Boonsuriyatham, Panthakan (2025) *Changepoint detection for net electricity demand modelling in Great Britain.* MPhil(R) thesis.

https://theses.gla.ac.uk/85614/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Changepoint Detection for Net Electricity Demand Modelling in Great Britain

Panthakan Boonsuriyatham

Submitted in fulfilment of the requirements for the Degree of

Master of Philosophy

School of Mathematics and Statistics

College of Science and Engineering

University of Glasgow

November 2024

Abstract

This thesis investigates a changepoint detection methodology applied to net electricity demand modelling in Great Britain. With the increasing integration of renewable energy sources and the growing complexity of electricity demand patterns, accurately identifying abrupt changes, or "changepoints," in demand data has become essential for reliable grid management. Initial analyses of national electricity demand data were conducted to explore the underlying features and factors influencing consumption patterns. By integrating Generalised Additive Models (GAMs) within a changepoint detection framework, this research introduces a flexible approach capable of capturing non-linear trends and seasonal variations without requiring extensive manual adjustments, allowing the model to adapt to diverse fluctuations in demand.

Traditional changepoint algorithms were evaluated, but the unique complexities of electricity demand data led to the development of a novel changepoint detection algorithm tailored to these demands. Simulation studies tested the proposed methodology under various mean shifts and noise levels, offering insights into how these parameters impact changepoint detection accuracy. The novel algorithm was subsequently applied to regional Grid Supply Point (GSP) electricity demand data, demonstrating how different demand patterns across geographic areas influence changepoint locations.

The findings underscore both the strengths and limitations of integrating model-based cost functions and GAMs within changepoint detection, particularly in managing daily and seasonal cycles, addressing computational constraints, and scaling across large datasets. This approach enhances the accuracy and flexibility of electricity demand modelling by effectively identifying abrupt changes, enabling a more robust response to demand variability.

Contents

Abstract **i**

Contents **ii**

List of Figures **v**

List of Tables **viii**

1 Introduction **1**

1.1 Electricity Industry in the UK 1

1.2 Electricity Demand 2

1.3 Objectives of Research 3

1.4 Literature Review 4

2 Modelling National Demand with Generalised Additive Model **11**

2.1 Introduction 11

2.2 Exploratory Analysis 13

2.2.1 Total Demand 13

2.2.2 Weather Features 16

2.2.3 Bank Holidays 22

2.3 Methodology - Generalised Additive Model 27

2.4 Evaluation Metrics 31

2.5 Study Design 32

2.6 Results 35

2.7 Discussion 43

2.8 Conclusion 46

3 Integrating Changepoint Detection Algorithms with Predictive Models **48**

3.1 Introduction 48

3.1.1 Notation 48

3.1.2 Cost Function 49

3.1.3 Penalty Term 49

3.2 Changepoint Detection Algorithms 50

3.2.1 At most one change (AMOC) 50

3.2.2 Binary Segmentation 51

3.2.3 Optimal Partitioning (OP) 52

3.2.4 Pruned Exact Linear Time (PELT) 53

3.3 Evaluation Metrics for Changepoint Detection Algorithms 55

3.4 Changepoint Detection Algorithm within a Predictive Model 57

3.5 Summary 61

4 A Simulation Study of Proposed Method **64**

4.1 Introduction 64

4.2 Simulation Study Design 66

4.3 Study 1: Model without smooth terms - Single changepoint 68

4.4 Study 2: Model without smooth terms - Multiple changepoints 73

4.5 Study 3: Model with smooth terms - Single changepoint 78

4.6 Study 4: Model with smooth terms - Multiple changepoints 80

4.7 Conclusion 84

5 Modelling Regional Demand with Changepoints **86**

5.1 Introduction 86

5.2 Exploratory Analysis 91

5.3 Variables and Models 92

5.4 Study Design 93

5.5 Results and evaluation 94

5.6 Discussion and Conclusion 109

5.6.1 Critical Analysis of Detected Changepoints 109

5.6.2 Methodological Limitations and Considerations 113
5.6.3 Implications and Future Directions 115

6 Conclusion 116
6.1 Key Findings 116
6.2 Strengths and Limitations 117
6.3 Implications and Future Work 120

A Approaches to Calculate a Cost Function in Simulation Study 122

References 127

List of Figures

2.1 Average half-hourly national demand 12

2.2 Embedded wind generation profile 13

2.3 Embedded solar generation profile 14

2.4 National electricity demand profile 14

2.5 National electricity demand profile in June and November 2014 . . . 15

2.6 Locations of hourly estimated weather features and population 18

2.7 2 metre temperature above the surface across Great Britain 18

2.8 2 metre temperature above the surface in certain cities 19

2.9 Different transformation of gridded population 21

2.10 National electricity demand profile during New Year season 24

2.11 National electricity demand profile during different bank holidays . . 24

2.12 National electricity demand profile during the New Year season across multiple years 26

2.13 Predictive errors from fitting the best GAM model to national electricity demand data 37

2.14 QQ-plot for GAM models from 2 years of training data 37

2.15 QQ-plot for GAM models from 2 years of training data after excluding Christmas and New Year seasons 40

2.16 ACF plot from a two-year training model (excluding Christmas and New Year) 40

2.17 Histogram of residuals from a two-year training model (excluding Christmas and New Year) 41

2.18 Scatter plot of predicted values against residuals from the model trained on two years of data, excluding the Christmas and New Year periods. 41

2.19 Scatter plot of predicted versus actual demand from the model trained on two years of data, excluding the Christmas and New Year periods. 42
2.20 Time series of residuals from the model trained on two years of data, excluding the Christmas and New Year periods. 42

4.1 Simulated dataset for single-changepoint simulations 65
4.2 Simulated dataset for multiple-changepoint simulations 66
4.3 Study 1: Frequency of simulations with no changepoint returned . . . 70
4.4 Study 1: Bias of changepoint location 71
4.5 Study 1: Mean Absolute Error of changepoint location 71
4.6 Study 1: Root Mean Squared Error of changepoint location 71
4.7 Study 1: Median Absolute Error of changepoint location 72
4.8 Study 1: Root Mean Squared Error of residuals 72
4.9 Study 2: Frequency of simulations with no changepoint detected . . . 75
4.10 Study 2: Covering metric of changepoint locations 76
4.11 Study 2: F1-score of changepoint locations 77
4.12 Study 3: Frequency of simulations with no changepoint detected . . . 79
4.13 Study 3: Root Mean Squared Error of changepoint location 80
4.14 Study 4: Frequency of simulations with no changepoint detected . . . 82
4.15 Study 4: Number of changepoints returned by the algorithm 83
4.16 Study 4: Covering metric of changepoint locations 83
4.17 Study 4: F1-score of changepoint locations 84

5.1 Geographical map of Grid Supply Point groups and IDs in Great Britain 88
5.2 Regional Demand Profile of Penn 91
5.3 Changepoint locations of the region OLDB_1 95
5.4 Reduction in BIC of the region OLDB_1 compared to the naive model 96
5.5 Changepoint locations of the region OLDB_1 using setup 8 97
5.6 Effect of day of the week on region OLDB_1 98
5.7 Effect of 2 metre temperature on region OLDB_1 98
5.8 Effect of 10 metre wind speed on region OLDB_1 98
5.9 Effect of surface solar radiation downwards on region OLDB_1 99

5.10 Changepoint locations of the region PENN_1 100
5.11 Reduction in BIC of the region PENN_1 compared to the naive model 101
5.12 Changepoint locations of the region PENN_1 using setup 6 102
5.13 Effect of day of the week on region PENN_1 103
5.14 Effect of 2 metre temperature on region PENN_1 103
5.15 Effect of 10 metre wind speed on region PENN_1 103
5.16 Effect of surface solar radiation downwards on region PENN_1 104
5.17 Changepoint locations of the region PENN_1 using setup 2 105
5.18 Changepoint locations of the region BOAG_P 106
5.19 Reduction in BIC of the region BOAG_P compared to the naive model 106
5.20 Changepoint locations of the region BOAG_P using setup 6 107
5.21 Effect of day of the week on region BOAG_P 108
5.22 Effect of 2 metre temperature on region BOAG_P 108
5.23 Effect of 10 metre wind speed on region BOAG_P 108
5.24 Effect of surface solar radiation downwards on region BOAG_P 109

List of Tables

2.1 Weather features 17
2.2 Bank holidays in Great Britain 23
2.3 Results from fitting GAM models to national electricity demand data 36
2.4 The predictive results from building GAM models 40

5.1 The GSP Groups in Great Britain and their corresponding Distribution Network Operators 89
5.2 Setup ID reference for configurable parameters within the algorithm . 93
5.3 Comparison of model parameters across two segments for two different setups 102

A.1 The evaluation metrics from the single-changepoint simulation. 125
A.2 The evaluation metrics from the multiple-changepoint simulation. . . 125

Chapter 1

Introduction

1.1 Electricity Industry in the UK

In 2021, investment in the UK's energy industries reached £16.5 billion, with the electricity sector accounting for 69% of this total [1]. The consistent rise in investment reflects the importance of the electricity sector in supporting both economic activity and the transition to cleaner energy sources. The supply of electricity within the UK can be described in three key stages: generation, transmission, and distribution.

Electricity generation involves converting primary energy sources into usable electricity. As of 2021, the primary contributors to UK electricity generation include gas (39.8%), wind and solar (24.9%), and nuclear power (14.9%). Each generation source employs distinct methods: gas is combusted to generate heat, wind turbines convert kinetic energy from wind, solar panels capture sunlight to generate electrical energy, and nuclear power relies on heat from nuclear fission to produce steam, which drives turbines to create electricity. Once electricity is generated, it moves from the generation stage to transmission, where it travels over long distances at high voltages to reduce power loss, before finally entering the distribution phase, where the voltage is gradually lowered for safe, local delivery to end-users.

Electricity is transmitted through a high-voltage network that links large power stations to regional distribution networks, ultimately delivering power to consumers.

[1] https://www.gov.uk/government/statistics/uk-energy-in-brief-2022, accessed on 14 June 2023

High-voltage transmission minimises power losses across long distances. England and Wales predominantly operate transmission networks at 400 and 275 kilovolts (kV), whereas Scotland's system runs at 132 kV, reflecting the unique infrastructure of different Transmission Owners (TOs). Electricity enters this network at Grid Entry Points (GEPs) and exits at Grid Supply Points (GSPs), where it is transferred to the distribution network.

The distribution phase commences as electricity leaves the transmission network through GSPs, where voltage is stepped down to extra-high voltage (EHV) levels, typically around 132 kV. Each GSP supplies several Bulk Supply Points (BSPs) that reduce voltage further to 33 kV. From here, BSPs provide electricity to primary substations, which lower the voltage to 11 kV. Finally, secondary substations reduce voltage to 400/230 volts for safe consumer use.

In the UK, 14 licensed Distribution Network Operators (DNOs) manage specific regional distribution areas, known as DNO License Areas or GSP Groups. These DNOs ensure efficient, reliable power distribution to end users across Britain, with each network adapting to regional needs and demand patterns.

1.2 Electricity Demand

Maintaining a stable electricity supply requires a precise balance between supply and demand, making accurate forecasting essential. Errors in forecasting can lead to either overproduction, which results in wastage since electricity cannot be stored cost-effectively, or underproduction, which risks shortages and blackouts. This challenge is compounded by daily and seasonal variations in demand. Typically, demand peaks during weekdays and is lower on weekends, with notable seasonal differences between summer and winter patterns.

Over the last decade, the UK's rising dependence on renewable energy sources, such as wind and solar, has added further complexity to demand forecasting. Renewables now account for around 43% of the UK's electricity generation [1], introducing variability that is contingent on weather conditions. To address fluctuations in re-

[1] https://www.carbonbrief.org/analysis-uk-electricity-from-fossil-fuels-drops-to-lowest-level-since-1957/, accessed on 26 October 2024

newable output, the demand for energy storage systems has grown (Mitali et al., 2022). These systems help store surplus energy when production is high and release it during periods of low production. However, development in efficient, large-scale storage has been gradual, highlighting an ongoing need for advancements in this area.

Smart grid technology offers potential solutions for managing variable energy production. These advanced systems use sensors and data-driven optimisation to enhance the flexibility and efficiency of electricity distribution (Ardito et al., 2013). However, the effectiveness of smart grids depends on the availability of real-time data on supply and demand, making the forecasting process even more critical. Additionally, the effects of climate change, such as heatwaves and storms, increasingly disrupt both energy production and demand, presenting further challenges to achieving accurate forecasts.

1.3 Objectives of Research

Given the increasing complexity in balancing electricity supply and demand, this research aims to deepen understanding at both national and regional levels. Electricity demand, as an aggregate measure, represents the total power consumed within a specific area, encompassing residential, commercial, and industrial needs. While this national-scale demand figure is valuable, it often obscures substantial regional differences. Population density, industrial activity, climate, and the extent of local renewable generation can all drive these differences. As a result, understanding regional, or local, demand patterns is crucial for accurate assessments of energy flow across the network, especially as the grid transitions to a decentralised model with variable generation sources.

Changepoints denote abrupt shifts in properties like mean or variance in time series data. Detecting changepoints in electricity demand data is essential for identifying and responding to shifts in consumption behaviour, often driven by societal changes, policy developments, or environmental factors. Changepoints can reveal impactful events, such as the effects of lockdowns, shifts in energy policy, or the influence of newly operational renewable facilities. Without accurately detecting these shifts, demand forecasting models may rely on outdated historical patterns, impairing the stability and efficiency of electricity supply. Identifying changepoints that signify shifts in demand patterns allows grid operators to anticipate and respond to demand fluctuations, optimising energy allocation, enhancing reliability, and supporting sustainable energy goals.

Creating an effective changepoint detection algorithm tailored to the complexities of local demand data is critical due to the unique, temporal variations in electricity use. Demand patterns are influenced not only by seasonal cycles but also by renewable energy integration, consumer habits, and industrial cycles. A robust changepoint detection algorithm helps energy providers proactively adapt to these shifts, enhancing grid reliability and stability. Such an algorithm also supports grid management by highlighting emerging trends, which can inform strategic investments in infrastructure. By accurately pinpointing changepoints, this approach strengthens demand forecasting, enhances resilience against fluctuations, and contributes to a more sustainable energy sector.

In light of these challenges and opportunities, the primary objectives of this research are as follows:

1. Develop a comprehensive understanding of national electricity demand and the impact of various influencing factors, as explored in Chapter 2.
2. Design an algorithm to detect changepoints that indicate shifts in demand patterns. This algorithm is developed in Chapter 3 and evaluated in Chapter 4.
3. Apply the algorithm to regional demand data to assess its effectiveness across diverse local conditions, as demonstrated in Chapter 5.

1.4 Literature Review

Electricity demand forecasting has seen significant advancements in the past decade. While both machine learning —such as neural networks and support vector machines— and statistical approaches, including multiple linear regression (MLR) and quantile regression, have been used for this task (Zhang et al., 2020), this thesis focuses on statistical methods due to their interpretability and transparency. Unlike machine learning models, which often function as "black boxes" with limited insight into underlying relationships, statistical models allow for clearer understanding of the effects of individual predictors. This characteristic is particularly important when modelling electricity demand, where interpretability aids not only in accurate forecasting but also in understanding structural changes in consumption behaviour.

MLR has been a longstanding technique in electricity demand forecasting and was used effectively by winning teams in the Global Energy Forecasting Competition (GEFCom) 2012 (Hong et al., 2014). For example, Tao's benchmark model, which employs predictors such as time, date, and temperature, has become a foundational approach in the literature (Hong, 2010). At the time, point forecasts like MLR—which offer a single anticipated value for future demand—were widely used in operational decision-making. However, more recent developments have shifted the focus toward probabilistic forecasting, which provides a range of possible outcomes with associated probabilities. This approach offers a more comprehensive understanding of forecast uncertainty, particularly important for robust grid management. Reflecting this shift, GEFCom 2014 introduced probabilistic forecasting tracks for electric load, wind power, solar power, and electricity prices (Hong et al., 2016). The winning team, Tololo, employed quantile regression—a form of probabilistic forecasting that estimates conditional quantiles rather than just the mean—alongside generalised additive models (GAMs) (Wood, 2017), highlighting the growing relevance of flexible, distributional modelling techniques (Gaillard et al., 2016). GAMs extend MLR by capturing nonlinear relationships between predictors and demand, making them particularly well-suited for complex forecasting tasks. Research by Hong and Fan (2016) documents the exponential growth in probabilistic demand forecasting, with studies typically categorised into short-term and long-term forecasting. Short-term forecasting, several hours up to a few days ahead, is driven by operational requirements, while long-term forecasting, weeks to years ahead, supports strategic planning and infrastructure investment.

Demand forecasting can be categorised into offline and online approaches. Offline forecasting analyses an entire dataset retrospectively to make predictions. As examples of offline forecasting, Luís et al. (2020) presents a study on offline electricity demand forecasting in Portugal using an ensemble of machine learning methods, while Jiménez-Herrera et al. (2020) proposes a new offline forecasting algorithm, "StreamNSP," based on clustering and nearest neighbours. In contrast, online forecasting operates in real-time, processing each data point as it becomes available to efficiently predict the next time step. Laouafi et al. (2017) applies online forecasting to very short-term electricity demand (minutes to a few hours ahead) by integrating various forecasting models, including classical techniques and artificial intelligence methods, using real-time load data from the French power system. Taylor (2003) also employs an online forecasting approach for short-term electricity demand, utilising the double seasonal Holt-Winters exponential smoothing method and the multiplicative double seasonal ARIMA model. While offline forecasting often achieves higher accuracy due to access to the full dataset, it lacks the flexibility to respond to real-time fluctuations. In contrast, online forecasting continuously updates predictions as new data becomes available, making it better suited for dynamic electricity demand management. Although online methods may sacrifice some accuracy due to limited historical context and potential noise in real-time data, they provide critical advantages in adaptability, responsiveness, and the ability to incorporate sudden changes in consumption patterns, weather conditions, or grid disturbances.

Weather conditions, particularly temperature and wind speed, significantly impact electricity demand, as demonstrated by models in GEFCom 2012 and 2014 (Hong et al., 2014), (Hong et al., 2016). Recent studies have also examined recency effects, where recent weather conditions disproportionately affect demand (Wang et al., 2016).

Holidays present another key variable in demand forecasting. Although studies have investigated the holiday effect on demand in regions like Germany (Ziel, 2018), similar research for the UK remains limited. With the UK consisting of distinct regions (England, Scotland, Wales, and Northern Ireland) with varying holidays, careful consideration of these regional factors is crucial.

Structural breaks—abrupt changes in the data-generating process—can severely impact forecasting accuracy. These breaks may arise from shifts in energy usage patterns, policy interventions, or technological disruptions. If not properly addressed, structural breaks lead to biased parameter estimates and model misspecification (Bai and Perron, 1998; Chow, 1960). In electricity demand forecasting, they often manifest as sudden changes in trend, variance, or seasonal behaviour.

Structural breaks in time series data signify points where underlying properties, such as mean, variance, or other parameters, shift due to external factors like policy changes, economic conditions, or technological advancements. Detecting these breaks is essential across various fields, as unaccounted structural changes can lead to misleading forecasts and analyses. For instance, Zeileis et al. (2002) highlighted the importance of identifying structural changes within linear regression models, introducing robust testing methods like the CUSUM and MOSUM tests to address such shifts in time series data. Similarly, Aue and Horváth (2013) adapted cumulative sum (CUSUM) statistics to capture breaks in mean, variance, and correlation structures, which are particularly challenging in data with serial dependence. In a review of structural break methodologies, Casini and Perron (2018) discussed advances in handling models with temporal dependencies, focusing on challenges in distinguishing structural breaks from unit root behaviour. Furthermore, Pesaran et al. (2006) proposed a Bayesian approach for forecasting time series with discrete structural breaks, using hierarchical models to account for the size and duration of these shifts, as demonstrated in their application to U.S. Treasury bill rates. Lastly, Perron (2005) reviewed methods for estimating and testing structural breaks, emphasising the importance of accounting for trend changes and long memory in accurately detecting these breaks, especially in economic and financial data. Collectively, these works underscore the necessity of structural break detection across disciplines and illustrate the complexity of break patterns in time series data, a need similarly relevant in fields like electricity demand forecasting.

Structural break techniques typically focus on detecting abrupt changes in the parameters of linear models, which makes them effective for certain applications but limited in scope. They may misinterpret regular seasonal patterns as structural shifts, resulting in excessive false positives. In contrast, changepoint detection methods are often likelihood-based and can detect broader shifts in the properties of a stochastic process, such as changes in mean or distribution. This flexibility makes changepoint detection particularly valuable in electricity demand forecasting, where it is essential to distinguish between true structural shifts in consumption and regular seasonal variations, thereby enhancing both forecasting accuracy and grid management.

Statistical Process Monitoring (SPM) provides a framework for real-time detection of changes in process behavior through control charts and related techniques (Joe Qin, 2003; Montgomery, 2020). Core SPM methods include Shewhart charts, CUSUM (Cumulative Sum) charts, and EWMA (Exponentially Weighted Moving Average) charts, each designed to detect shifts in process parameters such as mean and variance (Joe Qin, 2003; Montgomery, 2020). Shewhart charts are effective for identifying large, abrupt changes, while CUSUM and EWMA charts are more sensitive to small or gradual shifts (Montgomery, 2020).

In the context of electricity demand forecasting, SPM techniques face particular challenges due to seasonality, autocorrelation, and non-stationarity in demand data (Alwan and Roberts, 1988; Knoth and Schmid, 2004). Traditional control charts assume independent observations with constant mean and variance, assumptions frequently violated in electricity time series (Alwan and Roberts, 1988). To address these limitations, modified control charts have been developed, including residual-based charts that monitor forecast errors (Wardell et al., 1994; Yourstone and Montgomery, 1989), and time-series control charts that account for autocorrelation (Knoth and Schmid, 2004). Additionally, multivariate control charts such as Hotelling's T^2 can monitor multiple related series simultaneously, which is valuable in regional demand forecasting applications (Bersimis et al., 2007; Khazaie Poul et al., 2024).

The integration of SPM with changepoint detection methods offers complementary perspectives on process monitoring. While control charts provide continuous online monitoring with immediate alerts, changepoint detection methods can identify the precise timing and nature of structural changes retrospectively (Amiri and Allahyari, 2012; Hawkins et al., 2003). This combination is particularly relevant for electricity demand systems, where both real-time anomaly detection and post-hoc analysis of consumption pattern changes are essential for effective grid management (Khazaie Poul et al., 2024).

Changepoint detection methods aim to locate time points where the statistical properties of a series undergo abrupt changes. These methods can be broadly categorized into **offline** and **online** approaches, each serving distinct operational needs. Offline methods analyze complete datasets retrospectively to identify all changepoints simultaneously, prioritizing accuracy over computational speed. In contrast, online methods process data sequentially as new observations arrive, providing real-time changepoint identification at the cost of potentially reduced accuracy due to limited historical context.

Established offline approaches include Binary Segmentation (BS) (Scott and Knott, 1974), Optimal Partitioning (OP) (Jackson et al., 2005), and the Pruned Exact Linear Time (PELT) method (Killick et al., 2012), which leverage access to entire datasets for comprehensive analysis. Online methods, such as Bayesian online changepoint detection (Adams and MacKay, 2007), constant-memory algorithms (Wang et al., 2021), and kernel-based detection (Ferrari et al., 2023), sacrifice this comprehensive view for the ability to provide immediate alerts, making them valuable for grid operations requiring immediate responses to consumption anomalies. This thesis will only consider offline changepoint detection.

Despite their proven effectiveness in other domains, changepoint detection methods have seen limited application in electricity demand forecasting. This gap exists primarily due to the complexity of electricity demand time series, which exhibit multiple seasonal patterns (daily, weekly, annual), weather-dependent nonlinear relationships, and holiday effects that can be easily misinterpreted as structural breaks by standard changepoint algorithms. Traditional changepoint methods, designed for simpler time series, often generate excessive false positives when applied to demand data, as they may identify regular seasonal transitions or weather-related demand fluctuations as genuine structural changes. Additionally, the interpretability requirements in energy forecasting—where utilities need to understand *why* changes occur, not just *when*—favor domain-specific approaches that can distinguish between different types of structural shifts (policy changes, economic shifts, technological adoptions) rather than generic changepoint detection.

Although changepoint detection has not been widely used in electricity demand forecasting, it has demonstrated value in related domains. For instance, Yang et al. (2021) applied the PELT algorithm to mobility data during COVID-19, while Huang et al. (2017) analyzed precipitation shifts using changepoint methods. In predictive maintenance, Wang et al. (2023) applied binary segmentation to identify phases in battery degradation. These examples highlight the flexibility of changepoint methods and their potential for electricity demand analysis when appropriately adapted to handle the unique characteristics of energy time series.

Unlike existing changepoint models, this thesis incorporates model-based cost functions that account for seasonality and nonlinear demand patterns, building on approaches such as Sobhani et al. (2023), who developed a GAM-based changepoint detection algorithm to identify unrecorded load transfers. While their method—as well as the approach developed in this thesis—focuses on offline changepoint detection and does not incorporate real-time, online updating, standard algorithms like Binary Segmentation (BS), Optimal Partitioning (OP), and Pruned Exact Linear Time (PELT) remain widely used in many applications. This thesis expands on these ideas by proposing a novel framework that integrates GAMs with offline changepoint detection to capture both structural and parametric changes in electricity demand. Although not designed for real-time deployment, the framework lays a foundation that could be extended to online settings in future work. Chapter 3 develops this approach, with subsequent chapters evaluating its performance across simulated and real-world datasets.

Chapter 2

Modelling National Demand with Generalised Additive Model

2.1 Introduction

Short-term forecasting is crucial in the field of electrical energy due to its immediate impact on power system operation and energy markets. It helps grid operators optimise generation and demand response, ensuring grid stability and minimising costs. In energy markets, short-term forecasts guide bidding strategies, managing supply-demand imbalances. Long-term forecasting is less precise due to uncertainties like policy changes and technology shifts, making short-term forecasts more actionable for real-time decision-making. The aim of this chapter is to develop a statistical model for point forecasting the **total electricity demand** which is defined as the sum of national demand, estimated wind generation, and solar generation, which represents an estimate of the electricity consumption in Great Britain. In 2018, average demand per half hour decreased by 6504 MW (-16.71%) compared to that in 2006 as shown in Figure 2.1 which presents yearly averages. The factors contributing to the declining trend of total demand over time are the impact of improved energy efficiency of appliances and the decline of heavy industry [1] . Improved energy efficiency of appliances reduces overall energy consumption by decreasing the amount of energy required to

[1] https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/, accessed on 6 November 2024

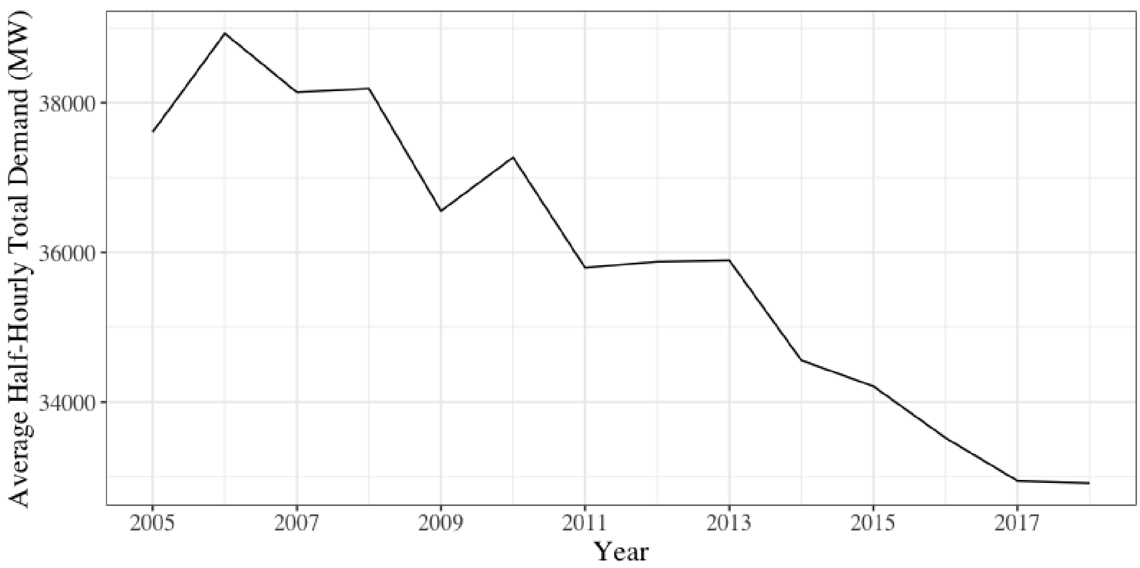

Figure 2.1: The available data on average half-hourly total demand spans from April 2005. Upon obtaining the remaining data from January to March 2005, it is possible that the average half-hourly total demand in 2005 may increase.

produce the same output. On the other hand, the decline of heavy industry, such as manufacturing and mining, reduces the demand for energy to power their operations. The combined effect of these factors leads to a significant reduction in the total demand, which has positive impacts on the environment, including the mitigation of greenhouse gas emissions and the effects of climate change.

This chapter starts with an overview of the data used in the study, encompassing both collected data from organisations and commonly known holiday data. The use of holidays and weather features as potential predictors will be explored in the model building process. The subsequent sections focus on analysing the available data to gain insight and establish connections with model building. The methodology and mathematical expressions used in model construction are discussed, followed by an exploration of methods for comparing and evaluating different models. The chapter progresses to outline the study design, detailing the approach for obtaining results, which are subsequently presented. The chapter concludes with a summary of key findings and implications.

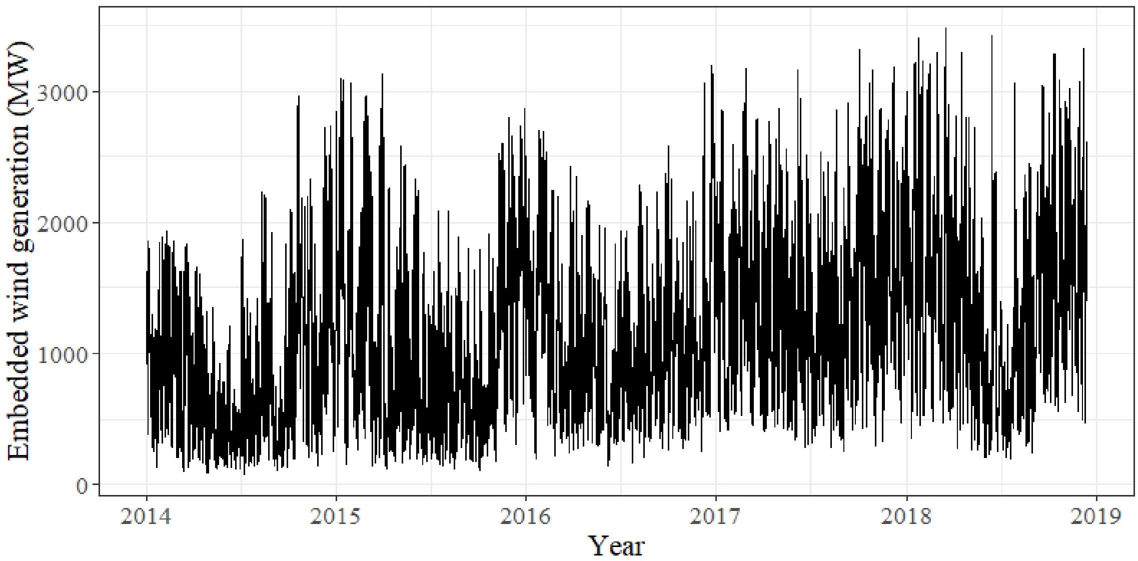

Figure 2.2: Embedded wind generation over time, showing short-term variations driven by fluctuating wind speeds. The overall magnitude of wind generation increases over time as more wind generation devices are installed.

2.2 Exploratory Analysis

2.2.1 Total Demand

The data for this chapter consists of national demand data sets collected by the National Grid Electricity System Operator (National Grid ESO) [1], from 2014 to 2018. The data include half-hourly national demand for Great Britain in units of Mega-Watts (MW). In addition, estimates of wind generation from wind farms which do not have Transmission System metering installed and solar generation from photovoltaic (PV) panels are collected in MW and will be considered in terms of total demand as renewable energy trends emerged during this period. They are generated from small-scale generators which are not measured directly by National Grid ESO. These sources contribute to negative demand as they are generation and not consumption.

Figure 2.2 presents the temporal trend in embedded wind generation estimated by National Grid ESO starting from 2014. The generation exhibits short-term variations due to uncontrollable wind speed. On the contrary, Figure 2.3 presents the temporal trend in estimated embedded solar generation over time estimated by National Grid ESO which an increasing amplitude pattern over time can be observed. The bell-shaped pattern observed each year represents the seasonal variation, with high

[1] https://data.nationalgrideso.com/, accessed on 7 March 2023

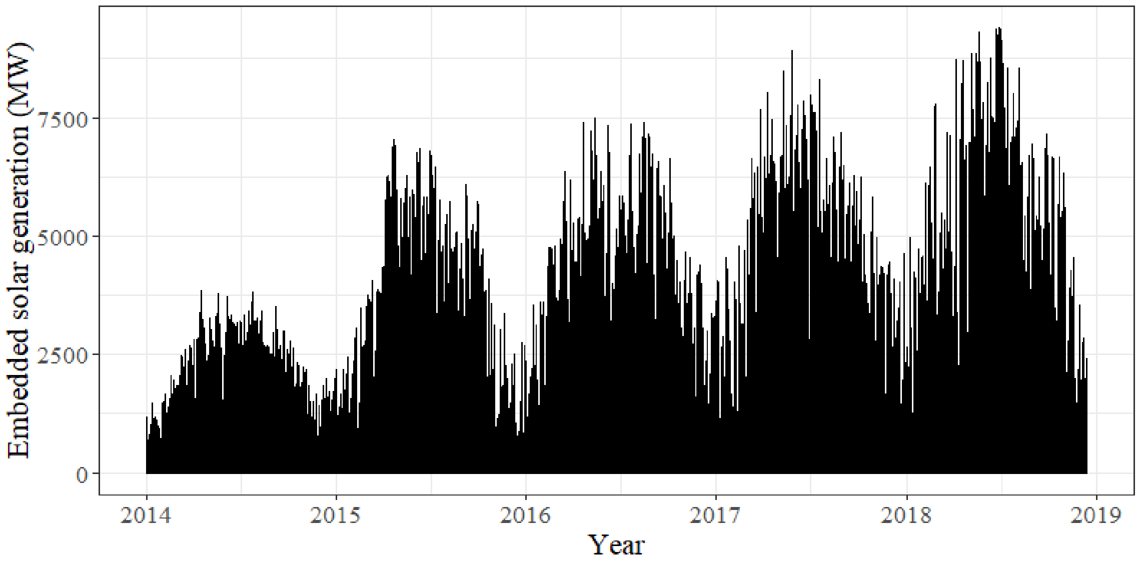

Figure 2.3: Embedded solar generation over time, showing an increasing amplitude pattern as more devices are installed. The annual bell-shaped pattern reflects seasonal variation, with higher generation in summer due to increased solar radiation. Regular drops to zero indicate nighttime periods. Both solar and wind generation display an upward trend over time, highlighting the expanding installed capacity of renewable energy.

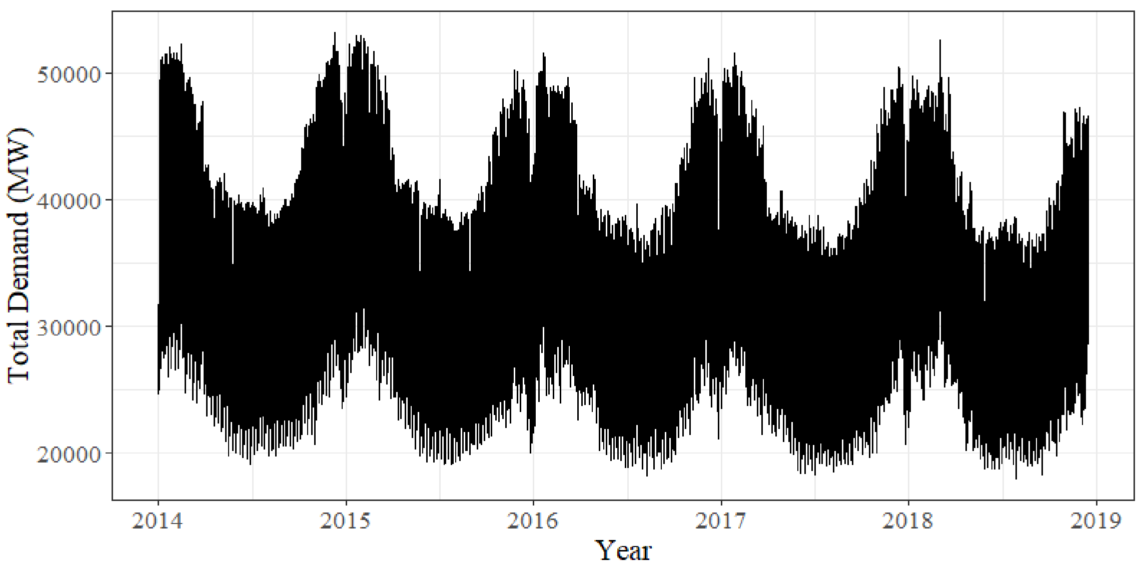

Figure 2.4: Total electricity demand over time, illustrating a general declining trend. Distinct drops in demand at the beginning and end of each year correspond to the Christmas and New Year holidays. The recurring seasonal pattern each year reflects variations, with lower demand in summer and higher demand in winter due to increased heating requirements.

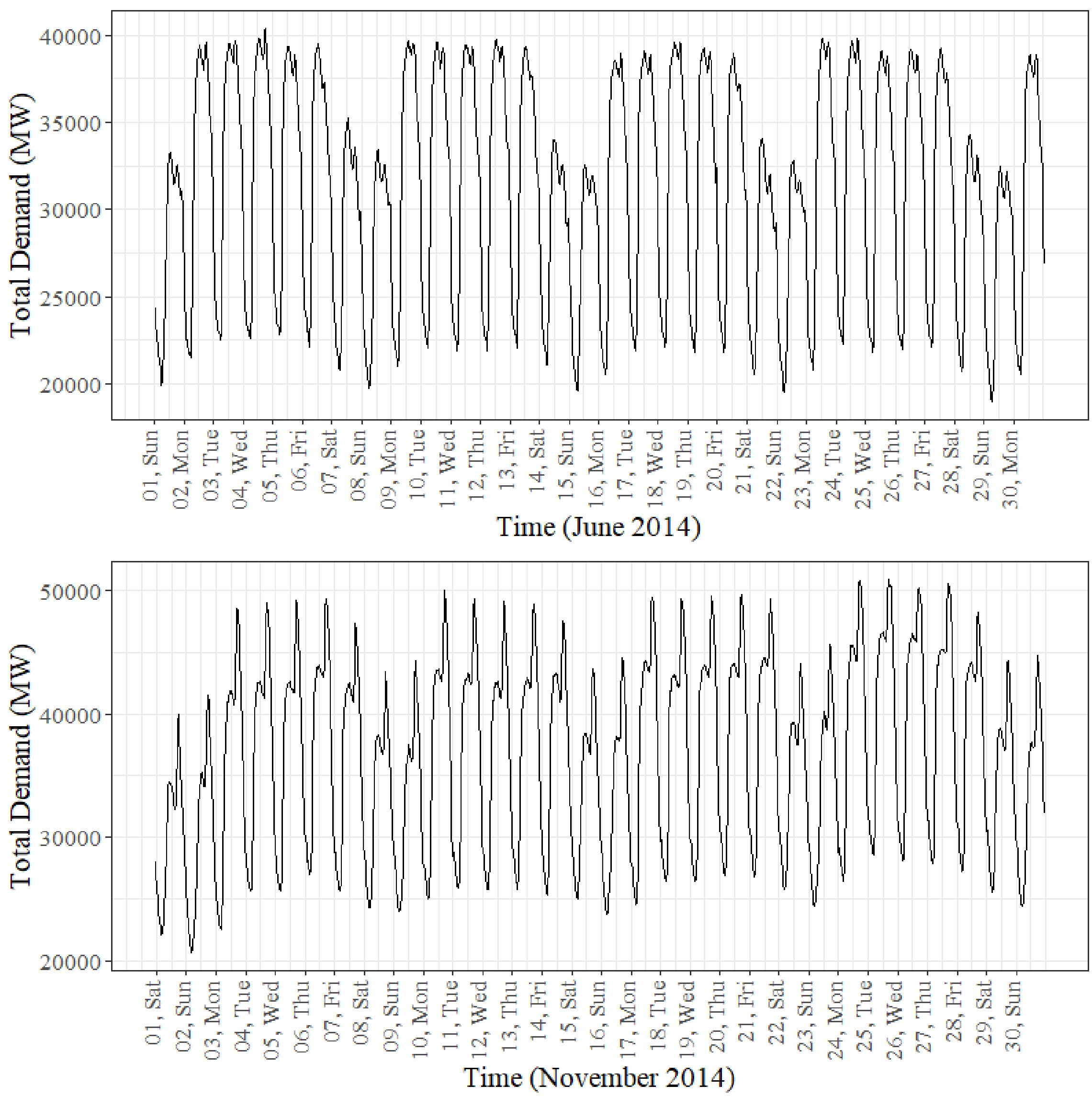

Figure 2.5: Total demand in June 2014 (top) and in November 2014 (bottom).

generation in summer due to the abundance of solar radiation. The values reach zero regularly, indicating night time each day. Both wind and solar generation exhibit an increasing trend over time, reflecting the growing installed capacity of renewable energy, with more devices installed over time. Figure 2.4 shows the plot of total demand over time. The seasonal pattern observed in each year represents the seasonal variation, with lower demand in summer and higher demand in winter when electrical heating is required.

