The School of Mathematics

Forecasting Short-term Electricity Demand in the UK

by

Jihong Ma, s2242121

July 2022

Supervised by
Prof. Simon Wood & Dr Maarya Sharif

University of Edinburgh – Own Work Declaration

This sheet must be filled in, signed and dated - your work will not be marked unless this is done.

Name: Jihong Ma
Matriculation Number: s2242121
Title of work: Forecasting Short-term Electricity Demand in the UK

I confirm that all this work is my own except where indicated, and that I have:

- Clearly referenced/listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Not sought or used the help of any external professional academic agencies for the work
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Complied with any other plagiarism criteria specified in the Course handbook

I understand that any false claim for this work will be penalised in accordance with the University regulations (https://teaching.maths.ed.ac.uk/main/msc-students/msc-programmes/statistics/data-science/assessment/academic-misconduct).

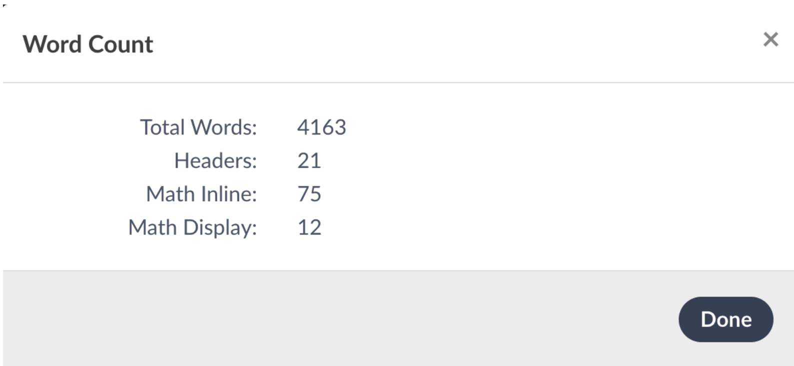
Signature

Jihong Ma

Date: 17 June 2022

Word Count

4163 words (including the executive summary, main text, references, and appendices).



Acknowledgments

I am grateful to the supervisors for this project, Prof. Simon Wood and Dr Maarya Sharif, for their support and guidance. Moreover, I want to thank UK National Grid Electricity System Operator and UK Meteorological Office as they make their data publicly available for research use.

Executive Summary

Background: Short-term load forecasting is always used to predict load demand between 1 hour and 1 week, and it provides support for planning the electricity load generation.

Research question: This report aims to provide an effective predictive model to forecast electricity demand in the UK for the following day. The generative additive model was used to achieve this aim.

Data: Half hourly UK load demand data from between 2011 to mid-2016 was retrieved from the UK National Grid Electricity System Operator, which corresponded with temperature data collected from the UK Meteorological Office.

Methods: A single generative additive model was applied to the data set. The most appropriate model was identified by minimising the values of the root mean square error, mean absolute percentage error, and Akaike information criterion.

Results: The final predictive model of load demand separates the interaction between one-day lagged load and time of day dependent on the level of the day of week, and also considers the interaction between temperature and exponentially lagged temperature. The root mean square error of our final model was 0.94 gigawatts (i.e. 940 megawatts), and the mean absolute percentage error was 1.98%.

Contents

List of Tables

List of Figures

1 Introduction

In recent years, determining how to effectively schedule electricity and manage the power system has become an increasingly important issue. In order to work towards a resolution, electric load forecasting (ELF) must first be understood and applied; ELF refers to a set of mathematical methods that systematically utilise known load data to predict the value of future load at a specific time. This is based on certain system operating characteristics, capacity increase decisions, natural conditions, and social impacts. Furthermore, load forecasting is extensive in its capability, as it can predict hourly, daily, weekly, and annual system demand. According to Wen et al. (2018), ELF can effectively contribute to the planning of power generation and the development of an electric power system, and can also help to preserve coal and fuel, thus enhancing the economic and social benefits of the power system.

This report focuses on the short-term load forecasting field. Short-term load forecasting helps regulate electricity generation to match demand, thus enabling targeted incentivisation to big energy users or small customers to increase or reduce their demand to match predicted supply. The aim of this report is to create an effective model to predict electricity demand in the UK 24 hours in advance, using half hourly load data from the UK National Grid and corresponding temperature data from the UK Meteorological Office. The predictive model considers the average accuracy of predictions, excluding any single prediction with a substantial error, so that power sales companies or other grid operators can minimise discrepancies between their reported and actual electricity consumption. The model's interpretability is another important aspect that has been considered to allow different grid operators to adjust the model based on their own needs.

2 Literature Review

Previous studies portray an array of model techniques that can be implemented to predict short-term electricity demand. As the data for electrical load are always historical, many statisticians promote the construction of time series models. For example, an Autoregressive Moving Average (ARMA) model was proposed by Huang and Shih (2003) to deal with the seasonal pattern in load data. However, time-series models have a high requirement for smoothness of the original data, as well as only being suitable for forecasts with relatively uniform load changes. This type of model lacks the ability to estimate the non-linear relationship between load and uncertain factors, such as the weather or holiday periods. To manage the non-linearity of load data, an approach involving neural networks has been proposed; Hippert et al. (2001) and Ferreira and da Silva (2007) both introduced artificial neural networks (ANN) to address the issue, whilst Kermanshahi (1998) utilised recurrent neural networks (RNN). Despite the high accuracy of neural networks at predicting future load, it should be noted that this is a form of black box machine learning, and that the model result can be difficult to interpret.

