Predictive Analytics Assignment Report

Syndicate 1 – Shuting, Yijie, Kuan, Nick, Bill

1. Executive Summary

For electricity producers, there is often a trade-off between system reliability and cost-effectiveness. From a supply-side perspective, these producers can minimize outages by setting a generous safety margin on top of the expected demand. This protects against sudden demand shocks but comes at an increased operational cost.

This ability to maintain reliability is becoming increasingly important as the world enters a new era of increasing effects of climate change that results in more severe weather events, but also new uses for electricity that were previously covered by alternative energy sources (for example, electric cars will place an additional load on electricity generators).

The solution outlined in this report approaches the problem from a demand-forecasting perspective. The idea is to develop models that predict the daily electricity demand well to establish a good baseline that energy producers can rely upon for their predictions, and but also to model the maximum electricity usage – the type of events that, if significant enough, can bring down the power grid and cause massive outages.

The solution proposed in this report contains a separate model for each of these two predictive use cases and iterates through many different approaches to arrive at the best model for each prediction objective. The performance of the model is also sufficiently good and uses a variety of model accuracy criteria for evaluation. Limitations and operational considerations are also included in this report, regarding the need for the integration of a good and reliable weather model, and outlines areas of improvement for the prediction accuracy of the model.

Looking to the future, the two models can be combined to inform both daily and strategic decisions. The output from the model can be combined with decision-making dashboards within the operating room, and the relationship between maximum and point forecasts can inform the mix of electricity generation methods used – where rapidly scalable sources of power can be used when expected demand fluctuates more.

2. Introduction

In July of 2022, many regions in China were struck by a severe heatwave. This caused the reduction of the level and strength of the Yangtze River, a major source of the country's hydroelectric power. The heatwave also made people turn on their air conditioning systems on overdrive. This culminated in both the reduction of supply and an increase in demand that resulted in the city of Shanghai making the extreme decision to turn off decorative lights along its iconic riverbank to prevent a failure of the electricity grid.

Typically, electricity companies and governments have some levers to expand the supply through increasing capacity. Long-term solutions may involve building more generation capacity but run the risk of underutilization – which may reduce operational efficiency. Another may involve making investments in flexible and scalable generation sources like wind and solar that can be ramped up or down according to demand. Alternatively, a demand-controlling approach may involve providing financial incentives to encourage certain individuals or large consumers of electricity to reduce their consumption and ease the pressure on the electric grid.

However, these interventions are costly, and may not be practical given time or resource constraints. An alternative approach involves improving planning and prediction to anticipate demand for the grid. This helps the generation plants to plan and use those costly interventions only when absolutely necessary. This is the main motivation for the development of such a model to predict electricity demand.

While predicting normal and steady state loads can help with overall operational efficiency and reduce costs for the electricity companies (which can be passed on as savings to customers), our model also accounts for predicting maximum/extreme values - times where the there is an unusually high demand for electricity. The value of capturing such missing values is to predict times when the grid may be close to capacity and give the Victorian government and energy providers the information to take action to prevent events like the one that happened in Shanghai.

3. Issues, Exploratory Data Features, and Additional Model Features

The data set after data cleaning process contains hourly information of trade price of electricity, total demand of electricity, wind speed, humidity and air temperature in VIC and range from January 2018 to June 2022. Also, three type of days indicators are included, the weekend indicator, the holiday indicator and the COVID indicator. Based on the data set, this project will focus on

discussing the relationship between total demand of electricity and those features that might affect it and forecasting the future electricity demand.

Figure 1 displays the STL decomposition for the total demand of electricity. There's still a strong seasonality in the season component. To reduce the seasonality, the multi-seasonal time series needed to be considered. Figure 2 is the decomposition for the multi-seasonal time series of electricity demand. According to the trend component, the total demand has an obviously decreasing trend before November 2021 and has an upward trend after that point. The continuously decreasing trend may due to residents' increasing environmental awareness (and transitioning to energy efficient appliances and smart home systems). As shown in Figure 1, during the pandemic period, the electricity demand is stable around 4700MW, indicating that electricity level dropped to a relatively low level. The Covid-19 lockdown brings a significantly negative impact on electricity demand with the reduction of the commercial and industrial activity. Also, total demand of electricity has a strong seasonality. From November to February is the peak season of high electricity demand every year, due to the high usage of air conditioners in summer.

