonstrates evidence of the lack of data in the upper range of temperature.

Figure 8: Scatter Plot mapping the approximate shape of the data provided



In figure 8, it is observed that extremely hot temperatures, particularly those above 30˚C, are lacking the appropriate relative abundance of input to produce a valid conclusion of demand but the data cannot be ignored because of its contextual significance. Extreme ranges in temperature are statistically significant because they portray a more realistic landscape of Victoria's current climate.

Overall, the association between these two variables was weak but under external recommendations, temperature was used. A main reason for following this recommendation was that if a conclusion was made, it would be sufficiently meaningful to the research question. To ascertain the fitness of each model, the R-squared score was considered to a reasonable extent as an objective measure of appropriateness.

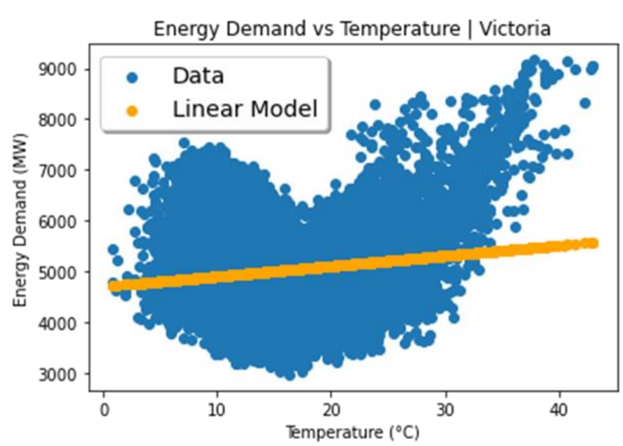
Analysis

Linear Regression Model

The first option used when attempting to model the effects of temperature on energy demand was a simple linear regression model. Here, the independent variable was the air temperature, while the energy demand was the response variable. Before a model was fitted to the data, normalisation to both features was done to prevent the difference in magnitude of each feature variable from skewing the model. After fitting the model, the resulting coefficients and metric were denormalized, giving the model shown in figure 9:

$$Energy\ Demand\ =\ 20.30\ \times\ Temperature\ +\ 4697.78$$

Figure 9: Linear Regression Model: Temperature vs Demand



This model had a mean absolute error of 696.63, and a root mean square error of 859.86. The goodness of fit was very low, with a testing R^2 of 0.0207. This indicated that only 2% of the variation in energy demand was explained by variations in temperature. Clearly the data is not well-fit by a simple linear model, something visible even by simple inspection of the data above. While regularisation may improve this model slightly, it is highly unlikely to improve the linear model to an acceptable standard, thus rejecting this model and proceeding to explore other non-linear models would be more appropriate.

Random Forest Regression Model

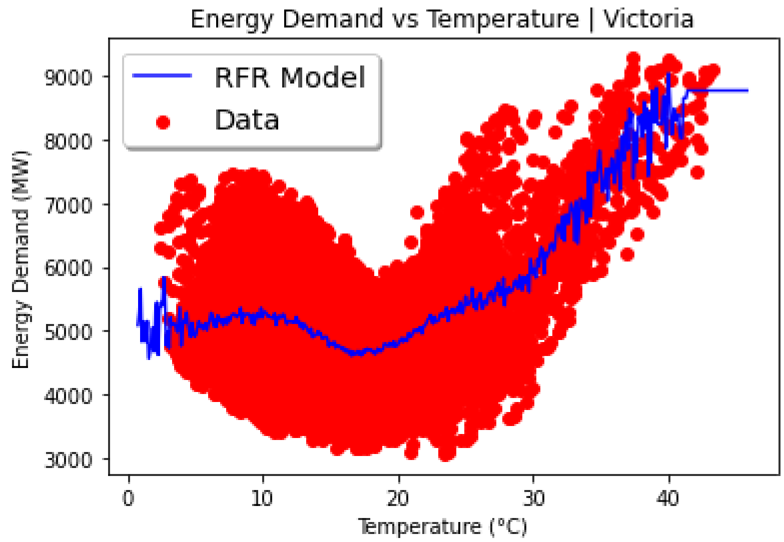
Another predictive model that was used was a random forests algorithm. This algorithm utilizes localised data to predict the outcome of the next predicted value. Sample data is taken from each region of temperatures in the dataset and decision trees regress the value of each data point to a specific range of values. When a random forest model is trained, multiple decision trees are constructed with each tree being characterised by the mean and mode of the trained values. Then a random forest prediction is made by regressing values in the testing set along the random forest constructed during the training stage, and taking the mean value of the regressed values as the overall prediction. Hence, extrapolation from this process alone is impossible, which creates issues in the context of this study, when temperatures in the upper range of the testing set are manipulated for forecasting purposes.

