A REGRESSION-BASED LONG-RANGE FORECASTING MODEL FOR IRELAND'S ELECTRICITY CONSUMPTION

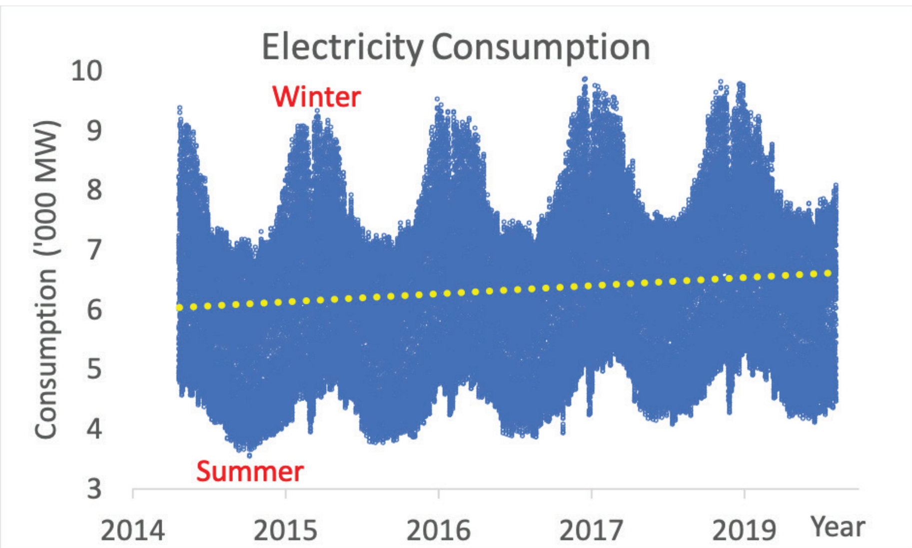
INTRODUCTION

Forecasting energy demand is vital and has only gotten more important over the recent decades as our natural resources get depleted and GHG emissions increase. Industry, as well as government agencies, need accurate energy forecasting systems to manage energy supply and demand.

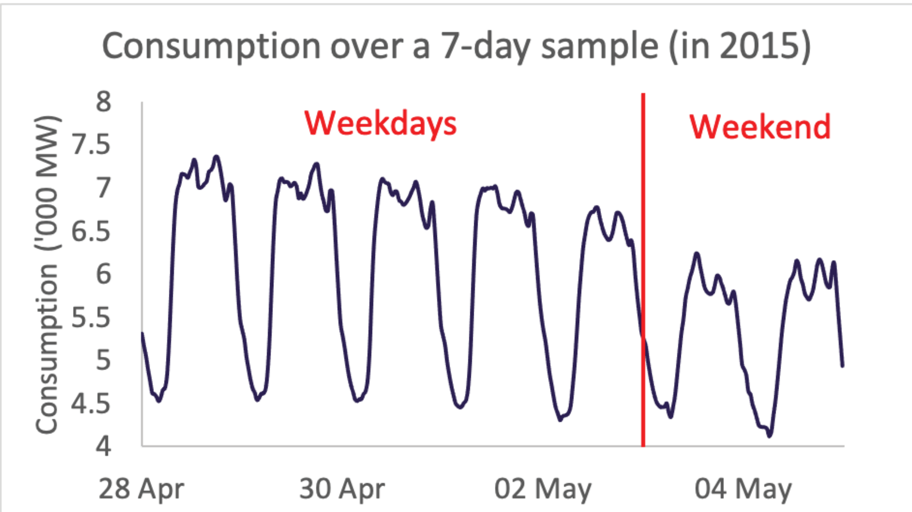
The goal of our project is to develop a novel forecast model to accurately predict the long-term electricity consumption based on multiple attributes: *date*, *time*, and *temperature*. The model developed is tested on unseen data covering a 3-month period and its accuracy has been verified.

