UNC CHARLOTTE

The WILLIAM STATES LEE COLLEGE of ENGINEERING

EMGT 6965

Energy Analytics

Proposal Final Report

Submitted by: Mahesh Alapati

I. INTRODUCTION

Electricity is one of the commodity that cannot be stored in a much efficient way, it has to be produced as it is consumed, so it is important for electricity producers to able to predict daily customers' electricity consumption as accurately as possible, if it is underestimated, factors like reliability and security will be affected to the energy suppliers. On the other side, if customer's consumption is overestimated, which leads to costly operation of the plant and make losses to the supplier.

The most of the literature and techniques to forecast load is devoted on hourly load consumption with temperature as a key driving factor. Although there are many other weather variables such as wind speed, precipitation amount, solar radiation, humidity and cloud cover, affects electricity demand but not been studied thoroughly in the load forecasting literature.

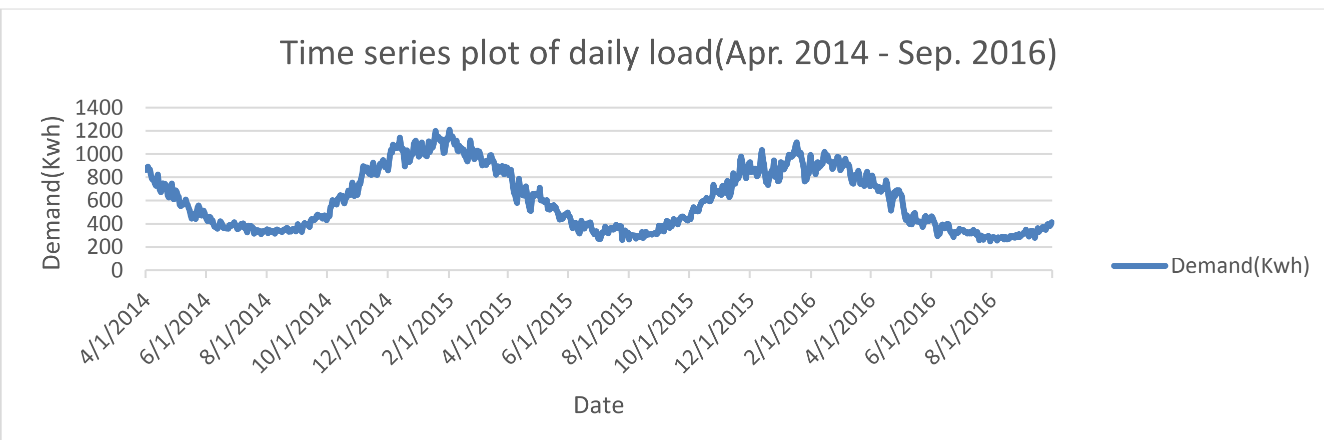
In this paper, I would like to examine influence of above mentioned variables in electricity demand and like to propose a methodology to forecast daily load using data of **Npower Forecasting Challenge 2017**.

