Contents

Contents 1

1 Executive Summary 3

2 Data Description 4
- 2.1 Overview of Datasets 4
- 2.2 Key Variables 4
- 2.3 Handling of Atypical Demand Periods 4

3 Linear Regression Model 6
- 3.1 Model Selection and Evaluation 6
  - 3.1.1 Initial Model Specification 7
- 3.2 Evaluation of Temperature Variables 8
  - 3.2.1 Selection of Temperature Smoothing Method 8
  - 3.2.2 Selection of Averaging Window 8
- 3.3 Introduction of Daylight Hours 9
- 3.4 Final Model Specification 9
- 3.5 Model Evaluation 9
  - 3.5.1 Residual Analysis 10
  - 3.5.2 Predictive Performance Analysis 11

4 Quantile Regression Model 13
- 4.1 Model Selection and Evaluation 13
  - 4.1.1 Pinball Loss Function 13
- 4.2 Final Model Specification 14
- 4.3 Model Evaluation 14

5 Scenario Analysis: Testing Alternative Weather Conditions 16
- 5.1 Motivation and Exploratory Analysis 16
- 5.2 Methodology 18
- 5.3 Results and Interpretation 19
  - 5.3.1 Linear Regression Model 19
  - 5.3.2 Quantile Regression Model 20
  - 5.3.3 Conclusion 21

6 Limitations and Future Considerations 22

A Initial Linear Regression Model Comparison 23

B Temperature Variable Selection 28

C Temperature Averaging Window Comparison 30

D Daylight Hours Model Comparison 32

E Quantile Selection Process 34

F Quantile Regression Model Metrics **36**

G Quantile Regression Model Selection **38**

Bibliography **40**

1 Executive Summary

The National Electricity System Operator (NESO) is responsible for ensuring the secure and reliable operation of Great Britain's electricity system. A key component of this role is forecasting peak electricity demand to support long-term infrastructure planning and balance system security with cost-efficiency.

NESO currently uses linear regression models to estimate peak evening demand during winter months, when the system is under greatest stress. These models incorporate weather-related variables and temporal indicators [1]. However, electricity usage also evolves over time due to broader structural factors—such as changes in population, technology adoption, and energy efficiency—that are not directly observable in the data [2]. To account for these long-term trends, NESO applies a 'year effect', which can be adjusted to project historical demand to a specified target year.

This report builds on NESO's approach by refining model inputs, evaluating alternative specifications, and testing model performance under different historical weather scenarios. Two complementary models were developed using historical winter data. The first refines the existing linear regression framework by incorporating additional predictors related to wind conditions and daylight hours.

A key refinement involved how temperature was modelled. Rather than relying solely on NESO's existing variables, the analysis found that averaging hourly air temperatures between 11:00 and 18:00 provided a more accurate predictor of daily peak demand. To better reflect the persistence of cold weather, a weighted average of this variable over recent days was also introduced. Together, these enhancements improved model fit, particularly under colder conditions.

While the linear model performs well overall, it tends to underestimate electricity demand on the highest-demand days. To address this, a second model was developed using quantile regression, which estimates the conditional 0.95 quantile of demand—that is, the level expected to be exceeded on only 5% of days, given the conditions on that day. This approach provides a more cautious estimate under high-demand scenarios and is better suited to assessing system risk.

To examine the influence of weather on peak demand, a scenario analysis was conducted for the winter of 2013/14. By substituting in colder weather conditions from earlier winters while holding structural factors constant, both models predicted substantially higher peak demand. This highlights the sensitivity of demand to meteorological variation and the importance of stress-testing demand forecasts under alternative conditions.

While both models perform well on historical data, several limitations should be considered when projecting further into the future. Firstly, the relationships captured by the models may be affected by developments such as the electrification of heating and transport, the growth of decentralised generation, behavioural change, and increasing climate uncertainty. Secondly, although a system is in place to estimate the 'year effect' for future planning, no equivalent exists for projecting demand sensitivity to temperature or wind—both of which are inputs required by the models. Addressing this gap may require the development of new methodologies beyond the scope of this report.

2 Data Description

2.1 Overview of Datasets

This analysis draws on two datasets. The primary dataset contains historical winter records of electricity demand, weather conditions, and temporal indicators, forming the basis for identifying relationships between electricity usage and external factors. A secondary dataset provides hourly temperature measurements, enabling refinement of temperature-based predictors.

The analysis is restricted to winter months, when peak demand and the risk of supply shortages are greatest. Each observation corresponds to electricity demand at 18:00, which consistently marks the highest point in the daily load curve. The data span multiple years, allowing for the study of yearly variation and long-term demand trends.

2.2 Key Variables

The main variables used in the modelling are summarised below:

- **Gross Demand (MW):** The response variable, representing total electricity demand at 18:00 across Great Britain, as reported by NESO. This is treated as the daily peak demand.
- **Temperature Variables:** Multiple temperature measures are considered, including NESO's existing formulations and newly constructed variables, to capture both short- and longer-term effects on demand.
- **Weather Conditions:** Estimated capacity factors for wind and solar generation, representing the proportion of installed capacity actively generating at 18:00 on a given day. These serve as proxies for wind speed and sunlight, which may influence electricity demand.
- **Temporal Factors:** Variables such as the day of the week and the number of days since 1st November, used to represent weekly and seasonal demand patterns.
- **Year Effect:** A categorical variable denoting the start year of each winter period, included to capture long-term structural changes in electricity use that are not directly observable.

These variables form the foundation for model development, supporting the identification of key drivers of peak demand and enabling more accurate and interpretable forecasts.

2.3 Handling of Atypical Demand Periods

Exploratory analysis identified a consistent reduction in electricity demand during the Christmas and New Year period, likely driven by reduced industrial activity, business closures, and changes in household routines. As shown in Figure 2.1, there is a clear and sustained dip in gross demand across all years during this period.

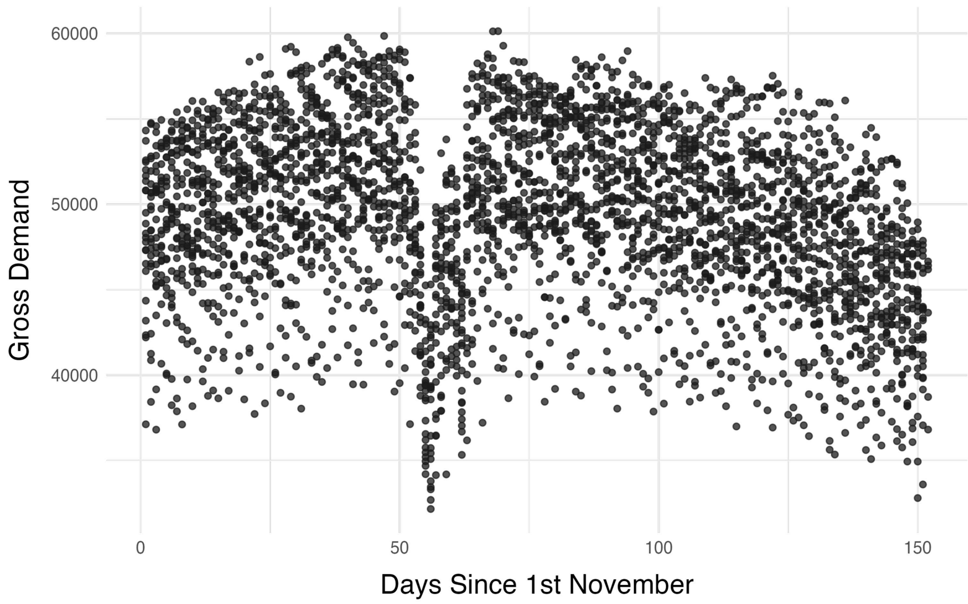

Figure 2.1: Gross electricity demand plotted against days since 1st November.

Since the aim is to model typical peak demand, including holiday observations could distort parameter estimates, as demand during this period is unrepresentative and may obscure key relationships with temperature and seasonality.

To mitigate this, all observations from 22nd December to 4th January were excluded prior to model fitting. This window was selected based on the pattern visible in Figure 2.1, and further justified by residual plots from preliminary models, which showed systematic deviations when the holiday period was included. Removing these observations ensures the model focuses on conditions most relevant to peak demand estimation and reduces the influence of atypical behaviour on the fitted relationships.

3 Linear Regression Model

Linear regression provides a natural starting point for modelling electricity demand due to its transparency, statistical robustness, and alignment with NESO's current methodology. It allows for the estimation of how demand responds to explanatory variables, while serving as a baseline for more advanced models. Importantly, linear regression also supports the inclusion of a 'year effect', enabling long-term structural changes in electricity usage to be modelled separately from weather-driven variation.

3.1 Model Selection and Evaluation

The linear regression model is developed iteratively, refining the set of explanatory variables to maximise predictive accuracy while maintaining interpretability. The initial model focuses on temperature-based predictors, with other variables introduced based on theoretical justification and empirical performance.

Temperature is a key driver of electricity demand during winter, as colder conditions increase heating requirements. However, demand does not depend solely on the ambient temperature at a specific point in time. Instead, it reflects both current and recent thermal conditions due to building heat retention and behavioural responses. To capture these dynamics, NESO defines two temperature variables based on the 18:00 demand being modelled.

The observed air temperature at hour h, denoted TA_h, reflects the external temperature at that specific time. A short-term average temperature, TO_{18}, is defined as the mean of the four hourly temperatures from 15:00 to 18:00:

$$TO_{18} = \frac{TA_{18} + TA_{17} + TA_{16} + TA_{15}}{4}$$

To incorporate persistent cooling effects, NESO defines an 'effective temperature', TE_{18}, by recursively averaging the current TO_{18} with the previous day's effective temperature, TE_{18-24}:

$$TE_{18} = \frac{TO_{18} + TE_{18-24}}{2}$$

This formulation allows the model to reflect both recent and longer-term thermal conditions, making TE_{18} a more stable and informative predictor of electricity demand.

A structured modelling process is used to balance predictive performance with simplicity. Beginning with temperature variables, additional covariates are introduced sequentially to capture systematic variation in demand. These include indicators for the day of the week, seasonal trend terms, and proxies for sunlight and wind conditions.