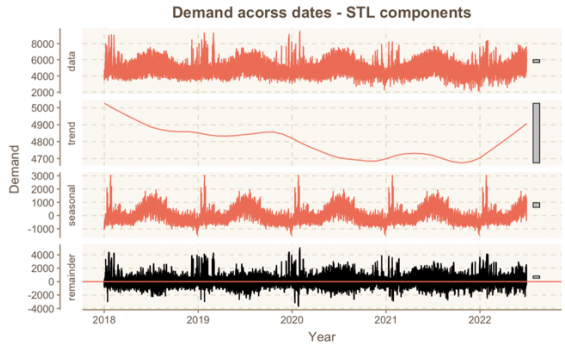

Figure 1 STL Decomposition

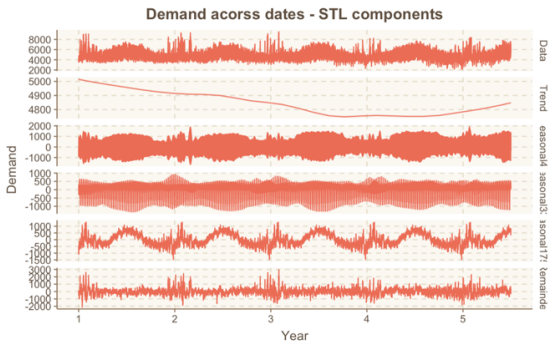

Figure 2 Multi-seasonal STL Decomposition

3.1 Price Elasticity

The trade price of electricity is an important factor when considering features that may affect the total demand of electricity. The linear relationship between total demand and trade price can be expressed as

$$TotalDemand = 0.004787 - 0.7275 TradePrice + \varepsilon$$

The price elasticity of demand (PED) can be calculated according to the equation

$$Elasticity = \beta \cdot \frac{x}{y} = -0.7275 \cdot \frac{81.36014}{4845.728} = -0.0122$$

Since the magnitude value of PED is less than 1, the electricity demand can be deduced as inelastic, which means large changes in price is accompanied by a small change of demand (consumers will

not significantly change their usage patterns to changes in price). The relationship between demand and price is also shown in Fig13 in appendix.

3.2 Date and Time Features

Features mentioned in the data set can be divided into two groups, one group being date & time features (including time of the day and type of the day), and the other group being weather features.

The trend chart shown in Figure 3 displays the variance of the total demand of electricity of 24 hours. The demand in different times of the day has some patterns from January 2018 to June 2022. The usage in the early morning period (12 AM-6 AM) decreases smoothly from around 5000MW to 4000MW. Compared to other periods, the nighttime period (5 PM-11 PM) normally has the highest demand among the different time periods. This pattern matches with the human activity pattern that people sleep in the mid of the night and come back home after 5 PM. People usually spend their leisure time with lifestyle appliances such as the TV, as well as turning on the light and air conditioning for the sleeping hours, leading the demand for electricity to peak. However, the pattern of total electricity demand has a great variance (length spanned by the points) in the lunchtime (from 11 am to 2 pm) across dates which might be attributed to the reduction of social activities during COVID-19 lockdown.

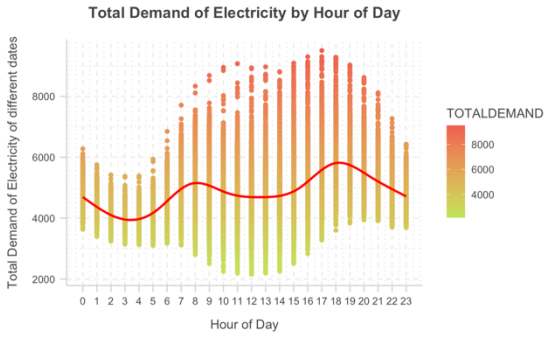

Figure 3 Relationship Between Demand and Time

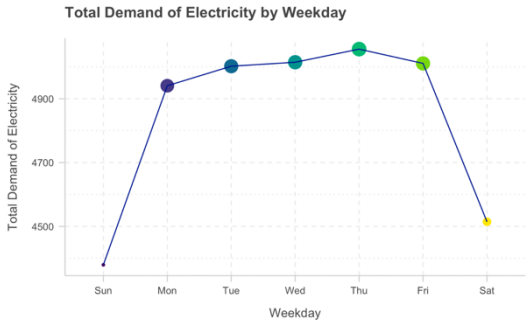

Figure 4 Relationship Between Demand and Weekday

Similar to Figure 3, Figure 4 shows the differences in average electricity demand between weekdays. We can see that electricity demand is generally lower on weekends than on weekdays (which may be indicative of the outdoor weekend leisure activities associated with the Melbourne culture). To figure out whether weekdays have a genuine impact on the total demand for electricity, we added the indicator *weekend to* divide data into weekdays and weekends. Figure 5 is the scatter plot presenting the electricity demand on weekdays and weekends. Two scatter plots have a similar trend throughout the time, again weekends normally have a lower electricity demand than

weekdays. We also illustrated the patterns of holidays or special dates. Figure 6 shows the electricity demand for holidays and normal days. While both scatter plots show similar seasonal patterns, on weekends/holidays, demand for electricity is lower, perhaps due to the reduced industrial/manufacturing activities.