Figure 2.5 presents the demand patterns for 2 sample months to illustrate the daily variation. June and November 2014 serve as examples to illustrate the differences between seasons, as they do not exhibit unusual patterns due to holidays (see Subsection 2.2.3). The figure reveals that weekdays and weekends exhibit similar shapes but can vary across seasons due to fluctuations in people's behaviour. In the summer, demand is typically higher during working hours on weekdays, while on weekends, it is higher in the morning and evening, albeit not as high as on weekdays. In contrast, during the winter season the overall demand is higher compared to summer, and the weekday profiles transition from low variation demand during working hours to a single peak in the evening. This pattern is also observed in the weekend profiles.

2.2.2 Weather Features

Weather features will be incorporated into the model-building process. These features can be found in the European Centre for Medium-Range Weather Forecasts (ECMWF) ReAnalysis 5 (ERA5) data [1], which provides hourly estimates of weather variables. The data used in this chapter covers the period from 2014 to 2018, which overlaps with the demand data. The weather variables consist of 9 features as shown in Table 2.1 (above the dashed line). The 2 metre dew point temperature refers to the temperature at which the air would become saturated with water vapour, leading to the formation of dew or fog. It is a measure of the amount of moisture present in the air at a height of 2 metres above the ground. The 2 metre temperature refers to the air temperature at a height of 2 metres above the ground. It is a measure of the thermal energy of the air in the lower part of the atmosphere. Temperature can significantly impact the amount of energy used because higher temperatures increase the demand for cooling, leading to increased energy consumption by air conditioning systems. Additionally, extreme cold temperatures can also raise energy usage as heating systems need to work harder to maintain indoor comfort levels.

The 100m and 10m U and V wind components refer to the horizontal wind speed measured at a height of 100 and 10 metres above the Earth's surface respectively. The U component represents the east-west wind speed, while the V component represents the north-south wind speed. Together, the U and V components provide a vector representation of the wind speed and direction at that altitude. By apply-

[1] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5, accessed on 27 March 2023

Weather feature	Short name	Unit
2 metre dew point temperature	d2m	K
2 metre temperature	t2m	K
100 metre U wind component	u100	m/s
100 metre V wind component	v100	m/s
10 metre U wind component	u10	m/s
10 metre V wind component	v10	m/s
Instantaneous 10 metre wind gust	i10fg	m/s
Snow depth	sd	m of water equivalent
Snowfall	sf	m of water equivalent
100 metre wind speed	w100	m/s
10 metre wind speed	w10	m/s

Table 2.1: Available weather features in the ERA5 data (above the dashed line) and additional calculated weather features (below the dashed line).

ing the Pythagorean theorem to the horizontal components of wind speed (i.e., the east-west and north-south components), one can calculate the wind speed (w100 and w10, Table 2.1 (below the dashed line)) at a given location, expressed as the magnitude of the wind vector. Instantaneous 10 metre wind gust is the maximum wind speed over a 10 minute interval. It is measured at a height of 10 meters above the ground and represents a short-lived burst of high wind speed that can occur during thunderstorms, squalls, or other weather events. Wind speed affects the amount of energy used by impacting the demand for heating and cooling. Higher wind speeds can increase the rate of heat transfer from buildings, leading to increased heating or cooling requirements. Snow depth is the vertical depth of snow on the ground and snowfall is the amount of snow that falls during a given period of time. It can affect the amount of energy used by increasing the demand for heating and insulation, as more energy is required to maintain comfortable indoor temperatures in colder regions with higher snow depths. Additionally, snow accumulation can hinder transportation and infrastructure, leading to increased energy consumption for snow removal and maintenance activities.

Figure 2.6 (left) presents the locations of the modelled weather data estimated at 1517 evenly distributed grid squares throughout the UK. The data consists of 41 longitude values ranging from -7.00 to 3.00 and 37 latitude values ranging from 50.00 to 59.00, with 0.25 increments (in degrees) in both dimensions. Since the time resolution of the weather data (hourly) differs from the electricity demand data, which is collected half-hourly, an interpolation of the weather data is required. This

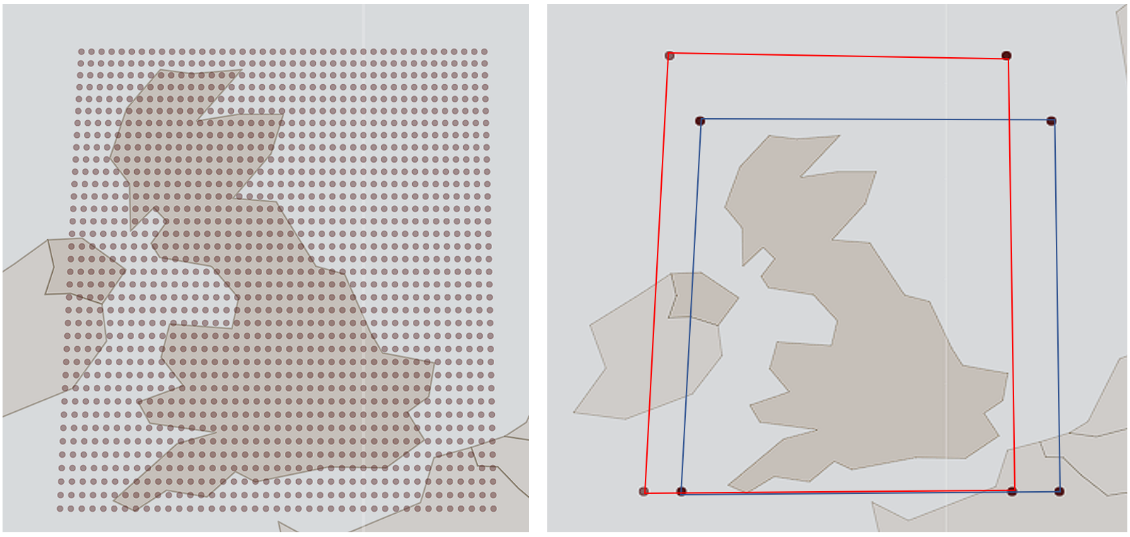

Figure 2.6: Available locations of hourly estimated weather features (left) and available area of (blue) hourly estimated weather features and (red) population (right).

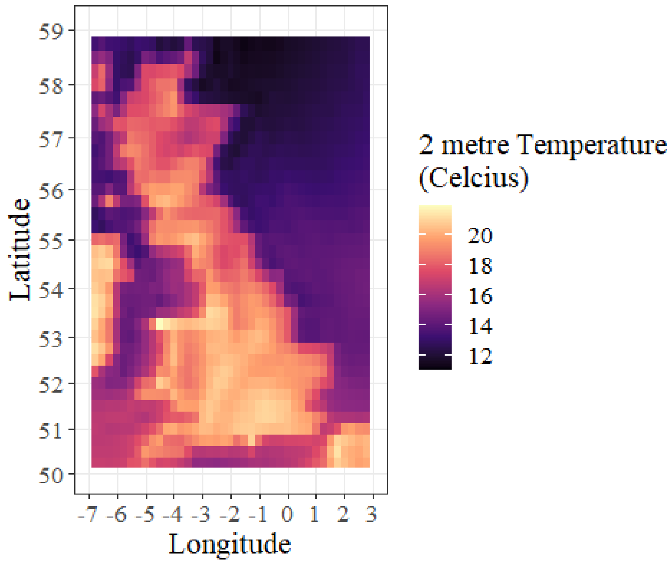

Figure 2.7: Air temperature at 2 metres above the surface on 1 July 2014 at 2PM.

is accomplished by using the linear interpolation of consecutive hourly estimates to fill in half-hourly gaps.

The correlation between weather patterns in geographically proximate areas, specifically the 2 metre temperature, is evident in Figure 2.7, as observed in the northern region where temperatures are cooler than the hotter southern region, especially the southeast. The figure shows the distinct shape of Great Britain and suggests a correlation between weather variables in closely situated locations. These 1517 data grid points of each weather variable are collected in a time series format at a time resolution of one hour. Figure 2.8 shows the 2-metre temperature in major cities of Great Britain in a time series format.

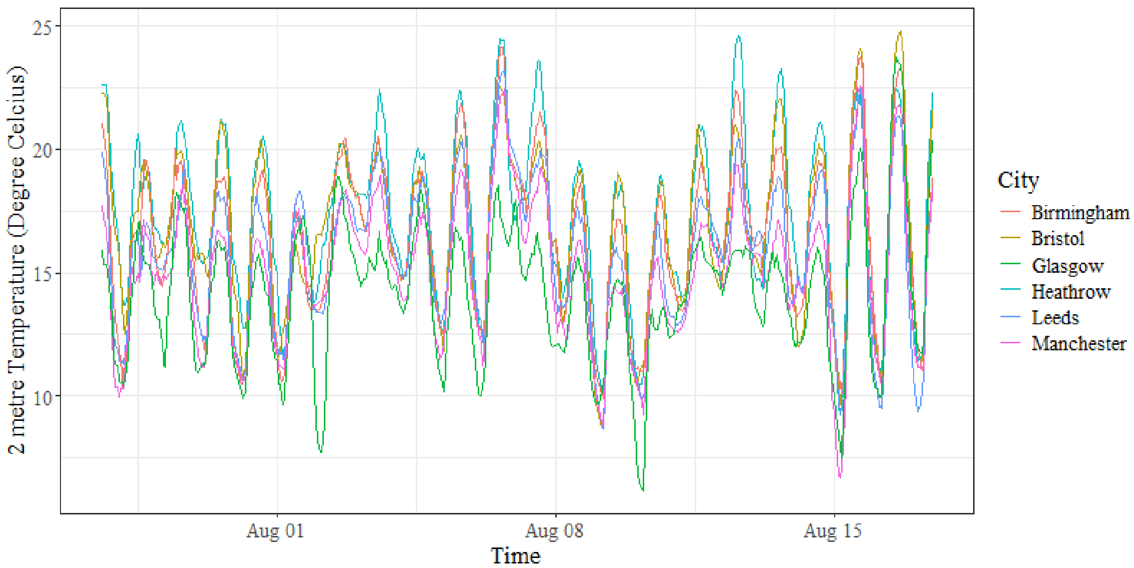

Figure 2.8: Air temperature at 2 meters above the surface of certain cities in 2016.

To address the misalignment between electricity demand—recorded as a single half-hourly value—and each weather feature—available as hourly, gridded values across the country—it is preferable to reduce each weather variable to a single representative value for each half-hour interval when fitting a model. A common approach is to take the spatial average of the weather variable across all grid points; however, this method implicitly treats each location equally, regardless of its relevance to electricity demand. For variables such as temperature, a more meaningful estimate can be obtained by using population-weighted averages, which give greater importance to areas with higher electricity usage, thereby better reflecting the true weather conditions experienced by the population.

By assigning higher weights to areas with larger populations, the average temperature calculation becomes more representative of the temperature experienced by the majority of Great Britain's inhabitants. This approach acknowledges that areas with higher population densities exert a greater influence on the overall temperature experienced by residents throughout Great Britain.

The UK gridded population 2011, based on Census 2011 data [1], is used to assign population weights to each grid point. The data set provides population data at a spatial resolution of 1 km $\times$ 1 km, which is different from the resolution of the weather data. The mapping product is based on the British National Grid (OSGB36

[1](https://data.gov.uk/dataset/ca2daae8-8f36-4279-b15d-78b0463c61db/uk-gridded-population-2011-based-on-census-2011-and-land-cover-map-2015, accessed on March 27, 2023

datum) [1], and the geometric conversion, which takes into account the curvature of the earth, can be performed using the `sf` package (Pebesma et al., 2021) in R. The resulting coordinates are rounded to match the resolution of the weather data in each grid square (0.25 increments in degrees of latitude/longitude), and multiple population data of the same grid square are summed up into one population for each corresponding grid. The longitude values range from -8.00 to 1.75 and the latitude values range from 50.00 to 60.75, causing the region of the coordinates to overlap in Figure 2.6 (right). The intersecting part covers the entire Great Britain, with longitude values ranging from -7.00 to 1.75 and latitude values ranging from 50.00 to 59.00, while non-land and Ireland areas will be excluded.

The distribution of the population across the grid cells was analysed in order to apply the population weight to the weather features. The spatial population distribution over the Great Britain is shown in Figure 2.9 (top left). The greater London area is found to be the population center, followed by the west midlands and north west area, which correspond to Birmingham and Manchester, respectively. A histogram of this distribution is presented in Figure 2.9 (top right). This distribution shows a significant right-skewness due to the presence of a few high population centers. While direct population weighting is a possible method, it may not be suitable for areas with sparse population density due to the limited consideration given to them. To address this, we applied a logarithmic transformation to the data to normalise the gridded population, which is illustrated in Figure 2.9 (middle right). Figure 2.9 (middle left) displays the spatial distribution after applying a logarithmic transformation. The major cities in the UK are identified as population centers, while certain northern parts of Scotland show no population presence. The transformed distribution shape is now closer to a normal distribution but exhibits left-skewness.

A Box-Cox transformation (Atkinson et al., 2021) is a technique used to transform a non-normal distribution into a normal distribution. The formula of this transfor-

[1] https://www.ordnancesurvey.co.uk/business-government/tools-support/os-net/coordinates, accessed on 27 March 2023

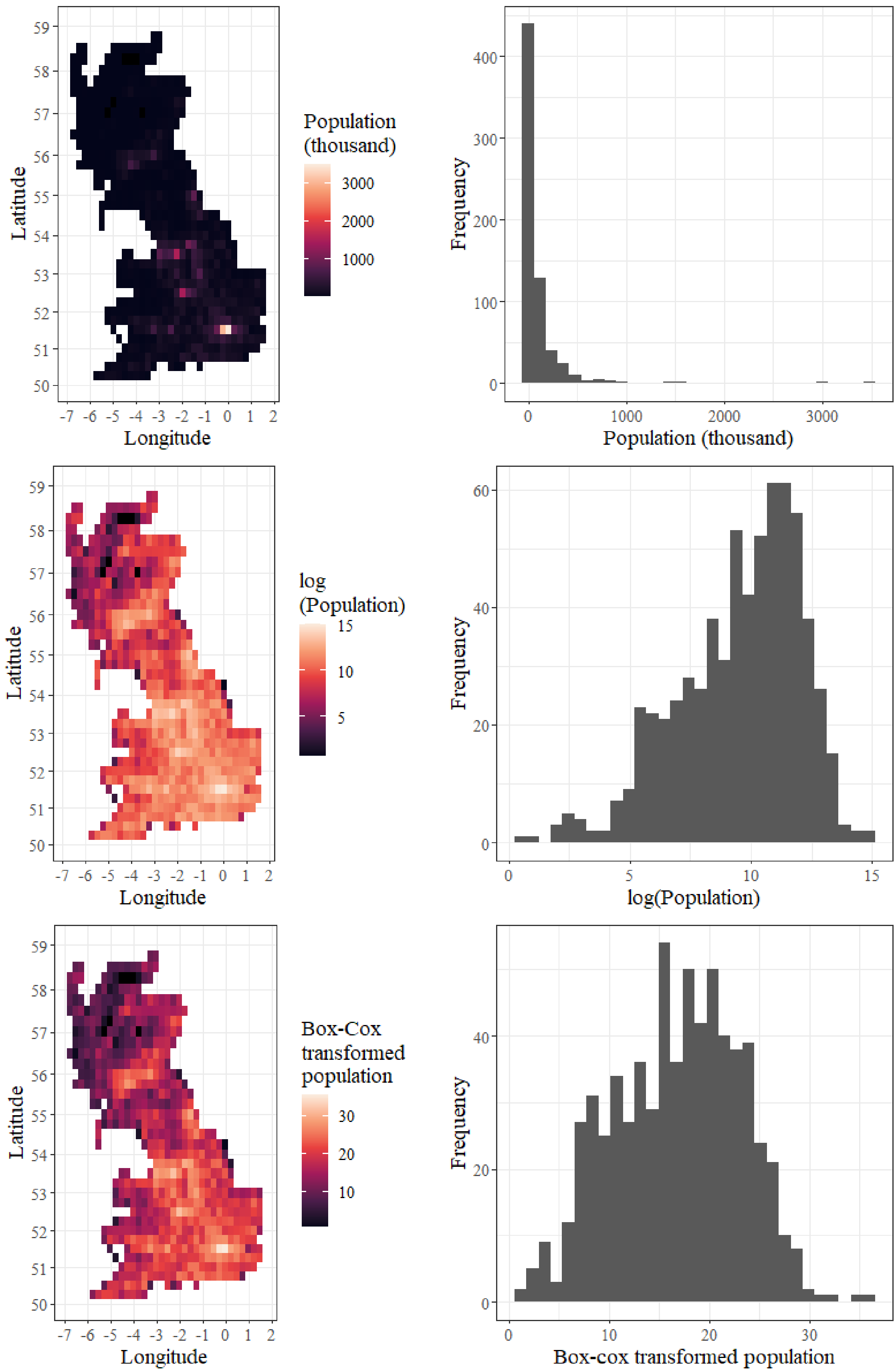

Figure 2.9: The gridded population of Great Britain in 2011, with the top left and top right plots displaying the raw population counts. The middle left and middle right plots depict the logarithmic transformation of the population, while the bottom left and bottom right plots display the Box-Cox transformed population using the optimal parameter. The plots on the left showcase the spatial distribution, while the plots on the right represent the distribution of each population grid cell.

mation to a variable y is

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0 \end{cases}$$

for a selected value of λ. Maximising the log-likelihood function enables the determination of the optimal value of λ that brings the distribution closest to a normal distribution. By applying the optimal λ value of 0.10 to the population data set, a Box-Cox transformed distribution of the Great Britain population is achieved, as depicted in Figure 2.9 (bottom left and right).

The result of applying the weight to a weather feature W_t at time period t can be calculated by

$$W_t = \frac{\sum_x p_x W_{x,t}}{\sum_x p_x}$$

when p_x is a weighted population of selected weight choice (no transformation/logarithmic/Box-Cox) to the grid cell x and $W_{x,t}$ is the value of the weather feature corresponding to the grid cell x at time period t. In order to determine the optimal transformation, this study considers three potential choices as population weight options for the 2 metre temperature in the model because the temperature is the most widely used predictor in the literature among the other weather features. Since temperature holds the most significant influence on demand, wherein lower temperatures result in higher demand and vice versa, the selection of population weight for temperature in relation to other weather features should accurately reflect the appropriate value for each individual weather characteristic. The population weight choice that has the best predictive performance on 2 metre temperature will be applied in assigning the weight to each weather feature.

2.2.3 Bank Holidays

This chapter also considers past Great Britain bank holidays from 2005 to 2018. The holidays differ between England and Wales and Scotland as shown in Table 2.2. Every year, Good Friday and Easter Monday fall on the same weekends but different dates due to the lunar calendar. The regions also have different summer bank holidays,

Bank holiday	England and Wales	Scotland	Day
New Year's day	✓	✓	1 January
2nd January		✓	2 January
Good Friday	✓	✓	Friday, varies
Easter Monday	✓		Monday, varies
Early May bank holiday	✓	✓	First Monday of May
Spring May bank holiday	✓	✓	Last Monday of May
Summer bank holiday	✓	✓	Monday of August
St Andrew's day		✓	30 November
Christmas day	✓	✓	25 December
Boxing day	✓	✓	26 December
The Wedding of Prince William and Catherine Middleton	✓	✓	29 April 2011
The Diamond Jubilee of Elizabeth II	✓	✓	5 June 2012
The Platinum Jubilee Bank Holiday	✓	✓	3 June 2022

Table 2.2: Yearly Bank Holidays (Top) and Special Bank Holidays (Bottom) in Great Britain.

with the last Monday of August for England and Wales and the first Monday of August for Scotland. When fixed-date bank holidays fall on weekends, the following working day becomes a substitute bank holiday except for 2nd January of Scotland. In certain years, special bank holidays were observed to mark significant events, such as the wedding of Prince William and Catherine Middleton in 2011. Additionally, the Spring May bank holiday was moved to June 4, 2012, in honor of the Diamond Jubilee of Elizabeth II, resulting in a four-day holiday.

Bank holidays can often be identified in a plot by unfamiliar patterns that are similar to weekends and they can be classified into three groups: Christmas Day, New Year's Day, and all others. On Christmas Day, there is typically a high demand in the morning and a peak at noon, while on New Year's Day, demand increases slightly between midnight and 1 AM. Similarly, on other bank holidays, demand is lower in the morning and peaks in the evening. During the Christmas season (around 20 December to 6 January), weekdays may behave like weekends, as shown in Figure 2.10, which depicts the total demand from mid-December to early January. The demand profiles in winter usually have a single peak in the evening, except for special days such as Christmas and New Year's Day, which have their own patterns every year, regardless of day of the week.

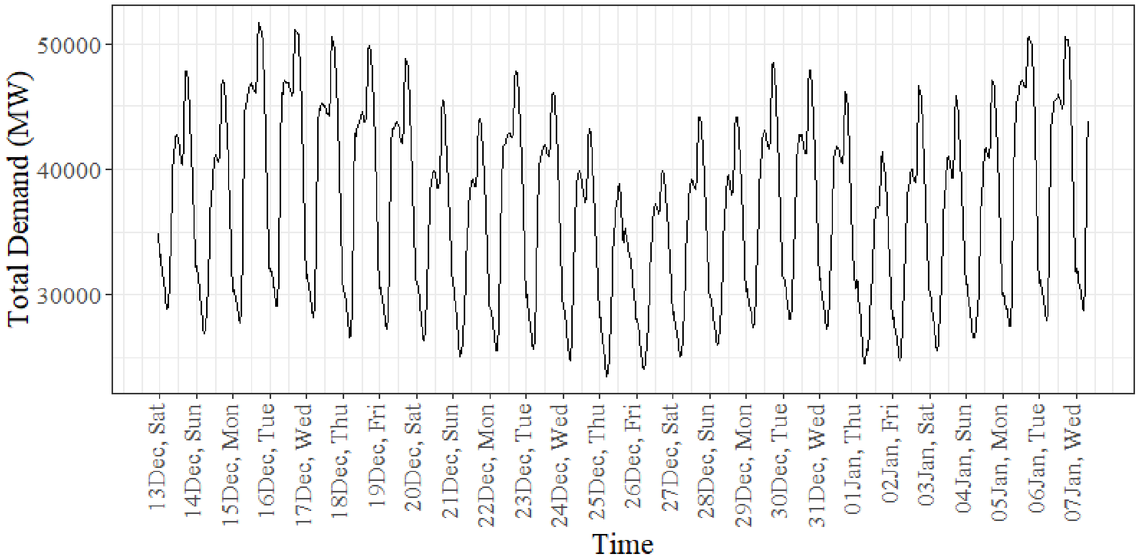

Figure 2.10: Total demand during 2015 New Year.

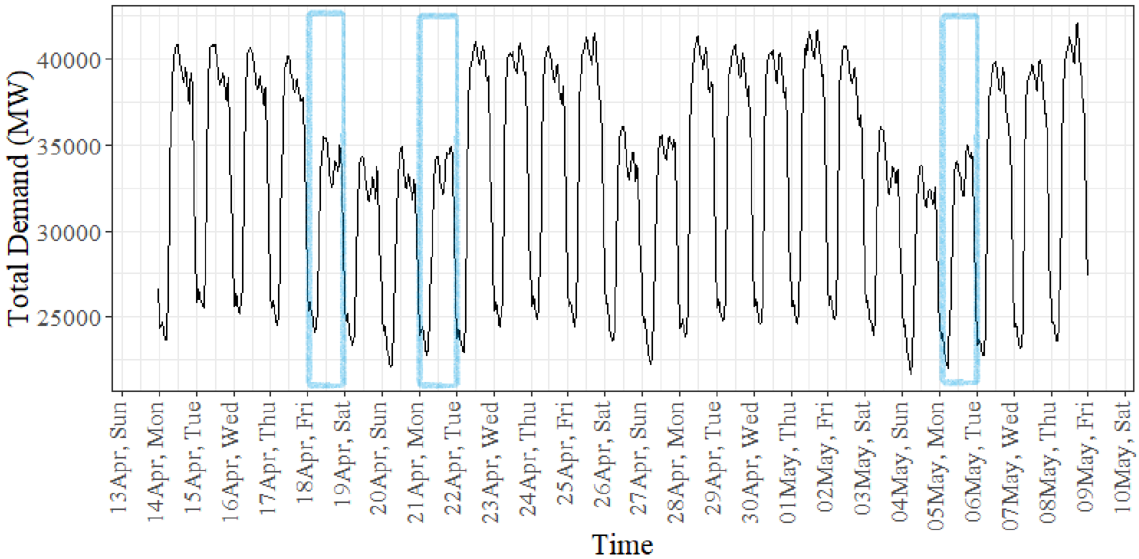

Figure 2.11: Total demand during April and May 2014. The bank holidays in this period are highlighted; 18 April is Good Friday, 21 April is Easter Monday, and 5 May is Early May bank holiday.

Figure 2.11 illustrates the demand patterns during April and May 2014, with Good Friday (18 April), Easter Monday (21 April), and Early May bank holiday (5 May) highlighted. The last weekdays of each week exhibit lower demand patterns compared to the other weekdays in the same week. This figure also shows that the total demand is higher during normal weekends compared to weekends with bank holidays on Friday and/or Monday. To investigate this, we define the terms **England and Wales vacation** and **Scotland vacation** as periods of three or more consecutive days off for each respective region. Typically, these vacations occur when a bank holiday falls on a Friday or Monday, allowing people to spend time off work consecutively during weekends. However, some non-holiday weekdays during the Christmas and New Year season are also considered vacations as people tend to take time off work to celebrate the season. The demand profiles from 20 December to 6 January in seven different years for all possible weekly patterns (starting on Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday) are displayed in Figure 2.12. The demand profiles on 25 December and 1 January are highlighted in the figure.

After inspection of the patterns, the selection of vacation days can be based on the day that Christmas and New Year's Day fall on. The first day of both vacations must have the following condition:

1. If 25 December falls on weekends (e.g. 2010 and 2011), so that the last day that most people work is the preceding Friday, hence the start day is Saturday of that weekend.

2. If 25 December falls on Monday or Tuesday (e.g. 2006 and 2012), the start day is the last Saturday before Christmas day. In the case that 25 December falls on Monday, the last day that most people work is the preceding Friday. When it falls on Tuesday, it is possible that people tend to stop working on Monday so they can take consecutive vacations starting from Saturday. A demand profile of 2012 shows an example of this behaviour.

3. The start day is 25 December otherwise. It is arguably to determine the starting vacation date when 25 December falls on Wednesday to be either the Saturday before or that Wednesday. Based on the evidence from the plot, I decide to set

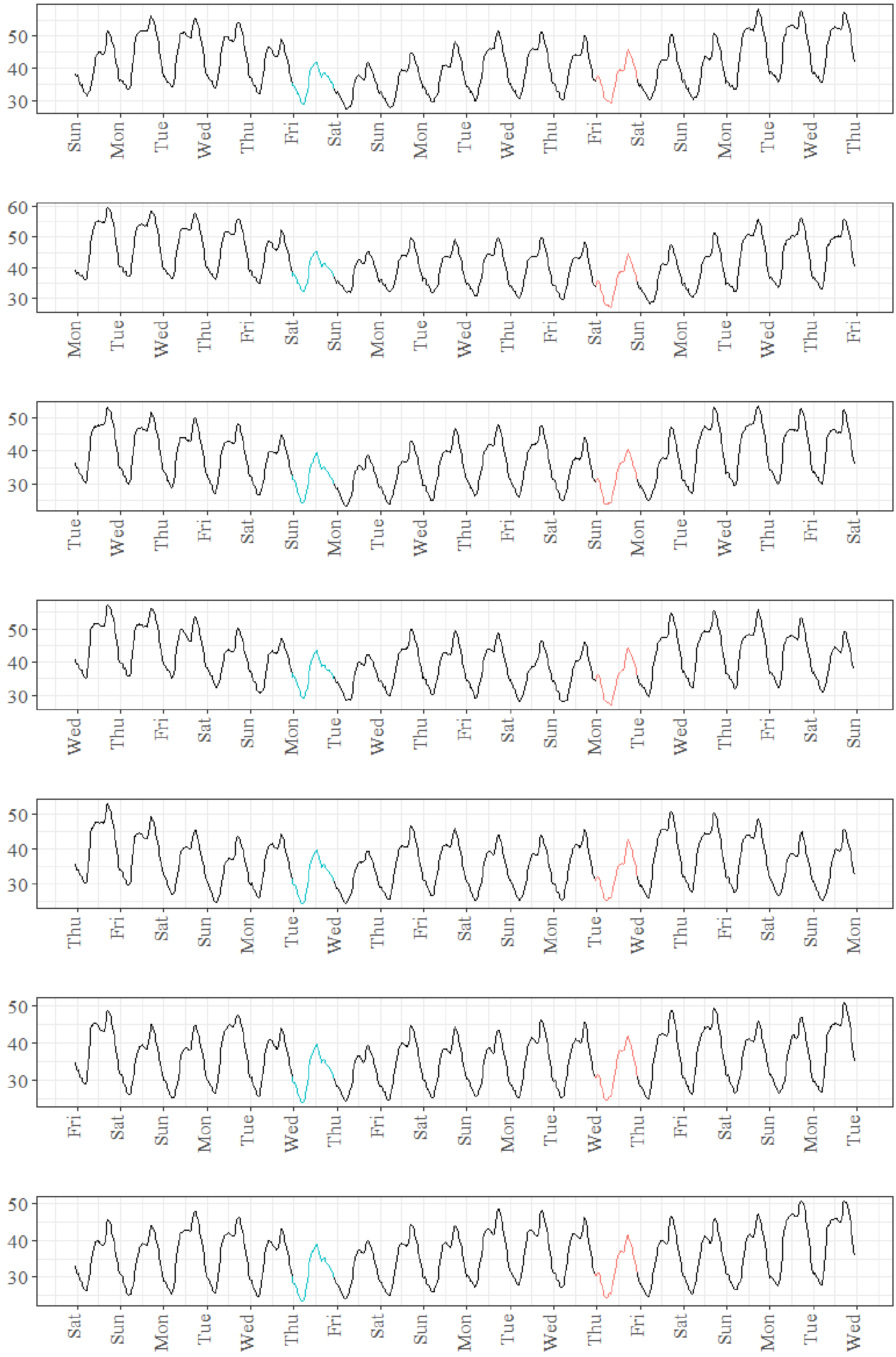

Figure 2.12: The total demand in GW from 20 December to 6 January in seven distinct years with cyan lines denote the demand on 25 December, while red lines represent 1 January. The demand profiles for the years 2009, 2010, 2011, 2006, 2012, 2013, and 2014 are shown from top to bottom, respectively.

the beginning of a vacation on Wednesday because the preceding Monday has a similar profile to working days.

The last day of the vacation is selected by the following:

1. If 1 January falls on weekends so that New Year's day is on next Monday because of the substitute holiday (e.g. 2011 and 2012), the last day is that Monday where the substitute New Year's day falls on. Note that Scotland's 2nd January is also a bank holiday but the substitute bank holiday does not apply when it falls on weekend.

2. If 1 January falls on Thursday or Friday (e.g. 2010 and 2015), the vacation's last day of both regions is the first Sunday after New Year's day. People tend to continue their vacation until the following weekends based on evidence in the plot.

3. If 1 January falls on Wednesday (e.g. 2014), the vacation's last day of Scotland is the first Sunday after New Year's day and of England and Wales is that Wednesday. This is because Scotland has the 2nd January bank holiday on Thursday so people tend to not work on Friday. However for England and Wales, the plot show that the demand profiles during the following Thursday and Friday are higher than the New Year's Day itself and the following weekends, so it is arguably that most people start to work on Thursday.

4. The last day is 1 January otherwise. This includes the case that New Year's Day falls on Monday and Tuesday.

2.3 Methodology - Generalised Additive Model

The Generalised Additive Model (GAM) (Wood, 2017) is selected to fit the data due to its flexibility in capturing daily demand profiles via non-linear covariate-response relationships. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be the response variable of total demand where $y_t \sim \text{EF}(\mu_t, \phi)$ for some exponential family (EF) distribution with mean μ_t and scale parameter ϕ. Let $\mathbf{X} = (x_{tj})_{n\times p}$ be the $n \times p$ design matrix where n is the number of observations and p is the number of covariates. The covariates of $\mathbf{X}$ are split into two block matrices; $\mathbf{X} = (\mathbf{A} : \mathbf{Z})$, where covariates in $\mathbf{A}$ are linear terms in the model and those in $\mathbf{Z} = (z_{tj})$ are treated as smooth terms. Let $\mathbf{A}_t$ be the t-th row of $\mathbf{A}$ with corresponding parameters $\boldsymbol{\gamma}$, and f_j be a smooth function of covariate z_j.

The model is in the form

$$g(\mu_t) = \mathbf{A}_t\boldsymbol{\gamma} + \sum_j L_{tj}(f_j(z_j)), \tag{2.1}$$

where the response variable y_t has expectation μ_t and $g(\cdot)$ is a monotonic link function, and L_{tj} is a bounded linear functional of f_j which is either $L_{tj}(f_j(z_j)) = f_j(z_{tj})$ or $L_{tj}(f_j(z_j)) = v_t f_j(z_{tj})$ when v_t is a varying coefficient variable that interacts with z_{tj}. If v_t is a factor variable, then the above formulation $v_t f_j(x_{tj})$ is replaced by separating it into m different functions, $f_j^{(1)}(z_{tj}), \ldots, f_j^{(m)}(z_{tj})$, for each level of the factor when m is the number of levels of v_t.

In both cases of L_{tj}, $f_j(z_{tj}) = \sum_{k=1}^{K_j} \alpha_{jk} b_{jk}(z_{tj})$ where K_j is the basis dimension, $b_{jk}(\cdot)$ are spline basis functions, and α_{jk} are coefficients. The matrix $\mathbf{Z}$ is now in the form $(\boldsymbol{\zeta}^{[1]} : \boldsymbol{\zeta}^{[2]} : \ldots)$ which is created by combining $\boldsymbol{\zeta}^{[j]}$ column-wise, where $\zeta_{tk}^{[j]} = L_{tj}(b_{jk}(z_{tj}))$ for each j. So far, the model equation 2.1 can be rewritten in the form

$$g(\mu_t) = \mathbf{D}_t\boldsymbol{\beta}, \tag{2.2}$$

where $\mathbf{D}_t$ is the t-th row of a whole model matrix $\mathbf{D}$ that contains spline basis functions and linear terms and $\boldsymbol{\beta}$ is the corresponding model coefficient vector containing $\boldsymbol{\gamma}$ and the individual smooth term coefficients $\{\alpha_{jk}\}$ vectors stacked row-wise.

To control the trade-off between the smoothness of each estimate f_j and fidelity to the data, a smoothing parameter, λ_j, is used. A larger value of λ_j leads to a smoother function, effectively reducing the complexity of the model. Smoother functions are less likely to overfit the training data but may fail to capture certain intricate patterns in the test data. On the other hand, a smaller value of λ_j allows the model to closely follow the training data, capturing more detailed variations but potentially leading to overfitting of the training data and possibly poor predictive performance.

The smoothing parameter to control the penalisation term is chosen to be the integral of the second derivative of each f_j, i.e., $\lambda_j \int f_j''(x)^2 dx$, where λ_j is the smoothing parameter corresponding to the smooth function f_j. In this study, a cubic regression spline is used as a smooth function. A total smoothing penalty for the model can be written as $\sum_j \lambda_j \boldsymbol{\beta}^{\mathbf{T}} \mathbf{S}_j \boldsymbol{\beta}$ where $\lambda_j \boldsymbol{\beta}^{\mathbf{T}} \mathbf{S}_j \boldsymbol{\beta}$ is the penalty for f_j. See Wood (2017) for the entries of $\mathbf{S}_j$.

The model coefficients can be estimated by maximisation of the penalised likelihood of the distribution $\mathrm{EF}(\mu_i, \phi)$ when $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$ is given where $l(\boldsymbol{\beta})$ is the likelihood function,

$$l_p(\boldsymbol{\beta}, \boldsymbol{\lambda}) = l(\boldsymbol{\beta}) - \frac{1}{2\phi} \sum_j \lambda_j \boldsymbol{\beta}^{\mathbf{T}} \mathbf{S}_j \boldsymbol{\beta},$$

which can be maximised via the penalised iteratively re-weighted least squares (PIRLS) algorithm as follows.

Algorithm

1. Initialise $\hat{\mu}_t = y_t + \delta_t$ and $\hat{\eta}_t = g(\hat{\mu}_t)$, where δ_t is usually zero or a small constant to ensure that $\hat{\eta}_t$ is finite.

2. Compute $v_t = g'(\hat{\mu}_t)(y_t - \hat{\mu}_t)/\alpha(\hat{\mu}_t) + \hat{\eta}_t$ and $w_t = \alpha(\hat{\mu}_t)/[g'(\hat{\mu}_t)^2 V(\hat{\mu}_t)]$.

3. Find $\hat{\boldsymbol{\beta}}$ to minimise the weighted least squares objective

$$\sum_t w_t (v_t - \mathbf{D}_t^{\mathbf{T}} \boldsymbol{\beta})^2 + \sum_j \lambda_j \boldsymbol{\beta}^{\mathbf{T}} \mathbf{S}_j \boldsymbol{\beta}$$

   and then update $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\mu}_t = g^{-1}(\hat{\eta}_t)$.

4. Repeat step 2 and 3 to convergence,

where $\mathbf{v} = \{v_1, \ldots, v_n\}$, $\mathbf{W}$ is a diagonal matrix that $W_{tt} = w_t$, $||\mathbf{a}||^2_{\mathbf{W}} = \mathbf{a}^{\mathbf{T}}\mathbf{W}\mathbf{a}$, $V(\eta)$ is the variance function determined by the exponential family distribution, and $\alpha(\mu_t) = 1 + (y_t - \mu_t)[V'(\mu_t)/V(\mu_t) + g''(\mu_t)/g'(\mu_t)]$.

The estimation of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$ can be performed by several ways. The generalised cross validation score (GCV) score is a guide for selecting appropriate levels of smoothing and model complexity and it is an approach that used to find the optimal value of $\boldsymbol{\lambda}$. A lower GCV score suggests a better-fitting model that captures the underlying patterns in the data while avoiding overfitting or underfitting. The goal is to find the $\boldsymbol{\lambda}$ that minimises the GCV score, striking a balance between goodness of fit and model complexity. The score is defined by

$$\mathrm{GCV}(\boldsymbol{\lambda}) = \frac{n||\mathbf{y} - \hat{\boldsymbol{\mu}}(\boldsymbol{\lambda})||^2}{[n - \mathrm{tr}(\mathbf{M}(\boldsymbol{\lambda}))]^2},$$

where $\mathbf{M}(\boldsymbol{\lambda})$ is the influence matrix for the model which is defined by $\mathbf{M}(\boldsymbol{\lambda}) = \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X} + \sum \lambda_j \mathbf{S}_j)^{-1}\mathbf{X}^\mathbf{T}$ and $\hat{\boldsymbol{\mu}} = \mathbf{M}(\boldsymbol{\lambda})\mathbf{y}$.

The function `gam`, which is utilised for fitting a GAM model, can be accessed in the library `mgcv` in the R programming language. Due to the large size of the data, the `bam` function developed by Wood et al. (2014) will be utilised instead of `gam`. This function has been optimised for large datasets by dividing the design matrix into subblocks and performing parallel computations. The use of `gam` to construct complex models can take up to 10 hours, while the implementation of `bam` can reduce the time to less than an hour.