Model refinements also include interaction terms, polynomial transformations, and alternative formulations of key variables. Variable selection is guided by forward and backward stepwise procedures, informed by statistical significance and model diagnostics.

Model performance is evaluated using the following criteria:

- **Adjusted R^2:** Measures the proportion of variance in demand explained by the model, adjusted for the number of predictors to penalise overfitting.
- **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC):** Penalised likelihood-based metrics that evaluate model fit while accounting for model complexity; lower values indicate better performance.
- **Root Mean Squared Error (RMSE) of Top 10%:** Assesses prediction error on the highest 10% of demand observations, reflecting the model's ability to forecast peak demand.
- **Residual Diagnostics:** Standard plots used to verify assumptions of linearity, homoscedasticity, normality, and independence in the residuals.

As all models use the same response variable, these metrics can be directly compared across specifications. The rationale for model construction and intermediate results are provided in Appendix A.

3.1.1 Initial Model Specification

The initial linear regression model incorporates effective temperature, day-of-week indicators, a quadratic seasonal trend, and a proxy for wind speed. This specification serves as a baseline for subsequent refinement. The model is defined as:

$$D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TE_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DSN_t + \delta_2 DSN_t^2 + \theta_1 WE_t + \epsilon_t$$

where t indexes individual days, and:

- $W_{i,t}$ is an indicator for winter i, equal to 1 if day t falls in winter i, and 0 otherwise. This captures the 'year effect'.
- $TE_t W_{i,t}$ introduces a winter-specific interaction, allowing the influence of effective temperature TE_t to vary across years.
- $DOW_{j,t}$ is an indicator for the day of the week (Sunday to Friday), with Saturday as the reference category.
- DSN_t and DSN_t^2 denote the number of days since 1st November and its square on day t, to capture seasonal variation.
- WE_t is the estimated wind generation capacity factor on day t, used as a proxy for wind conditions.
- $\alpha, \beta_i, \gamma_i, \omega_j, \delta_1, \delta_2, \theta_1$ are model parameters estimated via ordinary least squares.
- ϵ_t is the model's residual error term on day t.

3.2 Evaluation of Temperature Variables

Temperature is a key driver of electricity demand, making the choice of temperature variable critical to model performance. To assess whether alternative formulations could improve predictive accuracy relative to NESO's defined effective temperature (TE), several variants were evaluated. These were assessed based on their correlation with demand and their performance within the full linear regression model using adjusted R^2, AIC, BIC, and RMSE (focused on the top 10% of demand observations).

3.2.1 Selection of Temperature Smoothing Method

To account for lagged and cumulative temperature effects, several smoothing techniques were applied to the NESO-defined short-term average temperature (TO). The following approaches were considered:

- **Multi-day moving averages:** Simple averages over the past 3 or 7 days, capturing sustained cold spells.
- **Centred moving averages:** Symmetric averages that smooth fluctuations and account for the impact of short-term forecasts.
- **Daily temperature change:** The day-to-day difference in temperature, reflecting rapid shifts in weather.
- **Weighted averages:** Exponentially decaying averages that assign greater weight to more recent temperatures.

A weighted average incorporating temperatures from the two preceding days consistently outperformed the other smoothing methods across all evaluation metrics. It was therefore selected as the basis for further refinement. Full comparative results and details of the weightings are provided in Appendix B.

3.2.2 Selection of Averaging Window

The next consideration was the choice of base temperature variable. Several intraday averaging windows were tested, extending NESO's original definition of TO (average temperature from 15:00 to 18:00) to include both earlier and later hours.

The most effective formulation was the average temperature from 11:00 to 18:00:

$$T_{\text{11-18}} = \frac{TA_{11} + TA_{12} + \cdots + TA_{18}}{7}$$

This variant exhibited the strongest correlation with demand and delivered the best model performance when used within the weighted average structure. Full results are provided in Appendix C.

Accordingly, the final temperature variable, TW_t, applies decaying weights to the smoothed 11:00–18:00 average over three consecutive days:

$$TW_t = 0.5 \cdot T_{11\text{-}18,t} + 0.3 \cdot T_{11\text{-}18,t-1} + 0.2 \cdot T_{11\text{-}18,t-2}$$

Following the introduction of TW_t, the remainder of the model structure was re-evaluated, but no further changes were found to be necessary.

3.3 Introduction of Daylight Hours

The variable DSN, which measures the number of days since 1st November, primarily acts as a proxy for daylight availability. Since natural light levels can influence electricity demand, DSN was replaced with a directly computed daylight hours variable, DH, to improve both interpretability and predictive accuracy.

Daylight hours were calculated for the population-weighted centre of England, based on 2014 census data [3]. This location offers the most accurate available approximation of the demand-weighted centre of the NESO-managed electricity grid, as Scotland lies outside NESO's purview and no equivalent population centre was available for a combined England and Wales.

Model performance was assessed using the same criteria as in earlier specifications, with comparative results reported in Appendix D. Replacing DSN with DH led to a modest but consistent improvement in model fit, offering a more interpretable and direct measure of seasonal variation.

3.4 Final Model Specification

The final linear regression model incorporates the optimised temperature variable TW_t, a quadratic daylight hours term, weekday effects, relative wind generation, and winter-specific effects with temperature interactions. The full specification is given by:

$$D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TW_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DH_t + \delta_2 DH_t^2 + \theta_1 WE_t + \epsilon_t$$

This model captures both structural and weather-related influences on demand, while remaining interpretable and statistically robust.

3.5 Model Evaluation

Evaluating model performance is essential to ensure the reliability of any forecasting tool—particularly when it supports long-term infrastructure planning. Both statistical validity and predictive accuracy are assessed for the final model.

3.5.1 Residual Analysis

Residual diagnostics are used to evaluate whether the key assumptions of linear regression—linearity, homoscedasticity, normality, and the absence of influential outliers—are reasonably satisfied.

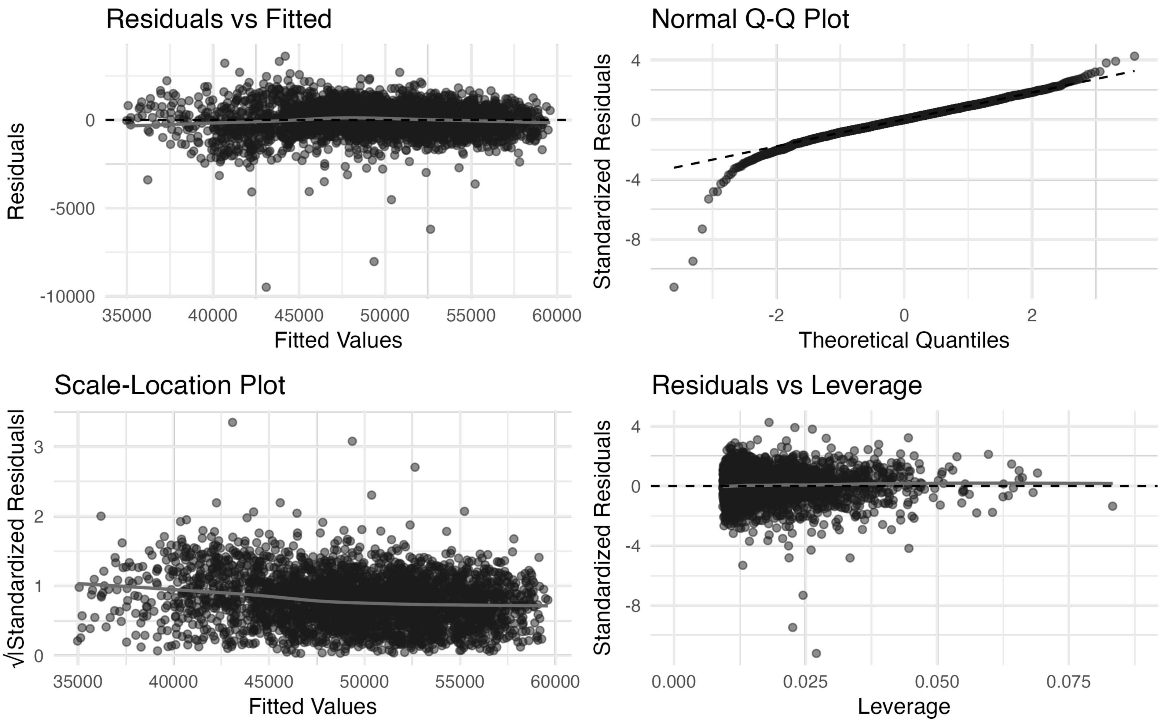

Figure 3.1: Residual diagnostics for the final linear regression model.

- **Residuals vs Fitted Values:** The residuals are centred around zero with no clear patterns, indicating that the assumption of linearity is broadly satisfied. Slight heteroscedasticity is observed, with higher variance at certain fitted values.
- **Normal Q-Q Plot:** Most residuals follow the theoretical quantiles closely, supporting the assumption of approximate normality. Some deviation is observed in the lower tail which may suggest skewness or outliers.
- **Scale-Location Plot:** The spread of residuals appears relatively constant, although a mild curvature supports the presence of minor heteroscedasticity.
- **Residuals vs Leverage:** No observations exceed the Cook's distance threshold, indicating that no single data point has undue influence on the model estimates.

Overall, these diagnostics suggest that the model offers a statistically valid approximation of the data. While minor departures from homoscedasticity and normality exist, they are not severe enough to undermine inference. Nevertheless, to improve robustness—particularly for peak demand estimation—a quantile regression model is introduced in the following chapter. Quantile regression relaxes many of the residual assumptions required by ordinary least squares (OLS) regression.

3.5.2 Predictive Performance Analysis

To assess the model's ability to replicate historical demand patterns, several graphical diagnostics compare predicted values to observed daily peak demand.