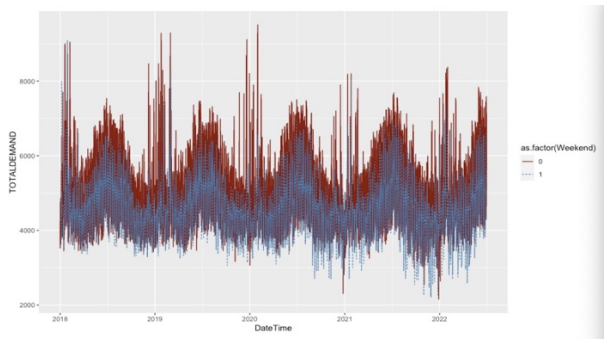

Figure 5 Relationship Between Demand and Type of Day

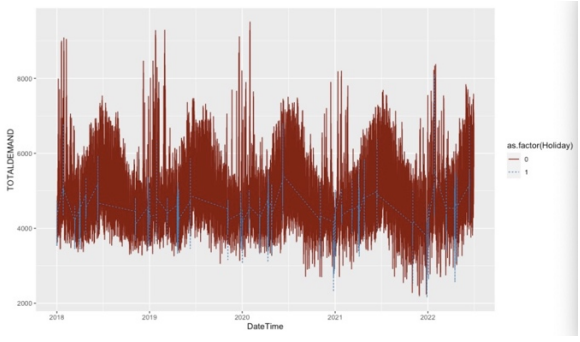

Figure 6 Relationship Between Demand and Holiday

3.3 Weather Features

Air temperature, relative humidity and wind speed are classified as weather features that may affect the total electricity demand. All three features are classified into four groups based on four quartiles to display a more clear visualization. Following figure 7, 8, 9 present the differences of total demand between different groups of air temperature, humidity, and wind speed respectively. Besides humidity and wind speed looks like having small influence on electricity demand, air temperature has a significant impact that can affect the behaviour of Melburnians to increase the settings on their air conditioning or heating appliances – which are significant components of electricity usage across households, offices, and other commercial areas.

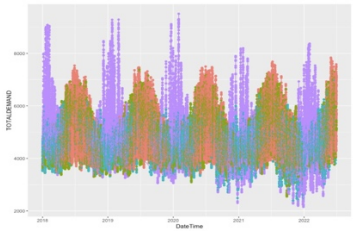

Figure 7 Air Temperature

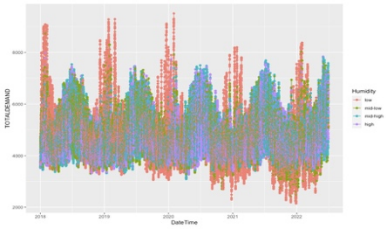

Figure 8 Humidity

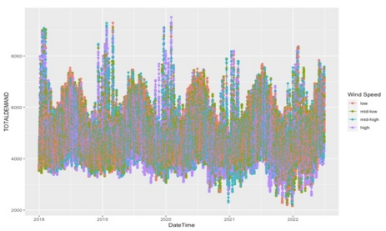

Figure 9 Wind Speed

4. Technical Analysis

4.1 Data Transformation

From the demand plot below, there is strong seasonality thus the seasonal difference is taken to the demand, and before the modelling process, and the order differencing is not needed after seasonal difference. However, given the significance of lags in the ACF diagram is decaying slowly, there is evidence that a strong level of seasonality still exists.

With the strong seasonality pattern in the data, the 'msts' series are used for the daily electricity demand data as it is a pattern of multi seasonal periods. To test the stationarity of the dataset, KPSS test and unitroot test is conducted and suggesting the first order difference should be taken. The ACF plot shown as figure 10, however, still suggesting there is seasonality that cannot be captured.

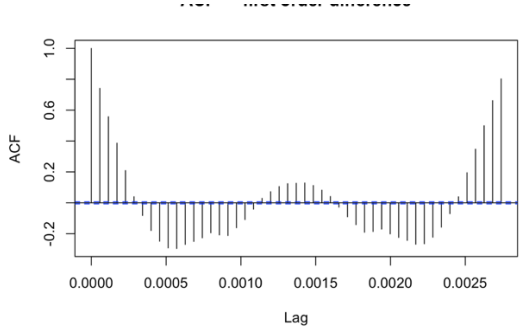

Figure 10: ACF Autocorrelation Plot

4.2 Features

During the modelling process, three set of features are used to fit the model to estimate the demand. The first set is from the result of data exploration, initially the selected features are air_temp which is one of the weather pattern indicators, weekend indicator as lower demand would be observed if it is weekend and COVID indicator as there is a sharp decrease of demand from the trend plot after COVID. The second set is all the features we care about included air_temp, humidity, wind_speed, price, weekend indicator, holiday indicator and COVID indicator. The third set is from fitting the tslm model, and by removing the insignificant variables which are air_temp and holiday indicator. Based on these three sets of features, the features would add into the modelling process after the base model is chosen.

5. Modelling Process

5.1 Model Construction and Selection

The baseline models are the time series linear models, one with all features included and the other one dropped the insignificant terms: air_temp and holiday. After dropping the insignificant terms, the AICc of the time series linear model is smaller, which is 908,599.

The first model constructed is the non-seasonal ARIMA model. For the choice of (p,d,q), d is set to 1 from the unitroot test result to make the msts data stationary. Based on the ACF and PACF plots shown in figure 11, p is first to be set as 1 as the first lag is significant. ACF plot is not that insightful in helping with determination of q. Though if both p, q need to be confirmed the ACF and PACF plots are not of much use, the modelling process begins with ARIMA(1, 1, 0).

