Extreme value modelling for episodes of poor air quality

Kyle Lambert-Betts, MSci (Hons.)

This thesis is submitted for the degree of

Doctor of Philosophy

at the School of Mathematical Sciences, Lancaster University

July, 2025

Abstract

Standard statistical modelling approaches are typically biased in tail behaviour estimation due to the modelling parameters being driven by the main body of the distribution. As such, the area of extreme value theory provides an asymptotically justified approach to model the probabilistic behaviour of rare events. Extreme value methods are used in a wide range of applications, for example they are often used for modelling storm surges in hydrology and heatwaves in medical statistics.

Most of our work has been motivated by the well-established negative impacts of poor air quality, more specifically, extreme episodes of ozone concentrations on human health. Our focus has been on a multivariate ozone dataset, which shows complex temporal and spatial trends. We begin by proposing the use of extreme value theory to validate numerical process-based model forecasts. Next, we examine the temporal dependence structure and propose a new measure to determine the order of an extreme Markov process. Then we present novel applications of the extreme Markov processes, simulating extremal chain behaviour of key health related scenarios and short-lead-time forecasts of extreme events. Finally, we present a spatial model to evaluate the risk of extreme events of ozone across Great Britain.

Acknowledgements

First of all, I would like to thank my supervisory team of Emma Eastoe, Ryan Hossaini and Oliver Wild, for their expert guidance and support throughout the PhD. I have felt fully supported and encouraged to pursue new research ideas and have wholeheartedly enjoyed the process due to their positivity. I am especially grateful to Emma for approaching me about the PhD opportunity during my Msci at Lancaster. Further, I appreciate the countless time you have spent reading through my many iterations of work and consistently providing insightful and thought-provoking feedback.

I would also like to acknowledge the support of the UK Meteorological Office for the provision of the meteorological data used in Chapter 4. I would like to thank my industrial supervisor Lucy Neal, for her vast meteorological expertise and patience with explaining new concepts, as well as providing detailed amendments to our project. I hope we have the opportunity in the near future to work on another project together.

I would like to thank the people of the School of Mathematical Sciences at Lancaster University for an enjoyable time, and in particular thanks to my fellow PhD students. Special mention to Alison, Björn, Isa, Jakub, Lauren, Luke, Marco, Noelia, Rebecca, Rowan, and Yasmin for making my on campus working life filled with joy and laughter.

For my time outside of work, I thank the members of the bar sports community for providing me an avenue to escape work and develop as a darts player. Special thanks to Mitchel, Nikhilean and Nathaniel for always making me feel welcome and

celebrating with me.

I would like to thank the James' for *adopting* me into their family, celebrating with me when things were going well and supporting me when things were tough. And finally, I would like to thank my family for being a constant source of encouragement and support during my many years at Lancaster.

Declaration

I declare that the work presented in this thesis is my own, except where stated otherwise, and has not been submitted elsewhere for the award of a higher degree.

The computing code used has mostly been my own, although have made use of the R libraries openair, rnaturalearth and elevatr to obtain the ozone data and geographic covariates and the R libraries ismev and evd.

Kyle Lambert-Betts

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Thesis outline

This thesis aims to use extreme value theory to model different characteristics of a pollutant contributing to poor air quality, tropospheric ozone, across the United Kingdom. The ideas and applications presented in this thesis could be used for policy making and protecting public health. The structure of this thesis is as follows:

Chapter 2 provides an overview of tropospheric ozone. We begin by discussing the formation of ozone and the impact high concentrations of ozone has on public health and infrastructure. Then we detail the changing nature of guidelines and legislations linked with managing air quality. We conclude this chapter by introducing the two datasets we will use throughout the thesis: ozone observations from the Automatic Rural and Urban Network (AURN) and ozone forecasts from the process-based numerical model of the UK Meteorological (Met) Office, the Air Quality in the Unified Model (AQUM).

Chapter 3 can be seen as a literature review introducing both the basic concepts of extreme value theory and the corresponding inference. We start by deriving the extreme value model under strict assumptions about the data. We then present the concept of extremal dependence and detail multivariate extreme value methods which can capture different types of extremal dependence structures.

In Chapter 4 we explore the capabilities of the AQUM forecasts at capturing the extreme values of ozone. We propose the use of univariate extreme value theory and hypothesis tests to investigate if the type of distributional tail is captured and if both the observations and forecasts come from the same distribution. This contrasts to existing methods which predominantly focus on the full distribution. Then we explore the joint behaviour of extreme ozone from AURN and AQUM and estimate probabilities of successfully forecasting extreme events using the bivariate extreme value distribution.

The question of how to analyse temporal dependence is tackled in Chapter 5. We propose the use of a kth-order Markov extremes model, as presented by Winter and Tawn (2017), to capture the asymptotic (in)dependence nature of extreme ozone episodes. In contrast, we propose a new measure to determine the order of the model which accounts for multiple characteristics of the process. After selecting the appropriate model order, we demonstrate the functionalities of the approach by forecasting through simulation the events of 2020.

In Chapter 6 we present a spatial model to evaluate the risk of extreme events of ozone. This is done by using a spatial Bayesian hierarchical model to capture the spatial variation in the marginal parameters of the generalised Pareto distribution (GPD). In contrast to previous work in the literature, we propose to fix the rate of exceedance and model the spatial structure of the thresholds through traditional geostatistical methods. Geographical covariates are incorporated into the spatial structure of the marginal GPD parameters and return level maps are generated through interpolation of the model.

Chapter 7 concludes by summarising the contributions of this thesis and discussing potential avenues for further work.

Chapter 2

Ozone background and data

2.1 Ozone

Ozone (O_3) is a natural constituent of the atmosphere and is present in both the stratosphere and the troposphere. Although ozone is transported between the stratosphere and the troposphere, the two are largely separate regimes (The Royal Society, 2008). Stratospheric ozone, i.e. the ozone layer, is produced following the photolysis of molecular oxygen and provides protection by filtering out dangerous ultraviolet (UV) radiation from the sun. Tropospheric ozone is not emitted directly into the air, but is a secondary, trans-boundary air pollutant (WHO, 2003). It is created in the air through photochemical reactions between anthropogenic nitrogen oxides ($NO_X = NO_2 + NO$) and volatile organic compounds (VOCs). These reactions occur when precursor pollutants emitted by combustion of fossil fuels by cars, power plants and other sources are in the presence of sunlight. When ozone is formed at the tropospheric level, it can remain at ground-level from between a few hours and a few days depending on the meteorological conditions. Slow-moving high-pressure systems with clear skies and elevated temperatures allow for increased photochemical reactions and the accumulation of ozone, while high winds and fast trans-boundary movement reduces the localised ozone concentration but can increase precursor and ozone concentrations downwind of the pollutant source (Guicherit and van Dop,

1977; Vukovich et al., 1977). Further, ozone concentrations vary spatially and temporally on various time scales due to the heterogeneity in ozone sources and sinks, meteorological variability, and trends in precursor emissions (Cooper et al., 2014).

Throughout the thesis we use *ozone* to refer to all levels of tropospheric ozone. However, three levels of ozone require definitions: *background ozone*, *ambient ozone* (air) and *peak-level ozone*. Background ozone refers to what the average concentration of ozone would be without any human influence. Ambient ozone (air) is the concentration (air) that occurs at a particular time and place outside of built structures. Peak-level ozone refers to the largest concentrations over a particular time, for example daily maxima.

2.2 Impacts

Many atmospheric pollutants, like ozone and nitrogen oxides, are not only important climate forcing agents (Andreae et al., 2005; Arneth et al., 2009; Forster et al., 2007) but pertain to negative effects on public health, agriculture and ecosystems if concentrations are high. The people who are most at risk of incurring health complications due to poor air quality include: sufferers of asthma, children, older adults, active outdoor workers and those with certain genetic characteristics and nutrient deficiencies. Short-term exposure to high ozone can lead to: coughing and scratchy throat, difficulty breathing and aggravate pre-existing breathing issues like asthma, emphysema and chronic bronchitis (Bell et al., 2004; Whitfield et al., 1996). Prolonged exposure to high concentrations of ozone can cause the development of asthma and other respiratory illnesses as well as potentially death (Bell et al., 2005). The risk of health implications is exacerbated during heatwaves due to the increase in emissions, which can result in a 20% increase in concentrations in urban areas (Hou and Wu, 2016). For an extensive review of health effects of ozone see Devlin et al. (1997). In the EU, approximately $21,400$ die prematurely each year due to

excess exposure to ozone (EEA and WHO, 2007) and approximately one million per annum globally. Furthermore, there is an estimated 7.7 million deaths per annum due to other air pollutants (Lelieveld et al., 2019; Vohra et al., 2021), mainly high concentrations of particulate matter less than $2.5\mu m$ in diameter ($PM_{2.5}$). It is estimated that poor air quality health implications cost the EU €166 billion annually (CE Delft, 2018).

Prolonged exposure to high concentrations of ozone can cause significant problems to the environment. Persistent exposure increases the risk of: biomass reduction in sensitive plant species due to reduced capability to photosynthesise, reduction in biodiversity and reduction in carbon uptake (Arneth et al., 2010; Mercado et al., 2009; Sitch et al., 2007; Pleijel et al., 2014; Simpson et al., 2014). These result in positive feedback which stimulates further global warming and in turn causes vegetation to become more susceptible to other environmental stresses such as high winds, extreme temperatures, pests and diseases (GCNTO, 2022). Further consequences include negative impacts on local water and nutrient cycles. Ozone sensitive crops have also been affected recently, resulting in a reduction in crop yield. For example, an estimated €6.7 billion was lost due to impacts on arable crops in the EU during 2000 (Holland et al., 2006).

The interactions between: changes in nitrogen deposition, increases in atmospheric CO_2 concentration, changes in aerosol burdens and increases in both background and peak-levels of ozone are causing greater risk of further negative effects to occur. Therefore, changing anthropogenic emissions is essential in improving air pollution and the impacts of climate change by changing, in a complex system, interacting feedbacks (Arneth et al., 2010; Raes et al., 2010). The following section provides details of legislation that provides support to improve air quality.

2.3 Legislation

The World Health Organisation (WHO) provides air quality guidelines (AQGs) to support regional and national actions to achieve a level of air quality that protects public health. These guidelines are based on the extensive body of scientific evidence relating air pollution to their health consequences, see WHO (2021). The guidelines were first introduced in 1987 and have been revised since, once in 2005 and again in September 2021. The latest revision was due to new-found evidence that showed even lower concentrations are having a greater impact on human health and well-being than once thought (Vohra et al., 2021).

Ozone guidelines were initially set at $120\,\mu g/m^3$, based on the daily maximum of eight-hour running mean (DM8), which got reduced to $100\,\mu g/m^3$ in 2005. This value did not change in the 2021 revision, however, an additional guideline was added stating that the six consecutive months of a year with the highest six-month running average should not exceed $60\,\mu g/m^3$ (Huangfu and Atkinson, 2020). The precursors to ozone did receive dramatic reductions to their guideline concentrations in the 2021 revision. These changes were influenced by the background ozone concentration doubling between the late 19th century and 1980 from $30\,\mu g/m^3$ to $60 - 70\,\mu g/m^3$, which has since increased by another $10 - 20\,\mu g/m^3$, and experiences of health risks at $< 70\,\mu g/m^3$ (Huangfu and Atkinson, 2020; WHO, 2021). Reduction in precursor concentrations will in turn reduce the background ozone and peak ozone concentrations. Furthermore, these revisions are required to help improve the air quality since 99% of the global population are living in places where the WHO air quality guideline levels are not being met (WHO, 2022).

Although the WHO has set these guidelines, each country sets their own air quality standards (National ambient air quality standards, NAAQS) to further protect the health of their citizens. These do vary between countries due to the approaches adopted to balance health risks, politics, and impacts to the economy and the environment. Consequently, the air quality standards differ resulting in different air quality indices and different threshold concentrations associated with each risk

band. Table 2.1 shows the differences between ozone standards in India, Singapore, USA and the UK.

India	Good 0-50	Satisfactory 51-100	Moderately Polluted 101-168	Poor 169-208	Very Poor 209-748	Severe >748

Singapore	Good 0-118	Moderate 119-157	Unhealthy 158-235	Very Unhealthy 236-785	Hazardous 786-1180

USA	Good 0-109	Moderate 110-140	Unhealthy 141-210	Very Unhealthy 211-400	Dangerous >400

UK	Low 0-100	Moderate 101-160	High 161-240	Very High >240

Table 2.1: Showing the differences of ozone standards across countries. Note: all of these values are based on the daily maximum of the 8-hour running mean (DM8), apart from 'Hazardous' in Singapore which is based on a one-hour average.

The UK uses the Daily Air Quality Index (DAQI), which is a four-band system approved by the Committee on Medical Effects of Air Pollutants (COMEAP). The system derives the risk index, scaled 1 (low) to 10 (very high), from concentrations of five key air quality pollutants: ozone, nitrogen oxide, sulphur dioxide, particulate matter less than $2.5\,\mu m$ in diameter ($PM_{2.5}$) and particulate matter less than $10\,\mu m$ in diameter (PM_{10}), (DEFRA and UK-AIR, 2022). Each pollutant has its own average period and threshold levels (Connolly et al., 2013). For ozone, the average period is 8-hours with thresholds of $100\,\mu g/m^3$, $160\,\mu g/m^3$ and $240\,\mu g/m^3$ for when the public health risk goes from low $(1-3)$ to moderate $(4-6)$, moderate to high $(7-9)$ and high to very high (10), respectively. DAQIs for all five pollutants can be found in Table 2.2. The overall index is calculated as the maximum of the five pollutant DAQIs. Defra (Department of Environment, Food and Rural Affairs) alongside the UK Met Office produces air quality forecasts so vulnerable individuals can plan their day based on the health advice associated with the forecasted risk level.

Band	**Index**	**O_3**	NO_2	SO_2	$PM_{2.5}$	PM_{10}
		Running 8h mean ($\mu g/m^3$)	Hourly mean ($\mu g/m^3$)	15 min mean ($\mu g/m^3$)	24h mean ($\mu g/m^3$)	24h mean ($\mu g/m^3$)
Low	**1**	**0 – 33**	0 – 67	0 – 88	0 – 11	0 – 16
	2	**34 – 65**	68 – 134	89 – 177	12 – 23	17 – 33
	3	**66 – 100**	135 – 200	178 – 266	24 – 35	34 – 50
Moderate	**4**	**101 – 120**	201 – 267	267 – 354	36 – 41	51 – 58
	5	**121 – 140**	268 – 334	355 – 443	42 – 47	59 – 66
	6	**141 – 160**	335 – 400	444 – 532	48 – 53	67 – 75
High	**7**	**161 – 187**	401 – 467	533 – 710	54 – 58	76 – 83
	8	**188 – 213**	468 – 534	711 – 887	59 – 64	84 – 91
	9	**214 – 240**	535 – 600	888 – 1064	65 – 70	92 – 100
Very High	**10**	**> 241**	> 601	> 1065	> 71	> 101

Table 2.2: Daily Air Quality Index (DAQI) for five pollutants: ozone, nitrogen dioxide, sulphur dioxide, $PM_{2.5}$ (particulate matter less than $2.5\,\mu m$ in diameter) and PM_{10} (particulate matter less than $10\,\mu m$ in diameter) respectively.

2.4 Ozone data

2.4.1 Observations - Automatic Rural and Urban Network (AURN)

To monitor and assess the air quality across the country, measurement sites are employed. Many networks exist to measure pollutant concentrations and meteorological activity across the UK, for example the London Atmospheric Emissions Inventory (LAEI, 2024) and the Scottish Air Quality Database (SAQD, 2024). However, our focus is on the Automatic Rural and Urban Network (AURN) (Defra, 2020) since it is the largest automatic monitoring network in the UK and is the main network used for reporting compliance against the Ambient Air Quality Directives. Further, it provides a nationwide dataset which includes sites within large cities that also have their own local networks (London, Manchester, etc). This network measures a wide range of pollutants: ozone, nitric oxide, nitrogen dioxide,

sulphur dioxide, carbon monoxide, $PM_{2.5}$ (hourly and daily), PM_{10} (hourly and daily), non-volatile $PM_{2.5}$, non-volatile PM_{10}, volatile $PM_{2.5}$, volatile PM_{10}. The network also provides numerical model estimates for several meteorological variables: wind speed, wind direction and temperature. The network has historical data which goes back to July 1972. However, not all sites are in operation for the entire period due to malfunctions, replacements and introduction of new sites. Further not all sites measure all variables.

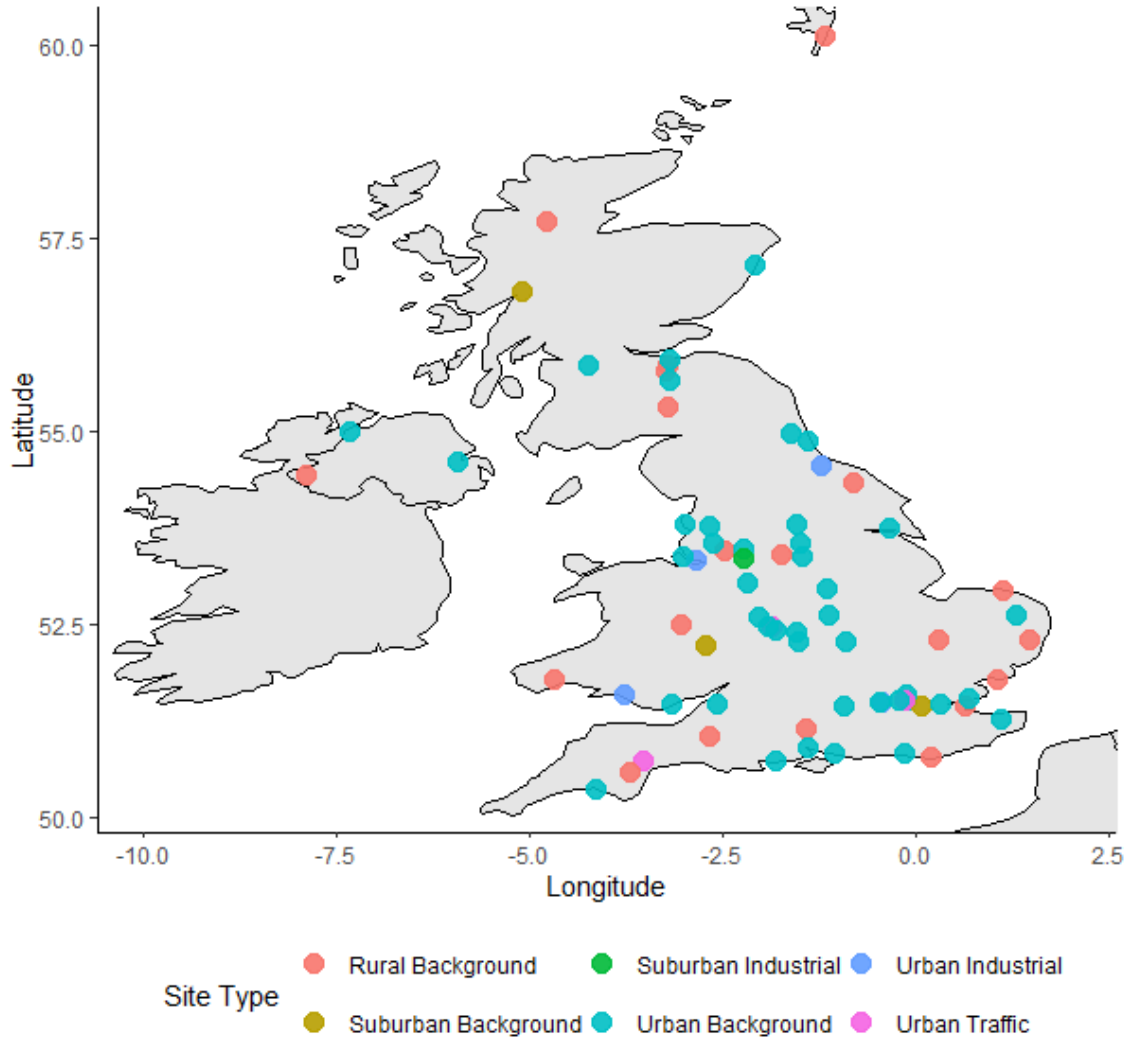


Figure 2.1: Map showing the AURN sites that have data available between 2011 – 2019 and their corresponding site types.

There is a total of 274 sites within the network and currently 171 active sites. Throughout this thesis our focus is on ozone and the years $2011 - 2019$ to avoid the need to account for decadal changes. This results in a sample of 72 AURN sites which can be seen in Figure 2.1. The sites are categorised into different types based on their location with respect to pollutant sources: rural background (RB), urban background (UB), suburban background (SUB), urban traffic (UT), urban industrial (UI) and suburban industrial (SUI), of which we have 41, 20, 3, 3, 4 and 1 respectively. For details about site type classification see `uk-air.defra.gov.uk/networks/site-types`. Here, we refer to sites that are not classified as rural background or urban background as *Other*. Figure 2.1 shows the relatively poor coverage of Scotland and Wales. The data can be downloaded from the website

`http://www.airquality.co.uk/archive/data_and_statistics.php`.

All but three sites have less than 10% of data missing during the measurement period. Some of the missingness is due to the monitoring sites malfunctioning and waiting for repair, however it does appear at random and does not hinder our ability to characterise certain times of the year. The missingness is low enough to still provide enough data to perform the analysis in this thesis.

	2011	2012	2013	2014	2015	2016	2017	2018	2019
All	0.417	0.210	0.394	0.212	0.149	0.211	0.153	0.875	0.589
RB	0.684	0.474	0.684	0.389	0.368	0.450	0.200	1.000	0.800
UB	0.281	0.088	0.289	0.179	0.077	0.125	0.146	0.854	0.488
Other	0.333	0.111	0.222	0.000	0.000	0.091	0.091	0.727	0.636

Table 2.3: The proportion of AURN sites that exceed the UK Air Quality Strategy objective, exceeding $100\,\mu g/m^3$ on more than 10 days.

Table 2.3 shows the proportion of sites by type which do not meet the UK Air Quality Strategy, i.e. for which there were at least 10 days of concentrations above $100\,\mu g/m^3$ a year. On average 56% of rural background sites observe at least 10 days above the moderate risk threshold, compared to 28% and 25% at urban background

and Other sites, respectively. The range of observed values differ between the three types as seen in Figure 2.2. The maximum observed value at rural background sites is 214.0 $\mu g/m^3$ compared to 191.2 $\mu g/m^3$ and 162.4 $\mu g/m^3$ at urban background sites and Other sites, respectively. Due to the limited number of sites of type SUB, UT, UI and SUI and the reduced number of moderate risk observations we focus on only rural background sites and urban background sites throughout the thesis, which will be depicted by ($\circ$) points and ($\triangle$) points in figures unless stated otherwise.

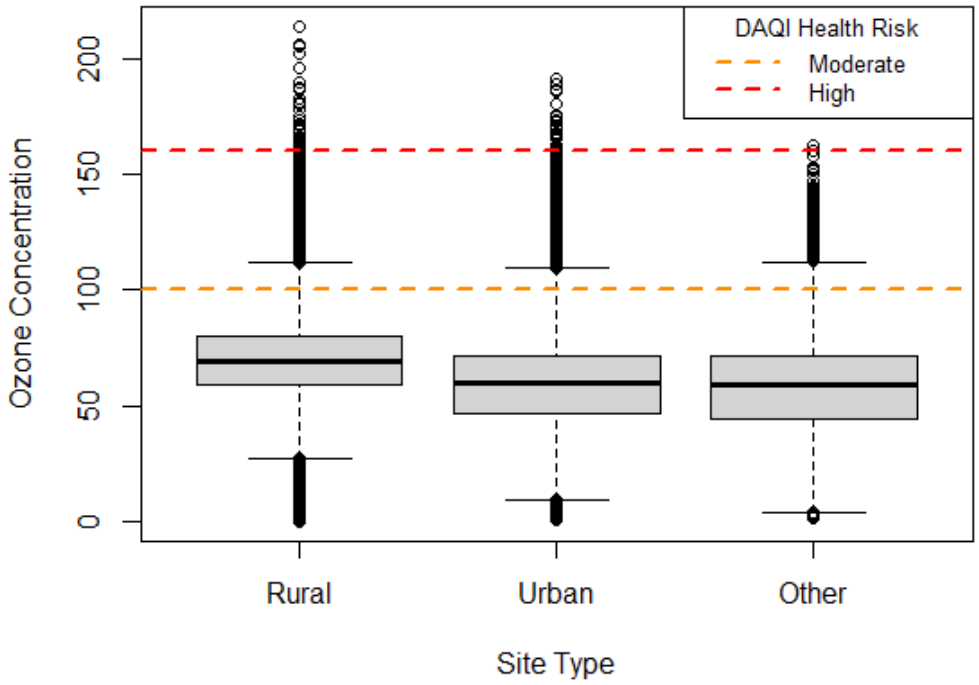


Figure 2.2: Box plot showing the distribution of DM8 ozone observations at each site type. Orange dashed line indicates the concentration of moderate risk level ($100 \mu g/m^3$) and red dashed line indicates the concentration of high-risk level ($160 \mu g/m^3$).