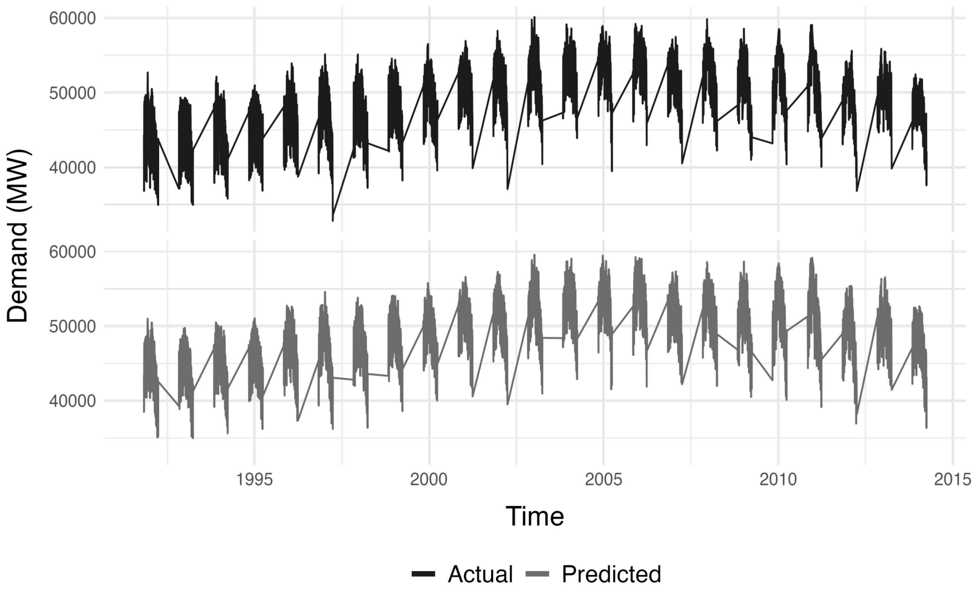

Figure 3.2: Actual and predicted daily peak electricity demand over time.

Figure 3.2 shows that the model successfully reproduces long-term trends and seasonal fluctuations in electricity demand, with predicted values closely tracking observed peaks and troughs throughout the sample period.

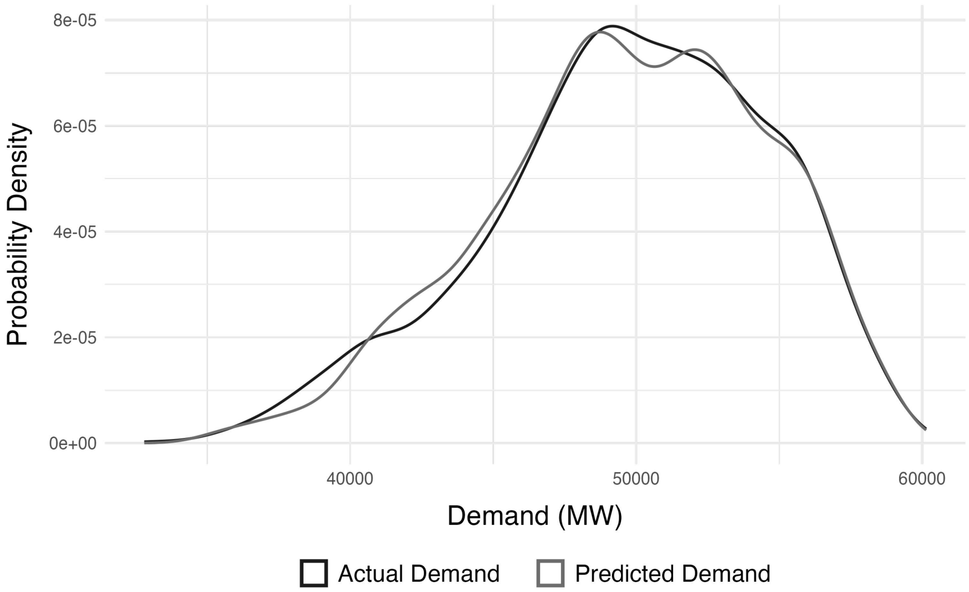

Figure 3.3: Probability density functions (PDFs) of actual and predicted daily peak electricity demand.

To evaluate how well the model captures the overall distribution of demand, Figure 3.3 compares the time-collapsed probability density functions (PDFs) of actual and predicted values. These

PDFs show the relative frequency of different demand levels across the full sample. The close alignment between the two curves indicates that the model accurately captures both the shape and spread of daily peak demand, supporting its overall goodness of fit.

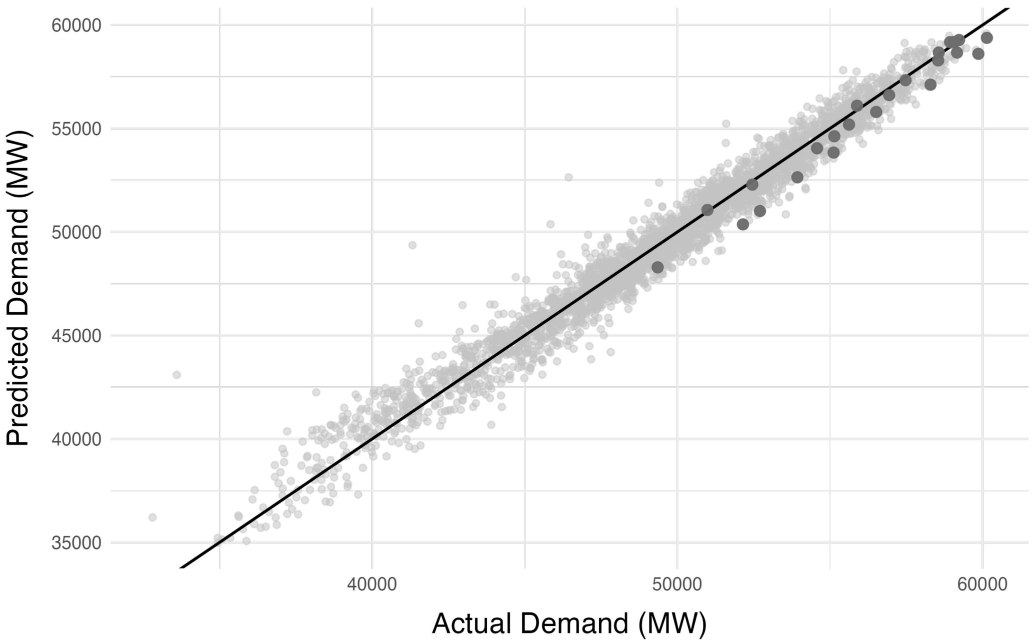

Figure 3.4: Scatter plot of actual vs predicted daily peak demand. Highlighted points indicate the peak demand in each winter. The black line represents $y = x$.

Finally, Figure 3.4 plots actual versus predicted values, with coloured points highlighting the highest daily demand in each winter. Most observations lie close to the identity line, indicating strong predictive performance. However, the winter peak values (highlighted points) consistently fall to the right of the line, suggesting systematic underestimation of maximum demand.

Given the importance of accurately forecasting maximum demand for NESO's operational planning, this motivates the development of a quantile regression model, introduced in the next chapter, which aims to improve predictive performance in the upper tail of the demand distribution.

4 Quantile Regression Model

The previous model used ordinary least squares (OLS) regression to estimate the conditional mean of electricity demand. While effective at capturing average patterns and broad seasonal trends, this approach may fall short when forecasting extreme demand events.

Quantile regression offers a more flexible alternative by estimating conditional quantiles of the response variable. Unlike OLS, it allows the influence of explanatory variables to vary across different points in the distribution, providing a more detailed view of how demand behaves under specific conditions [4]. This makes it particularly well suited to peak demand forecasting, where the drivers of extreme usage may differ from those affecting average behaviour.

To explore this, quantile regression models were estimated at multiple levels ($q = 0.90, 0.95, 0.98$), allowing for a comparison between greater focus on peak conditions and the statistical stability of model estimation. Higher quantiles place more weight on extreme outcomes but are fitted using fewer observations, leading to increased parameter uncertainty. The 0.95 quantile ($q = 0.95$) was selected as the most appropriate level, balancing predictive relevance with robustness. Further details of the quantile selection process are provided in Appendix E.

4.1 Model Selection and Evaluation

Quantile regression is evaluated using the Pinball Loss function, an asymmetric analogue to squared error loss that measures accuracy at a specified quantile. This makes it the natural criterion for model selection.

4.1.1 Pinball Loss Function

The Pinball Loss (or Check Loss) function quantifies error relative to a specified quantile [4]. For quantile q, it is defined as:

$$L_q(y, \hat{y}) = \begin{cases} q(y - \hat{y}), & \text{if } y \geq \hat{y} \\ (1 - q)(\hat{y} - y), & \text{if } y < \hat{y} \end{cases}$$

where y is the observed value and $\hat{y}$ is the predicted value at quantile q.

Lower Pinball Loss values indicate better predictive accuracy at the chosen quantile. Unlike squared error loss, which penalises over- and under-predictions equally, Pinball Loss accounts for the asymmetric nature of quantile estimation by weighting deviations differently depending on their position relative to the target quantile.

To complement Pinball Loss, Generalised Cross-Validation (GCV) is used to assess out-of-sample performance while adjusting for model complexity. In addition, quantile-adjusted versions of AIC and BIC—denoted AIC_q and BIC_q—are used to support model comparison. These are discussed further in Appendix F.

4.2 Final Model Specification

Various temperature formulations, temporal indicators, and weather-related variables were tested to identify the most accurate specification for forecasting the upper tail of electricity demand. Full comparative results are provided in Appendix G. The final model is given by:

$$Q_{0.95}(D_t) = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TW_t W_{i,t} + \theta_i WE_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DH_t + \delta_2 DH_t^2 + \epsilon_{0.95,t}$$

where:

- $Q_{0.95}(D_t)$ denotes the estimated 0.95 quantile of Gross Demand on day t, conditional on the predictor variables.
- $WE_t W_{i,t}$ introduces a winter-specific interaction with relative wind generation, allowing the effect of wind conditions on demand to vary across years.
- $\epsilon_{0.95,t}$ is the residual term associated with the 0.95 quantile, capturing distributional asymmetry and heteroscedasticity.

This model improves accuracy for high-demand scenarios and addresses the underestimation of extreme values observed in the linear regression model.