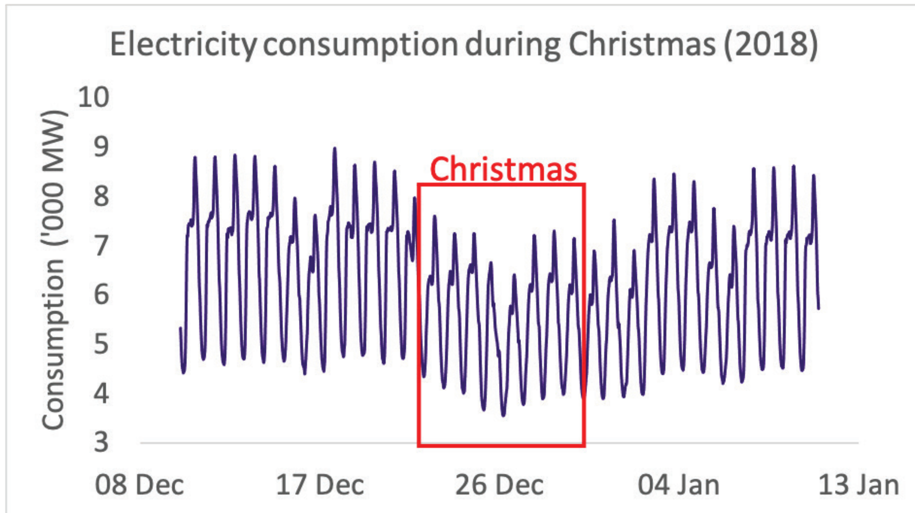
OBSERVATIONS

The proposed model is founded on some basic observations on the Irish electricity consumption data. The most basic underlying pattern in the Irish data is the positive trend and seasonality, as can be clearly seen in the below figure.



An investigation of the data reveals that there are other inherent patterns in the electricity consumption data due to three other attributes: *day of the week*, *time of the day* and *special days*.





METHOD & ASSUMPTIONS

The forecasting method we propose is based on the identification of these attributes. In other words, we decompose the time-series into its basic components. To serve this purpose, we follow the below steps:

1. Remove the *trend* from the data (i.e. detrending).
2. Identify the effects of *time* and *temperature* on electricity consumption.
3. Quantify the *day of the week* effect.
4. Assess the impact of various *holidays* on demand.

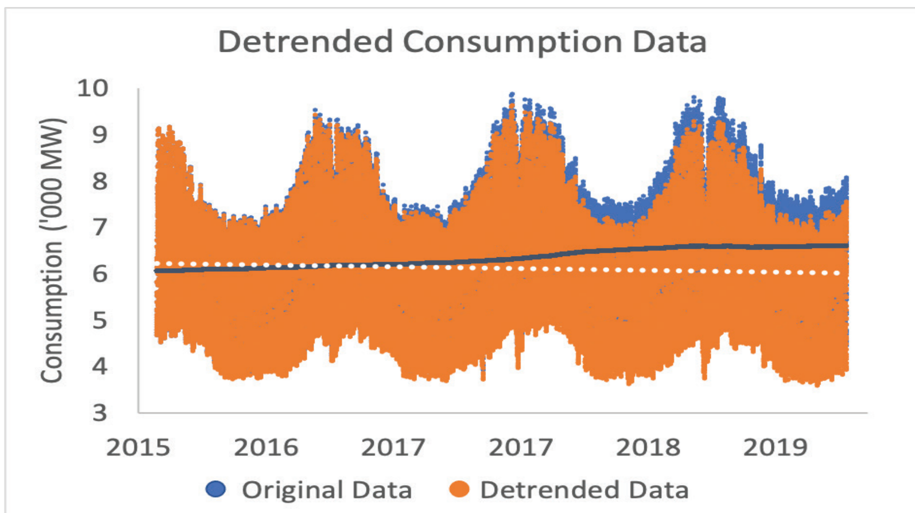
It is assumed that the above attributes are the only determinants of electricity consumption apart from random noise.