The application of GAMs in this study entails a set of underlying statistical assumptions that warrant careful consideration. Firstly, it is assumed that the response variable—total electricity demand—follows a Gaussian distribution, and that the identity link function appropriately relates the expected demand to the additive predictor structure.

Secondly, GAMs presuppose that the effects of covariates combine additively, with non-linear relationships adequately captured by smooth functions such as cubic regression splines. This additivity assumption facilitates the decomposition of complex temporal and weather-related influences into interpretable components.

Furthermore, the model assumes that the residuals are independently and identically distributed, with constant variance (homoscedasticity) across the range of fitted values. In the context of time series forecasting, this implies the absence of autocorrelation among errors over time, which is crucial for reliable inference.

Lastly, the validity of GAM estimation also depends on the absence of high-leverage observations or structural outliers that could distort the estimated smooth functions.

These assumptions are not guaranteed to hold in practice and will therefore be empirically assessed through diagnostic checks in the applied sections that follow. Where violations are identified, their implications will be discussed and potential remedies considered.

2.4 Evaluation Metrics

To compare forecast results between two techniques, an appropriate metric is necessary. In this study, out-of-sample evaluation is conducted to assess the predictive performance of each model on unseen data. The models are trained on historical data and evaluated on a holdout test set (see Section 2.5), ensuring that performance is measured based on the ability to forecast future observations rather than merely fitting historical data. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are popular metrics that have been widely used for this purpose. Let $\hat{y}_i$ be the predicted value, y_i the true value at time period i, and n the number of observations. Each metric can be calculated by:

$$\text{MAE} = \sum_{t=1}^{n} \frac{|\hat{y}_t - y_t|}{n},$$

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n} \frac{(\hat{y}_t - y_t)^2}{n}},$$

$$\text{Bias} = \sum_{t=1}^{n} \frac{\hat{y}_t - y_t}{n}.$$

The Mean Absolute Error (MAE) quantifies the average magnitude of errors between predicted and actual values. It treats all errors equally, without penalising larger errors more than smaller ones. In contrast, the Root Mean Square Error (RMSE) is similar to MAE but squares the errors before averaging, giving greater weight to larger errors. RMSE is a more sensitive measure of accuracy and is typically employed when larger errors have a more significant impact on the analysis.

Coverage probability evaluates the performance of statistical intervals, such as confidence or prediction intervals, in capturing the true parameter or future observations. A well-calibrated model will exhibit a balanced relationship between

MAE/RMSE and coverage probability. Specifically, when MAE/RMSE is minimised, the point predictions align more closely with the actual values, leading to an improvement in the coverage probability of the intervals. In contrast, larger MAE/RMSE values indicate that the predictions deviate further from the actual values, potentially resulting in lower coverage probability, as the intervals may fail to effectively capture the true parameter or observations.

When comparing predictive accuracy using different metrics, it can be challenging to determine which model performs better, particularly when the difference between the metrics is small. To address this, a formal statistical test, such as the Diebold-Mariano test, should be used. Developed by Diebold and Mariano (2002), the Diebold-Mariano (DM) test is a statistical method for comparing the accuracy of two competing forecasts under the null hypothesis that no significant difference exists between them.

The Diebold-Mariano test statistic is defined as follows:

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k)/n}}$$

where $\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$ and $d_t = |e_t|^p - |r_t|^p$, with e_t and r_t representing the residuals of the two forecasts, i.e., $y_t - \hat{y}_t$. When $p = 1$, d_t corresponds to the MAE, and when $p = 2$, it is associated with the MSE and RMSE. The autocovariance at lag k, γ_k, is $\frac{1}{n}\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d})$, where h represents the forecast horizon. Under the null hypothesis, the test statistic follows a standard normal distribution, $\mathrm{DM} \sim \mathrm{N}(0, 1)$.

2.5 Study Design

The statistical models are developed by training on data from 2014 to 2017 and tested on data from 2018. A model trained on a single dataset may become overly specialised and perform well only on similar examples. However, the real world is filled with variations and complexities. By using different training sets, we can expose the model to a broader range of data, enabling it to generalise better and perform well on unseen examples. Different models are therefore built on three separate training sets, consisting of four years of training (2014 to 2017), three years of training (2015 to 2017), and two years of training (2016 to 2017), to learn the relevance of older training data. The models' response variable represents total demand and assumes a Gaussian distribution for both the response variable and residuals.

The models are built by forward selection; model complexity is gradually increased by adding significant variables while disregarding non-significant ones at significance level $\alpha = 0.05$. The complexity of variables influencing demand profiles can be categorised based on evidence from previous sections. Day and time are the least complex features because their interpretations are straightforward: demand is higher during weekdays and lower during weekends, with higher demand during working hours or dinner time and lower demand during the night. Next is the group of holiday features, where demand during holidays is lower compared to normal days. The holiday profiles on weekdays exhibit different patterns from normal weekdays or holidays on weekends. Among the variables, weather features stand out as the most complex due to their indirect impact on demand profiles. Their values have varying impacts on the time of day or the season of the year. Understanding these intricate relationships between weather conditions and demand patterns requires a more nuanced analysis. The smooth terms of the following variables, from lower to higher complexity, will be taken into account in the following order during model development:

1. Day-of-week variable
2. Day-of-year variable
3. Half-hour-of-day
4. Interactions between half-hour-of-day and day-of-week
5. Interactions between half-hour-of-day and types of holidays (only England and Wales (1), only Scotland (2), both (3))
6. Interactions between half-hour-of-day and types of vacations (only England and Wales (1), only Scotland (2), both (3))
7. Interactions between half-hour-of-day and types of day before vacation (only England and Wales (1), only Scotland (2), both (3))

8. Interactions between half-hour-of-day and Christmas day

9. Interactions between half-hour-of-day and New Year day

10. Weighted 2 metre temperature (t2m) with different population weights (raw population (RP)/logarithmic transformation (log)/Box-Cox transformation (BC))

11. Weighted weather features corresponding to the previous best population weight (d2m/100w/10w/i10fg/sd/sf (see Table 2.1))

12. All significant weather features from previous model

13. Interactions between half-hour-of-day and weather features (d2m/t2m/100w/ 10w/i10fg/sd/sf (see Table 2.1))

14. Interactions between half-hour-of-day and all significant weather features from previous model

15. Rolling average 2 metre temperature (1 day/2 days/both)

The first four steps represent day and time features, steps 5 to 9 correspond to holiday features, and steps 10 to 15 relate to weather features. The selection of smooth terms in steps 11 and 12 is based on the reference in step 10, whereas the smooth term choices in steps 13 and 14 are based on the references in steps 11 and 12.

In order to determine the optimal variable selection for each model, we will begin with the previous best model and construct models for all potential variable choices. The base model will be built with only one predictor which is the linear trend. To test for significant differences between models, we will use the Diebold-Mariano test at a 95% confidence level, assuming the alternative hypothesis that the new model provides higher accuracy than the previous model. The accuracy of the models will be evaluated using two performance indicators, the root-mean-square error (RMSE) and the mean absolute error (MAE), with $p = 2$ and $p = 1$, respectively, and a forecast horizon of $h = 1$. The criteria for adding variable choices at each complexity level will be based on the p-value of the test. If any metric shows a significant improvement, the variable choice will be added.

2.6 Results

Table 2.3 displays the full results obtained from building models with 4 years, 3 years, and 2 years of training data. The best feature choice in each category is denoted in bold. The percentage bias beneath certain steps is computed based on the best model derived from that particular step. In cases where a step yields more than two top-performing models, the percentage bias is determined by the model with the least bias among that selection. In this section, we will discuss the use of day and time features, holidays, and weather features.

The fitting results demonstrate that the day and time feature (1 to 4) significantly improved the model performance across different sizes of training sets. These features include the dummy day-of-week variable (Sunday to Saturday), continuous day-of-year variable (from 0 to 1), discrete half-hour-of-day variable (1 to 48), and interactions between half-hour-of-day and day-of-week. Some holiday interactions (5 to 9), particularly the interaction between half-hour-of-day and types of holidays, also improved the model performance. However, no other holiday interactions had significant improvements on models.

Additionally, weather features (10 to 14) improved the performance of models, with population weight applied. The log transformation of population provided the best results in the error metrics compared to the other transformations. The 2 metre temperature and the interaction between half-hour-of-day and 2 metre temperature were suggested to be added in smooth terms. The recency effect, a cognitive bias that refers to the phenomenon where people tend to better remember and recall information or events that occurred most recently compared to information encountered earlier (15), also had significant results when adding the rolling temperature into the model. The results in different training sets varied, but each set indicated that a 2-day rolling average temperature must be applied to the model.

When comparing the results of different training sets, it was observed that the model trained for 2 years exhibited the best performance compared to the other two models. RMSE, MAE and bias from the models using 2 years of training outperform the other models while 4 years of training gives a better result than 3 years of training

Distribution family Link function		Gaussian identity					
Training length		4 years		3 years		2 years	
Variable choice		RMSE	MAE	RMSE	MAE	RMSE	MAE
1	Base	7334.84	6143.88	7404.66	6193.72	7401.68	6192.83
	Include	7136.48	5902.72	7206.93	5997.76	7203.83	5996.43
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
2	Include	6315.88	5618.20	6342.01	5679.90	6342.03	5657.58
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
3	Include	2506.37	1868.77	2569.42	1925.64	2567.76	1903.54
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
4	Include	2265.17	1621.99	2333.80	1699.45	2331.22	1657.16
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
5	Include	2232.27	1619.07	2302.15	1694.79	2300.16	1652.33
	p-value	0.00	0.15	0.00	0.04	0.00	0.03
6	Include	2229.53	1619.16	2300.36	1694.21	2298.91	1651.97
	p-value	0.04	0.54	0.13	0.29	0.22	0.35
7	Include	2229.05	1618.20	2301.84	1693.98	2300.12	1651.51
	p-value	0.16	0.07	0.22	0.09	0.47	0.11
8	Include	2226.14	1619.44	2299.03	1694.97	2297.28	1652.77
	p-value	0.15	0.62	0.16	0.58	0.23	0.67
9	Include	2225.66	1619.39	2299.45	1694.96	2297.87	1652.49
	p-value	0.14	0.60	0.19	0.58	0.22	0.58
10	RP	1872.89	1386.46	1926.86	1446.88	1825.78	1320.16
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	log	**1823.17**	**1357.19**	**1869.12**	**1404.98**	**1785.53**	**1295.81**
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	BC	1827.24	1359.56	1874.06	1408.76	1786.19	1296.10
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Percentage Bias	-1.46%		-2.01%		-0.99%	
11	w100	1803.67	1347.91	1855.69	1398.39	1784.03	1293.32
	p-value	0.00	0.00	0.00	0.00	0.05	0.00
	w10	1809.27	1351.36	1861.91	1401.99	1789.62	1297.64
	p-value	0.00	0.00	0.00	0.00	1.00	0.96
	d2m	1760.62	1306.05	1851.90	1405.33	1727.90	1243.28
	p-value	0.00	0.00	0.00	0.53	0.00	0.00
	i10fg	1811.02	1351.10	1865.46	1404.09	1798.96	1300.48
	p-value	0.00	0.00	0.00	0.09	1.00	1.00
	sd	1727.10	1304.11	**1750.89**	**1327.96**	**1701.88**	1244.62
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	sf	1819.84	1354.99	1882.63	1412.21	1820.70	1317.96
	p-value	0.00	0.00	1.00	1.00	1.00	1.00
12	All sig.	**1676.00**	**1248.51**	1757.63	1328.77	1716.00	**1237.41**
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
11-12	Percentage Bias	-1.17%		-1.58%		-0.42%	
13	w100	1668.32	1239.94	1718.56	1299.49	1686.18	1225.04
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	w10	1667.44	1239.20	1726.58	1304.92	1696.70	1232.64
	p-value	0.00	0.00	0.00	0.00	0.01	0.00
	d2m	1570.11	1140.03	1697.46	1272.05	1599.08	1122.12
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	t2m	**1557.83**	**1135.95**	**1612.42**	**1204.67**	**1533.68**	**1074.68**
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	i10fg	1662.55	1235.92	1735.52	1310.89	1713.41	1242.87
	p-value	0.00	0.00	0.00	0.00	1.00	0.24
	sd	1743.00	1268.00	1803.42	1339.10	1746.68	1245.60
	p-value	1.00	1.00	1.00	1.00	1.00	0.59
	sf	1672.99	1243.48	1762.98	1330.65	1748.47	1265.66
	p-value	0.08	0.00	1.00	0.89	1.00	1.00
14	All sig.	1576.61	1149.75	1698.48	1263.41	1587.25	1097.24
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
13-14	Percentage Bias	-0.98%		-1.63%		-0.47%	
15	1 day	1536.46	1107.67	1594.55	1185.79	1507.44	1045.53
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	2 days	**1528.78**	**1104.27**	1566.03	**1158.64**	1480.14	1020.61
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Both	1530.29	1108.29	**1562.66**	1162.07	**1466.52**	**1016.82**
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Percentage Bias	-0.91%		-1.51%		-0.17%	

Table 2.3: The results from building GAM models using a Gaussian family and an identity link function. These models were trained with varying lengths of training data to forecast the demand in 2018. The bold number is the predictor that achieves the best performance within the same level of comparison.

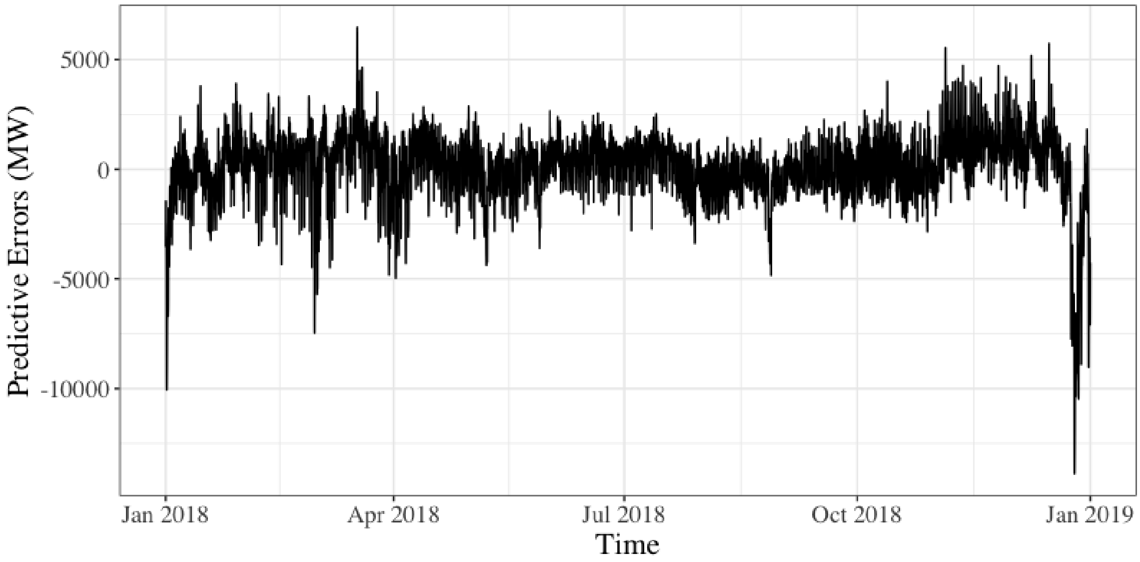

Figure 2.13: Predictive errors for 2018 from the best model with 2 years of training data. The model seems to underpredict the holidays, especially, the New Year and Christmas season.

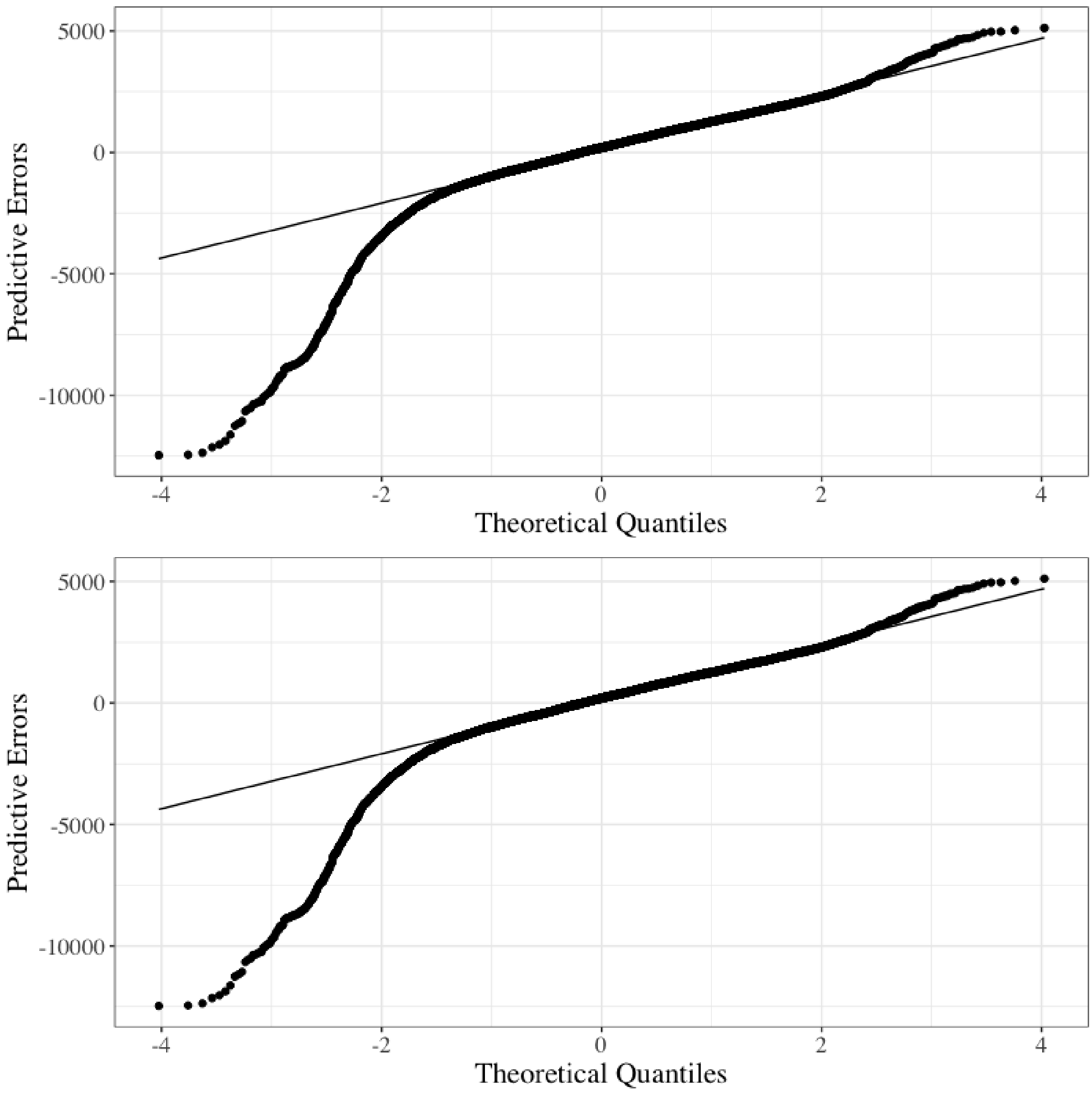

Figure 2.14: QQ-plot for models from 2 years of training of Gaussian family and identity link function (top) and gamma family and logarithm link function (bottom).

when comparing bias. The low bias of the best model using 2 years of training indicates that the model is able to capture the true relationships and patterns present in the data. Upon examining the assumptions of this particular model, a problem was identified related to the residuals during the new year and Christmas season, as depicted in Figure 2.13. Even when this variable is added to the models in terms of vacations, the plot clearly indicates the need to pay additional attention to the New Year and Christmas patterns. Additionally, the QQ-plot for this model, shown in Figure 2.14 (Top), reveals an inappropriate lower left tail for the normal assumption from the model. To try and address this issue, the data were re-modelled with a gamma distribution and logarithmic link function. This distribution with a log link works well for positive-only data with skewed errors. Despite refitting the three different training sets using the same criteria, the issues persisted, as illustrated in Figure 2.14 (Bottom). This issue shows a limitation of our models that cannot capture the demand profiles during Christmas and New Year seasons. One approach to fix the issues is to remove this season (20 December to 5 January) from the data and refit the model. A result of this is shown in Figure 2.15 which shows a substantial improvement in the lower left tail. The predictive results of the best additional models using the gamma and Gaussian models with and without the seasons are shown in Table 2.4. The table clearly indicates that eliminating seasonal factors contributes to enhanced predictive performance. When using a training dataset spanning 2 years, both RMSE and MAE exhibit reductions of roughly 15% and 8%, respectively, while the bias remains under 1%; i.e., the sum of the predictive error is less than 1% of the average true value.

Following model estimation, a comprehensive series of diagnostic checks were conducted to evaluate the validity of the underlying statistical assumptions outlined in Section 2.3. The autocorrelation function (ACF), Figure 2.16, of residuals reveals significant temporal dependence, with autocorrelation coefficients exceeding the 95% confidence bounds at multiple lags, most notably at lags 48 and 96, corresponding to daily and bi-daily cycles respectively (given the 30-minute sampling interval). This pattern indicates that the model fails to adequately capture the inherent temporal structure in electricity demand, suggesting the presence of unmodeled cyclical components.

The distribution of prediction errors, as illustrated in Figure 2.17, appears approximately normal and centred near zero, supporting the assumption of unbiased predictions overall. However, the residual scatter plot in Figure 2.18 reveals a systematic pattern of bias at lower predicted demand levels, where the model tends to underpredict actual demand. While some variation in error magnitude exists across the range of predicted values, the evidence for pronounced heteroscedasticity is limited. Instead, the observed pattern suggests a violation of the zero-mean assumption in the lower demand range, which may indicate structural issues in the model under specific load conditions.

The predicted versus actual demand comparison demonstrates a strong linear relationship with minimal systematic bias, as evidenced by the tight clustering of points along the diagonal in Figure 2.19. Nevertheless, the time series plot of prediction errors (Figure 2.20) reveals temporal clustering of large residuals, with intermittent periods of elevated forecasting errors reaching approximately 10% of actual demand. One particularly notable episode occurs in March 2018, where an extended sequence of positive errors indicates that the model persistently underpredicted demand over several days. This deviation likely reflects an external or structural factor—such as unseasonably cold weather or unexpected shifts in demand—that was not captured by the model inputs. Although such error magnitudes may be acceptable in operational settings, the observed clustering highlights the model's limited capacity to anticipate extreme demand events and reinforces the importance of incorporating mechanisms for detecting and adapting to such anomalies, including changepoint detection or regime-switching approaches.

These diagnostic findings indicate that despite achieving satisfactory predictive performance overall, the model exhibits systematic violations of key statistical assumptions, particularly independence and homoscedasticity. The presence of significant autocorrelation and heteroscedasticity in the residuals suggests that model refinement through alternative error structures, such as GARCH-type models for heteroscedasticity or the incorporation of additional temporal features, may enhance both statistical validity and predictive accuracy.

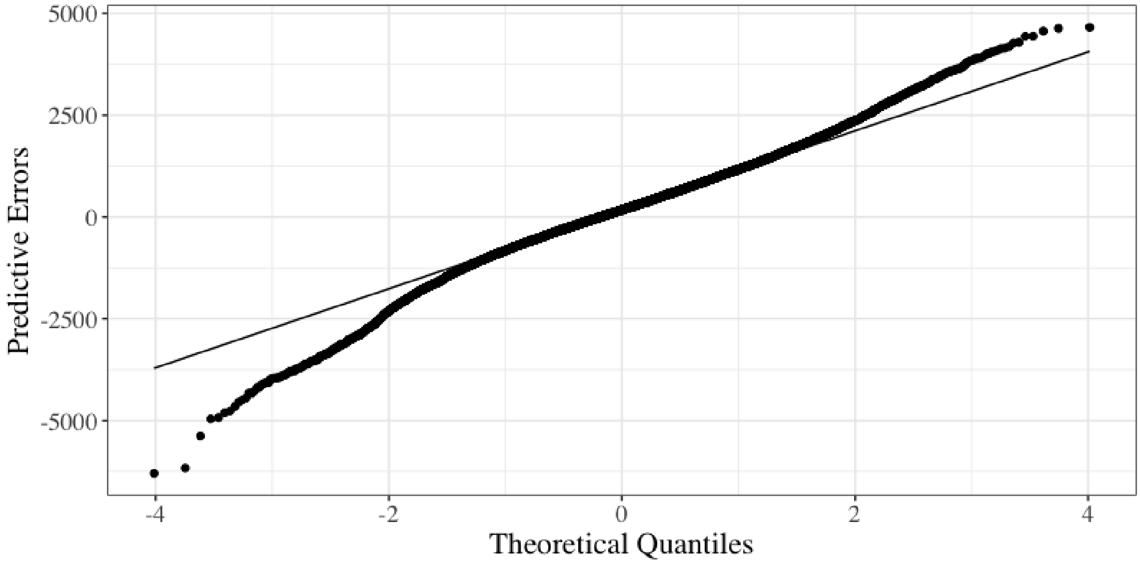

Figure 2.15: QQ-plot for models from 2 years of training after excluding Christmas and New Year seasons of Gaussian family and identity link function.

Data	Including Christmas and New Year seasons		Excluding Christmas and New Year seasons	
Family	Gaussian identity	Gamma log	Gaussian identity	Gamma log
RMSE	1466.52	1485.98	1258.28	1256.69
MAE	1016.82	1010.94	943.42	917.51
Percentage Bias	-0.17%	0.00%	-0.46%	-0.55%

Table 2.4: The final predictive results from building GAM models using a Gaussian family with an identity link function or a Gamma family with a logarithm link function. These models were trained with 2 years of training data to forecast the demand in 2018.

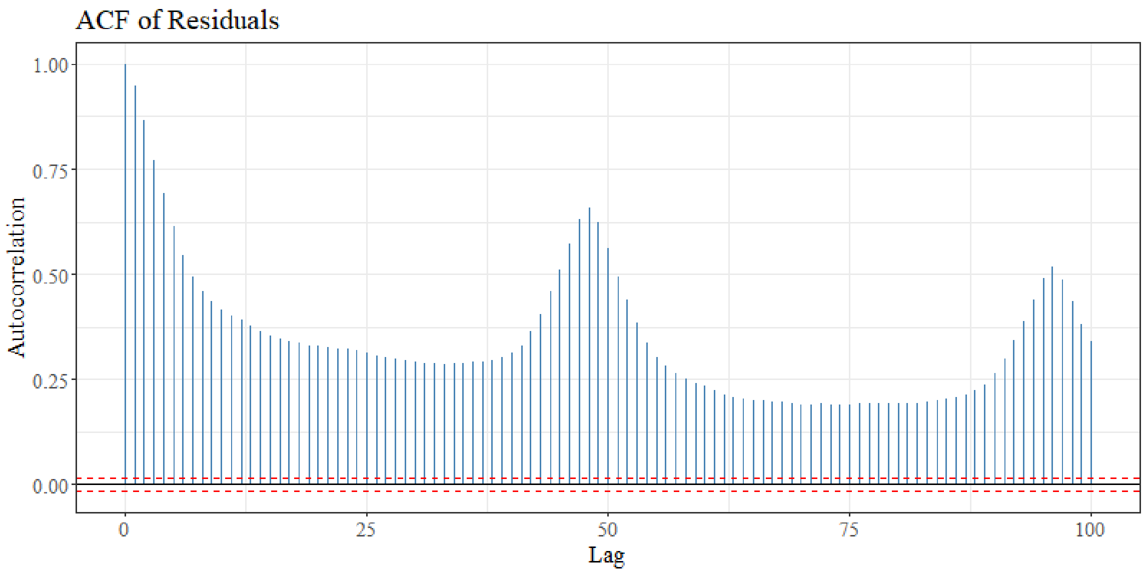

Figure 2.16: ACF plot from the two-year training model (excluding Christmas and New Year), fitted with a Gaussian family and identity link function. Lags 48 and 96 are corresponding to daily and bi-daily cycles respectively

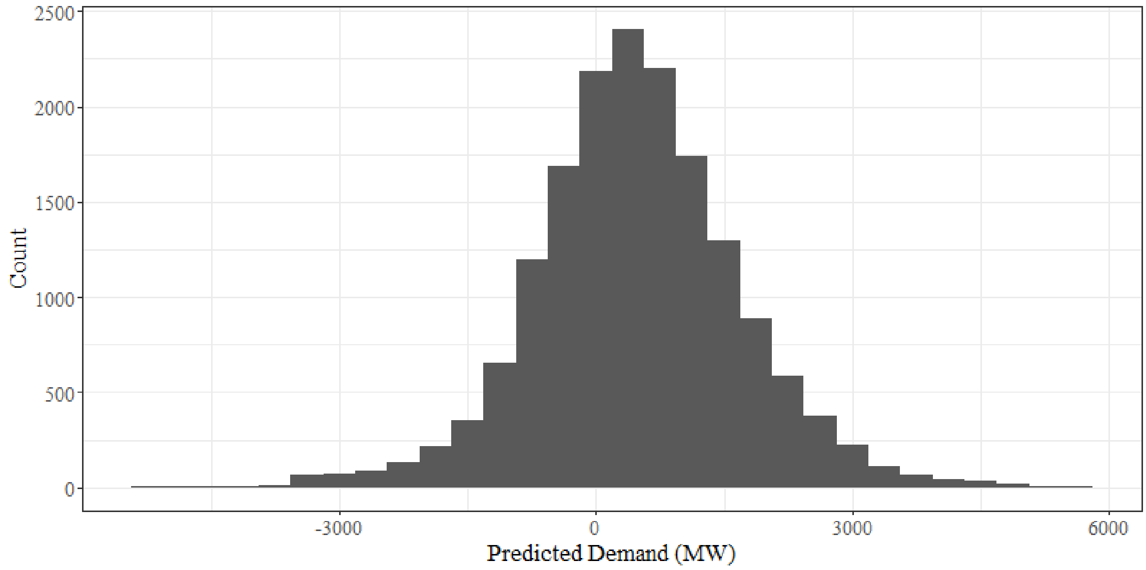

Figure 2.17: Histogram of residuals from the two-year training model (excluding Christmas and New Year), fitted with a Gaussian family and identity link function, showing an approximately normal distribution centered near zero.

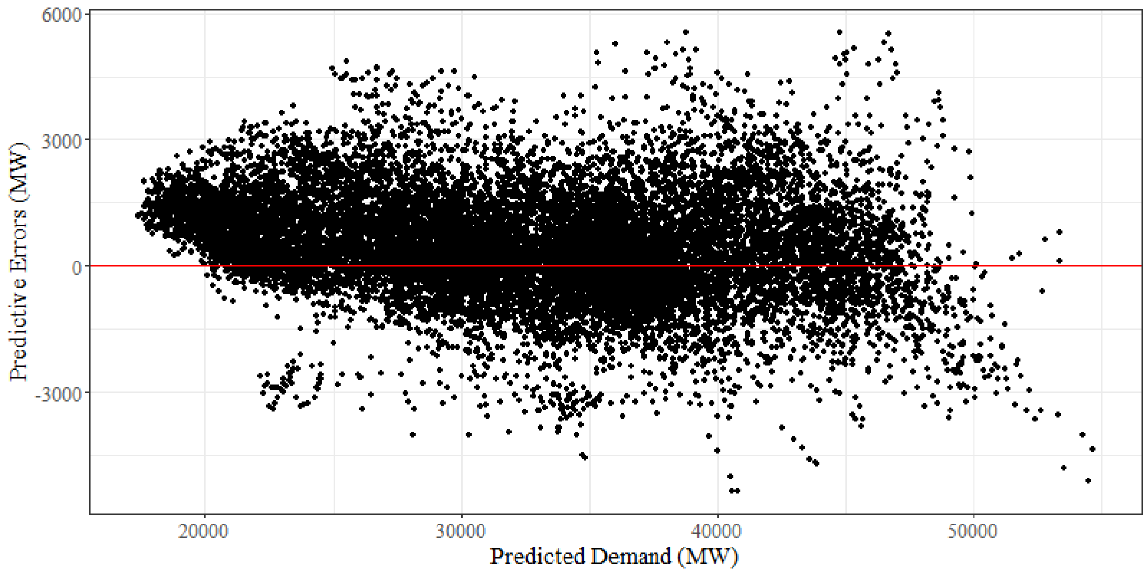

Figure 2.18: Scatter plot of predicted values against residuals from the model trained on two years of data, excluding the Christmas and New Year periods. This plot reveals pronounced heteroscedasticity, with prediction errors demonstrating substantially greater variance at higher predicted demand levels.

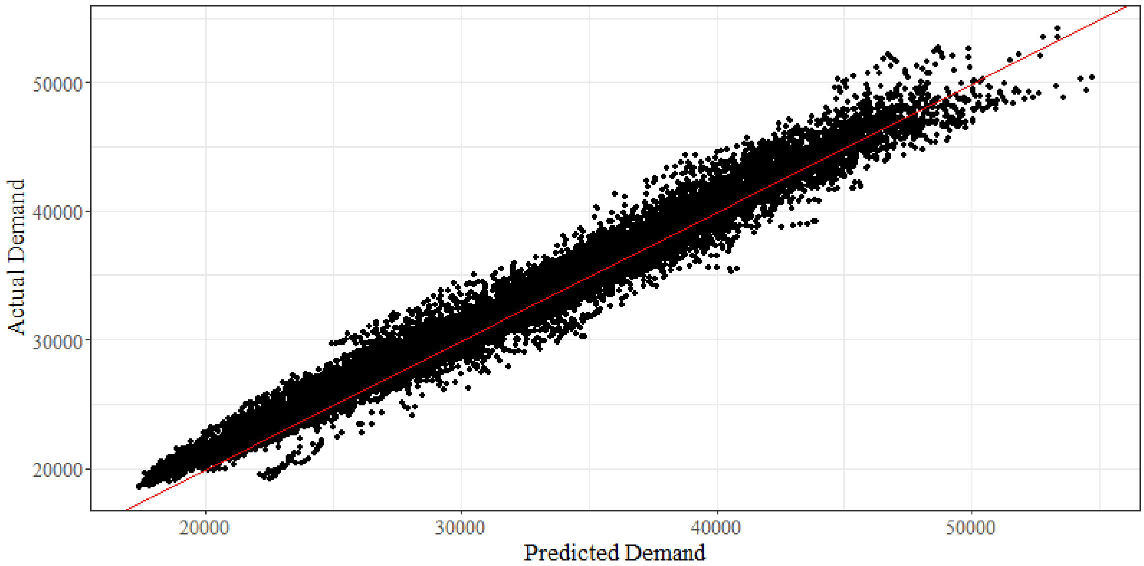

Figure 2.19: Scatter plot of predicted versus actual demand from the model trained on two years of data, excluding the Christmas and New Year periods. This demonstrates a strong linear relationship with minimal systematic bias.

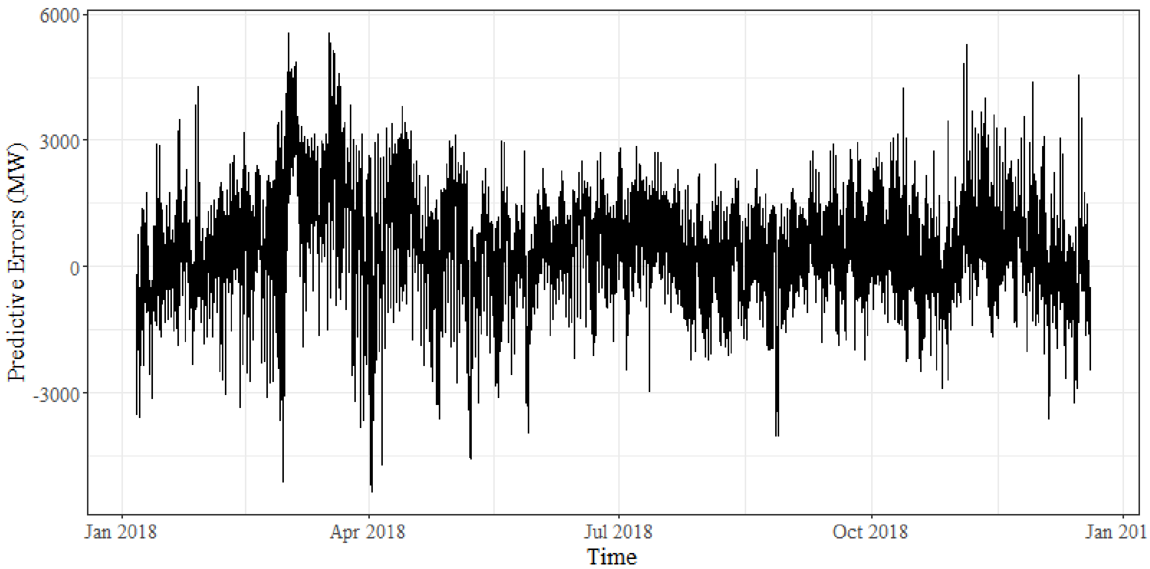

Figure 2.20: Time series of prediction errors from the model trained on two years of data, excluding the Christmas and New Year periods. Several clusters of large residuals are observed, with forecasting errors peaking at approximately 10% of actual demand. A sustained period of underprediction is particularly evident in March 2018, suggesting the presence of unmodelled external or seasonal effects.

2.7 Discussion

The diagnostic checks revealed two primary violations of the model's statistical assumptions: residual autocorrelation and heteroscedasticity. The autocorrelation function (ACF) of residuals exhibited significant coefficients at multiple lags—particularly at lags 48 and 96—corresponding to daily and bi-daily cycles under the half-hourly resolution. This indicates that the model fails to fully capture the cyclical temporal structure inherent in electricity demand. Additionally, the residuals demonstrated non-constant variance, with larger errors associated with higher predicted demand levels, indicating heteroscedasticity.

These findings have important implications for statistical inference and forecast reliability. Autocorrelated residuals suggest that prediction errors are not independent over time, potentially violating the assumptions underpinning the construction of confidence and prediction intervals. As a result, forecast uncertainty may be systematically underestimated, especially during periods of repeated or structured deviation (Wood, 2017). Heteroscedasticity further complicates this picture, as the model appears to generate wider error distributions during peak demand periods—times at which operational accuracy is most critical. This discrepancy may affect the credibility of forecasts during high-load intervals, where decision-making risks are elevated.

Several specific approaches can be employed to address the identified autocorrelation issues in GAM frameworks for electricity demand forecasting. The most straightforward method involves incorporating autoregressive components into the model structure. Marcjasz et al. (2019) demonstrated that GAMs enhanced with autoregressive post-processing significantly improve forecasting accuracy for electricity load data by capturing residual temporal dependencies. This can be implemented by adding lagged demand terms as additional predictors in the form $y_t = f(X_t) + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \varepsilon_t$, or by including smooth functions of lagged residuals to capture non-linear temporal patterns.

An alternative approach involves extending GAMs to Generalized Additive Mixed Models (GAMMs) that explicitly model correlation structures. Wood (2017) and Laurinec (2017) recommend incorporating AR(1) or AR(p) correlation structures of the form $\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$ where $\eta_t \sim N(0, \sigma^2)$. This approach can utilize continuous-time autoregressive processes for irregular time intervals and can be implemented through the correlation parameter in statistical software packages. Additionally, Simpson (2018) suggests that improving model specification can substantially reduce autocorrelation by better capturing underlying patterns through higher-order temporal interactions such as hour-of-day $\times$ day-of-week $\times$ month interactions, smooth functions of recent temperature history with varying lag periods, and cyclical basis functions to capture multi-seasonal patterns.

The heteroscedasticity observed in the residuals requires specific volatility modeling approaches commonly used in electricity demand forecasting. Koopman et al. (2007) and Janczura et al. (2013) demonstrate that GARCH models effectively handle time-varying variance in electricity markets. For demand forecasting, this involves implementing two-stage modeling by first fitting the GAM for the conditional mean, then applying GARCH(p,q) to the residuals using the conditional variance specification $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$. Regime-switching GARCH models may also be considered to account for different volatility patterns during peak versus off-peak periods.

Wood et al. (2016) introduced location-scale GAMs that simultaneously model both mean and variance, allowing variance to be modeled as a function of predictors through $\log(\sigma_i^2) = X_i\beta + Z_i b$. This approach enables smooth functions for variance that depend on time of day and demand level, providing a more integrated solution within the GAM framework. Furthermore, Gneiting (2011) recommends using robust error distributions for electricity forecasting applications, including t-distributions or skew-normal distributions instead of Gaussian assumptions, quantile regression approaches to model prediction intervals directly, and transformation techniques such as Box-Cox transformations to stabilize variance.