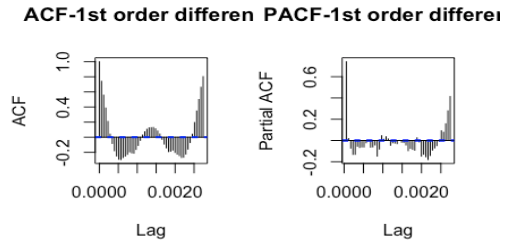

Figure 11: ACF and PACF Autocorrelation Plots

From the model selection criteria (Appendix: Table 1), the ARIMA(2,1,2) is the best model out of the six models we constructed, and all models fail the Box-Ljung test suggested all models are auto-correlated.

Then the dynamic regression model is constructed following the same sets of p,d,q we choose for naïve ARIMA models. But the dynamic regression models also try to include other information to predict the electricity demand. As stated above in the feature selection part, three sets of features are chosen in the modelling process, so in total 3*6=18 models (Appendix: Table 2, 3, 4) have been constructed. In correspondence to the naïve ARIMA model, the best in-sample performance model from using the dynamic regression to fit is still the ARIMA(2,1,2) model. And the most fit feature from three set of features is the time series model selected feature group which is the ARIMA(2,1,2) model with inclusion of COVID information, weather patterns indicators(humidity and wind_speed), price, weekend indicator.

The strong seasonality effect mentioned within the data transformation section cannot be eliminated even after conducting seasonal differencing, and due to the complexity of seasonal patterns, the Fourier transforms are used to capture the seasonality pattern in electricity demand. Then with the transforms of the data, the time series model is being fitted again. The number of Fourier terms we chose here is k = 3. Different values of k are being used to train the model, and 3 gives the smallest AICc of all k values tried, also when k = 3, the smoothness of seasonality pattern can be guaranteed. Refit with the time series model after dropping the insignificant terms, the AICc is 908559.3. Additionally, the Fourier transformation could be used to refit the dynamic regression model (Appendix: Table 5, 6, 7). From the model selection criteria statistics, ARIMA(2,1,2) is still the best performing model and out of three set of features, the time series model selected feature still gives the best performance, but with the Fourier transformation, the trained model is actually performing poorer than the model without Fourier transformation model.

Choosing from three different types of models:

Model 1: Time series linear regression model (covid, humidity, price, weekend and wind) + Fourier transformation

Model 2: Naive ARIMA(2, 1, 2)

Model 3: ARIMA(2, 1, 2) + (covid, humidity, price, weekend and wind)

From the table below, the result is that:

	AICc
Model 1	908559.3
Model 2	848555.5
Model 3	848355.7

Therefore, based on the model selection statistics AICc, the best fit of the in-sample model is the dynamic regression model with the predictors of COVID, weather patterns (humidity, wind_speed), price and weekend (an indicator variable).

5.2 Forecast assessment

To measure the performance of models, the out-of-sample forecast is weighted more than the in-sample model selection criteria when choosing models because the goal is to predict the electricity demand. The accuracy is tested against all models we proposed above.

The baseline model, time series model is constructed in two ways, with or without Fourier transformation and fit the regression model into all features and the removed the insignificant terms set of features. Checking the forecast performance accuracy, the basic time series linear regression model with all features using Fourier transformation gives the smallest RMSE and ACF1. However, the time series linear regression model with parameters excluding air_temp and holiday indicator into without Fourier transformation give the lowest MAPE and MASE (Appendix: Table 12).

For the naïve ARIMA model(Appendix: Table 8), the best forecast performance model is ARIMA(2,1,2) which is the same as the in-sample model. Checking with the dynamic regression models(Appendix: Table 9, 10, 11), the best fit model is still the ARIMA(2,1,2) and by including the other information to forecast with our demand, the first set of features which include all features give the best forecast performance, giving smallest RMSE and MAPE.

With the Fourier transformation to capture the seasonality, to refit the linear regression and the dynamic regression model, the dynamic regression models (Appendix: Table 13, 14, 15) generally have a higher accuracy in forecasting performance after the use of Fourier transformation, this is quite reasonable as the demand has high frequency of seasonality, the data is the half-hourly interval for each day, which corresponds with the choice of doing Fourier transformation. Besides above three chosen models based on AIC, one more model is added, named Model 4: ARIMA(2,1,2) + all features

	RMSE	MAPE	MASE	ACF1
Model 1	179.518798	146.8850	9.5839667	0.6074270
Model 2	180.8360	99.97252	1.9018066	0.6150588091
Model 3	161.26326	418.2992	2.1349908	0.3024101265
Model4	160.36884	414.3658	2.1300729	0.3049318218

So, to choose our model to do the point forecast, based on the in-sample model, from the model selection criteria the dynamic regression model with the predictors of COVID, weather patterns (humidity, wind_speed), price and weekend indicator would be the best fit model. But the out-of-sample forecast performance accuracy metrics gives different evaluation of the choice of model, which is the model with all features included. However, given the accuracy criteria, the RMSE of these two models do not differentiate to a large extent.

To confirm the predictive accuracy, the Diebold-Mariano Test is also done for five models (two time series linear regression models with Fourier transformation, Naïve ARIMA(2,1,2), dynamic regression ARIMA(2,1,2) with all features included and the dynamic regression ARIMA(2,1,2) with selected features), with the null hypothesis being model 1 predicts the same accurate as model 2, five models are compared and the alternative hypothesis is model 2 . And the output below suggest that though the accuracy forecast performance suggest the dynamic regression ARIMA(2,1,2) with all features would provide a better forecast result, the D-M test actually suggest that these two models have statistically equivalent accuracy.