2.4.2 Forecasts - Air Quality in the Unified Model (AQUM)

Numerical process-based forecast models are deterministic tools that use the underlying processes of meteorological variables to estimate the atmospheric composition at different levels. The UK Met Office air quality forecasts are produced using the on-line air quality model, Air Quality in the Unified Model (AQUM), which is a

configuration of the Met Office Unified Model (MetUM). AQUM currently operates with a 12km horizontal resolution grid covering the UK and Northwest Europe, with the native model grid being on a rotated-pole coordinate system with the North Pole at latitude 37.5 and longitude 177.5. For details of the mechanics behind the model see Savage et al. (2013). This numerical model produces hourly forecasts for up to five days of forecasts for each grid point. The model forecasts undergo bias correction known as Statistical Post-Processing of Observations (SPPO). At observation sites, this method is an extension of the Hybrid Forecast (HF) method described by Kang et al. (2008), where the HF at a single site i at time $t + \Delta t$ is given by,

$$HF_{i,t+\Delta t} = m_{i,t+\Delta t} + \mu_{i,t},$$

where $\mu_{i,t}$ is given by

$$\mu_{i,t} = 0.5(o_{i,t} - m_{i,t}) + 0.5[\text{median}_{j \in C}(o_{j,t} - m_{j,t})],$$

$o_{i,t}$ is the observed concentration at time t, $m_{i,t}$ and $m_{i,t+\Delta t}$ are the model predicted concentrations at time t and forecast time $t + \Delta t$ respectively and C denoting the classification of site i. The Ordinary Kriging technique (Denby et al., 2005) is then used to spread i, t onto a gridded field R_t. The final improved gridded forecast is therefore given by

$$SPPO_{t+\Delta t} = M_{t+\Delta t} + R_t$$

where $M_{t+\Delta t}$ is the raw model forecast at this future time $t+\Delta t$ and R_t is the gridded field of $\mu_{i,t}$. Further details of the SPPO procedure can be found in Neal et al. (2014). In the analysis of Chapter 4, we use the day one SPPO data from the operational air quality forecast from January 2012 to December 2019. This corresponds to the UK Met Office operational suites OS27-OS43, during which various improvements where made to the raw AQUM. From here on in, we refer to the sampled SPPO data as AQUM data.

2.4.3 Ozone season

With ozone being a photochemical reaction, high concentrations typically occur during low-moving high-pressure systems with clear skies and elevated temperatures. Consequently, showing strong seasonality in when high concentrations usually occur, this being in Spring and Summer. As such, restricting our datasets to only these periods allow for easier modelling since we would not need to account for seasonality. Therefore, defining a period known as the *ozone season*.

In this thesis we work with two definitions of ozone season. The first uses conventional definitions of Spring and Summer in the atmospheric literature, March to August, and will be used solely in Chapter 6. The second definition is a novel approach in defining such a season. We define the ozone season as the months where at least 2% of the monthly DM8 observations exceed the 95th quantile of the full series at each site. We obtain this definition through sensitivity checks of both the number of exceedances and the quantile defining an exceedance. This optimises the number of exceedances available and upholds the approximate independence modelling assumption needed in our analysis of Chapter 4. The second ozone season definition will be used in Chapter 4 and Chapter 5.

The second ozone season definition results in all but thirteen sites having ozone seasons starting in March, with those thirteen starting in April. Three sites have a season that ends in May or June, four ending in September and the rest in July or August. The shortest seasons are three months (at three sites) but more often seasons last for five or six months (32, 16 respectively), with two sites having a season lasting seven months. See Table A.1 in Appendix A for the ozone season for each sampled monitoring site.

Chapter 3

Theory

3.1 Introduction

Extreme value theory is a field developed throughout the twentieth century starting from asymptotic arguments derived by Tippett and Fisher (1928), formalised into statistical methods by Gumbel (1958) and greatly extended by the likes of Pickands (1975); Smith (1989); Ledford and Tawn (1996, 1997); Heffernan and Tawn (2004). The area of extreme value theory is motivated by wanting a way to accurately model the probabilistic behaviour of events that by definition are rare, i.e. the tails of a distribution where the number of observations are intrinsically small. Standard statistical modelling approaches are typically biased in tail behaviour estimation due to the modelling parameters being driven by the main body of the distribution where most observations lie. This creates limited confidence in the estimates for high quantiles and measures. There is often an interest in extrapolating beyond the observed upper limit of the data. Extreme value models provide an asymptotically justified approach for such extrapolation.

Section 3.2 reviews existing theory and inference for univariate extreme value approaches. In Section 3.3 we introduce multivariate extreme value theory and the concept of extremal dependence. We discuss an approach for modelling asymptotic independence, the joint tail model of Ledford and Tawn (1996, 1997), in Section

3.4. Note the Heffernan and Tawn (2004) model can capture both asymptotic independence and asymptotic dependence and is presented in Chapter 5. Section 3.5 details the bootstrapping approaches used throughout the thesis.

3.2 Univariate extreme value theory

3.2.1 Overview

Univariate extreme value distributions exist based on asymptotically derived theory that allows the analysis of extreme events. From using extreme value analysis, inference and prediction about rare events that have not yet been observed can be made. Two types of modelling approaches are discussed below in their simplest forms, modelling block maxima and modelling exceedances of a high threshold. With all approaches here, one can reparametrise to build in more complex structures, for example including covariate effects.

3.2.2 Block maxima approach

Theory

Let $X_1, \ldots, X_n$ be n independent and identically distributed (IID) random variables with common distribution function F, known as the *marginal distribution*. The *maximum order statistic* is defined as,

$$M_{X,n} = \max(X_1, \ldots, X_n).$$

For example, the daily maxima $M_{X,24}$ can be obtained from the hourly values $X_1, \ldots X_{24}$. We remark that any theory developed for the maxima can be applied to the minima through the relation

$$\min(X_1, \ldots, X_n) = \max(-X_1, \ldots, -X_n).$$

For notational simplicity the subscript X is dropped from $M_{X,n}$. Due to the independence assumption of $X_1, \ldots, X_n$, the distribution of M_n can be derived exactly as

$$\begin{aligned}\Pr(M_n \leq x) &= \Pr(X_1 \leq x, \ldots, X_n \leq x) \\ &= \Pr(X_1 \leq x) \ldots \Pr(X_n \leq x) \\ &= [F(x)]^n.\end{aligned}$$

Since F is unknown, we can study the behaviour of F^n as $n \to \infty$. However as $n \to \infty$, the maximum order statistic tends to a point mass x^* on the upper endpoint of F; the asymptotic distribution of M_n is degenerate. In the same way that the Central Limit Theorem (CLT) concerning the mean value of a sample, asymptotic results exist for the distribution of M_n. More precisely, if there exists sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that as $n \to \infty$

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x), \tag{3.2.1}$$

where G is a non-degenerate distribution function, then G belongs to the extreme value distribution family. Leadbetter et al. (1983) constructs and proves the Extremal Types Theorem, which states the limit distribution of the maximum values is one of the following forms:

$$\mathbf{I} : G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \quad -\infty < x < \infty;$$

$$\mathbf{II} : G(x) = \begin{cases} 0, & x \leq b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x > b; \end{cases}$$

$$\mathbf{III} : G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^{\alpha}\right]\right\}, & x < b, \\ 1, & x \geq b; \end{cases}$$

which correspond to the Gumbel, Fréchet and Negative Weibull families, respectively. These families possess the property of *max-stability*; a distribution G is max-stable if for every $n > 0$ there exists constants $A_n > 0$ and $B_n \in \mathbb{R}$ such that

$$G(A_n x + B_n) = [G(x)]^n .$$

A parametrisation is commonly used to unify the three distinct classes. The *generalised extreme value* (GEV) distribution, written as $\text{GEV}(\mu, \sigma, \xi)$, is defined as

$$G(x; \mu, \sigma, \xi) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \right\}, \tag{3.2.2}$$

where $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$ and $c_+ = \max(c, 0)$. The GEV distribution exists on the set $\{x : 1 + \xi(x-\mu)/\sigma > 0\}$ and the parameters (μ, σ, ξ) are known as the location, scale and shape parameters respectively. We say F is in the *domain of attraction* of G and the limit distribution of F is determined by the shape parameter ξ value, where

- $\xi > 0$ corresponds to the Fréchet distribution which has a heavy upper tail with an infinite upper endpoint and finite lower limit,
- $\xi \to 0$ corresponds to the Gumbel distribution which has an exponentially decaying tail and
- $\xi < 0$ corresponds to the Negative Weibull distribution which has a tail tending towards a finite upper limit and infinite lower endpoint.

Block size choice

The GEV distribution is used to model the distribution of block maxima and assumes that the limit of equation (3.2.1) holds for some finite n, where the data is split into k equal blocks of size n. The choice of block length amounts to a *trade-off* between bias and variance: too short a block size the approximation by the limit model in equation (3.2.2) is likely to be poor and lead to bias in estimation and

extrapolation due to taking the maximum of not sufficiently many observations and thus violating the independence assumption; too long a block will generate too few block maxima, leading to large estimation of variance. In practical applications the length of block n is usually given by the context, i.e. in many environmental applications taking annual maxima ensures the resulting maximum order statistic is stationary. Inference for the GEV model is typically carried out with likelihood-based or moment-based estimation with respect to the parameters (μ, σ, ξ). However, there exists no analytical solution, therefore numerical optimisation methods are employed.

Return levels

In practice, interest lies in the time to, or the severity of, the next sufficiently large extreme event. For a stationary series, the inference is expressed by *return levels* and *return periods*. The return period of level x_p is the expected time between exceedances of x_p. This relates to the $1/p$-block return level which is defined as the level for which the expected time to wait between block maxima exceedances is $1/p$ blocks. Therefore, the $1/p$ block return level x_p is the $1-p$ quantile of the GEV distribution for $0 < p < 1$. From this definition, the limiting distribution for block maxima, equation (3.2.2), can be rearranged to provide an expression for x_p,

$$\hat{x}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[1 - \{-\log(1-p)\}^{-\hat{\xi}} \right] & \text{if } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log\{-\log(1-p)\} & \text{if } \hat{\xi} = 0, \end{cases}$$

where $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ are the estimates of (μ, σ, ξ). Throughout the thesis, modelling parameters $\boldsymbol{\theta}$ are estimated using the maximum likelihood approach and are denoted by $\tilde{\boldsymbol{\theta}}$. Maximum likelihood estimates are not always obtainable due to violation of the usual regularity conditions resulting in the standard Gaussian asymptotic likelihood results to not be automatically applicable. Smith (1985) studied this and found:

- when $\xi > -0.5$, maximum likelihood estimators are regular, i.e. having the usual asymptotic properties,
- when $-1 < \xi < -0.5$, maximum likelihood estimates can be obtainable, but not have the usual asymptotic properties and
- when $\xi < -1$, maximum likelihood estimators are unlikely to be obtainable.

3.2.3 Threshold exceedance approach

Motivation

Although the block maxima approach is easily interpretable, it can be a wasteful approach to extreme values as there is a strong probability that other observations are large enough to be characterised as extreme, i.e. fall within the range of smallest and largest block maxima. Despite these tail values being more extreme than some block maxima, the block maxima approach will ignore these, resulting in the potential misclassification of the full tail behaviour of the distribution.

Figure 3.1 shows the time series of DM8 ozone concentrations measured at an AURN monitoring site in Aston Hill between 2011 and 2019. The solid red dots are the values that fall above $100\,\mu g/m^3$, whereas the blue triangles are the annual maxima as defined by the block-maxima approach. It is evident that some blocks contain more than one value above $100\,\mu g/m^3$. Consequently, threshold exceedance methods can offer a better alternative to block maxima methods by utilising the values that block maxima approaches would ignore, which could provide extra information about the tail behaviour.

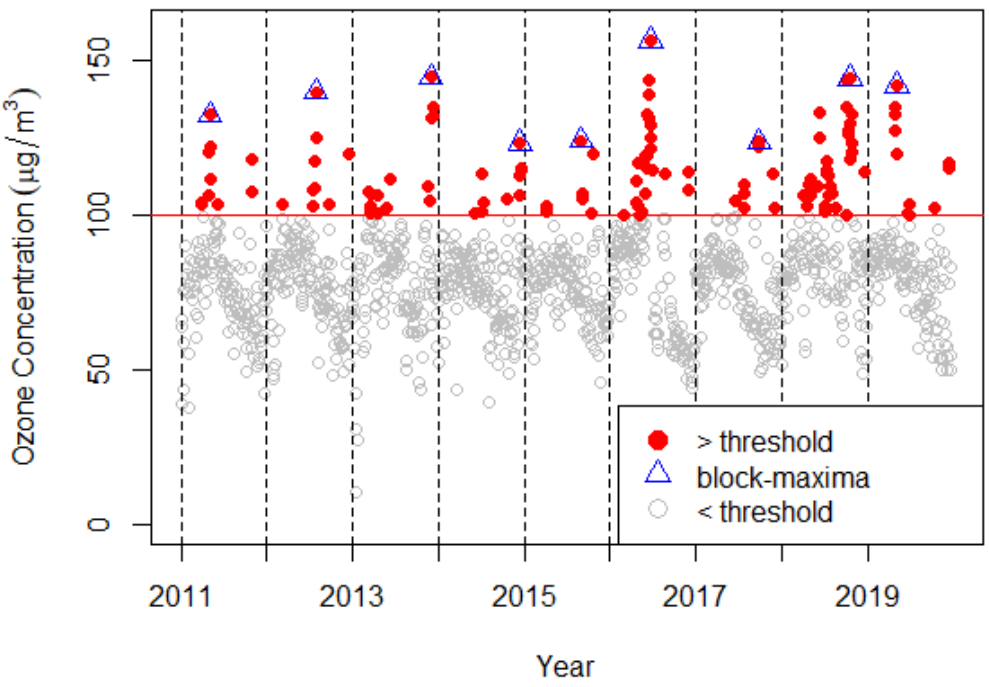


Figure 3.1: DM8 ozone concentration ($\mu g/m^3$) measured at an AURN monitoring site in Aston Hill during $2011 - 2019$. Solid red points are exceedances of $100\,\mu g/m^3$ and blue triangles are the annual maxima as defined by block maxima approach.

Theory

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed variables with common distribution function F. Assuming the asymptotic theory of equations (3.2.1) and (3.2.2) hold, then a sequence of point processes $P_1, P_2, \ldots$ can be constructed on $[0, 1] \times \mathbb{R}$ using P_n

$$P_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) ; i = 1, \ldots, n \right\}.$$

Consequently, by construction, we can learn about the tail of F by examining the limit behaviour of P_n as $n \to \infty$. The limit process is non-degenerate since the distribution of the normalised maxima is non-degenerate. Small points of the process

are normalised to the same value b_l, with

$$b_l = \lim_{n\to\infty} \frac{x^F - b_n}{a_n},$$

while large points of the process are retained in the limit process. Under these conditions on P_n, the limiting point process is defined on the set $[0,1] \times (b_l, \infty)$

$$P_n \to P \text{ as } n \to \infty, \tag{3.2.3}$$

where P is a non-homogeneous Poisson process (PP) with intensity function

$$\lambda(t,x) = \frac{1}{\sigma}\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-1-1/\xi},$$

for $(t,x) \in [0,1] \times (b_l, \infty)$. This limit result motivates that the behaviour of all large values from F are asymptotically determined by the characteristics of a_n, b_n and ξ, likewise with the block maxima approach. Assuming the conditions for limit in equation (3.2.3) hold, Pickands (1975) and Smith (1989) show for $x > 0$ and $X \sim F$

$$\Pr(X > u_n + a_n x \mid X > u_n) \to \left[1 + \xi\frac{x}{\psi}\right]_+^{-1/\xi} \tag{3.2.4}$$

as $n \to \infty$, where $u_n \to x^*$ as $n \to \infty$, $\psi > 0$ and $\xi \in \mathbb{R}$. The distribution function

$$G(x) = 1 - \left[1 + \xi\frac{x}{\psi}\right]_+^{-1/\xi}$$

for $x > 0$, corresponds to the generalised Pareto distribution (GPD), with scale parameter $\psi > 0$ and shape parameter $\xi \in \mathbb{R}$, denoted as GPD(ψ, ξ). The limit of equation (3.2.4) shows that as the threshold tends to the upper endpoint of the distribution, under weak conditions, the scaled excesses of the threshold tends to a GPD(ψ, ξ).

The limiting results of equation (3.2.4) motivates that for sufficiently large threshold

u_n,

$$\Pr(X > x \mid X > u) = \left[1 + \xi \left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}, \tag{3.2.5}$$

for $x > u$ with $u = u_n$, i.e.

$$(X - u) \mid X > u \sim \text{GPD}(\sigma_u, \xi). \tag{3.2.6}$$

The GPD possesses the *threshold stability property* which states that if $X - u$ is distributed as in equation (3.2.6), then for any larger threshold $v > u$

$$(X - v) \mid X > u \sim \text{GPD}(\sigma_u + \xi(v - u), \xi). \tag{3.2.7}$$

Thus, the shape parameter ξ is constant with threshold, i.e. is invariant to threshold choice, whereas the scale parameter $\sigma_v = \sigma_u + \xi(v - u)$ is not (Davison and Smith, 1990).

Threshold choice

Modelling extreme events identified by exceeding a high threshold u can provide better alternatives to the block maxima. However, the choice of threshold u directly affects the number of threshold exceedances which in turn creates an analogous issue to the block size in the block maxima approach, bias-variance trade-off. Choosing a low threshold increases the amount of data that is used which reduces the uncertainty in the estimation of the model parameters making the statistical inference more efficient. However, this increases the chance that the asymptotic basis of the model will break down and introduce bias. Setting a threshold too high will generate too few excesses leading to increased variance in the model parameter estimation. Although the choice of threshold using diagnostics has been studied by Tancredi et al. (2006) and Wadsworth and Tawn (2012) there are no specific rules in choosing the best threshold. This being said, there are two commonly used diagnostic plots to aid in threshold choice, namely the *mean residual life* (MRL) plots and *parameter*

stability plots (Scarrott and MacDonald, 2012).

In the construction of the MRL plot first consider a set of threshold exceedances $X_i \mid X_i > u$ for $i = 1, \dots n_u$, that follow a generalised Pareto distribution with parameters σ_u and ξ. Then, provided that $\xi < 1$, the expected value of an arbitrary X threshold excesses $X - u$ is given by

$$\mathbb{E}[X - u \mid X > u] = \frac{\sigma_u}{1 - \xi}.$$

With the assumption that the GPD is valid for the excesses of the threshold u, then it should be equally valid for all higher thresholds $v > u$. Therefore, for $v > u$,

$$\mathbb{E}[X - v \mid X > v] = \frac{\sigma_u + \xi(v - u)}{1 - \xi}, \tag{3.2.8}$$

if $\xi < 1$. We call u a suitable threshold for the modelling data points above if the mean excesses in equation (3.2.8) is linear in v for $v > u$.

In the construction of the parameter stability plots, first assume X follows a GPD above the threshold u as in equation (3.2.6) then for any higher threshold $v \geq u$, X above the higher threshold v follows the distribution given in equation (3.2.7). This shows the shape parameter is constant for the higher threshold, but the scale parameter varies with threshold.

To assess parameter stability we reparametrise the scale parameter for the higher threshold v to the modified scale σ^* such that $\sigma^* = \sigma_v - \xi v$. Consequently, both shape parameter ξ and the modified scale σ^* are threshold invariant for suitably high threshold. The choice of threshold from the parameter stability is determined by the lowest possible value for u such that the estimates of the modified scale and shape parameters remain approximately constant above this level. Figure 3.2 provides a demonstration of parameter stability plots. These plots indicate that, in the case of the DM8 ozone observations at Aston Hill, the lowest possible value for u is approximately the 90th quantile, subject to estimation uncertainty depicted by the 95% confidence intervals.

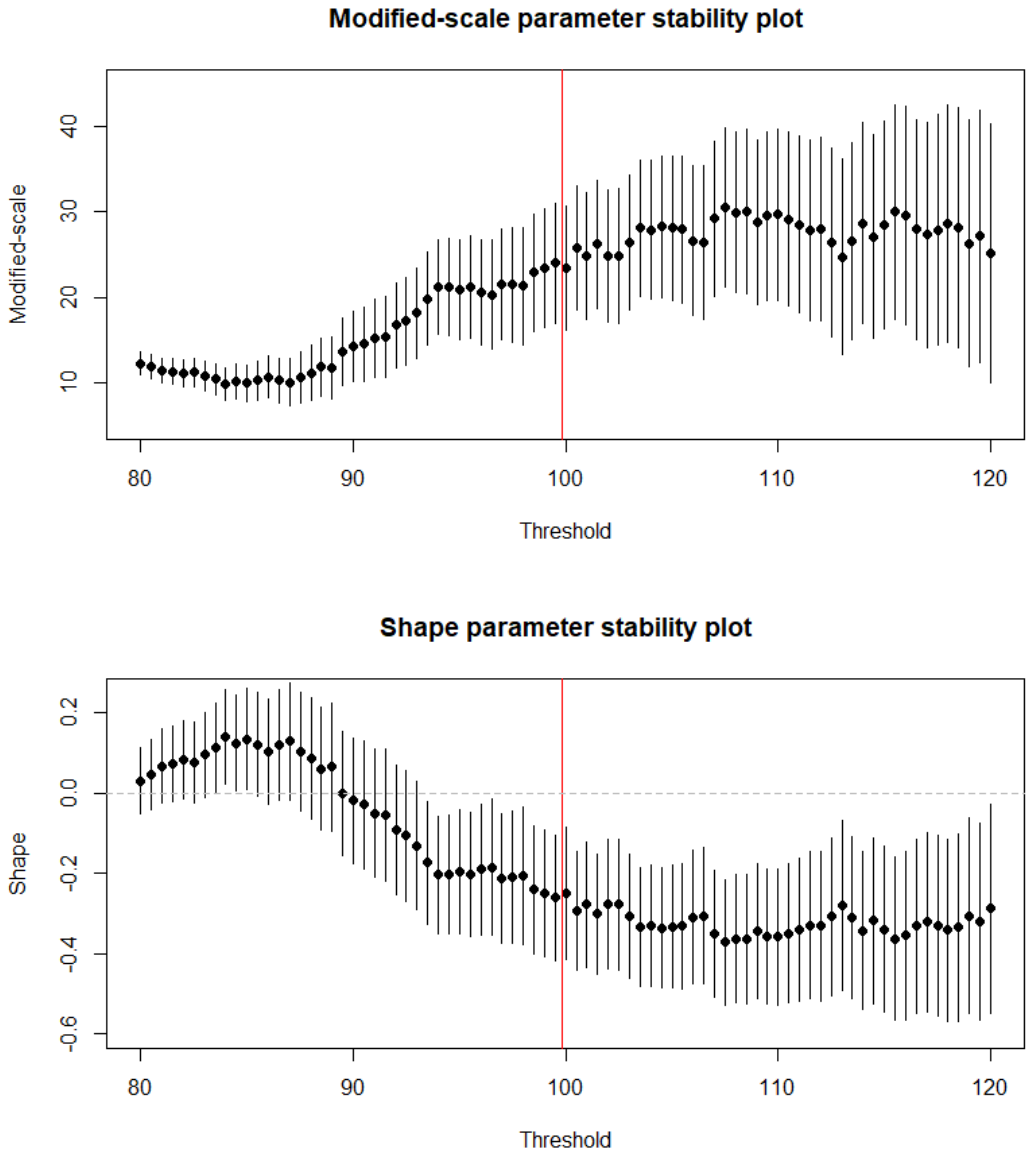


Figure 3.2: Parameter stability plots for (top) modified-scale parameter (bottom) shape parameter at Aston Hill, where the black vertical lines indicate the 95% confidence intervals and the red line indicates the empirical 90th quantile.

Return levels

Return levels are computed using a similar process to the block maxima approach, however we must undo the conditioning in equation (3.2.5) since the data are conditional upon exceeding a sufficiently high threshold u. To undo this conditioning,

we multiple by the *rate of exceedance* $\varphi_u = \Pr(X > u)$ such that for $x > u$,

$$\begin{aligned}\Pr(X > x) &= \Pr(X > u) \Pr(X > x \mid X > u) \\ &= \varphi_u \left[1 + \xi \left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}. \end{aligned} \tag{3.2.9}$$

Inverting the unconditional probability distribution function given in equation (3.2.9) gives the return value that is exceeded once every m observations

$$\hat{x}_m = \begin{cases} u + \frac{\hat{\sigma}_u}{\hat{\xi}} \left[(m\varphi_u)^{\xi} - 1\right] & \text{for } \hat{\xi} \neq 0 \\ u + \hat{\sigma}_u \log(m\varphi_u) & \text{for } \hat{\xi} = 0, \end{cases} \tag{3.2.10}$$

where $m > \varphi_u^{-1}$ and $(\hat{\sigma}_u, \hat{\xi})$ are the maximum likelihood estimates of (σ_u, ξ). Since the estimates have no closed forms, we use numerical optimisation to maximise the likelihood.