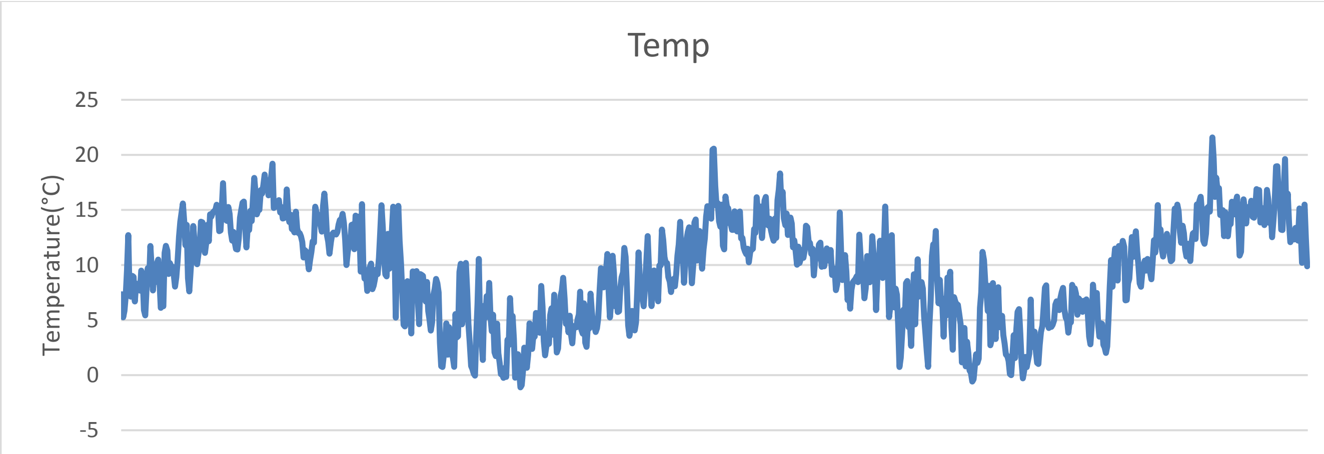
II. DATA

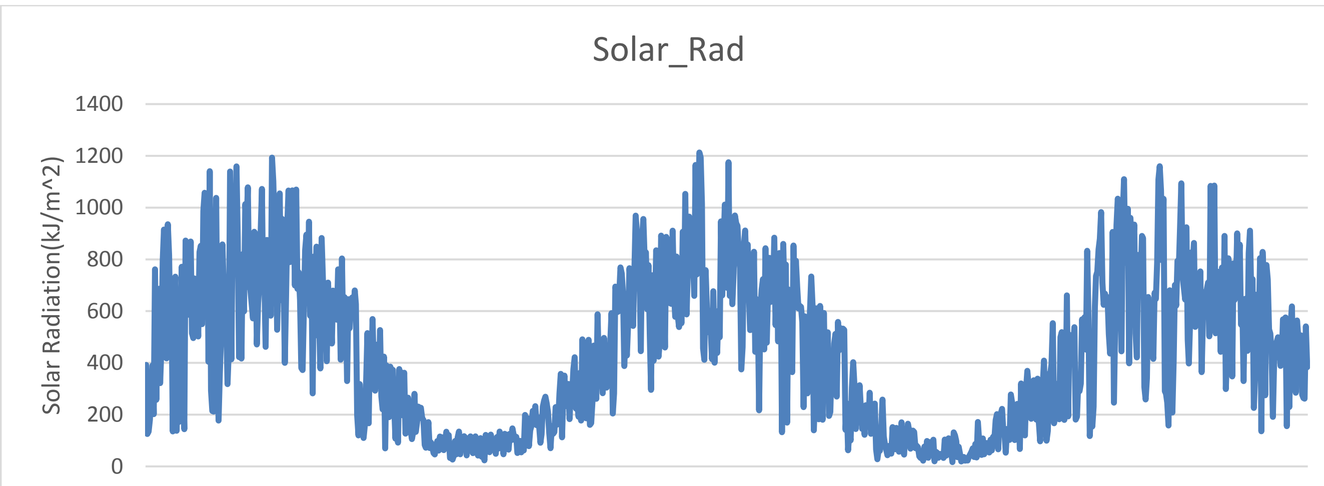
A. Data Description

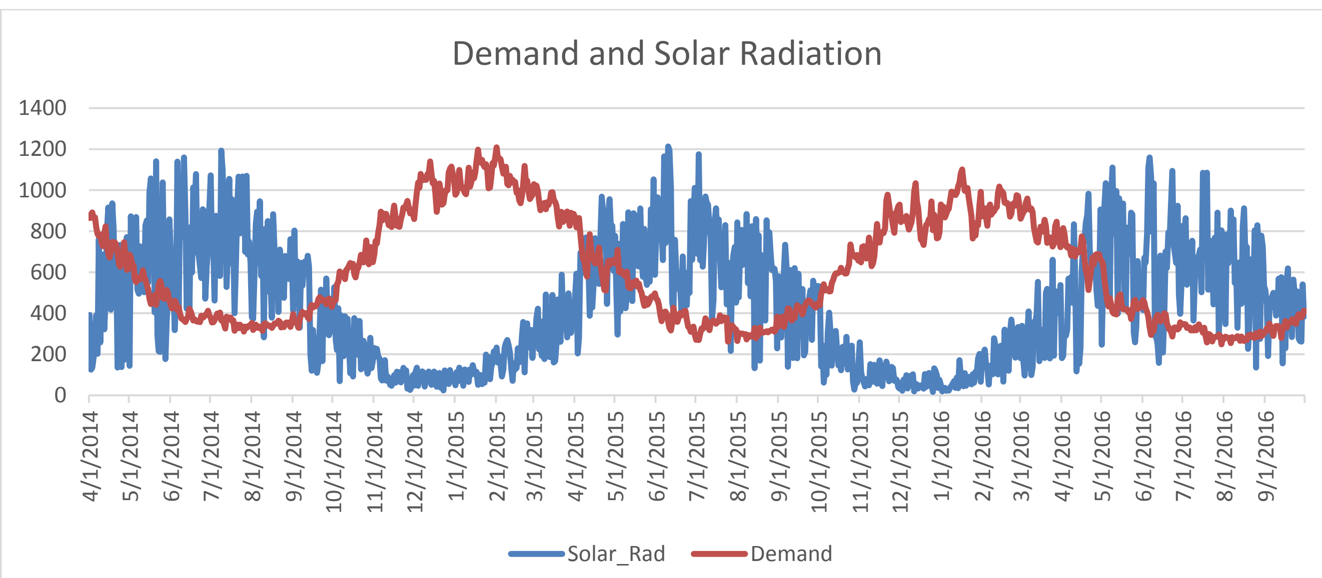
The data used in the case study includes 2.5 years of daily load and weather data from April 2014 to September 2016. The weather data includes temperature, wind speed, precipitation amount, solar radiation, humidity and clover cover from the location North East England.

Figure 1 shows the actual daily consumption data from 4/1/2014 to 9 /30/2016




Temp
Temperature(°C)
25
20
15
10
5
0
-5


Solar_Rad
Solar Radiation(kJ/m^2)
1400
1200
1000
800
600
400
200
0


Demand and Solar Radiation
1400
1200
1000
800
600
400
200
0
4/1/2014
5/1/2014
6/1/2014
7/1/2014
8/1/2014
9/1/2014
10/1/2014
11/1/2014
12/1/2014
1/1/2015
2/1/2015
3/1/2015
4/1/2015
5/1/2015
6/1/2015
7/1/2015
8/1/2015
9/1/2015
10/1/2015
11/1/2015
12/1/2015
1/1/2016
2/1/2016
3/1/2016
4/1/2016
5/1/2016
6/1/2016
7/1/2016
8/1/2016
9/1/2016
Solar_Rad
Demand