To overcome both issues of non-linear patterns in the data and interpretability, a generalised additive model (GAM) can be used for load forecasting. In recent studies, several statisticians have applied GAM to short-term load forecasting problems in various parts of the world: Fan and Hyndman (2011) focused on load forecast in Australia, whilst Pierrot and Goude (2011) applied models to French loads. Both of these previous studies fitted separate models for each half hourly time period. In a more recent paper by Wood et al. (2015), an improvement was shown in the prediction of French load 24 hours in advance. The authors had treated the French load data as a whole, and had proposed a single GAM model rather than using 48 separate models in the prediction.

The present report focuses on half hourly grid data in the UK to create a predictive model, with an aim of predicting electricity demand 24 hours beforehand. To achieve this aim, the objective of this report is to fit a single GAM model with smooth functions of proper covariates and interactions. Several models will be proposed before determining a final effective model based on assessments of model validation and model accuracy. In addition, guidance will be provided for the engineers responsible for operational grid management in regard to which variables or interactions are most significantly important in forecasting future load.

3 Data

3.1 Data Overview

Half hourly UK load demand data from 2011 to mid-2016 was collected from the UK National Grid Electricity System Operator (ESO). Data from the winter holiday periods were excluded to ensure a more convenient prediction process. As well as load information measured directly from the grid, the data set also incorporated temperature information correlating to each time slot, which was taken from the UK Meteorological Office.

Overall, the data set included 90,419 observations, each of which corresponds to a half hourly time point. Each time point had eleven variables, as described in Appendix A.

As the predictive model was intended for load demand 24 hours earlier, it was decided that the variable "load" would be used as the response variable. The process of choosing independent variables formulating our model will be provided in Section 5. Since the scales of values in current grid load and lagged grid load were large, we transformed the unit from megawatt (MW) to gigawatt (GW) (i.e. 1000 MW = 1 GW).

3.2 Exploratory Data Analysis

After assessing the basic structure of the UK load data, focus was placed on identifying any patterns in the data. Figure 1 presents the annual pattern in the UK load demand, whereby each year is represented by a different colour for greater clarity. The load demand at the start or end of each year is larger than in the middle of the year; essentially, this highlights that the load demand in winter and summer are different. Additionally, a descending trend can be observed over the six year period.

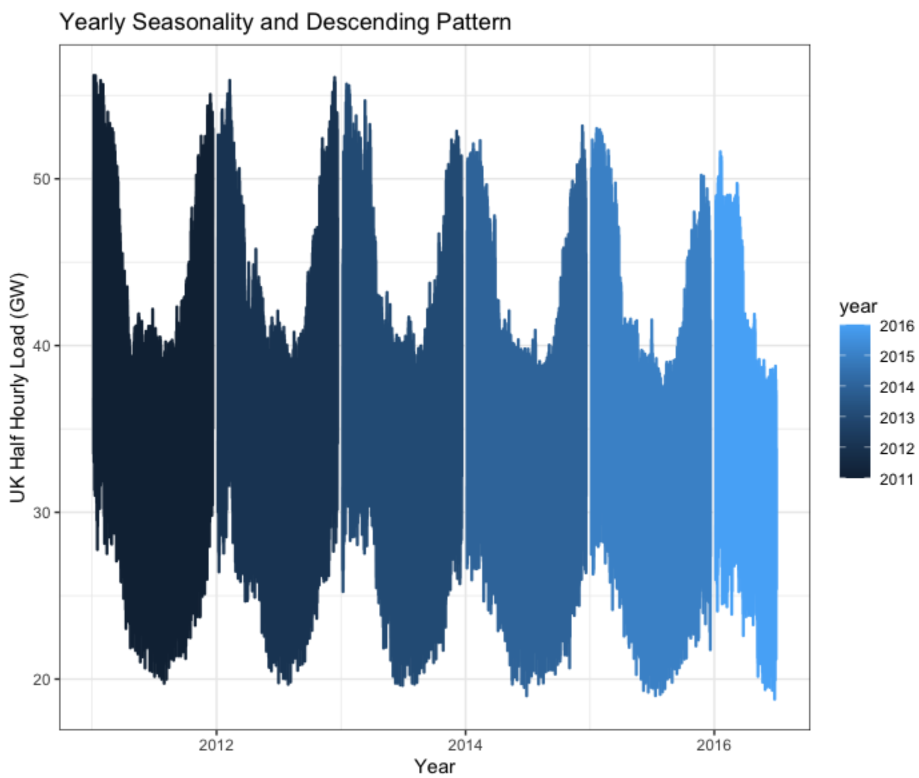


Figure 1: Yearly pattern in electric load in the UK

Due to the identification of this seasonal difference, data was extracted from November and June to determine whether any other patterns were present. As shown in Figure 2, the weekly patterns are similar in winter and summer. The load demands appear similar during the weekdays, but decrease at weekends. However, the daily patterns in the two seasons show distinct differences: the peak load appears at around 5pm to 6pm in the winter, whereas the peak appears at noon in the summer.

