Linear Relationship Analysis of Peak Energy Demand against Weather and Time

s2152592, s2277508, & s2279407

28th March 2025

Word Count: 4375 + 19

Executive Summary

Preventing electricity shortfalls is essential for avoiding region-wide blackouts, economic losses, and disruptions to daily life. Aimed at aiding NESO's plan for averting energy shortfalls in Great Britain, this report covers statistical analysis of the relationship between electricity demand, time, and weather. Multiple models were tested and fit using historical data throughout every winter period from 1991 to 2013, covering: electricity demand, estimated solar and wind generation at 6 pm, and hourly temperatures. The analysis was primarily limited by the lack of more extreme temperature occurrences and the absence of information regarding other factors that may be important to consider, such as energy usage by region. To simplify the model, the dates were grouped by month and split into weekends and workdays. While weather and seasonal demand patterns are naturally related, the model was further adjusted in an attempt to isolate weather-specific from time-specific impacts as accurately as possible. Then, through a variety of linear model fitting techniques and comparisons of performance metrics, we arrived at the following model for energy demand:

$$\text{Gross Demand}_{\text{TE}} \sim \text{Year} + \text{Year}^2 + \text{Year}^3 + \text{Month} + \text{Weekend} + \text{TE}_1 + \text{TE}_1^2 + \text{Wind} + \text{Solar}.$$

To find the effect of different weather conditions on the model's predicted maximum annual demand, the weather conditions of previous years were used on the derived model for the 2013/2014 winter period. The importance of the year outweighed that of the weather conditions, as the peak annual demand prediction varied slightly despite the different weather conditions, and also differed noticeably from the actual peak values during those years. Then, to assess the model's predictive ability using the dataset's known values, predictions of maximum demand 5 and 10 years ahead were made for every year the data would allow. Both predictions performed especially poorly in earlier years with less data; although they did increase in accuracy with additional year data, there were still noticeable fluctuations from the true values. The 95% prediction intervals were quite wide, especially for the significantly wider 10-year predictions. A lack of knowledge about the conditions affecting generation in future years makes accurate future energy demand forecasting incredibly difficult, with predicting innovations in technology, unprecedented events, changes in climate, consumer behaviour, and the breadth of energy generation all being beyond the scope of this report. Future research suggestions include improving confounding, considering different data groupings, and using a multivariate statistical approach. Researching and incorporating more details into the dataset, such as the inclusion of different temperature measurements, resonance-specific factors (e.g. population density), and the breakdown of energy types, may result in a more practical model.

1 Introduction

NESO, the National Energy Service Operator, is responsible for managing and mitigating risk in the supply and demand of the electrical network across Great Britain, along with overseeing long-term electricity production planning decisions. Crucially, to forecast the energy market dynamics, they calculate predictions of the peak daily electricity demand using linear regression models; further, using their "year effect" statistic to appropriately rescale historic data to suitably anticipate future changes such as climate change, carbon reduction policies, population increases, new technology, and more. This equips them with the tools to mitigate potential electricity shortfalls through proposals for new grid generation based on these predictions.

The objective of this report is to aid NESO in their supply analysis and predictions through the development of a new linear regression model to anticipate daily peak energy demand, and subsequently, annual peak energy demand. This is done by proposing, fitting, and testing various models using the historical data that we have available and comparing them using appropriate performance metrics. Care is taken in model selection due to the highly confounding nature of some of our dataset's covariates; then analysing the relationship between them, attempting to mitigate the impact of confounding on the final model. Additionally, using our dataset, model validity is assessed using 10-fold cross-validation and by studying the variation in maximum annual demand under changes in weather conditions. We conclude with a discussion of our model's limitations for long-term forecasting, in the 5-10 year scale.

2 Dataset Overview

Description

As temperatures go down, there is an increase in heating, leading directly to an increase in electricity demand. This heating demand is also associated with the time of year, as not only do some households have pre-set timings for their heating rather than using a temperature-controlled system, but some opt to turn their heating off during the warmer months. Additionally, days get shorter during the winter months, further increasing electricity usage through lighting. Thus, to perform an analysis on energy demand, a data set covering the following variables will be primarily used (with any measurements done at 6 pm on the given day):

- Weekday: the day of the week, from Sunday to Saturday, indexed from 0 to 6.
- Month: the month of the year, from November to March.
- Start year: the year of the winter period's start.
- Date: from November 1991 to March 2014.
- Wind: the estimated capacity factor of wind generation, based on the installed wind generation on the 1^{st} January 2015.
- Solar S: the estimated capacity factor of solar generation, based on the installed solar generation on the 1^{st} January 2015.
- Gross demand: the electricity demand for all of GB recorded by NESO.
- Temperature: British population-weighted average temperature from MERRA1.
- TO: the average temperature between 3 pm and 6 pm.
- TE: the average of TO and TE on the previous day; the temperature metric used by NESO for demand estimates.

A secondary dataset containing 24-hourly temperature measurements for every day from 1991 to 2015 was used to investigate an alternative temperature metric to TE. Our dataset has the following limitations, affecting the applicability and accuracy of our statistical analysis:

- Only the gross energy demand at 6 pm is included, which is assumed to always be when peak daily demand occurs.
- The main dataset only covers winter months; however, as energy demand in Great Britain during this period is significantly higher, periods where electricity shortfall is highest are covered. [2]
- Data only covers temperature occurrences between -6°C and 18°C (and only between -5°C and 15°C for the main dataset), limiting analysis for more extreme occurrences.
- Lack of information regarding other impacting variables or factors, such as usage by population density.
- Potential measurement errors.

Variable Groupings

In our investigation of the relationship between demand and TE, we consider weekdays as factor variables in two groups: weekends (Saturday and Sunday) and work days (Monday to Friday). Analysing the data of each weekday separately, their effects on the relationship is similar within the same group but differ from the other group. Specifically, the linear relationships for weekends have gradients in the -567 to -549 range and intercepts between 46,700 and 47,500, while for weekdays, they have higher intercepts (52,100 to 53,300) and higher gradients (-400 to -510). While some weekdays differed more in effect on the relationship than their group (e.g. Friday with gradient -506.5 (d.p.1)), we opted to only have two groups to not overfit (Table 3), visually represented below in Figure 1.