APPLICATION

Data Sources: • Temperature Data - Met Éireann
• Electricity Demand - EirGrid

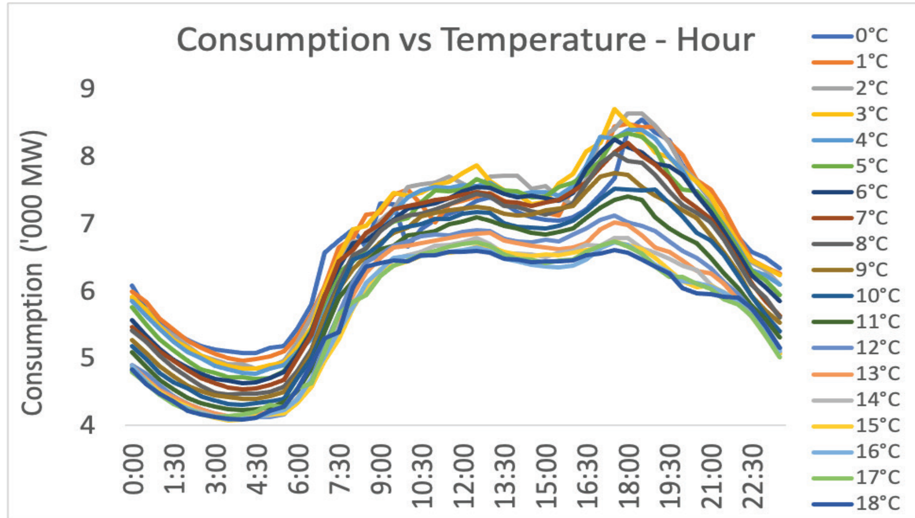
1. Detrending the data

We calculate the moving average of consumption with a length of one-year and subtract it from the original data to remove the trend.

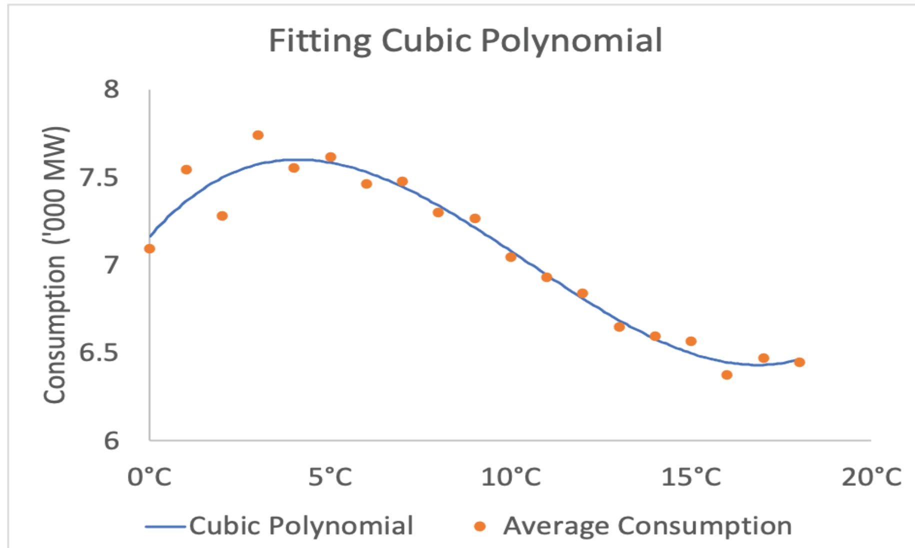


2. Identifying the effects of *time* and *temperature*

The below plots clearly show that demand is a function of *time* and *temperature*.



We fit a cubic polynomial, $f_h(t)=a_0+a_1t+a_2t^2+a_3t^3$, for each half-hour and determine the best coefficients $a_{0,1,2,3}$. In this notation, h represents *time* periods (i.e. half-hours) and t represents *temperatures*.



3. Quantifying the *day of the week* effect

The below table presents the calculated **correlations** between the days of the week. The results show that weekdays have almost perfect correlations, while the correlations of Saturday and Sunday with the weekdays are relatively low.

Correlation Table

	Mon	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*
Tue	0.999					
Wed	0.998	1.000				
Thu	0.998	1.000	1.000			
Fri	0.994	0.997	0.997	0.997		
Sat	0.971	0.965	0.963	0.964	0.963	
Sun	0.928	0.912	0.910	0.911	0.898	0.976

To further investigate the differences between the mean values of weekday-consumptions we use **One-Way Anova** for testing the below hypothesis