4.3 Model Evaluation

Unlike linear regression, quantile regression does not rely on assumptions such as homoscedasticity or normally distributed residuals. As a result, evaluation focuses entirely on predictive performance, with particular attention to high-demand cases.

Since the model estimates the conditional 0.95 quantile of electricity demand, predicted values are expected to exceed actual values on approximately 95% of days. This is not a flaw but a defining feature, as the model is constructed to reflect upper-tail demand behaviour.

Figure 4.1 compares actual and predicted values. As expected, the majority of points lie below the identity line $y = x$, supporting the interpretation of the model as offering an upper-bound estimate of daily peak demand.

The coloured points highlight the highest daily demand in each winter. Compared to the linear regression results (Figure 3.4), these peak values are captured more accurately. Many are now slightly overestimated, which is preferable in the context of NESO's objective to safeguard against underestimating demand during extreme conditions. Overall, the quantile regression model complements the linear approach by providing a more cautious estimate of peak usage, better supporting NESO's emphasis on system adequacy and risk management.

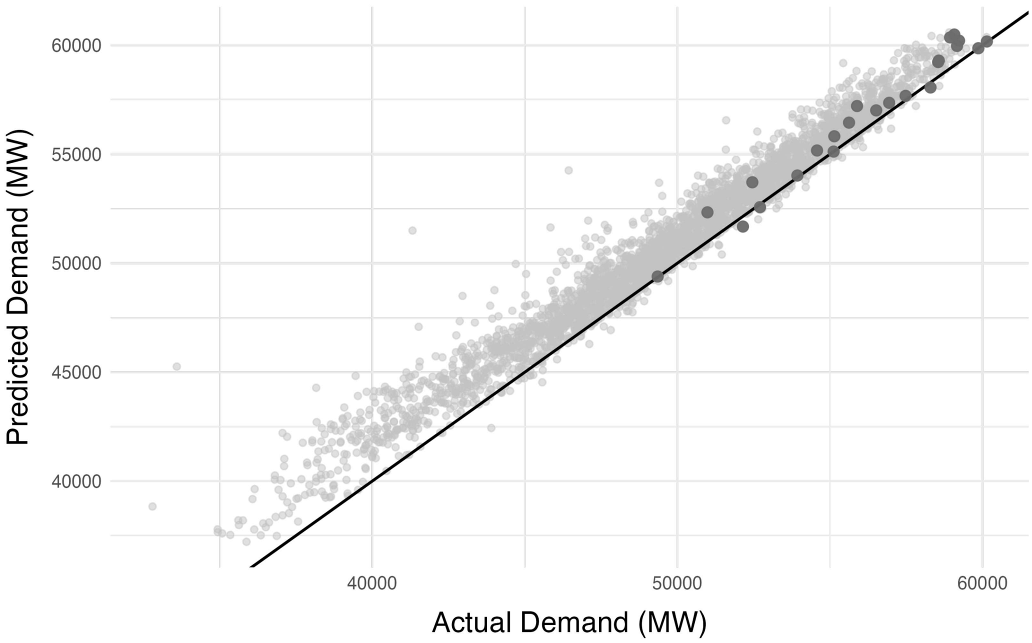

Figure 4.1: Scatter plot of actual vs predicted daily peak demand using the 0.95 quantile regression model. Highlighted points indicate the peak day in each winter. The black line represents $y = x$.

5 Scenario Analysis: Testing Alternative Weather Conditions

5.1 Motivation and Exploratory Analysis

To explore how peak electricity demand in the 2013/14 winter might have differed under alternative weather conditions, a scenario analysis is conducted using historical meteorological data. This approach isolates the impact of weather variation on peak demand while holding structural factors—represented by the 'year effect'—constant.

Throughout this section, winters are referred to by the year in which they begin. For example, the 2013/14 winter is referred to as 2013.

The 2013 winter is of particular interest due to a marked drop in peak electricity demand compared to previous years. Figure 5.1 shows the maximum daily demand for each winter, with the 2013 value noticeably lower than those observed throughout the 2000s and early 2010s.

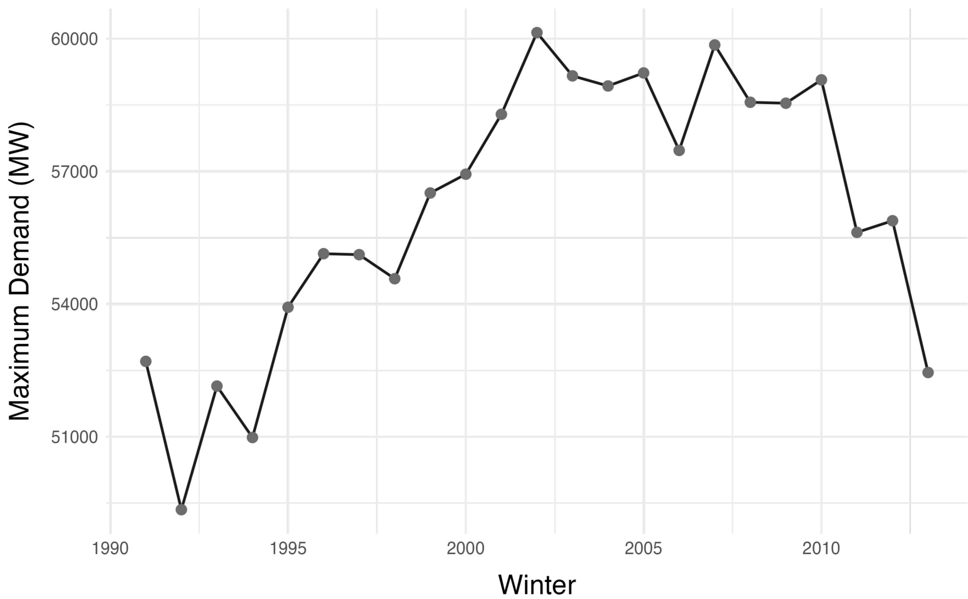

Figure 5.1: Maximum daily peak electricity demand for each winter.

To assess how weather conditions contributed to the reduced demand observed in 2013, Figure 5.2 presents the 7-day rolling average of the 18:00 British population-weighted temperature across all winters. Figure 5.3 shows the equivalent series for wind capacity factor, used as a proxy for wind speed.

The 2013 temperature trace remains consistently above those of most other winters, indicating relatively mild conditions. In contrast, wind conditions in 2013 fall within the typical range observed in other years, without systematic deviation. This suggests that temperature is the more likely driver of the unusually low demand observed during the 2013 winter.

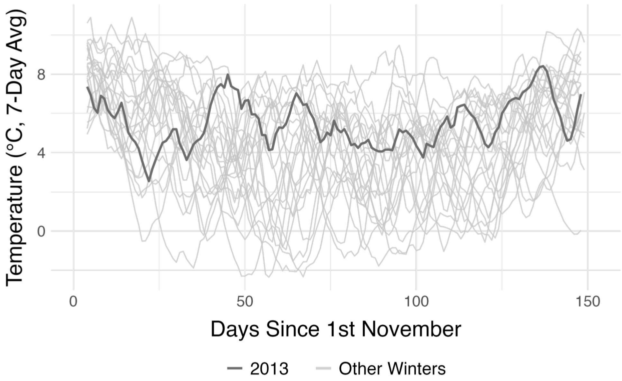

Figure 5.2: 7-day rolling average of temperature across winters, with the 2013 winter highlighted.

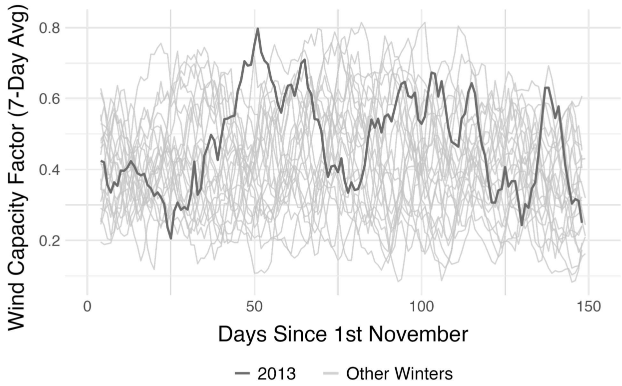

Figure 5.3: 7-day rolling average of wind capacity factor across winters, with the 2013 winter highlighted.

To examine conditions during periods when peak demand is most likely to occur, Figure 5.4 shows the distribution of the coldest 10% of 18:00 temperatures for each winter. The 2013 winter again stands out as one of the mildest in the dataset, with its coldest days generally warmer than those of other years.

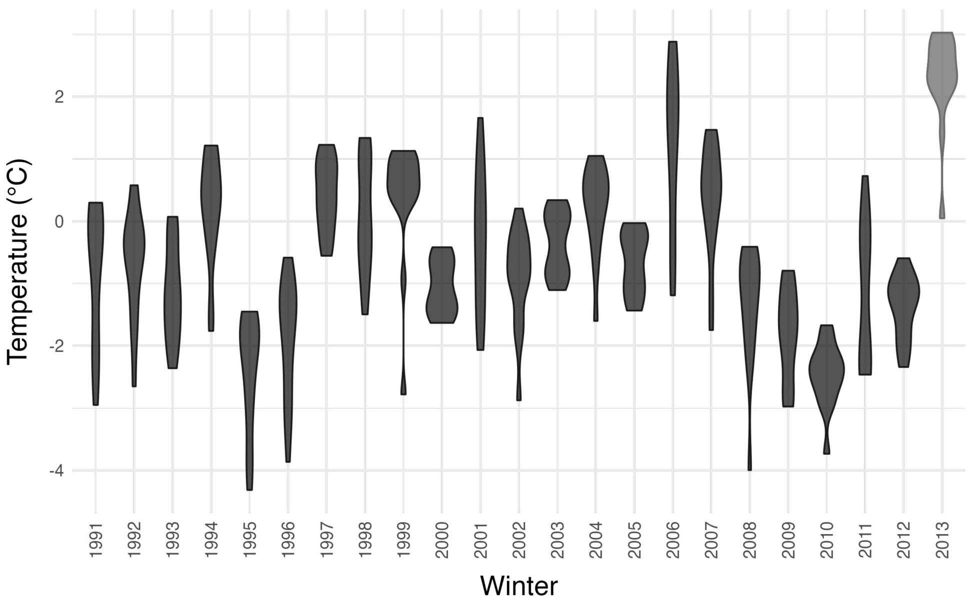

Figure 5.4: Distribution of the 10% coldest 18:00 temperatures in each winter.