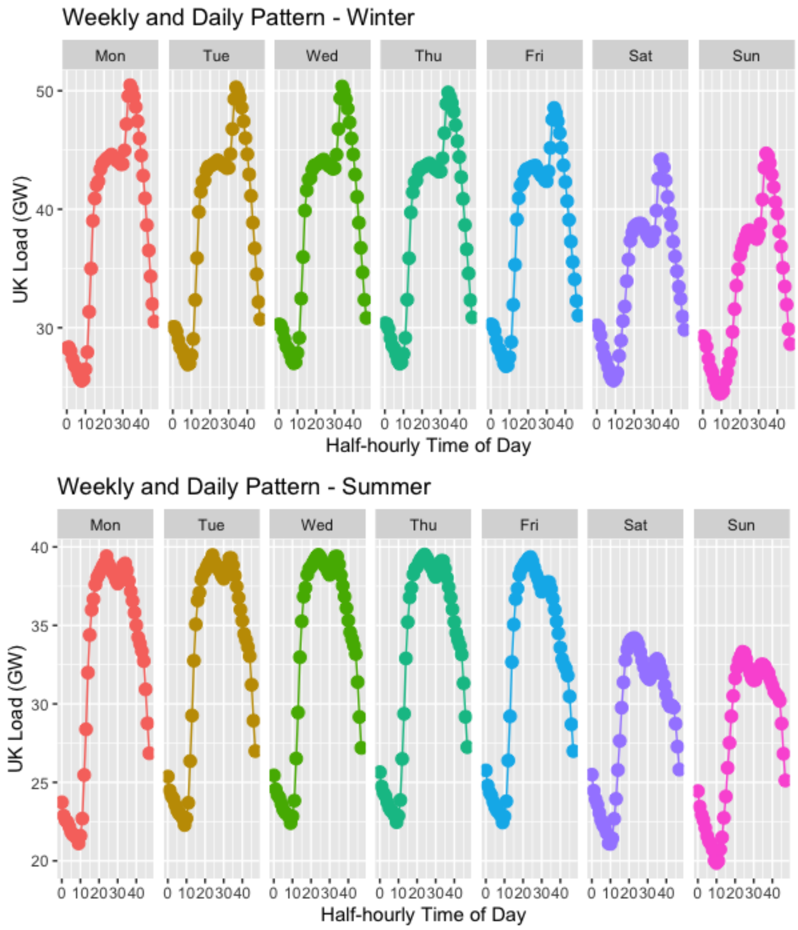


Figure 2: Weekly and daily patterns of electric load in winter and summer

Based on Figure 2,it is evident that temperature affects the load forecast, and so, Figure 3 was created to better visualise this relationship between temperature and load. Subsequently, a non-linear pattern exists between temperature and load; also, the load is generally larger at lower temperatures, which is similar to the phenomenon that was observed in the yearly pattern.

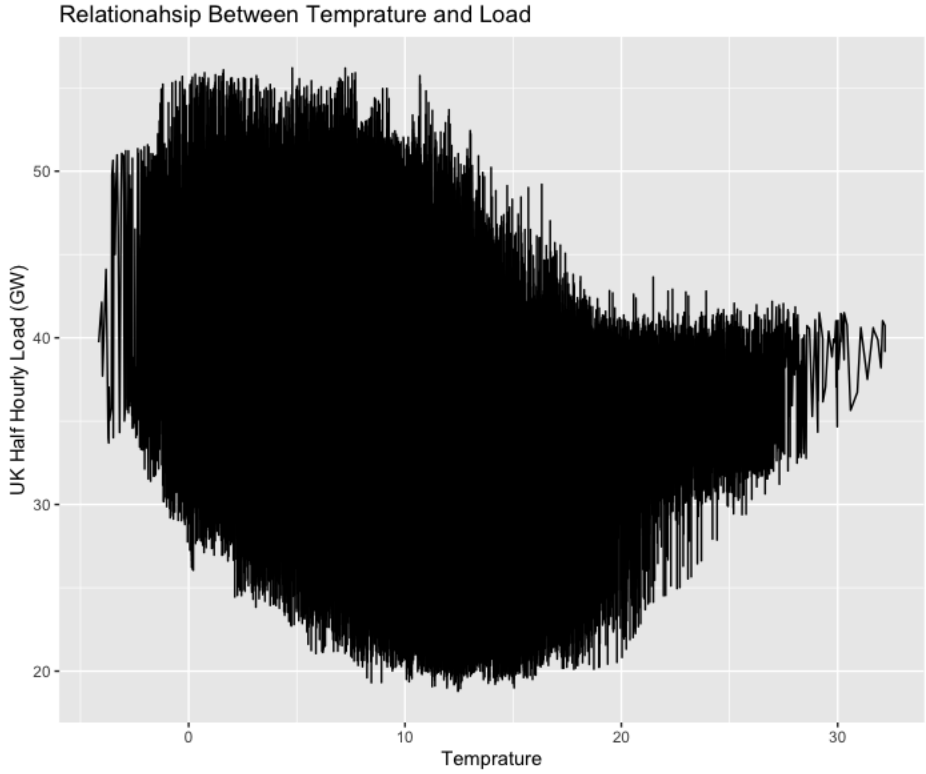


Figure 3: Relationship between temperature and electric load

Before constructing models, a correlation matrix was implemented to identify whether any of the features showed strong correlation. The plot of the correlation matrix is provided in Figure 4. From the plot, the following outcomes were found:

- • The feature “load48” has very high correlation with the response variable “load” (0.93). This is very intuitive in the time series as ”load48” is a lagged variable of ”load”, and it represents the corresponding electric load one day earlier. Therefore, it was decided that the lagged load feature would be taken into account in the model.