First Model	Second Model	D-M test p-value
Model 3	Model 5	<2.2e-16
Model 3	Model 1	<2.2e-16
Model 3	Model 2	<2.2e-16
Model 3	Model 4	1

Model 5 mentioned in table is the Time series linear regression model with all features and fourier terms. By comparing the p-value for DM test, ARIMA(2,1,2) with all features and ARIMA(2,1,2) with selected significant features are deduced to have the smallest forecast performance. Given the AIC values mentioned above, the final model would be the dynamic regression ARIMA(2,1,2) with inclusion of all features (Model 4) to predict the point forecasts for the week 1–7 July 2022.

To choose the best fit model to forecast the maximum electricity demand, coverage test is applied. Four different distribution (Normal, Student t, Skewed normal, Skewed student t) are fitted into the residuals of all five models mentioned above and student t distribution gives the smallest AIC. The best fit model is chosen based on the phat value given cc-test and the statistics are shown in Table below.

	p_hat	**LR**	**p-value**
Model 1	0.0269437	171.9822	2.73e-39
Model 2	0.08499475	1906.122	0
Model 3	0.08459945	1892.827	0
Model 4	0.08505985	1910.56	0
Model 5	0.0269337	171.9822	2.73e-39

The best fit model should be the one that has the phat values closest to 0.01. Unfortunately, none of the model reached phat = 0.01. With phat equals to 0.0269337, Time series linear regression model with all features and fourier term (Model 5) is chosen as the best fit model to predict the maximum demand. However, the small p-value suggests that it fails to make 99% of the data falls within the predicted maximum demand.

6. Discussion of Key Results

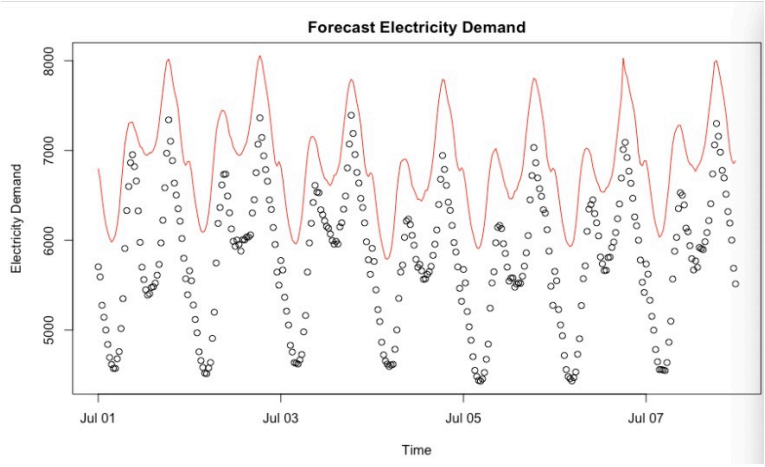

Figure 12 Forecasted Electricity Demand

ARIMA(2,1,2) model with all features without fourier term is selected to do the point forecasting and Time series linear regression model with all features including fourier term is chosen to forecast the maximum electricity demand. Figure 12 is the forecast plot of electricity demand for the following week, where the black dots represent the point forecasts, and red lines are the forecasted maximum electricity demand under 99% circumstances. The mean electricity demand is 6000, the lowest demand each day is about 4500 and the highest is about 7000. It shows that demand consumption is expected to be higher on July 1st and July 2nd or on July 7th and July 8th than on the other days. The forecasted variance is likely stable, and there is also a regular pattern in the plot in line with the hourly and daily seasonality. Above all, there seems to be no extreme irregularity in the forecast plot, for which the result could be used as a fair reference.

The difference between the peak of the point forecasts and the maximum demand also provides some valuable information. On the days where the gap is high, like on the eve of 5th of July, there is about a 14% difference between the point forecasts and the maximum demand. Practically, to avoid widespread outages, the electricity producer should consider adding a secondary source of generation that can flexibly make up this difference on demand. For example, in other nations, hydroelectric and geothermal power are often used as flexible generation sources that can be scaled up and down in a cost-effective manner. On days where this difference is expected to be high, the company can modify its operations to use these flexible sources to face unexpected shocks.

7. **Limitations and Conclusions**

However, several limitations must be considered before using this model for deployment. One limitation that is associated with the results is that the model used to generate point forecasts failed the KS test, where the residuals (while they visually resemble a normal distribution in Figure 14) are statistically not normally distributed and indicates that the model is not appropriate for performing point forecasts. While all the models tried failed the KS test, this is the best available model – and further iterating on this model with some of the ideas discussed below will be good next steps.

A key component of the model's predictive power is the incorporation of weather data for predicting electricity demand. While this is an important source of decision-making information, out-of-sample predictions would require integration with weather data source providers, or the building of a custom weather prediction engine. The model is also coupled with weather predictions, so there is an inherent risk associated with faulty predictions from the underlying

weather model. This introduces additional complexity and/or cost but is a good initiative for increasing performance.