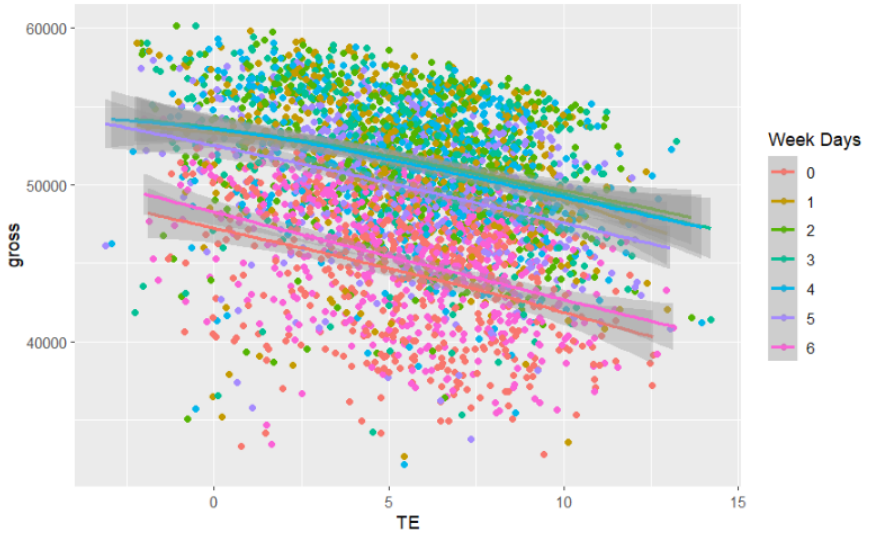


Figure 1: Plot of the energy demand by TE by weekday

To similarly avoid overfitting and retain model simplicity, we grouped dates together by their month, as factor variables, rather than the specific day of the month. To analyse yearly differences, we examined the effect of the start year rather than the full year, as the model only includes winter periods at year-end. Additionally, we treated it as numeric rather than a factor, as the quantity of years could lead to overfitting and overcomplicate analysis. We assume these previous groupings will hold validly for our alternative temperature metrics as well.

3 Model Selection

Linear Regression

A linear relationship model is an equation, linear in some β_j coefficients, between some X_j predictor variables and a Y response variable, with ε as a residual term; its general form is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon.$$

Linear regression is then the method that finds an unbiased estimate for each of the *regression coefficients* β_i that minimises the residual sum of squares, as below, for n data points.

$$\mathbf{RSS} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j (X_i)_j \right) \right)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 .$$

In our situation, we wish to construct a linear relationship between predictor variables of Year, Month index, Day, TE, Wind, and Solar to the response variable Gross Demand. However, we cannot naïvely identify this relationship as some of the covariates may be highly correlated or confound each other, leading to biased coefficient estimates and misleading conclusions.

Weather-Time Covariate Relationship Analysis

Ideally, there would be no relationship between any of our covariates in our model, but this is not possible as there is no foolproof method of completely removing all confounding impact on our model. However, since future information regarding the Time Covariates (Year, Month, Day) can be directly obtained, i.e from a calendar, we aim to only mitigate them from confounding the Weather Covariates (TE, Wind, Solar) as we do not reliably have their future values. For ease of notation, we assign the Weather Covariates as W_i for $i \in \{1, 2, 3\}$ and the Time Covariates as T_i for $i \in \{1, 2, 3\}$, carefully noting that both T_2 & T_3 are factor variables $(T_2)_j$ for $i \in \{0, 1, 2, 3, 4\}$ and $(T_3)_j$ for $i \in \{0, 1\}$.

We analyse the relationships between each of the Weather Covariates with all of the Time Covariates by creating box plots of each W_i against T_i, from these we infer a handful of different relationships between each W_i and all of the T_1, T_2, & T_3. As both of T_2 & T_3 are factor variables, proposed models that contain transformations of them would not change anything; so we investigate relationships for T_1 such as polynomial[1], logarithm, exponential, and trigonometric functions aiming to capture the non-linear nature of each relationship.

To select the best proposed relationship for each W_i, we introduce model performance metrics R^2, Cross-Validation Mean-Square-Error (CV-MSE), & AIC to compare them against each other.

- The R^2 coefficient (ranging from 0 to 1) measures how well the predictor explains variance in the response, with 1 meaning perfect explanation and 0 meaning none, i.e it describes the "goodness of fit" for a model to the historic data given its predictor variables. However, note that a high R^2 does not necessarily imply a good model, as it could be misleading when faced with overfitting. Below is the formula for R^2, where $\bar{Y}$ is the mean of the observed values.

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2}$$

- The CV-MSE is a metric of a model's predictive performance on unseen data, where a lower CV-MSE means better predictive accuracy. We first perform k-fold cross-validation wherein we train the model on $(k-1)$ folds of the data and predict the data from the remaining fold. Then we calculate the MSE, mean-squared-error, that being the average squared difference between the actual and predicted values, for this remaining fold. This is iterated k times, and the average of these is taken to be the CV-MSE. This metric can only be used to compare models on the same dataset with the same response variable. We used $k = 10$ for this report.

- The AIC is an error metric of regression; lower is better, but it increases with additional model complexity. It is calculated by multiplying the number of predictor variables, K, by 2; then subtracting 2 times the maximum likelihood, L. This penalises models with additional predictor variables, mitigating overfitting by disincentivising unnecessary complexity.

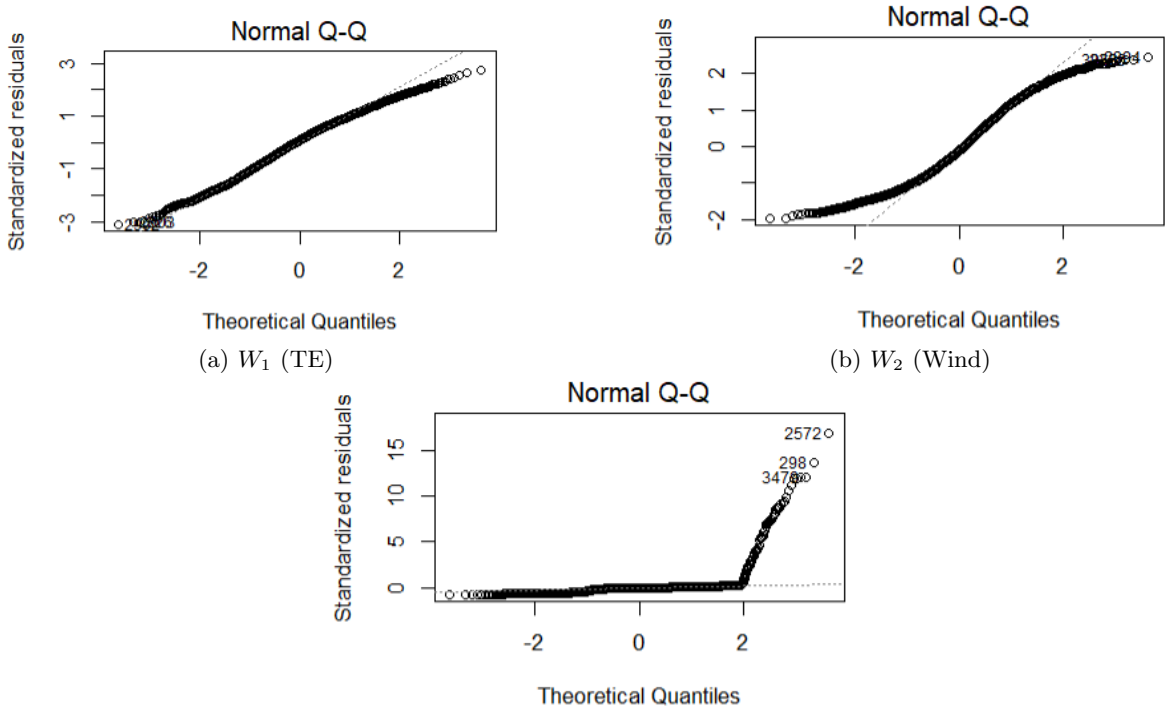
Now, the validity of these metrics depends on the distribution of the residuals, as is shown in a QQ-plot; so before we analyse the performance metrics for each proposed model for each W_i, we