Additional model enhancement strategies may address the underlying structural issues identified in the residual analysis. Hamilton (1989) and Janczura and Weron (2010) show that electricity demand often exhibits regime-switching behavior, which can be addressed through Hidden Markov Models (HMMs) combined with GAMs to identify different demand regimes, threshold autoregressive models that switch between different model specifications based on demand levels, and change-point detection algorithms. Marcjasz et al. (2020) demonstrate that ensemble approaches combining multiple GAM specifications can improve robustness through model averaging techniques weighted by historical performance and adaptive model selection based on recent forecast accuracy.

The incorporation of additional external factors may also improve model performance and reduce assumption violations. Taylor (2010) emphasizes the importance of including economic indicators such as industrial production and GDP growth for longer-term patterns, social media sentiment or search trends data for behavioral shifts, and real-time weather forecast uncertainties rather than point estimates. These enhancements may help capture the exogenous shocks and behavioral changes that contribute to the observed residual patterns.

The residual patterns suggest that these assumption violations are not uniformly distributed across time. Instead, they appear concentrated around specific dates and demand regimes, such as holiday periods and seasonal peaks. These intervals likely involve atypical human behavior or exogenous shocks—such as changes in industrial activity, school closures, or extreme weather—not fully accounted for by the covariates included in the current model. Thus, the performance degradation is not a general failure of the GAM framework but rather a limitation in adapting to structurally distinct events.

Despite these limitations, the Generalized Additive Model remains a flexible and interpretable framework for modeling electricity demand (Wood, 2017). Its capacity to incorporate smooth temporal and weather-related effects without rigid parametric assumptions makes it particularly attractive for exploratory forecasting. While it may not fully capture all structural anomalies, its performance on regular days is strong, and the model serves as a reliable foundation for further extension. The identified limitations motivate a more targeted investigation of structural shifts in demand, which is the focus of the following chapter on changepoint detection.

2.8 Conclusion

The development of predictive models for forecasting total electricity demand in Great Britain has proven to be a complex task, reflecting the dynamic nature of consumption driven by temporal, seasonal, and behavioural factors. This chapter has uncovered intricate patterns of electricity usage, underscoring the importance of modelling time-of-day, day-of-week, and seasonal variations. Peak demand typically coincides with periods of heightened human activity—particularly weekday working hours—while demand subsides overnight and on holidays.

The diurnal and weekly demand cycles reveal consistent behavioural routines, while seasonal effects highlight the influence of heating in winter and the relative reduction in demand during summer months. A more granular analysis reveals subtle shifts in daily profiles, such as morning and evening peaks during summer and a pronounced evening peak in winter, further demonstrating the behavioural underpinnings of energy use.

Weather-related variables, particularly temperature and snow depth, emerged as key drivers of demand variability. Incorporating these factors—especially through population-weighted averaging and logarithmic transformation—improved the model's predictive power. The analysis also confirmed the importance of temporal persistence effects, with rolling average temperatures capturing behavioural adaptation to recent conditions.

Model comparison revealed that training on two years of data yielded the best predictive performance. This is likely due to the relevance of recent patterns and reduced overfitting to outdated trends. However, despite strong overall accuracy, residual analysis revealed systematic deviations during holiday periods such as Christmas and New Year. Even with explicit holiday indicators and vacation definitions, the model struggled to capture these exceptional demand behaviours.

Excluding these periods improved residual normality, highlighting the unique structure of holiday demand. Nevertheless, the presence of autocorrelation and heteroscedasticity in the residuals suggests the need for more flexible modelling strategies—particularly those that account for structural shifts in the data.

Building on these findings, the next chapter explores the application of change-point detection algorithms to electricity load data. By identifying and segmenting structural shifts, these methods aim to enhance forecasting accuracy and provide deeper insights into evolving demand behaviour.

Chapter 3

Integrating Changepoint Detection Algorithms with Predictive Models

3.1 Introduction

This chapter focuses on the development and evaluation of changepoint detection techniques specifically designed for time series data, with a particular emphasis on detecting abrupt changes related to mean shifts and seasonal variations in electricity load data. Traditional methods, such as Binary Segmentation and Pruned Exact Linear Time (PELT), excel at identifying sudden shifts in statistical properties but often struggle with datasets containing regular seasonal or trend-driven patterns. To address this limitation, this work enhances changepoint detection algorithms by integrating predictive models that effectively capture underlying trends and seasonality. This integration aims to improve the accuracy of identifying meaningful structural breaks while reducing false detections caused by routine patterns. This chapter first describes existing methods, then introduces a new approach that incorporates predictive models to better manage the complexities of time series data.

3.1.1 Notation

For an ordered sequence of data $y_{1:n} = (y_1, \ldots, y_n)$ of length n. Assume the data contains a number of changepoints, m, with positions, $\tau_{1:m} = (\tau_1, \ldots, \tau_m)$ where

$\tau_i \in \{1, 2, \ldots, n-1\}$ for all $i \in \{1, \ldots, m\}$. We define $\tau_0 = 0$ and $\tau_{m+1} = n$ and assume that the changepoints are ordered such that $\tau_i < \tau_j$ if and only if $i < j$. It follows that these m changepoints split the data into $m+1$ segments where the i-th segment contains $y_{(\tau_{i-1}+1:\tau_i)}$.

3.1.2 Cost Function

A cost function is a real-valued function that represents a measure of goodness of fit. A cost function is commonly used for parameter estimation, where the parameters that produce the best fit are those that minimise the chosen cost function. One commonly used cost function in the changepoint literature is twice the negative log-likelihood (Chen and Gupta, 2000) which is, for example, if y_t is normally distributed with mean μ and common variance σ^2, the cost of the segment $y_{a:b}$ is

$$\mathcal{C}(y_{a:b}) = -2 \log L(y_{a:b}, \mu, \sigma)$$

where

$$L(y_{a:b}, \mu, \sigma) = \frac{1}{(\sqrt{2\pi\sigma^2})^{b-a+1}} \exp(-\frac{1}{2}\sum_{t=a}^{b}((y_t - \mu)/\sigma)^2).$$

Other cost functions such as quadratic loss and cumulative sums can also be used, (Inclan and Tiao, 1994); (Rigaill, 2015).

3.1.3 Penalty Term

The introduction of a penalty parameter, denoted as β, serves as a mechanism to mitigate overfitting. Together with cost function $\mathcal{C}$, a common approach to identify m changepoints is to minimise

$$\sum_{i=1}^{m+1}[\mathcal{C}(y_{(\tau_{i-1}+1:\tau_i)})] + \beta m. \tag{3.1}$$

The penalty term controls the trade-off between complexity and fit to training data. It discourages overfitting by penalising an excessive number of model parameters. Overfitting itself arises when a model becomes too specific to the training data, losing its ability to generalise well to new, unseen data points. Through the inclusion of the penalty term βm, the model is nudged towards achieving an equilibrium: adeptly fitting the training data while also reigning in its complexity to ensure robust generalisation. Examples of the penalty parameters β include Akaike's information criterion (AIC; Akaike (1974)) where $\beta = 2p$ and Schwarz information criterion (SIC, or also known as BIC; Schwarz (1978)) where $\beta = p \log n$ and p is the number of additional parameters introduced by adding a changepoint.

3.2 Changepoint Detection Algorithms

3.2.1 At most one change (AMOC)

The AMOC algorithm is designed to detect changepoints where there is at most one change in the underlying statistical properties of the time series data. This algorithm finds a single τ that minimises the expression

$$\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{\tau+1:n})$$

with the additional constraint that

$$\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{\tau+1:n}) + \beta < \mathcal{C}(y_{1:n}), \tag{3.2}$$

where $\mathcal{C}$ is a cost function and β is a penalty term. In practice, we can compute the prospective τ in term of $\underset{1<\tau<n}{\operatorname{argmin}}\,[\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{\tau+1:n})]$. If τ that minimises the minimisation term satisfies (3.2), τ is the changepoint. If it does not, then this algorithm does not detect any changepoints. The algorithm can be described in Algorithm 1.

When a time series presents multiple changepoints, a singular detection approach becomes insufficient. Consequently, we require algorithms capable of identifying multiple changepoints within the series. An approach that is commonly used to identify multiple changepoints is to find the values of m and $\tau_1, \ldots, \tau_m$ that minimise

$$\sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1:\tau_i)})] + \beta f(m), \tag{3.3}$$

Algorithm 1 At Most One Change

Require: A set of data $(y_1, \ldots, y_n)$.
Require: A penalty constant β.
Require: A cost function $\mathcal{C}$.
Set $\tau^* = \underset{1<\tau<n}{\text{argmin}}\, [\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{\tau+1:n})]$.
if $\mathcal{C}(y_{1:\tau^*}) + \mathcal{C}(y_{\tau^*+1:n}) + \beta < \mathcal{C}(y_{1:n})$ **then**
  A changepoint is detected at τ^*.
else
  No changepoint is detected. NULL is returned.
end if

where $\mathcal{C}$ is a cost function and $\beta f(m)$ is a penalty to prevent overfitting. The most common choice is one that is linear in the number of changepoints; $f(m) = m$. Given that the number of changepoints is known, say m, then the number of possible combinations is $\binom{n}{m}$. In the worst case, when the number of true changepoints, m, is unknown, there are 2^n possible combinations in total to calculate the minimisation of (3.3). Obviously, evaluating every possible combination is impractical.

This challenge constitutes a complex problem, which has been addressed through various methodologies within the literature.

3.2.2 Binary Segmentation

Binary segmentation, or BS, is arguably the most established search method used within the changepoint literature. This algorithm extends a single changepoint method into multiple changepoints by repeating a single changepoint method on different subsets of the sequence. It begins by initially testing if a τ that satisfies (3.2) exists. If none exists, then no changepoint is detected. Otherwise, assuming that $\tau^{(0)}$ is selected by the maximum likelihood estimate, the data will be split into two segments; $y_{1:\tau^{(0)}}$ and $y_{\tau^{(0)}:n}$. The same method will then be used iteratively to both segments until no further changepoints are detected. The advantage of this method is the computation cost, which is $\mathcal{O}(n \log n)$, but with a limitation that is not guaranteed to find the global minimum of (3.3). The algorithm is described in Algorithm 2.

Algorithm 2 Binary Segmentation

Require: A set of data $(y_1, \dots, y_n)$.
Require: A penalty constant β.
Require: A cost function $\mathcal{C}$.
Require: A maximum number of changepoints Q.
Let $\mathcal{T} = \emptyset$ be the set of changepoints.
Set $\mathcal{S} = \{[1, n]\}$.
while $|\mathcal{T}| < Q$ and $\mathcal{S} \neq \emptyset$ **do**
Set $[a', b'] = \underset{[a,b]\in\mathcal{S}}{\text{argmin}} \left[\mathcal{C}(y_{(1:(a-1))}) + \mathcal{C}(y_{a:\tau^{[a,b]}}) + \mathcal{C}(y_{\tau^{[a,b]}+1:b}) + \mathcal{C}(y_{((b+1):n)})\right]$
where $\tau^{[a,b]}$ is calculated using AMOC with a penalty β on $y_{a:b}$.
if $\tau^{[a,b]}$ =NULL for all $[a, b] \in \mathcal{S}$ **then**
No further changepoint will be detected. Set $\mathcal{S} = \emptyset$.
else
Add $\tau^{(0)} = \tau^{[a',b']}$ to $\mathcal{T}$
Add $[a, \tau^{[a,b]}]$ and $[\tau^{[a,b]} + 1, b]$ to $\mathcal{S}$.
Remove $[a, b]$ from $\mathcal{S}$.
end if
end while
Return $\mathcal{T}$ as the output of the algorithm.

3.2.3 Optimal Partitioning (OP)

This method aims to find the values of m and $\tau_1, \dots, \tau_m$ that minimise

$$\sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1:\tau_i)}) + \beta] \tag{3.4}$$

which is equivalent to (3.3) where $f(m) = m$. Let $s \in \{1, 2, \dots, n\}$ be an index and let us consider a set of the first s sequence $y_{1:s}$. Denote the set of possible vector of changepoints for $y_{1:s}$ by $\mathcal{T}_s = \{\boldsymbol{\tau} \mid 0 = \tau_0 < \tau_1 < \cdots < \tau_m < \tau_{m+1} = s\}$ and let

$$F(s) = \min_{\boldsymbol{\tau}\in\mathcal{T}_s} \left\{ \sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1:\tau_i)}) + \beta] \right\}$$

be the minimisation of (3.4) for $y_{1:s}$.

Following Jackson et al. (2005), the OP method begins by first conditioning on the last point of change and it follows that

$$\begin{aligned} F(s) &= \min_{\boldsymbol{\tau} \in \mathcal{T}_s} \left\{ \sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1:\tau_i)}) + \beta] \right\} \\ &= \min_t \left\{ \min_{\boldsymbol{\tau} \in \mathcal{T}_t} \sum_{i=1}^{m} [\mathcal{C}(y_{(\tau_{i-1}+1:\tau_i)}) + \beta] + \mathcal{C}(y_{(t+1):n}) + \beta \right\} \\ &= \min_t \left\{ F(t) + \mathcal{C}(y_{(t+1):n}) + \beta \right\} \end{aligned}$$

when $F(0) = -\beta$. This gives a recursion that provides the minimal cost for $y_{1:s}$ in terms of the minimal cost for $y_{1:t}$ where $t < s$. The recursion can be solved in turn for $s = 1, 2, \ldots, n$ with the computation cost $\mathcal{O}(n^2)$. The pseudocode can be described in Algorithm 3.

Algorithm 3 Optimal Partitioning

Require: A set of data $(y_1, \ldots, y_n)$.
Require: A penalty constant β.
Require: A cost function $\mathcal{C}$.
  Set $F(0) = -\beta$.
  Set $cp(0) = \emptyset$, where $cp(m)$ is the set of changepoints for $(y_1, \ldots, y_m)$.
  for $\tau^* = 1, \ldots, n$ **do**
    Calculate $F(\tau^*) = \min_{0 \le \tau < \tau^*} [F(\tau) + \mathcal{C}(y_{(\tau+1):\tau^*}) + \beta]$.
    Let $\tau' = \arg \left\{ \min_{0 \le \tau < \tau^*} [F(\tau) + \mathcal{C}(y_{(\tau+1):\tau^*}) + \beta] \right\}$.
    Set $cp(\tau^*) = [cp(\tau'), \tau']$.
  end for
  Changepoints are recorded in $cp(n)$.

3.2.4 Pruned Exact Linear Time (PELT)

This algorithm aims to improve the computational efficiency of the OP method while ensuring that the method finds the global minimum of (3.4). PELT removes those values of τ that can never be minima in each iteration. Killick et al. (2012) proposed Theorem 1 for a reduction process.

Theorem 1 *Assume that introducing a changepoint into a sequence of observations reduces the cost, $\mathcal{C}$, of the sequence. More formally, assume that there exists a constant K such that*

$$\mathcal{C}(y_{t+1:s}) + \mathcal{C}(y_{s+1:T}) + K < \mathcal{C}(y_{t+1:T}), \tag{3.5}$$

for all $t < s < T$*. Then if*

$$F(t) + \mathcal{C}(y_{(t+1):s}) + K \geq F(s) \tag{3.6}$$

holds, t *can never be the optimal last changepoint prior to* $T > s$*.*

If (3.6) holds, then the optimal segmentation with the most recent changepoint prior to T being at s will be more optimal than any that has this most recent changepoint at t for any $T > s$. The condition in Theorem 1 that discards t from future consideration helps reduce computation cost, as it removes irrelevant candidate changepoints from the calculation. The algorithm is described in Algorithm 4. The worst-case complexity of the algorithm is $\mathcal{O}(n^2)$ when no discarding takes place, and $\mathcal{O}(n)$ under a mild condition.

Algorithm 4 PELT

Require: A set of data $(y_1, \ldots, y_n)$.
Require: A measure of fit $\mathcal{C}$.
Require: A penalty constant β.
Require: A constant K that satisfies (3.5).
  Set $F(0) = -\beta$.
  Set $cp(0) = \emptyset$, where $cp(m)$ is the set of changepoints for $(y_1, \ldots, y_m)$.
  Set $R_1 = \{0\}$.
  for $\tau^* = 1, \ldots, n$ **do**
    Calculate $F(\tau^*) = \min_{\tau \in R_{\tau^*}} [F(\tau) + \mathcal{C}(y_{(\tau+1):\tau^*}) + \beta]$.
    Let $\tau^1 = \arg\left\{\min_{\tau \in R_{\tau^*}} [F(\tau) + \mathcal{C}(y_{(\tau+1):\tau^*}) + \beta]\right\}$.
    Set $cp(\tau^*) = [cp(\tau^1), \tau^1]$.
    Set $R_{\tau^*+1} = \{\tau^* \cup \{\tau \in R_{\tau^*} \mid F(\tau) + \mathcal{C}(y_{\tau+1:\tau^*}) + K < F(\tau^*)\}\}$.
  end for
  Changepoints are recorded in $cp(n)$.

Killick et al. (2012) noted that PELT boasts higher accuracy compared to binary segmentation, whereas binary segmentation demonstrates computational efficiency over PELT. Meanwhile, van den Burg and Williams (2020) introduced numerous time series from various application domains, specifically curated for evaluating these algorithms along with others not covered in this chapter. Surprisingly, the study found that the binary segmentation algorithm outperformed others in terms of accuracy, which contrasts Killick et al. (2012)'s findings. However, this superior performance did not prove statistically significant when compared to the other algorithms tested.

3.3 Evaluation Metrics for Changepoint Detection Algorithms

Evaluation metrics provide a quantitative measure of how well a changepoint detection algorithm performs in identifying true changes in a dataset while minimising false detections. A good evaluation metric helps us understand the strengths and limitations of an algorithm, guiding its selection for specific applications. It also allows for comparisons between different algorithms, aiding in the advancement and refinement of changepoint detection techniques. Examples of metrics for a single changepoint detection algorithm include

1. Bias; $\frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_i - \tau)$, measures how much the average prediction of the algorithm differs from the true changepoint it is trying to predict,

2. Mean absolute error (MAE); $\frac{1}{n}\sum_{i=1}^{n}|\hat{\tau}_i - \tau|$, measures the average size of prediction errors, regardless of their direction. It provides a clear understanding of the typical deviation between predictions and the true changepoint. MAE is straightforward to interpret because it shares the same units as the original data,

3. Median absolute error (MedAE); $\mathrm{Mdn}_{i=1}^{n}(\hat{\tau}_i - \tau)$, measures the central location of prediction errors, irrespective of their direction. It is similar to MAE but provides more reliable measure of central tendency and a robust evaluation when the data is skewed.

4. Root mean squared error (RMSE); $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_i - \tau)^2}$, is similar to MAE but it penalises larger errors more heavily due to the squaring of residuals. RMSE is useful for understanding the spread of errors, as it is on the same scale as the predicted changepoint,

where τ is assumed to be known as the true changepoint, n is the number of simulations that return a predicted changepoint in a simulation study, i.e., excluding the simulations where no changepoint is detected, and $\hat{\tau}_i$ is the changepoint output from the i-th trial that detects a changepoint.

In cases where there is only a single changepoint, τ, within a dataset, the described evaluation metrics can be employed. However, the computation becomes less straightforward when dealing with multiple changes, as it is not always the case that only one changepoint exists in general.

Two primary approaches are used to evaluate the effectiveness of changepoint detection algorithms: clustering-based and classification-based metrics, as discussed by van den Burg and Williams (2020). These approaches do not assume a fixed number of changepoints, allowing for more flexible and realistic assessments of detection performance.

One example of clustering metrics is the covering metric. Let G represent the partition of segments created by true changepoints and S denote the partition of segments created by predicted changepoints from an algorithm. The **covering metric** $C(G, S)$ is defined as follows:

$$C(G, S) = \frac{1}{n} \sum_{A \in G} |A| \cdot \max_{A' \in S} J(A, A')$$

where

$$J(A, A') = \frac{|A \cap A'|}{|A \cup A'|}.$$

It is important to note that G and S need not have the same cardinality. The resulting metric satisfies $0 \leq C(G, S) \leq 1$, where a value closer to 1 indicates superior performance, while proximity to 0 suggests a poorer performance.

An example of classification metrics is the F_1**-score**, which combines both precision, denoted by P, and recall, denoted by R. The F_1-score is defined as:

$$F_1 = \frac{2PR}{P + R}$$

Here,

$$P = \frac{|\mathrm{TP}(T, X)|}{|X|}$$

and

$$R = \frac{|\mathrm{TP}(T, X)|}{|T|}$$

where T is the set of true chagepoints, X is the set of predicted changepoints, and $\mathrm{TP}(T, X) = \{\tau \in T \mid \exists x \in X, |\tau - x| \leq M\}$ be the set of true positives for T compared to predictions X where $M \geq 0$ is the allowed margin of error. Note that T and X need not have the same cardinality.

3.4 Changepoint Detection Algorithm within a Predictive Model

When considering the chosen cost function in terms of negative log-likelihood, it is valuable to understand the different forms of the function based on the type of change one aims to detect. This discussion will first delve into three scenarios: a single mean shift, a single variance shift, and a combined mean-and-variance shift, where multiple shifts can be appropriately applied within the same conceptual framework. Here, the cost function of the data where there is no changepoint is

$$\mathcal{C}(y_{1:n}) = -2 \log L(y_{1:n}, \mu, \sigma)$$

where μ and σ^2 denote the mean and the variance of $y_{1:n}$, respectively. The maximum likelihood estimators (MLEs) of μ and σ^2 are given by $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$, respectively. These estimators are derived under the assumption that the data points y_t are independently and identically distributed (i.i.d.) and follow a normal distribution, $y_t \sim \mathcal{N}(\mu, \sigma^2)$.

For a mean shift, let us assume a shift occurs at timepoint τ. The cost function depends on whether the variance is known or unknown (Chen and Gupta, 2000). In the case where the variance is known, the cost functions of segments before and after the timepoint τ are:

$$\mathcal{C}(y_{1:\tau}) = -2 \log L(y_{1:\tau}, \mu_1, \sigma)$$

and

$$\mathcal{C}(y_{\tau+1:n}) = -2 \log L(y_{\tau+1:n}, \mu_2, \sigma),$$

as stated in Killick et al. (2012)'s changepoint detection package. Here, the variance is assumed to be 1, as this makes no difference in an optimisation problem. The number of additional parameters p to be used in the penalty term is 2—specifically, an extra τ and an extra mean parameter.

On the other hand, if the variance is unknown, the cost functions of segments before and after the timepoint τ are:

$$\mathcal{C}(y_{1:\tau}) = -2\log L(y_{1:\tau}, \mu_1, \sigma)$$

and

$$\mathcal{C}(y_{\tau+1:n}) = -2\log L(y_{\tau+1:n}, \mu_2, \sigma),$$

where μ_1 represents the mean of $y_{1:\tau}$, μ_2 is the mean of $y_{\tau+1:n}$, and

$$\hat{\sigma}^2 = \frac{1}{n}\left[\sum_{t=1}^{\tau}(y_t-\mu_1)^2 + \sum_{t=\tau+1}^{n}(y_t-\mu_2)^2\right].$$

The number of additional parameters p to be used in the penalty term is 3, which are an extra mean parameter, the new variance σ^2, and τ. Note that the differences between the known variance case and the unknown variance case are the estimated variance $\hat{\sigma}^2$ when the variance is unknown and the number of additional parameters, while the estimated means μ_1 and μ_2 are identical between cases.

Now, considering a variance shift where a change occurs at timepoint τ, and we assume no change in mean (meaning the mean remains constant across the data equal to μ). Let σ_1^2 represent the variance of $y_{1:\tau}$, and σ_2^2 denote the variance of $y_{\tau+1:n}$. The cost functions of segments before and after the timepoint τ are

$$\mathcal{C}(y_{1:\tau}) = -2\log L(y_{1:\tau}, \mu, \sigma_1)$$

and

$$\mathcal{C}(y_{\tau+1:n}) = -2\log L(y_{\tau+1:n}, \mu, \sigma_2)$$

where

$$\hat{\sigma}_1^2 = \frac{\sum_{i=t}^{\tau}(y_t-\mu)^2}{\tau}, \quad \hat{\sigma}_2^2 = \frac{\sum_{t=\tau+1}^{n}(y_t-\mu)^2}{n-\tau}$$

represent the variance of $y_{1:\tau}$ and $y_{\tau+1:n}$, respectively. In this scenario, the number of additional parameters p contributing to the penalty term is 2—an additional σ and τ.

For a combined mean-and-variance shift occurring at timepoint τ, let $y_{1:\tau}$ have mean μ_1 and variance σ_1^2, $y_{\tau+1:n}$ have mean μ_2 and variance σ_2^2. The cost functions of segments before and after the timepoint τ are

$$\mathcal{C}(y_{1:\tau}) = -2\log L(y_{1:\tau}, \mu_1, \sigma_1)$$

and

$$\mathcal{C}(y_{\tau+1:n}) = -2\log L(y_{\tau+1:n}, \mu_2, \sigma_2)$$

where

$$\hat{\sigma}_1^2 = \frac{\sum_{t=1}^{\tau}(y_t - \mu_1)^2}{\tau}, \quad \hat{\sigma}_2^2 = \frac{\sum_{t=\tau+1}^{n}(y_t - \mu_2)^2}{n-\tau}$$

represent the variance of $y_{1:\tau}$ and $y_{\tau+1:n}$, respectively. In this case, the number of additional parameters p contributing to the penalty term is 3—an extra mean parameter, σ, and τ.

There are additional options available for defining a cost function, one of which is going to be a novel method, which involves constructing a selected statistical model aimed at detecting structural breaks. Let $\hat{y}_t^{(a,b)}$ represent the predicted value of y_t obtained from a predictive model built using $y_{a:b}$. Continuing with the cost function $\mathcal{C}$ as twice the negative log-likelihood, $-2\ell_{a:b}$. As the likelihood function represents the probability of observing the given data under a specific statistical model, the formula for the likelihood function varies depending on the distribution assumed for the data. For Gaussian family and linear regression model case, the log-likelihood is defined as follows:

$$\ell_{a:b} = -\frac{b-a+1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2.$$

Here $\hat{\sigma}^2 = \frac{1}{b-a+1}\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2$ is the estimated variance using maximum likelihood method. Simplifying, we find:

$$\ell_{a:b} = -\frac{1}{2}\left((b-a+1)\log\left(\frac{2\pi}{b-a+1}\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2\right) + (b-a+1)\right),$$

leading to the expression for the cost function:

$$\mathcal{C}(y_{a:b}) = (b-a+1)\log\left(\frac{2\pi}{b-a+1}\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2\right) + (b-a+1). \tag{3.7}$$

It is important to note that the term $\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2$ is equivalent to the residual sum of squares (RSS) of the model fitted to the data segment $y_{a:b}$. It is also worth noting that if the length of segment that is used to fit the model, $b-a+1$, is less than the rank of the model, we cannot evaluate the cost function of the segment. This issue can be resolved by considering only segments that are long enough to allow the algorithm to calculate the cost.

In this option, if a linear model has p estimated parameters, the total number of estimated parameters typically becomes $p' = 2p + 1$. This includes $2p$ estimated parameters from the linear models fitting both $y_{1:\tau^\star}$ and $y_{\tau^\star+1:n}$, along with 1 estimated parameter $\tau^\star$. However, there are cases where the rank of the data is not high enough, resulting in the fitted model having a rank lower than p, hence $p' < 2p + 1$ in such instances.

Each predictive model used within the changepoint detection framework is fit using a distinct cost function tailored to optimally estimate model parameters. For example, Generalised Additive Models (GAMs) are fit using a Restricted Maximum Likelihood (REML) approach, which not only estimates model coefficients but also optimally selects smoothing parameters to best capture non-linear trends and seasonal patterns in the data. The REML approach is particularly effective for GAMs as it balances the model's fit with the smoothness of its components, avoiding overfitting while preserving the model's ability to represent complex patterns.

This fitting process differs from the cost function used in the changepoint detection algorithm, which is primarily designed to capture shifts in statistical properties between segments rather than to optimise the smoothness of trends or seasonality.

The changepoint algorithm's cost function, as defined in Equation 3.7, is therefore more straightforward, focusing on segment-level changes in overall structure rather than optimising the predictive precision within each segment. This dual approach allows the changepoint detection algorithm to focus on locating structural breaks efficiently, while the predictive models maintain a high degree of fit for the data's underlying trends and seasonal patterns.

For the novel method, the Binary Segmentation algorithm is applied to this type of cost function. While PELT, an enhanced version of Optimal Partitioning (OP), typically requires partitioning the time series into multiple segments, it becomes computationally intensive when combined with model-based cost functions. A standard cost function, which relies solely on statistical properties like mean and variance, is relatively inexpensive. However, a model-based cost function incurs high computational costs because each partition requires fitting a model to calculate the cost, which becomes impractical when handling large datasets. Repeatedly fitting models across multiple partitions leads to prohibitive computational times.

In contrast, Binary Segmentation is less affected by these computational demands, as it sequentially identifies changepoints and splits the series into smaller segments at each step. This segmentation process reduces the segment size with each split, thereby lowering the computational burden of model fitting. For these reasons, Binary Segmentation with the cost function described in Equation 3.7 is chosen as the novel approach in this study.

3.5 Summary

This chapter introduced an innovative approach to changepoint detection by integrating changepoint algorithms within a predictive model framework for time series data. The primary focus is to address the limitations of traditional changepoint detection methods, which often misidentify seasonal or trend-driven patterns as structural changes. By incorporating predictive models, particularly the Generalised Additive Models (GAMs) introduced in the previous chapter, the proposed methodology aims to capture these regular patterns, enhancing the accuracy and robustness of change-

Algorithm 5 Proposed Algorithm
Binary Segmentation for a predictive model with conditional Gaussian distributions

Require: A set of data $(y_1, \ldots, y_n)$.
Require: A penalty constant β.
Require: A maximum number of changepoints Q.
Require: A predictive model.
Let $\mathcal{T} = \emptyset$ be the set of changepoints.
Set $\mathcal{S} = \{[1, n]\}$.
Set $\mathcal{C}(y_{a:b}) = (b-a+1)\log\left(\frac{2\pi}{b-a+1}\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2\right) + (b-a+1)$ where $\hat{y}_t^{(a,b)}$ is the predicted value of y_t obtained from the predictive model built using $y_{a:b}$.
while $|\mathcal{T}| < Q$ and $\mathcal{S} \neq \emptyset$ **do**
Set $[a', b'] = \underset{[a,b]\in\mathcal{S}}{\operatorname{argmin}}\left[\mathcal{C}(y_{(1:(a-1))}) + \mathcal{C}(y_{a:\tau^{[a,b]}}) + \mathcal{C}(y_{\tau^{[a,b]}+1:b}) + \mathcal{C}(y_{((b+1):n)})\right]$
where $\tau^{[a,b]}$ is calculated using AMOC with a penalty β on $y_{a:b}$.
if $\tau^{[a,b]}$ =NULL for all $[a, b] \in \mathcal{S}$ **then**
No further changepoint will be detected. Set $\mathcal{S} = \emptyset$.
else
Add $\tau^{(0)} = \tau^{[a',b']}$ to $\mathcal{T}$
Add $[a, \tau^{[a,b]}]$ and $[\tau^{[a,b]} + 1, b]$ to $\mathcal{S}$.
Remove $[a, b]$ from $\mathcal{S}$.
end if
end while
Return $\mathcal{T}$ as the output of the algorithm.

point detection in complex time series data, such as electricity load.

The chapter began with a discussion of essential components for changepoint detection: the cost function, which quantifies the quality of a given segmentation, and the penalty term, which prevents overfitting by balancing model complexity with fit quality. Key algorithms for changepoint detection, including At Most One Change (AMOC), Binary Segmentation (BS), Optimal Partitioning (OP), and Pruned Exact Linear Time (PELT), were reviewed. Their strengths and limitations were analysed, particularly in the context of handling cost functions that incorporate model-based components.

Recognising the computational demands associated with fitting models repeatedly across multiple segments, Binary Segmentation was chosen for the novel approach due to its efficiency in processing large datasets. This method reduces segment sizes incrementally, minimising computation costs while retaining accuracy in identifying changepoints. A novel cost function tailored for this method was defined, based on the Gaussian log-likelihood for model segments, ensuring that model fitting aligns with changepoint detection objectives. This combination of Binary Segmentation with the model-based cost function introduced in Equation 3.7 constitutes the core of the new methodology.

The proposed algorithm is summarised in Algorithm 5, establishing a foundation for the following chapters. Chapter 4 will present simulation studies to evaluate the efficiency and accuracy of this novel approach, and Chapter 5 will demonstrate its application on real-world electricity demand data.

Chapter 4

A Simulation Study of Proposed Method

4.1 Introduction

The objective of this simulation study is to evaluate the effectiveness of algorithms when utilising a designed cost function that incorporates model building. By simulating data with characteristics akin to real-world scenarios, we can assess the algorithms' performance and their potential utility with the new cost function. Our goal is to gain insights into the efficacy of these algorithms and identify potential enhancements for their application to real-life datasets.

Let $t \in \{1, \ldots, 2000\}$ and $\epsilon_t \sim N(0, \sigma^2)$. The time series, given by

$$y_t = \sin\left(\frac{2\pi t}{4.8}\right) + \cos\left(\frac{2\pi t}{4.8 \times 7}\right) + \sin\left(\frac{2\pi t}{4.8 \times 365}\right) + \epsilon_t + \mu \cdot I(t > 1000), \quad (4.1)$$

as shown in Figure 4.1, is used in the study for the single-changepoint simulation where the true changepoint is $\tau = 1000$. Here,

$$I(t > K) = \begin{cases} 0 & \text{for } t \leq K \\ 1 & \text{for } t > K \end{cases}.$$

On the other hand, for the multiple changepoints study, rather than exploring vari-

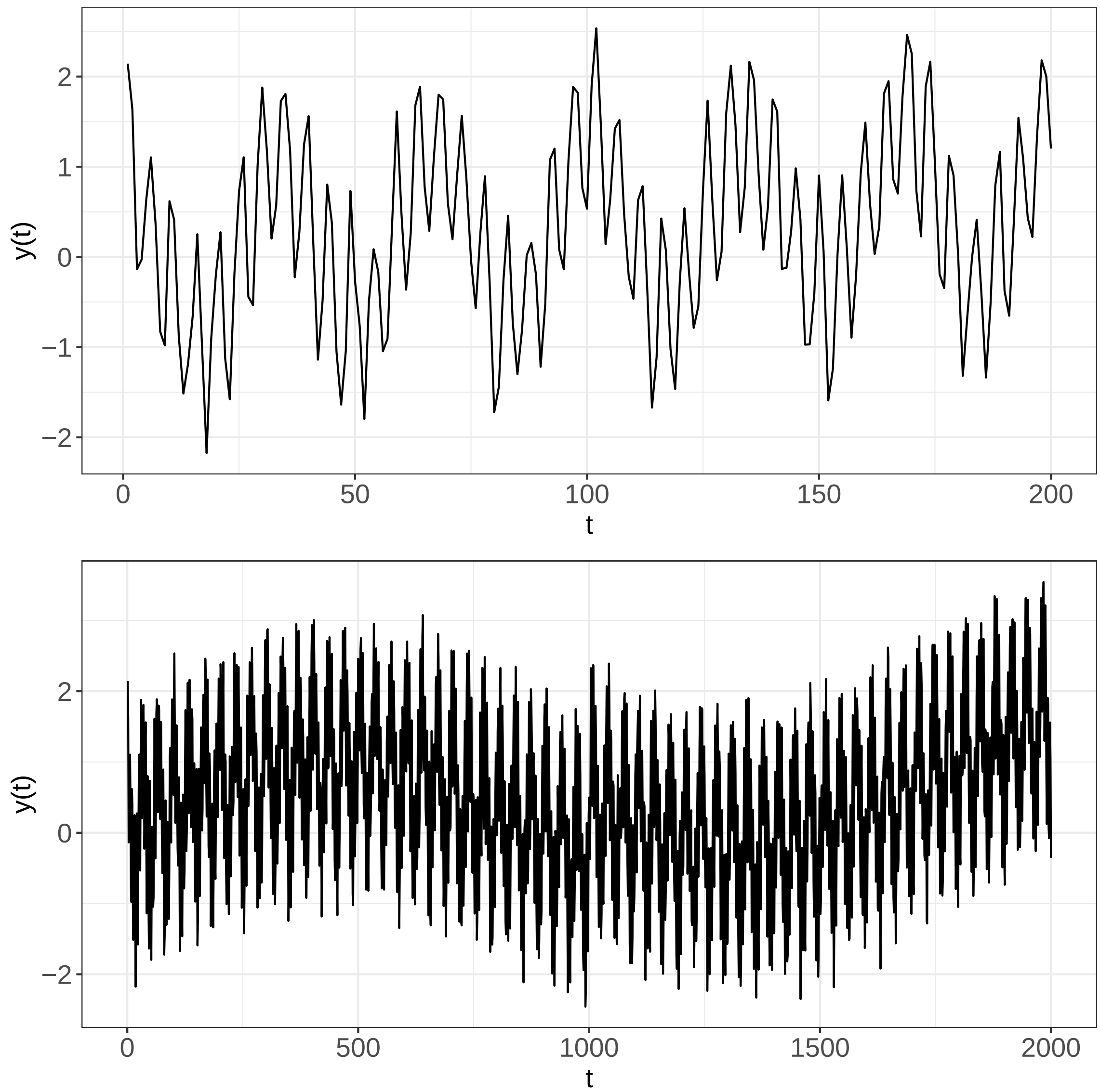

Figure 4.1: The simulation of the time series used in this section employs $\mu = 1.21$ and $\sigma = 0.21$, with features representing daily, weekly, and yearly seasonality applied to the simulation. The true changepoint is $\tau = 1000$. (Top) Shows the first 200 time points where μ has no effect, and (Bottom) displays the entire time series.

ous numbers of changepoints, we narrow our investigation to cases with two changepoints. This decision stems from the understanding that detecting a higher number of changepoints operates similarly but requires multiple iterations. The time series given by

$$y_t = \sin\left(\frac{2\pi t}{4.8}\right) + \cos\left(\frac{2\pi t}{4.8 \times 7}\right) + \sin\left(\frac{2\pi t}{4.8 \times 365}\right) + \epsilon_t + \mu_1 \cdot I(t > 700) + \mu_2 \cdot I(t > 1300), \tag{4.2}$$

as shown in Figure 4.2, is utilised to investigate the algorithm where the true changepoints are $\tau_1 = 700$ and $\tau_2 = 1300$.

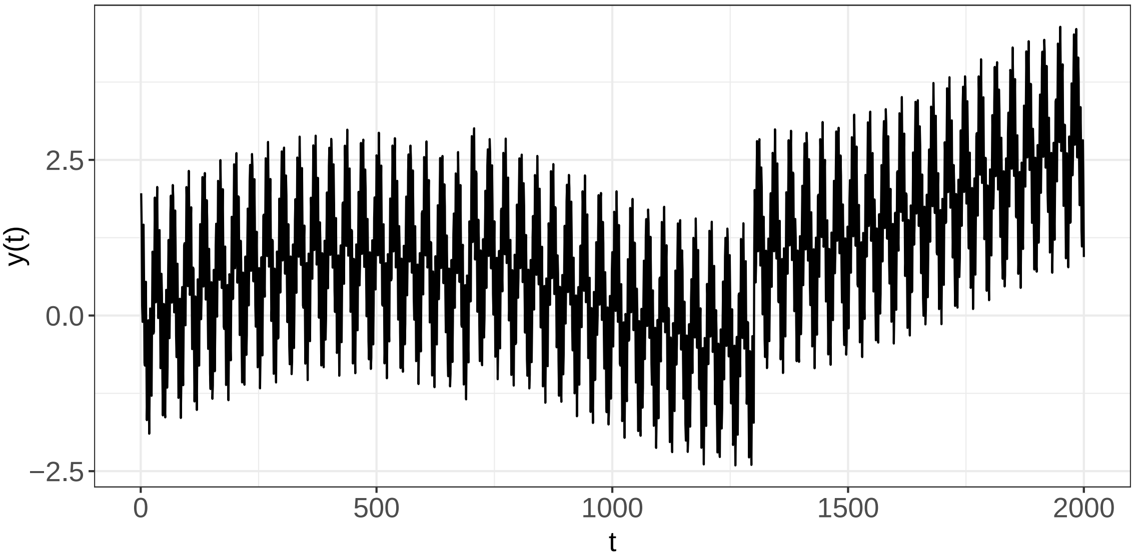

Figure 4.2: The simulation of the time series used in this section employs $\mu_1 = 0.51, \mu_2 = 1.51$ and $\sigma = 0.01$, with features representing daily, weekly, and yearly seasonality applied to the simulation. The true changepoints are $\tau_1 = 700$ and $\tau_2 = 1300$

4.2 Simulation Study Design

To identify changepoints, the At Most One Change (AMOC: 1) and Binary Segmentation (BS: 5) algorithms will be used, with AMOC detecting a single changepoint and BS identifying multiple changepoints.