Another limitation of this model is that there is a strong seasonality component, even after differencing to multiple orders. This indicates that there is either an unexplained component masquerading itself as seasonality or that the underlying seasonality patterns are too complex to predict perfectly using this model. This is where alternative data sources could be used to augment the model's predictive power. Building a contextual understanding of the largest consumers of power, and their usage patterns can help with predicting electricity demand. For example, while weather and office working patterns (weekdays/weekend splits) can help predict the air conditioning needs for residential and office electricity demand, having a measurable indicator for when industrial users will ramp up their usage will be key to building a better model.

Nonetheless, the models constructed in this project were able to perform sufficiently well to predict the expected maximum demand for electricity. The models we arrived at were also specialised in predicting either the regular point forecasts, or the maximum demand. Having a separate model for each component allows for each model to focus on a specific prediction task and achieve sufficiently good predictions for their respective niche. In a future where new uses for electricity gradually increase (like electric cars), this model will need to be able to consider new information to stay up to date with how electricity is being used (for example, increased travel demand due to special events)

The model that estimates maximum values will have tremendous advantages for the power generation companies as they will be able to maintain a higher power grid reliability and reduce the cost and disruption associated with outages – and the model that predicts regular point forecasts can help inform, to some extent, the baseline electricity generation level to minimise waste.

8. Appendix:

Table 1: Naive ARIMA model

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	857072.1	857072.1	857090.4	0
ARIMA(2,1,0)	856752.4	856752.4	856779.9	0
ARIMA(2,1,1)	856744.1	856744.1	856780.7	0
ARIMA(1,1,1)	848630.9	848630.9	848658.3	0
ARIMA(2,1,2)	848555.5	848555.5	848601.3	0
ARIMA(1,1,2)	856739.2	856739.2	856775.8	0

Table 2: dynamic regression With all features

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	856909	856909	856991.4	0
ARIMA(2,1,0)	856596.6	856596.6	856688.2	0
ARIMA(2,1,1)	856587.9	856587.9	856688.6	0
ARIMA(1,1,1)	848433.8	848433.8	848525.3	0
ARIMA(2,1,2)	848370.1	848370.1	848480	0
ARIMA(1,1,2)	856583.5	856583.5	856684.2	0

Table 3: dynamic regression with Data exploration selected features

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	857074.1	857074.1	857119.9	0
ARIMA(2,1,0)	856755.2	856755.2	856810.1	0
ARIMA(2,1,1)	856746.1	856746.1	856810.2	0
ARIMA(1,1,1)	848721.4	848721.4	848776.3	0
ARIMA(2,1,2)	848516.4	848516.4	848589.7	0
ARIMA(1,1,2)	856742	856742.1	856806.2	0

Table 4: dynamic regression with Time series model selected features

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	856958	856958	857022.1	0
ARIMA(2,1,0)	856646.2	856646.2	856719.4	0
ARIMA(2,1,1)	856638.5	856638.5	856720.9	0
ARIMA(1,1,1)	848600.5	848600.5	848673.8	0

ARIMA(2,1,2)	848355.7	848355.7	848447.3	0
ARIMA(1,1,2)	856632	856632	856714.4	0

Table 5: Fourier with all features

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	856921	856921	857058.3	0
ARIMA(2,1,0)	856608.6	856608.6	856755.1	0
ARIMA(2,1,1)	856599.9	856599.9	856755.5	0
ARIMA(1,1,1)	848443.5	848443.5	848590	0
ARIMA(2,1,2)	848388.5	848388.5	848553.3	0
ARIMA(1,1,2)	856595.5	856595.5	856751.2	0

Table 6: Fourier with Time series model selected features

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	856969.9	856969.9	857089	0
ARIMA(2,1,0)	856658.2	856658.2	856786.4	0
ARIMA(2,1,1)	856650	856650	856787.4	0
ARIMA(1,1,1)	848463.4	848463.4	848591.6	0
ARIMA(2,1,2)	848364.3	848364.3	848510.9	0
ARIMA(1,1,2)	856644	856644	856781.4	0

Table 7:Fourier with Data exploration selected features

	AIC	AICc	BIC	LB p-value
ARIMA(1,1,0)	857086.1	857086.1	857186.9	0
ARIMA(2,1,0)	856767.2	856767.2	856877.1	0
ARIMA(2,1,1)	856758	856758	856877	0
ARIMA(1,1,1)	848595.5	848595.5	848705.4	0
ARIMA(2,1,2)	848529.6	848529.6	848657.8	0
ARIMA(1,1,2)	856754	856754	856873.1	0

Table 81: Naive ARIMA forecast performance accuracy

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	332.9048	1240.62627	3.6872355	0.6151979490
ARIMA(2,1,0)	327.2759	1210.71503	3.6065216	0.6152089457

ARIMA(2,1,1)	328.9457	1219.60931	3.6304804	0.6152097425
ARIMA(1,1,1)	180.8361	99.97483	1.9017947	0.6150643113
ARIMA(2,1,2)	180.8360	99.97252	1.9018066	0.6150588091
ARIMA(1,1,2)	325.7832	1202.74742	3.5851000	0.6152071192