[1] including fractional exponents

inspected the corresponding QQ-plots. From this, we saw that the characteristic features of the plots across the proposed relationships for each W_i were quite consistent, so we take the following models.

$$\begin{aligned} W_1 &\sim T_1 + T_1^2 + T_1^3 \quad + T_1^{\frac{1}{2}} \qquad\quad + T_2 + T_3 \\ W_2 &\sim T_1 + T_1^2 + T_1^3 \qquad\quad + T_1^{\frac{1}{3}} + T_2 + T_3 \\ W_3 &\sim T_1 + T_1^2 + T_1^3 \quad + T_1^{\frac{1}{2}} \quad + T_1^{\frac{1}{3}} + T_2 + T_3 \end{aligned} \tag{1}$$

Thus, now we make our analysis based on the plots in Figure 2 for each W_i.



(a) W_1 (TE)

(b) W_2 (Wind)

(c) W_3 (Solar)

Figure 2: QQ-Plots of each proposed relationship model for the Weather Covariates

In the above, we can see that, for (a), the residuals display a strong alignment with the theoretical quantities, indicating the assumption that the distribution has mean zero and constant variance holds well; for (b), we observe some deviation from our assumptions on residual distribution in the tails, however the residuals still exhibit reasonable alignment with the theoretical quantities in the central region; & for (c), there is a significant deviation from distribution assumptions in the upper tail, leading to a highly skewed residual distribution likely with non-constant variance.

So in the case of W_1 & W_2, we can reasonably employ the discussed performance metrics for model comparison; however, W_3 appears more problematic. While strict zero mean and constant variance assumptions are ideal for statistical inference, our primary focus is on relative model performance to reduce the effect of confounding over the absolute interpretability of the residuals. So, for the purpose of this report, we can maintain that our model selection metrics are useful in identifying the best relative fit among competing models, whilst acknowledging their limitations.

Thus, we employ our performance metrics to select the best model for each W_i. In doing so, we see that the relationships with the highest R^2 and lowest CV-MSE & AIC are those stated in Equation 1. Now, to reduce unnecessary complexity from these models, we employ AIC Backwards-Selection. This technique iteratively removes predictor variables, calculating the corresponding AIC for each new model, identifying which removal results in the lowest AIC, refitting the model under this removal, and repeating this process until AIC no longer decreases. Thus, we arrive at the following relationships.

$$\begin{aligned} W_1 &\sim T_1 + T_1^2 + T_1^3 + T_2 \\ W_2 &\sim T_1 + T_1^2 \quad\quad + T_2 \\ W_3 &\sim T_1 \quad\quad\quad + T_2 + T_3 \end{aligned} \tag{2}$$

Full Model

Now, proceeding similarly, we inspected box plots of Gross Demand against each T_i, noticing obvious non-linear trends, particularly in T_1, and scatter plots of Gross Demand against each W_i, which proved minimally insightful. In construction of proposal models, we make sure to include all T_i transforms from Equation 2 to mitigate confounding while also considering additional adjustments to the other predictor variables, such as polynomial[2], logarithmic, exponential, and trigonometric functions where physically appropriate. Note that transforms of W_3 were not considered due to the problematic nature of the solar data, which contained a large number of zero values, making meaningful transformations challenging.

Regarding the full model, we now need to take greater care to ensure the assumptions of our residual distribution, notably zero mean and constant variance, are upheld to ensure accurate parameter estimates, reliable predictions, and valid confidence intervals. So, when finalizing the model, we account for these residual properties to ensure statistical validity as done in section 3. However, for the sake of brevity, we proceed by applying our performance metrics to facilitate model selection without currently concerning ourselves with residual distribution.

We find that the candidate model that maximises the R^2, i.e that best fits the historic data, and minimises the CV-MSE and AIC out of all proposed models is as follows.

$$\text{Gross Demand}_{\text{TE}} \sim T_1 + T_1^2 + T_1^3 + T_1^{\frac{1}{2}} + T_1^{\frac{1}{3}} + T_2 + T_3 + W_1 + W_1^2 + W_2 + W_3$$

To reduce non-essential complexity from our model, we used AIC Backwards Selection, making sure to retain any T_i transforms from Equation 2 even if dropped, in an attempt to reduce the impact of confounding. Arriving at the following for our full model.

$$\text{Gross Demand}_{\text{TE}} \sim T_1 + T_1^2 + T_1^3 + T_2 + T_3 + W_1 + W_1^2 + W_2 + W_3$$

Alternative Temperature Metric

In NESO's formulation of the model for energy demand, they use the temperature metric TE. In this section, we explore alternatives to this, designated TA_i for the i^{th} Temperature Alternative metric, to see if they could improve our relative model performance compared to using TE.

We propose two alternatives to TE, notably TA_1, which is a weighted average of $w_1\times$ maximum, $w_2\times$ minimum, and $w_3\times$ mean temperature for a given day, and TA_2, which is a moving average of TA_1 over the current day and the 2 days prior. As before, we propose multiple different relationships for each TA_i in terms of all T_i, seeing again that the characteristic features of the QQ-plots are consistent across all respective candidate models, with different weightings. Thus, we take the following models for our residual analysis.

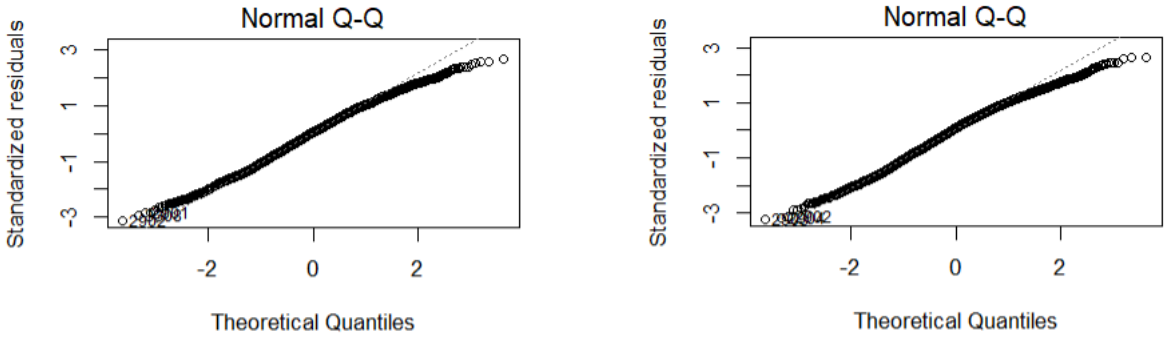
$$\begin{aligned} \text{TA}_1 &\sim T_1 + T_1^2 + T_1^3 + T_1^{\frac{1}{3}} + T_2 + T_3 \\ \text{TA}_2 &\sim T_1 + T_1^2 + T_1^3 + T_1^{\frac{1}{3}} + T_2 + T_3 \end{aligned} \tag{3}$$