3.3 Multivariate extreme value theory

3.3.1 Motivation

For many types of data either or both of the independent and identically distributed modelling assumptions are not realistic. For example, if temperature is high then there is a higher probability that air quality is of unhealthy levels than if the temperature was low. Contextually, the definition of multivariate extreme events will naturally lead into the type of modelling approach to use (Barnett, 1976). Two types of modelling approaches are discussed below, modelling using componentwise maxima; the maximum of each variable, which is a natural extension of the univariate block maxima approach; and the multivariate threshold approaches, an extension of the univariate threshold approaches. In Section 3.3.3 both approaches are outlined with a semi-parametric conditional approach given in Chapter 5.

3.3.2 Copula theory

In multivariate modelling, it is important to consider the dependence properties of the variables. Copulas provide a method to separate out the marginal behaviour from the dependence structure, allowing one to model these features separately (Joe, 1997).

By Sklar's theorem (Sklar, 1959), if $\mathbf{X} = (X_1, \ldots, X_d)$ has joint distribution F, and has continuous marginal distributions $X_i \sim F_i$, for $i = 1, \ldots, d$, then there exists a unique copula C such that

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}.$$

The copula C is a joint distribution function with the property that every marginal distribution is uniform on the interval $[0, 1]$. This result and the probability integral transform allows for transformation between different marginal distributions without changing the dependence structure. Consequently, if F_i has an inverse distribution F_i^{-1}, then by the probability integral transform, $U = F_i(X) \sim \text{Uniform}(0, 1)$ with $F_i^{-1}(U) \sim F_i$. So for example, we can obtain a copula with standard Fréchet margins by

$$F(x_1, \ldots, x_d) = C\left\{-\frac{1}{\log F_1(x_1)}, \ldots, -\frac{1}{\log F_d(x_d)}\right\}.$$

These results are often used in multivariate extreme value theory as transforming margins can highlight features associated with the extreme values. For example, when transforming onto Fréchet margins the largest values are magnified, as seen in Figure 3.3 (bottom left). To estimate the extremal dependence the variables are transformed onto common margins prior to modelling the dependence structure. Following Coles and Tawn (1991), we model the marginal distributions of two random variables X_1 and X_2 as

$$F_{X_i}(x) = \begin{cases} 1 - [1 - \tilde{F}_{X_i}(u_i)] \left[1 + \xi_i \left(\frac{x - u_i}{\sigma_{u,i}}\right)\right]_+^{-1/\xi_i}, & \text{for } x > u_i, \\ \tilde{F}_{X_i}(x), & \text{for } x \leq u_i, \end{cases} \tag{3.3.1}$$

where $\tilde{F}_{X_i}(x)$ is the empirical marginal cumulative distribution function of X_i and $i = 1, 2$.

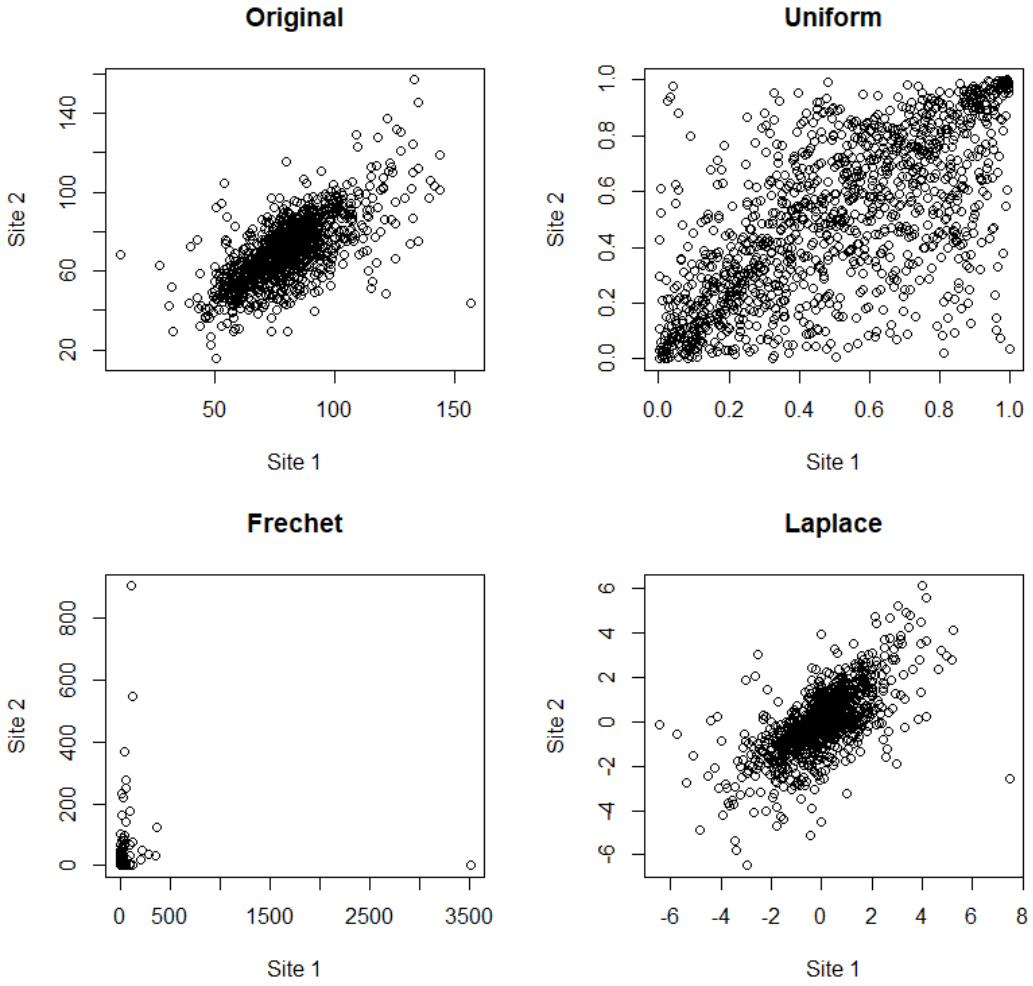


Figure 3.3: Scatter plots of DM8 ozone data from two AURN sites on (top left) the original scale (top right) uniform scale (bottom left) Fréchet scale and (bottom right) Laplace scale.

Having estimated the marginal structure, which can be obtained using methods such as, standard likelihood approaches, L-moments or the empirical cumulative distribution function, an appropriate choice of transformation onto common margins is required before characterising the extremal dependence. At different stages of the thesis we require Fréchet and Laplace margins. The probability integral transformation is used to obtain the appropriate marginal distributions. For

standard Fréchet margins,

$$X_{F,i} = -1/\log\{F_{X_i}(X_i)\}, \tag{3.3.2}$$

which gives $\Pr(X_{F,i} \leq x_F) = \exp(-1/x_F)$. For standard Laplace margins,

$$X_{L,i} = \begin{cases} \log\{2F_{X_i}(X_i)\}, & \text{if } F_{X_i}(X_i) < 1/2, \\ -\log\{2[1 - F_{X_i}(X_i)]\}, & \text{if } F_{X_i}(X_i) \geq 1/2. \end{cases} \tag{3.3.3}$$

3.3.3 Theory

Componentwise maxima

Let $(X_{j,1}, \ldots, X_{j,d})$, where $j = 1, \ldots, n$, be d-dimensional vectors which for each j has joint distribution function G and is independent over j. We denote the vector of *componentwise maxima* by $\mathbf{M}_n = (M_{n,1} \ldots, M_{n,d})$, where $M_{n,k} = \max_{j \in \{1,\ldots,n\}} X_{j,k}$ for $k = 1, \ldots, d$. Analogous to the univariate case, if there exists normalising constants $a_{n,k} > 0$ and $b_{n,k}$ for $k = 1, \ldots, d$ such that

$$\Pr\left(\frac{M_{n,1} - b_{n,1}}{a_{n,1}} \leq z_1, \ldots, \frac{M_{n,d} - b_{n,d}}{a_{n,d}} \leq z_d\right) \to F(z_1, \ldots, z_d),$$

as $n \to \infty$ where the limiting distribution function F is non-degenerate in each margin, then F is a d-dimensional *multivariate extreme value* distribution. Each marginal Z_k follows a GEV distribution with parameters (μ_k, σ_k, ξ_k), as given in equation (3.2.2). Common practice in studying the dependence structure of F is to assume that each margin follows a common distribution, for example the standard Fréchet distribution, i.e. GEV$(1, 1, 1)$, $\exp(-1/z_k)$ for $z_k > 0$. Choosing standard Fréchet marginal distributions is the usual choice for modelling componentwise maxima as it emphasises the largest values. Consider the vector of random variables $Z_1, \ldots, Z_d$ with common Fréchet marginal distributions. The multivariate extreme

value distribution of F is written as

$$F(z_1, \ldots, z_d) = \exp\{-V(z_1, \ldots, z_d)\} \text{ for } z_k > 0, \tag{3.3.4}$$

where $k = 1, \ldots, d$ and the *exponent measure* V on the unit simplex, $\mathcal{S}_{d-1} = \{\boldsymbol{\omega} \in [0,1]^d : \sum_{k=1}^{d} \omega_k = 1\}$, is defined as

$$V(z_1, \ldots, z_d) = d \int_{\mathcal{S}_{d-1}} \max_{k=1,\ldots,d} \left(\frac{\omega_k}{z_k} \right) dH(\boldsymbol{\omega}), \tag{3.3.5}$$

with the *spectral measure* H, an arbitrary distribution function on $[0,1]$, satisfying the moment constraint

$$\int_{\mathcal{S}_{d-1}} w_k dH(\boldsymbol{\omega}) = \frac{1}{d}.$$

In the bivariate case, consider the pair of random variables (Z_1, Z_2) with common Fréchet marginal distributions, then the exponent measure V is given by

$$V(z_1, z_2) = 2 \int_0^1 \max \left\{ \frac{\omega}{z_1}, \frac{1-\omega}{z_2} \right\} dH(\omega). \tag{3.3.6}$$

The pair is independent if $V(z_1, z_2) = z_1^{-1} + z_2^{-1}$ and $H(\{0\}) = H(\{1\}) = 1/2$; perfectly dependent if $V(z_1, z_2) = \max(z_1^{-1}, z_2^{-1})$ and $H(\{1/2\}) = 1$.

Several parametric models belong to the class of multivariate extreme value distributions, including the logistic distribution (Gumbel, 1960), for which the exponent measure is

$$V(z_1, \ldots, z_d) = \left(\sum_{k=1}^{d} z_k^{-\alpha} \right)^{1/\alpha}, \tag{3.3.7}$$

for $\alpha \in (0, 1]$. For this model, taking $\alpha = 1$ leads to independence between the variables, while as $\alpha \to 0$ we approach exact dependence. Another parametric model is the Dirichlet distribution (Coles and Tawn, 1991) which confines all mass to the interior of $\mathcal{S}_d$ while allowing for marginal bivariate asymmetry. In the bivariate case,

the exponent measure for the Dirichlet distribution has the form

$$V(z_1, z_2) = 1/z_1[1 - \text{Be}(q; \alpha + 1, \beta)] + 1/z_2\text{Be}(q; \alpha, \beta + 1) \tag{3.3.8}$$

where $q = \alpha z_1/(\alpha z_1 + \beta z_2)$ and $\text{Be}(q; \alpha, \beta)$ is the beta distribution function evaluated at q. Complete dependence is obtained in the limit as $\alpha = \beta$ tends to infinity. Independence is obtained as $\alpha = \beta$ approaches zero and when one of α, β is fixed and the other approaches zero.

Other parametric models belonging to the class of multivariate extreme value distributions include: the asymmetric logistic model (Tawn, 1988, 1990), the negative logistic model (Galambos, 1975), the negative asymmetric logistic model (Joe, 1990), the Hüsler-Reiss distributions (Hüsler and Reiss, 1989), the bilogistic model (Smith, 1990a), the negative bilogistic model (Coles and Tawn, 1994) and the asymmetric mixed model (Tawn, 1988).

The link between V and H

The exponent measure V and the spectral measure H are related through the definition of V in equation (3.3.5). Coles and Tawn (1991) show that these measures are linked by further relations. Let $h_{j,c}$ be the class of densities for H, then for $c = \{i_1, \ldots, i_m\}$ the spectral density is given by

$$\frac{\partial V}{\partial z_{i_1} \ldots \partial z_{i_m}} = -m \left(\sum_{j=1}^{m} z_{i_j} \right)^{-(m+1)} h_{m,c} \left(\frac{z_{i_1}}{\sum z_{i_j}}, \ldots, \frac{z_{i_m}}{\sum z_{i_j}} \right) \tag{3.3.9}$$

on $\{\mathbf{z} \in \mathbb{R}_+^p : z_\alpha = 0 \text{ if } \alpha \notin c\}$. To show how this is obtained, we focus on the bivariate case where the exponent measure is given by equation (3.3.6). Assuming H is differentiable everywhere, we can rewrite the integral in terms of ω,

$$V(z_1, z_2) = 2 \int_0^1 \max \left\{ \frac{\omega}{z_1}, \frac{1 - \omega}{z_2} \right\} h(\omega) d\omega.$$

We can simplify the integral by splitting the interval into two regions: the part where $\omega/z_1 > (1-\omega)/z_2$ and the part where $(1-\omega)/z_2 > \omega/w_1$. To do this, we need to solve the equation $\omega/z_1 = (1-\omega)/z_2$ for ω, which is $\omega = z_1/(z_1+z_2)$. As such, the integral becomes

$$V(z_1, z_2) = 2\int_0^{\frac{z_1}{z_1+z_2}} \frac{\omega}{z_1} h(\omega) d\omega + 2\int_{\frac{z_1}{z_1+z_2}}^1 \frac{1-\omega}{z_2} h(\omega) d\omega. \tag{3.3.10}$$

Computing the integral in equation (3.3.10) and differentiating twice with respect to z_1 and z_2 yields the required result of

$$\frac{\partial^2}{\partial z_1 \partial z_2} V(z_1, z_2) = -\frac{2}{(z_1+z_2)^3} h(\omega). \tag{3.3.11}$$

To demonstrate how to obtain the spectral density using the result of Coles and Tawn (1991), lets consider the 2-dimensional case of the logistic model with exponent measure given in equation (3.3.7). By equation (3.3.11), using ω as the pseudo-coordinate space where $\omega = z_1/(z_1+z_2)$,

$$\frac{\partial V}{\partial z_1 \partial z_2} = (z_1+z_2)^{-3} \alpha \left(1/\alpha - 1\right) \left[\omega(1-\omega)\right]^{-(\alpha+1)} \left[\omega^{-\alpha} + (1-\omega)^{-\alpha}\right]^{1/\alpha - 2}.$$

Consequently, the bivariate logistic dependence model has the spectral density given by,

$$h(\omega) = (\alpha - 1)[\omega(1-\omega)]^{-(\alpha+1)} \left[\omega^{-\alpha} + (1-\omega)^{-\alpha}\right]^{1/\alpha - 2}. \tag{3.3.12}$$

Generalising to the d-dimensional case yields

$$h_{p,c_p}(\boldsymbol{\omega}) = \left[\prod_{j=1}^{p-1} (j\alpha - 1)\right] \left(\prod_{j=1}^{p} \omega_j\right)^{-(\alpha+1)} \left(\sum_{j=1}^{p} \omega_j^{-\alpha}\right)^{1/\alpha - p}, \tag{3.3.13}$$

for $h_{j,c} \equiv 0$ and $j < p$. If $\alpha = 1$ in the exponent measure then the variables are independent, which corresponds to $h_{1,(i)} \equiv 1$ for all i and $h_{j,c} \equiv 0$ for all $j > 1$, therefore all the mass is confined to the vertices of S_p.

Multivariate threshold model

Suppose that $\mathbf{X_1}, \ldots, \mathbf{X_n}$ are independent realisations of a random d-dimensional vector $\mathbf{X}_i = (X_1, \ldots, X_d)$ with joint distribution function F. For suitable thresholds, $u_1, \ldots, u_d$ the marginal distributions of F, F_i, each approximately takes the form

$$F_i(x) \approx 1 - \varphi \left\{ 1 + \frac{\xi(x-u)}{\sigma} \right\}^{-1/\xi},$$

for $x > u$, with parameter sets $(\varphi_i, \sigma_i, \xi_i)$, for $i = 1, 2, \ldots, d$. We choose to transform $(X_1, \ldots, X_d)$ onto the unit Fréchet to allow us to use the simplest expression of the multivariate extreme value distribution. Therefore, define $Z_i = -1/\log(F_i(X_i))$ for each i, such that $P(Z_i \leq z) = \exp(-1/z)$ for $0 < z < \infty$. Then,

$$F_*(z_1, \ldots, z_d) = F(x_1, \ldots, x_d),$$

where F_* denotes the joint distribution function of $(Z_1, \ldots, Z_d)$. Resnick (1987) gives the following necessary and sufficient conditions for the above to hold,

$$\lim_{t\to\infty} \frac{\log F_*(tz_1, \ldots, tz_d)}{\log F_*(t, \ldots, t)} = \lim_{t\to\infty} \frac{1 - F_*(tz_1, \ldots, tz_d)}{1 - F_*(t, \ldots, t)} = \frac{\log G_*(tz_1, \ldots, tz_d)}{\log G_*(t, \ldots, t)}, \quad (3.3.14)$$

where G_* is a multivariate extreme value distribution that has Fréchet marginal distributions such that F_* is in the domain of attraction of G_*. Therefore we approximate

$$F_*(z_1, \ldots, z_d) \approx G_*(z_1, \ldots, z_d) = \exp\{-V(z_1, \ldots, z_d)\} \quad (3.3.15)$$

for $z_i > r_i$, where the r_i are high thresholds and V takes the form in equation (3.3.5) (Pickands, 1981; Coles and Tawn, 1991, 1994). Alternatively, for large n, the homogeneity of order -1 property of V allows us to make the following

approximation,

$$\begin{aligned} F_*(z_1, \ldots, z_d) &= F_*^n(z_1, \ldots, z_n)^{1/n} \\ &\approx [\exp\{-V(z_1/n, \ldots, z_d/n)\}]^{1/n} \\ &= \exp\{-V(z_1, \ldots, z_d)\}. \end{aligned}$$

The approximation in equation (3.3.4) motivates a model in which joint exceedances of suitably high thresholds allow the dependence structure of F_* to be that of an exact multivariate extreme value distribution. Note that when the marginal variables are independent, the joint distribution for F is given by the product of the marginal distributions for $x_i > u_i$.

3.3.4 Asymptotic dependence modelling

Dependence structure between variables is a popular feature of interest since it describes how variables interact. As such, selecting a suitable distribution to model the structure is vital, as different distributions can only capture certain types of dependence structures. For example, Clayton and Gumbel distributions can model independence and positive dependence whereas the Frank, Gaussian and Student-t distributions can model negative dependence, independence and positive dependence, but omits *tail dependence*.

In modelling multivariate extremes, the tail dependence structures are a critical consideration and determining the class of tail dependence is vital in model selection. The tail (extremal) dependence class, of the random variables (X_1, X_2) is determined by the probability of one variable being extreme conditional on the other also being extreme:

$$\lim_{x \to x^*} \chi(x) = \lim_{x \to x^*} \Pr(X_2 > x \mid X_1 > x) \to \begin{cases} 0, & \text{if asymptotically independent} \\ p > 0, & \text{if asymptotically dependent} \end{cases} \tag{3.3.16}$$

where x^* is the upper endpoint of F_X and $p \in [0, 1]$ is the strength of asymptotic

dependence if X_1 and X_2 are on common margins (Coles et al., 1999). Coles et al. (1999) introduced a measure to determine the strength of asymptotic independence, $\bar{\chi}$. This is given by

$$\bar{\chi} = \lim_{x \to x^*} \frac{2 \log \Pr(X_1 > x)}{\log \Pr(X_1 > x, X_2 > x)}, \tag{3.3.17}$$

which takes values $[-1, 1]$. The pair (X_1, X_2) are said to be asymptotically independent with positive association if $0 < \bar{\chi} < 1$, asymptotically independent with negative association if $-1 \leq \bar{\chi} < 0$ and independent if $\bar{\chi} = 0$.

3.4 Modelling asymptotic independence

The previous joint tail estimation methods are based on the multivariate extreme value distribution, which require the assumption that the data is either asymptotically dependent or exactly independent. However, in many environmental applications the data exhibits asymptotic independence. As such a broader range of models have been constructed by Ledford and Tawn (1996, 1997) that includes the case of asymptotic independence.

Ledford and Tawn (1996) specifies that, under broad conditions, the joint survivor function of an arbitrary random pair (X_1, X_2) with unit Fréchet marginal distributions satisfies, for large u,

$$\Pr(X_1 > u, X_2 > u) \sim \mathcal{L}(u) \Pr(X_1 > u)^{1/\eta}, \tag{3.4.1}$$

where $\mathcal{L}(u)$ is slowly varying as $u \to \infty$, that is, $\mathcal{L}(u)$ satisfies $\mathcal{L}(tu)/\mathcal{L}(u) \to 1$ as $u \to \infty$ for all fixed $t > 0$, and $\eta \in (0, 1]$ is known as the *coefficient of tail dependence*. The coefficient η describes the type of limiting dependence between the variables, and $\mathcal{L}(u)$ is the relative strength of the dependence given a value of η. The motivation for this model is detailed in Ledford and Tawn (1996), which investigates the score statistic for the bivariate normal distribution with correlation $\rho < 1$ and compares the probabilities of lying in the jointly large $(x_1 > u, x_2 > u)$ region for

asymptotic independence and exact independence. The bounding cases occur when $\eta \to 0$ and $\eta = 1$ with $\mathcal{L} = 1$, corresponding to perfect negative dependence and perfect positive dependence respectively, with $\eta = 1/2$ and $\mathcal{L}(u) = 1$ corresponding to exact independence. When $\eta = 1$ and $\mathcal{L}(u) \to c > 0$ as $t \to \infty$, the variables are classified as asymptotically dependent, and the variables are asymptotically independent when $0 < \eta < 1$. The asymptotic conditions are obtained through the conditional probability,

$$\Pr(X_1 > u | X_2 > u) \sim \mathcal{L}(u) u^{1-1/\eta}. \tag{3.4.2}$$

Ledford and Tawn (1996) identified three classes of positive dependence through equation (3.4.2), namely: asymptotic dependence ($\eta = 1$ and $\mathcal{L}(u) \nrightarrow 0$), positive association ($1/2 < \eta < 1$) and near independence ($\eta = 1/2$ and $\mathcal{L}(u) \geq 1$).
An example of asymptotic dependence is the upper tail of a bivariate extreme value distribution. The distribution function for this distribution is $F(x_1, x_2) = \exp\{-V(x_1, x_2)\}$ with V as defined in equation (3.3.5). Using the Taylor series expansion of $\exp(x)$ and the homogeneity of order -1 property of V the joint probability is given by,

$$\begin{aligned}
\Pr(X_1 > u, X_2 > u) = \bar{F}(u, u) &= 1 - \bar{F}_1(u) - \bar{F}_2(u) + \bar{F}_{12}(u, u), \\
&= 1 - 2\exp(-1/u) + \exp\{-V(u, u)\} \\
&\sim u^{-1}[2 - V(1, 1)] = [2 - V(1, 1)] \Pr(X_1 > u).
\end{aligned}$$

Consequently, asymptotic dependence is achieved when $\eta = 1$ and the slowly varying function $\mathcal{L}(u)$ takes the form $2 - V(1, 1)$ while being in the range of $(0, 1]$, exact independence when $V(1, 1) = 2$ and perfect dependence when $V(1, 1) = 1$.
We are interested in performing inferences on events in the region $(u, \infty) \times (u, \infty)$. In order to do this we can rewrite the joint probability of (X_1, X_2) being greater

than some large value u as,

$$\Pr(X_1 > u, X_2 > u) = \Pr(\min(X_1, X_2) > u), \tag{3.4.3}$$

allowing us to reduce the bivariate survivor function to a univariate variable, say $T = \min(X_1, X_2)$. Thus,

$$\Pr(T > u) = \Pr(X_1 > u, X_2 > u) \sim \mathcal{L}(u)u^{-1/\eta} \tag{3.4.4}$$

as $u \to \infty$. Coles and Tawn (1994) coined T as the *structure variable*. For a structure variable that satisfies the survivor function in equation (3.4.4), exceedances of a high threshold u satisfy

$$\Pr(T > u + t | T > u) \sim \frac{\mathcal{L}(u+t)}{\mathcal{L}(u)}(1 + t/u)^{-1/\eta} \sim (1 + t/u)^{-1/\eta} \tag{3.4.5}$$

since $\mathcal{L}$ is slowly varying. We can express this conditional probability in the form of a univariate generalised Pareto distribution,

$$\Pr(T > u + t | T > u) \sim (1 + \xi t/\sigma_u)^{-1/\xi}. \tag{3.4.6}$$

Equating equations (3.4.5) and (3.4.6) we see $\xi = \eta$ and $\sigma_u = \eta u$, which provides a method in estimating the coefficient of tail dependence. Furthermore, χ and $\bar{\chi}$ can be obtained by the relationships derived by Ledford and Tawn (1996, 1997) between η and $\mathcal{L}(u)$:

$$\bar{\chi} = 2\eta - 1,$$

$$\chi = \begin{cases} c & \text{if } \bar{\chi} = 1 \text{ and } \mathcal{L}(u) \to c > 0 \text{ as } t \to \infty, \\ 0 & \text{if } \bar{\chi} = 1 \text{ and } \mathcal{L}(u) \to 0 \text{ as } t \to \infty, \\ 0 & \text{if } \bar{\chi} < 1. \end{cases}$$

3.5 Bootstrapping

Throughout the thesis, estimates of the uncertainty of extremal quantities and statistical summaries will be derived by non-parametric bootstrapping (Efron, 1979), due to the simplistic nature compared to other methods, e.g. cross-validation and the delta method (Doob, 1935; Stone, 1974; Oehlert, 1992), which are computationally intensive and require asymptotic assumptions. Bootstrap sampling algorithms aim to generate N *pseudosamples* (bootstrap samples) $\mathbf{x}_B^{(1)}, \ldots, \mathbf{x}_B^{(N)}$ based on the original data $x_1, \ldots, x_n$. For each $i = 1, \ldots, N$, $\mathbf{x}_B^{(i)}$ is created through sampling of the original data with replacement until $x_B^{(i)}$ has length n. For each $\mathbf{x}_B^{(i)}$ we compute the measure of interest $\hat{\theta}_B^{(i)}$. With the vector of estimates $\hat{\boldsymbol{\theta}}_B$ we approximate the sampling distribution of $\hat{\theta}$ by taking the 2.5% and 97.5% quantiles of $\hat{\boldsymbol{\theta}}_B$ to construct the 95% bootstrapped confidence interval of $\hat{\theta}$.