- The two features corresponding to time of day and temperature are both relatively highly correlated with the response. These two correlations have been explained in terms of the daily pattern shown in Figure 2 and the non-linear pattern shown in Figure 3, respectively. These two features were also selected to be included in the model.

- The feature ”temp” and ”temp95” are significantly highly correlated with each other. This is also very intuitive,as the derived formula of the exponentially smoothed lagged temperature (temp95) is based on the temperature (temp) - see Appendix A. As temperature was already to be included in the model, it was concluded that the interaction between the temperature and exponentially lagged temperature should also be tested.

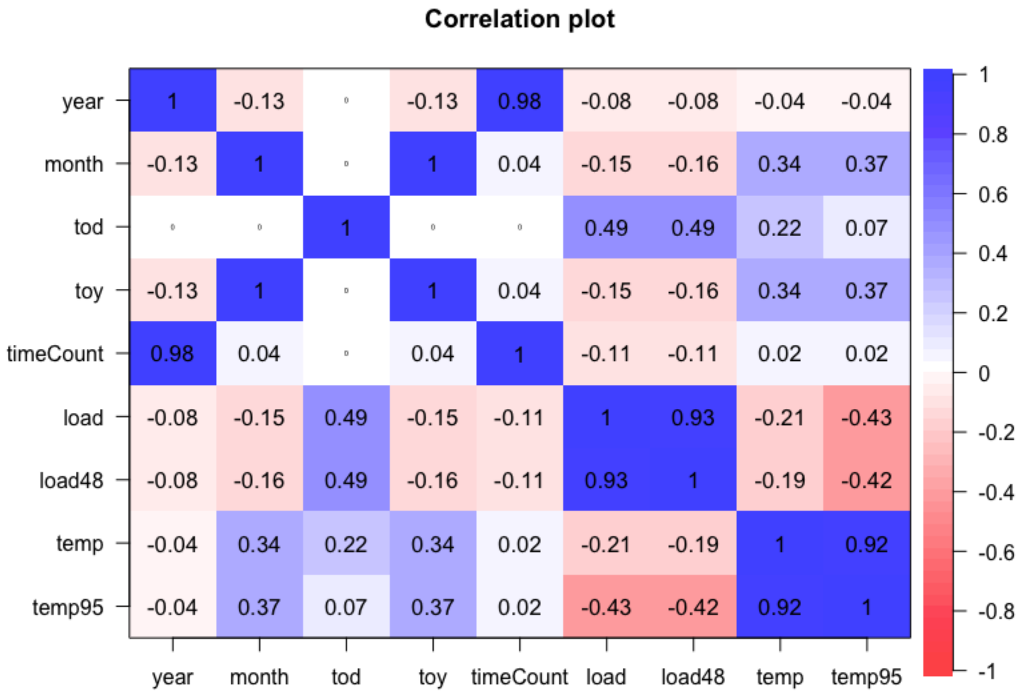


Figure 4: Correlation matrix map

4 Method

To accurately predict the UK demand for electricity, the GAM technique was employed. In the following section, the general framework of GAM is provided, as well as a description of how GAM can be used to estimate the parameters.

4.1 GAM: Basic Structure

GAM is a generalised linear model with a linear predictor that involves a sum of smooth functions of predictors (Wood, 2017, Chapter 4). The general formula for a GAM model is as follows:

$$g(\mu_i) = \boldsymbol{A_i}\boldsymbol{\theta} + \sum_j f_j(x_{ji}) \tag{4.1}$$

- $\mu_i = E(y_i)$, and the response y_i belongs to the exponential family
- $g()$ is a link function (e.g. link/logistic/etc.)
- $\boldsymbol{A_i}$ is the model matrix, $\boldsymbol{\theta}$ is the corresponding parameter vector
- f_j are the smooth functions we want to estimate

4.2 GAM: Estimation

GAM is based on the sum of the smooth functions. Each smooth function is represented using a basis expansion of several basis functions, which can be viewed as regression splines. Hence, GAM is a model that uses splines, and each basis function b_k has a coefficient β_k. Therefore, the resultant spline or smooth function can be written as a sum of the weighted basis functions evaluated at values of x:

$$s(x) = f_j(x) = \sum_{k=1}^{K} \beta_k b_k(x) = \boldsymbol{\beta}^T \mathbf{b}(x) \tag{4.2}$$

GAM is capable of handling non-linear patterns, and so, when fitting a non-linear model, there are two aspects that should be considered: (1). fitting a model which can capture most of the relationship in the data; (2). avoiding fitting model to noise (i.e. overfitting). In GAM, each smooth function that maps the data includes a feature called wiggliness, which should be controlled and penalised so as to avoid overfitting. The formula for wiggliness is as follows:

$$\int f_j''(x)^2 dx = \int \boldsymbol{\beta}^T \boldsymbol{b}''(x) \boldsymbol{b}''(x)^T \boldsymbol{\beta} dx = \boldsymbol{\beta}^T \boldsymbol{S}_j \boldsymbol{\beta} \tag{4.3}$$

Furthermore, controlling the wiggliness can be achieved by taking into account the number of basis functions in a smooth function, and utilising the smoothing parameter λ_j. After setting the wiggliness penalty, the model coefficients $\boldsymbol{\beta}$ can be estimated by maximising penalised likelihood, as shown below:

$$\mathcal{L}_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \frac{1}{2} \lambda_j \boldsymbol{\beta}^T \boldsymbol{S}_j \boldsymbol{\beta} \tag{4.4}$$

Where:

- l is the log-likelihood for $\boldsymbol{\beta}$ and represents how well the GAM captures the patterns in the data.