Now we begin analysis of the corresponding QQ plots for each TA_i, in figure 3

In both cases, we see clear alignment with the theoretical quantities, showing that our residual distribution assumption holds and that we can employ our performance metrics to compare our potential models. Through which, we find optimal weightings of $w_1 = 0.7, w_2 = 0.15$, and $w_3 = 0.15$, subsequently arriving at the relationships presented in Equation 3; performing AIC Backwards Selection, we achieve the reduced complexity models in Equation 4 below.

$$\begin{aligned} \text{TA}_1 &\sim T_1 + T_1^2 + T_2 \\ \text{TA}_2 &\sim T_1 + T_1^2 + T_2 \end{aligned} \tag{4}$$

[2]again, including fractional exponents



(a) TA_1 (b) TA_2

Figure 3: QQ-Plots of each proposed relationship model for the alternative temperature metrics

We can refit our full model for each TA_i whilst again making sure not to drop any relevant T_i transforms, attempting to prevent confounding in the full model. As in the "Full Model" subsection, we chose the candidate models that maximise the R^2 coefficient and minimise the CV-MSE and AIC out of all potential models, given in Equation 5.

$$\begin{aligned} \text{Gross Demand}_{TA_1} &\sim T_1 + T_1^2 + T_1^3 + T_2 + T_3 + TA_1 + TA_1^2 + W_2 + W_3 \\ \text{Gross Demand}_{TA_2} &\sim T_1 + T_1^2 + T_1^3 + T_2 + T_3 + TA_2 + TA_2^2 + W_2 + W_3 \end{aligned} \tag{5}$$

Now we can explicitly state and compare the performance metrics of each model for Gross Demand_k for $k \in \{TE, TA_1, TA_2\}$.

- For $k = TE$ we have $R^2 = 0.7629$, CV-MSE = 6250120, & AIC = 64315.73;
- for $k = TA_1$ we have $R^2 = 0.7567$, CV-MSE = 6427051, & AIC = 64404.83;
- and for $k = TA_2$ we have $R^2 = 0.758$, CV-MSE = 6388823, & AIC = 64386.94.

Based on the above, we conclude that the NESO's modified 2-day rolling average form of the temperature metric, TE, creates the relative best model when compared to our investigative temperature metric alternatives.

Residual Analysis

As stated in the previous section, now that we have selected the relative best model, we take greater care to assess how closely the residual behaviour follows our independence, zero mean, and constant variance assumptions.

Upon analysing the residual plots for our demand model, we observe that the residuals are centred around zero but exhibit increasing variance with higher fitted values, indicating heteroscedasticity. The QQ-plot reveals a heavy lower tail, suggesting deviations from normality. Additionally, a plot of residuals against time shows strong periodic behaviour, violating the independence assumption, confirmed further with a plot of residual autocorrelation (Figures 4 and 7).

To address heteroscedasticity, we first applied a Box-Cox transformation, which is commonly used to stabilize variance and improve residual normality [1]. Based on our likelihood plot (Figure 5), the Box-Cox transformed response variable takes the form $(\text{Gross Demand}_{TE} - 1)^2/2$.

However, residual plots following this transformation showed minimal qualitative improvement (Figure 6). Furthermore, since models with different response variable transformations cannot be directly compared using the same evaluation metrics, we opted not to proceed with the transformed demand model.

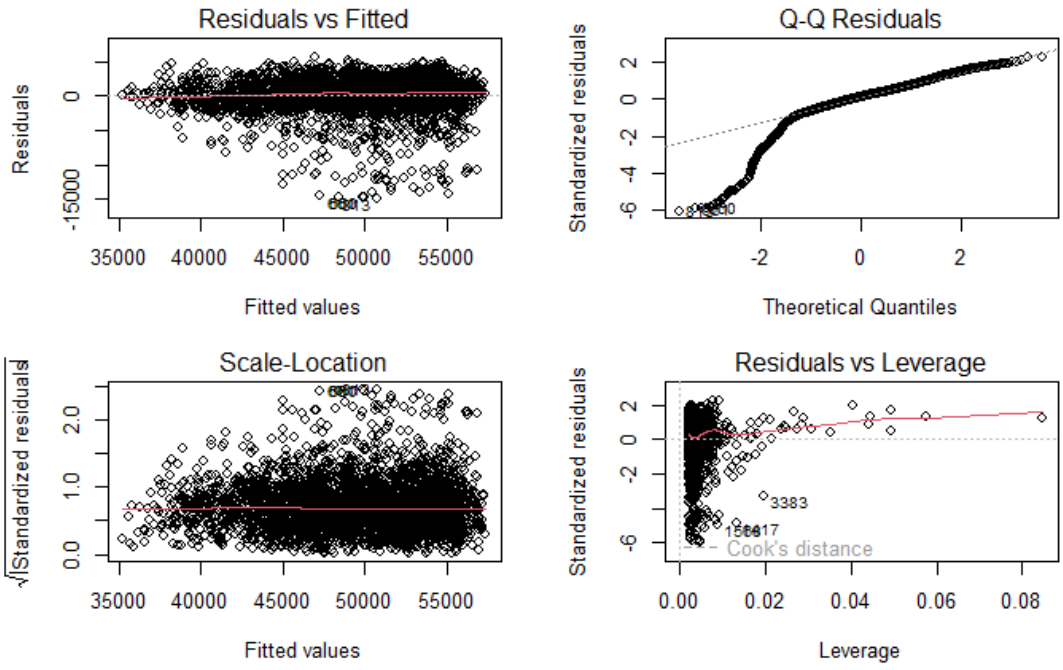


Figure 4: Residual plots of the Gross Demand$_{\mathrm{TE}}$ model.

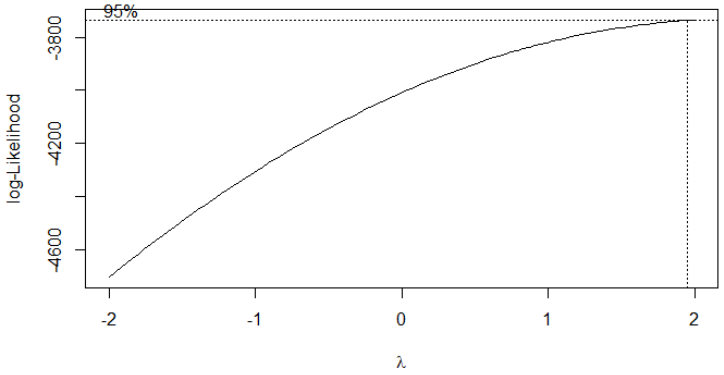


Figure 5: Box-Cox likelihood plot for the Gross Demand$_{\mathrm{TE}}$ model

Block Bootstrapping

To mitigate the impact of violated assumptions on residuals, we recalculated the model parameter estimates and their confidence intervals using block bootstrapping. This method accounts for the time-dependent structure in the residuals present in Figure 7, further reinforced with the autocorrelation plot in Figure 8, which standard bootstrapping would otherwise ignore.