In Chapter 4 we require sampling of bivariate data ($\mathbf{X}_1$, $\mathbf{X}_2$) which is done by sampling pairs of data, i.e. we sample with replacement (x_{1k}, x_{2k}) where k denotes an arbitrary point in the time series. In Chapter 5 we require a different approach as we need to maintain the temporal structure. One approach would be to split the data into b equally sized blocks and sample with replacement the blocks until $\mathbf{x}_B$ is of length n. However, a more technical approach is adopted and is explained in Section 5.3.4.

Chapter 4

Evaluation of process-based extreme forecasts

4.1 Introduction

Air quality models (AQMs) are numerical models designed to emulate the underlying processes of meteorological variables and estimate the degree of emissions. Air quality models can provide deterministic real-time forecasts which are used to assess the risks from poor air quality to both the environment and public health and to aid in policy making. These models have become increasingly complex and sophisticated due to the importance of including extra layers such as the atmospheric composition (Eyring et al., 2013; Kukkonen et al., 2012). Model validation is a key aspect of the modelling process, but it is specifically important for poor air quality episodes since there is strong evidence linking short-term exposure to poor air quality and respiratory issues and mortality (EEA and WHO, 2007; Ji et al., 2011; Lavigne et al., 2014; Lelieveld et al., 2019; Vohra et al., 2021). As such, increasing awareness of extreme events and introducing new ways to assess the performance of AQMs at the extremal level is vital in providing sufficient public warnings.

The most common method of model validation is a direct comparison between the model output and measurement data, for instance monitoring network mea-

surements or satellite observations. There is a vast selection of statistical tools available to perform such a comparison. These include: summary statistics, which capture the variability and summaries the bias and linear relationship in the mean behaviour (Borrego et al., 2008; Thunis P, 2013; Lin et al., 2017; Neal et al., 2014), categorical metrics, which are used as a measure of model skill and to evaluate risk of health-threatening events (Kang et al., 2005; Savage et al., 2013) and statistical distance measures, which examines the difference between the underlying probability distributions and the significance of such difference (Düsterhus and Hense, 2012). Examples of each type can be seen in Table 4.1.

Measure	Type	Formula
Root-mean-squared error (RMSE)	Summary	$\sqrt{\sum_{i=1}^{n} \frac{(m_i - o_i)^2}{n}}$
Hit rate	Categorical	$\Pr(m > x \mid o > x), x \in \mathbb{R}$
Kullback-Leibler divergence (KLD)	Distance	$\sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$

Table 4.1: A selection of different statistical tools available to compare model output and measurement data, where m is the model output, o is the observations and Q and P are the respective probability distributions.

These statistical tools for model validation can be extended, for example Jolliff et al. (2009) proposed the use of RMSE normalised by the standard deviation of the observations in order to provide a standardised metric to compare model outputs. Further, Thunis et al. (2012) introduced a new method to evaluate air quality models through deriving a consistent set of Model Performance Criteria (MPC) for four statistical indicators. The MPC is calculated based on the observation uncertainty, i.e. the statistical indicators are normalised by the uncertainty in the observations. However, as the MPC are based on a comparison of data characteristics with their uncertainty, the MPC will vary by site type, pollutant, geographic area etc.
Although such measures provide a solid basis for air quality model validation, AQMs rarely consider extreme events in their model validation process and are optimised to predict the mean behaviour of the process. Since events (extremes) occurring in the

tails are typically rare, RMSE and correlation coefficients can mask the behaviour of the extreme values, especially if the data is of high temporal resolution. The use of categorical metrics is an initial step in exploring the capabilities of capturing the behaviour in the extremes as they can be constructed to measure accuracy in forecasting extremes. However, these metrics quantify only a very specific aspect of the model performance, and there are other characteristics of the predicted data that it would be useful to explore.

This chapter will look at two air quality datasets for the UK, the Automatic Urban and Rural Network (AURN) and the Air Quality in the Unified Model (AQUM), as described in Section 2.4. The latter is a numerical process-based forecasting model that includes a post-processing bias correction technique (Neal et al., 2014). This technique has seen an improvement in the forecasting skill of ozone during July 2012 to July 2013, with Pearson correlation coefficient increasing from 0.64 to 0.76, bias decreasing from 7.02 to $0.53\,\mu g/m^3$ and the RMSE decreasing from 20.85 to $15.42\,\mu g/m^3$ (Neal et al., 2014). However, there have been no investigations into the behaviour of extremes in these bias-corrected forecasts.

We demonstrate how extreme value theory can be used as a validation tool for the AQUM ozone forecasts. There are some previous examples of similar studies in the literature. Contzen et al. (2022) uses a clustering block maxima approach to validate the upper tail of the AWI-ESM global climate model. Weller et al. (2012) used bivariate extreme value theory to investigate the pineapple express phenomenon and validate Weather Research and Forecasting regional climate model, driven by reanalysis. In contrast, we explore the marginal characteristics of extreme ozone events by using the peaks-over-threshold (POT) approach to fit a generalised Pareto distribution (GPD) (Smith, 1989; Davison and Smith, 1990) and estimate return levels. We also use multivariate extreme value methods to identify the class of extremal dependence between the observations and forecasts as this provides clarity to the forecasting capability of AQUM at the extremal level. Finally, we estimate hit rates of exceeding relevant thresholds, which is the conditional probability

$\Pr(X_{AQUM} > x \mid X_{AURN} > x)$ for some sufficiently large x, through modelling the joint probabilistic behaviour of AURN and AQUM.

The chapter is structured as follows. After introducing the methodology in Section 3.2, we present the ozone datasets used in this study in Section 3.3. The results from applying extreme value theory to the data are presented in Sections 3.4 and 3.5. A section on conclusions finalises the chapter.

4.2 Extreme value theory background

Extreme value theory provides a framework for studying the behaviour of rare events, i.e. the upper tail of a probability distribution where the number of observations is intrinsically small. We provide a brief review of the methodology with focus on the bivariate case.

4.2.1 Univariate threshold exceedances

An extreme value analysis does not require knowledge of the underlying distribution and allows for asymptotically justified consideration of only data deemed extreme. Two types of modelling approaches exist: modelling block maxima and modelling exceedances of a high threshold. If a random variable X is in the domain of attraction of an extreme value distribution, that is there exists sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that $\lim_{n\to\infty} F^n(a_n x + b_n) \to G(x)$ where F is the marginal distribution of X and G is the generalised extreme value distribution, then in the limit the conditional distribution of $X|X > u$ follows a generalised Pareto distribution (GPD), as given in equation (3.2.5).

The GPD parameters can be estimated and used to obtain the T-year return level, a useful quantity to estimate as it allows one to extrapolate beyond the observed data period and is often used to guide legislation changes. The T-year return level is the value expected to be exceeded once every T-years, under the assumption of no system change, which can be estimated using equation (3.2.10).

4.2.2 Bivariate extremes and tail dependence

For random variables (X_1, X_2) the covariance and correlation measures are useful dependence measures. However, they do not capture the tail dependence as they are dominated by the main body of the distribution. As such, we categorise the joint upper tail differently to the whole joint distribution. The tail dependence class of the random variables (X_1, X_2), given they are on common margins, can be quantified by the measure χ, established by Coles et al. (1999) and is given by equation (3.3.16). Whilst Coles et al. (1999) provides a measure of asymptotic independence $\bar{\chi}$, Ledford and Tawn (1996, 1997) developed an early formation for different tail dependences. Assuming X_1 and X_2 have standard Fréchet marginal distributions so that $\Pr(X_i < x) = \exp(-1/x)$ for $x > 0$ and $i = 1, 2$, then the joint tail behaviour can be modelled using the model of Ledford and Tawn (1996), as described in Section 3.4 and by equation (3.4.1).

The coefficient of tail dependence η determines the strength of association in the asymptotically independent case and indicates that the pair (X_1, X_2) are asymptotically dependent when $\eta = 1$. Consequently, Ledford and Tawn (1996, 1997) developed a hypothesis test using a likelihood ratio test to determine the type of tail dependence testing $H_0 : \eta = 1$ against $H_1 : \eta \neq 1$.

4.2.3 Spectral measure

The extremal dependence structure can be characterised by the spectral measure (Fougères, 2004). A commonly used tail characterisation of $\mathbf{X}$ is in terms of its pseudo-polar coordinates $(R, \mathbf{W})$, the radial and angular components respectively. Consider the random vector $\mathbf{X} = (X_1, X_2)$, with standard Fréchet marginal distributions. As such the radial and angular components, for an arbitrary norm $||\cdot||$, are defined as $R = ||\mathbf{X}||$ and $\mathbf{W} = \mathbf{X}/R$, with $R > 0$ and $\mathbf{W} \in \mathcal{S} = \{(\omega_1, \omega_2) \in [0, 1]^2 : \omega_1 + \omega_2 = 1\}$, the unit simplex. Here we take the L_1 norm, i.e. $||\mathbf{x}|| = x_1 + x_2$. Then under the assumption of multivariate regular variation, that is if there exists a sequence a_n such that on Borel subsets of $\mathcal{C} = [\mathbf{0}, \infty) \backslash \{\mathbf{0}\}$

$$n \Pr\left(\frac{\mathbf{X}}{a_n} \in \cdot\right) \to \nu(\cdot) \tag{4.2.1}$$

as $a_n \to \infty$ and $n \to \infty$ where ν is a measure (Resnick, 2007),

$$\lim_{n\to\infty} \Pr\left(\frac{R}{a_n} > r, \mathbf{W} \in B\right) \to r^{-1} H(B) \tag{4.2.2}$$

for $r \geq 1$, where H is the spectral measure that satisfies the moment constraint

$$\int_{\mathcal{S}} \omega(1-\omega) dH(\omega) = \frac{1}{2}. \tag{4.2.3}$$

Throughout we will assume that H is differentiable such that a spectral density h exists. Note that under this definition the pair is: independent if $H(\{0\}) = H(\{1\}) = 1/2$, perfectly dependent if $H(\{1/2\}) = 1$, asymptotically independent if all the mass in H lies uniformly on the vertices, and asymptotically dependent if the mass in H lies away from the axes. This formulation of H links to the *exponent measure* V of the bivariate extreme value distribution (BVEVD) which is defined as,

$$F(x_1, x_2) = \exp\{-V(x_1, x_2)\} \text{ for } x_k > 0, \tag{4.2.4}$$

where $k = 1, 2$ and V is defined as

$$V(x_1, x_2) = \int_{\mathcal{S}} \max\left(\frac{\omega}{x_1}, \frac{1-\omega}{x_2}\right) dH(\omega). \tag{4.2.5}$$

Coles and Tawn (1991) provide justification for the use of the BVEVD to model joint exceedances of sufficiently large thresholds (u_1, u_2). As such, to obtain estimates for the dependence parameters $\boldsymbol{\theta}$ contained in the model for V we have to maximise, using numerical optimization, the censored likelihood

$$L(\boldsymbol{\theta}; \mathbf{x_1}, \mathbf{x_2}) = \prod_{i=1}^{n} \varphi(\boldsymbol{\theta}; (x_{1i}, x_{2i})), \tag{4.2.6}$$

where

$$\varphi(\boldsymbol{\theta};(x_1,x_2)) = \begin{cases} \frac{\partial^2 F}{\partial x_1 \partial x_2}\Big|_{(x_1,x_2)} & \text{if } (x_1,x_2) \in R_{1,1}, \\ \frac{\partial F}{\partial x_1}\Big|_{(x_1,u_2)} & \text{if } (x_1,x_2) \in R_{1,0}, \\ \frac{\partial F}{\partial x_2}\Big|_{(u_1,x_2)} & \text{if } (x_1,x_2) \in R_{0,1}, \\ F(u_1,u_2) & \text{if } (x_1,x_2) \in R_{0,0}, \end{cases}$$

and $R_{0,0} = (-\infty, u_1)\times(-\infty, u_2)$, $R_{1,0} = [u_1, \infty)\times(-\infty, u_2)$, $R_{0,1} = (\infty, u_1)\times[u_2, \infty)$, $R_{1,1} = [u_1, \infty) \times [u_2, \infty)$, (Coles, 2001). Each term is derived from the joint tail approximation in equation (4.2.4).

4.3 Ozone data pre-processing

The hourly ozone observations from AURN, described in Section 2.4.1, and the hourly forecasts from AQUM, described in Section 2.4.2, are aggregated to the daily maximum of the 8-hour running mean scale (DM8). At each of the 61 sampled AURN sites, the observations are then matched to the corresponding forecasts obtained by using the nearest-neighbour algorithm, i.e. matched to the closest grid cell. Since we are interested in the period where most exceedances occur, we discard all pairs which do not fall within the site-specific ozone season, as defined in Section 2.4.3, and discard all days that are not complete, i.e. either or both observations and forecasts have no data available. We use subscript $AURN$ and $AQUM$ to denote which dataset we are using and subscript j to denote the jth site.

4.4 Investigating the marginals

The AQUM forecasts determine the publicly displayed Daily Air Quality Index (DAQI), as such it would be desirable if AQUM captures the observations well, especially at higher concentrations due to the greater public health risk. We begin

by examining the differences between AURN and AQUM, that is, the extent to which extreme AQUM forecasts correspond to the extreme AURN observations. We first follow standard methodology by performing marginal quantile comparisons and evaluating metrics above and below the 90th quantile to motivate the use of extreme value methods.

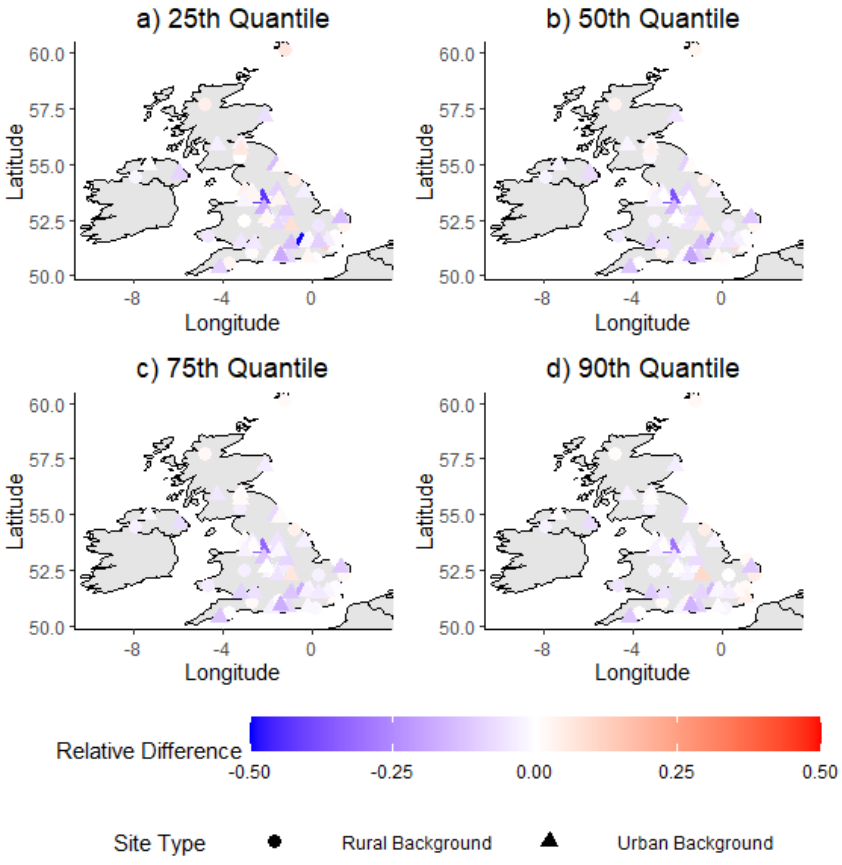


Figure 4.1: Relative difference in the marginal quantiles of the pairwise AURN DM8 ozone data and the AQUM DM8 ozone forecasts.

4.4.1 Marginal characteristics

We first look at the marginal characteristics of AURN and AQUM at each site separately. Figure 4.1 shows the relative difference; defined as $(X_{AURN} - X_{AQUM})/X_{AURN}$; in marginal 25th, 50th, 75th and 90th quantiles between AURN observations and AQUM forecasts. Negative values indicate AQUM forecasts

overestimates the observed ozone concentration whereas positive values indicate an underestimation, assuming the AURN measurements are all correct. On average the forecasts are closer to the observations in Scotland and North England than the rest of the UK, as seen in Figure 4.1, however the common trend is AQUM overestimates the ozone distributions. The largest overestimations occur at UB sites, with all four overestimations of at least 15% occurring at UB sites.

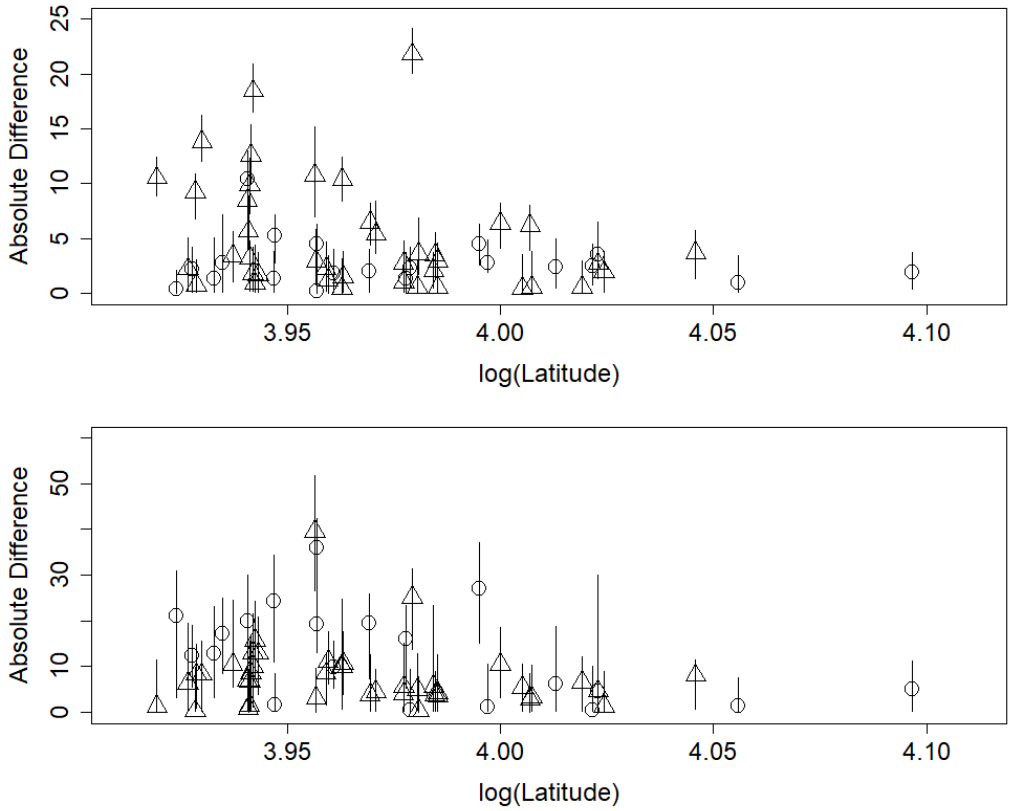


Figure 4.2: Absolute difference in the marginal 90th and 99th quantiles of AURN observations and AQUM forecasts, respectively, with 95% bootstrapped confidence intervals. (○) points denote rural background sites and (△) points denote urban background sites.

Across all common site types the average of the absolute (relative) difference of the four estimated quantiles is $2.27 - 2.72\,\mu g/m^3$ (-0.18% to 1.13%) and $4.54 - 5.01\,\mu g/m^3$ (-4.20% to -3.10%) for RB sites and UB sites respectively, as seen in Figure 4.2. The maximum underestimation of $10.80\,\mu g/m^3$ (relative difference 9.84%) occurs at an UB site in Nottingham, whereas the largest overestimation of

$21.78\,\mu g/m^3$ (relative difference 30.24%) occurs at an UB site in Manchester. The difference between AURN and AQUM increases with quantile, however the sampling uncertainty in the estimates also increase as depicted by the 95% bootstrapped confidence intervals in Figure 4.2. The average absolute difference is $12.81\,\mu g/m^3$ for RB sites and $7.66\,\mu g/m^3$ for UB sites.

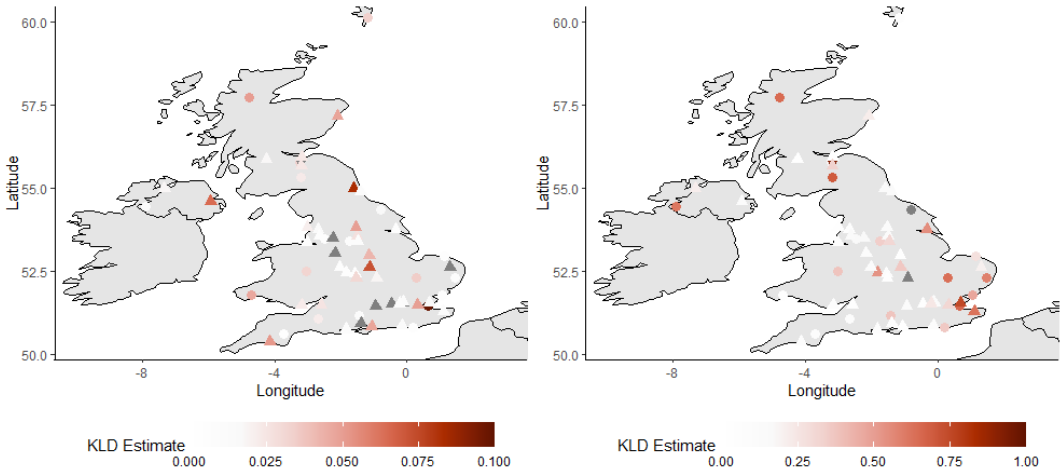


Figure 4.3: Map showing the KLD estimate between the AURN observations and the AQUM forecasts for below and above the marginal 90th quantile, respectively. (○) points denote rural background sites and (△) points denote urban background sites. Note: sites in grey indicate values above the upper endpoint of the colouring scale.

Next, we evaluate the Kullback-Leibler divergence (KLD) and root-mean squared error (RMSE) metrics above and below the site-specific 90th quantile of AURN (denoted $q_{90,AURN}$). The latter metric indicate the capability of the model to capture the observations conditional on position in a time series, whereas the former is a measure of closeness between two probability distributions. Figure 4.3 shows the KLD estimate between the probability distributions of AURN observations and AQUM forecasts below and above the marginal 90th quantile of AURN respectively, where a value of zero indicates perfect match between the two probability distributions. The KLD estimates for below $q_{90,AURN}$ indicates that AQUM approximately captures the distribution of ozone, with an average of 0.029 and 0.064 for RB and UB sites respectively. However, the probability distribution tails (above $q_{90,AURN}$) of the rural background sites in North England and Scotland,

as well as all sites in East Anglia and South-East England, are poorly captured as indicated by larger KLD estimates. On average the tails of RB sites are captured more poorly than the UB sites, with average KLD estimates of 0.437 and 0.278 respectively.