Table 92: Dynamic regression model with all features

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	171.72994	534.6041	1.8694656	0.3064102818
ARIMA(2,1,0)	171.72761	535.5560	1.8856358	0.3117966519
ARIMA(2,1,1)	171.27360	532.4522	1.8757738	0.3117966519
ARIMA(1,1,1)	161.00541	419.6592	2.1643121	0.3111540076
ARIMA(2,1,2)	160.36884	414.3658	2.1300729	0.3049318218
ARIMA(1,1,2)	172.18968	537.7154	1.8965857	0.3117937337

Table 10: Dynamic regression model with selected features using time series model

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	172.10733	537.2606	1.8698568	0.3062255251
ARIMA(2,1,0)	172.18245	538.8964	1.8870627	0.3117452258
ARIMA(2,1,1)	171.66436	535.3572	1.8766730	0.3117511247
ARIMA(1,1,1)	161.40710	422.0093	2.1577006	0.3107191816
ARIMA(2,1,2)	161.26326	418.2992	2.1349908	0.3024101265
ARIMA(1,1,2)	172.69348	541.3782	1.8985640	0.3117396064

Table 11: Dynamic regression model with selected features using data exploration features

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	172.10733	537.2606	1.8698568	0.3062255251
ARIMA(2,1,0)	172.18245	538.8964	1.8870627	0.3117452258
ARIMA(2,1,1)	171.66436	535.3572	1.8766730	0.3117511247
ARIMA(1,1,1)	161.40710	422.0093	2.1577006	0.3107191816
ARIMA(2,1,2)	161.26326	418.2992	2.1349908	0.3024101265
ARIMA(1,1,2)	172.69348	541.3782	1.8985640	0.3117396064

Table 12: Time series model performance

	RMSE	MAPE	MASE	ACF1
All features tslm model	179.588947	144.44720	9.6433223	0.608050970
Tslm model exclude air_temp and holiday	179.540275	142.62856	9.6389653	0.608050210
All features tslm model + Fourier terms	179.518798	146.8850	9.5839667	0.6074270
All features tslm model + Fourier terms	179.499281	146.4304	9.5787949	0.6074253

Table 13: Fourier transformation and dynamic regression model with all features with Fourier terms added

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	171.81136	535.2014	1.8705106	0.3064037459
ARIMA(2,1,0)	171.73591	535.6123	1.8858011	0.3118087493
ARIMA(2,1,1)	171.23385	532.1513	1.8752455	0.3117910908
ARIMA(1,1,1)	160.93239	419.4356	2.1634958	0.3111518677
ARIMA(2,1,2)	160.71920	418.3199	2.1416997	0.3066325560
ARIMA(1,1,2)	172.18478	537.6718	1.8965765	0.3118063056

Table 14: Fourier transformation and dynamic regression model with selected features using time series model significant features

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	172.72472	541.0404	1.8855030	0.307365301
ARIMA(2,1,0)	172.75189	542.2091	1.9021942	0.312760703
ARIMA(2,1,1)	172.22431	538.6222	1.8915708	0.312772755
ARIMA(1,1,1)	160.63073	417.3759	2.1577756	0.311151375
ARIMA(2,1,2)	160.78108	418.8806	2.1330369	0.304282114
ARIMA(1,1,2)	173.23589	544.4707	1.9133726	0.312781936

Table 15: Fourier transformation and Dynamic regression model with data exploration features

	RMSE	MAPE	MASE	ACF1
ARIMA(1,1,0)	172.15984	537.6365	1.8705572	0.306226364
ARIMA(2,1,0)	172.14497	538.6239	1.8865626	0.311746463
ARIMA(2,1,1)	171.52555	534.3536	1.8745019	0.311748881
ARIMA(1,1,1)	161.43248	421.7771	2.1687952	0.311038636
ARIMA(2,1,2)	161.18257	420.2530	2.1423593	0.305378664
ARIMA(1,1,2)	172.65617	541.1102	1.8980856	0.311737380