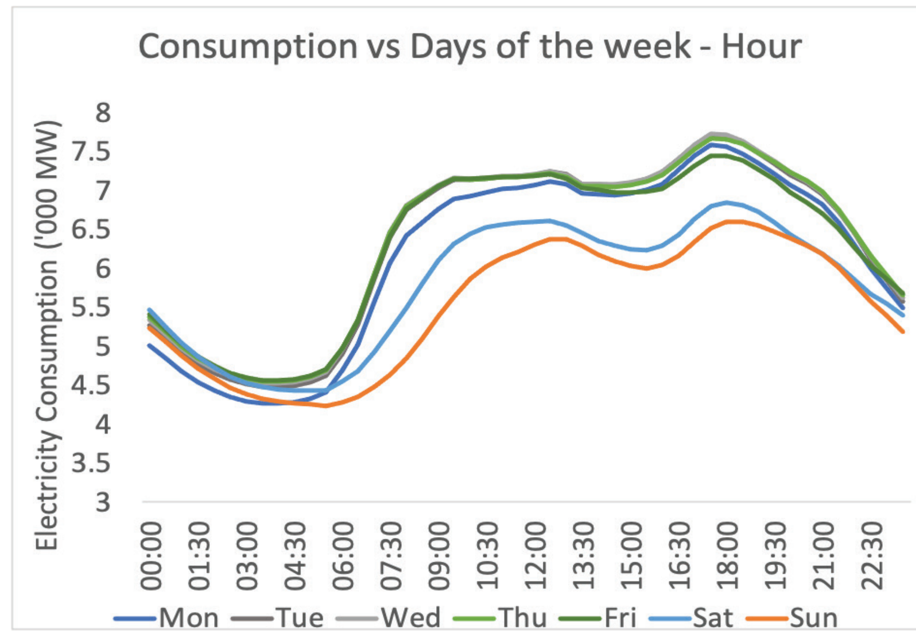
$$H_0: \mu_{Mon} = \mu_{Tue} = \mu_{Wed} = \mu_{Thu} = \mu_{Fri}$$

$$H_A: \textit{the means are not all equal}$$

The observed P-value is 0.875 (at α=0.05). Therefore we conclude that there is not enough support to reject H_0 and retain the hypothesis that weekday-consumptions can be treated in the same manner.

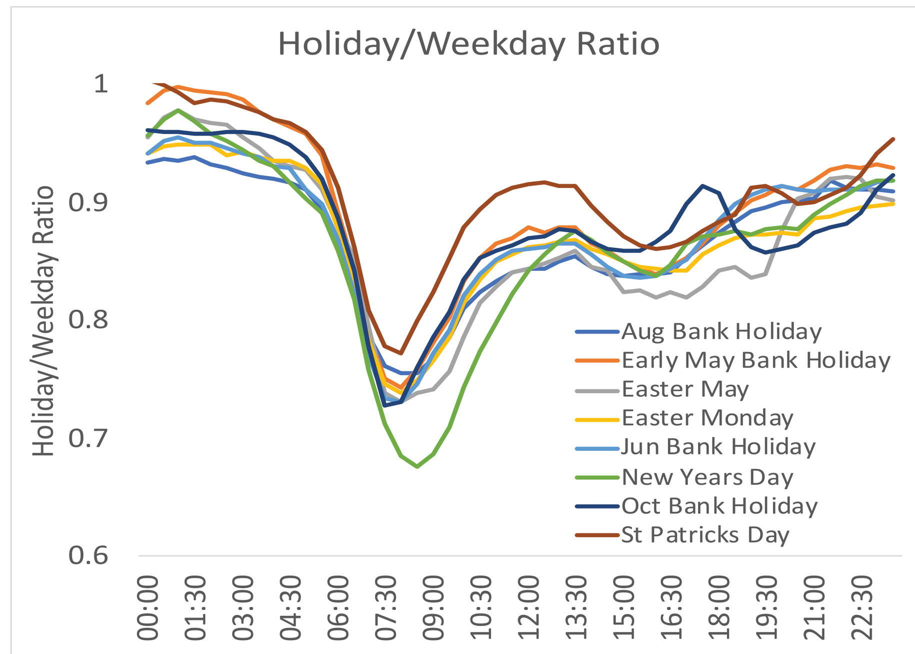
When we add *Saturday* to the set of weekdays, the P-value is 0.056, which is a border value. And for *Sunday*, a similar analysis yields a P-value of 0.0005, which clearly indicates that there is enough support to reject H_0 and provisionally accept H_A.

The average consumption for each half-hour, for any given day of the week, is plotted in the following graph, and as expected, plots are clustered into three groups: *Weekdays*, *Saturday*, *Sunday*. In order to quantify the *day* effect, we calculate the ratios of the electricity consumptions on Saturday and Sunday to an average weekday at each half-hour.