In block bootstrapping, we first partition the time-ordered data into equal-length blocks, of length two weeks, to preserve temporal dependencies in the residuals while ensuring a sufficient number of blocks for resampling. This approach assumes that the residuals are approximately stationary across time, meaning their statistical properties (such as mean and variance) remain constant over time. We justify this assumption by observing the periodic behaviour of residuals centred around some constant mean, present in the plot (Figure 7).

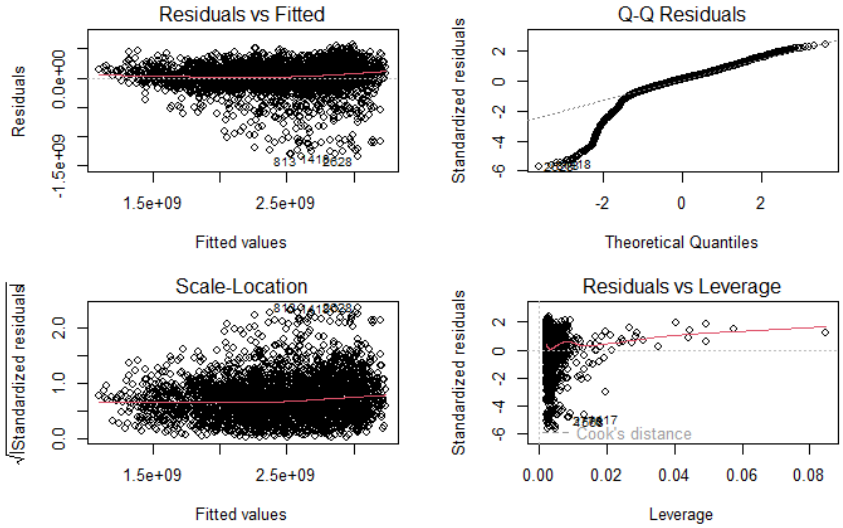


Figure 6: Residual plots of the (Gross Demand$_{\text{TE}}$ - 1)^2/2 model

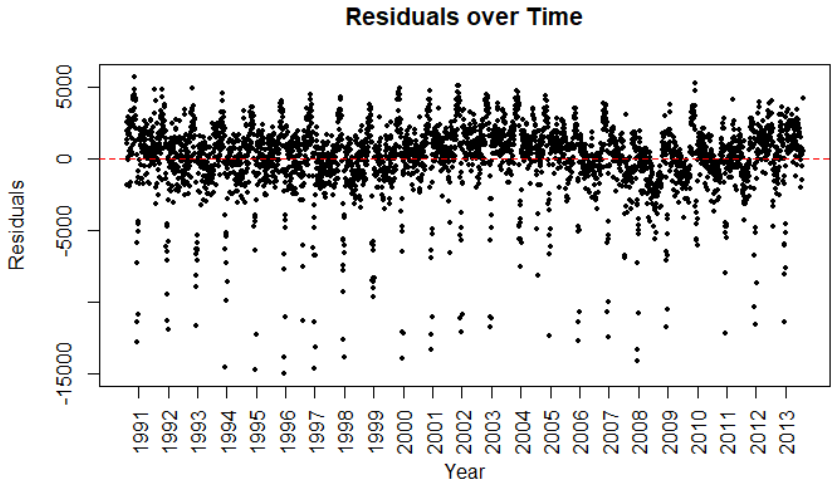


Figure 7: Residuals against time of the Gross Demand$_{\text{TE}}$ model

Within these blocks, we compute the residuals from the original model:

$$\hat{e}_i = y_i - \hat{\beta}^T \mathbf{x}_i \, .$$

Next, we randomly sample these residuals with replacement to create a new set of bootstrapped residuals:

$$\hat{e}_1^*, \hat{e}_2^*, ..., \hat{e}_n^* \, .$$

We then generate bootstrapped response values by adding these residuals to the predicted values from the original model:

$$y_i^* = \hat{\beta}^T \mathbf{x}_i + \hat{e}_i^* \, .$$

For each bootstrapped dataset, we re-estimate the model coefficients. Repeating this process R times results in R samples of the coefficient estimates. The final bootstrapped coefficient estimates, $\hat{\beta}^*$, are the mean of these samples, while the 2.5^{th} and 97.5^{th} percentiles provide the 95% confidence intervals. Additionally, we recalculate the adjusted R^2 based on the residuals from these new coefficient estimates. Applying this approach to our dataset yields Table 4.

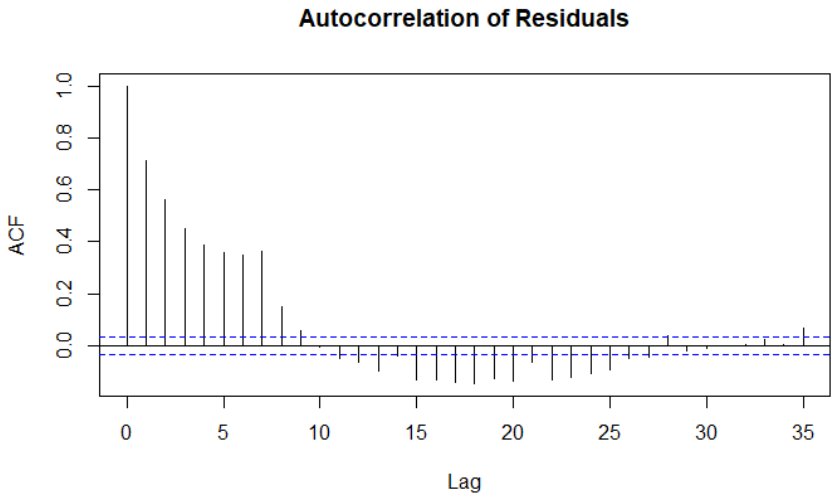


Figure 8: Residual autocorrelation plot of the Gross Demand$_{\text{TE}}$ model

Interpretation of Regression Coefficients

With our full model selected and residuals block bootstrapped to ensure accuracy of our regression coefficients, we now briefly analyse them regarding how they influence the Gross Demand.

The intercept captures baseline information on January and the Workdays, so the coefficients for the remaining months and the weekend represent deviations from this reference point. This is particularly interesting as November has a positive value, indicating higher energy demand than January, while all the rest have negative values. This potentially indicates our mitigation of weather-time confounding may have been successful. As expected, there is a positive relationship of Gross Demand with the Year, and interestingly, also Year^2, with a negative cubic correction. The weekend effect is strongly negative on energy demand from the work-week, likely due to reduced industrial and commercial activity.