In addition, random forest algorithms do not need to be regularised or cross validated. Cross validation is unnecessary because during the training phase, each decision tree chooses the next option which would result in the least sum of squared errors, optimising the parameters that would typically require a manual cross-validation to find. Regularisation does not always need to be applied because random forests use a method called bootstrapping by default, which lowers variance as much as possible without increasing bias in this process. However, in the event that variance and bias is required to be balanced, regularisation can be applied, albeit not by using a typical penalty cost function. In an extremely simplified sense, rather than allowing a full reduction of variance that cannot be explained by the model, (objectively measured as the sum of squared errors or residual sum of squares) the degree of reduction in variance can be fine tuned to optimize accuracy as a means of regularisation. (Wundervald, 2019) Normalisation is unnecessary because although a random forest acts globally on the defined training and testing sets of data, each decision tree partitions the training data during the fitting stage, greatly reducing the effects that the magnitude of feature variables has on the predicted outcome.

It is vital to note that random forest algorithms can be (and typically are) used to find feature importance of variables but for the purposes of this project, such a procedure is unnecessary because the project's scope of concern only analyses the relationship between air temperature and demand.

In this project, when the random forest regression (RFR) algorithm was used, it was found that the root mean square error and mean absolute error respectively was 773.603 and 635.443. Moreover, as shown by a visual representation of the model in the next figure, the regression line oscillates frantically, either citing poor performance of random forest algorithms at the outer ranges of the data due to the model's inability to extrapolate or a lack of input from the training set around the regions of strong fluctuation .

Figure 10: Energy demand in Victoria for the corresponding temperature range



The red scatter plot represents the training data on which the model was fitted and the blue line represents the predicted demand according to the model. As observed, the ends of the regression line are oscillating heavily, and this characteristic could be attributed to either reasons mentioned above. Also note that beyond the maximum temperature of the training set, regression ceases to occur, as shown by the line plateauing when the extreme temperature is reached, indicating evidence of the random forest algorithm's inability to extrapolate. Another crucial metric that has been previously mentioned as a means of comparison is the R2 score. This model's R2 score is 0.222, indicating a better, albeit poor, fit to the data than the linear model.

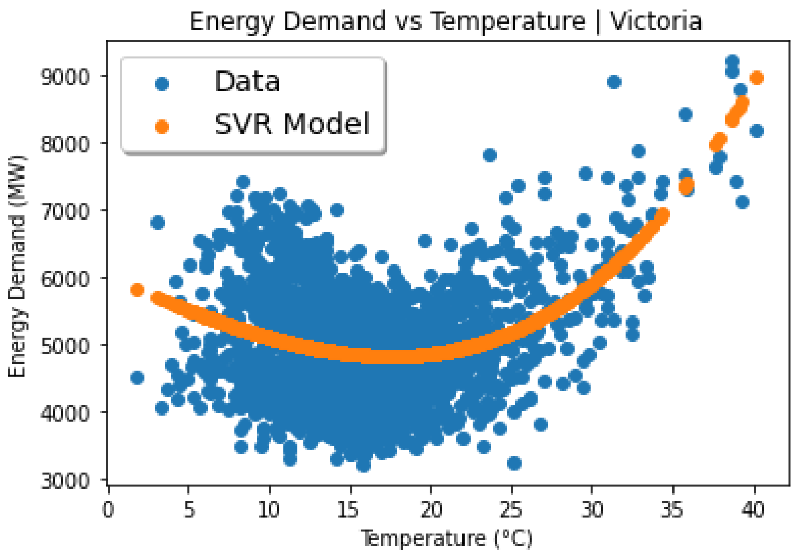
Support Vector Machine Regression Model

Support Vector Machine Regression, or SVR, is a powerful regression modelling technique that allows for the use of support vector machines to fit non-linear models. SVR is flexible and customisable and can be applied to many situations and use-cases by altering parameters and kernels to fit specific data.

One of the largest drawbacks of SVR modelling is the large computational cost of fitting and large amounts of data. Cross-validating models to find the optimal parameters involves fitting many SVR models, and is thus even more expensive and time consuming. Because of this factor, SVR modelling is typically reserved for projects involving fewer than 10,000 data points.

This limitation proved troublesome when it came to fit a non-linear SVR model to the data, which consisted of over 90,000 data points when restricted to Victorian data from 2015-2020. In order to work around this issue, a random sample of 10,000 data points was taken from the dataset. This sample was split randomly into train-test sets (80% train, 20% test). An SVR model was fitted to the training data with a cubic polynomial kernel. This kernel allowed for the avoidance of oscillations on the extreme ends of the model that the RFR model encountered, and ensured that the model behaved as expected when temperatures grew beyond the upper limits of the dataset. This became an important factor when using the model to forecast into the future.