For a model without smooth terms, this model will be used in a model-building process:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{4.8}\right) + \beta_3 \cos\left(\frac{2\pi t}{4.8}\right) + \beta_4 \sin\left(\frac{2\pi t}{4.8 \times 7}\right) + \beta_5 \cos\left(\frac{2\pi t}{4.8 \times 7}\right) + \beta_6 \sin\left(\frac{2\pi t}{4.8 \times 365}\right) + \beta_7 \cos\left(\frac{2\pi t}{4.8 \times 365}\right) + \epsilon_t \quad (4.3)$$

where $t \in \{1, \ldots, 2000\}$ and $\epsilon_t \sim N(0, \sigma^2)$. This linear model is chosen to fit the simulated data as it captures every characteristic of the dataset. The AIC penalty is applied in the algorithm. On the other hand, for a model formula with smooth terms, the following model formula will be used:

$$g(E(y_t)) = \beta_0 + \beta_1 t + f_1\left(\left\{\frac{t}{4.8}\right\}\right) + f_2\left(\left\{\frac{t}{4.8 \times 7}\right\}\right) + f_3\left(\left\{\frac{t}{4.8 \times 365}\right\}\right) + \epsilon_t, \quad (4.4)$$

where $t \in \{1, \ldots, 2000\}$, $\epsilon_t \sim N(0, \sigma^2)$, g represents a link function, and $\{\cdot\}$ denotes the fractional part function, i.e., $\{x\} = x - \lfloor x \rfloor$. The functions f_1, f_2, and f_3 are smooth functions. In accordance with the optimal choice identified in the previous chapter, we assume that y_t follows a Gaussian distribution with an identity link function, and cubic regression splines are employed to fit the models.

To calculate the cost of a segment $y_{a:b}$, the formula 3.7

$$\mathcal{C}(y_{a:b}) = (b-a+1)\log\left(\frac{2\pi}{b-a+1}\sum_{t=a}^{b}(y_t - \hat{y}_t^{(a,b)})^2\right) + (b-a+1)$$

is used, where $\hat{y}_t^{(a,b)}$ are the fitted values from a model constructed using $y_{a:b}$. Several methods in R can compute these fitted values; some yield more accurate results but may be computationally slower, while faster methods can produce less precise estimations.

There are various methods for calculating the cost of a segment, including approaches that can improve computational efficiency when the cost needs to be calculated repeatedly as segments expand with each iteration. Five approaches for calculating this cost are outlined in Appendix A. However, these alternatives do not outperform the simplest method—repeatedly refitting the model—which consistently provides the most reliable results.

An **undetected changepoint** is defined as a true changepoint that the algorithm **fails to detect** within a small window around its actual location (corresponding to the variable M of $\text{TP}(T, X)$ as described in Section 3.3). In this chapter, the window is set to 5 time steps, which approximately corresponds to one day (with 4.8 time steps equalling exactly one day). Separately, there are cases where the algorithm **fails to return any changepoint** at all, regardless of whether that segment contains a true changepoint or not. This occurs when $\mathcal{C}(y_{1:\tau^*}) + \mathcal{C}(y_{\tau^*+1:n}) + \beta \geq \mathcal{C}(y_{1:n})$ for all $\tau^\star$ within the search range, meaning the cost difference between competing segmentations does not exceed the specified detection threshold for any candidate location.

4.3 Study 1: Model without smooth terms - Single changepoint

The time series in Equation 4.1 is utilised to investigate the algorithm. For this example, the chosen parameters in Equation 4.3 are $\beta_2 = \beta_5 = \beta_6 = 1$ and $\beta_0 = \beta_1 = \beta_3 = \beta_4 = \beta_7 = 0$

The standard deviation σ of the error term and the mean shift μ are selected from the set $0.01, 0.21, \ldots, 2.01$ in each simulation. The time series of each combination will be generated 100 times. A lower variance is expected to yield better predictive performance. When the variance of the random term is high, the expectation is that this term will dominate the data, making it challenging or impossible to identify a correct changepoint. On the other hand, lower values of μ are expected to make it more challenging to identify the true changepoint, while higher values are expected to be estimated more reliably.

We then evaluate the accuracy of the algorithm not only using standard metrics on location accuracy, but also by assessing its ability to detect the correct number of changepoints. This is crucial because identifying the right number of changepoints reflects whether the algorithm captures the true underlying structure of the data. As a real world application, detecting too few changepoints may overlook significant shifts, while detecting too many may indicate overfitting to noise. In simulation studies, where the true number and location of changepoints are known, this provides a clear benchmark for evaluating algorithmic performance and robustness. The results from this section are displayed in the tables, where each metric is presented for every combination of μ (the size of the mean shift) and σ (the standard deviation of the noise).

Figure 4.3 illustrates the frequency of simulations, out of 100, where the algorithm fails to return any changepoint. This occurs when $\mathcal{C}(y_{1:\tau^*}) + \mathcal{C}(y_{\tau^*+1:n}) + \beta \geq \mathcal{C}(y_{1:n})$ for all $\tau^\star$ within the search range, as outlined in Algorithm 1. The results indicate that higher values of σ or lower values of μ pose challenges for the algorithm in detecting changepoints, leading to a higher count of simulations where no changepoint is detected, particularly observed in the bottom-right corner of the table. Conversely, lower σ values or higher μ values facilitate easier identification of changepoints, resulting in fewer instances where the algorithm fails to return a changepoint, notably seen in the top-left corner.

Figure 4.4 presents the average bias observed in the changepoint locations across simulations. The results reveal that lower σ values or higher μ values tend to facilitate the identification of change points with a bias of 0, as notably observed in the upper left corner of the table. Conversely, higher σ values or lower μ values present challenges, resulting in biases that appear to exhibit a random sign. The reasons for the variations in bias, with some combinations showing positive biases and others negative, are not clearly discernible. This variation may be attributed to the stochastic nature of the process, particularly considering the limited number of observations used in calculating the mean.

Figure 4.5 presents the average Mean Absolute Error (MAE) observed in the changepoint locations across simulations, while Figure 4.6 shows the Root Mean Squared Error (RMSE) observed in the changepoint locations across simulations. The results indicate that lower σ values or higher μ values tend to facilitate the identification of changepoints with an MAE of 0, as notably observed in the top-left corner of the MAE table. Conversely, higher σ values or lower μ values present challenges, resulting in a larger value of MAE. This suggests that when identifying the correct changepoint becomes difficult, the algorithm does not select the location randomly but tends to choose locations at one end or the other instead (as the real changepoint is located in the middle). This behavior is influenced by the denominator $b-a+1$ in the cost function. When the segment length is short, $b-a+1$ is low, leading to instability caused by this denominator, which increases the chance that a change point may be incorrectly detected close to one end of the time series. Figure 4.7 shows the Median Absolute Error, supporting the observation that the distribution of predicted changepoints is skewed towards the bottom-left of the table.

After a changepoint is predicted, the RMSE of residuals from models for the two segments of the time series is examined in Figure 4.8. It is observed that the size of RMSE depends primarily on σ rather than on the size of the mean shift μ. This sug-

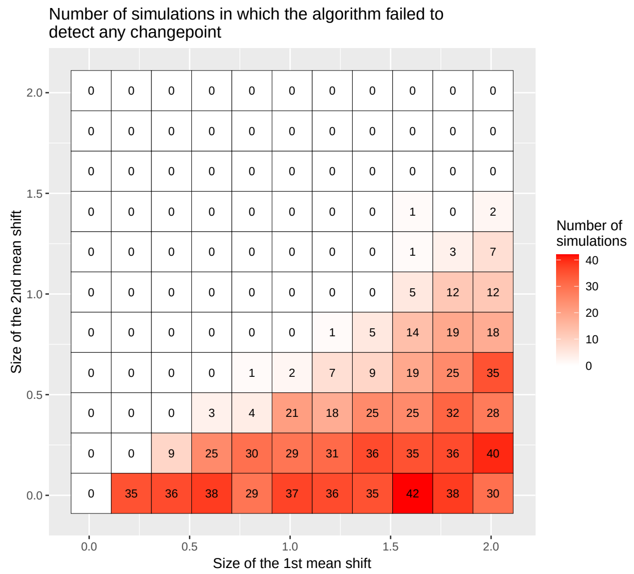

Figure 4.3: Number of simulations (out of 100) in which the algorithm failed to return any changepoint, with σ on the x-axis and μ on the y-axis.

gests that when noise levels are high, prediction error is dominated by the noise itself, making the exact location of detected changepoints relatively unimportant in terms of RMSE. In other words, even if changepoint detection accuracy is compromised in high-noise conditions, the RMSE remains largely unaffected, as the fitting model's performance is constrained by the noise rather than by the changepoint locations. This is because the calculated RMSE reflects the square root of the residual sum of squares over n, which applies when the model fits the data well overall.

In conclusion, when the shift in the data is significantly larger than the noise level, we typically observe a clear indication of where the true change occurs. Conversely, when the shift is relatively small compared to the noise level, identifying the true changepoint becomes challenging, often leading to difficulty in detecting any changepoint using the algorithm. In practical examples where the true location of the changepoint is unknown, this limitation of the algorithm becomes apparent. The algorithm endeavors to compute the cost across all potential possibilities, yet its effectiveness is hindered when the data itself is inherently noisy.

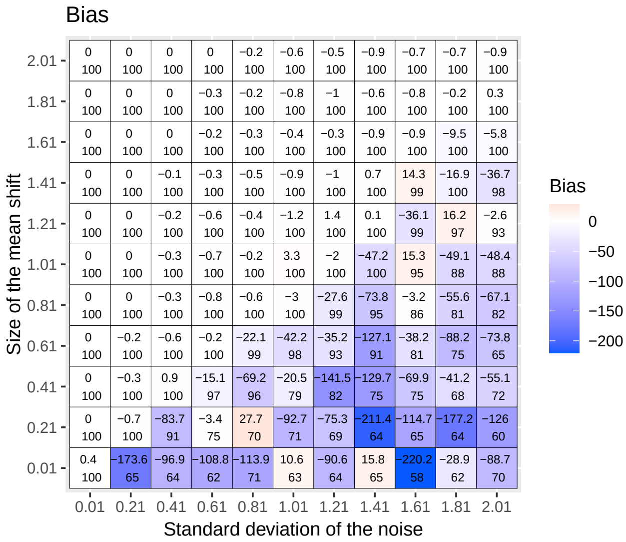

Figure 4.4: The average bias is depicted for 100 simulations with σ on the x-axis and μ on the y-axis. In each grid, the bias is shown at the top, with the corresponding count used to calculate the average displayed at the bottom.

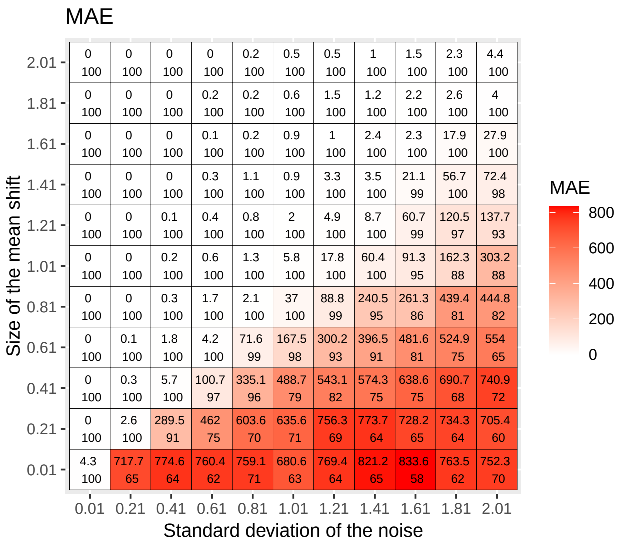

Figure 4.5: The average MAE is depicted for 100 simulations with σ on the x-axis and μ on the y-axis. In each grid, the MAE is shown at the top, with the corresponding count used to calculate the average displayed at the bottom.

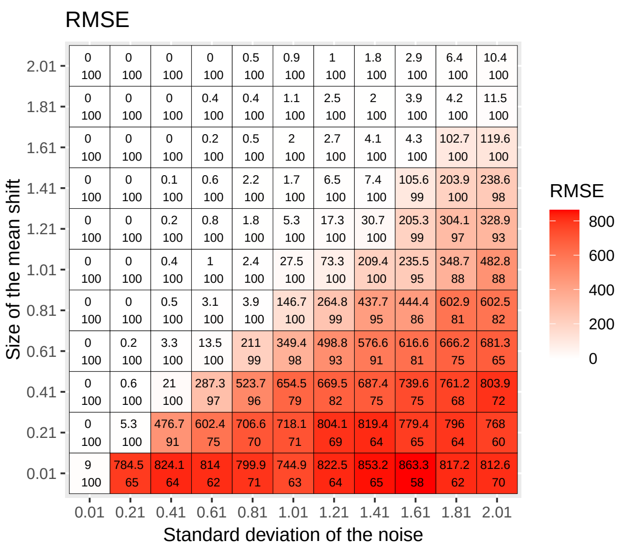

Figure 4.6: RMSE is depicted for 100 simulations with σ on the x-axis and μ on the y-axis. In each grid, the RMSE is shown at the top, with the corresponding count used to calculate the average displayed at the bottom.

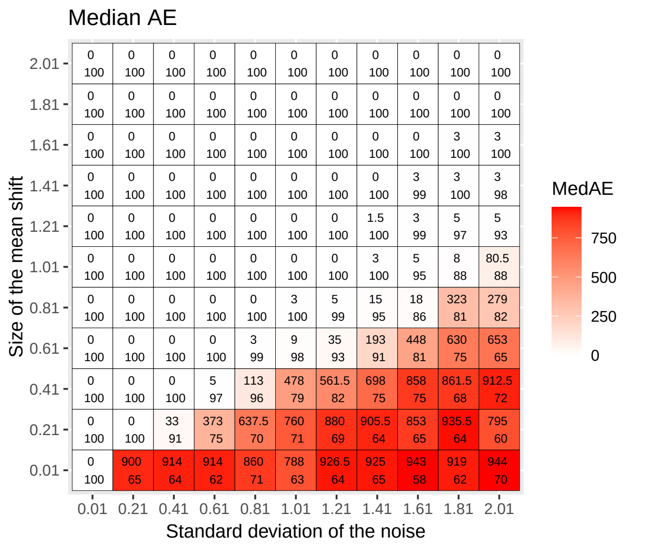

Figure 4.7: Median absolute error is depicted for 100 simulations with σ on the x-axis and μ on the y-axis. In each grid, the median absolute error is shown at the top, with the corresponding count used to calculate the average displayed at the bottom.

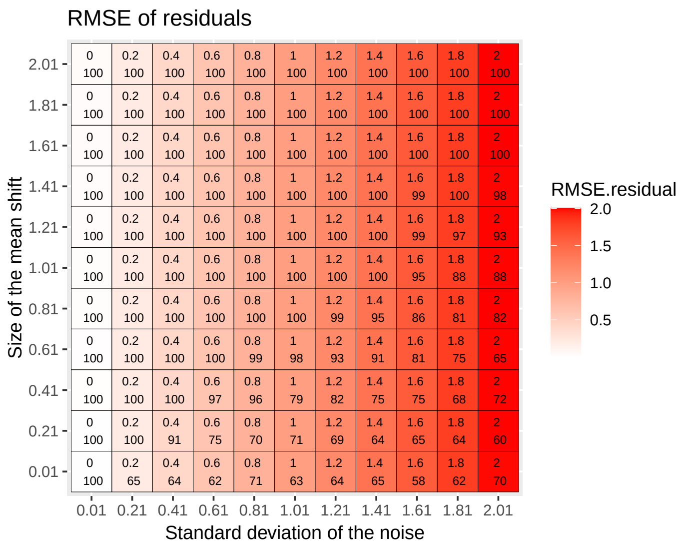

Figure 4.8: The RMSE of residuals is depicted for 100 simulations with σ on the x-axis and μ on the y-axis. In each grid, the RMSE of residuals is shown at the top, with the corresponding count used to calculate the average displayed at the bottom.

4.4 Study 2: Model without smooth terms - Multiple changepoints

In the previous section, we examined the scenario involving a single changepoint within a dataset. Now, our focus shifts to studying multiple changepoints. The time series in Equation 4.2 is utilised to investigate the algorithm. From the model equation 4.4, $\beta_2 = \beta_5 = \beta_6 = 1$ and $\beta_0 = \beta_1 = \beta_3 = \beta_4 = \beta_7 = 0$. The standard deviation σ of the error term is selected from the set $0.01, 1.01, 2.01$, representing low, medium, and high levels of noise, while the sizes of the mean shifts μ_1 and μ_2 are chosen from the set $0.01, 0.21, \ldots, 2.01$ in each simulation study. For each combination of parameters, 50 time series are generated—fewer than the 100 replications used in the previous section. This reduction is due to the higher computational cost associated with the more complex multiple changepoint setting, which makes it infeasible to apply the algorithm 100 times per configuration. Nevertheless, we consider 50 replications sufficient to reveal meaningful insights. The binary segmentation algorithm is employed with a maximum of five changepoints to allow for some flexibility beyond the two true changepoints, using an AIC penalty to guide selection.

Nevertheless, the results are expected to be consistent with those of the previous simulation study. A lower variance is expected to lead to better predictive performance. When the variance of the random term is high, it is anticipated that this term will dominate the data, making it challenging or even impossible to identify the correct changepoints. On the other hand, lower values of μ_1 or μ_2 are expected to make it more difficult to distinguish between segments, while higher values are expected to be estimated more reliably, both in terms of the locations and the numbers of changes.

The results from this section are displayed in the tables, where each metric is presented for every combination of μ_1, μ_2 (the sizes of the mean shifts) and σ (the standard deviation of the noise).

Figure 4.9 illustrates the frequency of simulations, out of 50, where the algorithm fails to return any changepoint. The results indicate that higher values of σ together with lower values of μ_1 or μ_2 pose challenges for the algorithm in detecting changepoints, leading to a higher count of simulations where no changepoint is detected. Conversely, lower σ values and higher μ_1 or μ_2 values facilitate easier identification of changepoints, resulting in fewer instances where the algorithm fails to return a changepoint.

Figure 4.10 presents the covering metric (See 3.3) across simulations. The results show that when the value of σ is low, the algorithm almost identifies changepoint locations correctly. When the value of σ gets larger, it becomes more challenging for the algorithm to identify the true location. Especially when one of the shifts is lower than the other, the algorithm may not be able to detect the lower one as a changepoint, resulting in the detection of only one changepoint. Figure 4.11 shows similar results for the F1-score (See 3.3). However, if a detected changepoint falls outside the error margin, it will not be counted as a true positive. As a result, when there are no true positives, precision and recall will both be 0, rendering the F1-score undefined.

In summary, when a shift in the data significantly exceeds the noise level, we usually observe a clear indication of where the true change occurs. Conversely, when a shift is relatively small compared to the noise level, identifying the true changepoint becomes challenging, often resulting in difficulty detecting any changepoint using the algorithm. As expected, the results are also consistent with those of the previous simulation study.

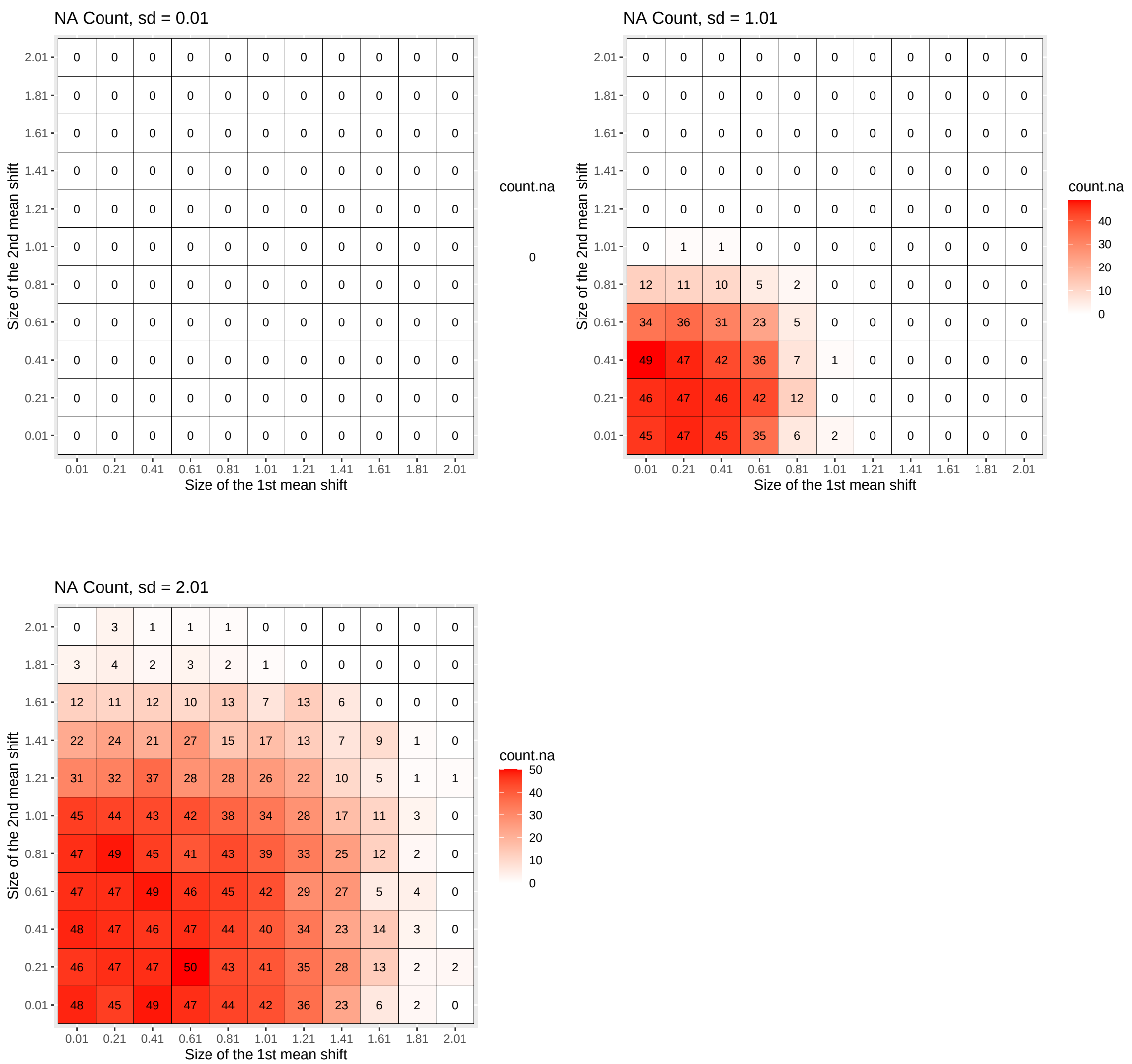

Figure 4.9: Number of simulations (out of 50) in which the algorithm failed to detect any changepoint, across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the top left, σ equals 0.01. In the top right, σ equals 1.01. In the bottom left, σ equals 2.01.

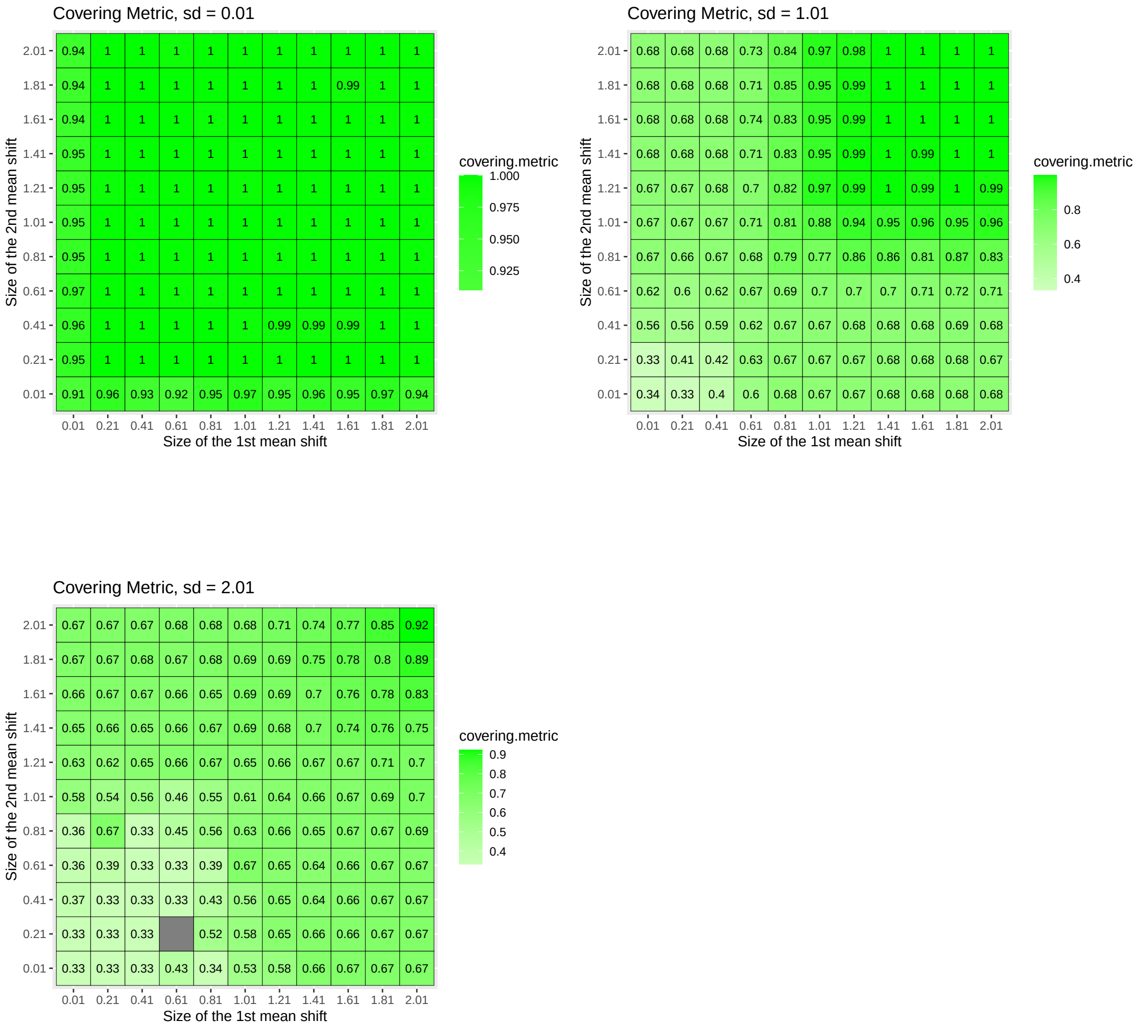

Figure 4.10: Covering metric from 50 simulations across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the top left, σ equals 0.01. In the top right, σ equals 1.01. In the bottom left, σ equals 2.01.

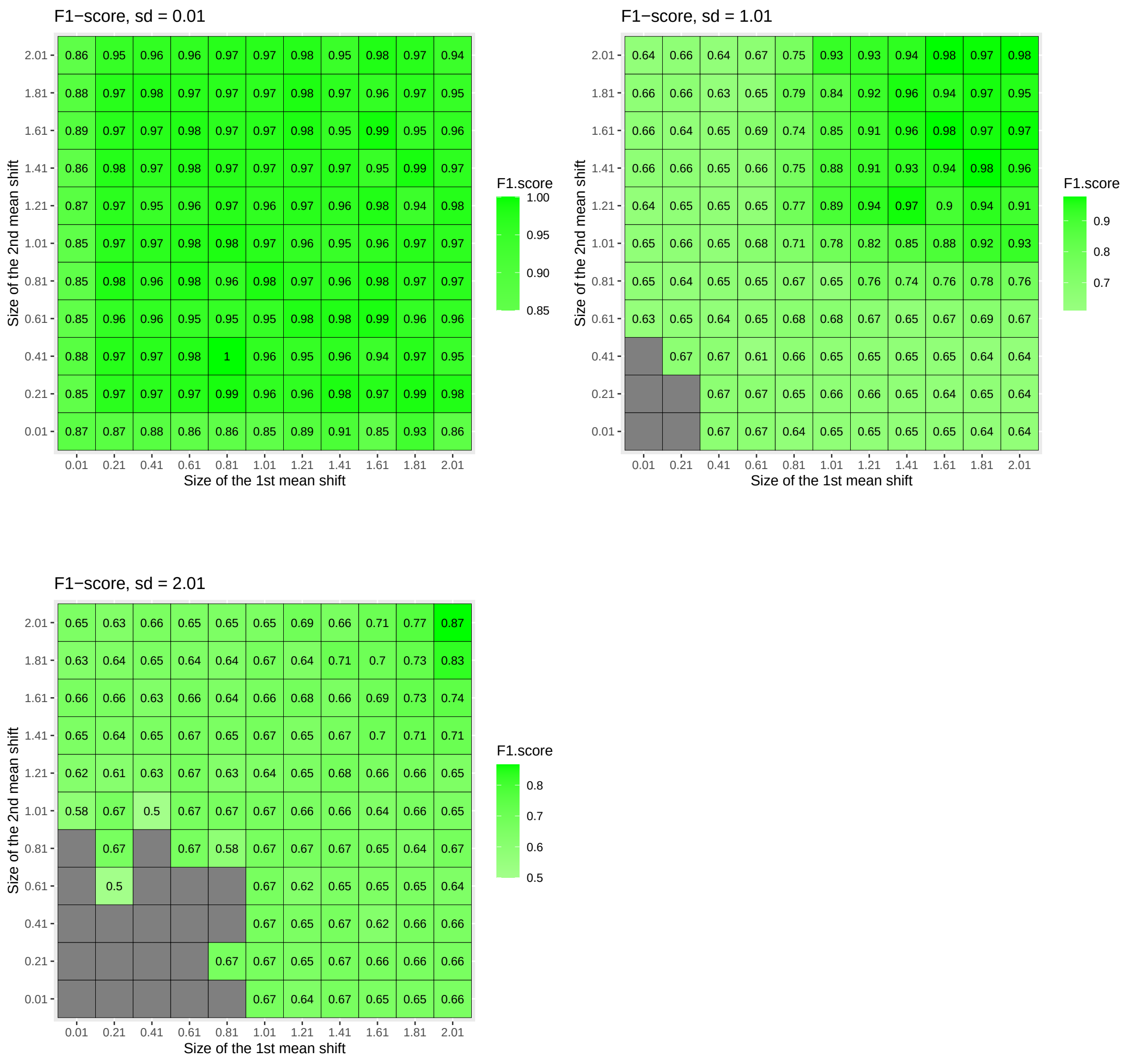

Figure 4.11: F1-score from 50 simulations across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the top left, σ equals 0.01. In the top right, σ equals 1.01. In the bottom left, σ equals 2.01.

4.5 Study 3: Model with smooth terms - Single changepoint

Instead of relying solely on a linear model, the Generalised Additive Model (GAM) offers a more flexible approach to capturing seasonality, particularly beneficial when dealing with real-life data. The time series in Equation 4.1 is utilised to investigate the algorithm with the model equation 4.4 in this example.

The standard deviation σ of the error term and the mean shift μ are selected from the set $0.01, 0.21, \ldots, 2.01$ in each simulation study. The time series of each combination will be generated 100 times. A lower variance is expected to yield better predictive performance. When the variance of the random term is high, the expectation is that this term will dominate the data, making it challenging or impossible to identify a correct changepoint. On the other hand, lower values of μ are expected to make it more challenging to identify the true changepoint, while higher values are expected to be estimated more reliably. Refitting via `gam` function approach (B1) will be used to build a model to obtain the most accurate results.

The results from this section are displayed in the tables, where each metric is presented for every combination of μ (the size of the mean shift) and σ (the standard deviation of the noise).

Figure 4.12 illustrates the frequency of simulations, out of 100, where the algorithm fails to return any changepoint. This occurs when $\mathcal{C}(y_{1:\tau^*}) + \mathcal{C}(y_{\tau^*+1:n}) + \beta \geq \mathcal{C}(y_{1:n})$ for all $\tau^\star$ within the search range, as outlined in Algorithm 1. The results indicate that higher values of σ or lower values of μ pose challenges for the algorithm in detecting changepoints, leading to a higher count of simulations where no changepoint is detected, particularly observed in the bottom-right corner of the table. Conversely, lower σ values or higher μ values facilitate easier identification of changepoints, resulting in fewer instances where the algorithm fails to return a changepoint, notably seen in the top-left corner. However, the inclusion of smooth terms makes changepoint detection more challenging. Their flexibility allows them to absorb gradual changes into the model structure, potentially masking abrupt shifts.

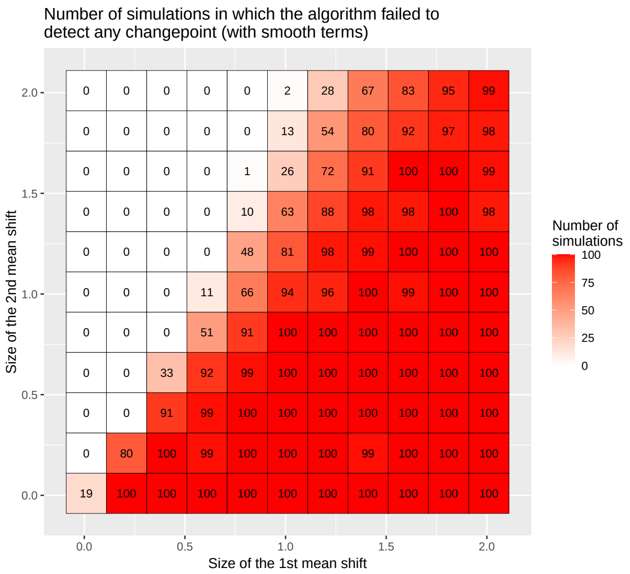

Figure 4.12: Number of simulations (out of 100) in which the algorithm failed to detect any changepoint, with σ on the x-axis and μ on the y-axis.

As a result, detecting changepoints becomes difficult unless the change in mean (μ) is sufficiently large or the noise level (σ) is low enough to make the shift statistically distinguishable.

Figure 4.13 shows the Root Mean Squared Error (RMSE) observed in the changepoint locations across simulations. The results indicate that lower σ values or higher μ values tend to facilitate the identification of changepoints with an RMSE of 0, as notably observed in the top-left corner of the RMSE table. Conversely, higher σ values or lower μ values present challenges, resulting in a larger value of RMSE.

In conclusion, when the shift in the data is significantly larger than the noise level, we typically observe a clear indication of where the true change occurs. Conversely, when the shift is relatively small compared to the noise level, identifying the true changepoint becomes challenging, often leading to difficulty in detecting any changepoint using the algorithm. In practical examples where the true location of the changepoint is unknown, this limitation of the algorithm becomes apparent. The algorithm endeavors to compute the cost across all potential possibilities, yet its effectiveness is hindered when the data itself is inherently noisy.

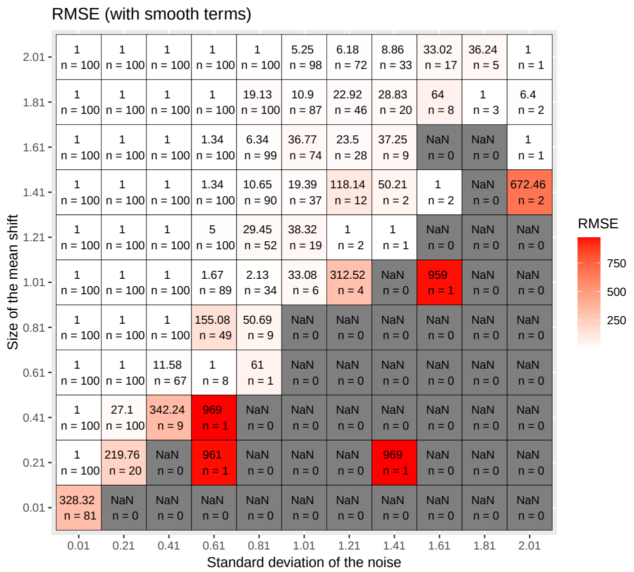

Figure 4.13: RMSE is depicted for 100 simulations with σ on the x-axis and μ on the y-axis. In each grid, the RMSE is shown at the top, with the corresponding count used to calculate the average displayed at the bottom.

4.6 Study 4: Model with smooth terms - Multiple changepoints

In the previous section, we examined the scenario involving a single changepoint within a dataset. Now, our focus shifts to studying multiple changepoints. The time series in Equation 4.2 is utilised to investigate the algorithm with the model equation 4.4 in this example. The standard deviation σ of the error term is selected from the set $0.01, 1.01, 2.01$, representing low, medium, and high levels of noise, while the sizes of the mean shifts μ_1 and μ_2 are chosen from the set $0.01, 0.21, \ldots, 2.01$ in each simulation study. The time series of each combination will be generated 100 times. The binary segmentation algorithm, with a maximum of 10 changepoints, is selected to allow for a few additional changepoints beyond 2, using an AIC penalty.

Nevertheless, the results are expected to be consistent with those of the previous simulation study. A lower variance is expected to lead to better predictive performance. When the variance of the random term is high, it is anticipated that this term will dominate the data, making it challenging or even impossible to identify the correct changepoints. On the other hand, lower values of μ_1 or μ_2 are expected to make it more difficult to distinguish between segments, while higher values are expected to be estimated more reliably, both in terms of the locations and the numbers of changes.

The results from this section are presented in the tables, where each metric is shown for every combination of μ_1 and μ_2 (the sizes of the mean shifts) and two values (0.01 and 1.01) of σ (the standard deviation of the noise), as a value of 2.01 is observed to be too noisy to extract meaningful insights from the simulation.

Figure 4.14 illustrates the frequency of simulations, out of 100, in which the algorithm fails to return any changepoint. The results indicate that higher values of σ combined with lower values of μ_1 or μ_2 pose challenges for the algorithm in detecting changepoints, leading to a higher count of simulations where no changepoint is detected. Conversely, lower values of σ and higher values of μ_1 or μ_2 facilitate the easier identification of changepoints, resulting in fewer instances where the algorithm fails to return a changepoint. On the other hand, Figure 4.15 shows the number of changepoints returned by the algorithm in simulations where at least one changepoint is returned. When the value of σ is low, the algorithm tends to return more changepoints than the true number, averaging around five changepoints. However, when the value of σ is higher, the algorithm tends to return one or two changepoints, depending on the size of the mean shifts.

It is important to note that in neither scenario does the algorithm correctly identify the true number of changepoints, which is two. When σ is low, the algorithm substantially overestimates by returning approximately five changepoints on average. This occurs because the GAM's flexibility allows it to fit local variations in the data, leading to overfitting where minor fluctuations are incorrectly identified as structural breaks. Conversely, when σ is high, the algorithm tends to underestimate by returning only one changepoint. In this case, the noise obscures the second true changepoint, causing the GAM to absorb the gradual transition into its smooth terms rather than detecting it as an abrupt change. This illustrates a fundamental challenge with using flexible models for changepoint detection: the signal-to-noise ratio critically affects whether the model overfits to noise (when σ is low) or overlooks genuine structural changes (when σ is high). This sensitivity has important implications for real-world applications where the true number of changepoints is unknown and must be inferred from the data alone.

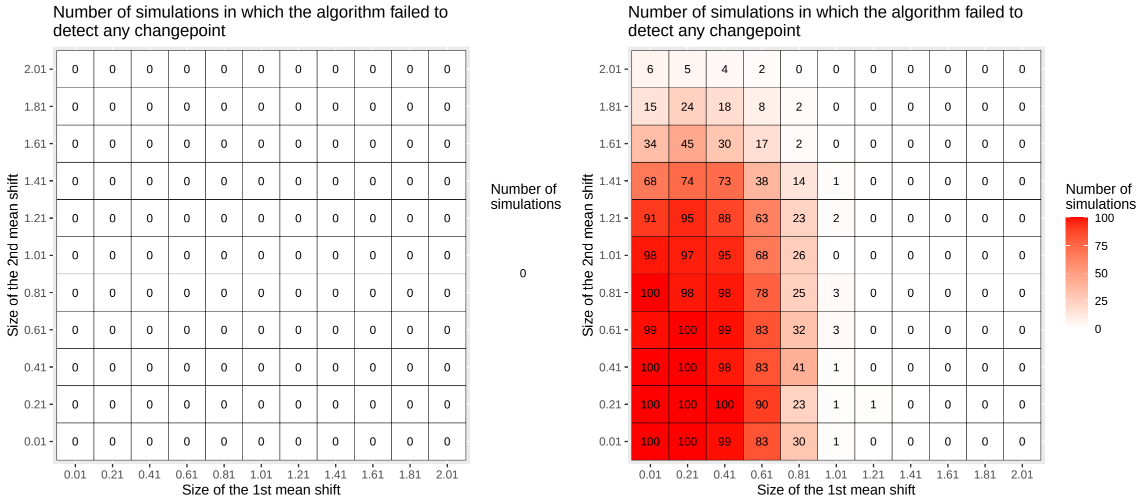

Figure 4.14: Number of simulations (out of 100) in which the algorithm failed to detect any changepoint, across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the left, σ equals 0.01. In the right, σ equals 1.01.