Taken together, these findings suggest that the low peak demand observed in 2013 was primarily weather-driven—most notably due to higher-than-usual temperatures. This motivates the use of counterfactual simulations in which the 2013 winter is re-evaluated under historical weather conditions from other years, while maintaining the structural context of that year.

5.2 Methodology

To estimate how peak demand in the 2013 winter might have differed under alternative meteorological conditions, a series of simulations are conducted using historical weather data. The structural context is fixed to that of 2013, while daily weather inputs are substituted from earlier winters spanning 1991 to 2012.

The following weather-related variables are replaced in each scenario:

- Weighted temperature average (TW_t)
- Daylight hours (DH_t) and squared daylight hours (DH_t^2)
- Estimated relative wind generation (WE_t)

The winter label remains set to 2013 in all simulations, meaning the model continues to use the 2013-specific coefficients (e.g. β_{2013}, γ_{2013}, θ_{2013}). This ensures that differences in predicted demand reflect changes in weather alone, without interference from long-term structural variation.

As the models are based on time series data, residuals may exhibit autocorrelation—that is, correlation between observations across time. To assess this, the autocorrelation functions (ACFs) of residuals from both models are examined.

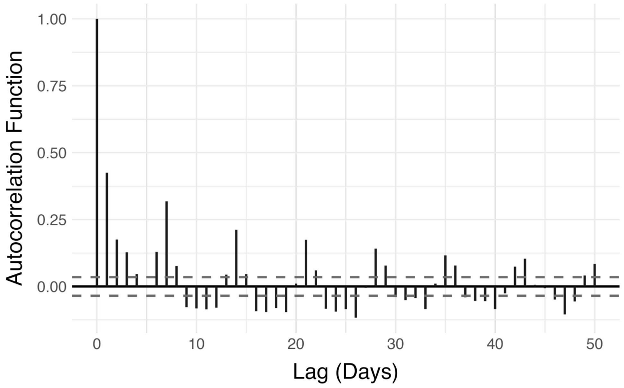

Figure 5.5: ACF of residuals in the linear regression model.

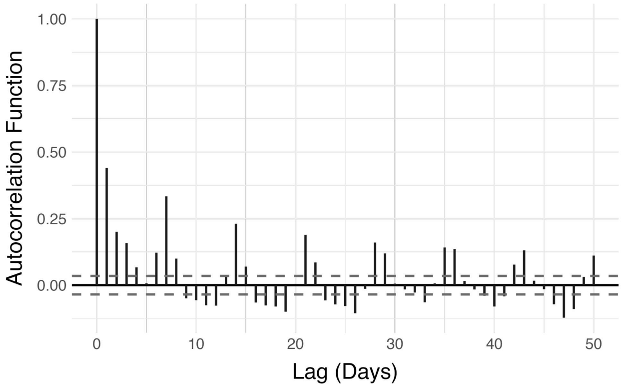

Figure 5.6: ACF of residuals in the quantile regression model.

Figures 5.5 and 5.6 indicate short-run temporal dependence, with autocorrelation diminishing substantially at a 5-day lag. To account for this, block bootstrapping with a block length of 5 days is used, preserving local dependence during resampling.

The simulation procedure for each historical winter proceeds as follows:

1. Exclude 2013 data and resample in 5-day blocks to generate 1,000 bootstrap datasets.
2. Fit the selected model (linear or quantile) to each resampled dataset, creating a matrix of estimated coefficients. Insert the 2013-specific coefficients into each matrix.

3. For each historical winter, replace the weather-related variables in the 2013 dataset with those from that winter, keeping all other inputs fixed.
4. Use the coefficient matrices to predict demand under each scenario, recording the maximum value in each bootstrap iteration.
5. For each scenario, calculate the mean of predicted maxima and construct 95% confidence or prediction intervals using the empirical distribution.

This procedure is repeated for each historical winter, allowing the sensitivity of 2013 peak demand to alternative weather conditions to be quantified under a fixed structural framework.

5.3 Results and Interpretation

5.3.1 Linear Regression Model

The linear regression model was applied to each scenario using substituted weather conditions from historical winters. Figure 5.7 presents the predicted means and 95% prediction intervals for the 2013 winter's maximum daily demand, with each interval corresponding to weather inputs from a different year. The black dashed line indicates the actual peak demand observed in 2013.

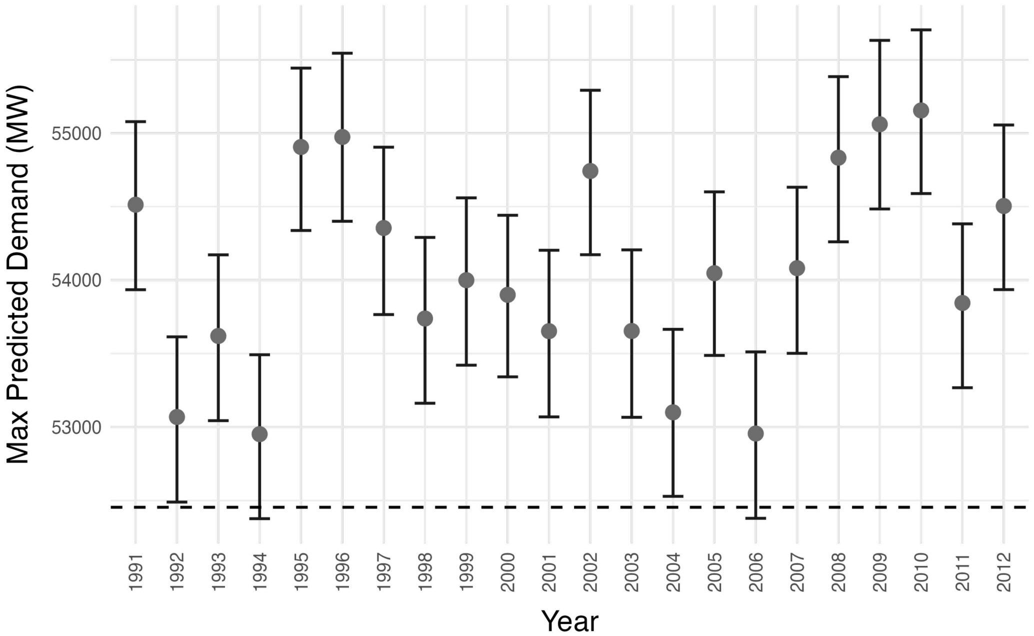

Figure 5.7: Predicted means and 95% prediction intervals for 2013 maximum daily demand under historical weather conditions, based on the linear regression model. The dashed line shows the actual 2013 peak demand.

In nearly all scenarios, the predicted peak demand exceeds the value observed in 2013, supporting the hypothesis that unusually mild weather significantly reduced electricity usage. The highest predicted demands occur under weather inputs from 2010, which was a winter that had extended periods of cold and snowfall [5].

5.3.2 Quantile Regression Model

A parallel analysis was conducted using the 0.95 quantile regression model. In this version, relative wind generation inputs were also substituted, alongside temperature and daylight hours. This model estimates the conditional 0.95 quantile of electricity demand, offering a more conservative view of potential peak conditions.

Figure 5.8 presents the predicted maximum of the 0.95 quantile for the 2013 winter under each historical weather scenario. The 95% confidence intervals, obtained via block bootstrapping, reflect uncertainty in the quantile estimates due to sampling variability.

The dashed black line shows the modelled 0.95 quantile for the specific day on which 2013's actual peak occurred. While it does not represent the maximum demand directly, it provides a more meaningful point of comparison, as the quantile model does not estimate the true peak but rather the conditional upper bound of demand.

For completeness, the simulation was also run using 2013's actual weather conditions. This result is included as the final point on the plot (labelled 2013), enabling a direct comparison with historical weather scenarios under an identical structural setting.

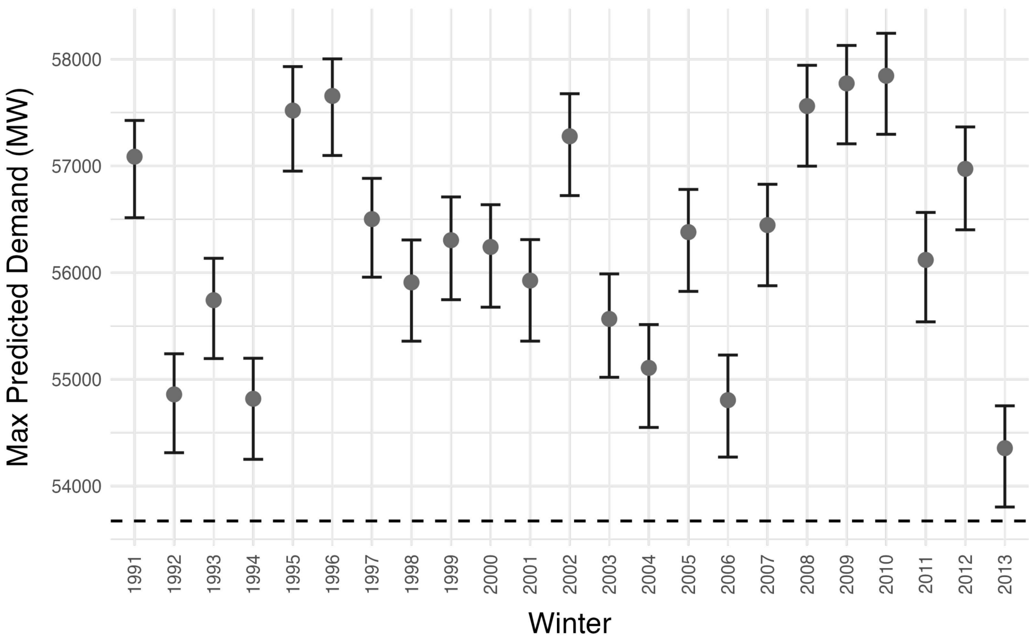

Figure 5.8: Predicted maximum of the conditional 0.95 quantile for 2013 daily demand under historical weather conditions, based on the quantile regression model, with 95% confidence intervals. The dashed line represents the modelled 0.95 quantile on the actual 2013 peak demand day. The final point (2013) reflects the bootstrapped estimate using the full 2013 weather scenario.

