Forecasting United Kingdom Electricity Demand

Aditya Prabaswara Mardjikoen*

School of Mathematics, University of Edinburgh, Edinburgh, U.K.

* A.P.Mardjikoen@sms.ed.ac.uk

1. Introduction

The United Kingdom (UK) increased production of renewable energy (wind, solar and biomass) and reduced production of fossil fuels [1]. Nevertheless, renewable energy sources (such as wind, wave and solar) cannot be predicted and controlled one day in advance [2]. Further, technological advancements such as electric cars and the smart grid make forecasting electricity demand difficult. Recently, the study of electricity load forecasting [3, 4, 5] has garnered a lot of interests from academics and businesses alike. Electricity supply management is vital to meeting demand at all times. Our goal is to build a model for forecasting UK electricity demand one day in advance for each half-hour of the day using the UK half-hourly electricity load data.

2. Data

The dataset was collected by the UK national grid operator and contains 90419 observations of electricity load (in megawatts) and average daily temperature (in Celcius), which were measured every half hour from January 2011 to June 2016. There are no data on electricity load and average daily temperature during the Christmas/New Year holiday period.

3. Model

- **(Response)** Define L_i as the grid load in megawatts at the ith half-hour period.
- **(Covariates)** Let toy = time of year (as proportion), t = cumulative time (scaled to [0,1]), day = day of the week (integer 1-7), month = month of the year (integer 1-12), $\overline{T}$ = average daily temperature, year = year, and θ = exponential smoothing of the real temperature T_i. We define $\theta_i = \sum_{j=1} T_{i-48j}(0.95)^j$ with T_{i-48j} is T_i at the ith half-hour period j days before.
- **(Penalized Regression Spline)** We denoted $f_q(x_{q,i})$ as the smooth functions for the covariates x_q. f_1, f_3, f_6, f_7, f_8 are cubic regression spline; f_2, f_4, f_5 are cyclic cubic regression spline; f_9 is tensor product of cubic regression splines (cyclic in L_{i-48}); f_{10} is tensor product of cubic regression splines (cyclic in $\overline{T}_i$); f_{11} is tensor product of cubic regression splines (cyclic in θ_i).
- **(Generalized Additive Model)** We denoted ϵ_i as the model error term and proposed the following generalized additive model (GAM) [6, 7].

Short-Term Electricity Load Forecasting Model

$$L_i = f_1(L_{i-48})\text{day}_i + f_2(\text{toy}_i) + f_3(t_i) + f_4(\text{day}_i) + f_5(\text{month}_i) + f_6(\overline{T}_i) + f_7(\theta_i)$$
$$+ f_8(\text{year}_i) + f_9(\text{day}_i, L_{i-48}) + f_{10}(\text{day}_i, \overline{T}_i) + f_{11}(\text{day}_i, \theta_i) + \epsilon_i,$$
$$\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2).$$

4. Model Evaluation

Let N and $\hat{L}_i$ be the number of observations estimated in the forecasting period and the forecasted L_i respectively. We use the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) metrics for model evaluation [8]:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{L_i - \hat{L}_i}{N}\right|$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(L_i - \hat{L}_i)^2}$$

We used the dataset from January 2011 to December 2015 as the training dataset, and dataset from January 2016 to June 2016 as the test dataset. The model was fitted using the training dataset at each 48 half-hours, while the test dataset is used for model evaluation.

5. Results

Our analysis suggests:

- The forecasted error during summer is the highest between midnight until morning.
- The highest forecasted error through the rest of the half-hour period of the day happened during spring.
- Even though the MAPE during weekdays is higher than during weekends, however the RMSE during weekends can be higher than during weekdays in the morning.
- The RMSE in Friday is the highest over the entire half hour period of the day.
- The MAPE during morning is the worse in Sunday.
- The average predicted error and the typical error is roughly around 1.68% and 774 megawatts respectively between January 2016 until the end of June 2016.
- The average predicted error and the typical error is roughly around 1.14% and 572 megawatts respectively between January 2011 until December 2015.