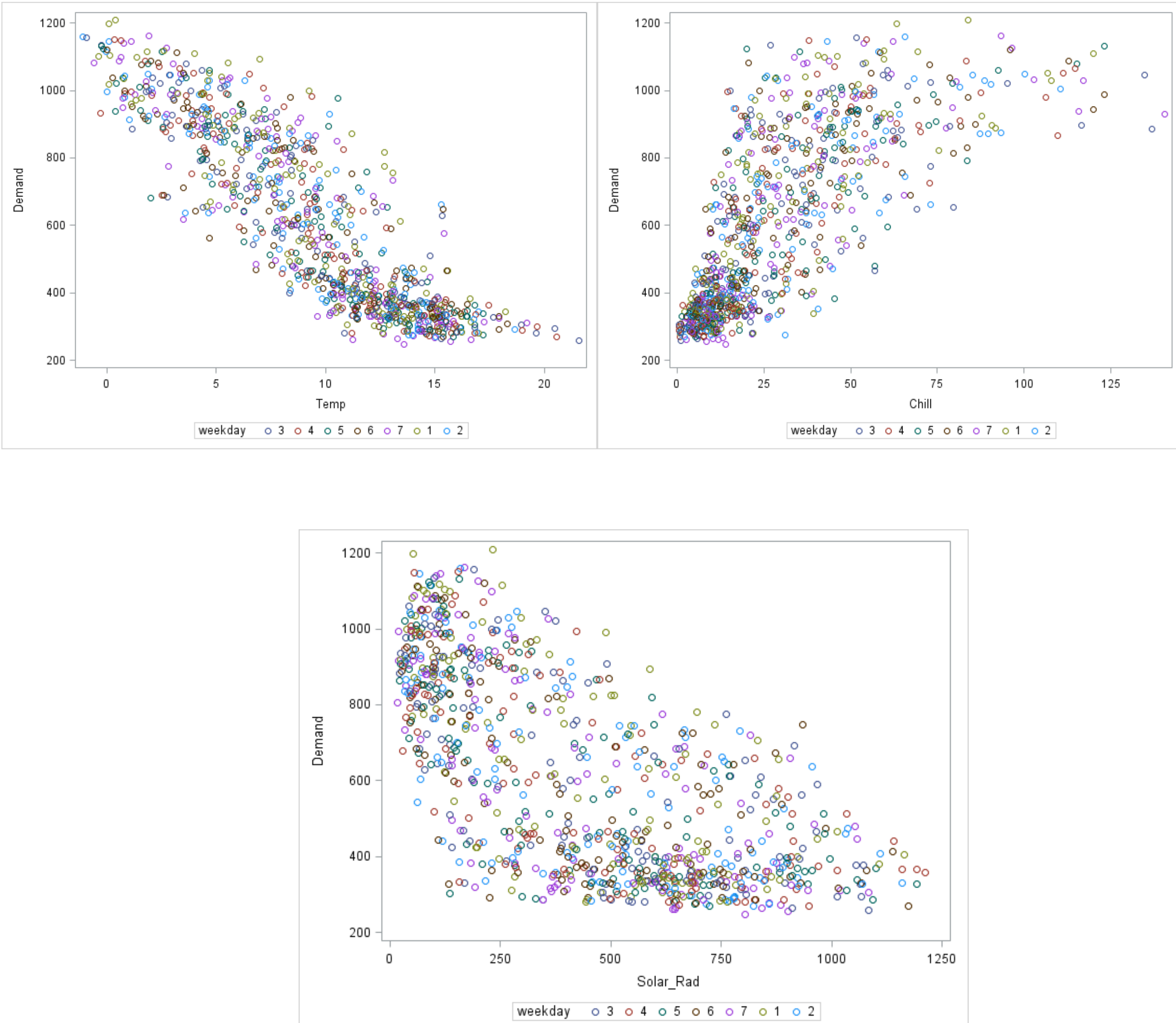


Figure 2. Scatter plots of daily energy consumption (Kwh) and weather variables by weekday

	Temp	Wind Speed	Precip Amount	Chill	Solar Rad	Humidity	Cloud Cover	Demand
Temp	1							
Wind Speed	-0.22	1						
Precip Amount	-0.11	0	1					
Chill	-0.76	0.71	0.09	1				
Solar Rad	0.4	-0.06	-0.29	-0.31	1			
Humidity	-0.27	-0.09	0.43	0.16	-0.72	1		
Cloud Cover	-0.22	-0.09	0.26	0.13	-0.67	0.67	1	
Demand	-0.81	0.22	0.04	0.7	-0.37	0.3	0.22	1

Table 1 Correlation coefficient between response variable and predictor variables

B. Explanatory Data Analysis

From Fig. 1 shows demand varies per seasonal pattern according to UK. Demand is high from December to February because of winter season. However, demand gradually decreases from March to August even though it is summer, because given location might be dominated by solar energy production.

From time series plot of demand and solar radiation, it is shown that if sunshine is more, then there is less demand of electricity. So, in this case study, solar radiation variable plays very prominent role in forecasting consumer's electricity on daily. Correlation coefficient between response variable and predictor variables are tabulated in table 1.