Figure 4.16 presents the coverage metric across simulations. The results show that when the value of σ is low, the algorithm almost correctly identifies the changepoint locations. As the value of σ increases, it becomes more challenging for the algorithm to pinpoint the true locations. This is especially true when one of the shifts is smaller than the other, as the algorithm may fail to detect the smaller shift as a changepoint, resulting in the detection of only one changepoint. Figure 4.17 shows similar results for the F1-score. However, if a detected changepoint falls outside the error margin, it will not be counted as a true positive. As a result, when there are no true positives, both precision and recall will be zero, rendering the F1-score undefined.

In summary, when a shift in the data significantly exceeds the noise level, we typically observe a clear indication of where the true change occurs. Conversely, when a shift is relatively small compared to the noise level, identifying the true changepoint becomes challenging, often resulting in difficulties detecting any changepoints using the algorithm. As expected, these results are consistent with those of the previous simulation study.

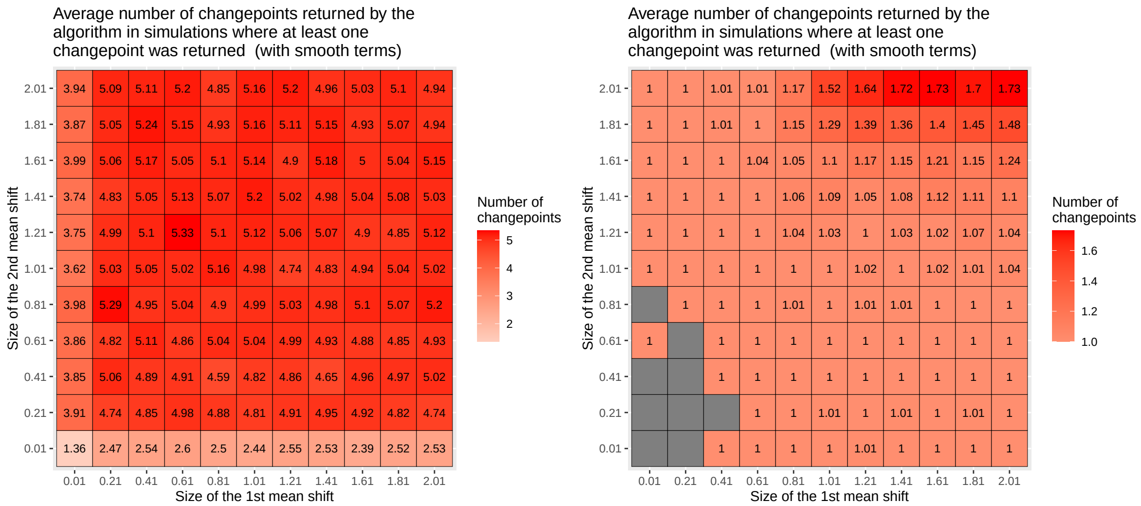

Figure 4.15: Average number of changepoints returned by the algorithm in simulations where at least one changepoint was returned (out of 100 simulations), across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the left, σ equals 0.01. In the right, σ equals 1.01.

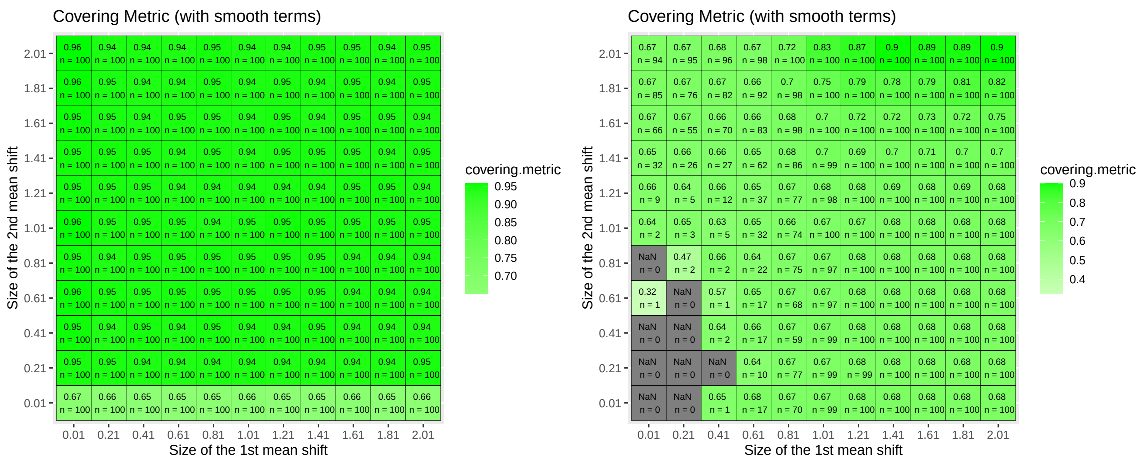

Figure 4.16: Covering metric from 100 simulations across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the left, σ equals 0.01. In the right, σ equals 1.01.

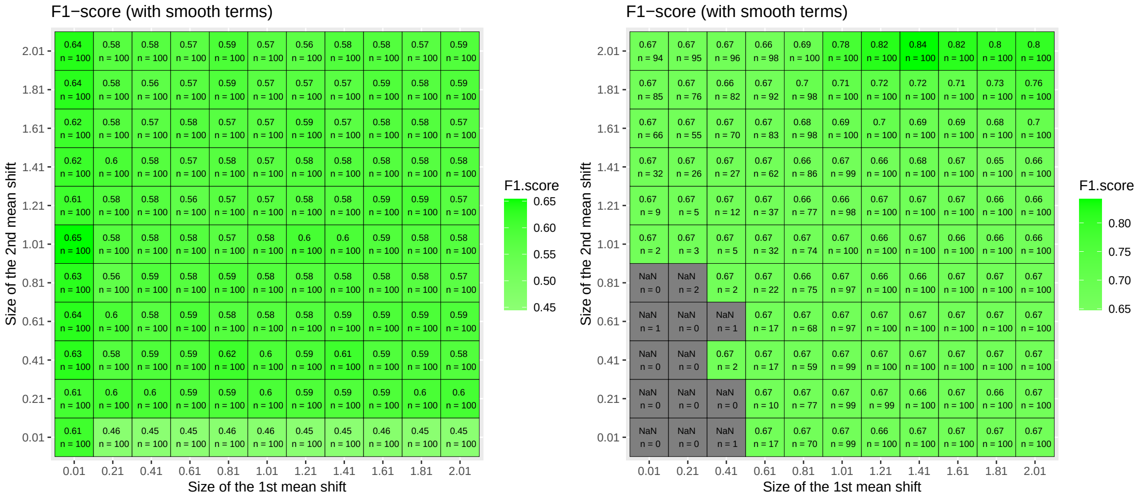

Figure 4.17: F1-score from 100 simulations across different values of σ with μ_1 on the x-axis and μ_2 on the y-axis. In the left, σ equals 0.01. In the right, σ equals 1.01.

4.7 Conclusion

In this chapter, we conducted an extensive simulation study to assess the performance of changepoint detection algorithms integrated with model-based cost functions. By simulating data with varying levels of mean shifts μ and noise σ, we explored how these parameters affect the algorithms' accuracy and computational efficiency. The results demonstrate that changepoint detection accuracy depends significantly on the relative magnitude of the mean shift to the noise level. When μ is large relative to σ, changepoints are more easily identified, resulting in lower bias, reduced Mean Absolute Error (MAE), and lower Root Mean Squared Error (RMSE) in the changepoint locations. However, when σ is high and μ is relatively small, RMSE becomes dominated by the noise, meaning that, from a practical perspective, the exact detection of changepoints has a limited impact on prediction accuracy. In these cases, high noise levels reduce the importance of accurately locating changepoints, as prediction error is primarily driven by the noise rather than by the changepoint locations themselves.

In both single and multiple changepoint simulations, the effect of σ and μ aligns with expectations: lower values of σ and higher values of μ improve changepoint detection accuracy. Conversely, when noise is high relative to the mean shift, the algorithm may fail to return any changepoint at all, as the cost difference does not exceed the detection threshold. When using flexible models such as GAMs, an additional challenge emerges: the algorithm's sensitivity to the signal-to-noise ratio affects not only whether changepoints are detected, but also how many are identified. Low noise can lead to overfitting, where the model's flexibility causes it to identify spurious changepoints in minor fluctuations, while high noise can cause genuine structural changes to be absorbed into smooth terms and overlooked.

These findings have critical implications for real-world applications. When applying changepoint detection to electricity load data in Chapter 5, we cannot know the true number or location of changepoints in advance. The simulations reveal that the detected number of changepoints is highly sensitive to data characteristics, and there is no single noise level at which the algorithm reliably identifies the correct structure. Therefore, results from real data must be interpreted cautiously, with awareness that the algorithm may overestimate, underestimate, or miss changepoints entirely depending on local signal-to-noise conditions. This understanding will inform our interpretation of detected changepoints in the electricity load analysis and guide decisions about model validation and robustness checks.

Chapter 5

Modelling Regional Demand with Changepoints

5.1 Introduction

National demand refers to the total electricity demand observed on the national grid, net of embedded generation such as small-scale renewables, accounting for consumption from households, businesses, industries, and losses. It provides a broad picture of electricity demand on a national scale, reflecting the load that needs to be met by large-scale generators connected to the transmission system. In contrast, regional net demand represents electricity demand on a more local level, similarly net of embedded generation such as rooftop solar and small wind farms, which do not directly participate in the wholesale market. The key distinction between national net demand and regional net demand lies in their geographical scope: national net demand reflects grid-wide consumption across the entire country, whereas regional net demand captures localised consumption within specific areas. Both measures exclude embedded generation, but differ in the scale of the demand they represent.

Understanding the distinction between national net demand and regional net demand is essential for effective grid management and planning. At the regional level, a single Grid Supply Point (GSP) can be heavily influenced by specific factors, such as a large industrial consumer, a concentration of certain types of customers (e.g., residential or commercial), or significant embedded generation from local renewable sources like rooftop solar. These regional variations can create unique demand profiles that are not necessarily representative of the national picture. By isolating regional net demand, grid operators and planners gain valuable insight into localised demand characteristics, enabling them to address specific regional needs, optimise grid stability, and anticipate potential issues, such as local overloads or voltage fluctuations, that may not be apparent at the national level. This localised view allows for targeted interventions and tailored demand management strategies, contributing to a more resilient and efficient energy system overall.

Grid Supply Points (GSPs) are critical nodes within the electricity distribution network that connect the high-voltage transmission system, managed by National Grid, to lower-voltage distribution networks operated by Distribution Network Operators (DNOs). These points are essential for transferring electricity from the national transmission network to regional distribution networks, ensuring power reaches homes and businesses efficiently. A GSP Group consists of a collection of Grid Supply Points within a specific geographic area managed by a DNO. These groups represent defined boundaries where electricity flows are measured and settled, aiding in the balancing of supply and demand within the region. GSP Groups are identified by unique alphanumeric codes, which are used for electricity market settlements and help streamline the distribution process.

Each Grid Supply Point within a GSP Group is assigned a unique GSP ID. This identifier serves as a specific reference for tracking electricity distribution, monitoring flow data, and facilitating the settlement processes. GSP IDs are crucial for market participants, such as suppliers and DNOs, to ensure accurate billing, monitoring, and energy balancing across the network. Together, the GSP Group and GSP ID play vital roles in maintaining the efficient and reliable management of electricity distribution. Figure 5.1 (Right) presents a map of Great Britain, highlighting 14 distinct GSP groups, each of which includes several GSP IDs. Similarly, Figure 5.1 (Left) provides maps of GSP Group E and P, showing their unique GSP IDs.

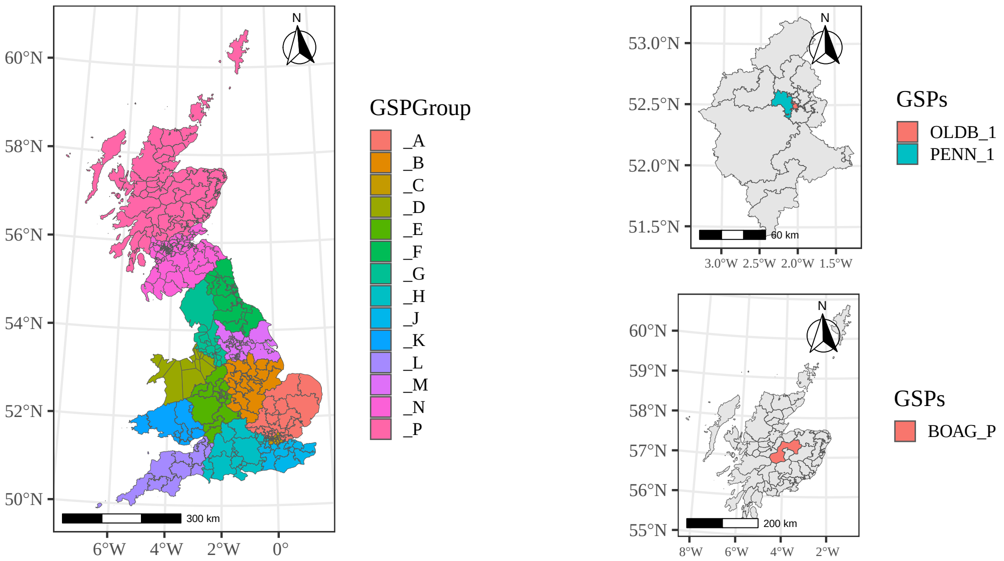

Figure 5.1: Map of the 14 GSP (Grid Supply Point) Groups in Great Britain (Left). The Distribution Network is divided into 14 distinct geographical areas across the country. The labels refer to the different areas where the area names and their respective DNOs are shown in Table 5.1. While these areas are geographically separate, they are not electrically isolated, allowing electricity to flow between them. Meters are placed at the boundaries to measure the volumes of electricity transferred between areas. Map of the 21 GSP IDs within GSP Group E (Top Right). The areas refer to the different GSP IDs, while the coloured areas (Penn (PENN_1) and Oldbury (OLDB_1)) indicate those being studied in this chapter. Some IDs appear smaller in area due to the presence of towns or cities with higher population densities. Despite their smaller size, these regions have total electricity demand comparable to that of the larger, less densely populated areas. Map of the 63 GSP IDs within GSP Group P (Bottom Right). The areas refer to the different GSP IDs, while the coloured area (Boat of Garten (BOAG_P)) is also studied in this chapter.

GSP Group ID	Area	DNO Company
_A	East England	Eastern Power Networks plc
_B	East Midlands	National Grid Electricity Distribution (East Midlands) plc
_C	London	London Power Networks plc
_D	North Wales, Merseyside and Cheshire	SP Manweb plc
_E	West Midlands	National Grid Electricity Distribution (West Midlands) plc
_F	North East England	Northern Powergrid (Northeast) plc
_G	North West England	Electricity North West Limited
_H	Southern England	Southern Electric Power Distribution plc
_J	South East England	South Eastern Power Networks plc
_K	South Wales	National Grid Electricity Distribution (South Wales) plc
_L	South West England	National Grid Electricity Distribution (South West) plc
_M	Yorkshire	Northern Powergrid (Yorkshire) plc
_N	South and Central Scotland	SP Distribution plc
_P	North Scotland	Scottish Hydro-Electric Power Distribution plc

Table 5.1: The 14 GSP Groups in Great Britain and their corresponding Distribution Network Operators (DNOs).

The goal of this chapter is to analyse and interpret the locations of detected changepoints in various real-world net demand examples from Grid Supply Points across Great Britain. Despite accounting for known drivers of demand—such as calendar effects, temperature, wind speed, and solar radiation—electricity demand time series remain vulnerable to abrupt structural shifts that conventional modelling approaches may not capture.

These net demand changes can arise from multiple sources. Operational factors include network reorganisation, such as reconfiguring connections, opening or closing switches to ensure safe and stable operation, or temporarily rerouting electricity flows during maintenance activities. Although many of these actions are recorded, the records are often inaccessible, incomplete, or subject to human error, even within the same company. Additionally, faults on the electricity network can cause abrupt changes in electricity demand. More significantly, changepoints may result from unmodelled behavioural responses and external events, such as the COVID-19 lockdown, policy interventions, or large-scale societal changes that alter consumption patterns. Technological developments also play a crucial role, including the adoption of embedded generation technologies, the commissioning of new solar or wind farms, or the integration of distributed energy resources that fundamentally alter regional demand profiles.

Even with well-specified models that incorporate established demand drivers, the functional relationships between predictors and electricity consumption can evolve over time, leading to shifts in underlying model parameters. The challenge is compounded by the fact that historical demand data were often poorly recorded, with limited documentation of operational changes or infrastructure modifications, making retrospective analysis more difficult.

Therefore, changepoint detection remains an essential tool for identifying latent shifts in demand patterns that are not captured by specified model features. By systematically detecting these structural breaks, we can provide critical insights for both demand forecasting accuracy and operational grid management, enabling more robust responses to evolving electricity consumption patterns in an increasingly complex energy landscape.

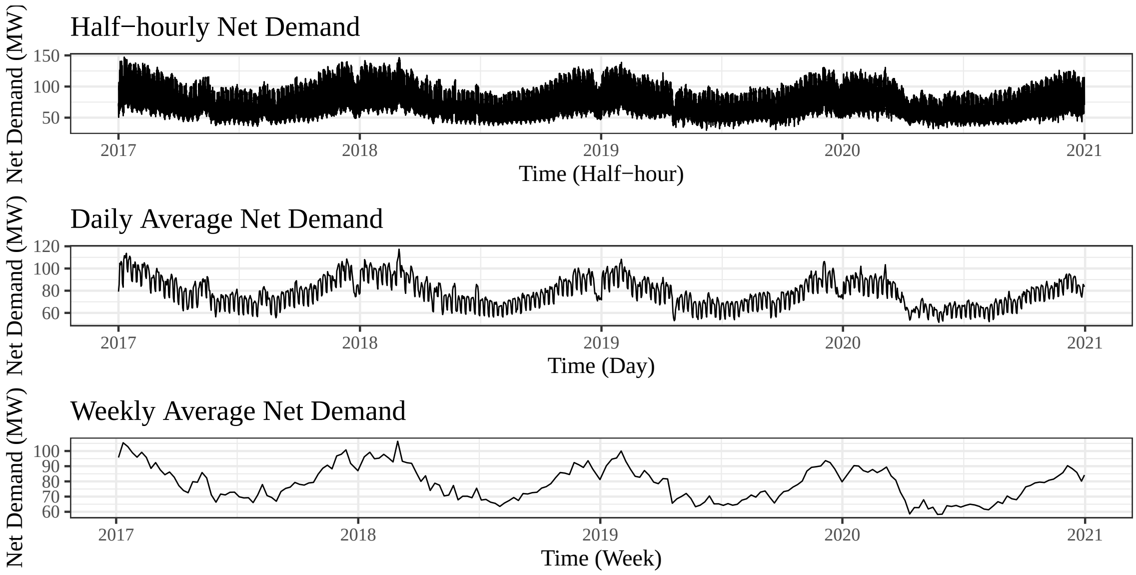

Figure 5.2: Net demand Profile of Penn (GSP ID PENN_1, from GSP Group E), January 2017 to December 2020. Half-hourly net demand Profile from raw data (top), daily average (middle), and weekly average (bottom).

5.2 Exploratory Analysis

The data for this chapter includes net demand datasets for the Grid Supply Point (GSP) IDs, collected by Elexon [1], spanning from 2017 to late 2020. The datasets provide half-hourly net demand measurements in megawatts (MW). Weather data is also incorporated into the model-building process, which can be accessed in the European Centre for Medium-Range Weather Forecasts (ECMWF) ReAnalysis 5 (ERA5) data [2], which provides hourly estimates of weather variables. This weather data consists of 41 longitude values ranging from -7.00 to 3.00, and 37 latitude values ranging from 50.00 to 59.00, with increments of 0.25 degrees in both directions. Bilinear interpolation is applied to assign weather features to each GSP ID, based on the four surrounding data points. The weather features include 2-meter temperature, 10-meter wind speed, and surface downward solar radiation.

In this chapter, GSP Group E, located in the West Midlands, is selected as the candidate for study. Figure 5.2 displays the net demand of GSP ID PENN_1 from January 2016 to December 2020. The original data, consisting of a half-hourly net demand profile, is shown at the top of the figure. Applying the Binary Segmentation (Algorithm 2) directly to this large dataset can be time-consuming due to its size. To

[1] https://www.elexon.co.uk/data/open-settlement-data/, accessed on 14 March 2022
[2] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5, accessed on 27 March 2023

address this, the daily and weekly averages of the net demand are also considered, as shown in the middle and bottom of the figure, respectively. However, the weekly average oversimplifies the data, making it difficult to detect changepoints. Therefore, the chapter proceeds with the daily average, which strikes a balance by improving computation time while preserving key data characteristics, such as seasonal variations and weekly patterns.

The daily data reveals a clear weekly pattern, with higher demand on weekdays and lower demand on weekends. Net demand is lower in the middle of the year compared to the beginning and end, primarily due to higher electricity consumption for heating during the winter. In contrast, in regions where air conditioning drives demand, summer consumption often exceeds winter levels. Additionally, increased solar energy generation during the summer helps further reduce net demand. During the Christmas season, at the end of each year, net demand drops significantly compared to the surrounding weeks, likely due to the holiday period when many people go on vacation. Additionally, there is a noticeable decline in average demand after March 2020, which corresponds to the UK's first coronavirus lockdown. Demand gradually rises afterward as the lockdown restrictions are lifted.

5.3 Variables and Models

The Generalised Additive Model (GAM) is utilised in Binary Segmentation (Algorithm 2) for greater flexibility compared to a standard linear model. It incorporates day of the week, 2-meter temperature, 10-meter wind speed, and surface solar radiation downwards as smooth terms, along with a linear trend and sinusoidal components to capture yearly seasonal patterns as the demand is influenced by these factors. Weekdays typically see higher energy consumption due to increased activity in commercial and industrial sectors, while weekends may experience lower demand. Temperature significantly impacts energy usage; higher 2-meter temperatures in summer lead to increased air conditioning use, while lower winter temperatures drive up heating requirements. Wind speed affects the availability of wind energy; higher speeds can enhance generation. Surface solar radiation also plays a crucial

Setup ID	With linear trend		Without linear trend	
Minseglen	AIC Penalty	BIC Penalty	AIC Penalty	BIC Penalty
30	1	2	5	6
20	3	4	7	8

Table 5.2: Setup ID reference for the study.

role; sunny days with high radiation reduce the need for traditional energy sources as solar panels generate more power, whereas cloudy days may increase demand for conventional generation methods.

5.4 Study Design

In this chapter, eight different configurations of the Binary Segmentation (Algorithm 2) will be applied to selected real data examples, as shown in Table 5.2. These configurations vary across two penalty types (AIC and BIC), two minimum segment lengths (minseglen) of 20 and 30, and two models (with and without a linear trend). Binary Segmentation (BS) will be used, with no maximum number of changepoints and a search increment of 1. The best configuration will be determined by computing the AIC and BIC on the final model, as other evaluation metrics that minimise the residuals, including MAE and RMSE, tend to overfit by selecting too many changepoints, leading to an overly complex model. The ideal number of changepoints should not be excessive. For instance, when the algorithm selects as many changepoints as possible, the length of each segment becomes equal to the minimum segment length, which may not accurately reflect real changes in the data. While some changes are evident, others may not be. For example, there might be a change in a smooth term, such as demand being affected by weather during a certain time of year, and the algorithm should be able to capture those.

The choice between AIC and BIC depends on the goals of the analysis. AIC aims to find a model that balances goodness of fit with complexity, often favouring models that explain the data well, but this can sometimes lead to overfitting. In contrast, BIC applies a stronger penalty for model complexity, particularly as the sample size increases, making it more conservative and less prone to overfitting. Consequently, AIC is typically preferred when accuracy is the priority, especially with smaller datasets, while BIC is preferred when the goal is to identify the true underlying model, particularly with larger datasets.

The selection of minimum segment length is important to prevent the identification of excessively short segments that may not represent meaningful structural changes. However, if set too high, the model may fail to detect genuine changes occurring over shorter periods. Due to time constraints, it is not feasible to test every possible minimum segment length, so lengths of 20 and 30 are chosen—where 30 corresponds to approximately one month, and 20 is the minimum length required to build a model of this complexity. Using less data than this would make it impossible for the model to fit properly. Excluding the linear trend from the model can also be contentious, as keeping it in may lead to significantly different results, particularly with short data segments. In such cases, the model might misinterpret sinusoidal patterns, such as natural seasonal fluctuations, as changepoints, incorrectly treating the regular increase and decrease as separate linear trends.

5.5 Results and evaluation

An example of the results is Oldbury (OLDB_1 as a GSP ID) from GSP Group E. Figure 5.3 illustrates the changepoint locations predicted by the algorithm using the eight different setups. On the left-hand side, setups 1, 3, 5, and 7 appear to overfit by predicting too many changepoints, suggesting that the AIC penalty is not as strong as the BIC because it favours models that explain the data well despite having smaller segments. In contrast, the BIC-based setups perform better, predicting fewer, more reasonable changepoints. The most common changepoint locations occur around the Christmas period and in March 2020, coinciding with the start of the coronavirus lockdown. When comparing the BIC values of the final models, setup 8 emerges as the best-performing method, achieving the greatest reduction in BIC (11.35%), as shown in Figure 5.4.

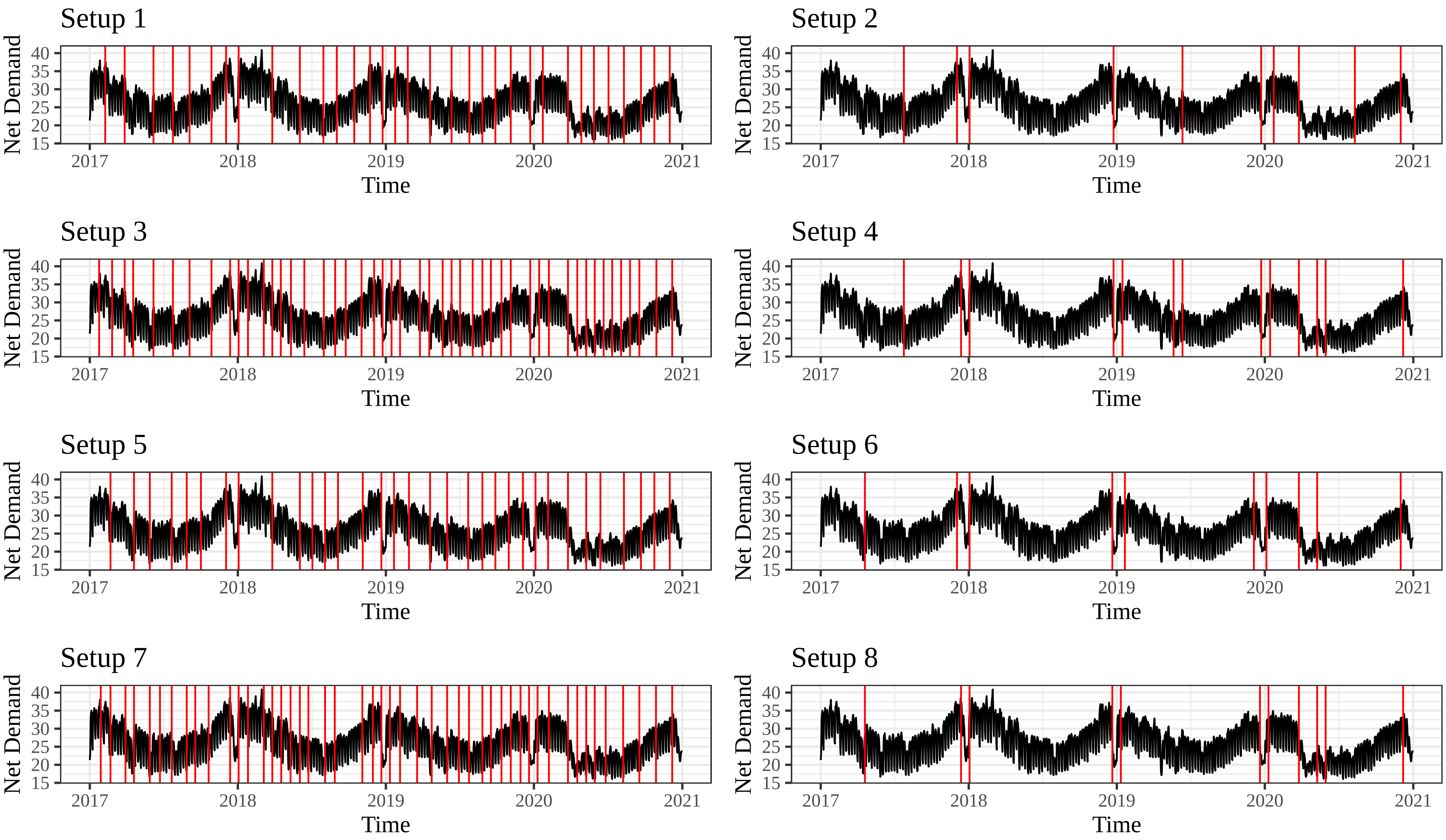

Figure 5.3: Changepoint locations of the region OLDB_1 predicted by the algorithm using different setups related to Table 5.2.

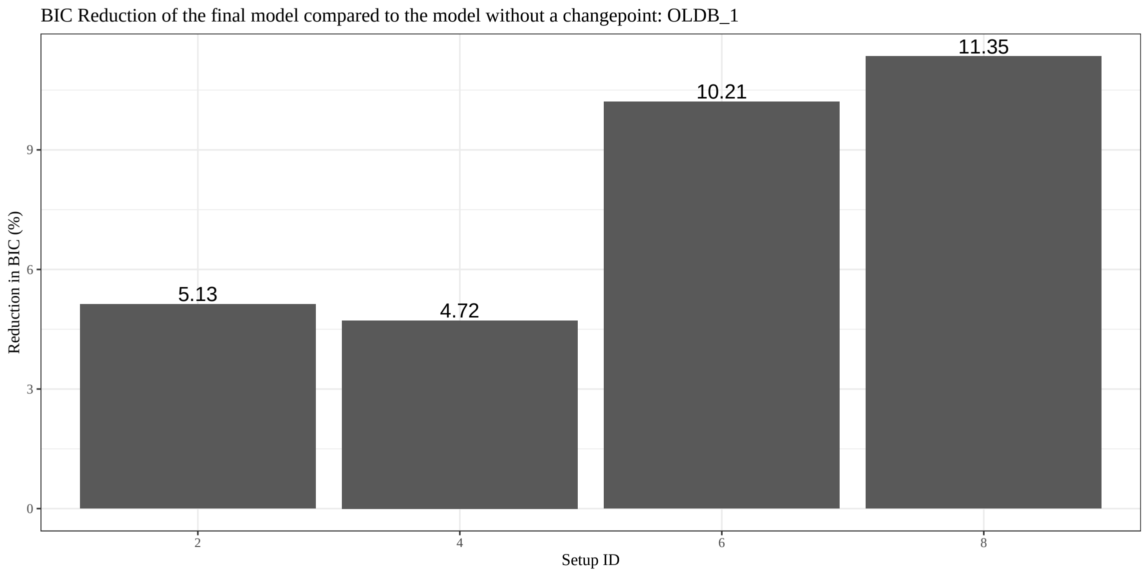

Figure 5.4: Reduction in BIC of the region OLDB_1 for the final model compared to the model without changepoints.

Figure 5.5 provides a detailed view of the location and demand profile from Setup 8, which features 12 segments of varying lengths. Notably, Setups 3, 5, 7, and 12 represent shorter segments corresponding to the Christmas period each year, covering only a few weeks. A closer analysis of the smooth terms will be conducted for Segments 2, 4, 6, and 11, as they contain sufficient data for meaningful analysis.

Figure 5.6 illustrates the effect of smooth terms based on the day of the week across different segments. The behaviour of these smooth terms is consistent, with no unusual patterns, reflecting a typical weekday effect due to the adequate amount of data in each segment. In contrast, Figure 5.7 shows the impact of smooth terms for 2-meter temperature (t2m). The figure highlights that when temperatures are extremely low or high, there is insufficient data for calculating the smooth terms with low uncertainty. However, around 290 K (approximately 17°C), the smooth term effect is minimal, indicating this as an ideal, comfortable temperature. Lower temperatures lead to increased heating demand, while higher temperatures correspond with a rise in air conditioning demand, as suggested by the plot.

Notably, segment 11, which corresponds to the lockdown period, shows reduced evidence of air conditioning demand. This segment's distinctive parameter behaviour is identified by the changepoint detection algorithm, underscoring the value of changepoint analysis. By identifying such shifts in parameters, the algorithm enables us to detect key changes in demand patterns over time, which can inform more responsive

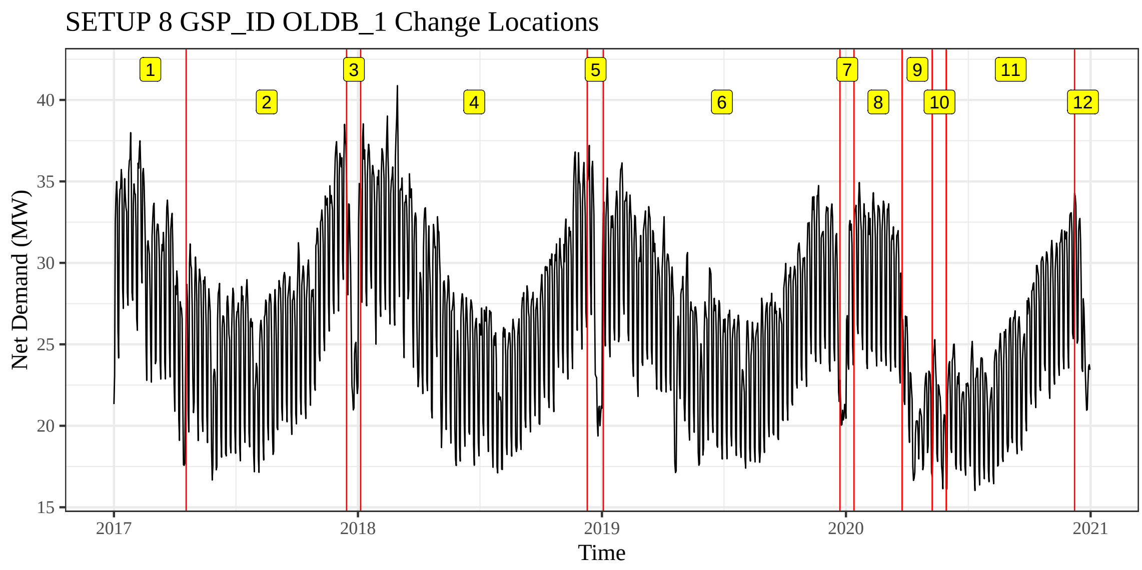

Figure 5.5: The demand profile of OLDB_1 and its changepoint locations, as identified by the algorithm using Setup 8. Each segment is labeled to indicate its identity. Segments 3, 5, 7, and 12 correspond to the Christmas period. Segments 9, 10, 11, and 12 represent the period following the coronavirus lockdown. While changes between segments 1 and 2, as well as between segments 9, 10, and 11, may not be visually apparent, they could indicate changes in model parameters. Although the algorithm is designed to detect behavioral changes that may not be easily observed, no changes may be detected during the Christmas period if the data is excluded or modeled carefully. However, this cannot be done in the model-building process, as it is not feasible to build a separate model for non-Christmas segments; the algorithm calculates the cost uniformly across segments, and a Christmas segment is too short to build a model efficiently.

and adaptive forecasting models.

Figure 5.8 demonstrates the effect of smooth terms for 10-meter wind speed (10w), revealing a slightly positive linear relationship between wind speed and the smooth term effect. As wind speeds increase, the cooling effect becomes more pronounced, similar to the effects of low temperatures. Conversely, weaker winds result in reduced cooling effects. Finally, Figure 5.9 shows the smooth term effects for surface solar radiation downward (ssrd). As ssrd values increase, indicating more solar energy reaching the Earth's surface, the smooth term effect diminishes, particularly in Segment 11. This suggests that higher ssrd levels enable individuals to generate more energy independently, thereby reducing overall demand.

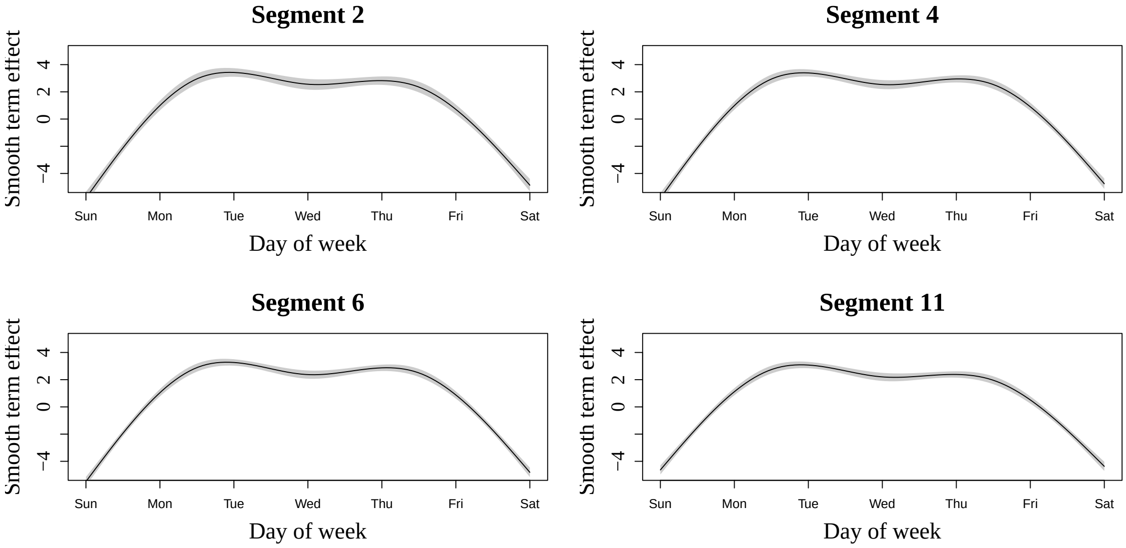

Figure 5.6: Figure showing the effect of the day of the week for segments 2, 4, 6, and 11 on region OLDB_1.

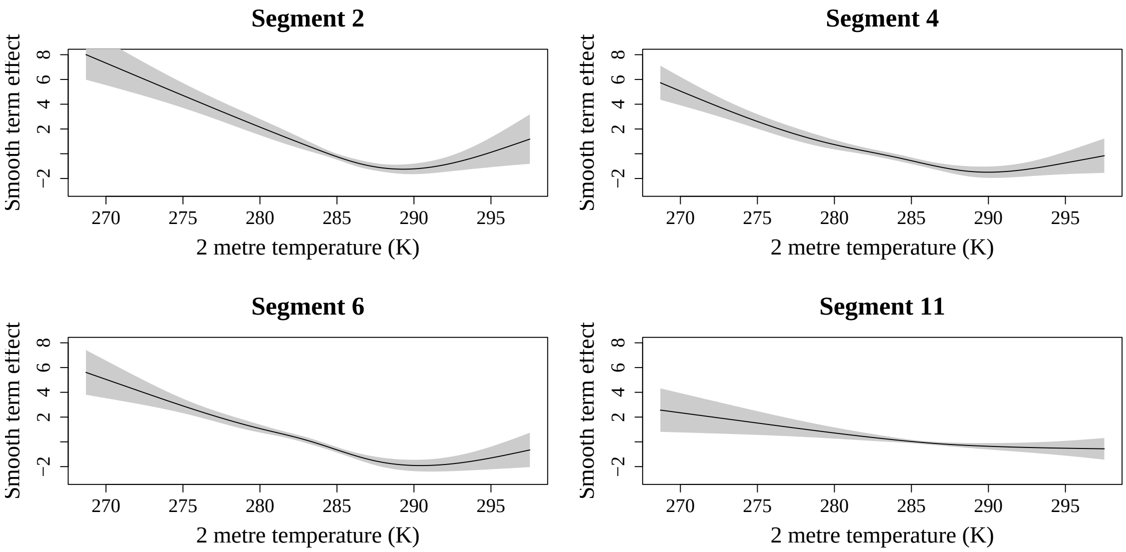

Figure 5.7: Figure showing the effect of the 2 metre temperature for segments 2, 4, 6, and 11 on region OLDB_1.