Figure 11: Support Vector Regression Model



The model produced had an R^2 of 0.181, a mean absolute error of 656.451, and a root mean square error of 791.459. While these metrics appear worse than those of the RFR model, the SVR model has some advantages, such as its more intuitive behaviour at the extremes of the data, and lower risk of overfitting due to the cubic kernel's stability.

Forecasting

The introductory research surrounding this project has indicated that due to the effects of climate change, a 2 degree increase in average global temperatures will be observed by the year 2050. In order to account for the effect that this increase in temperature will have on the demand for energy, the models explored were trained on historical data. The temperatures from 1st January 2019 to 1st January 2020 were increased by 2˚C to simulate a crude representation of the climate from the year 2050, after global temperatures have increased. Inputting the new dataset into the SVR and random forests regression models outputted a set of predicted energy demand values from which the impact on energy demand could be gleaned. In the case of the RFR model, the total annual energy demand was 2.93% higher than in 2019, while the SVR model predicted a 2.46% increase.

While these figures appear insignificant, there are many effects and nuances that this analysis does not take into account. The true impacts of climate change on temperatures are likely to be localised and affect different regions in volatile and diverse ways: Australia will likely see vastly different changes in climate when compared to other countries such as Sweden, for instance. While the average global temperature may rise by 2 °C by 2050, this number could vary wildly in different regions nationwide as well. Also, many technological and societal changes will occur in the coming decades, which will likewise unsettle many of the assumptions asserted in this report. Since these models have only been trained on historical data from within the last decade, they are ill-equipped to forecast so far into the future. In order to gain a true indication of the effects that climate change will have on the energy industry, a much more rigorous process of climate modelling will have to be undertaken.

Conclusion

In an attempt to forecast expected energy demand, the chosen predictive models generally forecasted a 2-3% increase in Victorian energy demand on an annual basis when factoring in the effects of rudimentary climate change. Since this project utilised an extremely simplified model that only takes temperature into account, the results would not remotely reflect a realistic figure range as many other variables relating to energy demand would also need to be forecasted, such as population change, technological efficiency, and disruption from extreme weather events, which would be far too intensive and a near impossible of a task to achieve with the current resources provided.

Additionally, the theoretical results from the models are considered to be highly inaccurate due to the simplicity of the methodology, so relative judgements had to be made, rendering it unreliable and hindering its reproducibility and repeatability when trying to obtain a precise result. A potential improvement that could be made to increase accuracy is if the parameters of the models were optimised further, a procedure which is achieved through incorporating algorithms such as Bayesian optimisation, random searching, grid searching and further regularisation to see which would yield the most accurate results. Other techniques to consider using are Gaussian processes, particularly those that are proficient at forecasting future data based on training data and are unsupervised.

In conclusion, while this investigation into the effects of climate change on electricity demand was quite thorough and intricate, there is so much more to be done to further understand this unique relationship. Proceeding into the future, interest and knowledge in these issues would continue to expand as governments try to meet the growing demand for energy without exacerbating the effects of climate change. Hence, this topic must be explored in further detail, with the intention of preparing the energy sector for the consequences of the future measures taken by governments to curb the destruction caused by climate change.

Bibliography

Australian Energy Market Commission. (n.d.). *National Electricity Market.* https://www.aemc.gov.au/energy-system/electricity/electricity-system/NEM

Doyle K. (2018, February 12). *Heatwaves send temperatures soaring in many parts of Australia, but what are they?* https://www.abc.net.au/news/2018-02-12/heatwaves-occur-regularly-but-where-do-they-come-from/9416800#:~:text=The%20outback%20gets%20so%20hot,the%20earth%20and%20the%20sun.&text=Because%20our%20planet%20is%20a,equator%20than%20at%20the%20poles.

Herring D. (2012, March 6). *Climate Change: Global Temperature Projections.* Climate.gov. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature-projections#:~:text=Results%20from%20a%20wide%20range,gases%20that%20human%20activities%20produce.

Raj, A. (2020 June), *A Quick and Dirty Guide to Random Forest Regression*, Towards Data Science, https://towardsdatascience.com/a-quick-and-dirty-guide-to-random-forest-regression-52ca0af157f8

Science Museum. (2020, January 28). *Electrifying: The story of lighting our homes.* https://www.sciencemuseum.org.uk/objects-and-stories/everyday-wonders/electric-lighting-home

The Open University. (2016). *Future energy demand and supply*. https://www.open.edu/openlearn/nature-environment/environmental-studies/future-energy-demand-and-supply/content-section-4.2.1/?printable=1

Victoria State Government. (2021, May). *Energy Sector Pledge.* https://www.climatechange.vic.gov.au/victorian-government-action-on-climate-change/Energy-Sector-Pledge-Accessible.pdf

Wundervald, B., Parnell, A., Domijan, K. (2019, May) *Regularisation methods in Random Forests,* National University of Ireland Maynooth.https://brunaw.com/slides/CASI/presentation.html#1