The quantile regression model consistently produces higher demand estimates than the linear model, reflecting its focus on upper-tail behaviour. Despite this, both models display a consistent pattern: colder historical winters result in higher predicted peak demands.

For 2013, the predicted 0.95 quantile lies toward the lower end of the historical range, reinforcing the conclusion that mild weather played a significant role in reducing peak demand that winter.

These results highlight the value of quantile regression in capturing plausible high-demand scenarios. By focusing explicitly on tail risk, it offers a useful complement to mean-based approaches and strengthens the foundation for robust system planning.

5.3.3 Conclusion

Taken together, the findings highlight the sensitivity of electricity demand to weather variability and the importance of incorporating this uncertainty into long-term forecasting. In the context of a changing climate, such scenario-based analyses are critical for ensuring that future infrastructure decisions remain resilient under a wide range of conditions.

6 Limitations and Future Considerations

While the models developed in this study perform well under historical conditions, several limitations should be considered when applying them to forecast peak electricity demand over a 5–10 year horizon:

- **Structural changes in electricity use:** The increasing electrification of heating and transport—such as the uptake of heat pumps and electric vehicles—may alter demand patterns in ways not captured by historical data.
- **Changing temperature sensitivity:** The relationship between temperature and electricity demand may evolve over time due to improved insulation, more efficient heating systems, or behavioural adaptation.
- **Weather uncertainty:** Scenario simulations rely on historical weather to approximate future variability. However, climate change may introduce novel extremes or shifts in seasonality, reducing the reliability of this assumption.
- **Policy and market shifts:** Future changes in energy policy, pricing structures, or demand-side response schemes could significantly influence consumption patterns in ways not currently modelled.
- **Growth in decentralised generation:** The increasing adoption of rooftop solar, battery storage, and other behind-the-meter technologies may reduce visible grid demand and reshape daily profiles—particularly during daylight hours.

These limitations highlight the need for caution in long-term planning and support the use of adaptive, regularly updated modelling frameworks.

Additionally, the inclusion of interaction terms between the 'year effect' and variables such as temperature and wind in the models introduces further challenges for future projection. While NESO currently uses ACS peak demand to estimate year effects, there is no established methodology for forecasting how these interactions will evolve [6]. Addressing this gap may require the development of new techniques to estimate how weather sensitivity itself changes over time.

Appendix A

Initial Linear Regression Model Comparison

This appendix outlines the construction of the initial linear regression model, developed iteratively to arrive at a statistically sound and interpretable specification. Explanatory variables were selected based on both theoretical relevance and empirical performance. As noted earlier, observations during the Christmas and New Year period were removed prior to model fitting to avoid distortion from atypical behaviour.

Each model specification was evaluated using the following metrics.

Adjusted R^2: This measures the proportion of variance in the dependent variable explained by the model, adjusted for the number of predictors. It is calculated as:

$$R^2_{\text{adj}} = 1 - \left(\frac{(1 - R^2)(n-1)}{n-p-1}\right)$$

where n is the number of observations and p is the number of predictors. Higher values indicate stronger explanatory power while accounting for model complexity.

AIC and **BIC:** These criteria evaluate model quality by penalising unnecessary complexity relative to fit. They are given by:

$$\text{AIC} = 2p - 2\ln(\hat{L}), \quad \text{BIC} = p\ln(n) - 2\ln(\hat{L}),$$

where $\hat{L}$ is the maximum likelihood of the model, p is the number of parameters, and n is the number of observations. Lower values are preferred indicating a better balance between goodness-of-fit and parsimony.

RMSE (Top 10%): This measures the model's predictive accuracy on high-demand days. It is computed as:

$$\text{RMSE}_{\text{peak}} = \sqrt{\frac{1}{m}\sum_{i \in \text{Top 10\%}} (y_i - \hat{y}_i)^2},$$

where m is the number of observations in the top 10% of demand. Lower values indicate better performance in capturing extreme demand scenarios.

Inclusion of Year Effect

The initial model uses NESO's current temperature variable TE_t. The first modelling decision involved incorporating winter-specific dummy variables, $W_{i,t}$, to capture long-term structural changes in electricity demand. These dummies represent the 'year effect', which adjusts for broad changes such as population growth, technological adoption, and energy efficiency improvements—factors not directly included in the model but which influence demand over time. NESO calculates these year effects for long-term planning purposes [6].

An initial model included only year effects and effective temperature:

$$D_t = \alpha + \sum_{i=1990}^{N} \beta_i W_{i,t} + \delta \cdot TE_t + \epsilon_t$$

However, this assumes a constant effect of temperature across all winters. In reality, the relationship between temperature and demand may evolve due to improved insulation, shifts in heating technology, or behavioural change. To account for this, interaction terms between temperature and each winter were introduced:

$$\textbf{Model A:} \quad D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TE_t W_{i,t}) + \epsilon_t$$

This specification allows the temperature-demand relationship to vary across years. Table A.1 presents the results of these sequential model enhancements.

Model Specification	**Adjusted R^2**	**AIC**	**BIC**	**Top 10% RMSE**
$W_{i,t}$	0.327	61842.51	61988.07	4012.54
$W_{i,t} + TE_t$	0.466	61107.83	61259.45	3573.68
$W_{i,t} + TE_t \times W_{i,t}$	0.470	61104.58	61389.61	3547.22

Table A.1: Model performance with year effects and temperature interactions.

This comparison confirms that introducing year-specific temperature effects provides a modest but consistent improvement across all metrics. Accordingly, Model A is retained as the baseline for subsequent refinement.

Inclusion of Temporal and Seasonal Controls

Electricity demand exhibits systematic weekly patterns, with generally lower consumption at weekends. To account for this, day-of-week indicator variables $DOW_{j,t}$ (Sunday to Friday, with Saturday as the reference category) were added to the model:

$$D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TE_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \epsilon_t$$

To represent the gradual seasonal changes in demand over the winter period, the number of days since 1st November (DSN_t) was introduced. Exploratory analysis, shown in Figure A.1, revealed a clear non-linear relationship between DSN and demand, motivating the inclusion of a squared term:

$$\textbf{Model B:} \quad D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TE_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DSN_t + \delta_2 DSN_t^2 + \epsilon_t$$

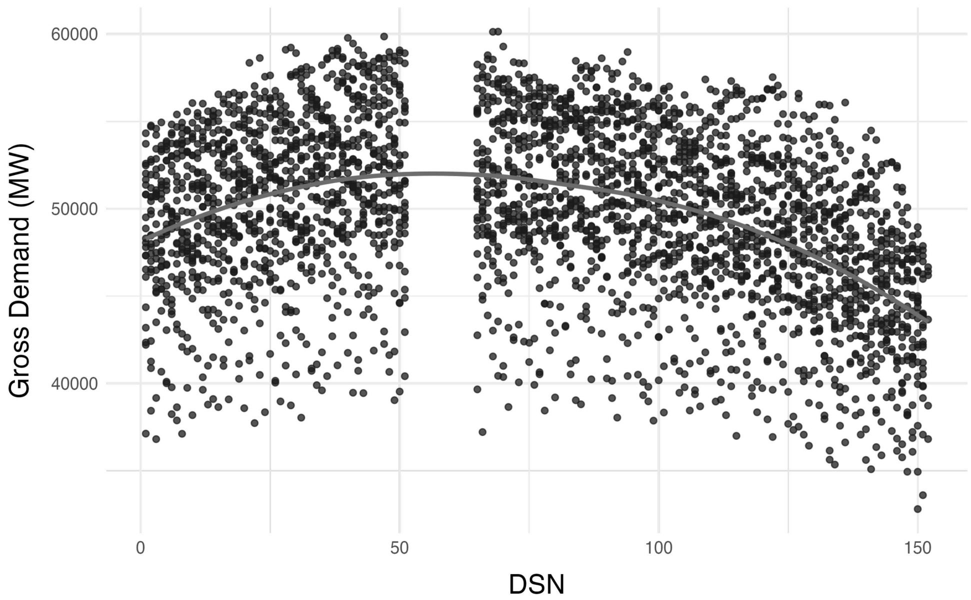

Figure A.1: Gross electricity demand plotted against days since 1st November (DSN).

Model Specification	Adjusted R^2	AIC	BIC	Top 10% RMSE
Model A + $DOW_{j,t}$	0.867	56718.24	57039.67	1776.42
Above + DSN_t	0.939	54232.16	54559.65	1201.29
Above + DSN_t^2	0.968	52208.58	52542.14	873.64

Table A.2: Model performance with temporal and seasonal variables.

Table A.2 presents the effect of adding day-of-week and seasonal terms on model performance. The inclusion of these temporal and seasonal controls leads to a marked improvement in fit, particularly for high-demand days. We therefore proceed with Model B for further refinement.

Inclusion of Relative Solar and Wind Generation

As a final refinement, estimated relative solar (SE_t) and wind (WE_t) generation were considered as additional covariates. These variables act as proxies for prevailing weather conditions—sunlight and wind speed—which may influence electricity demand, particularly for lighting and heating.

Model Specification	Adjusted R^2	AIC	BIC	Top 10% RMSE
Model B + SE_t	0.968	52197.70	52537.32	871.87
Model B + WE_t	0.969	52076.35	52415.97	855.39
Model B + $SE_t + W_{i,t}$	0.969	52063.60	52409.28	873.09
Model B + $SE_t \times W_{i,t}$	0.968	52206.63	52546.25	853.41

Table A.3: Model performance with added solar and wind covariates.

Both variables were tested individually and with interactions to assess whether their effects varied by winter. Table A.3 presents some of the notable results from extending Model B to include these terms.