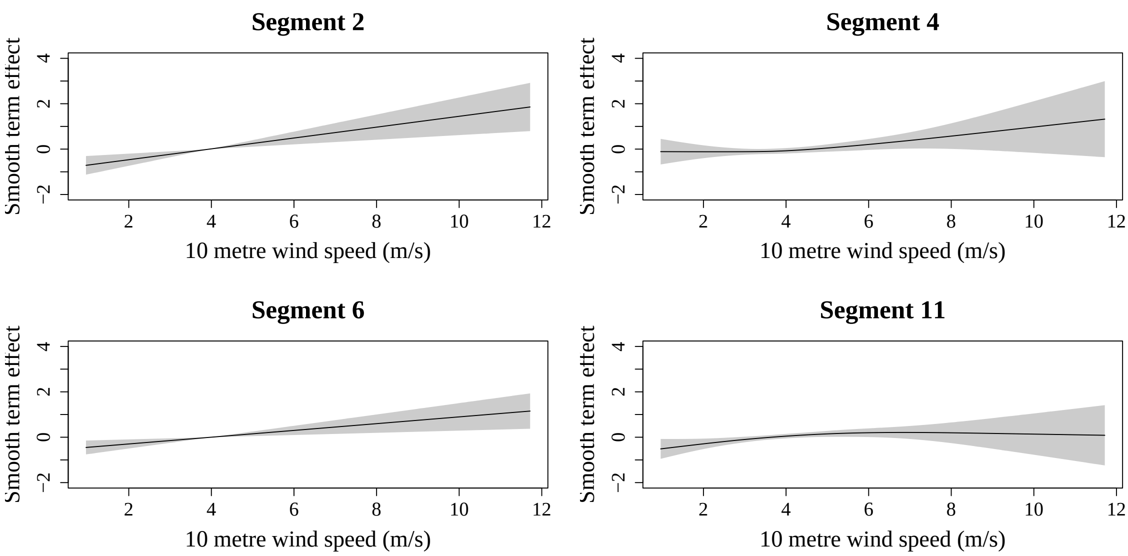

Figure 5.8: Figure showing the effect of the 10 metre wind speed for segments 2, 4, 6, and 11 on region OLDB_1.

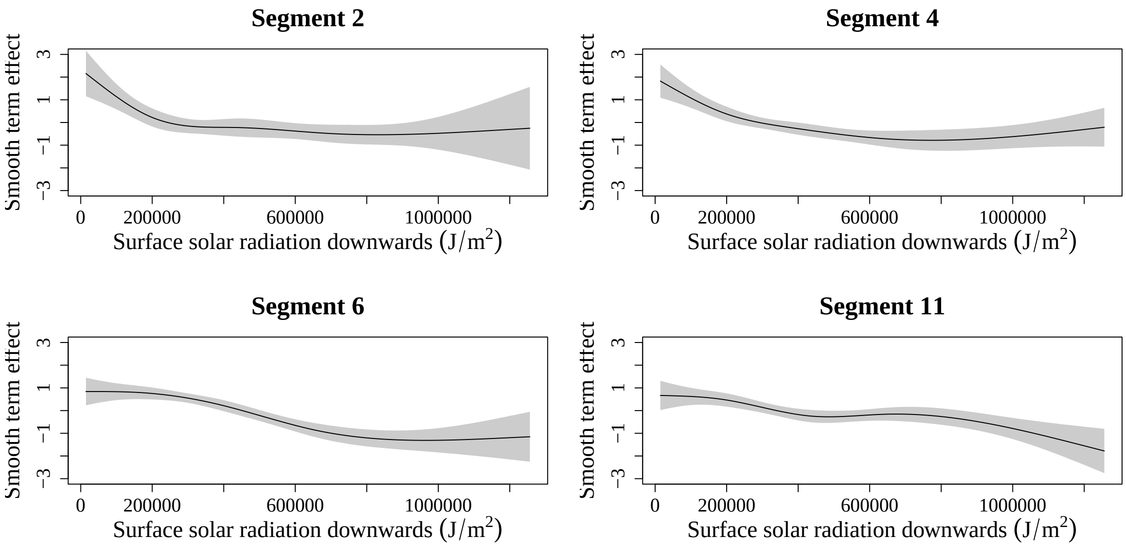

Figure 5.9: Figure showing the effect of the surface solar radiation downwards for segments 2, 4, 6, and 11 on region OLDB_1.

Another example of the results is PENN_1, which is also a GSP ID from GSP Group E. Figure 5.10 illustrates the changepoint locations predicted by the algorithm using the eight different setups. Again, on the left-hand side, setups 1, 3, 5, and 7 appear to overfit by predicting too many changepoints, suggesting that BIC-based setups perform better by predicting fewer, more reasonable changepoints. The most common changepoint locations occur around March 2020, when demand begins to decrease, coinciding with the start of the coronavirus lockdown. Unlike OLDB_1, the changes during the Christmas period are not as pronounced, particularly when the linear trend is included. When comparing the BIC values of the final models, setup 6 emerges as the best-performing method, achieving the greatest reduction in BIC (5.76%), as shown in Figure 5.11. Interestingly, setup 4 does not improve the BIC; in fact, it worsens it by increasing the BIC from the starting model (by 1.3%). In this case, having no changepoint is preferable to having changepoints, based on the criterion of maximising the BIC reduction percentage.

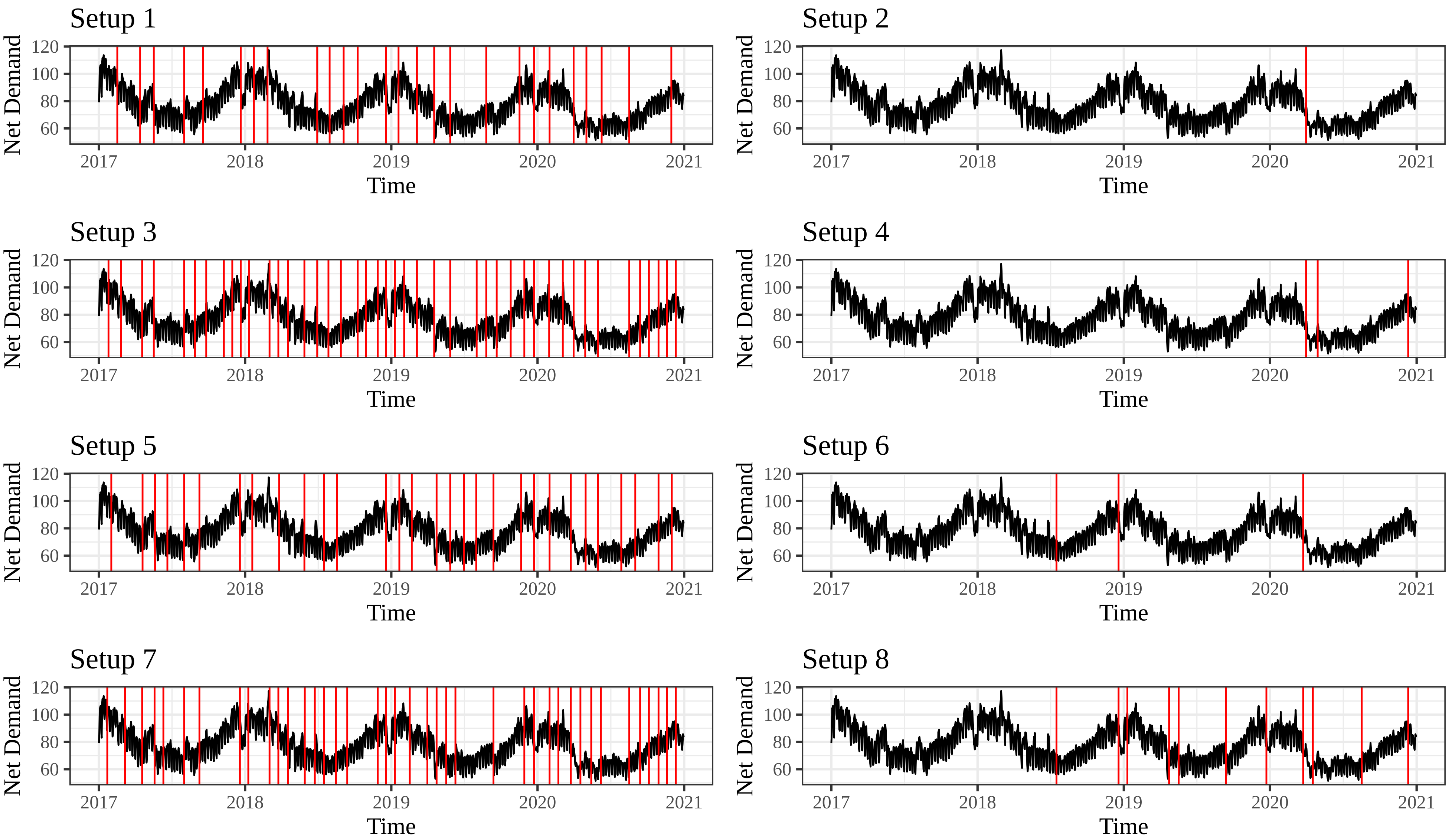

Figure 5.10: Changepoint locations of the region PENN_1 predicted by the algorithm using different setups related to Table 5.2.

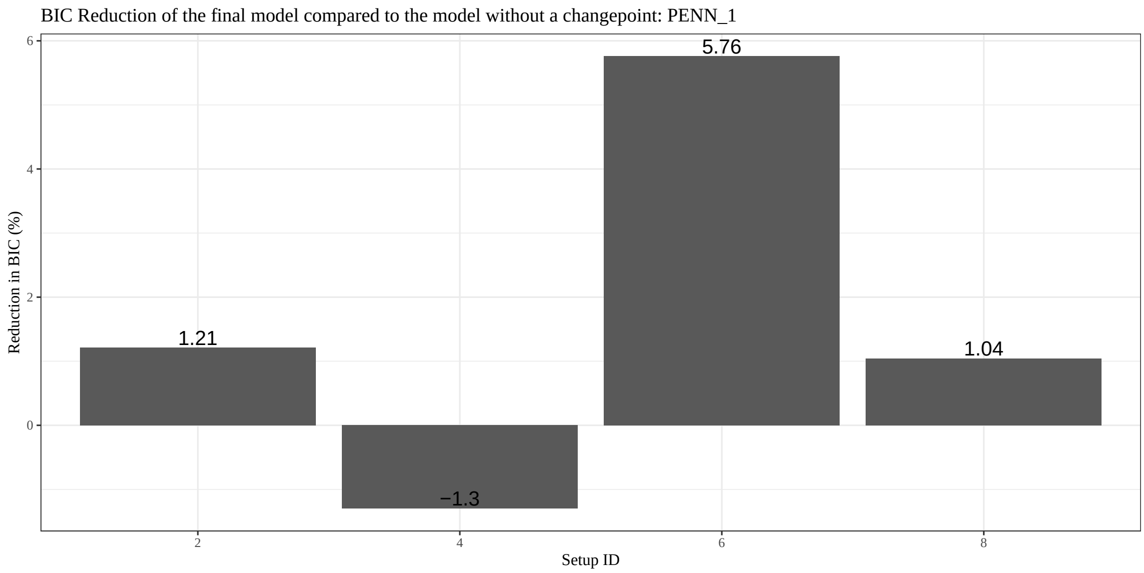

Figure 5.11: Reduction in BIC of the region PENN_1 for the final model compared to the model without changepoints.

Figure 5.12 offers a detailed view of the location and demand profile from setup 6, which includes four segments of varying lengths. This number of segments is lower than that of OLDB_1, emphasising the geological differences that may influence behavior, such as variations in the concentration of industries or businesses in the area. Segment 4 pertains to the period following the coronavirus lockdown.

Figures 5.13 and 5.14 display the impact of smooth terms based on the day of the week and 2-meter temperature (t2m) across different segments, respectively. The behavior of the smooth terms is largely consistent, except for segment 4, which exhibits a different t2m effect due to changes in behavior. This also shows a similar shift in the temperature smooth term in the prior example, further highlighting a consistent behavioural change during segment 4.

Figure 5.15 highlights the influence of smooth terms for 10-meter wind speed (10w), showing a slightly positive linear relationship between the effect of smooth terms and wind speed. Segment 3 appears to exhibit a different response, with greater uncertainty at higher wind speeds. Figure 5.16 illustrates the impact of smooth terms for surface solar radiation downward (ssrd). The reason for supporting the effect of the smooth terms is similar to the previous example.

Next, consider setup 2 of PENN_1, as shown in Figure 5.10, which is examined more closely in Figure 5.17. Segment 2 from this setup, referred to as Segment A, is comparable to segment 4 of setup 6, referred to as Segment B. Segment A

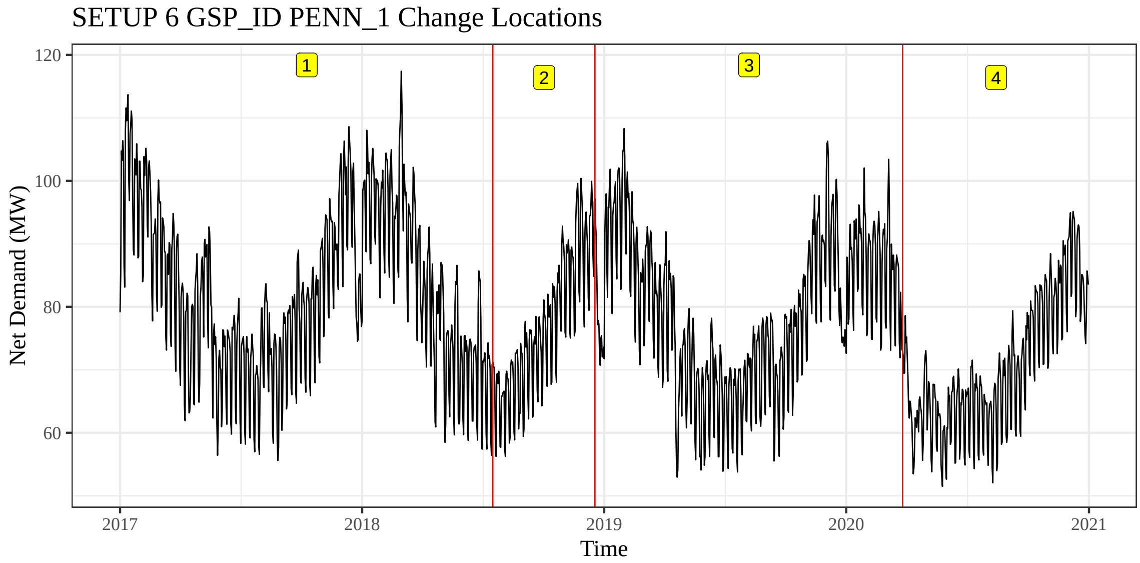

Figure 5.12: The demand profile of PENN_1 and its changepoint locations, identified by the algorithm using setup 6, are displayed here. Each segment is clearly labeled to indicate its identity. The number of segments is lower compared to OLDB_1, reflecting the geological differences that may influence behavior, including variations in the number of industries or businesses in the area. Segment 4 represents the period following the coronavirus lockdown. Although a change in segment 3 may appear evident to the eye, the algorithm does not identify a changepoint in this segment.

Segment	Segment A			Segment B		
Variable	Estimate	Conf. Interval	P-value	Estimate	Conf. Interval	P-value
Intercept term	13.10	(-11.48, 37.68)	0.30	70.95	(70.48, 71.42)	<0.01
Linear trend (MW/week)	0.31	(0.17, 0.45)	<0.01		-	
Yearly seasonality (Sine)	0.30	(-1.17, 1.77)	0.69	-2.61	(-3.37, -1.85)	<0.01
Yearly seasonality (Cosine)	3.41	(1.98, 4.84)	<0.01	5.20	(3.98, 6.42)	<0.01

Table 5.3: Comparison of model parameters, including estimated value, 95% confidence interval, and P-value, between segment 2 of setup 2 (Segment A) and segment 4 of setup 6 (Segment B) for PENN_1. Notably, the intercept term for Segment A is not significant at 95% confidence level, whereas it is significant for Segment B. This result indicates that Segment B, a model with a simpler structure, performs better as all non-smooth variables within the model are statistically significant at the 95% confidence level.

consists of 275 data points, while Segment B has 281, just under a week's difference. The algorithm configurations for setups 2 and 6 are identical, except that setup 2 incorporates a linear trend in the model, whereas setup 6 does not. Their model parameters are investigated and compared. With the inclusion of a linear trend, the algorithm does not detect changes in segment 1 of setup 2 but does identify a change when the linear trend is excluded in setup 6. The estimated coefficients for the non-smooth terms are shown in Table 5.3. The yearly seasonal components include both sine and cosine terms with the same frequency. Notably, in segment B—where the linear trend is excluded—the confidence intervals are narrower than in segment A, where the linear trend is included. However, the intercept term in Segment A is not significant at 95% confidence level, while it is significant for Segment B. Despite this, there are indications that a genuine linear trend may be present in Segment A.

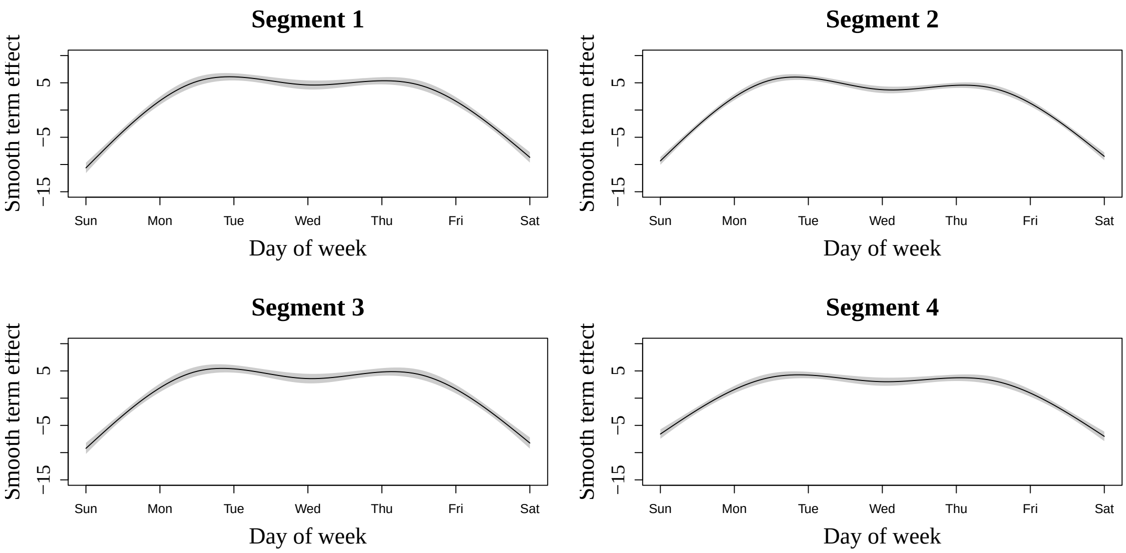

Figure 5.13: Figure showing the effect of the day of the week for segments 1, 2, 3, and 4 on region PENN_1.

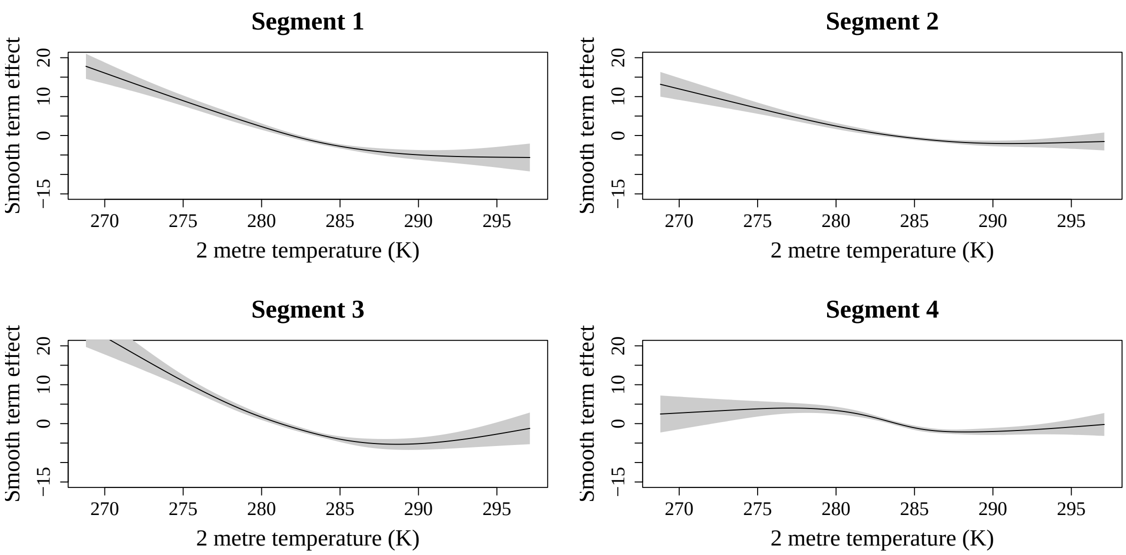

Figure 5.14: Figure showing the effect of the 2 metre temperature for segments 1, 2, 3, and 4 on region PENN_1.

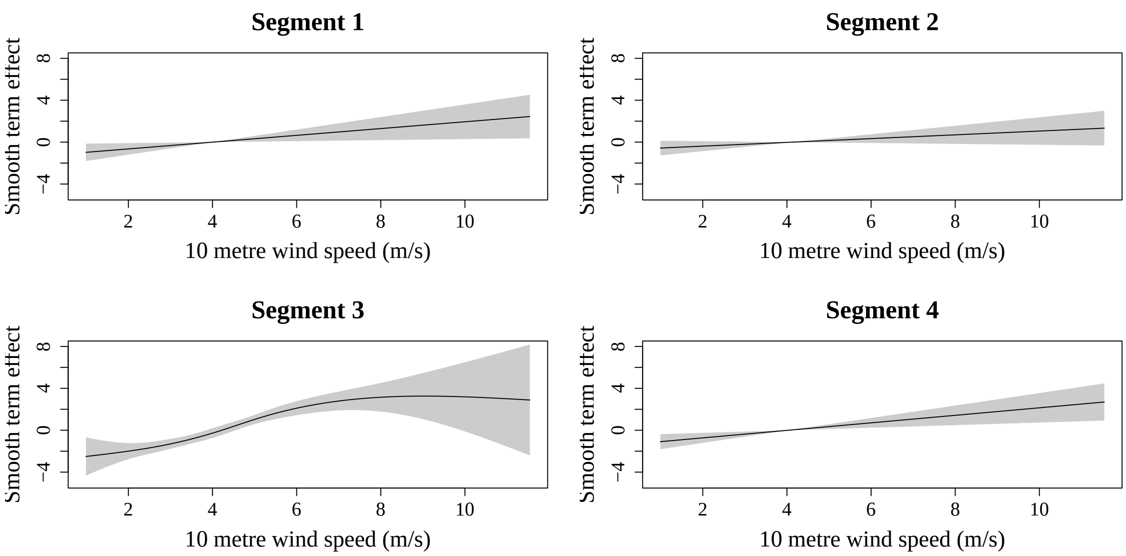

Figure 5.15: Figure showing the effect of the 10 metre wind speed for segments 1, 2, 3, and 4 on region PENN_1.

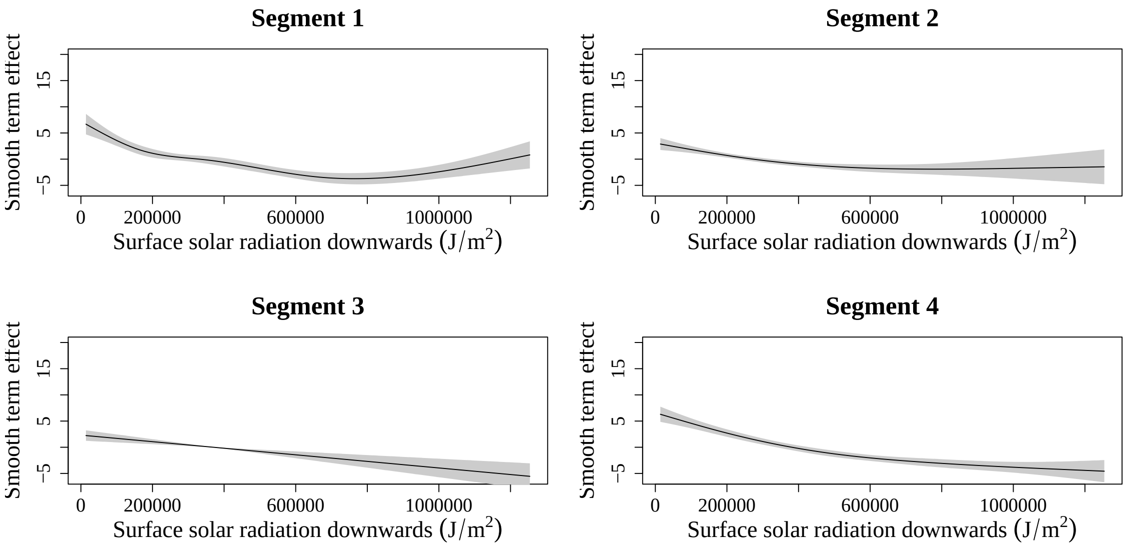

Figure 5.16: Figure showing the effect of the surface solar radiation downwards for segments 1, 2, 3, and 4 on region PENN_1.

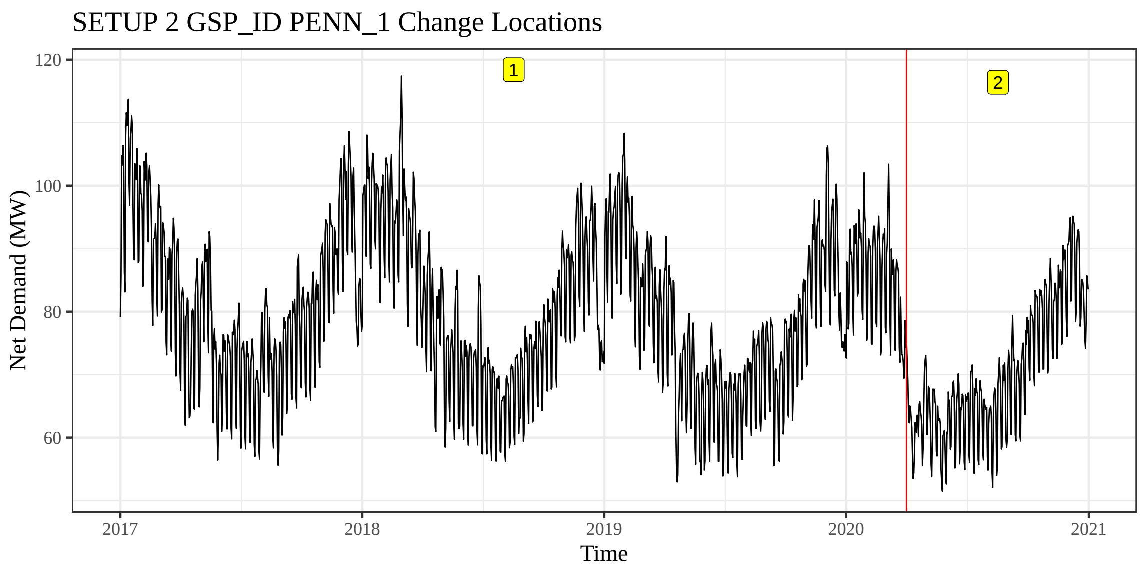

Figure 5.17: The demand profile of PENN_1 and its changepoint locations, identified by the algorithm using setup 2, are presented here. Each segment is clearly labeled for easy identification. Segment 2 in this figure is comparable to segment 4 in Figure 5.12; however, segment 2 contains 275 data points, while segment 4 has 281 data points. With the inclusion of a linear trend, the algorithm does not identify any changes in segment 1, although it does find a change when the linear trend is excluded.

Finally, BOAG_P is presented, a GSP ID representing Boat of Garten from GSP Group P in North Scotland. Figure 5.18 displays the changepoint locations predicted by the algorithm using four different BIC penalty setups referenced in Table 5.2. The most frequent changepoint locations occur around September 2017, where the change is subtle, February 2019, where the change is more noticeable, and September 2019, where there is a sharp shift in net demand characteristics. The algorithm does not identify any changepoints during the Christmas period. When comparing BIC values for the final models, setup 6 is the best-performing, achieving the largest BIC reduction (27.29%), as shown in Figure 5.19. This represents a significant improvement compared to the previous two examples in percentage terms.

Figure 5.20 provides a detailed view of the location and demand profile from setup 6, which is divided into six segments of varying lengths. Figure 5.21 shows the effect of smooth terms based on the day of the week across these segments. The behaviour of the smooth terms varies slightly; segments 1 and 2, being adjacent, display similar day-of-week patterns with typical net demand profiles, whereas segments 5 and 6 show no such similarities. In these latter segments, the day of the week no longer affects net demand, and they exhibit abnormal demand behaviour.

Figure 5.22 illustrates the effect of smooth terms for 2-meter temperature (t2m) across segments. Although segments 1 and 2 are adjacent, they show differing tem-

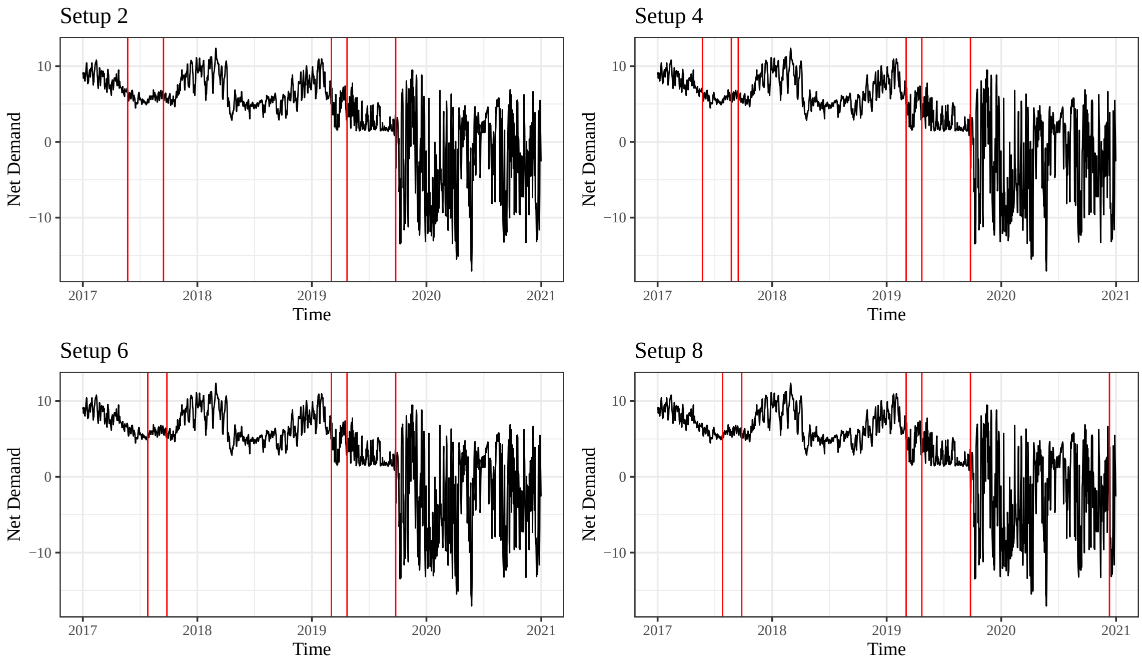

Figure 5.18: Changepoint locations of the region BOAG_P predicted by the algorithm using different setups related to Table 5.2.

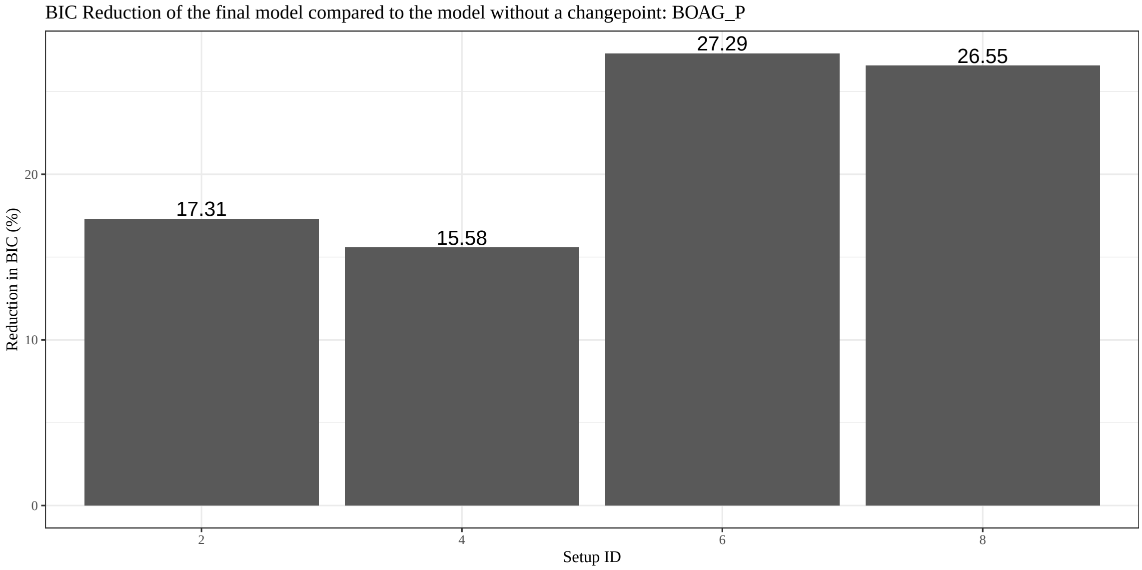

Figure 5.19: Reduction in BIC of the region BOAG_P for the final model compared to the model without changepoints.

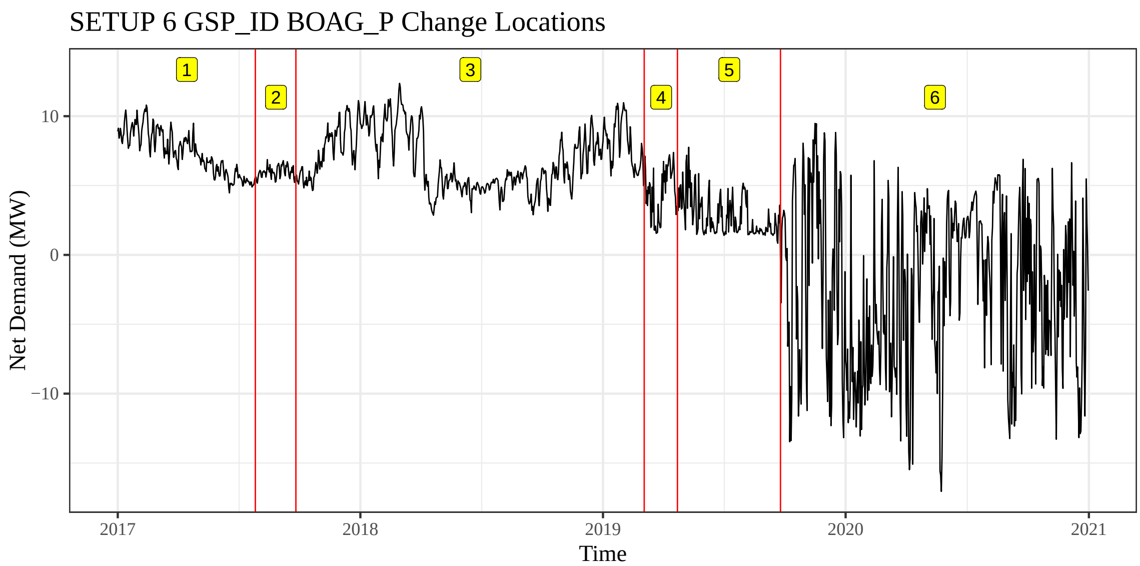

Figure 5.20: The demand profile of BOAG_P and its changepoint locations, identified by the algorithm using setup 6, are displayed here. Each segment is clearly labeled to indicate its identity. Segment 6 begins to show negative demand, meaning that generation exceeds usage. This suggests that a wind farm may have been constructed during that period.

perature effects, alongside changes in demand variance, prompting the algorithm to detect a changepoint between them. Segment 6 presents an anomalous effect, likely due to its irregular net demand profile.

Figure 5.23 highlights the influence of smooth terms for 10-meter wind speed (10w), revealing a slightly positive linear relationship between wind speed and smooth terms in segments 1 and 2, but not in segments 5 and 6. This provides strong evidence that a wind farm was operational during segment 6. As instantaneous wind speed increases, the smooth term's effect decreases because wind farms generate more energy at higher wind speeds, reducing net demand. However, when wind speed reaches the cut-off point for wind farms, they automatically shut down for safety reasons, causing an increase in the smooth term effect due to the cessation of energy generation. It can also be concluded that wind farms impacted segment 5, as the behaviour of the smooth terms in this segment is similar to that of segment 6.

Finally, Figure 5.24 displays the effect of smooth terms for surface solar radiation downward (ssrd), revealing a slightly negative linear relationship. However, segment 6 behaves differently, likely due to the substantial impact of wind farms on net demand during this period.

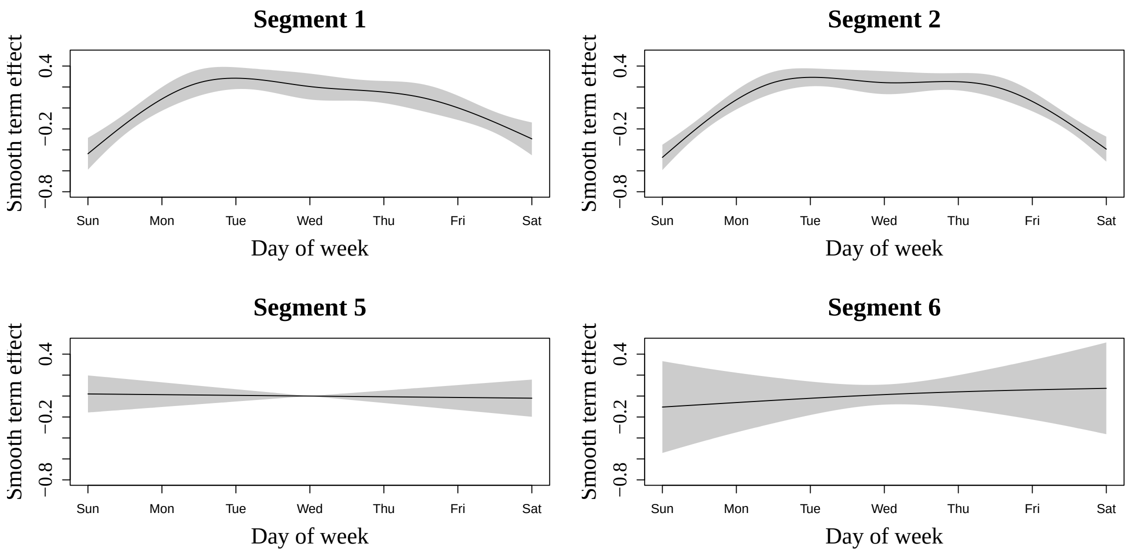

Figure 5.21: Figure showing the effect of the day of the week for segments 1, 2, 5, and 6 on region BOAG_P.

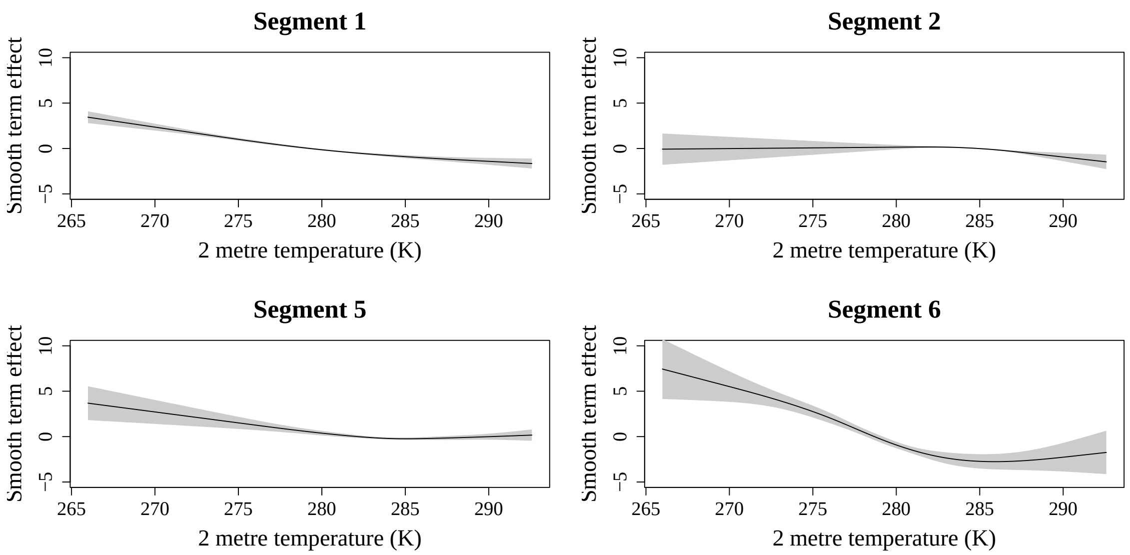

Figure 5.22: Figure showing the effect of the 2 metre temperature for segments 1, 2, 5, and 6 on region BOAG_P.