The estimation of the smoothing parameter λ_j is also very important,an estimation that is too large will result in the smooth function depicting a straight line, whereas an estimation that is too small will lead to overfitting. In light of this, there are several ways to optimise λ_j, including generalised cross validation (GCV), and restricted maximum likelihood (REML). In practical terms, REML is numerical stabler, and is more resistant to severe overfitting on rare occasions (Wood, 2017, Chapter 6). Optimisation of λ_j was achieved by minimising the following (Wood et al. (2015)):

$$\mathcal{V}_r(\boldsymbol{\lambda}) = \frac{\left\| \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\lambda \right\|^2 + \hat{\boldsymbol{\beta}}_\lambda^{\mathrm{T}} \mathbf{S}_\lambda \hat{\boldsymbol{\beta}}_\lambda}{2\phi} + \frac{n - M_p}{2} \log(2\pi\phi) + \frac{\log \left| \mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{S}_\lambda \right| - \log |\mathbf{S}_\lambda|_+}{2} \tag{4.5}$$

Where:

- $\mathbf{S}$ is the precision matrices of Gaussian random effects
- $|\mathbf{S}_\lambda|_+$ is the product of the strictly positive eigenvalues of $\mathbf{S}_\lambda$
- M_p is the degree of rank deficiency of $\mathbf{S}_\lambda$

Another factor that can alter the wiggliness of the smooth function is the number of basis functions: too few basis functions will result in an insufficiently wiggly spline that is unable to capture the relationships present in the data, whereas too many basis functions will result in overfitting of the model. In this report, the model defined $k = dim(\boldsymbol{\beta})$ as the maximum degree of freedom, whereby k should be large enough to be able to capture non-linear patterns in the model. Once penalisation is applied to the smooth function, the variability of $\boldsymbol{\beta}$ becomes restricted, thus meaning an alternative definition of degrees of freedom to k is required. This results in effective degrees of freedom (EDF), which should be less than k under penalisation; in addition, checking basis dimensions should be contained in the model checking process.

5 Models and Results

The general framework of the model to predict UK load demand one day ahead using GAM method is shown below:

$$L_i = \beta_0 + \sum_j^p f_j(x_{ji}) + \epsilon_i \tag{5.1}$$

- L_i is the grid load to be forecasted in at i^{th} half hour period, supposing load follows a Gaussian distribution
- f_j denotes the smooth functions, assuming p smooth functions in the model
- $\boldsymbol{x_i} = x_{1i}, ..., x_{pi}$ denotes the explanatory variables mapped to f_j at i^{th} half hour period
- β_0 is the intercept, and ϵ_i is the residual error that is identically and independently distributed with $N(0, \sigma^2)$

All models in this report were created using R package *mgcv* - see (Wood, 2017, Chapter 7) for details.

5.1 Model Preparation

The data were separated into training and validation sets, whereby data from 2011 to 2015 were assigned as the training set, and the remaining data from 2016 were assigned as the validation set. Several models were trained using the training set; however, prior to modelling, useful variables were chosen in the models.

As the UK load data set is a time series set, the electricity load at the current time will be highly related to the load at the same time days before. The relationship is displayed in Figure 4, where the lagged load variable shows strong correlation with the response load variable.

Figure 2 shows weekly and daily patterns in the data. The daily pattern can be modelled using the numerical feature of time of day (i.e. 48 time intervals in a day), whilst the weekly pattern can be modelled using categorical variable day of week (i.e. 7 levels, one for each day of the week).

Meanwhile, Figure 3 highlights a non-linear relationship between temperature and electricity load. The temperature shows strong correlation with the exponentially smoothed lagged temperature, and therefore, the significance of the interaction between these two features was tested through the model.

The subsequent models, detailed in Section 5.2, contain the following variables:

- the lagged grid load 48 half hours previously;
- the time of day, corresponding to 48 half hours in a day;
- the day of the week;
- the average daily temperature;
- the exponentially smoothed lagged temperature dependent on the average daily temperature

Each explanatory variable was modelled using regression splines (i.e. smooth functions). Specifically, default thin plate splines were employed, as such splines to select knot placements or basis functions were unnecessary. Furthermore, thin plate splines can accommodate any number of predictors and provide suitable flexibility to allow the user to choose the order of derivative in wiggliness measurement (Wood, 2017, Chapter 5).

After fitting the model at each time, residual checking and basis dimension checking were conducted. By ensuring that the residuals for each model were normally and randomly distributed, the regression model would be a good fit. It was also important to ensure that the basis dimension was large enough to explain the patterns in the data.

Three models were trained by the training set, which were then applied yo the validation set. The models were then compared and selected from one to another by minimising the prediction errors. The best model was determined by finding the smallest values of the root mean square error (RMSE) and the mean absolute percentage error (MAPE) of the validation set (Appendix B shows the formula for these two selection criteria). In addition, explained deviance was used to validate the model performances.