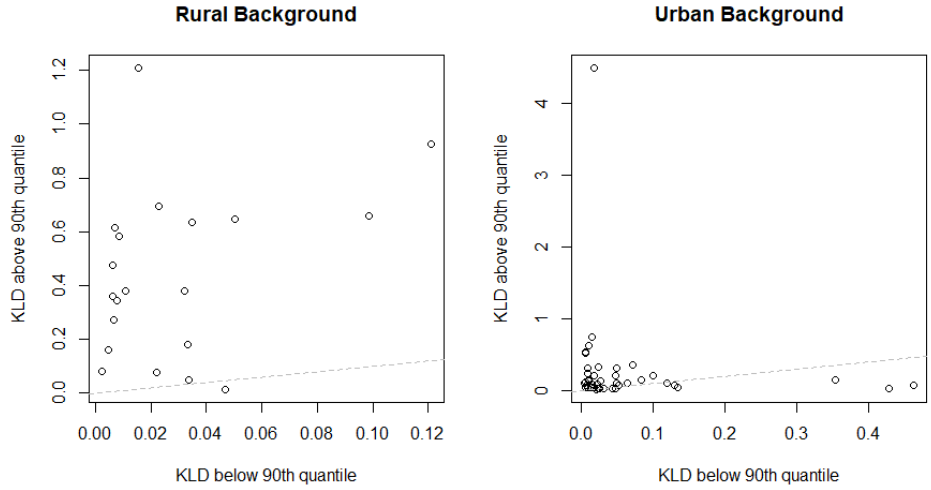


Figure 4.4: Scatterplot of the KLD estimate above and below the marginal 90th quantile at rural background sites and urban background sites, respectively. The dotted line is the $y = x$ line.

The scatterplot of KLD estimates, as seen in Figure 4.4, highlights the difference between the captured marginal distribution below $q_{90,AURN}$ and above $q_{90,AURN}$. One key observation is, the smaller the discrepancy below the marginal threshold at urban background sites the larger the discrepancy range above the marginal threshold. This indicates the skewness in the forecasted distribution, i.e. AQUM struggles to capture the upper tail behaviour of ozone at urban background sites more so when the distribution below $q_{90,AURN}$ is captured well.

Figure 4.5 shows the RMSE given above and below the site-specific $q_{90,AURN}$. The across site average RMSE is 12.62 $\mu g/m^3$ for below $q_{90,AURN}$ and 17.54 $\mu g/m^3$ above. No obvious spatial pattern can be observed in the RMSE below the 90th quantile, however above the 90th quantile we see larger values on the east-coast and across South England. Further, RB sites on average see better forecasts below $q_{90,AURN}$ but poorer forecasts above. Across all RB sites the average bias in the data above

$q_{90,AURN}$ is 12.72 $\mu g/m^3$ compared to 6.28 $\mu g/m^3$ at UB sites, indicating AQUM is more than 2-fold underestimating the largest concentrations in rural areas compared to urban background areas. Note, AQUM provides accurate forecasts for when AURN observations are below their respective 90th quantiles.

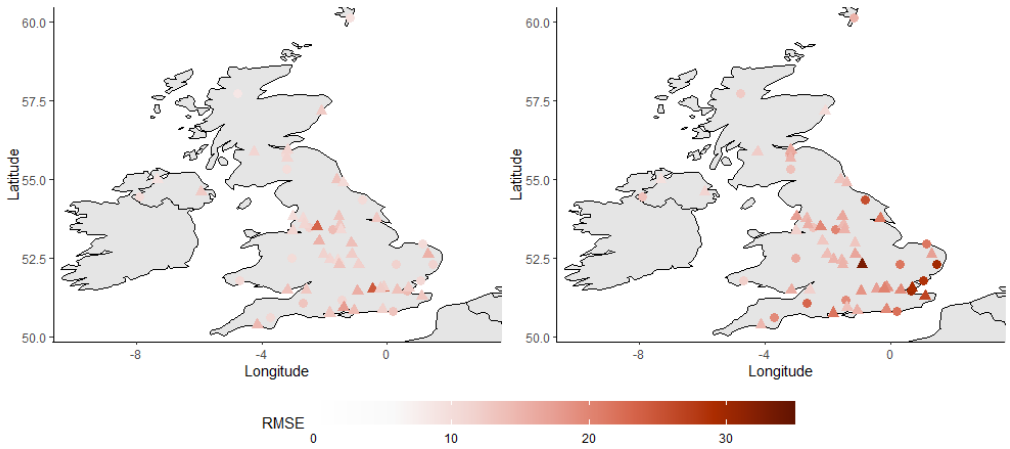


Figure 4.5: Root-mean squared error (in $\mu g/m^3$) between AURN observations and AQUM forecasts conditional on (left) AURN being below the site-specific 90th quantile and (right) AURN being above the site-specific 90th quantile. (○) points denote rural background sites and (△) points denote urban background sites.

4.4.2 Modelling threshold exceedances

While the bulk seems to be captured relatively well, the metrics studied in Section 4.4.1 indicate relatively poor matchings between AURN observations and AQUM forecasts for higher quantiles, i.e. the marginal tails. As such, as a preliminary step in examining the upper tail dependence in (AURN, AQUM), the generalised Pareto distribution (GPD) is used to estimate the margins separately. We fit a GPD to the exceedances of a suitably high threshold at each site and for each dataset separately. The suitably high threshold is determined through stability plots (Coles, 2001) and is taken to be the site-specific dataset-specific marginal 90th quantile.

Figure 4.6 shows the difference between the maximum likelihood estimates of the scale and shape parameters for AURN and AQUM. There is a large discrepancy in the magnitude of the scale parameter where in some instances the AQUM estimates are more than half the AURN estimates. Further, the AQUM forecasts misidentifies

the type of tail structure at 39% of sites, i.e. estimates of the shape parameter are of opposite sign. However, it is encouraging that the estimates of the shape parameter ξ are of the same sign at 61% of sites. These plots would suggest that on average the threshold exceedances come from differently scaled but similarly shaped distributions.

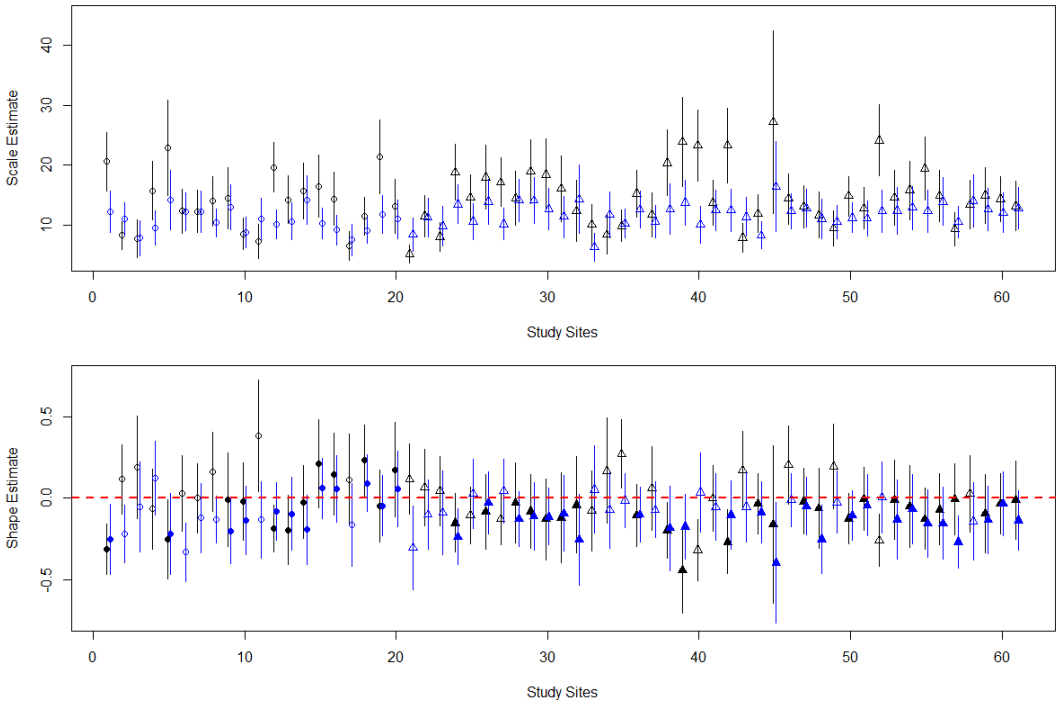


Figure 4.6: Plot showing the maximum likelihood estimates of the (top) scale parameter and (bottom) shape parameter of the generalised Pareto distribution fitted to (black) the AURN observations and (blue) the AQUM forecasts, with vertical lines representing the 95% confidence intervals. Solid symbols indicate the same sign of shape parameter and hollow symbols indicate opposite sign between the AURN and AQUM estimates. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

Figures 4.7 and 4.8 show the differences between the estimated one- and ten-year return levels from the individually fitted generalised Pareto distributions between AURN and AQUM. The one(ten)-year return level estimates of AQUM underestimates by at most 43.51(98.62) $\mu g/m^3$ and overestimates at most 23.13(16.36) $\mu g/m^3$. On average RB sites have the largest return level estimates and consequently the largest differences observed. Rural background sites on average have wider confidence intervals than urban background sites, which could be attributed to the

larger shape parameters and wider ranges of observed concentrations. The one-year estimates from AQUM are typically under the lower bound of the confidence intervals of AURN at rural background sites. Whereas on average the confidence intervals overlap for the ten-year return level estimates. Further, sites which are located on the east-coast have larger underestimations with the greatest being in East Anglia (site numbers 15, 16 and 46 in Figures 4.6 and 4.7) and in Nottingham (site number 45 in Figures 4.6 and 4.7).

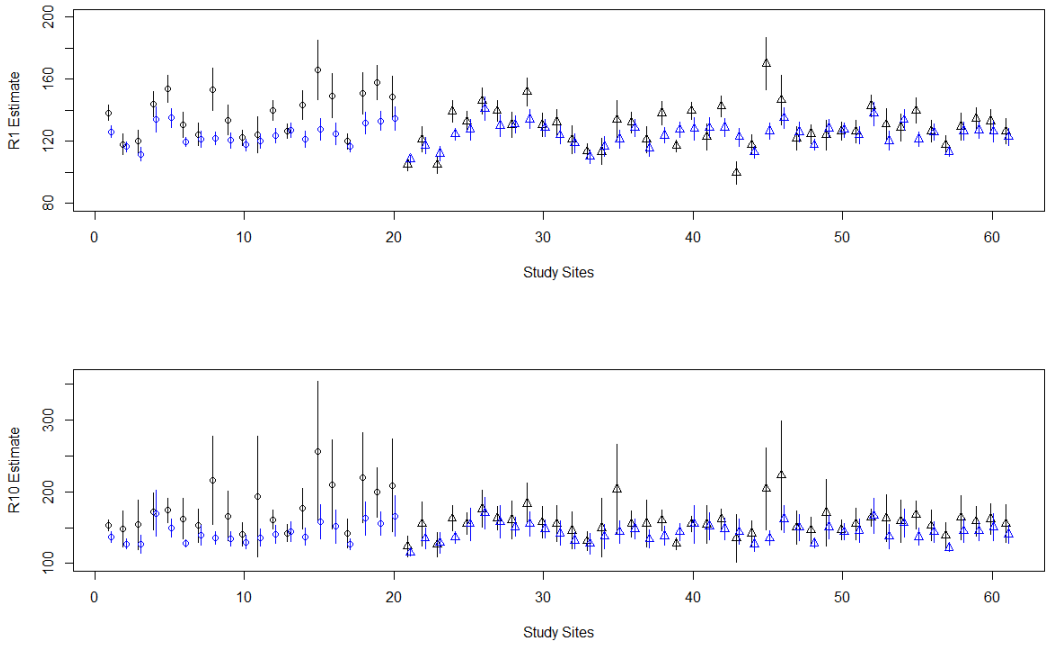


Figure 4.7: Plot showing the estimated (top) one-year return levels and (bottom) ten-year return levels from the individually fitted generalised Pareto distribution to (black) the AURN observations and (blue) the AQUM forecasts, with vertical lines representing the 95% confidence intervals. (○) points denote rural background sites and (△) points denote urban background sites.

To examine the GPD model fits further we pool the data across the two datasets and fit three separate models. These being a model in which the two datasets have common scale and shape parameters (M1), one in which the datasets have a common shape parameter but different scale parameter (M2), and one in which both scale and shape parameters are *free* between the datasets (M3). Since these are nested (M1 in

M2, M2 in M3) we can then do a likelihood ratio test (LRT) between these models to see whether a single set of parameters is sufficient to describe both distributions, or whether you need separate scale parameters, or whether you need the full flexibility of separate scale and shape parameters.

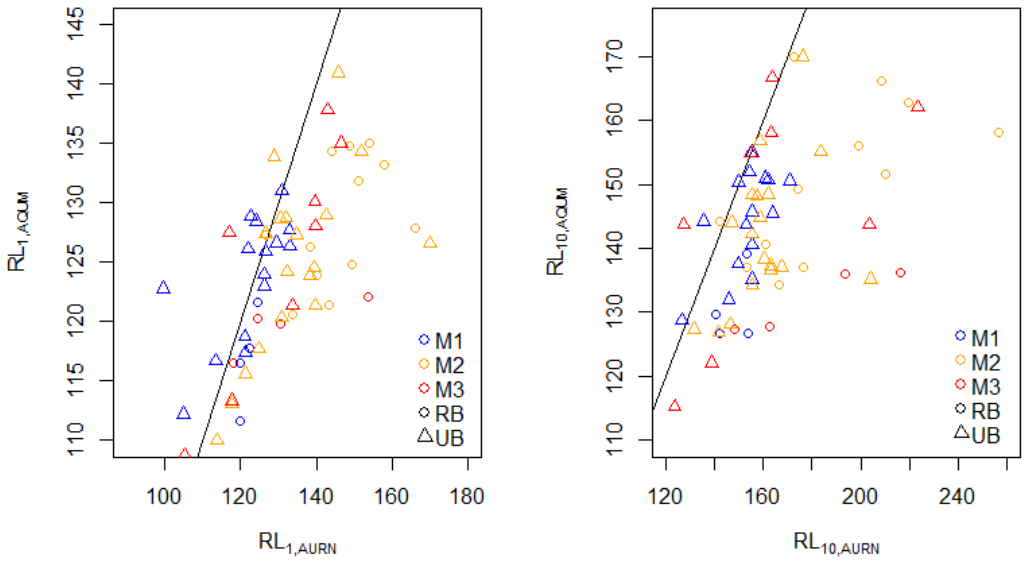


Figure 4.8: Scatterplot comparing the estimates of the (left) one-year return levels and (right) ten-year return levels using the AURN observations and AQUM forecasts. Colours denote the outcome of the LRT. (○) points denote rural background sites and (△) points denote urban background sites.

The LRT indicated that 80% of RB sites required either the scale or both parameters to be free and 63% of UB sites required either the scale or both parameters. This indicates that the observations and forecasts have the same marginal distribution at 20% of RB sites and 37% of UB sites. These results are reflected in the return level estimates shown in Figure 4.8.

4.5 Investigating and modelling tail dependence

The LRT showed that the marginal distributions of the extreme observations and forecasts differed at 69% of the studied sites. However, to estimate the probability of joint exceedance, that is where both AURN and AQUM are extreme at the

same time, we investigate the upper tail dependence structure. Probability integral transformations are applied in each margin by using the fitted GP distributions above the marginal threshold and the empirical distribution function below. That is, we model the marginal distribution of X as

$$F_X(x) = \begin{cases} 1 - \varphi_u \left[1 + \hat{\xi}\left(\frac{x-u}{\hat{\sigma}_u}\right)\right]_+^{-1/\hat{\xi}}, & \text{for } x > u, \\ \tilde{F}_X(x), & \text{for } x \leq u, \end{cases}$$

where $c_+ = \max(c, 0)$, u is a suitably high threshold, $\varphi_u = \Pr(X > u)$, $\tilde{F}_X(x)$ is the empirical marginal cumulative distribution function of X and $\hat{\sigma}_u > 0$ and $\hat{\xi} \in \mathbb{R}$ are the maximum likelihood estimates of the scale and shape parameters respectively. Then to obtain unit Fréchet margins, we use the probability integral transformation $\tilde{X} = -1/\log F_X(x)$. This allows the identification and modelling of tail dependence through the use of the bivariate extreme value distribution model, as given in equation (4.2.4).

4.5.1 Extremal dependence estimation

To assess whether the AQUM forecasts and AURN observations exhibit asymptotic dependence we use the hypothesis test of Ledford and Tawn (1996). For exceedances of the marginal 90th quantile, this results in p-values greater than 0.05 for all but two sites, Lerwick (LERW) and Portsmouth (PMTH), suggesting $\eta = 1$, that is the joint tail exhibits asymptotic dependence for all sites except Lerwick and Portsmouth. Increasing to the marginal 95th quantile, no sites indicate asymptotic independence at up to significance level of 0.1.

We estimate the strength of asymptotic dependence $\hat{\chi}(q_{95})$ where q_{95} is the 95th empirical quantile in each margin. Across all sites this ranges between 0.228 $(0.013, 0.313)$ and 0.731 $(0.670, 0.822)$ where the 95% bootstrapped confidence intervals are given in parentheses. On average UB sites having stronger asymptotic dependence which indicates closer joint tails and thus better forecasts of AQUM.

The $\hat{\chi}(q_{95})$ estimates for Lerwick and Portsmouth are respectively $0.271\,(0, 0.630)$ and $0.434\,(0.171, 0.698)$, where the 95% confidence intervals are given in parentheses. The estimates at Portsmouth suggest asymptotic dependence, whereas at Lerwick the estimate and confidence intervals suggest weak asymptotic dependence to asymptotic independence. Figure 4.9 (middle) shows the estimated $\chi(q_{95})$ across all sites. The sites along the coast on average have lower tail dependence and thus suggest poorer forecasts of extreme ozone events. This could be due to the resolution of AQUM compared to the observations, i.e. picking up a sea point with different ozone characteristics rather than a point over land. Further, sites that are closer together have similar tail dependence. This could be expected since local atmospheric conditions are more likely to be similar, however site type may dominant the level of similarity.

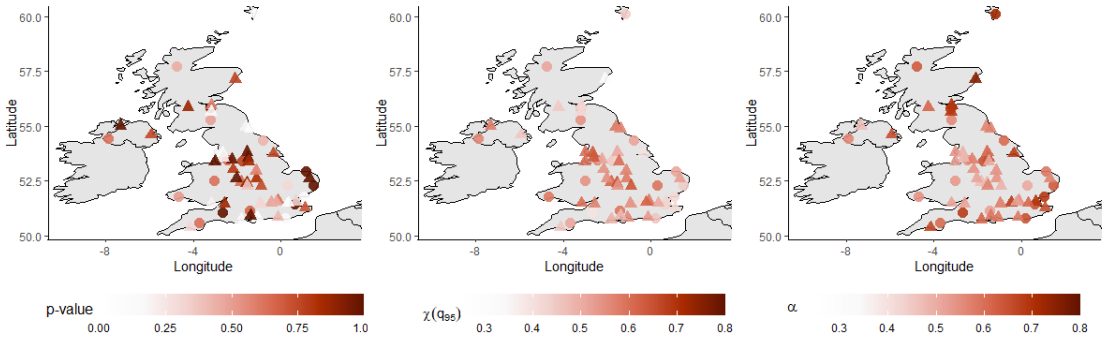


Figure 4.9: Maps showing (left) p-values from the L&T hypothesis test (middle) χ evaluated at the 95th quantile and (right) estimate of the dependence parameter α from the logistic spectral measure model, in equation (4.5.1). ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

4.5.2 Models for the spectral measure

The extremal dependence measure and the hypothesis test indicate the tails of AURN and AQUM are asymptotically dependent. As such we use the established procedure detailed in Section 4.2.3 and Coles (2001) to estimate the dependence parameters contained in the model for V. We start by fitting a one-parameter

model whose spectral density is given by

$$h(\omega;\alpha) = \frac{1}{2}(\alpha^{-1}-1)\{\omega(1-\omega)\}^{-1-1/\alpha}\left[\omega^{-1/\alpha}+(1-\omega)^{-1/\alpha}\right]^{\alpha-2}, \qquad (4.5.1)$$

which corresponds to the logistic model with dependence parameter α for the exponent measure of the bivariate extreme value distribution (Gumbel, 1960). Estimating the parameters of the spectral density, we obtain estimates for α ranging from 0.469 to 0.756. These values indicate weak to moderate asymptotic dependence; as $\alpha \to 1$ the variables become independent and as $\alpha \to 0$ the variables are perfectly dependent. Figure 4.9 (right) presents the estimated dependence parameter α at each site. The sites closer to the coastline exhibit weaker tail dependence compared to those further inland, which is in agreement with $\hat{\chi}(u_{95})$.

We start to increase the complexity of the spectral measure by using the two-parameter Hüsler-Reiss distributions (Hüsler and Reiss, 1989). Then we increase the flexibility of the spectral measure by accounting for asymmetry, i.e. removing the exchangeability of X_1 and X_2. We use several different models that account for asymmetry: asymmetric logistic model (Tawn, 1988, 1990), the negative logistic model (Galambos, 1975), the negative asymmetric logistic model (Joe, 1990) and the Dirichlet distribution (Coles and Tawn, 1991).

Figure 4.10 shows the estimated spectral densities $h(\omega)$ at two sites: Aston Hill (AH), a rural background site, and Wirral Tranmere (TRAN), an urban background site. These figures provide a visual representation of the dependence structure and provide a visual diagnostic for the goodness-of-fit of the models. The asymmetry in the dependence structure at Wirral Tranmere is captured well by Dirichlet model and the asymmetric logistic model. At Aston Hill, the dependence structure is less obvious, and each model captures a different feature. For instance, the Hüsler-Reiss model captures the two peaks, however puts equal mass between them, since it is a symmetric model, whereas the Dirichlet model captures the asymmetry and the decreasing frequency of larger ω.

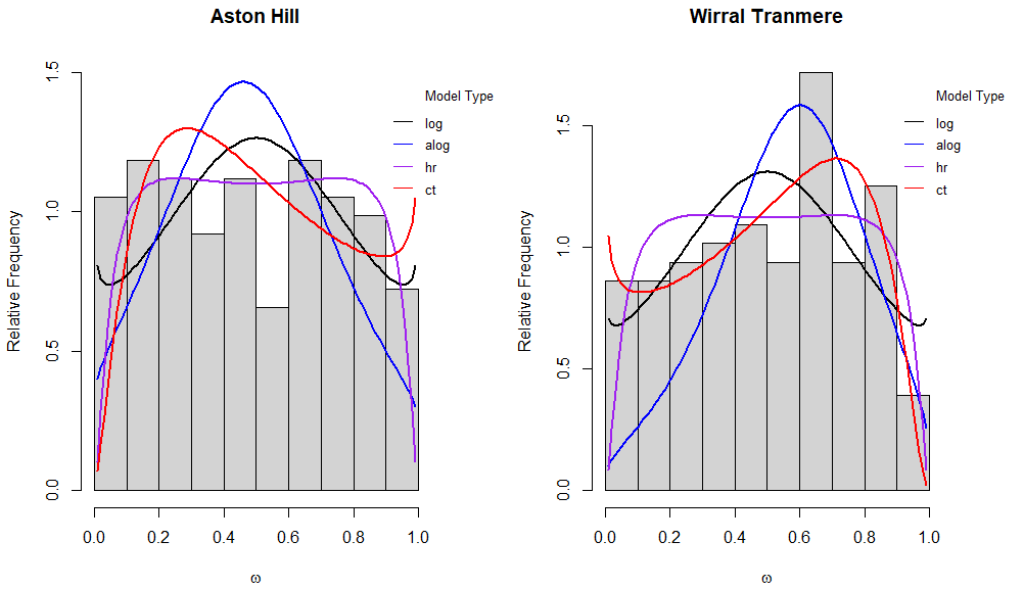


Figure 4.10: Histogram of ω values associated with the exceedances of the marginal 90th quantile at Aston Hill, a rural background site and Wirral Tranmere, an urban background site, respectively. Lines show estimates of $h(\omega)$ for models: *log* (logistic), *alog* (asymmetric logistic), *hr* (Hüslser-Reiss) and *ct* (Dirichlet).

To determine which model is preferable we chose the model which has the lowest Akaike information criterion (AIC). This results in a Dirichlet distribution model for Aston Hill and an asymmetric logistic model for Tranmere Wirral. Across all sites this results in 1 asymmetric logistic model, 7 negative logistic models, 14 Dirichlet distribution models, 19 logistic models and 20 Hüsler-Reiss distribution models. Sites close together require the same type of model, with the majority of sites on the east-coast requiring a logistic model, as seen in Figure 4.11. Figure 4.11 also presents the empirical proportion of joint exceedances ($n_{u,v}$) to the marginal AURN exceedances (n_u), often referred to as the empirical hit rate, which will be used to assess model fit in the following section.