Including either solar and/or wind generation as additive terms improves model fit marginally. However, the interaction terms with winter dummies yield minimal further improvement, suggesting limited year-on-year variation in the relationship between these variables and demand.

Figure A.2 and Figure A.3 plot the residuals from Model B against relative solar and wind generation, respectively. A slight upward trend is visible in the wind residuals, indicating a positive association between wind generation and unexplained variation in demand. This residual pattern suggests that wind generation may contain predictive information not fully captured by the current Model B supporting its inclusion in an extended model. Conversely, the residuals plotted against solar generation exhibit no discernible structure or trend, implying that solar generation contributes little additional explanatory power beyond what is already accounted for.

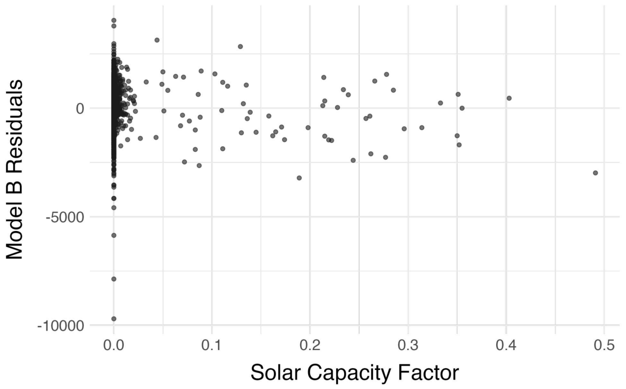

Figure A.2: Residuals of Model B against relative solar generation.

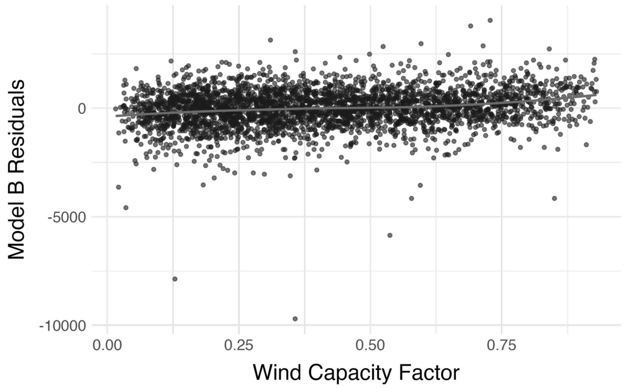

Figure A.3: Residuals of Model B against relative wind generation.

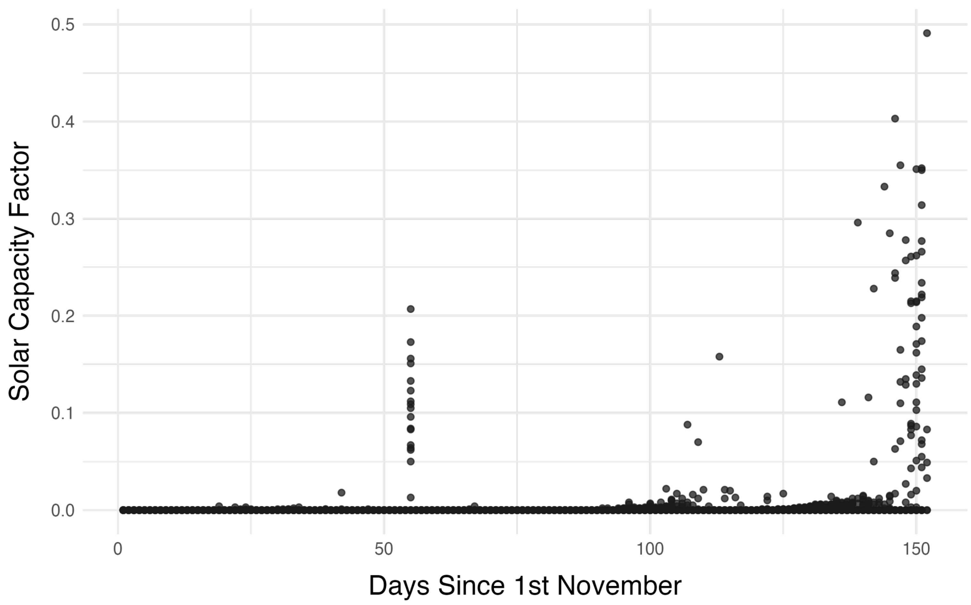

Figure A.4: Relative solar generation plotted against days since 1st November (DSN).

Further concerns arise from the distribution of relative solar generation across the winter period.

As shown in Figure A.4, solar output is negligible for most of the core winter months, with a sharp increase only in late February and March. There is also a spurious spike at day 55, potentially reflecting data quality issues. Since daily peak demand typically occurs after sunset, solar has little direct effect on evening loads. Any potential influence is likely to be small, indirect, or limited to a narrow part of the sample.

Taken together, these findings suggest that while solar generation marginally improves model fit, it does not provide a consistent or interpretable contribution to explaining winter peak demand. To avoid overfitting and maintain a simpler, more interpretable model, it was excluded from the final specification.

Wind generation, by contrast, demonstrates clearer and more consistent predictive value. Its inclusion captures residual variation not accounted for by temperature alone and aligns with the broader objective of improving performance under typical winter conditions relevant to NESO.

The final specification of the initial linear regression model is therefore given by:

$$D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TE_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DSN_t + \delta_2 DSN_t^2 + \theta_1 WE_t + \epsilon_t$$

Appendix B

Temperature Variable Selection

This appendix evaluates alternative temperature formulations for inclusion in the linear regression model. The objective was to identify a variable that most effectively captures the relationship between temperature and daily peak electricity demand—particularly during periods of cold weather.

All model specifications build on the baseline structure developed in Appendix A, with only the temperature variable substituted in each case.

NESO's effective temperature variable, TE_t, provides a theoretically grounded measure of thermal persistence by incorporating lagged effects through recursive averaging. However, alternative smoothing techniques may offer improved empirical performance by more flexibly capturing recent and cumulative cold exposure.

Formulations Considered

All alternative temperature formulations were initially constructed using NESO's short-term average temperature, TO_t, as the base variable. The following smoothing methods were tested:

- **Multi-day moving averages:** Simple rolling averages over ranges between the previous 3 and 7 days, designed to capturing the persistence of recent cold spells.
- **Centred moving averages:** Symmetric rolling means centred on day t, offering smoothed temperature trajectories and account for short-term weather forecasts.
- **Temperature change:** The first difference of temperature, $TO_t - TO_{t-1}$, capturing abrupt shifts in weather conditions.
- **Weighted averages:** Exponentially decaying weights applied to the current and two preceding days, assigning greater influence to recent temperatures while preserving a memory of past conditions. The best-performing scheme was:

$$TW_t = 0.5 \cdot TO_t + 0.3 \cdot TO_{t-1} + 0.2 \cdot TO_{t-2}$$

Evaluation Criteria

Each temperature variable replaced TE_t in the full linear regression model. Once again the resulting models were evaluated using adjusted R^2, AIC, BIC, and the Root Mean Squared Error (RMSE) on the top 10% of demand observations.

Results

Table B.1 summarises the performance of each formulation under the four evaluation metrics.

Temperature Variable	Adjusted R^2	AIC	BIC	Top 10% RMSE
NESO's TE_t	0.969	52076.35	52415.97	855.39
3-day Moving Average	0.959	52981.32	53320.94	986.19
7-day Moving Average	0.956	53057.61	53397.23	998.09
3-day Centred Average	0.965	52499.77	52839.39	914.28
Temperature Change	0.919	55124.91	55464.53	1381.45
Weighted Average (0.5, 0.3, 0.2)	0.969	52040.47	52380.09	849.32

Table B.1: Comparison of alternative temperature variables.

The weighted temperature average, TW_t, achieved the best overall performance across all evaluation criteria. It matched the explanatory power of NESO's original TE_t but delivered improved fit and reduced error on peak-demand days—particularly as measured by RMSE in the upper tail.

This suggests that TW_t more effectively captures both the immediate influence of recent temperatures and the lingering effects of cold spells. Accordingly, it was selected as the preferred temperature smoothing method used in the final linear and quantile regression models.

Appendix C

Temperature Averaging Window Comparison

This appendix evaluates alternative intraday temperature averaging windows to identify the most effective short-term temperature measure for predicting daily peak electricity demand. The goal was to determine the time range around 18:00 that best captures the thermal conditions influencing demand.

Formulations Considered

All temperature variables were derived from hourly, population-weighted air temperature data, sourced from the MERRA-1 dataset. The candidate variables represent variations on NESO's baseline definition of TO, which averages hourly temperatures from 15:00 to 18:00. The following represent some of the extensions that were tested:

- **NESO Baseline:** $TO = T_{\text{15-18}}$
- **Backward Extensions:** $T_{\text{14-18}}$, $T_{\text{13-18}}$, $T_{\text{12-18}}$, $T_{\text{11-18}}$
- **Forward Extensions:** $T_{\text{15-19}}$, $T_{\text{15-20}}$

These alternatives were selected to test whether including a longer historical temperature window prior to the 18:00 peak, or a brief extension beyond it, would improve model performance.

Evaluation Criteria

Each candidate variable was used as the base input in the construction of the smoothed temperature variable TW_t, following the optimal weighting scheme identified in Appendix B. Each version of TW_t was then incorporated into the full linear regression model, and its performance evaluated using adjusted R^2, AIC, BIC, and the Root Mean Squared Error (RMSE) on the top 10% of demand observations. In addition, the Pearson correlation between each base temperature variable and electricity demand was calculated to assess explanatory strength in isolation.

Backward and forward extensions were incrementally tested, and the process was halted once further extensions no longer yielded improvements across these metrics.

Results

Table C.1 summarises the results for each intraday temperature average.

Temperature Variable	**Correlation**	**Adjusted** R^2	**AIC**	**BIC**
$T_{11\text{-}18}$	-0.2646	0.969	52031.08	52370.70
$T_{12\text{-}18}$	-0.2639	0.969	52050.78	52390.40
$T_{13\text{-}18}$	-0.2627	0.969	52071.46	52411.08
$T_{14\text{-}18}$	-0.2608	0.969	52108.74	52448.36
$T_{15\text{-}18}$ (TO)	-0.2572	0.968	52172.08	52511.70
$T_{15\text{-}19}$	-0.2461	0.968	52224.80	52564.42
$T_{15\text{-}20}$	-0.2362	0.967	52273.09	52612.71

Table C.1: Comparison of alternative weighted intraday temperature averages.