5.2 Best Model Selection

The formula for Model 1 (M1) is as follows:

$$L_i = \beta_0 + f_1(L_{i-48}) + f_2(I_i) + f_3(T_i) + f_4(T_{95_i}) + f_5(T_i, T_{95_i}) + e_i \tag{5.2}$$

Where:

- L_{i-48} is the grid load lagged 48 half-hours
- I_i is the half-hour period of the day (coded 0-47)
- T is the temperature
- T_{95} is the corresponding exponentially lagged temperature
- f_j, $j = 1, 2, 3, 4$ are univariate smooth functions represented as thin plate splines
- f_5 is the pure smooth interactions of thin plate splines

After running the summary table for M1, the pure smooth interaction term was found to be of suitable significance to be included in the model. The weekly pattern was then added to the model, resulting in Model 2 (M2). The formula for M2 is as follows:

$$L_i = \beta_0 + \gamma_{j(i)} + f_{j(i)}(L_{i-48}) + g_1(I_i) + g_2(T_i, T_{95_i}) + e_i \tag{5.3}$$

Where:

- observation i is from day j of the week ($j = 1, ..., 7$)
- each γ_j denotes a fixed effect for each level of the day j of week (i.e. fixed effect when the day is Monday, Tuesday, or etc.) as the weekly pattern is represented as a linear term
- the seven f_j and g_1 are thin plate splines
- g_2 is isotropic smooth interaction of thin plate splines, as the units of temperature and exponentially lagged temperature are the same

M2 works on the basis that the smooth function of one-day lagged depends on whether the days are Monday, Tuesday, etc.

M1 and M2 were applied to the validation set, allowing a comparison of the deviance explained from the results of each model. A comparison of the corresponding RMSE and MAPE of the two models was also completed. The results are shown in Table 1:

Table 1: Comparison of M1 and M2

Model	$RMSE(GW)$	$MAPE(\%)$	Deviance Explained(%)
M1	2.56	5.55	89.4
M2	1.53	3.40	96.7

The values for RMSE and MAPE of M2 are smaller than those of M1, and M2 explains greater deviance than M1. This indicates that separate smooth of lagged load based on the level of days of the week distinctly improves the model performance.

Despite the response itself, the correlation matrix provided in Figure 4 shows that the one-day lagged grid load is correlated with the time of day. This is intuitive as the grid load for one day previously also displays a daily pattern. Thus, it was decided to interact the one-day lagged load and time of day, and separate the smooth of interaction based on the level of day of the week. The formula for Model 3 (M3) is as follows:

$$L_i = \beta_0 + \gamma_{j(i)} + f_{j(i)}(L_{i-48}, I_i) + g_1(T_i, T_{95_i}) + e_i \tag{5.4}$$

Where:

- the seven f_j are smooth interactions represented by tensor products of thin plate splines, as the units for one-day lagged load and time of day are different
- g_1 is isotropic smooth interaction

The idea of M3 is that the daily pattern of the load in the previous day varies in different days of week.

The comparison between M3 and the previous two models is presented in Table 2. The results show that RMSE and MAPE are smaller for M3, whilst the explained deviance is larger, indicating that M3 is an improvement of M1 and M2.

Table 2: Comparison of M1, M2, and M3

Model	$RMSE(GW)$	$MAPE(\%)$	Deviance Explained(%)
M1	2.56	5.55	89.4
M2	1.53	3.40	96.7
M3	0.94	1.98	99.2

As almost 100% of the deviance can be explained by M3, it was deemed unnecessary to add any new smooth to the model. The residuals of M3 were checked, as shown in Figure 5.

The residual plots indicate that the residuals are not independently and randomly distributed, as they do not show random scattering around 0. Therefore, autocorrelation exists. Rather than assuming independent residuals, it was assumed that e_i is an AR(1) residual where $e_i = \rho e_{i-1} + \epsilon_i$ and ϵ_i follows $N(0, \sigma^2)$. The AR parameter was estimated to be 0.97, due to the AIC and REML score hits the minimum at 0.97 - see (Wood, 2017, p. 396) for instructions. In Figure 6, the left plot displays the ACF for residuals ignoring auto-correlation, whilst the right plot shows the ACF for residuals with AR(0.97) errors. The standardised residuals in the right plot do not appear to show correlation.

However, after adding autocorrelation, a worse fit on the validation set was found (Table 3. As a result, the autocorrelation model was not included as the final model.

Table 3: Comparison of model performance with and without autocorrelation

Model	$RMSE(GW)$	$MAPE(\%)$
$\rho = 0$ fitted	0.70	1.3
$\rho = 0$ validation	0.94	1.98
$\rho = 0.97$ fitted	0.82	1.7
$\rho = 0.97$ validation	1.02	2.2