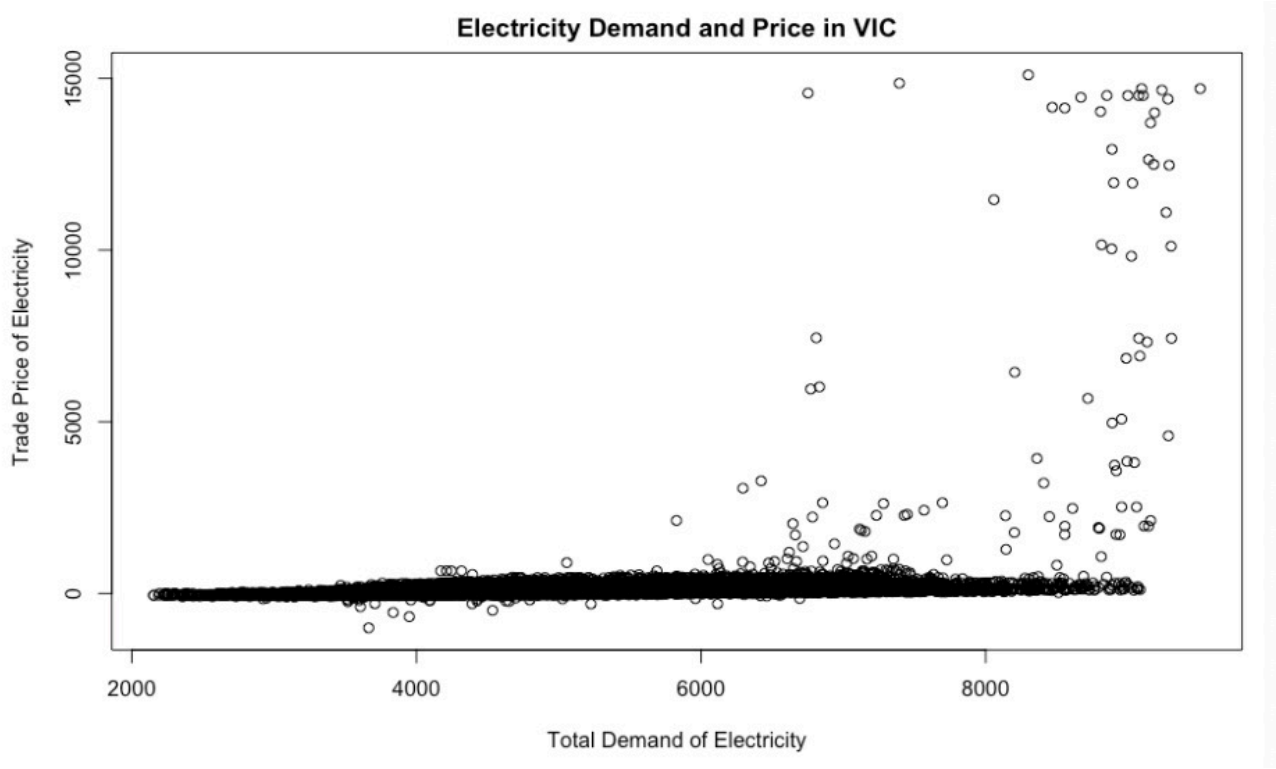

Figure 13 Relationship between demand and price

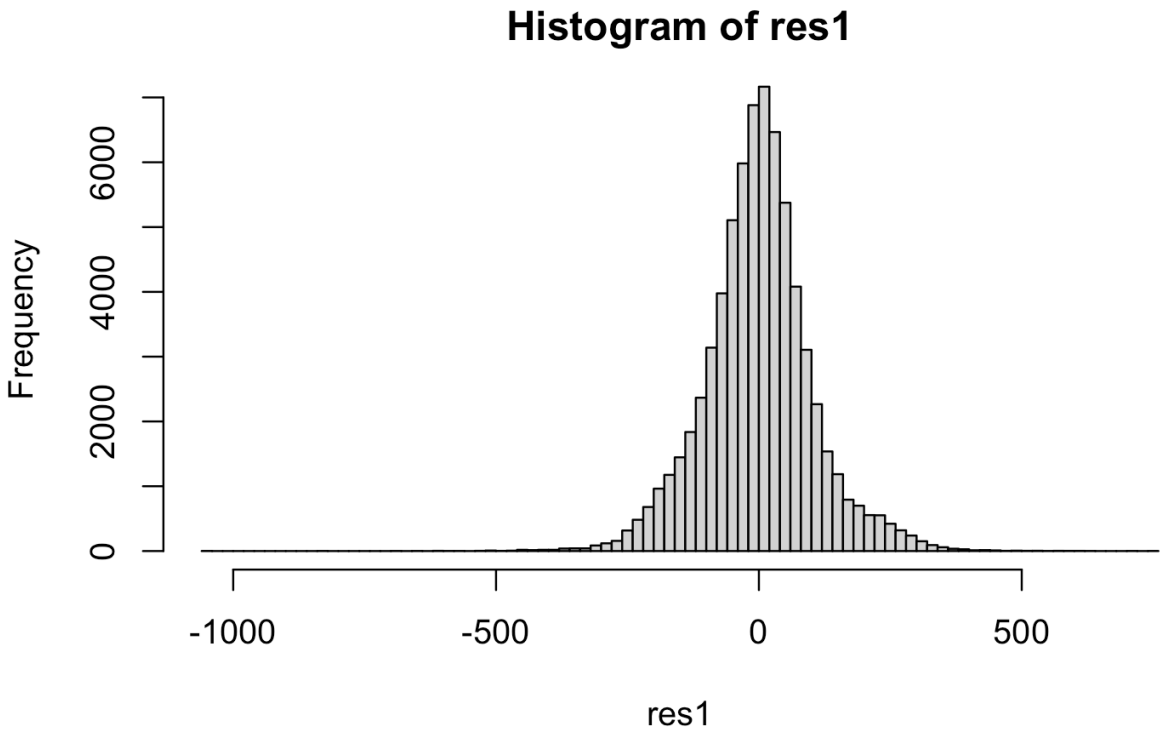

Figure 14: residual plot of the final model

Reference for Introduction & Motivating Idea: https://www.france24.com/en/asia-pacific/20220822-lights-out-for-shanghai-s-bund-as-china-heatwave-sparks-power-cuts