Fig. 2 shows the scatter plot of temperature vs. load, chill vs. load and solar radiation vs. load from the given data respectively. The reason to select only three variables (temp, chill, solar radiation) because there is strong similarity observed with the demand.

The major differences between Temperature and Chill are listed below:

- Temperature

➢ The measure of degree of hotness or coldness of a body
➢ Increase in temperature in summer will result in increase in load and decrease in temperature will result in decrease in not only average daily load but also will lower the peak demand.

- Wind chill index

➢ The warmer air surrounding the body then acts as insulator preventing the further heat loss. But if the wind blows then the colder air takes the place of the warmer air thus causing further heat loss.

III. MODELLING PROCESS

In this study, two techniques have been implemented such as Multiple Linear Regression (MLR), and Artificial Neural Networks (ANN). Among the above, MLR analysis has been widely used in load forecasting. Below sections discuss its implementation.

The length of data used to for parameter estimation and error statistic calculation:

Testing	**Validation**	**Forecast**
April 2014 -Mar 2015 & Oct.2015 – Mar. 2016	Apr. 2015 – Sep. 2015	Apr.2016 – Sep.2016

Table 1 Length of data used for calculating MAPE

A. *Multiple Linear Regression*

The general linear regression model can define as**:**

$$E(Load) = \beta 0 + \beta 1 \times X\,1 + \beta 2 \times X\,2 + \beta 3 \times X\,3 + \ldots\ldots\ldots + \beta p - 1 \times Xp - 1 + ei$$

Where β0… βp-1 are the coefficients, X1, X2... Xp-1 are the predictor variables. In this paper, they are referred as Y= Load, X1, X2 as weather variables such as temperature, solar radiation etc.

In the model selection process, first forward selection strategy is followed. The first model with all quantitative variables such as temperature, wind chill, solar radiation and interactions between weekday and temperature variables. The model specified as below:

DEPENDENT VARIABLE: Load
CLASSIFICATION VARIABLES: Hour, Month, Day, Weekday
QUANTITATIVE VARIABLES: Temp, Tlag 1-8, Chill, Chilllag1-7
MAIN EFFECTS: Hour, Month, Day, Temp, Chill, Weekday, Tlag 1-8, Clag1-8.

B. *Artificial Neural Networks*

ANN approach is used to perform nonlinear modeling for the relationship between the load and weather variables. The algorithm is tested using npower forecasting challenge data. The algorithm has been implemented in MATLAB tool.

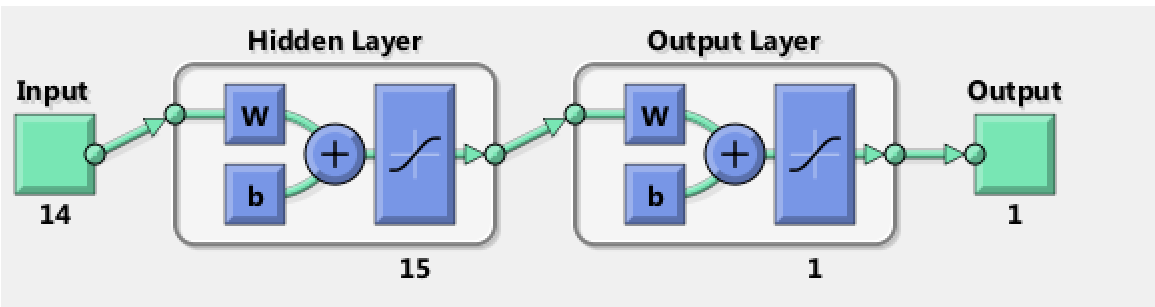


Figure 1 ANN structure in Matlab

The input variables include: Temp, Wind Chill, Weekday, Day, Weekend, Day and Year. Only weekday has been modified as categorical variables in the MATLAB.