The weighted average temperature between 11:00 and 18:00, denoted $T_{11\text{-}18}$, achieved the highest correlation with demand and consistently outperformed other formulations across all statistical criteria. It was therefore adopted as the optimal intraday measure of short-term temperature exposure.

The smoothed temperature variable TW_t was subsequently redefined as the weighted average of $T_{11\text{-}18}$ over three consecutive days:

$$TW_t = 0.5 \cdot T_{11\text{-}18,t} + 0.3 \cdot T_{11\text{-}18,t-1} + 0.2 \cdot T_{11\text{-}18,t-2}$$

This formulation was adopted as the final temperature input for all subsequent modelling and scenario analysis.

Appendix D

Daylight Hours Model Comparison

This appendix compares two model specifications for representing gradual seasonal variation in electricity demand across the winter period. The first uses the original seasonal trend variable—days since 1st November (DSN_t)—while the second substitutes this with calculated daylight hours (DH_t) and its square. The aim is to evaluate whether daylight hours offer a more interpretable and effective representation of seasonal patterns.

Both models build on the final linear regression specification developed in Appendix A, incorporating the new temperature variable TW_t, winter fixed effects, temperature–year interactions, day-of-week indicators, and wind generation. The only difference lies in the formulation of the seasonal trend component:

$$\textbf{Model C:}\quad D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TW_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DSN_t + \delta_2 DSN_t^2 + \theta_1 WE_t + \epsilon_t$$

$$\textbf{Model D:}\quad D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TW_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \eta_1 DH_t + \eta_2 DH_t^2 + \theta_1 WE_t + \epsilon_t$$

Daylight hours were calculated for the population-weighted centre of England using 2014 census data, as outlined in the main report. This location was selected to approximate the spatial distribution of electricity demand across the grid, providing a representative measure of day length for the NESO network.

Model Evaluation

Model performance was assessed using adjusted R^2, AIC, BIC, and the Root Mean Squared Error (RMSE) on the top 10% of demand observations. The results for the two models is summarised in Table D.1.

Model	**Adjusted R^2**	**AIC**	**BIC**	**Top 10% RMSE**
Model C (with DSN_t)	0.970	52040.47	52380.09	850.5777
Model D (with DH_t)	0.970	52031.08	52370.70	849.3226

Table D.1: Comparison of seasonal trend specifications.

Replacing the proxy variable DSN_t with daylight hours yielded modest but consistent improvements across all evaluation metrics. Although the gains are relatively small, the use of physically grounded daylight hours offers clearer interpretability and is directly tied to a known driver of electricity usage—natural light availability.

Given its empirical performance and conceptual clarity, the daylight hours specification Model D was selected for the final model used in forecasting and scenario analysis. Therefore the final linear regression model is

$$D_t = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TW_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \eta_1 DH_t + \eta_2 DH_t^2 + \theta_1 WE_t + \epsilon_t$$

Appendix E

Quantile Selection Process

This appendix outlines the selection process for determining the appropriate quantile to model high electricity demand using quantile regression. The objective was to balance the operational relevance of extreme demand scenarios with the statistical stability required for reliable forecasting.

Quantile regression models were fitted at $q = 0.90$, 0.95, and 0.98. Higher quantiles provide greater focus on rare, high-demand events, but rely on fewer observations, which can lead to unstable parameter estimates and poor calibration.

Evaluation Criterion

Each model was evaluated based on its calibration using the observed exceedance rate. For a model targeting quantile q, the proportion of observed values that exceed the predicted quantile should approximate $1-q$. For example, a well-calibrated 0.95 quantile model should be exceeded by approximately 5% of the data.

This approach provides a consistent and interpretable basis for comparing model performance across quantiles.

Results

The dataset comprises 3,180 daily observations. Table E.1 compares the expected and actual number of exceedances for each candidate quantile.

Quantile q	**Expected Exceedances**	**Observed Exceedances**	**Exceedance Rate**
0.90	318	289	0.091
0.95	159	185	0.058
0.98	64	36	0.011

Table E.1: Calibration performance across candidate quantiles.

The 0.90 quantile model is well-calibrated, with an observed exceedance rate close to the expected 10%. However, it may not sufficiently capture the extreme demand scenarios most relevant to NESO's operational planning. At the other end, the 0.98 quantile model substantially underestimates the expected exceedance rate, with only 36 exceedances observed versus an expected

64. This under-calibration reflects the limited number of data points in the upper tail, resulting in greater estimation uncertainty.

The 0.95 quantile model strikes an appropriate balance. It records 185 exceedances versus the expected 159, corresponding to an exceedance rate of 5.8%. This modest over-calibration is acceptable and preferable to the instability observed at higher quantiles. Moreover, the sample size in this region remains sufficient for robust model estimation.

Therefore, based on calibration performance and statistical stability, the 0.95 quantile ($q = 0.95$) was selected as the preferred level for quantile regression modelling. It provides focused insight into high-demand conditions while maintaining the reliability required for forecasting and stress-testing in NESO's long-term planning framework.

Appendix F

Quantile Regression Model Metrics

This appendix outlines the evaluation metrics used to assess and compare quantile regression models. Unlike ordinary least squares, quantile regression does not minimise squared error, so traditional fit measures such as R^2 are not appropriate. Instead, performance is assessed using metrics based on the Pinball Loss function, which is specifically designed for evaluating predictive accuracy at a chosen quantile.

Generalised Cross-Validation (GCV)

Generalised Cross-Validation (GCV) provides an estimate of out-of-sample prediction error while accounting for model complexity. It is a computationally efficient alternative to k-fold cross-validation and is particularly well-suited to quantile regression, where residual variance is not constant [4].

$$GCV = \frac{L_q}{(1 - p/n)^2}$$

where:

- L_q is the total Pinball Loss for the model at quantile q,
- p is the number of estimated model parameters,
- n is the number of observations.

Lower GCV values indicate better generalisation performance, with an implicit penalty for overfitting as model complexity increases.

Quantile Information Criteria: AIC$_q$ and BIC$_q$

To support model selection and comparison, quantile-adapted versions of Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are also used. These replace the standard log-likelihood term with the model's Pinball Loss [4]:

$$AIC_q = 2L_q + 2p, \qquad BIC_q = 2L_q + p\log(n)$$

where:

- L_q is the total Pinball Loss,

- p is the number of model parameters,
- n is the sample size.

Both criteria penalise complexity to discourage overfitting. AIC_q applies a linear penalty, while BIC_q imposes a stronger, logarithmic penalty. Lower values indicate more parsimonious models with better fit to the data.

Summary

The combined use of Pinball Loss, GCV, AIC_q, and BIC_q provides a statistically rigorous framework for evaluating quantile regression models. These metrics were used throughout the model selection process to ensure that the final specification achieves a robust balance between predictive accuracy, simplicity, and reliability. Full results are presented in Appendix G.

Appendix G

Quantile Regression Model Selection

This appendix outlines the model selection process for the 0.95 quantile regression model used to estimate peak electricity demand. The objective was to identify a specification that balances predictive accuracy, statistical robustness, and interpretability—focusing on high-demand scenarios of particular relevance to NESO.

Evaluation Criteria

Quantile regression estimates the conditional quantile of the response variable rather than its conditional mean. As such, traditional OLS fit metrics are not appropriate. Instead, the following criteria—detailed in Appendix F—were used:

- **Pinball Loss:** The primary loss function in quantile regression, measuring prediction accuracy at the target quantile.
- **Generalised Cross-Validation (GCV):** A computationally efficient estimate of out-of-sample error that adjusts for model complexity.
- **Quantile AIC and BIC (AIC_q, BIC_q):** Penalised information criteria adapted for quantile regression by substituting in Pinball Loss.

Model Specifications Considered

All candidate models build on the linear regression structure but are refitted to estimate the conditional 0.95 quantile. Each uses the weighted temperature variable TW_t, previously identified as the most effective predictor of heating-related demand.

The models below differ by their treatment of daylight hours and the inclusion of weather covariates:

Model A: $Q_{0.95}(D_t) = \alpha + \sum_{i=1990}^{N} (\beta_i W_{i,t} + \gamma_i TW_t W_{i,t}) + \sum_{j=1}^{6} \omega_j DOW_{j,t} + \delta_1 DH_t + \delta_2 DH_t^2 + \epsilon_{0.95,t}$

Model B: $Q_{0.95}(D_t) =$ Model A, with DH_t^2 removed

Model C: $Q_{0.95}(D_t) =$ Model A, with DH_t removed

$$\textbf{Model D:} \quad Q_{0.95}(D_t) = \text{Model A} + \theta_2 WE_t$$

$$\textbf{Model E:} \quad Q_{0.95}(D_t) = \text{Model A} + \sum_{i=1990}^{N} \theta_i WE_t W_{i,t}$$

Additionally, the various temperature variables considered earlier were re-evaluated for these models, but TW_t continued to show the best performance.

Results

Table G.1 reports the performance of each model across the evaluation criteria.

Model	Pinball Loss	GCV	AIC$_q$	BIC$_q$
Model A	74.40	76.99	13812.12	14139.61
Model B	75.00	77.57	13835.66	14157.09
Model C	74.42	76.96	13810.83	14132.25
Model D	72.04	74.60	13711.50	14045.05
Model E	70.73	74.29	13697.27	14164.24

Table G.1: Comparison of 0.95 quantile regression model specifications.

Model E, which includes year-specific interactions with relative wind generation, achieves the best overall performance—recording the lowest Pinball Loss, GCV, and AIC$_q$. While its BIC$_q$ is marginally higher due to the increased complexity, the improvement in predictive accuracy justifies the added terms.

These results support the inclusion of wind-related variables when forecasting high-demand events. Accordingly, Model E was selected as the final specification for scenario analysis and peak demand forecasting using the 0.95 quantile regression framework.