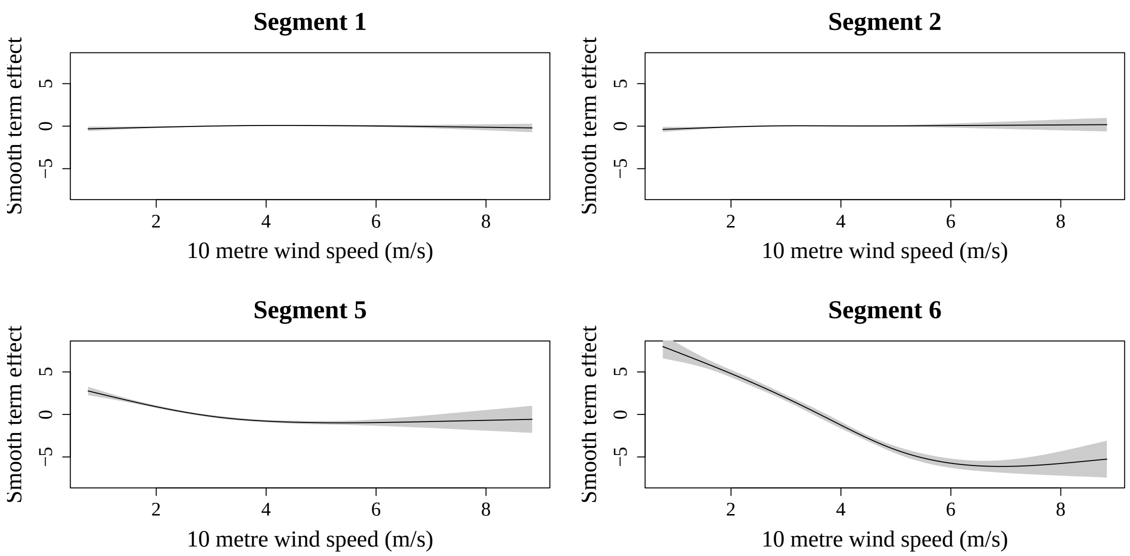

Figure 5.23: Figure showing the effect of the 10 metre wind speed for segments 1, 2, 5, and 6 on region BOAG_P.

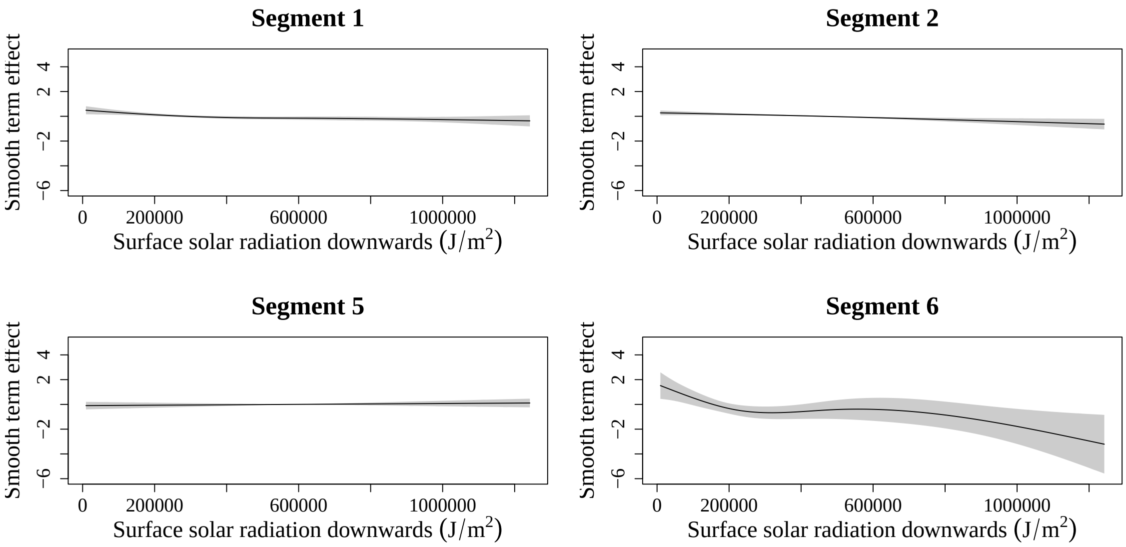

Figure 5.24: Figure showing the effect of the surface solar radiation downwards for segments 1, 2, 5, and 6 on region BOAG_P.

5.6 Discussion and Conclusion

In this chapter, we applied changepoint detection algorithms to real-world net demand data, focussing on various Grid Supply Points (GSPs) across Great Britain. The study effectively identified changepoints in net demand profiles using different configurations of Binary Segmentation with both AIC and BIC penalty terms. The results demonstrated the algorithm's ability to capture significant shifts in electricity demand patterns, particularly during notable events such as the Christmas period and the onset of the COVID-19 lockdown.

5.6.1 Critical Analysis of Detected Changepoints

The changepoint detection algorithm identified various significant shifts in electricity demand across the analysed GSPs. These changepoints can be broadly categorised into several groups based on their underlying causes and interpretation: seasonal and holiday effects, pandemic-related disruptions, infrastructure development, and weather-driven variations.

Seasonal and Holiday Effects

Changepoints were detected during the Christmas–New Year period in some GSPs (notably Oldbury:OLDB_1), typically occurring around December 24–26, resulting in multiple short segments of only a few weeks. However, these detections should be interpreted differently from true structural breaks. Holiday periods represent **recurrent annual patterns** rather than genuine structural changes in the demand-generating process. The shifts in electricity consumption during these periods—such as industrial shutdowns and increased residential heating—occur predictably every year and are therefore part of the normal seasonal cycle.

When the algorithm detects holiday periods as changepoints, this indicates that the model specification was **insufficient to fully capture these recurrent patterns**. In cases where holidays were flagged as changepoints (as in OLDB_1), the GAM framework's seasonal components and other covariates failed to adequately represent the abrupt but predictable nature of holiday demand shifts. Conversely, for GSPs where holiday periods were **not** detected as changepoints (such as Penn:PENN_1 and Boat of Garten:BOAG_P), the model successfully learnt and incorporated these recurring patterns into its baseline representation, correctly treating them as expected variations rather than structural breaks.

This distinction is crucial for interpretation: the detection of holiday-related changepoints represents a **model limitation** rather than a methodological success. An ideal model would fully capture recurrent seasonal and holiday patterns, reserving changepoint detection for truly anomalous events that represent fundamental shifts in the underlying demand process. The variable performance across GSPs suggests that model adequacy differs by location, potentially due to varying degrees of regularity in holiday consumption patterns or differences in the dominant consumer types (residential versus commercial/industrial) across regions.

COVID-19 Pandemic Impact

The most dramatic and widespread changepoints were detected in March 2020, coinciding precisely with the UK government's announcement of the first national lockdown on March 23, 2020. These changepoints appeared across all analysed GSPs (OLDB_1, PENN_1, and BOAG_P), reflecting the unprecedented and uniform impact of pandemic restrictions on electricity demand patterns.

The lockdown measures resulted in fundamental changes to electricity consumption behaviour. Analysis of the smooth terms revealed distinctive parameter shifts during the lockdown period. Most notably, segment 11 of setup 8 in OLDB_1 and segment 4 of setup 6 in PENN_1 (both corresponding to the post-lockdown period) showed reduced evidence of air conditioning demand in the temperature smooth term effect. This behavioural change—where the typical response to high temperatures was diminished—reflects the shift in building occupancy patterns as commercial facilities remained largely empty whilst residential consumption increased.

The magnitude and consistency of these shifts across multiple GSPs validate the algorithm's detection capabilities for large-scale societal disruptions. These represent genuine structural breaks in demand patterns, as the fundamental relationship between weather variables and consumption changed during this period due to altered human behaviour rather than system infrastructure changes.

Infrastructure Development and Renewable Generation

Some detected changepoints reflect genuine structural transformations in the electricity system infrastructure. The most compelling evidence comes from BOAG_P (Boat of Garten), which showed changepoints in September 2017, February 2019, and September 2019. Analysis of segment 5 and particularly segment 6 from setup 6 revealed dramatic changes in the relationship between wind speed and net demand.

The smooth term analysis for 10-metre wind speed in BOAG_P provides strong evidence of **wind farm commissioning**. In early segments (1 and 2), wind speed showed a slightly positive linear relationship with net demand—higher wind speeds increased cooling effects and thus demand. However, in segments 5 and 6, this relationship fundamentally changed: as wind speed increased, net demand **decreased** due to wind farm generation offsetting consumption. Furthermore, the characteristic pattern of demand increase at very high wind speeds (when wind farms shut down for safety at cut-off speeds) provides conclusive evidence of embedded wind generation.

This infrastructure-related changepoint represents a **legitimate structural break** that the algorithm correctly identified. When significant renewable capacity comes online, it fundamentally alters net demand patterns by introducing weather-dependent generation that directly reduces observed demand. The net demand measured at a

GSP represents gross consumption minus embedded generation; thus, new wind farms create a structural break in the relationship between weather variables (particularly wind speed) and observed net demand.

The BOAG_P example also demonstrated changes in day-of-week effects, with segments 5 and 6 showing no typical weekday patterns, suggesting that the dominant demand characteristics shifted from consumption-driven (with clear weekday/weekend differences) to generation-dominated patterns (where renewable output follows weather rather than human schedules).

Similarly, changes in solar capacity, battery storage systems, or major industrial operations (such as factory closures or openings of energy-intensive facilities like data centres) would constitute genuine structural changes. These infrastructure-related changepoints are typically permanent or long-lasting alterations to the system, unlike recurrent seasonal patterns.

Weather-Related Variations and Model Specification

Beyond the clear categories above, some changepoints appear associated with weather variations or represent ambiguous cases. For instance, PENN_1 showed differences between adjacent segments in temperature effects and wind speed responses that may reflect either genuine behavioural shifts or limitations in the smooth functional forms used in the GAM framework.

The comparison between setup 2 and setup 6 for PENN_1 (Penn) illustrates the sensitivity of changepoint detection to model specification. When a linear trend was included (setup 2), the algorithm did not detect changes in segment 1, but when the linear trend was excluded (setup 6), a changepoint was identified. This highlights that some detected changepoints may reflect **model specification choices** rather than genuine structural breaks. The relationship between weather and electricity demand may exhibit non-linearities, threshold effects, or interaction effects that create apparent structural breaks when the smooth GAM structure cannot fully accommodate these patterns.

Implications for Interpretation

The analysis reveals that detected changepoints fall along a spectrum of interpretability. At one end are clear model limitations, such as holiday detections in OLDB_1, which represent recurrent patterns the model should learn rather than genuine structural breaks. Pandemic impacts across all GSPs represent genuine behavioural changes, reflecting real shifts in consumption patterns driven by altered human behaviour during lockdown periods. Infrastructure transformations, exemplified by wind farm commissioning in BOAG_P, represent permanent system changes that fundamentally alter the relationship between weather variables and net demand. Finally, some weather-related or specification-dependent changepoints fall into ambiguous cases that require domain knowledge and external validation to distinguish between genuine shifts and model artefacts.

The challenge lies in distinguishing genuine structural breaks from model artefacts. Cross-referencing detected changepoints with known external events—such as policy announcements, infrastructure commissioning dates, or extreme weather records—is essential for proper interpretation. The BOAG_P example demonstrates that when domain knowledge (understanding wind farm operation) is applied to smooth term analysis, infrastructure changepoints can be confidently validated. This contextual investigation is particularly important for changepoints that do not coincide with obvious external events, as they may represent either subtle but genuine shifts in the demand-generating process or inadequacies in the model's ability to capture complex relationships between covariates and demand.

5.6.2 Methodological Limitations and Considerations

Despite the promising results, several limitations should be addressed. First, the minimum segment length was set to 20 or 30. If multiple changepoints occur within segments shorter than twice this length, the algorithm may fail to detect them, as those segments would be excluded from the search. Moreover, if the model becomes too complex, the algorithm may struggle with small segments due to insufficient data, highlighting the need to adjust the minimum segment length appropriately in such cases.

Another limitation relates to the reliability of the detected changepoints. Some true changepoints may be missed if their associated cost slightly exceeds the detection threshold, while other false positives may be identified. To improve the algorithm's sensitivity, adjusting the penalty size—either by multiplying it with a constant or varying it based on the data length—could help fine-tune the algorithm's responsiveness to real changepoints.

Among the algorithm setups tested, BIC-based methods consistently outperformed others, yielding more parsimonious models with fewer but more meaningful changepoints. This suggests that while AIC may overfit the data, BIC strikes a better balance between model complexity and accuracy, particularly for larger datasets. The preference for BIC-based approaches is further justified by the fact that many detected changepoints correspond to verifiable external events, suggesting that the more conservative BIC penalty effectively filters out spurious detections while retaining genuine structural breaks.

An important limitation of the current approach is the incomplete treatment of temporal correlation in electricity demand data. While the GAM framework captures seasonal patterns through sinusoidal components and accounts for day-of-week effects, it does not explicitly model the strong autocorrelation inherent in electricity demand time series. Electricity consumption exhibits significant temporal dependence, where today's demand is strongly correlated with recent historical values, and weather effects may persist beyond the immediate observation period. The changepoint detection algorithm assumes that observations within each segment are conditionally independent given the covariates, which may be violated when substantial temporal correlation remains in the residuals.

This limitation could potentially affect the accuracy of changepoint locations and lead to the detection of spurious changepoints where the model struggles to capture smooth temporal transitions. The frequent detection of changepoints during predictable periods, such as the Christmas season, may partly reflect this issue, as the model attempts to accommodate temporal patterns that are not adequately captured by the specified covariates. However, it is worth noting that in the case of holiday periods, the abrupt nature of the behavioral change (e.g., simultaneous factory shutdowns) may justify changepoint detection even in the presence of improved temporal modeling. Future research should investigate incorporating autoregressive components, lagged variables, or dynamic GAM structures that allow model parameters to evolve smoothly over time, potentially reducing the need for abrupt changepoints and providing more nuanced insights into the temporal evolution of electricity demand patterns.

5.6.3 Implications and Future Directions

The ability to distinguish between normal seasonal variations and abrupt shifts in demand offers valuable insights for energy management and planning. The successful detection of pandemic-related changepoints, in particular, demonstrates the algorithm's potential for identifying unforeseen disruptions to electricity systems, which could inform contingency planning and grid management strategies during future crises.

Future research could expand on this analysis by incorporating additional explanatory variables, such as economic activity indicators, fuel prices, or more granular weather data, to further improve the model's predictive power and reduce the occurrence of changepoints that merely reflect model inadequacy. Additionally, applying the changepoint detection framework to other regions or countries could offer a more comprehensive understanding of how different factors influence regional electricity demand across varying policy contexts and climatic conditions. Such insights would contribute to the development of more effective energy distribution strategies and support sustainability goals, particularly as electricity systems undergo transitions toward renewable generation and electrification of heating and transport.

Chapter 6

Conclusion

6.1 Key Findings

The study achieved significant insights and validation regarding the application of changepoint detection in modelling national electricity demand. The numerical results support the utility and robustness of the proposed approach:

1. **Comprehensive Modelling of National Demand**: This study provided a detailed understanding of national electricity demand in Chapter 2, accounting for variables such as temperature, wind speed, holiday patterns, and daily activity cycles. The model's ability to capture demand dynamics was confirmed with a Root Mean Square Error (RMSE) of 1528.78 when incorporating comprehensive predictors such as time, weather, and holiday effects, compared to an RMSE of 7334.84 for a 4-year baseline model without these features—representing a reduction in RMSE of 79.16%. These results underscore the importance of including both predictable elements, such as time-of-day usage patterns, and less predictable influences, like weather variations, for accuracy in demand forecasting.

2. **Development of a Changepoint Detection Algorithm**: A key contribution of this research is the creation of a changepoint detection algorithm optimised for complex, seasonally varying time series data, as discussed in Chapters 3 and 4. By integrating Generalised Additive Models (GAMs) with Binary Segmentation, the framework effectively differentiates between structural breaks and routine seasonal

variations. The approach was validated in its ability to detect multiple changepoints accurately; for instance, in simulations with reasonable mean shifts relative to noise levels, the algorithm successfully identified changepoints, achieving covering metrics and F1-scores close to 1.

3. **Regional Demand Analysis and Algorithm Application**: Applying the developed algorithm to regional demand data in Chapter 5 demonstrated its adaptability to varying local conditions. This regional analysis confirmed the algorithm's effectiveness in detecting changepoints in localised demand trends, influenced by factors like renewable energy integration or regional industrial shifts, and highlighted significant variation in demand characteristics across the UK. Compared to a baseline model without changepoints, OLDB_1 shows a reduction in BIC of 11.35%, PENN_1 shows a reduction of 5.76%, and BOAG_P shows the most substantial reduction of 27.29%, likely reflecting the impact of newly established wind farms in the region.

This thesis has presented an innovative framework for changepoint detection within the context of electricity demand modelling and forecasting. Given the dynamic nature of electricity demand patterns, influenced by seasonal, societal, and economic factors, detecting structural changes in demand data is essential for enabling timely adaptation to shifting consumption behaviours and supporting a resilient, efficient electricity grid.

6.2 Strengths and Limitations

Integrating Generalised Additive Models (GAMs) within the changepoint detection framework is a central strength of this work, enabling the automatic capture of complex non-linear trends and seasonal patterns across applications ranging from national to regional datasets. This approach allows the model to detect changepoints effectively without requiring extensive pre-specification of trends or seasonality, as GAMs inherently adjust for these underlying temporal patterns. As a result, the model can focus solely on identifying genuine changes, reducing the risk of misinterpreting regular patterns as changepoints. This capability not only simplifies the modelling process but also enhances accuracy, providing a reliable basis for detecting true changes in demand. Nonetheless, several limitations must also be acknowledged, both in the general methodology and in its specific application to Grid Supply Point (GSP) demand data.

One primary limitation lies in the requirements for minimum segment length and search increment. Setting a minimum segment length ensures each segment has enough data for accurate model fitting, but it also restricts the model's capacity to detect rapid, successive changepoints within short intervals. If the segment length is set too low, the algorithm may struggle to construct a reliable model due to inadequate data within each segment. Additionally, when the model lacks sufficient complexity to capture underlying patterns, a low segment length can introduce instability in the cost function. This instability arises from the segment length in the denominator, causing fluctuations in the cost function and reducing the reliability of changepoint detection.

Conversely, a longer segment length improves computational efficiency by reducing the frequency of model fitting; however, it risks missing changepoints within shorter segments (those shorter than twice the minimum length), potentially allowing important shifts to go undetected. Although reducing the minimum segment length would permit more granular detection—such as at a half-hourly level—this would come at the cost of significantly increased computational demands.

Similarly, adjusting the search increment—how frequently the model updates as it progresses through the data—also impacts efficiency and precision. A larger search increment, such as updating every three days rather than one day, speeds up computation substantially, though at the cost of some precision. If computational time is not a primary concern, setting the search increment to one unit (e.g., one day or one half-hour) maximises accuracy, allowing for finer detection of changepoints. However, this approach may not be feasible for applications with extensive data and time constraints.

Scalability and model selection also present notable challenges. Currently, the framework requires manual adjustments of model parameters for each dataset, including the specification of day-of-week effects, the periodic length of sinusoidal components for seasonal patterns, variable selection for weather and holiday effects, and other domain-specific features relevant to electricity demand forecasting. This manual parameter tuning process limits scalability when handling hundreds or thousands of time series, as each dataset may require customised model specifications rather than employing a universal model formulation. Automating the model selection process could improve the framework's efficiency and usability in large-scale applications where manual parameter adjustments for individual time series would be impractical. However, balancing model complexity is crucial: a simpler model may fail to capture data intricacies, leading to overfitting, whereas a more complex model would increase computational demands and require a longer minimum segment length to accommodate additional variables.

Feature selection adds further complexity, particularly with features like the New Year season that do not appear in every segment. Attempting to fit a model with a New Year variable in segments that lack relevant data can cause errors. To address this, one could use different models—for instance, including the New Year variable only where relevant. However, this approach necessitates careful adjustments to the cost function and penalty choices to account for the variation across models.

The assumption of Gaussian distribution is another simplification with inherent limitations. While this assumption facilitates model interpretation, real-world data often deviate from normality, especially in cases of skewed or heavy-tailed distributions. Such deviations may reduce the model's ability to capture extreme events or shifts outside typical demand patterns, thus impacting accuracy.

Computational time is a practical constraint, particularly with large datasets. The iterative model fitting required for accurate changepoint identification imposes a high computational burden, making the approach less suitable for scenarios that require rapid or real-time results.

In the specific application to GSP demand data, daily averages are used rather than half-hourly data to manage computational demands. While daily averages allow for more efficient analysis, they limit the model's applicability in contexts that require fine-grained, sub-daily detail, which could better support short-term operational adjustments. Additionally, while the model has been validated in offline settings, it is not currently adapted for online changepoint detection, where real-time processing is necessary to identify changepoints in streaming data. This limitation is particularly relevant for dynamic electricity systems, where real-time adjustments are critical.

While this methodology marks a significant advancement in changepoint detection for electricity demand forecasting, these limitations highlight that no model is flawless. Recognising these aspects encourages future refinements to enhance the model's scalability, efficiency, and robustness in capturing meaningful structural changes.

6.3 Implications and Future Work

The results of this study offer important implications for electricity demand forecasting and grid management, particularly in the context of integrating more renewable energy sources and managing variability in demand. Enhanced changepoint detection not only improves forecasting precision but also supports more efficient energy allocation, infrastructure planning, and system stability.

Several directions for future research and enhancement are suggested:

1. **Online Changepoint Detection for Real-Time Data Streams**: To enable changepoint detection in an online setting, the framework could be adapted to process new incoming data using a moving time window as information arrives. This approach aims to allow for early detection of shifts in demand, enabling the model to respond to changes in real time. By sliding the data window forward as new data streams in, the model could become more responsive to abrupt shifts, which is particularly relevant for sudden events.

2. **Multivariate and Cross-Regional Detection Capabilities**: The current framework processes each time series independently, analysing individual locations or grid cells in isolation without considering spatial correlations or interdependencies between regions. Extending the model to simultaneously detect changepoints across multiple regions or variables could enhance its robustness in complex, interconnected systems. For instance, an increase in load in one region might correspond with a decrease in another, reflecting a shift in consumption patterns that would benefit from a multivariate approach that captures these cross-regional dynamics. By detecting common or connected changepoints across multiple grid cells and regions,

the model could better capture these spatial interdependencies and improve overall grid resilience through a more holistic understanding of demand shifts across the electricity network.

3. **Expanding Predictive Models Beyond GAMs**: While this thesis used GAMs due to their flexibility in modelling non-linear trends, other predictive models could be incorporated to broaden the framework's applicability. Other models, ARIMA, and machine learning techniques like neural networks may offer advantages for different types of time series data. While machine learning methods may handle non-linearity well, they often lack interpretability and can be computationally intensive, requiring careful consideration when balancing predictive power with efficiency.

In conclusion, this thesis provides a novel approach to changepoint detection within electricity demand forecasting, addressing a critical need for adaptive and regionally nuanced models in energy management. The proposed framework offers advancements in forecasting methodology with potential for broader application in sectors reliant on time series data. Future work incorporating online detection, multivariate analysis, and alternative predictive models promises to further enhance the UK's ability to manage a sustainable, resilient energy landscape.

Appendix A

Approaches to Calculate a Cost Function in Simulation Study

Five approaches will be studied in this appendix to find the best approach to apply to the real data.

Model without smooth terms

For a model without smooth terms, this model will be used in a model-building process:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{4.8}\right) + \beta_3 \cos\left(\frac{2\pi t}{4.8}\right) + \beta_4 \sin\left(\frac{2\pi t}{4.8 \times 7}\right) + \\ \beta_5 \cos\left(\frac{2\pi t}{4.8 \times 7}\right) + \beta_6 \sin\left(\frac{2\pi t}{4.8 \times 365}\right) + \beta_7 \cos\left(\frac{2\pi t}{4.8 \times 365}\right) + \epsilon_t \quad \text{(A.1)}$$

where $t \in \{1, \ldots, 2000\}$ and $\epsilon_t \sim N(0, \sigma^2)$. This linear model is chosen to fit the simulated data as it captures every characteristic of the dataset. The AIC penalty is applied in the algorithm. The methods to calculate the cost using this formula are:

A1 **Refitting via `lm` function**

In each grid search step of the AMOC algorithm, two linear models are built to fit $y_{1:\tau}$ and $y_{\tau+1:n}$ from the beginning using the `lm` function in R. This approach is expected to be the slowest because it refits two models in each iteration while being the best in terms of accuracy.

A2 **Updating QR factorisation**

Instead of refitting new models in each iteration, an updating approach could be considered to help improve the overall computational cost. Golub and Van Loan (1996) provides detailed explanations on how to obtain the new QR factorisation $QR = A \in \mathbb{R}^{m \times n}$ when a row is appended or removed from the model matrix A using rotations. Thus, the QR factorisation only needs to be updated rather than refitting a model from scratch. This is expected to be faster than the refitting approach while yielding the same fitted values.

A3 **Calculate the model matrix directly**

Instead of using the `lm` function in R to fit a model, a direct calculation can be performed to avoid unnecessary processes behind the function call. Specifically, the fitted values can be calculated by $\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y$, where X is the design matrix, and y is the response vector. This method should perform faster than fitting via `lm` and provide the same results. However, a function implemented with a C backend, which is `lm`, can achieve greater efficiency due to C's ability to directly manage memory and execute compiled machine code, which results in faster execution compared to R's interpreted environment.

A4 **Refitting via `bam` function**

Similar to refitting via the `lm` function, the `bam` function will be used instead, without utilising any smooth terms, with a Gaussian family and an identity link function. The `bam` function is designed for building a model for a large dataset, which could be applied to the real data later.

A5 **Updating model using `bam.update`**

Instead of refitting with `bam`, `bam.update` (Wood, 2017) offers an updating approach from an old model object and a new set of data. Repeating this process multiple times should significantly reduce the computational cost compared to refitting via the `bam` function.

Model with Smooth Terms

For a model formula with smooth terms, the following model formula will be used:

$$g(E(y_t)) = \beta_0 + \beta_1 t + f_1\left(\left\{\frac{t}{4.8}\right\}\right) + f_2\left(\left\{\frac{t}{4.8 \times 7}\right\}\right) + f_3\left(\left\{\frac{t}{4.8 \times 365}\right\}\right) + \epsilon_t, \quad \text{(A.2)}$$

where $t \in \{1, \ldots, 2000\}$, $\epsilon_t \sim N(0, \sigma^2)$, g represents a link function, and $\{\cdot\}$ denotes the fractional part function, i.e., $\{x\} = x - \lfloor x \rfloor$. The functions f_1, f_2, and f_3 are smooth functions. In accordance with the optimal choice identified in the previous chapter, we assume that y_t follows a Gaussian distribution with an identity link function, and cubic regression splines are employed to fit the models.

Two approaches, B1 and B2, will be implemented using this model formula, corresponding to approaches A4 and A5, respectively.

B1 **Refitting via `bam` function**

Similar to refitting via the `lm` function, the `bam` function will be used instead, without utilising any smooth terms, with a Gaussian family and an identity link function. The `bam` function is designed for building a model for a large dataset, which could be applied to the real data later.

B2 **Updating model using `bam.update`**

Instead of refitting with `bam`, `bam.update` (Wood, 2017) offers an updating approach from an old model object and a new set of data. Repeating this process multiple times should significantly reduce the computational cost compared to refitting via the `bam` function.

Comparisons between methodologies

In this section, different methods (A1, A2, A3, A4, A5, B1, B2) will be compared to identify the most appropriate method for constructing the cost function, which will later be implemented with real data. The simulated data from Equations 4.1 and 4.2, with $\mu = 1.41, \mu_1 = \mu_2 = 2$, will be generated five times using different random noise, which will be regenerated for each simulation with $\sigma^2 = 1.01^2$ in the

Method	Bias	MAE	RMSE	Time (s)
A1	-1	6.6	9.85	0.45
A2	-3.2	6	10.33	135.60
A3	-1	6.6	9.85	0.14
A4	-1	6.6	9.85	2.94
A5	-1	6.6	9.85	2.03
B1	-	-	-	33.2
B2	941	941	941	3.39

Table A.1: The evaluation metrics from the single-changepoint simulation.

Method	Covering Metric	F1-score	Time (s)
A1	0.89	0.57	45.35
A2	0.89	0.57	588.69
A3	0.80	0.61	13.13
A4	0.89	0.57	401.61
A5	0.89	0.57	245.54
B1	0.99	0.92	1226.42
B2	0.33	-	82.01

Table A.2: The evaluation metrics from the multiple-changepoint simulation.

single-changepoint simulation and $\sigma^2 = 0.01^2$ in the multiple-changepoint simulation. The average accuracy and computation time of each method will then be compared.

For the single-changepoint simulation, where the true changepoint is located at 1000, the results are presented in Table A.1, while for the multiple-changepoint simulation, with true changepoints at 700 and 1300, the results are shown in Table A.2.

The results indicate that method A3 outperforms other methods in terms of computation time for the single-changepoint simulation, while method A2 has the longest computation time due to the computational expense of performing multiple QR decompositions. All methods achieve optimal accuracy, except for B1 and B2, which fail to identify a changepoint. This failure is likely due to the added model complexity from incorporating smooth terms, which increases the model's flexibility. Introducing smooth terms significantly increases computation time. While refitting the model fails to detect a changepoint, `bam.update` operates faster but does not accurately identify the changepoint; as the model updates repeatedly, its approximation increasingly diverges from the rebuilt model, ultimately resulting in a substantially different model structure.

For the multiple-changepoint simulation, method A3 is the fastest, although this speed comes at a slight reduction in accuracy. If accuracy is paramount, method A1 remains the best choice, delivering the highest accuracy and the fastest computation time among methods that achieve comparable accuracy. The addition of smooth terms in the multiple-changepoint scenario yields results consistent with those of the single-changepoint study. Notably, method B1 excels in prediction, as demonstrated by its near-perfect average coverage metric and F1-score, surpassing the performance of other methods.

In conclusion, for enhanced model-fitting capability, method B1 is recommended for use with real-world data, despite its higher computation time, which is considered a worthwhile trade-off.

References

Adams, R. P. and D. J. C. MacKay (2007). Bayesian online changepoint detection. 9

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723. 50

Alwan, L. C. and H. V. Roberts (1988). Time-series modeling for statistical process control. Journal of Business & Economic Statistics 6(1), 87–95. 8

Amiri, A. and S. Allahyari (2012). Change point approach for control chart post-signal diagnostics: A literature review. Quality and Reliability Engineering International 28(7), 673–685. 8

Ardito, L., G. Procaccianti, G. Menga, and M. Morisio (2013). Smart grid technologies in Europe: An overview. Energies 6(1), 251–281. 3

Atkinson, A. C., M. Riani, and A. Corbellini (2021). The Box–Cox transformation: Review and extensions. Statistical Science 36(2), 239 – 255. 20

Aue, A. and L. Horváth (2013). Structural breaks in time series. Journal of Time Series Analysis 34(1), 1–16. 7

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. Econometrica 66(1), 47–78. 7

Bersimis, S., M. Panaretos, and S. Psarakis (2007). Multivariate statistical process control charts and the problem of interpretation. Quality and Reliability Engineering International 23(5), 517–543. 8

Casini, A. and P. Perron (2018). Structural breaks in time series. 7

Chen, J. and A. Gupta (2000). Parametric Statistical Change Point Analysis. Deutsche Mathematiker-Vereiniung. Birkhäuser. 49, 57

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica 28(3), 591–605. 7

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics 20(1), 134–144. 32

Ferrari, A., C. Richard, A. Bourrier, and I. Bouchikhi (2023). Online change-point detection with kernels. Pattern Recognition 133, 109022. 9

Gaillard, P., Y. Goude, and R. Nedellec (2016). Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. International Journal of Forecasting 32(3), 1038–1050. 5

Gneiting, T. (2011). Making and evaluating point forecasts. Journal of the American Statistical Association 106(494), 746–762. 44

Golub, G. H. and C. F. Van Loan (1996). Matrix Computations (3rd ed.). Johns Hopkins Studies in Mathematical Sciences. Baltimore, MD: Johns Hopkins University Press. 123

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57(2), 357–384. 44

Hawkins, D. M., P. Qiu, and C. Kang (2003). The changepoint model for statistical process control. Journal of Quality Technology 35(4), 355–366. 8

Hong, T. (2010). Short Term Electric Load Forecasting. Ph. D. thesis, North Carolina State University, Raleigh, NC. Doctor of Philosophy in Operations Research and Electrical Engineering. 5

Hong, T. and S. Fan (2016). Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting 32(3), 914–938. 5

Hong, T., P. Pinson, and S. Fan (2014). Global energy forecasting competition 2012. International Journal of Forecasting 30(2), 357–363. 5, 6

Hong, T., P. Pinson, S. Fan, H. Zareipour, M. Tahanan, A. Troccoli, and R. J. Hyndman (2016). Global energy forecasting competition 2014 and beyond. International Journal of Forecasting 32(3), 896–913. 5, 6

Huang, H., J. M. Winter, E. C. Osterberg, R. M. Horton, and B. Beckage (2017). Total and extreme precipitation changes over the northeastern United States. Journal of Hydrometeorology 18(6), 1783 – 1798. 10

Inclan, C. and G. C. Tiao (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association 89(427), 913–923. 49

Jackson, B., J. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai (2005). An algorithm for optimal partitioning of data on an interval. IEEE Signal Processing Letters 12(2), 105–108. 9, 52

Janczura, J., S. Truck, and R. Weron (2013). Modeling electricity price spikes: A comparison between regime-switching models. In F. E. Benth, V. Kholodnyi, and P. Laurence (Eds.), Quantitative Energy Finance, pp. 147–176. New York: Springer. 44

Janczura, J. and R. Weron (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Economics 32(5), 1059–1073. 44

Jiménez-Herrera, P., L. Melgar-García, G. Asencio-Cortés, and A. Troncoso (2020). A new forecasting algorithm based on neighbors for streaming electricity time series. In E. A. de la Cal, J. R. Villar Flecha, H. Quintián, and E. Corchado (Eds.), Hybrid Artificial Intelligent Systems, Cham, pp. 522–533. Springer International Publishing. 6

Joe Qin, S. (2003). Statistical process monitoring: basics and beyond. Journal of Chemometrics 17(8-9), 480–502. 8

Khazaie Poul, M., H. Farughi, and Y. Samimi (2024). Spatiotemporal monitoring of electricity consumption by using residual-based control charts. Journal of Energy Management and Technology 8(1), 35–51. 8

Killick, R., P. Fearnhead, and I. A. Eckley (2012). Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association 107(500), 1590–1598. 9, 53, 54, 57

Knoth, S. and W. Schmid (2004). Control charts for time series: A review. In H.-J. Lenz and P. W. Wilrich (Eds.), Frontiers in Statistical Quality Control 7, pp. 210–236. Berlin: Springer. 8

Koopman, S. J., M. Ooms, and M. A. Carnero (2007). Periodic seasonal reg-arfima–garch models for daily electricity spot prices. Journal of the American Statistical Association 102(477), 16–27. 44

Laouafi, A., M. Mordjaoui, S. Haddad, T. E. Boukelia, and A. Ganouche (2017). Online electricity demand forecasting based on an effective forecast combination methodology. Electric Power Systems Research 148, 35–47. 6

Laurinec, P. (2017). Analyzing double seasonal time series with GAM in R. Personal blog, accessed 1 Oct 2025. 43

Luís, G., J. Esteves, and N. P. da Silva (2020). Energy forecasting using an ensamble of machine learning methods trained only with electricity data. In 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), pp. 449–453. 6

Marcjasz, G., B. Uniejewski, and R. Weron (2019). On the importance of the long-term seasonal component in day-ahead electricity price forecasting with narx neural networks. International Journal of Forecasting 35(4), 1525–1538. 43

Marcjasz, G., B. Uniejewski, and R. Weron (2020). Beating the naïve: Combining lasso with naïve intraday electricity price forecasts. Energies 13(7), 1667. 45

Mitali, J., S. Dhinakaran, and A. Mohamad (2022). Energy storage systems: a review. Energy Storage and Saving 1(3), 166–216. 3

Montgomery, D. C. (2020). Introduction to Statistical Quality Control (8th ed.). Hoboken, NJ: John Wiley & Sons. Comprehensive textbook on SPC, including Shewhart, CUSUM, and EWMA charts. 8

Pebesma, E., T. Mailund, et al. (2021). sf: Simple Features for R. R package version 1.0-2. 20

Perron, P. (2005). Dealing with structural breaks. Palgrave handbook of econometrics, volume 1: econometric theory. 7

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. The Review of Economic Studies 73(4), 1057–1084. 7

Rigaill, G. (2015). A pruned dynamic programming algorithm to recover the best segmentations with 1 to k_{max} change-points. Journal de la Société Française de Statistique 156(4), 180–205. 49

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics 6(2), 461 – 464. 50

Scott, A. and M. Knott (1974). A cluster analysis method for grouping means in the analysis of variance. Biometrics 30, 507. 9

Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. Frontiers in Ecology and Evolution 6, 149. 44

Sobhani, M., P. Wang, and T. Hong (2023). Detecting load transfers. IEEE Transactions on Smart Grid 14(2), 1367–1375. 10

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. Journal of the Operational Research Society 54(8), 799–805. 6

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. European Journal of Operational Research 204(1), 139–152. 45

van den Burg, G. J. J. and C. K. I. Williams (2020). An evaluation of change point detection algorithms. ArXiv abs/2003.06222. 54, 56

Wang, P., B. Liu, and T. Hong (2016). Electric load forecasting with recency effect: A big data approach. International Journal of Forecasting 32(3), 585–597. 6

Wang, R., M. Zhu, X. Zhang, and H. Pham (2023). Lithium-ion battery remaining useful life prediction using a two-phase degradation model with a dynamic change point. Journal of Energy Storage 59, 106457. 10

Wang, Z., X. Lin, A. Mishra, and R. Sriharsha (2021). Online changepoint detection on a budget. In 2021 International Conference on Data Mining Workshops (ICDMW), pp. 414–420. 9

Wardell, D. G., H. Moskowitz, and R. D. Plante (1994). Run-length distributions of special-cause control charts for correlated processes. Technometrics 36(1), 3–17. 8

Wood, S. N. (2017). Generalized Additive Models: An Introduction with R (2 ed.). New York: Chapman and Hall/CRC. 5, 27, 29, 43, 45, 123, 124

Wood, S. N., Y. Goude, and S. Shaw (2014). Generalized additive models for large data sets. Journal of the Royal Statistical Society Series C: Applied Statistics 64(1), 139–155. 30

Wood, S. N., N. Pya, and B. Säfken (2016). Smoothing parameter and model selection for general smooth models. Journal of the American Statistical Association 111(516), 1548–1563. 44

Yang, M., C. Han, Y. Cui, and Y. Zhao (2021). Covid-19 and mobility in tourism cities: A statistical change-point detection approach. Journal of Hospitality and Tourism Management 47, 256–261. 10

Yourstone, S. A. and D. C. Montgomery (1989). A time-series approach to discrete real-time process quality control. Quality and Reliability Engineering International 5(4), 309–317. 8

Zeileis, A., F. Leisch, K. Hornik, and C. Kleiber (2002). Strucchange: An R package for testing for structural change in linear regression models. Journal of Statistical Software 7, 1–38. 7

Zhang, G., L. Lei, X. Yang, and J. Li (2020). Criteria for classifying forecasting methods. International Journal of Forecasting 36(1), 167–177. 4

Ziel, F. (2018). Modeling public holidays in load forecasting: a German case study. Journal of Modern Power Systems and Clean Energy 6(2), 191–207. 6