IV. Results and Conclusion

In this case study, two different techniques have been implemented for preliminary analysis and chosen combination forecast produced better forecast results compared to individual forecasts for this npower forecast challenge 2017 of round 1 data based on a percentage of MAPE among other models.

Model	MAPE (Validation) %	MAPE (Forecast)%
MLR	14.13	11.8
ANN	8.92	7.93
Combination	8.52	7.23

Table 2 Result Analysis of two models

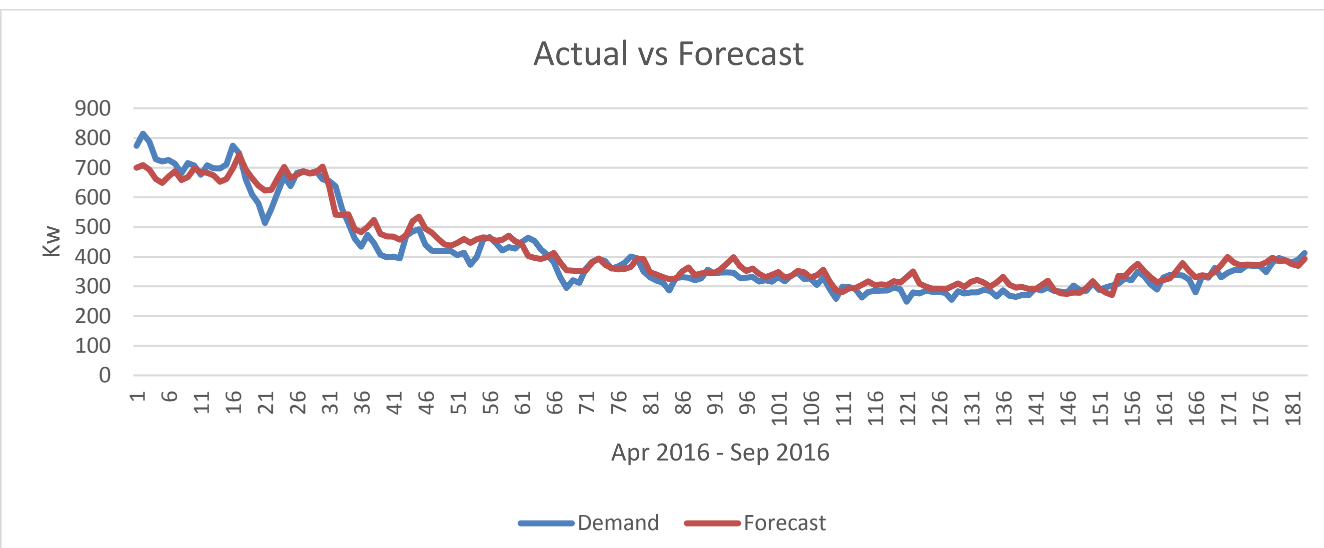


Figure 2 Graphical representation of actual vs forecasted demand

V. Conclusion

1. ANN or combination forecast is the best fit model for the npower Round-1 data.
2. Need to know why MLR is not giving better forecast.
3. It is always good to do exploratory analysis on predictor variables vs response variable.

VI. References

- Tao Hong, Pu Wang and H. Lee Willis," A Naïve Multiple Linear Regression Benchmark for Short Term Load Forecasting," Power and Energy Society General Meeting, 2011 IEEE.
- Rothe, Wadhwani and Wadhwani," Short Term Load Forecasting Using Multi Parameter
- Regression" International Journal of Computer Science and Information Security, Vol. 6, No. 2, 2009.
- https://www.npowerjobs.com/graduates/forecasting-challenge
- https://www.mathworks.com/help/stats/classificationknn-class.html
- Hesham K Alfares, Mohammad Nazeeruddin," Electric load forecasting: literature survey and classification of methods", International Journal of Systems Science.
- Tao Hong, Hong, T. (2010). Short-term electric load forecasting. North Carolina State University.
- Lee and Cha "SHORT-TERM LOAD FORECASTING USING AN ARTIFICIAL NEURALNETWORK", Transactions. Stems, Vol. 7. No. 1, February 1992