Regarding the Weather Covariates, Gross Demand decreases as TE increases, indicating that warmer temperatures reduce energy demand, likely by lowering heating needs; a negative TE^2 coefficient suggests that, at extreme temperatures, the reduction becomes more pronounced. In energy generation, solar appears to have the most substantial effect, having a strong negative relationship with Gross Demand, likely due to the increased temperature with higher levels of sun. In contrast, wind energy displays a positive relationship with Gross Demand, potentially due to periods of higher winds, causing higher wind energy generation, corresponding to increased levels of heating regionally.

From our model, we obtain an adjusted R^2 coefficient of 0.7501, up to the variability introduced by bootstrapping. Specifically, this means that our model explains approximately 75% of the variance of the full model, which is a good level of explanatory power. The remaining 25% of the variance is likely due to short term fluctuations, or behavioural shifts not fully captured by our choice of model. Despite this, an R^2 of 0.75 is quite reasonable for a model dealing with real-world energy data, as electricity demand is influenced by many unpredictable factors. This relatively high value suggests that our model effectively captures a majority of the influential factors, in our historic data, driving gross electricity demand across the UK.

4 Model Evaluation

Assessing Climate Impact with Fixed Year Baseline

To investigate the impact of the changing climate on maximum annual peak demand, we predict daily peak demand using historic Weather Covariate data while fixing the year, month, and weekend data

to that of 2013/2014. Given the residual issues discussed earlier, we apply prediction bootstrapping to enhance the accuracy of our predictions and prediction intervals.

To perform prediction bootstrapping at a point ($\mathbf{x_p}$), we first calculate the prediction y_p using the model parameters obtained from block bootstrapping previously. Moving onto the prediction interval, as laid out in block bootstrapping, we first create R bootstrapped response variables ($\hat{\mathbf{y}}_r^*$), and from this R coefficient samples ($\hat{\beta}_r^*$); we then sample a residual from the original model (e^*), and calculate an error term:

$$\hat{\delta}_r^* = (\hat{\beta}_r^*)^T \mathbf{x_p} - (\hat{\beta}^T \mathbf{x_p} + e^*) \, .$$

This error term is calculated M times with M sampled residuals and for all R bootstrapped coefficient samples, resulting in RM error samples. The 95% prediction interval estimate is then determined by taking the 2.5^{th} and 97.5^{th} percentiles of the error samples and adding them to the prediction y_p.

Year	Coefficient Estimate	Lower CI	Upper CI
1991	51916.79	47199.70	56517.20
1992	52275.41	47223.64	56822.11
1993	53864.16	48792.56	57004.43
1994	52644.58	47196.40	57247.51
1995	53225.61	47699.61	57629.29
1996	53091.31	48138.02	57743.04
1997	52774.51	46849.13	57310.79
1998	52844.75	47459.47	57209.36
1999	52753.37	47166.47	57255.10
2000	52951.71	48099.73	57385.47
2001	52499.13	47904.75	56825.61
2002	53489.65	48508.04	57967.75
2003	53506.74	47526.10	56859.61
2004	52075.97	47240.80	56856.29
2005	52519.53	47079.21	56718.19
2006	52566.31	47783.42	57232.69
2007	51584.87	46544.46	56049.31
2008	52840.27	46899.84	56892.45
2009	53100.22	47483.44	57310.86
2010	53317.41	48356.07	57609.44
2011	52657.48	46276.02	57119.83
2012	51776.59	46942.20	56415.80
2013	53272.37	48960.34	56701.56

Table 1: Bootstrapped 95% prediction intervals for annual maximum peak demand fixing year, month, and weekend data to 2013/2014, and varying historic climate data through each year.

Taking these daily predictions, we find the maximum demand within each year, leading to Figure (9) and Table (1). Although there is some fluctuation in the predicted maximum demand for the 2013 Winter period, there seems to be no obvious trend, with the prediction being a similar value for all the different years. This aligns with the distribution of TE, wind, and solar for each year being generally consistent, with no obvious trend (Figure 11). When comparing the actual maximum demand for each year to that year's prediction using the 2013 model, it is clear to see that the effects of the year are substantial compared to that of the climate.

5-10 Year Prediction

Before commenting on the model's predictive ability for future times, we first investigate the model's predictive ability 5 and 10 years in the future, within the historic data. To do this, we take a start-year cut-off and remove historic data after this start year, then use the historic data to predict demand D-years in the future (with prediction bootstrapping). This is repeated for all start-year cut-offs from

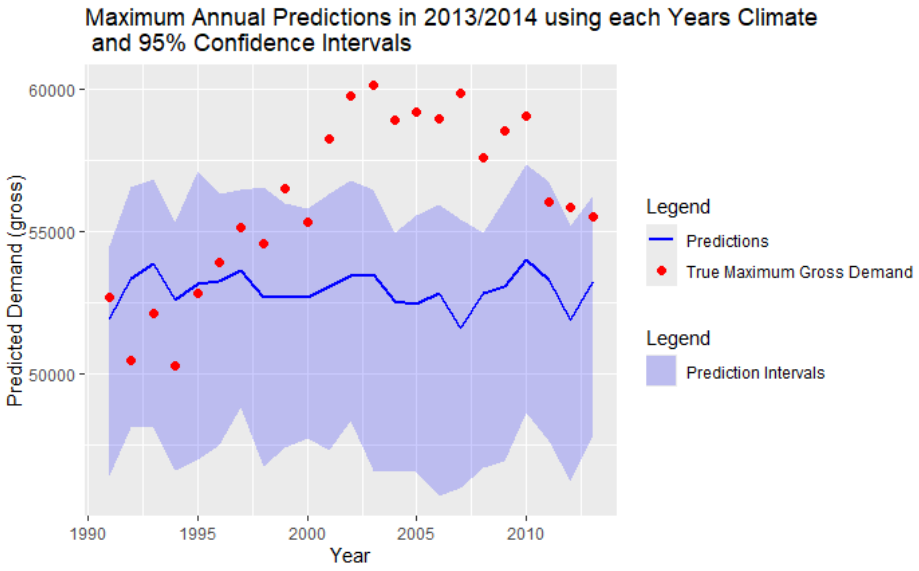


Figure 9: Bootstrapped 95% prediction intervals for annual maximum peak demand fixing year, month, and weekend data to 2013/2014, and varying historic climate data through each year.

1992 to $(2013-D)$, as we do not have historic demand data beyond a start-year of 2013 to compare our predictions to.

Due to our focus being on the maximum demand, we perform this process for each day and extract the predicted annual peak demand. We then compare these predictions with the true maximum annual peak demand from the historic data, resulting in Figures 10 and 12, and Tables 2 and 5.