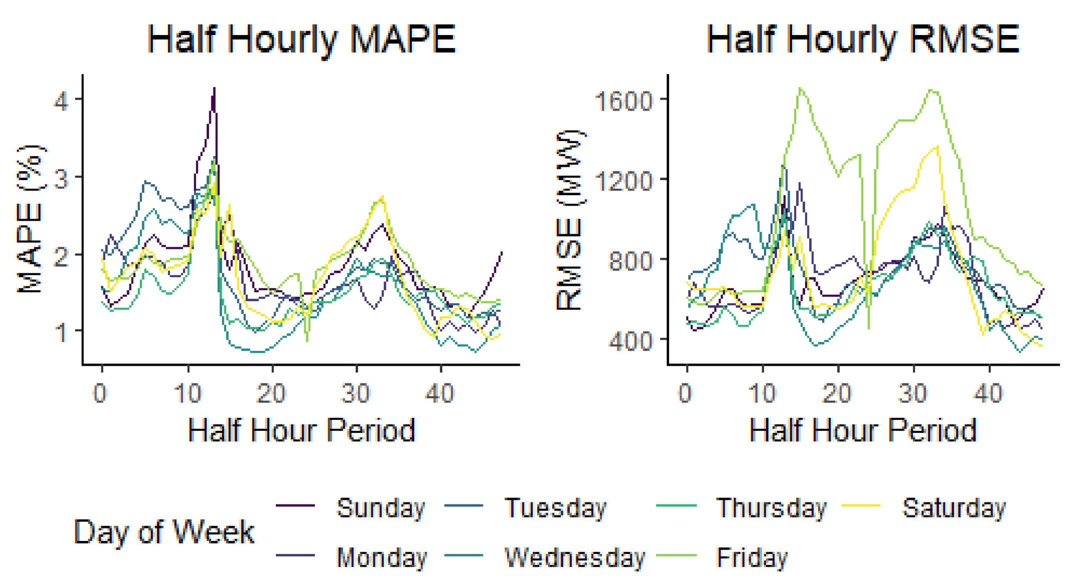

Fig 1. **Half-hourly residuals based on day of week.**

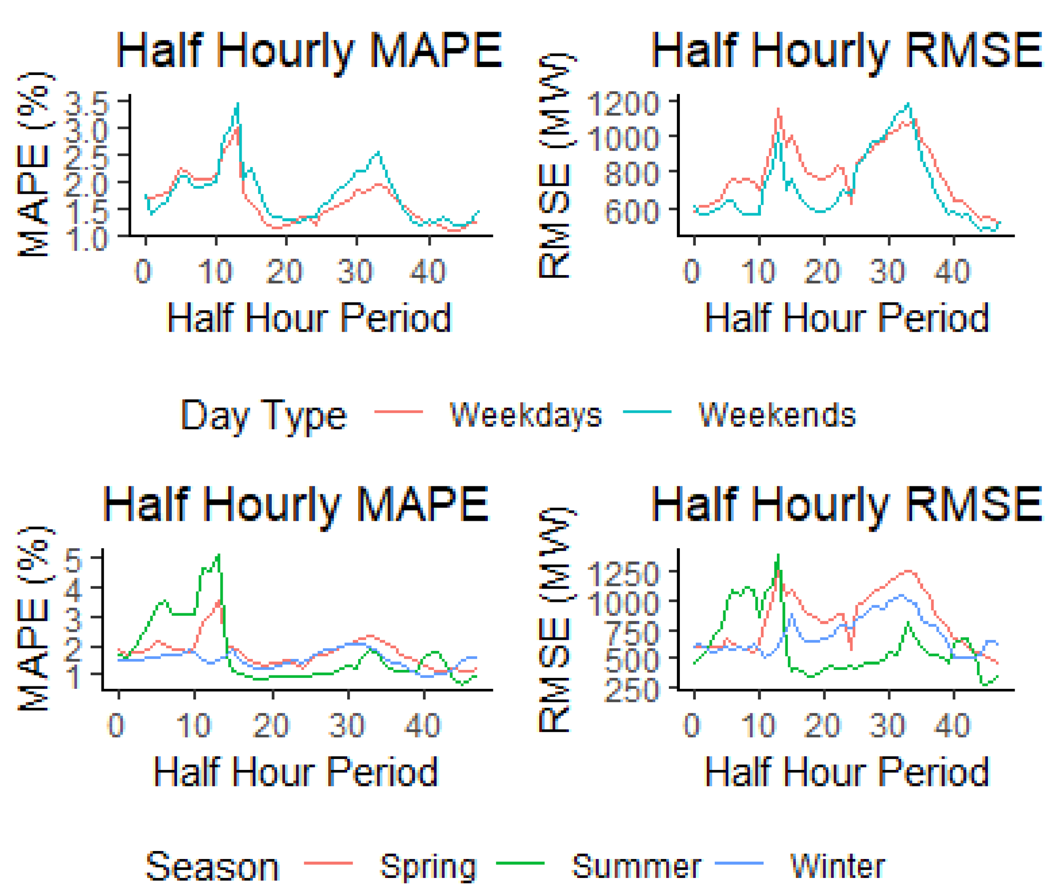

Fig 2. **Half-hourly residuals based on day type and season.**

Dataset	MAPE %	RMSE (MW)
Training	1.14	571.78
Test	1.68	774.27

Tbl 1. **Model fit in the training and test datasets.**

6. References

[1] IA Grant Wilson, J. Sharma, S.and Day, and N. Godfrey. Calculating great britain's half-hourly electrical demand from publicly available data. *Energy Strategy Reviews*, 38:100743, 2021.

[2] G. Boyle. *Renewable electricity and the grid: The challenge of variability*. Earthscan, 2009.

[3] S.N. Wood, Y. Goude, and S. Shaw. Generalized additive models for large data sets. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 64(1):139–155, 2015.

[4] Y. Goude, R. Nedellec, and N. Kong. Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE Transactions on Smart Grid*, 5(1):440–446, 2014.

[5] A. Pierrot and Y. Goude. Short-term electricity load forecasting with generalized additive models. *Proceedings of ISAP power*, 2011, 2011.

[6] S.N. Wood. *Generalized Additive Models: An Introduction with R*. Texts in Statistical Science. London: Chapman & Hall, second edition, 2017.

[7] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, second edition, 2009.

[8] S. Kim and H. Kim. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679, 2016.