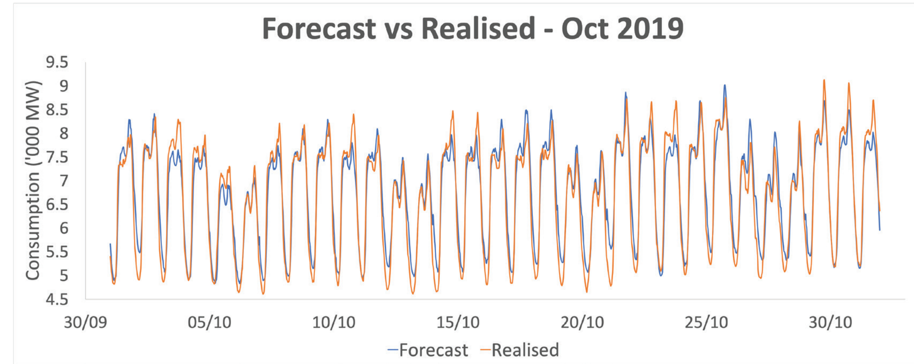
4. Assessing the impact of *holidays*

To identify the impact of holidays on demand, we compare average demands during holidays with weekdays. We observe a decrease of up to 30% at certain hours. As previously described, we calculate the ratios for each holiday at each time period.

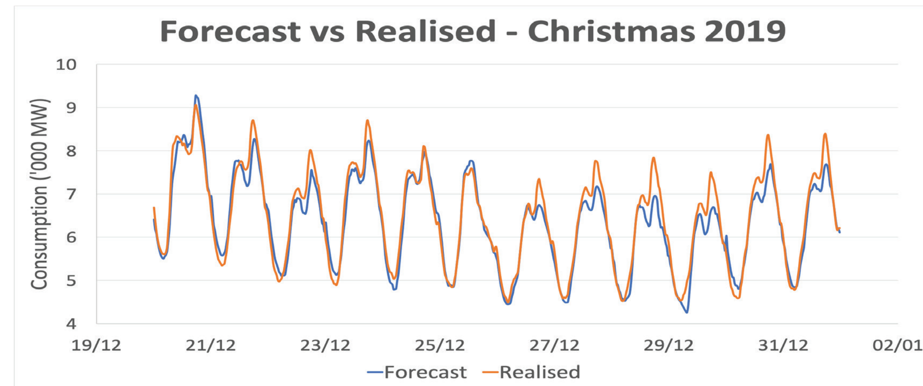


RESULTS & CONCLUSION

The described forecast model is implemented in the **statistical programming framework R**. The electricity consumption and temperature data between Jan 2015 - Sep 2019 are used for training the forecast model. The resultant model is tested on the unseen data for the 3 month period covering Oct 2019 - Dec 2019. The below graph depicts the actual electricity consumption and the superimposed forecasts for Oct 2019.



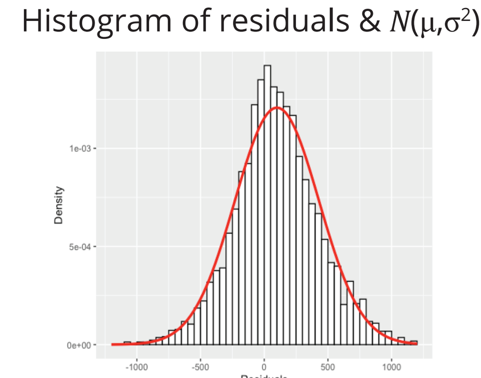
The below graph provides an indication of the long-term performance of the proposed forecast model over the Christmas week in 2019. Note that the training data used in building the forecast model do not include any observations after the end of September 2019.



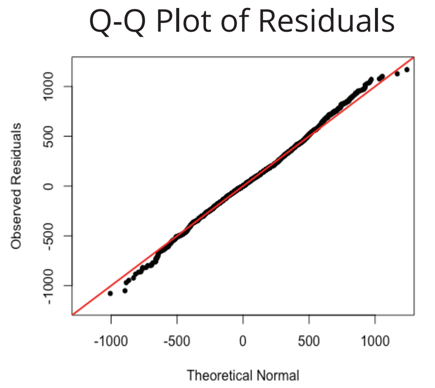
The forecast accuracy of the model on the unseen data, over Oct-Dec 2019, is **3.8%** with respect to **MAPE** (*mean absolute % error*).

Our investigation on **residuals** reveals that they follow a **normal distribution** which is also confirmed by the **Q-Q Plot**.

Histogram of residuals & $N(\mu,\sigma^2)$



Q-Q Plot of Residuals



Although the residual mean is close to 0 (i.e. $\mu_{Residual}$=99.3 MW), indicating an unbiased estimator, we believe that the model can still benefit from the introduction of additional explanatory economic variables, which will be investigated in the future.