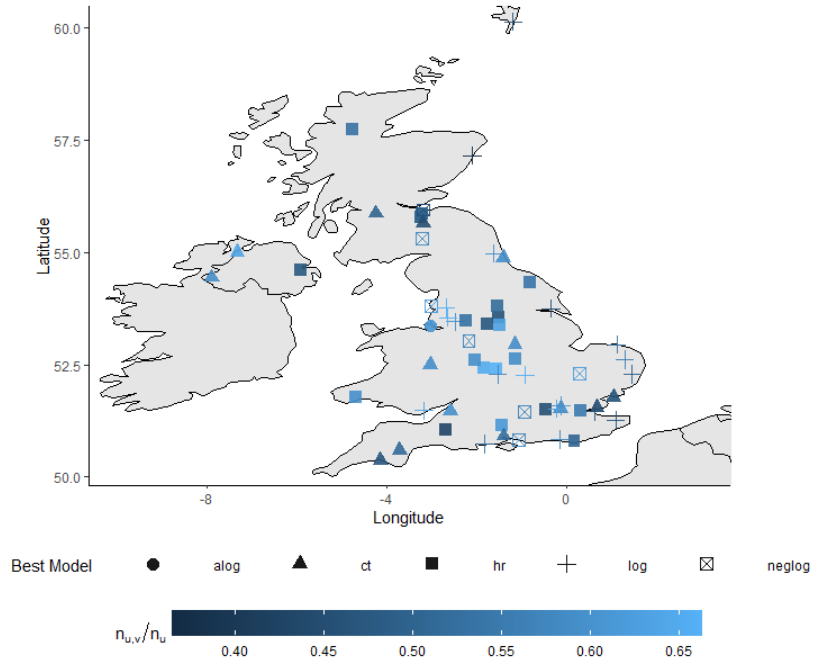


Figure 4.11: Map showing the preferable model at each site determined by the AIC value and coloured by the empirical proportion of joint exceedances to the marginal exceedances. Model acronyms: *log* (logistic), *alog* (asymmetric logistic), *neglog* (negative logistic), *ct* (Dirichlet), *hr* (Hüsler-Reiss).

4.5.3 Conditional metrics for categorical prediction

Conditional metrics, such as false-alarm ratio and hit rate, are frequently used to evaluate the forecasting capability of a model and highlight the risk of health-threatening events. Consequently, this can indicate the need for further risk management and can drive changes to policies. As such, we demonstrate the capabilities of the bivariate extreme value model in capturing the tail behaviour of AQUM forecasts and AURN observations by obtaining model-based estimates of the hit rate, the conditional probability $\Pr(X_{AQUM,j} > x \mid X_{AURN,j} > x)$ where $x \geq q_{90,j}$. To obtain such estimates, we use the survivor function of the bivariate extreme value model associated with the spectral measure chosen by minimising AIC. We set x to be 100, 120, 140 and 160 $\mu g/m^3$ which correspond to the moderate to high (4 to 7) risk bands of the Daily Air Quality Index (DEFRA and UK-AIR, 2022).

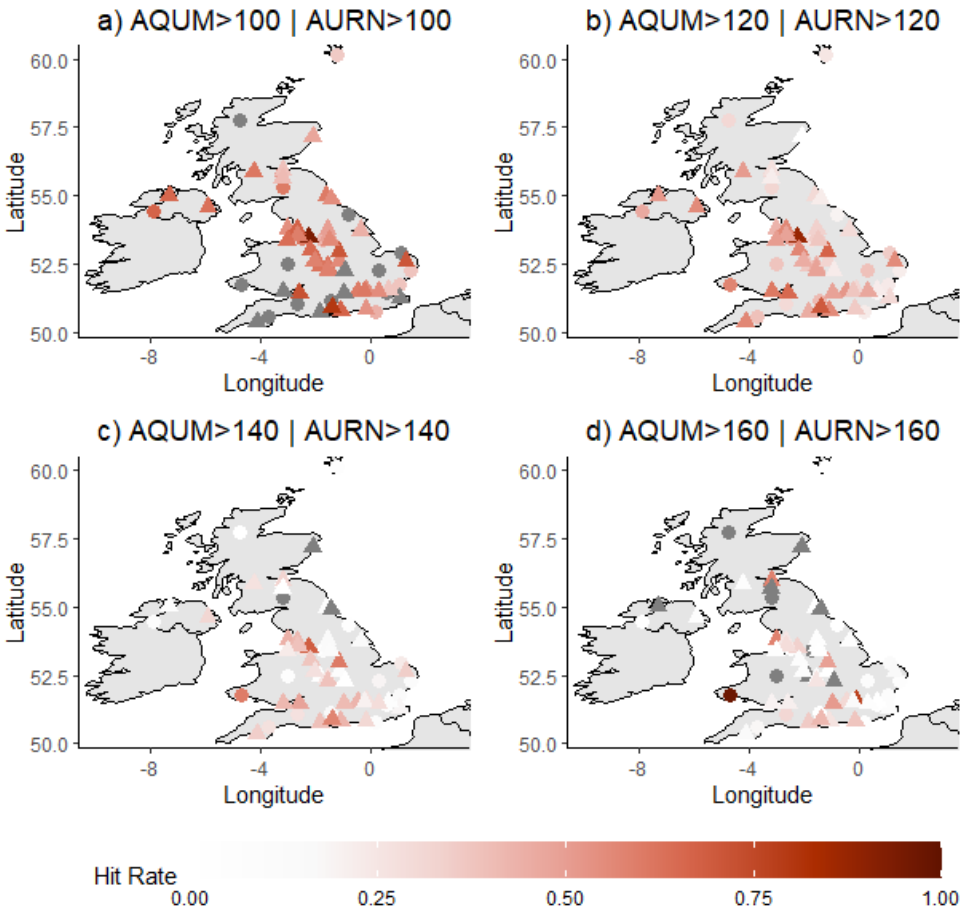


Figure 4.12: Map showing the hit rate of AQUM, that is AQUM correctly forecasting the concentration being above a) $100\,\mu g/m^3$, b) $120\,\mu g/m^3$, c) $140\,\mu g/m^3$ and d) $160\,\mu g/m^3$ given the AURN observation is above $100\,\mu g/m^3$, $120\,\mu g/m^3$, $140\,\mu g/m^3$ and $160\,\mu g/m^3$ respectively. Grey indicates that: (a) the marginal 90th quantile is larger than $100\,\mu g/m^3$, (b, c, d) the chosen value is above the upper endpoint of either marginal distributions. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

Figure 4.12 shows the forecasting skill of AQUM through the estimated hit rates. Of the 61 modelled sites, 19 have marginal thresholds above $100\,\mu g/m^3$, depicted in grey in top left panel. For $100\,\mu g/m^3$, the average estimated hit rate at UB sites is $0.599\ (0.376, 0.939)$ and at RB sites is $0.472\ (0.355, 0.670)$, where the range is given in parentheses. Increasing the concentration to $120\,\mu g/m^3$ decreases the hit rate probability to an average of $0.447\ (0.008, 0.876)$ at UB sites and $0.322\ (0.178, 0.573)$ at RB sites. These continue to reduce to 0.212 and 0.126 above $160\,\mu g/m^3$ for UB and RB sites respectively. There are five sites that see an increase in hit rate when

comparing the estimate for above $140\,\mu g/m^3$ and above $160\,\mu g/m^3$. This could be explained by $160\,\mu g/m^3$ being above the maximum value observed so little to no empirical information is guiding the estimation and the misclassification of the marginal shape parameter could be influencing these enlarging hit rate.

As a way to assess the model fit, we compare the model-based estimates of hit rate to the empirical estimates for x equal to the marginal 90th quantile, $x = 120\,\mu g/m^3$ and $x = 140\,\mu g/m^3$. For the first case, the empirical hit rate ranges between 0.365 and 0.663 whereas the model-based estimates range between 0.370 and 0.638. The difference between the empirical and model-based estimates is no more than 4.70% with a median difference of 0.015. Increasing the concentration to $120\,\mu g/m^3$ the median difference becomes -0.036, indicating a slight overestimation of hit rate by the bivariate extreme value distribution models. The overestimation more than doubles to 0.088 when increasing the concentration to $140\,\mu g/m^3$. This indicates that the model captures the dependence structure well, however improvements could be made at higher quantiles.

4.6 Conclusions

In this study, we have demonstrated several ways in which extreme value theory can be used to validate numerical process-based forecasts. We first investigated the tails of the marginal distributions by modelling the threshold exceedances of both AURN and AQUM by a generalised Pareto distribution. We then contrasted the parameter estimates and found that AQUM underestimated the magnitude of the scale parameter on average 3.013 and misidentifies the tail structure at 39% of sites. To assess if AQUM forecasts and AURN observations come from the same distribution we constructed likelihood ratio tests. These concluded that the forecasts and observations at 20% of RB sites and 37% of UB sites come from the same distribution.

To probabilistically estimate conditional probabilities, such as the hit rate, we

require a statistical model for the joint tail behaviour of AURN and AQUM. After estimating the extremal dependence measure χ and performing a likelihood ratio test (Ledford and Tawn, 1996) we concluded the joint tail exhibits asymptotic dependence. The bivariate extreme value distribution is fitted to the joint tail of AURN and AQUM. The survivor function of the fitted bivariate extremes model is used to obtain estimates of conditional probabilities. Such conditional probability estimates indicate that when ozone events become more extreme and pose greater risk to public health, AQUM struggles to accurately capture the magnitude particularly for observed events at least $120\,\mu g/m^3$. Comparing the model-based estimates and empirical estimates of hit rate showed that a bivariate extreme value distribution captures the dependence structure well, however improvements for higher quantiles are needed.

One extension of this work would be to explore the dependence structure between day T and day $T+\Delta T$ to extend the evaluation into an on-line operational validation tool. This could be done by a new post-processing technique; given that an exceedance has been observed on day T then forecasts on day $T + \Delta T$ could be corrected by an amount obtained from a similar bivariate extremes value model. Further extensions include: allowing the threshold to vary across the seasons, fit a non-stationary model to the hourly ozone measurements and forecasts by having the GPD parameters as Fourier series of time, and include meteorological covariates such as temperature and wind direction in the marginal parameters. Another extension could be to use the Heffernan and Tawn (2004) conditional extremes model to capture the temporal dependence in the observations, and contrast the structure to the AQUM forecasts. The temporal dependence structure of the observations is modelled using the conditional extremes model in Chapter 5 and is applied to allow for short-lead-time extreme event forecasts to be made, see Section 5.5.

Chapter 5

Modelling the temporal dependence structure of extreme ozone events

5.1 Introduction

Air pollution episodes are periods of time where the concentration of a pollutant is unusually elevated. Such episodes draw witness to only the largest values and are often termed *extreme events*. Exposure to these extreme events poses risks: development of respiratory issues for example ozone alone attributed to $254,000$ added deaths globally from chronic obstructive pulmonary disease in 2015 (Cohen et al., 2017; Murray et al., 2020), reduction in photosynthesis and damaging crop yields (Dingenen et al., 2009) and structural integrity of buildings (Brimblecombe and Grossi, 2007). These events can add stress to health care systems and has cost the EU approximately €166 billion in social costs during 2018 across 432 EU cities (CE Delft, 2018). Therefore, it is important to quantify and predict the risk of extreme events to mitigate these impacts.

Due to public health risks, warning systems are commonly used to forecast the short-term risk of poor air quality. The UK uses the Daily Air Quality Index (DAQI) which quantifies the health risk associated with concentrations of sulphur dioxide, nitrogen oxide, ozone, $PM_{2.5}$ and PM_{10}, (DEFRA and UK-AIR, 2022). The index is a ten-

point scale divided into four bands: low $(1-3)$, moderate $(4-6)$, high $(7-9)$ and very high (10). The DAQI is defined for each pollutant as the average concentration over 15-minute, 1-hour, 8-hour, 24-hour and 24-hour intervals respectively, (Connolly et al., 2013). The published value is taken to be the maximum of the five pollutant DAQIs. This method of computing and presenting health risk has been approved by the Committee on Medical Effects of Air Pollutants (COMEAP, 2011). For ozone, the risk level increases from low to moderate at $100\,\mu g/m^3$, from moderate to high at $160\,\mu g/m^3$ and from high to very high at $240\,\mu g/m^3$, see Table 2.2 for full DAQI breakdown for each pollutant.

The DAQI is formed from forecast concentrations to provide warnings for upcoming air pollution episodes. Short-lead-time forecasts of pollutant concentrations typically involve complex numerical models that describe chemical and physical processes and their interactions. These forecasts are deterministic and are usually optimised to predict the mean behaviour of the process. However, these models rarely consider extreme events in their calibration process. Therefore, a more bespoke modelling approach is needed to capture extreme events.

Extreme value theory provides such an approach by allowing for the characterisation of the tail behaviour of any distribution. Let X_t be a time-series representing the concentration of a pollutant with marginal distributions given by F_X. Assuming that the series is stationary, the generalised Pareto distribution can be fitted to the independent exceedances of a high threshold u_X, see Sections 3.2.3 and 5.2.1 (Pickands, 1975; Leadbetter et al., 1983). However, the independence assumption may not be appropriate as extremes often occur in *clusters*. Here a cluster is defined by the *runs method* as a group of threshold exceedances in which any two consecutive cluster members are separated by at most $l-1$ non-exceedances where l is referred to as the run length (Smith and Weissman, 1994). Modelling clusters is more complicated due to the need to capture both the marginal distribution and the dependence structure of the extremes. Joe (1997) suggests modelling the marginal distribution of all exceedances and the dependence structure separately.

Before describing a modelling approach for the within-cluster dependence structure of clusters of extremes we first need to define the two classes of extremal dependence: *asymptotic dependence* and *asymptotic independence*. The extremal dependence class, of the random variables (X_0, X_t) at lag t, is determined by the probability of one variable being extreme conditional on the other also being extreme,

$$\lim_{x\to x^*} \chi_t(x) = \lim_{x\to x^*} \Pr(X_t > x \mid X_0 > x) \to \begin{cases} 0, & \text{if asymptotically independent} \\ p_t > 0, & \text{if asymptotically dependent} \end{cases} \tag{5.1.1}$$

where x^* is the upper endpoint of F_X and $\lim_{x\to x^*} \chi_t(x) = \chi_t = p_t \in [0, 1]$ determines the strength of asymptotic dependence (Coles et al., 1999). We term $\chi_t(x)$ as the sub-asymptotic extremal dependence measure. To determine the strength of asymptotic independence, the measure $\bar{\chi}_t$ was introduced by Coles et al. (1999) and is given by

$$\bar{\chi}_t = \lim_{x\to x^*} \frac{2\log \Pr(X_0 > x)}{\log \Pr(X_0 > x, X_t > x)}, \tag{5.1.2}$$

which takes values $[-1, 1]$. The pair (X_0, X_t) are said to be asymptotically independent with positive association if $0 < \bar{\chi}_t < 1$, asymptotically independent with negative association if $-1 \leq \bar{\chi}_t < 0$ and independent if $\bar{\chi}_t = 0$.

Both measures together, $(\chi_t, \bar{\chi}_t)$, provide a summary of the extremal dependence between X_0 and X_t. For example, if $\bar{\chi}_t = 1$, the variables are asymptotically dependent, and the value of χ_t summarises the strength of extremal dependence. Whereas, if $\bar{\chi}_t < 1$, then $\chi = 0$ and the variables are asymptotically independent, and the value of $\bar{\chi}_t$ is the strength of extremal dependence.

Identifying the type of extremal dependence structure is critical in model selection as not all models can describe asymptotic (in)dependence and misclassification could lead to underestimation of the dependence structure and measures of interest. Such measures include D_v, the number of exceedances of v in a cluster, $\pi(i, v)$ the distribution of D_v, $\theta(v)$, the sub-asymptotic extremal index introduced by Ledford and Tawn (2003) which can be computed as the reciprocal of the mean cluster size

$\pi(\cdot, v)$, and the measure given in Winter and Tawn (2017),

$$\psi_T(i,v) = 1 - \exp\left\{-\theta(v)[1-\tilde{F}(u_X)]n_T\Pi(i,v)\left[1+\xi\left(\frac{v_X-u_X}{\sigma_{u_X}}\right)\right]_+^{-1/\xi}\right\}, \quad (5.1.3)$$

where n_T is the number of observations in time period T, $\tilde{F}$ is the empirical distribution and $\Pi(i,v) = \sum_{j=i}^{\infty}\pi(j,v)$. This measure can be interpreted as the probability of observing at least one cluster containing at least i exceedances in a given time period. These measures are termed *cluster functionals*.

A wide range of models are used to capture extremal temporal dependence. Smith et al. (1997) use a first-order Markov chain approach for modelling threshold exceedances; however, they assume asymptotic dependence at lag one which consequently implies asymptotic dependence at all lags. Extensions of this approach to kth-order Markov chains can be found in Fawcett and Walshaw (2006) and Ribatet et al. (2009), however, their models are still limited by an assumption of asymptotic dependence at all lags. Winter and Tawn (2017) addresses this weakness by introducing a model based on the conditional model for multivariate extremes (Heffernan and Tawn, 2004) which can account for both asymptotic dependence and asymptotic independence. As such, we seek to take advantage of the kth-order Markov extremes model to provide good probabilistic estimates of extreme events and provide more information on the risk of an extreme event of ozone than the current DAQI system. The advantage of the Winter and Tawn method is it has the capability to look at the full cluster behaviour. Further we aim to demonstrate the forecasting capabilities of the model by simulating extreme events from an initial exceedance using the forward simulation procedure given by Rootzén (1988).

The chapter is structured as follows. Section 5.2 sets out the asymptotically justified modelling approach for kth-order chains based on the Heffernan and Tawn (2004) conditional model. The parameter inference is explored in Section 5.3.1. The simulation algorithm is given in Section 5.3.2 with a new diagnostic to choose the model order in Section 5.3.3. Section 5.4 details the modelling results for our ozone

data set from across Great Britain. Section 5.5 demonstrates the forecasting skill of the kth-order Markov extremes model by forecasting all extreme events that occurred during 2020. Conclusions and further work are presented in Section 5.6.

5.2 Modelling temporal dependence

The approach taken to modelling clusters of extremes is based on the concept of a Markov process. By the Markov property, the probability distribution of future states of the process conditioned on both the past and present states depends only on the present state. The Markov property holds in a model if the values in any state depend only on the immediately preceding or a small number of immediately preceding states. Under the assumption that a time series $\{X_t\}$ follows a kth-order Markov process, the joint density $f_{1:n}$ of $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ can be written as

$$f_{1:n}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = f_{1:k}(\mathbf{x}_1, \ldots, \mathbf{x}_k) \prod_{t=1}^{n-k} f_{k+1|1:k}(\mathbf{x}_{t+k} \mid \mathbf{x}_{t:t+k-1}), \tag{5.2.1}$$

where $f_{k+1|1:k}(\cdot \mid \cdot)$ is the conditional density function of $Y_{k+1} \mid \mathbf{Y}_{1:k}$. Consequently we can model the extremes of the whole joint distribution by analysing the extremes of $(X_t, \ldots, X_{t+k})$ for $t = 1, \ldots, n-k$ and the conditional distribution of $X_{t+k} \mid (X_t, \ldots, X_{t+k-1})$. Adopting a copula framework (see Section 3.3.2 for details) allows us to model the marginal distributions and the joint distribution separately (Joe, 1997). We provide the modelling framework for the marginal exceedances of threshold u_X in Section 5.2.1. The extremal dependence modelling framework is presented in Section 5.2.3 with the asymptotic justifications outlined prior in Section 5.2.2.

5.2.1 Marginal modelling

As $\{X_t\}$ is a stationary series, the marginal distributions F_X are identical. The most common approach to modelling the marginal distributions of extreme values

is to fit a generalized Pareto distribution (GPD), equation (3.2.5), to the peaks over threshold data, i.e. exceedances of a high threshold u_X (Pickands, 1975; Leadbetter et al., 1983). Following Coles and Tawn (1991), we model the marginal distribution of X_t as

$$F_X(x) = \begin{cases} 1 - G_{u_X}(x)[1 - \tilde{F}(u_X)], & \text{for } x > u_X, \\ \tilde{F}(x), & \text{for } x \leq u_X, \end{cases}$$

where $G_{u_X}(x)$ is the generalised Pareto distribution, $\tilde{F}(x)$ is the empirical marginal cumulative distribution function of $\{X_t\}_{t=1}^n$ and $1 - \tilde{F}(u_X)$ is the threshold exceedance rate.

Having estimated the marginal structure through standard likelihood approaches an appropriate choice of transformation onto common margins is required before characterising the extremal dependence. Following the work of Keef et al. (2013), we use the probability integral transform to transform X_t, $t = 1, \ldots, n$ onto Laplace margins, denoted by subscript L,

$$X_{t,L} = \begin{cases} \log\{2F(X_t)\}, & \text{if } F(X_t) < 1/2, \\ -\log\{2[1 - F(X_t)]\}, & \text{if } F(X_t) \geq 1/2 \end{cases}$$

which allows us to use the Normal distribution assumption in our inference, see Section 5.3 for details. The symmetry of the Laplace distribution ensures the limiting dependence model is unchanged with respect to t and the threshold and captures the exponential upper tail of the Gumbel distribution required for modelling positive dependence while allowing for negatively associated variables to be incorporated into the model parsimoniously (Heffernan and Tawn, 2004; Heffernan and Resnick, 2007). For variables on Laplace margins, we denote the associated threshold as u_L.

5.2.2 Asymptotics for conditional extremes

Heffernan and Tawn (2004) proposed an asymptotically justified conditional approach for modelling the extremes of a vector $\mathbf{X}_{1:m,L}$ where $m \in \mathbb{N}$ and each

variable has Laplace margins. To investigate the conditional distribution $\mathbb{P}(\mathbf{X}_{1:m,L} \leq \mathbf{x} \,|\, X_{0,L} > u_L)$ for large u_L we require the distribution to be non-degenerate as $u_L \to \infty$, therefore an appropriate normalisation is required. Heffernan and Resnick (2007) proposed that $\mathbf{X}_{1:m,L}$ is linearly normalized as a function of either $X_{0,L}$ or u_L, however for statistical simplicity the Heffernan and Tawn (2004) approach to normalize by $X_{0,L}$ is used.

Heffernan and Tawn (2004) assume there exists functions $a : \mathbb{R} \to \mathbb{R}^m$ and $b : \mathbb{R} \to \mathbb{R}^m_+$, such that

$$\mathbb{P}\left(\frac{\mathbf{X}_{1:m,L} - a(X_{0,L})}{b(X_{0,L})} \leq \mathbf{z}_{1:m}, X_{0,L} - u_L > x \;\middle|\; X_{0,L} > u_L\right) \to G_{1:m}(\mathbf{z}_{1:m}) \exp(-x), \tag{5.2.2}$$

as $u_L \to \infty$, where $G_{1:m}$ is a joint distribution function that is non-degenerate in each margin, i.e. for $j = 1, \ldots, m$ the jth margin G_j of $G_{1:m}$ is non-degenerate and $\mathbf{z} \in \mathbb{R}^m$. Heffernan and Resnick (2007) show that a and b must be regularly varying functions, i.e. a decaying function of some order. A simple form for a and b that holds for a very broad range of copulas was found to be

$$a(X_{0,L}) = \boldsymbol{\alpha}_{1:m} X_{0,L} \;\text{ and }\; b(X_{0,L}) = X_{0,L}^{\boldsymbol{\beta}_{1:m}} \tag{5.2.3}$$

where $\boldsymbol{\alpha}_{1:m} = (\alpha_1, \ldots, \alpha_m) \in [-1, 1]^m$ and $\boldsymbol{\beta}_{1:m} = (\beta_1, \ldots, \beta_m) \in (-\infty, 1)^m$. This canonical parametric subfamily of a and b provides a flexible family for statistical modelling.

Keef et al. (2013) provides details on how to interpret the type of extremal dependence from the dependence parameters $\boldsymbol{\alpha}_{1:m}$ and $\boldsymbol{\beta}_{1:m}$. For $1 \leq j \leq m$, the pair $(X_{0,L}, X_{j,L})$ are independent when $\alpha_j = \beta_j = 0$, asymptotically positive dependent when $\alpha_j = 1$ and $\beta_j = 0$, asymptotically negative dependent when $\alpha_j = -1$ and $\beta_j = 0$, and asymptotically independent when $\alpha_j < 1$.

5.2.3 Temporal dependence modelling

The within cluster dependence model of Winter and Tawn (2017) has two components: (a) a conditional model for the joint behaviour of terms $2, \ldots, k$ given the initial term ($k = 1$); (b) for $t = k + 1, \ldots$, a conditional model for each term given the preceding k terms.