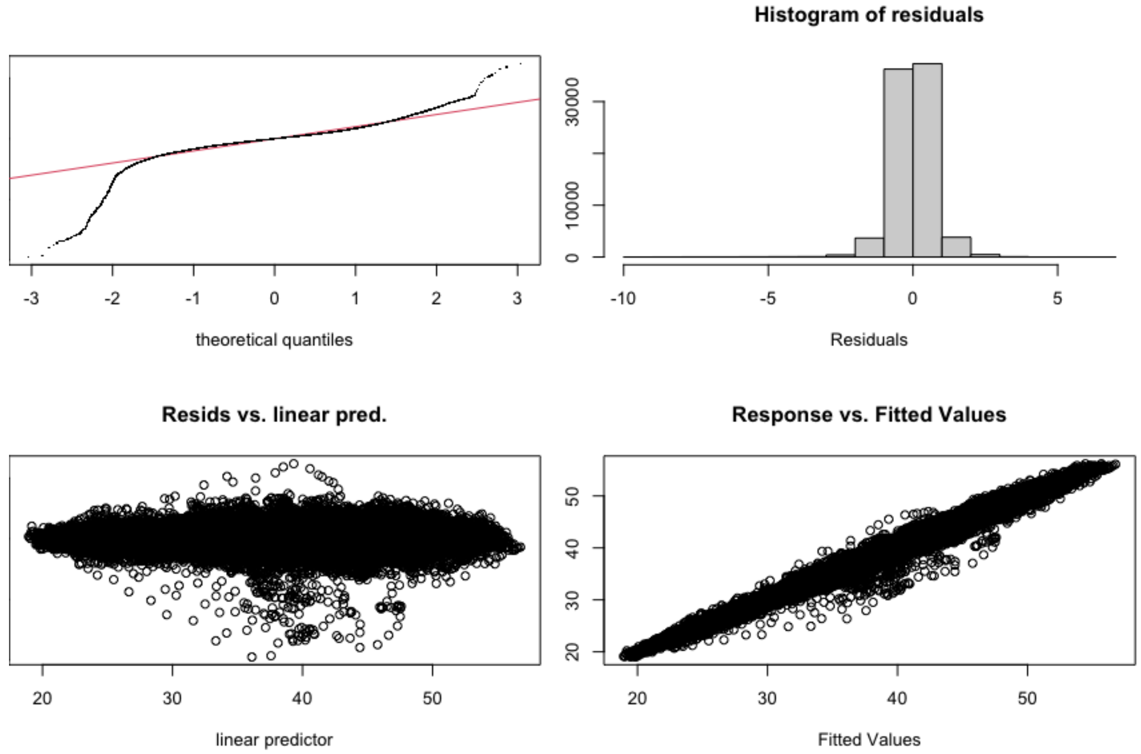


Figure 5: Diagnostic residual plots for M3

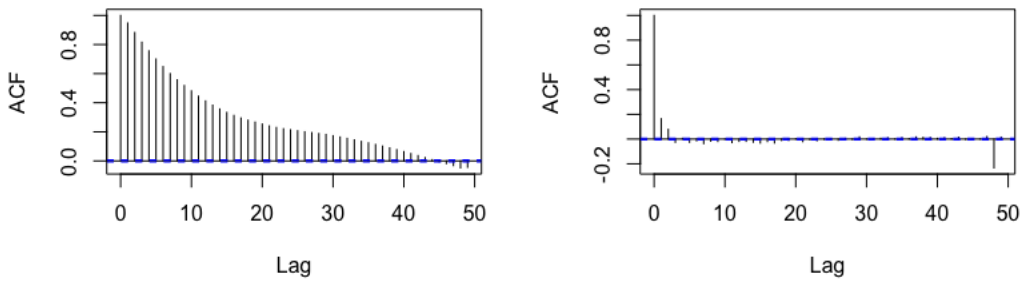


Figure 6: ACF plot for $\rho = 0$ and $\rho = 0.97$

To further confirm the final model selection, AIC was utilised - see Appendix B for details. A smaller value for AIC represents a better model. Table 4 shows the AIC values for the M1, M2, and M3, with the third model clearly showing the smallest AIC value. This result confirmed M3 as being the best predictive model in this study.

Table 4: Model comparison using AIC

Model	*df*	*AIC*
M1	191.76	383531.1
M2	124.90	287351.8
M3	955.33	175467.8

5.3 Predictive Results

M3 was finally applied over grid load data from January 2016 to June 2016 (i.e. the validation set) to assess the model accuracy for load demand prediction 24 hours in advance. The RMSE of the final model was 0.94 gigawatts (i.e. 940 megawatts), and 1.98% MAPE, which means the average difference between the forested load demand and the actual load demand was only 1.98%.

To determine the performance of the final model, the predicted load demand for the first 75 days in 2016 was plotted along with the observed load for this period (Figure 7). There is almost complete overlapping of the actual and predicted load, meaning that the fitted GAM accurately predicted one-day ahead grid load demand in the UK.