Both 5 and 10 year ahead predictions perform very poorly in the first few years, most likely due to an absence of previous data. From 2005 onwards, the 5-year predictions tend closer to the actual values, though the prediction interval somewhat overestimates. Then from around 2010, there starts to be a downward trend, with demand predictions being considerably lower than the true values. In contrast, the 10-year predictions, though closer to the actual peak than before, start to overestimate from 2010. For most years, the 10-year prediction interval is also more than twice as wide as the 5-year one. This tendency for over- and under-estimating with large prediction intervals does not allow for proper planning of electricity transmission and generation, where shortfall prevention is imperative, while also making sure not to over-inflate the budget.

While the model's performance in predicting future peak demand for the pre-2014 data points towards a lack of accuracy in future demand prediction, our lack of knowledge about the future further exacerbates this inability. Firstly, unprecedented changes in climate may occur, increasing the time periods where peak demand is probable, which will decrease relevance of the analysis done using previous weather conditions; this might also change the behaviours of energy usage. There may be a change in installed solar and wind generation, those of which belonging to individual households are significantly harder to collect data about. Innovations in technology could lead to an increase in electricity usage, or instead, optimize electricity generation such that planned or existing infrastructures must be scrapped. Other unpredictable changes or events, such as a switch to a 4-day work week or a global pandemic such as COVID-19, are also likely to affect the model's reliability. As such, it is hard to make predictions about future peak demand.

Year	Prediction	Lower CI	Upper CI	Max Gross
2001	73991.62	34972.74	133647.24	58291.5
2002	47377.82	21165.10	89879.27	59780.5
2003	40680.48	19612.39	74874.90	60137.5
2004	46043.26	30814.34	76033.35	58932.0
2005	53467.84	32127.67	66732.06	59226.0
2006	61488.19	46917.82	79729.56	58970.5
2007	60737.39	52694.67	76598.82	59859.0
2008	62630.16	56114.45	80838.69	57630.5
2009	60694.10	59313.26	76732.33	58563.0
2010	57533.26	50663.61	69344.49	59870.0
2011	53510.96	43845.42	64148.41	56050.0
2012	48677.94	35423.19	65139.63	55884.5
2013	45622.54	39503.75	61013.23	55521.0

Table 2: Bootstrapped 95% prediction interval estimates of predictions 5 years in the future using start-year cut-off data, alongside maximum gross values.

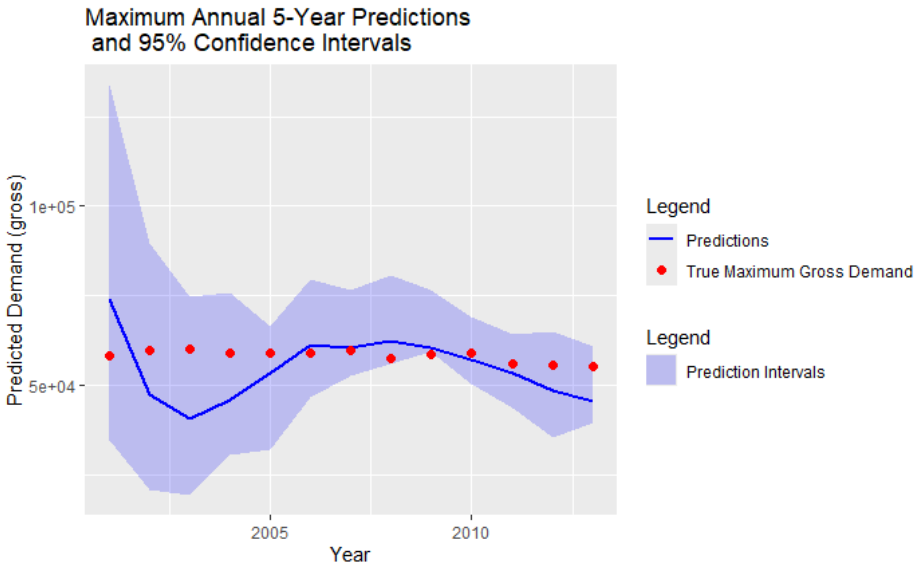


Figure 10: Bootstrapped 95% prediction interval estimates of predictions 5 years in the future using start-year cut-off data, alongside maximum gross values.

5 Conclusion

After assessing and attempting to mitigate important confounding factors, and evaluating various temperature metrics, we arrived at our full model that performed best out of all candidate models in metrics R^2, AIC, and CV-MSE, further finding NESO's temperature metric to be most suitable.

Upon fixing the year to 2013, we found climate behaviours incurred minimal fluctuations in demand compared to the true peak demand of each year. Then when assessing the model's ability of 5 and 10 year ahead predictions with known annual maximums, we found that both may not have sufficient accuracy to aid NESO in electricity demand security planning. This poor prediction ability is only exacerbated by the magnitude of uncertainty about future conditions regarding weather, technology, and even consumer behaviour.

Improving on the dataset's limitations, such as gathering more data and using different statistical analysis methods, could help improve future research. In particular, better confounding between covariates, analysing using different data groupings, considering a different temperature metric (as

opposed to the population-weighted average from MERRA1), and breaking down model analysis by energy type (e.g. appliances), may lead to an increase in model accuracy. Research into the effects of region-specific factors could also improve the model, while also allowing better planning for shortfalls (e.g. more energy storage structure in certain areas).

References

[1] Anthony C Atkinson, Marco Riani and Aldo Corbellini. 'The box–cox transformation: Review and extensions'. In: *Stat. Sci.* 36.2 (May 2021).

[2] Amy Wilson. *Statistical Case Studies 2024/25 Semester 2 Part 2*. Accessed: 25/02/2025. The University of Edinburgh, Feb. 2025.

Appendix

Weekday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Intercept	53,816.15	53,775.93	53,904.85	53,804.45	52,628.30	48,297.94	47,393.38
Gradient	-446.21	-402.93	-436.68	-450.54	-506.49	-566.33	-549.37

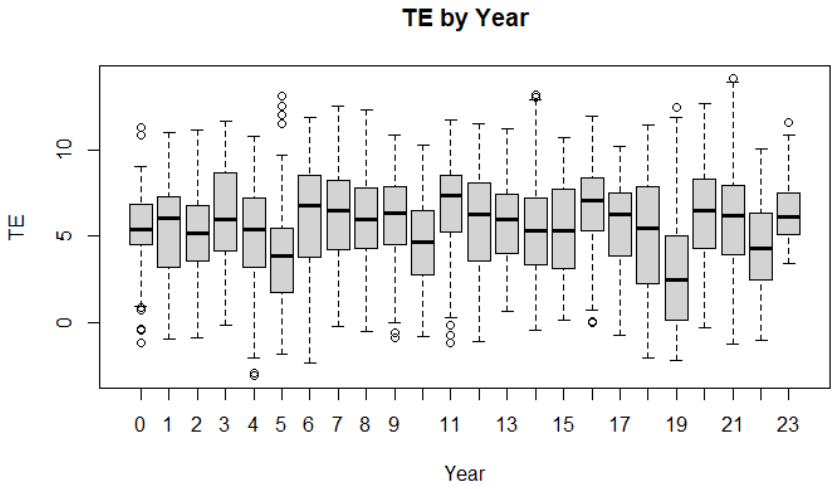
Table 3: Gradients and intercepts of Demand vs TE linear relationship for each weekday (2 d.p.)