Component (a) uses the conditional extremes approach of Heffernan and Tawn (2004) and Heffernan and Resnick (2007). Similarly the limiting form of the conditional distribution (5.2.2) is assumed to hold exactly for all values of $X_{t,L} > u_L$ with $m = k$ such that the normalizing functions a and b can be given by the forms in equation (5.2.3). This results in

$$\mathbf{X}_{t+1:t+k,L} \mid (X_{t,L} > u_L) = \boldsymbol{\alpha}_{1:k} X_{t,L} + (X_{t,L})^{\boldsymbol{\beta}_{1:k}} \mathbf{Z}_{1:k}, \tag{5.2.4}$$

for $\boldsymbol{\alpha}_{1:k} \in [-1, 1]^k$, $\boldsymbol{\beta}_{1:k} \in [0, 1)^k$ and where $\mathbf{Z}_{1:k}$ is a random variable, which is independent of t and $X_{t,L}$ since the process is assumed to be stationary, with joint density $g_{1:k}$ and distribution function $G_{1:k}$. As the recurrence relationship in equation (5.2.4) is assumed to hold only for $X_{t,L} > u_L$, the processes generated under this model will return from an extreme state to the body of the distribution, i.e. will eventually fall back below the marginal threshold, Smith (1992) and Papastathopoulos et al. (2017).

For component (b), model (5.2.4) is assumed to hold exactly for $X_{t,L} = x_{t,L} > u_L$ such that,

$$X_{t+k,L} \mid (\mathbf{X}_{t:t+k-1,L} = \mathbf{x}_{t:t+k-1,L}) = \alpha_k X_{t,L} + (X_{t,L})^{\beta_k} Z_{k|1:k-1}, \tag{5.2.5}$$

where the conditional distribution of $Z_{k|1:k-1}$ is obtained from $g_{1:k}$ using Bayes Theorem,

$$G_{k|1:k-1}(z \mid \mathbf{z}_{1:k-1}) = \int_{-\infty}^{z} g_{k|1:k-1}(r \mid \mathbf{z}_{1:k-1}) dr. \tag{5.2.6}$$

Consequently estimates for future events X_{t+k+j} for $j = 1, \ldots, m$ can be obtained without explicitly evaluating $\alpha_{k+1:m}$, $\beta_{k+1:m}$ and $G_{k+1:m}$.

5.3 Inference

5.3.1 Parameter inference

Following Winter and Tawn (2017) we use a three-step inference procedure where we estimate the marginal parameters (σ_{u_X}, ξ), the dependence parameters $(\boldsymbol{\alpha}_{1:k}, \boldsymbol{\beta}_{1:k})$ and then the distribution $(G_{k|1:k-1})$. Winter and Tawn (2017) provide justification to how one can separate out the inference for the margins from the inference for the dependence. The likelihood of the model can be written as

$$f_{1:n}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \approx \prod_{t:x_{t,L}>u_L} \frac{c^{X_L}_{1:k+1}(\mathbf{x}_{t:t+k,L})}{c^{X_L}_{1:k}(\mathbf{x}_{t:t+k-1,L})} \prod_{t:x_t>u_X} f(x_t), \tag{5.3.1}$$

where $c^{X_L}_{1:k+1}$ and $c^{X_L}_{1:k}$ are the copula densities for $\mathbf{X}_{t:t+k,L}$ and $\mathbf{X}_{t:t+k-1,L}$ respectively and $x_{t,L} = F^{-1}_{X_L}(F_X(x_t))$ for all t. For the second product of the likelihood we use standard maximum likelihood estimation for (σ_{u_X}, ξ) using all the threshold exceedances of u_X. The first product term requires the conditional distribution of $X_{t+k} \mid \mathbf{X}_{t:t+k-1}$ given in equation (5.2.5) and since $G_{1:k}$ and its marginals do not take any finite parametric form, we make a temporary assumption that $Z_{1:k}$ are independent Normal variables with $Z_j \sim N(\mu_j, \gamma_j^2)$ for $j = 1, \ldots, k$ (Keef et al., 2013). Under this assumption,

$$X_{t+j,L} \,|\, \{X_{t,L} = x_L\} \sim N\left(\alpha_j x_L + \mu_j (x_L)^{\beta_j}, \gamma_j^2 (x_L)^{2\beta_j}\right) \quad \text{for} \quad x_L > u_L,$$

where $j = 1, \ldots, k$ for all t resulting in the likelihood

$$L(\boldsymbol{\alpha}_{1:k}, \boldsymbol{\beta}_{1:k}, \boldsymbol{\mu}_{1:k}, \boldsymbol{\gamma}_{1:k}) = \prod_{j=1}^{k} \prod_{i=1}^{n_u} \frac{1}{\sqrt{2\pi\gamma_j^2 (x_{i,L})^{2\beta_j}}} \exp\left\{ -\frac{(x_{i,L} - \mu_{ji})^2}{2\gamma_j^2 (x_{i,L})^{2\beta_j}} \right\}, \tag{5.3.2}$$

where n_u is the number of exceedances by $\{X_t\}$ of u_X, and $\mu_{ji} = \alpha_j x_{i,L} + \mu_j (x_{i,L})^{\beta_j}$ for $j = 1, \ldots, k$ and $i = 1, \ldots, n_u$. Maximization of likelihood L gives estimates $(\hat{\boldsymbol{\alpha}}_{1:k}, \hat{\boldsymbol{\beta}}_{1:k}, \hat{\boldsymbol{\mu}}_{1:k}, \hat{\boldsymbol{\gamma}}_{1:k})$.

The conditional distribution function $G_{k|1:k-1}$ is estimated non-parametrically by first obtaining a non-parametric estimate of the joint density function $g_{1:k}$. We use a similar kernel density approach to Papastathopoulos and Tawn (2013). We first obtain fitted residuals from the IID model $\hat{\mathbf{z}}_{1:k}^{(i)}, i = 1, \ldots, n_u$ from $\mathbf{Z}_{1:k}$. Specifically, let $t_1, \ldots, t_{n_u}$ be the indices of $t = 1, \ldots, n$ where $x_t > u_X$. We invert equation (5.2.4) such that, for $i = 1, \ldots, n_u$

$$\hat{\mathbf{z}}_{1:k}^{(i)} = \frac{\mathbf{x}_{t_i+1:t_i+k,L} - \hat{\boldsymbol{\alpha}}_{1:k} x_{t_i,L} - \hat{\boldsymbol{\mu}}_{1:k} (x_{t_i,L})^{\hat{\boldsymbol{\beta}}_{1:k}}}{\hat{\boldsymbol{\gamma}}_{1:k} (x_{t_i,L})^{\hat{\boldsymbol{\beta}}_{1:k}}}. \tag{5.3.3}$$

For each $j = 1, \ldots, k$, the sample $\hat{\mathbf{z}}_j = (\hat{\mathbf{z}}_j^{(i)}, i = 1, \ldots, n_u)$ has zero mean and unit variance.

We estimate the joint density $g_{1:k}$ by fitting a multivariate kernel density estimation methods to the fitted residuals, $\hat{\mathbf{z}}_{1:k}$,

$$\tilde{g}_{1:k}(\mathbf{z}) = \tilde{g}(z_1, \ldots, z_k) = \frac{1}{n_u} \sum_{i=1}^{n_u} K_{\mathbf{H}} \left(\mathbf{z} - \hat{\mathbf{z}}_{1:k}^{(i)} \right) \tag{5.3.4}$$

where $K_{\mathbf{H}}$ is the multivariate Normal kernel function centred on $\mathbf{0}$ and $\mathbf{H}$ is a symmetric positive definite matrix of bandwidths which control the roughness of the smoothing. Winter and Tawn (2017) use the standard independent multivariate normal kernel. Instead we use the approach of Liu and West (2001) who shrink the variance of a kernel to reduce the risk of inflating the variance of the estimator which arises when too large a bandwidth is used. This reduces the risk of misleading estimates as the samples will be less varied. Here we use a shrinkage factor $c_j = \sqrt{1 - h_j^2}$ where h_j is the bandwidth associated with lag j. As such equation (5.3.4) becomes

$$\tilde{g}_{1:k}(\mathbf{z}) = \tilde{g}(z_1, \ldots, z_k) = \frac{1}{n_u} \sum_{i=1}^{n_u} \prod_{j=1}^{k} \frac{1}{h_j} \phi \left(\frac{z_j - \zeta_j^{(i)}}{h_j} \right), \tag{5.3.5}$$

where for $j = 1, \ldots, k$, $\zeta_j^{(i)} = c_j \hat{z}_j^{(i)} + (1 - c_j)\mu_{z,j}$ and $h_j = \left(\frac{4\sigma_{z,j}^5}{3n_{u_X}}\right)^{1/5}$, $\mu_{z,j}$ and $\sigma_{z,j}$ are the mean and standard deviation of $\hat{z}_j$ respectively (Scott, 1992) and $\phi(\cdot)$ is the standard Normal density function. It follows that our non-parametric estimate of the conditional distribution function $G_{k|1:k-1}$ is

$$\hat{G}_{k|1:k-1}(z \mid \mathbf{z}_{1:k-1}) = \sum_{i=1}^{n_u} \omega_i \Phi\left(\frac{z - \zeta_k^{(i)}}{h_k}\right),$$

where the weights

$$\omega_i = \prod_{j=1}^{k-1} \phi\left(\frac{z_j - \zeta_j^{(i)}}{h_j}\right) \Big/ \sum_{r=1}^{n_u} \prod_{j=1}^{k-1} \phi\left(\frac{z_j - \zeta_j^{(r)}}{h_j}\right) \qquad i = 1, \ldots, n_u, \tag{5.3.6}$$

satisfy $0 \leq \omega_i \leq 1$, $\sum_{i=1}^{n_u} w_i = 1$.

5.3.2 Extremal chain simulation algorithm

The model given in Section 5.2 describes the stochastic evolution of the chain over $k+m$ days following the initial exceedance. With the estimates of the model parameters, we can use a simulation process to derive estimates of the cluster functionals outlined in Section 5.1. The following simulation process follows the method outlined in Winter and Tawn (2017). Starting with a simulation of the initial value in the cluster, the remaining values of the cluster can be simulated by iterative application of the model (5.2.5). This creates realisations of a tail chain with kth-order structure. Unless stated otherwise, we initialise the simulated cluster by generating a starting exceedance $X_0^{(v)} = v + E_1$ where E_1 is simulated from the standard exponential distribution. Conditionally on $X_0^{(v)}$ the next $k-1$ observations are simulated jointly by

$$\mathbf{X}_{1:k-1}^{(v)} = \hat{\boldsymbol{\alpha}}_{1:k-1} X_0^{(v)} + \hat{\boldsymbol{\mu}}_{1:k-1} \left(X_0^{(v)}\right)^{\hat{\boldsymbol{\beta}}_{1:k-1}} + \hat{\boldsymbol{\gamma}}_{1:k-1} \left(X_0^{(v)}\right)^{\hat{\boldsymbol{\beta}}_{1:k-1}} \mathbf{Z}_{1:k-1}, \tag{5.3.7}$$

where $\mathbf{Z}_{1:k-1}$ is sampled independently from $\hat{g}_{1:k-1}$ and the marginal of $\hat{g}_{1:k}$ is given by equation (5.3.5). At all subsequent time-steps $m \geq k$, the transition kernel of the extremal tail chain is used,

$$\mathbf{X}_m^{(v)} = \hat{\boldsymbol{\alpha}}_k X_{m-k}^{(v)} + \hat{\boldsymbol{\mu}}_k \left(X_{m-k}^{(v)} \right)^{\hat{\boldsymbol{\beta}}_k} + \hat{\boldsymbol{\gamma}}_k \left(X_{m-k}^{(v)} \right)^{\hat{\boldsymbol{\beta}}_k} \mathbf{Z}_{m|m-k+1:m-1}, \tag{5.3.8}$$

where $\mathbf{Z}_{m|m-k+1:m-1}$ values are sampled independently from $\hat{G}_{k|1:k-1}$. Following this approach provides a tail chain $X_0^{(v)}, \ldots, X_m^{(v)}$ on Laplace margins with kth-order temporal dependence structure. The tail chain is immediately terminated when $X_m^{(v)} < 0$, i.e. the simulated process falls below the marginal median, as the transition (5.3.8) cannot be evaluated since $\beta_k < 1$.

By repeating the steps above n_{sim} times we obtain a sample of tail chains of size n_{sim} where each has the desired kth-order temporal dependence structure. From the sample, we average over the chain lengths to obtain a model-based estimate for $\theta(v)$. We obtain an estimate for $\chi_j(v)$ by evaluating the distribution of D_v. Estimates for $\psi_T(i, v)$ naturally follows by evaluating equation (5.1.3).

5.3.3 Extremal Markov process order selection

A standard approach to estimate the order of a Markov chain is to use the partial auto-correlation function (PACF) to identify the largest lag at which the PACF is significantly different from zero (Chatfield, 2003). This may not necessarily be appropriate for an extremal Markov process as the PACF inference is dominated by the data in the body of the distribution and the extremes may exhibit different dependence structure to data from the body of the distribution.

Winter and Tawn (2017) developed alternative diagnostics for the Markov process order which assess the stability of cluster functionals over a range of $v > u_L$. The resulting visual diagnostics are similar to the threshold stability plots for extremal parameters by Coles (2001). If no particular cluster functional is of interest then

they suggest that the best estimate of the extremal Markov process order is $\hat{\tau}$ where

$$\hat{\tau} = \min \left\{ \tau > 0 : \sum_{v \in V} \{ \mid \hat{\theta}^{(\tau)}(v) - \tilde{\theta}(v) \mid + \frac{1}{n_J} \sum_{j \in J} \mid \hat{\chi}_j^{(\tau)}(v) - \tilde{\chi}_j(v) \mid < \epsilon \} \right\}, \quad (5.3.9)$$

for a choice of $\epsilon > 0$ where $\tilde{\theta}(v)$ and $\hat{\theta}^{(\tau)}(v)$ are the empirical and model-based sub-asymptotic extremal index estimates, $\tilde{\chi}_j(v)$ and $\hat{\chi}_j^{(\tau)}(v)$ are the empirical and model-based sub-asymptotic extremal dependence measure estimates, J is a set of lags, V is a set of thresholds and n_J is the length of J.

We propose a new diagnostic that accounts for multiple metrics. This new diagnostic adapts recent work of Knutti et al. (2017) which is used to compute ensemble model estimates using a weighted sum of multiple model outputs. The diagnostic normalises the difference between the model-based estimate and the empirical estimate of a metric over a range of levels $v > u_L$ from the τth-order model by the difference between the τth-order metric estimate and all other order metric estimates. More precisely, the best estimate of the extremal Markov process order is $\hat{\tau} = \max(\omega_\tau)$, where ω_τ is the weight of the model of τth-order given by

$$\omega_\tau = \frac{\sum_{i=1}^{M} \exp\left(-D_{i\tau}^2\right)}{M + \sum_{i=1}^{M} \sum_{j \neq \tau} \exp\left(-S_{ij\tau}^2\right)}, \quad (5.3.10)$$

where $D_{i\tau}$ is a distance measure between the empirical estimate and τth-order model-based estimate of the ith metric, $S_{ij\tau}$ is the difference between the τth- and jth-order model-based estimate of the ith metric and M is the number of metrics.

5.3.4 Bootstrapping

Throughout, estimates of the uncertainty of extremal quantities, including the dependence parameters and cluster functionals, will be derived by bootstrapping. As we are modelling the temporal dependence of a single variable X_t we require a resampling approach that keeps the temporal dependence features of the original data. Thus, we construct bootstrap samples by splitting the original data into

periods of exceedances (clusters) and non-exceedances. Then we iteratively pick randomly from the sets of non-exceedances and exceedances until we have a new sample with the same length as the original data. For each bootstrap sample the marginal and dependence characteristics can then be assessed using the approaches above.

To estimate the uncertainty in the estimates of the temporal dependence and the cluster functionals we take the 2.5% and 97.5% quantiles of the bootstrap estimates to form a 95% *block-bootstrapped confidence interval*. When the bootstrapping approach becomes too computationally intensive, we construct uncertainty bounds by taking the standard error of a smaller number of bootstrap replicates (here used 20) and construct a symmetric confidence interval, termed as *modified block-bootstrapped confidence interval*.

5.4 Data analysis

5.4.1 Observational data

The hourly ozone observations from AURN, described in Section 2.4.1, are aggregated to the daily maximum of the 8-hour running mean scale (DM8). We restrict ourselves to the site-specific ozone seasons, as defined in Section 2.4.3, to focus on the time of year where the levels of ozone are usually highest. The sampled monitoring sites and their corresponding marginal 90th quantile are shown in Figure 5.1. Throughout the analysis of Chapter 5 we demonstrate the modelling approach on two sites, Aston Hill (AH, rural background) and Wirral Tranmere (TRAN, urban background). These two sites approximately represent all sites of respective type by capturing the average distribution of ozone, the extremal index and the temporal dependence.

As we are interested in modelling the temporal dependence structure, we first identify clusters using the empirical runs estimator, defined in Section 5.1, with a run length of one day. A run length of one day was chosen to optimise our data and

to be in-line with other meteorological and air quality definitions, such as heatwaves. As such, this results in an average of 56 clusters at each site with a median cluster length of two days. Further the longest duration event ranges from five to 21 days across sites (median of nine days). Across the sites, the proportion of clusters that lasted one day ranges between 0.37 and 0.66 with the average being 0.50, indicating that a run length of one day is appropriate.

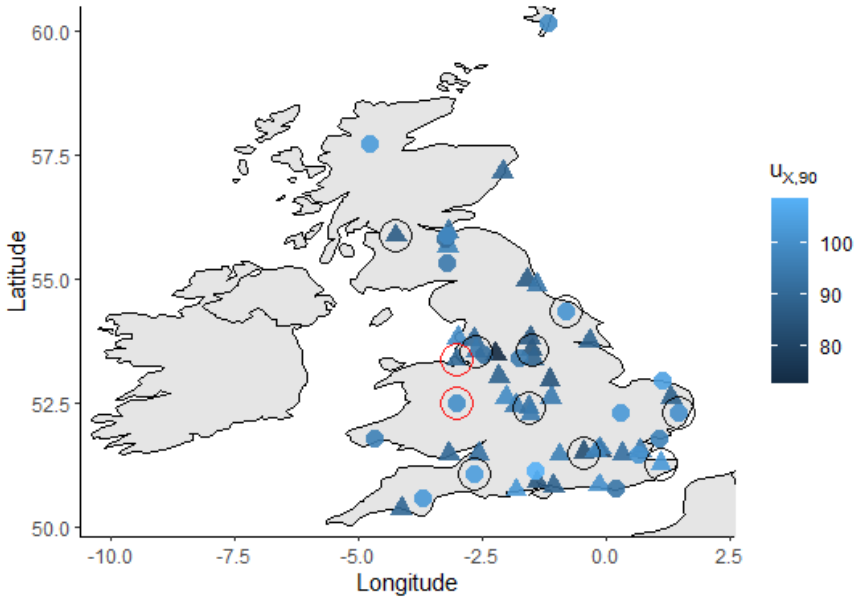


Figure 5.1: Map showing the modelling threshold (lighter colour denotes higher concentration) of the sampled urban background sites (▲) and rural background sites (●). Sites circled are (in black) the randomly selected case study sites we use to verify the model and (in red) the sites we use to demonstrate the modelling approach.

5.4.2 Marginal modelling

To fit a stationary generalised Pareto distribution, we first identify an appropriate marginal modelling threshold at each site by examining the stability in the shape estimates with threshold, Coles (2001). Such stability plots indicate that it is sensible to use the site-specific 90th quantile as our marginal modelling threshold at all sites. This results in a threshold ranging from 72.75 $\mu g/m^3$ to 108.56 $\mu g/m^3$ with nine rural background and three urban background sites having thresholds above 100 $\mu g/m^3$.

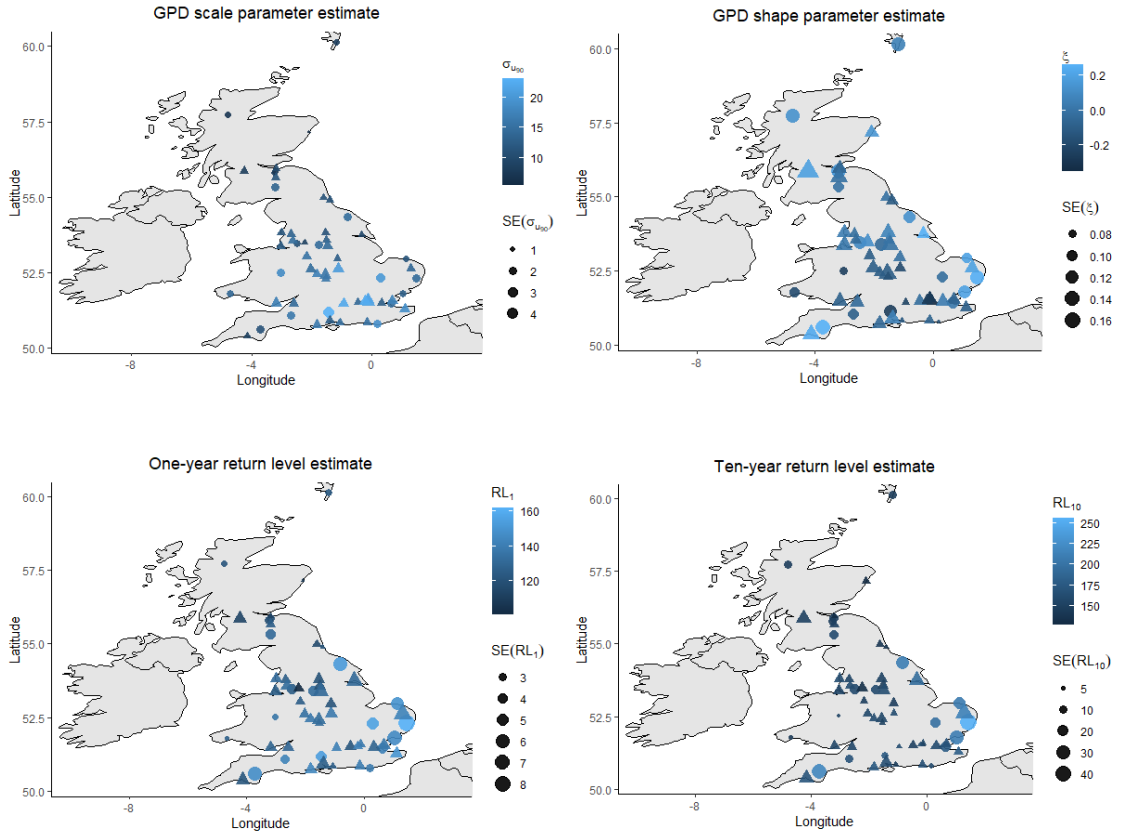


Figure 5.2: Maps showing (top) the MLEs of the marginal GPD scale parameter (left) and shape parameter (right) and (bottom) the one-year return levels (left) and ten-year return levels (right). The size of the points indicate the standard error of the estimate. (○) points denote rural background sites and (△) points denote urban background sites.

The GPD parameter estimates for each site are displayed in Figure 5.2 (top). The largest estimates of the scale parameter occur in central England and around London, with estimates decreasing with latitude. The shape parameter estimates have a more complex pattern, with heavier tails around the coast and lighter tails in the Midlands. Only 42% of rural background sites have a negative shape parameter compared to 71% of urban background sites. One- and ten-year return level estimates are also displayed in Figure 5.2 (bottom). Larger return levels occur in the south and at the east-coast sites. The one-year return levels indicate that four rural background and one urban background site are expected to exceed $160\,\mu g/m^3$ (DAQI of 7, meaning a high health risk) on average once every year. Whereas

the ten-year return level estimates suggest all but one site (Aberdeen, ABD) are expected to experience an exceedance of $160\,\mu g/m^3$, and eight sites are expected to see an exceedance of $240\,\mu g/m^3$ (DAQI of 10, meaning a very high health risk) once every ten years.

5.4.3 Conditional modelling

Figure 5.3 shows the estimates of $\chi_j(v)$ for the case where v is set to the 90th quantile ($u_{90,L}$). The estimates indicate strong asymptotic dependence at lags one and two across all sampled monitoring sites. As the lag increases the strength of asymptotic dependence decreases. Upon examining the corresponding confidence intervals (not shown here), the type of dependence for higher lags switches to asymptotic independence since the confidence intervals span zero. For some sites in the Midlands, the confidence intervals suggest that asymptotic independence occurs as early as lag two or three.

The dependence parameter estimates for the Heffernan and Tawn (2004) conditional model are shown in Figures 5.4 and 5.5 for lags $j = 1, \ldots, 8$. Lag one and two estimates indicate asymptotic dependence for most sites with some Midlands sites indicating asymptotic independence as the α estimates are not one, although the confidence intervals do include one (not shown here). Sites on the south east-coast and Midlands on average have smaller negative α estimates with some sites indicating negative asymptotic independence. The β estimates in Wales and south-west England remain marginally positive over the eight lags whereas the rest of the UK has negative estimates. This implies that Wales and south-west England on average see longer events than the rest of the UK since a negative β value implies that all of the conditional quantiles of X_{t+k} converge to the same value as X_t. This is consistent with the sub-asymptotic extremal index estimates and that fewer longer events are observed in these regions.