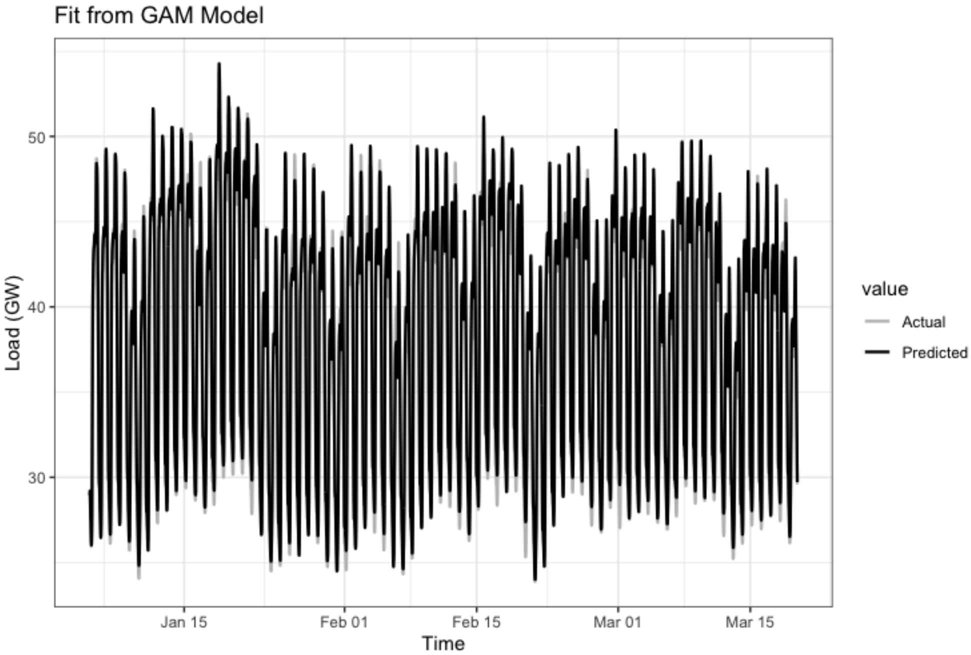


Figure 7: Fitted load on the first 75 days of data in 2016

6 Conclusion

Grid load demand has shown a decreasing trend over recent years, with a differing pattern between weekly and weekdays and weekends, as well as a different daily pattern dependent on the season. A curved pattern was also identified between temperature and grid load. Overall, this report found that the features that affect grid load demand are non-linear. To effectively deal with the non-linear patterns and maintain model interpretability, a single GAM was proposed to predict UK load demand 24 hours ahead.

Both the short-term effect and the temperature effect were considered when constructing this model: (1) the interaction between one-day lagged load and time of day was carefully separated based on whether it was a weekday, thus taking into account how the daily pattern varies in a week; (2) in regard to the main effect of temperature and exponentially smoothed temperature, consideration was granted to how their interaction would affect the load forecast. The final model used to validate the model accuracy of the data from 2016 performed very well, with small values of RMSE and MAPE, thereby indicating a minimal prediction error between the forecast and the observed.

However, certain limitations of the predictive model should be noted, with further improvements suggested below:

- the UK half-hourly data used to construct the model did not include data for winter holiday periods. Grid load during this time may differ from other periods during the year, and so the current model may result in high prediction error if it was used to forecast the load demand during winter holidays. Most people may stay at home during the Christmas period in the UK so that the electricity demand may be higher, especially the cold weather also indicates more people stay at home (Figure 3);

- as well as temperature, there may be other meteorological effects that affect the load demand, such as cloud cover, or precipitation. Therefore, more meteorological factors could be incorporated into future studies to further increase the accuracy of the predictive model.

References

Fan, S. and Hyndman, R. J. (2011). Short-term load forecasting based on a semi-parametric additive model. *IEEE transactions on power systems*, 27(1):134–141.

Ferreira, V. H. and da Silva, A. P. A. (2007). Toward estimating autonomous neural network-based electric load forecasters. *IEEE Transactions on Power Systems*, 22(4):1554–1562.

Hippert, H. S., Pedreira, C. E., and Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, 16(1):44–55.

Huang, S.-J. and Shih, K.-R. (2003). Short-term load forecasting via arma model identification including non-gaussian process considerations. *IEEE Transactions on power systems*, 18(2):673–679.

Kermanshahi, B. (1998). Recurrent neural network for forecasting next 10 years loads of nine japanese utilities. *Neurocomputing*, 23(1-3):125–133.

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

Wen, L., Zhou, K., Yang, S., and Li, L. (2018). Compression of smart meter big data: A survey. *Renewable and Sustainable Energy Reviews*, 91:59–69.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition.* CRC Press LLC, Philadelphia, PA.

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

Appendices

A Data: Feature Description

The following are the descriptions of 11 features in the UK half-hourly data used in this report:

- date: date and time of measurements;
- year: year of corresponding date;
- month: month of year, coded 1-12;
- tod: time of day, coded 0-47 (i.e. 48 half hours in a day);
- toy: time of year, as proportion;
- timeCount: cumulative time scaling;
- load: grid load in megawatts;
- load48: grid load 48 half hours previously;
- temp: average daily temperature;
- temp95: exponentially smoothed lagged temperature, the formula at time t is as following: $\sum_{i=1} T_{t-48i}\rho_i$, where T_t represents temperature at time t, T_{t-48i} is the temperature i days before, and $\rho = 0.95$.
- dow: day of week (i.e. Mon, Tue, etc.).

B Methods: Statistics for Model Comparison

Root Mean Square Error: The RMSE statistic is the standard deviation of residuals, the distance between the fitted data and the observed values. In other words, RMSE tells you how spread out the prediction errors are. The formula for RMSE is the following:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_1)^2}{n}}$$

where $\hat{y}_i$ represents the predicted values; y_i denotes the observed values; and n is the number of observations.

Mean Absolute Percentage Error: The MAPE is another statistic used to measure the prediction error of one model, and the formula is as follows:

$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{n}$$

where A is the actual values; and F is the forecasted values.

Akaike Information Criterion: The AIC statistic is a method can be used to measure the model performance, and provides a means of model selection. In GAM, the formula for AIC is as follows:

$$AIC = -2l(\hat{\boldsymbol{\beta}}) + 2EDF$$

where $l(\hat{\boldsymbol{\beta}})$ is a model's maximum likelihood estimation; and EDF is the effective degrees of freedom in a GAM model.

C R code

The R code for generating the results and models described in the report can be found using the following link: https://github.com/candice616/Dissertation_1.git

I did not include the whole piece of code in the report as it may mess up the report.