Coefficient	**Estimate**	**Lower CI**	**Upper CI**
Intercept	47670.477820	46192.22904	48748.480460
Year	618.360544	331.53138	984.933727
$\text{Year}^{(2)}$	39.416229	7.30042	68.842178
$\text{Year}^{(3)}$	-2.571255	-3.42991	-1.739357
MonthFebruary	-1007.976683	-1360.36429	-730.72664
MonthMarch	-2857.322043	-3263.03522	-2504.608438
MonthNovember	594.320534	229.84602	1000.758126
MonthDecember	-1200.763795	-1815.42067	-619.60891
WeekendYes	-6402.523437	-6726.723056	-6246.694999
TE	-234.593300	-456.98346	-45.754675
$\text{TE}^{(2)}$	-22.827825	-35.69860	-7.411137
Wind	842.755927	252.53828	1524.811652
Solar_S	-14854.791173	-20919.31915	-8977.637963
Adjusted r^2		0.7501981	

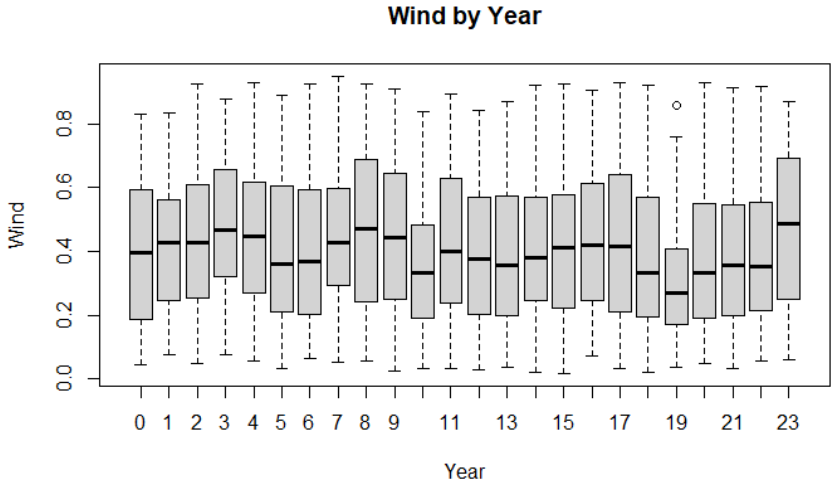
Table 4: Bootstrapped coefficient estimates, their 95% confidence intervals, and an adjusted r^2 value from these coefficient estimates.

Year	Prediction	Lower CI	Upper CI	Max Gross
2005	138021.443	-119055.08	355597.70	58970.5
2006	10517.339	-167208.95	180283.94	59859.0
2007	-1829.824	-115326.89	128687.51	57603.5
2008	30540.780	-46485.22	149344.28	58563.0
2009	43164.506	-23413.92	115953.53	59070.0
2010	65819.923	20804.07	112419.92	56050.0
2011	70327.205	32319.53	120223.84	55884.5
2012	65341.365	39363.79	93275.88	55521.0
2013	60837.720	43160.61	91480.87	51856.0

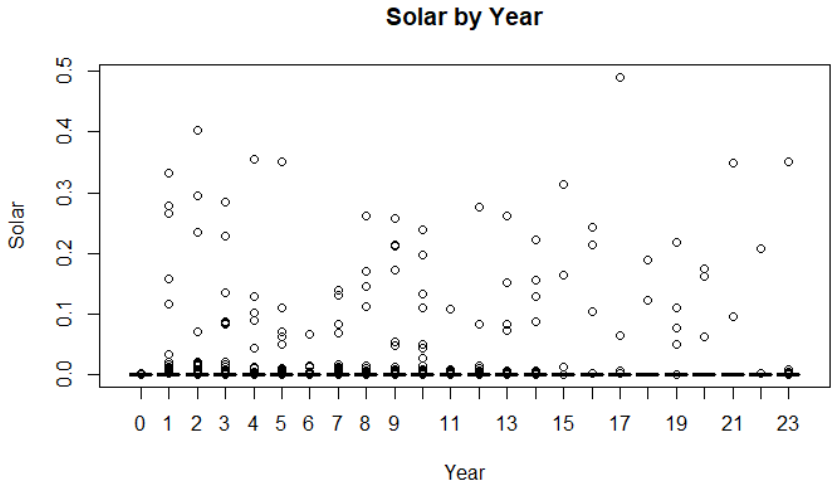
Table 5: Bootstrapped 95% prediction interval estimates of predictions 10 years in the future using start-year cutoff data, alongside maximum gross values.



(a) Box plot of TE by year



(b) Box plot of wind by year)



(c) Box plot of solar by year)

Figure 11: Box plots of the climate covariates by year

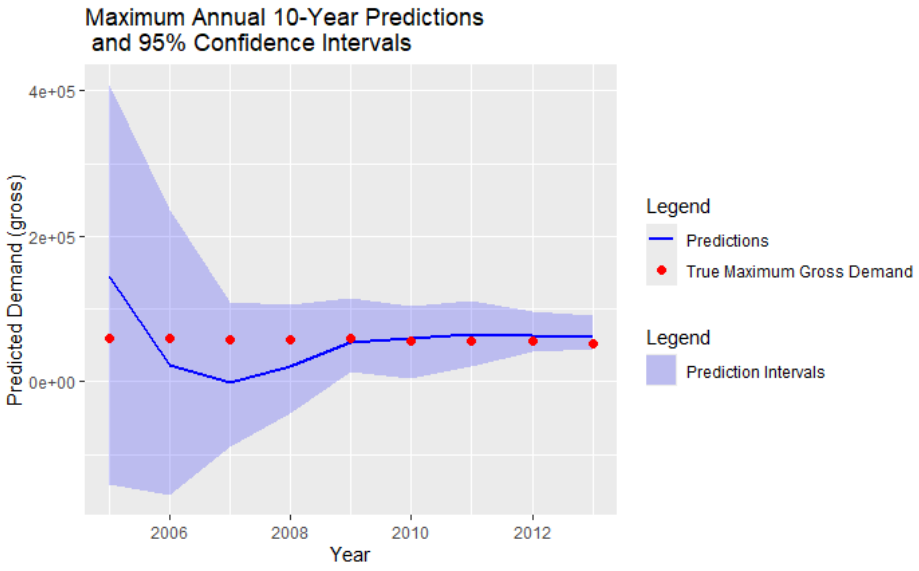


Figure 12: Bootstrapped 95% prediction interval estimates of predictions 5 years in the future using start-year cutoff data, alongside maximum gross values.