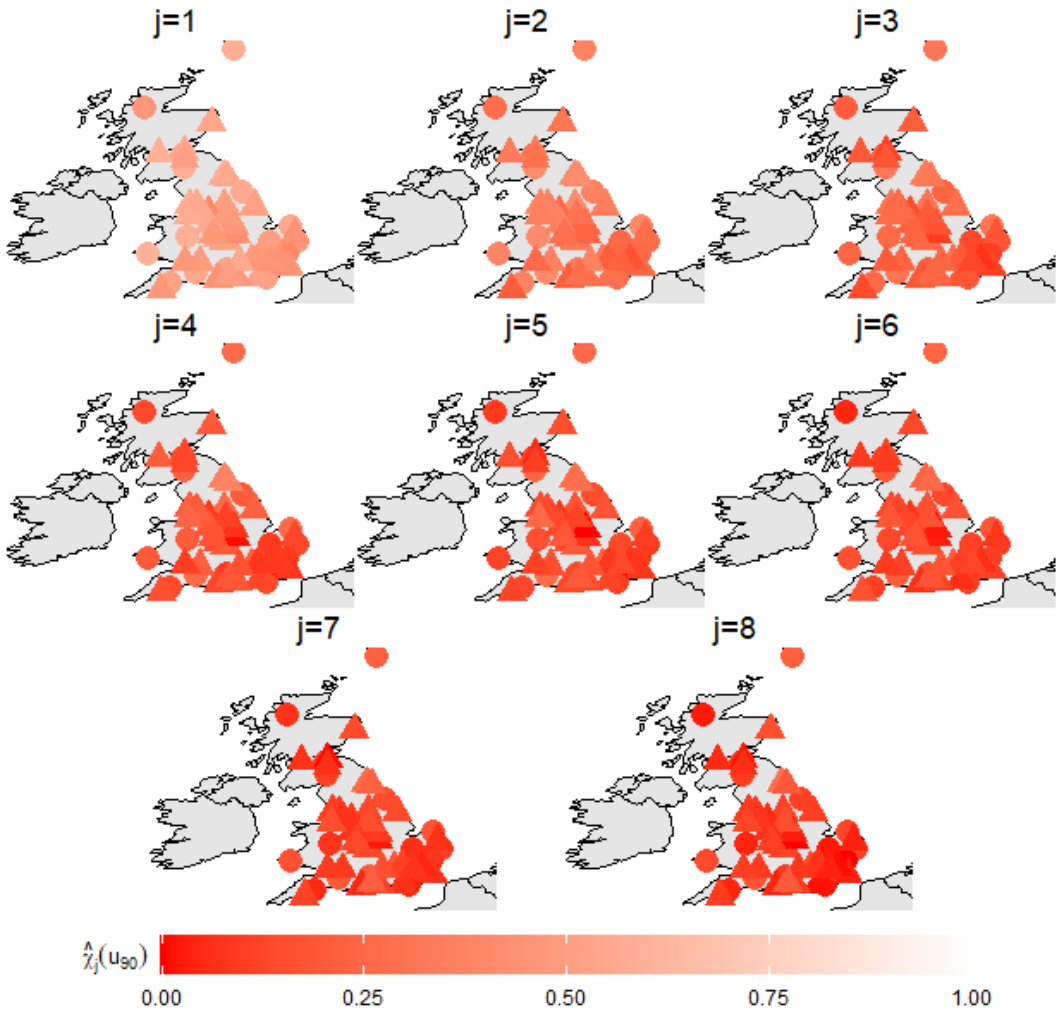


Figure 5.3: The extremal dependence measure $\chi_j(v)$ evaluated at the 90th quantile for $j = 1, \ldots, 8$. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

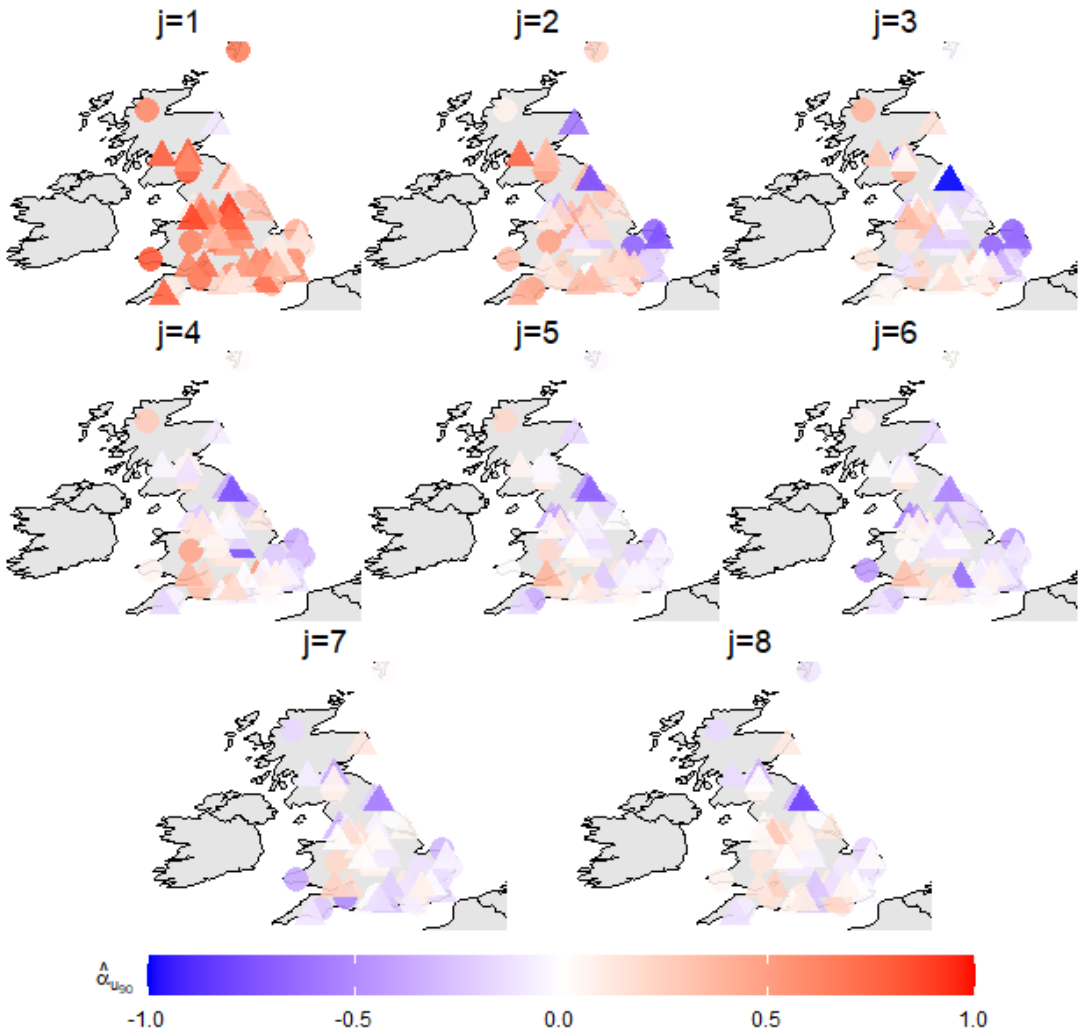


Figure 5.4: Estimates for α_j for $j = 1, \ldots, 8$ at all sampled AURN sites. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

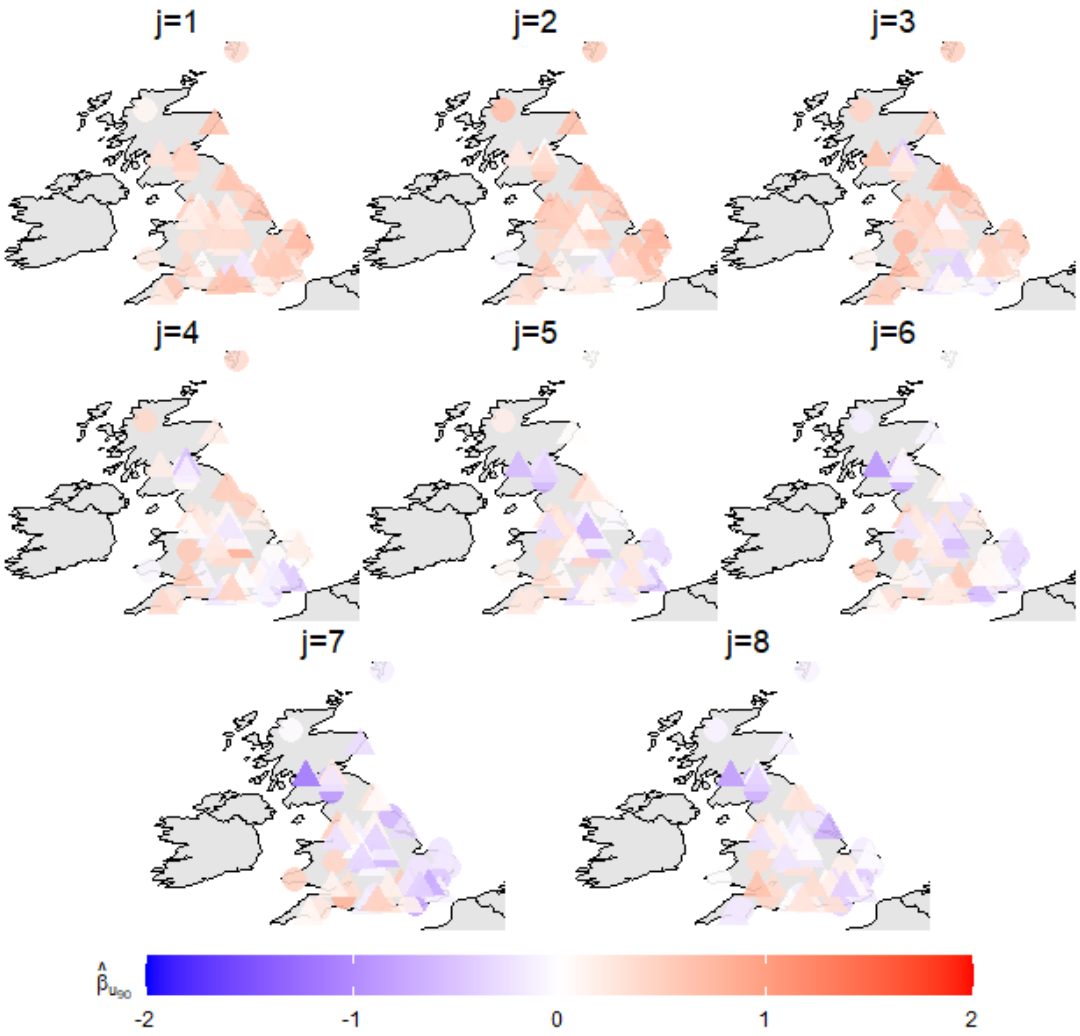


Figure 5.5: Estimates for β_j for $j = 1, \ldots, 8$ at all sampled AURN sites. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

5.4.4 Order selection

We use the algorithm in Section 5.3.2 to simulate 5000 extremal chains from the fitted kth-order Markov extremes models with $k = 1, \ldots, 8$ and look at how the model-based estimates of the sub-asymptotic extremal dependence measure $\chi_j(v)$ and $\psi(i, v) = \psi_1(i, v)$ compare to the empirical-based estimates, Figures 5.6 and 5.7. In these figures we set v to be the marginal 90th quantile ($u_{90,L}$) and 95th quantile ($u_{95,L}$). At Aston Hill, all kth-order Markov extremes models overestimate the short

(3) and medium $(4-5)$ range extremal dependence $\chi_j(u_{90})$ with $k=1$ performing the poorest and $k=5$ performing the best. For long-range dependence structure (≥ 6) there is little difference between $k=5$, 6, 7, 8 and the empirical estimate. For $v=u_{95,L}$, all models slightly underestimate the longer dependence structure which could be due to the limited number of such observed events causing the empirical estimate to flat-line. At Wirral Tranmere, the simplest model $(k=1)$ estimates are closest to the empirical and the more complex models $(k=6,\ 7,\ 8)$ do just as well, whereas the models of order $k=2,\ 3,\ 4$ perform the worst at $v=u_{90,L}$. The best order at $v=u_{95,L}$ is $k=2$ with the worst performing being $k=7$ or 8. Note all model-based estimates fall within the empirical block bootstrapped 95% confidence intervals indicating good model fits.

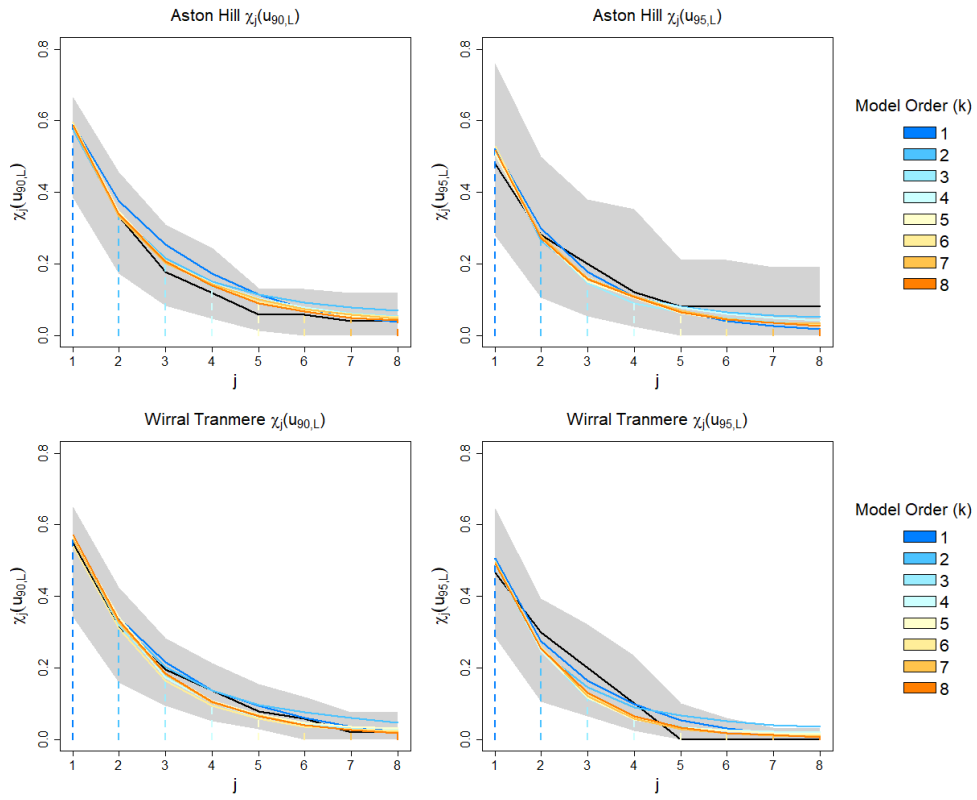


Figure 5.6: Estimates of the sub-asymptotic extremal dependence measure $\chi_j(v)$ for $j=1,\ldots,8$ at (top row) Aston Hill and (bottom row) Wirral Tranmere with $v=u_{90,L}$ and $v=u_{95,L}$ respectively. Lines show the empirical estimator (black) and model-based estimates for different order Markov models (rainbow). The grey regions are the 95% block-bootstrapped confidence intervals of the empirical estimate.

Figure 5.7 shows that the estimates for $\psi(i, u_{90,L})$ at Aston Hill overestimate the probability of observing an event at least $i = 1, \ldots, 8$ days long in an ozone season. The model of order $k = 2$ performs the worst, whereas the model of order $k = 1$ performs the best. This is the same at Wirral Tranmere, however the models of order $k = 1, 6, 7, 8$ captures the probability of seeing events at least 6 days long well. The estimates of $\psi(i, u_{95,L})$ model-based estimates are closer to the empirical estimates for both sites with $k = 4$ being the closest to the empirical estimates and $k = 1$ being the poorest at Aston Hill and at Wirral Tranmere the best is $k = 1$ and poorest is $k = 2$ or $k = 8$.

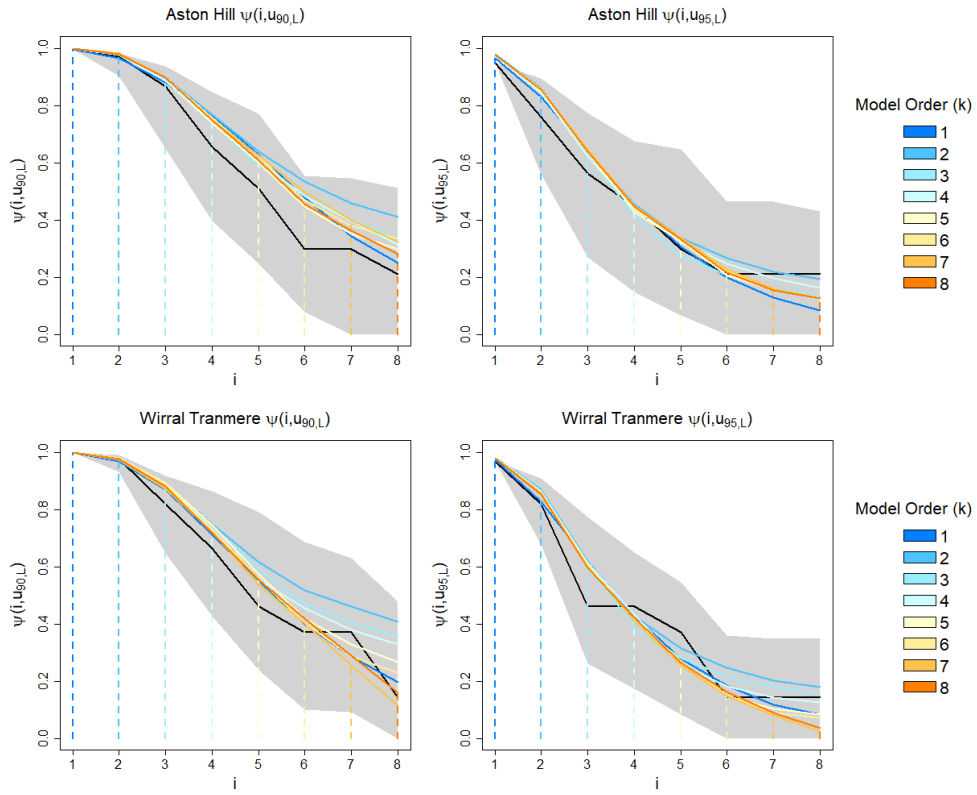


Figure 5.7: Estimates of $\psi(i, v)$ for $i = 1, \ldots, 8$ at (top row) Aston Hill (bottom row) and Wirral Tranmere, with $v = u_{90,L}$ and $v = u_{95,L}$ respectively. Lines show the empirical estimator (black) and model-based estimates for different order Markov models (rainbow). The grey regions are the 95% block-bootstrapped confidence intervals of the empirical estimate.

These figures indicate that, for Aston Hill, the best model would be order 1, 4, or 5 depending on short-, medium- or long-dependence structures and modelling

threshold, whereas at Wirral Tranmere the best model would be $k = 1$. Such visual diagnostics provide a subjective determination of the best model order. Thus to obtain a more objective choice, we use equation (5.3.10) over the cluster functionals $\chi_j(v)$ and $\psi(i, v)$ evaluated at $v = u_{90,L}, \ldots, u_{95,L}$ and across the first five lags to determine $\hat{\tau}$ for each sampled site. This indicates that Aston Hill requires a model of order $k = 4$ and it confirms that Wirral Tranmere requires a model of order $k = 1$. Across all the sampled sites we find that 22 $(36\%, 64\%)$, 13 $(38\%, 62\%)$, 5 $(20\%, 80\%)$, 5 $(60\%, 40\%)$, 3 $(0\%, 100\%)$, 4 $(50\%, 50\%)$, 3 $(0\%, 100\%)$, 2 $(0\%, 100\%)$ sites required models of order $k = 1, \ldots, 8$ respectively, where values in parentheses are the rural background and urban background percentage breakdowns of each order.

	Lon.	Lat.	$\theta(v)$	$\chi_j(v)$	$\psi(i,v)$	$\mathrm{W}_{\theta,\chi}$	$\mathrm{W}_{\theta,\psi}$	$\tau_{\omega,1}$	$\tau_{\omega,2}$
BAR3	-1.51044	53.56292	5	7	2	7	7	2	2
CANT	1.09806	51.27399	2	2	1	2	2	1	1
COAL	-1.56023	52.41156	8	1	1	8	8	1	1
GLKP	-4.24363	55.86578	2	7	1	2	2	1	1
HIL	-0.46086	51.49633	7	6	7	7	7	2	7
WIG5	-2.63814	53.54914	1	5	1	1	1	1	1
MACK	-2.68345	51.05625	1	6	4	1	1	4	1
HM	-0.80855	54.33494	3	7	2	3	3	2	2
SIB	1.46350	52.29440	2	1	1	2	2	1	1

Table 5.1: Determining the best model order at the nine case-study sites using different measures: Absolute difference in $\theta(v)$, mean absolute error in $\chi_j(v)$, mean absolute error in $\psi(i, v)$, Winter and Tawn measure, Winter and Tawn measure replacing $\chi_j(v)$ with $\psi(i, v)$, equation (5.3.10) looking at $\chi_j(v)$ and $\psi(i, v)$ and equation (5.3.10) looking over $\theta(v)$, $\chi_j(v)$ and $\psi(i, v)$ respectively where we evaluate at $v = u_{90,L}, \ldots, u_{95,L}$ and where appropriate average is taken over the first five lags.

We do find that the best model order changes depending on the statistical measure used. We show how the order differs for seven statistical measures at the case study sites, see Table 5.1. There appears to be some consistency between the model order determined by the difference in $\theta(v)$ and the Winter and Tawn measure, equation (5.3.9). In fact, 48 sites have the same model order across the two measures. However, one would suggest the Winter and Tawn measure is biased

towards $\theta(v)$ since the contributions from $\chi_j(v)$ and $\psi(i,v)$ do not alter the model order at majority (84%) of sites and their corresponding model orders frequently differ (81% and 75% respectively) to that of $\theta(v)$. Further, to suggest their measure is biased is when exchanging $\chi_j(v)$ for $\psi(i,v)$ as the model order for all sites remain the same.

Figure 5.8 shows the best model order for each site, estimated by equation (5.3.10). As one may expect sites which are close together are of similar order. Sites in the south-west on average require more complex models with 3 sites in Greater London having $\hat{\tau} = 8$. The Midlands typically require models of order $\hat{\tau} = 1, 2$ with a few requiring more complex models.

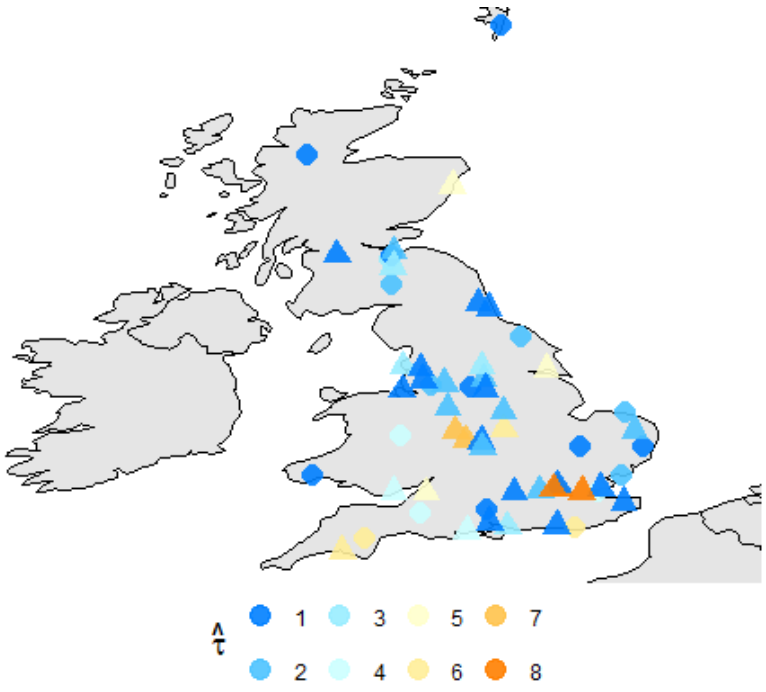


Figure 5.8: Map showing the best model order ($\hat{\tau}$) determined by equation (5.3.10) using the cluster functionals $\chi_j(v)$ and $\psi(i,v)$. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

5.4.5 Model verification

In statistical modelling, it is natural to ask how well the model performs on the trained dataset. To assess the performance, we compare the empirical estimates of

the cluster functionals $\theta(v)$, $\chi_j(v)$ and $\psi(i,v)$ to the model-based estimates at each site. We use the difference measure for $\theta(v)$ and the mean absolute error (MAE) for $\chi_j(v)$ and $\psi(i,v)$ (averaging over the first five lags). Further, we look at how the model performs at difference quantiles, $v = u_{90,L}$ and $v = u_{95,L}$.

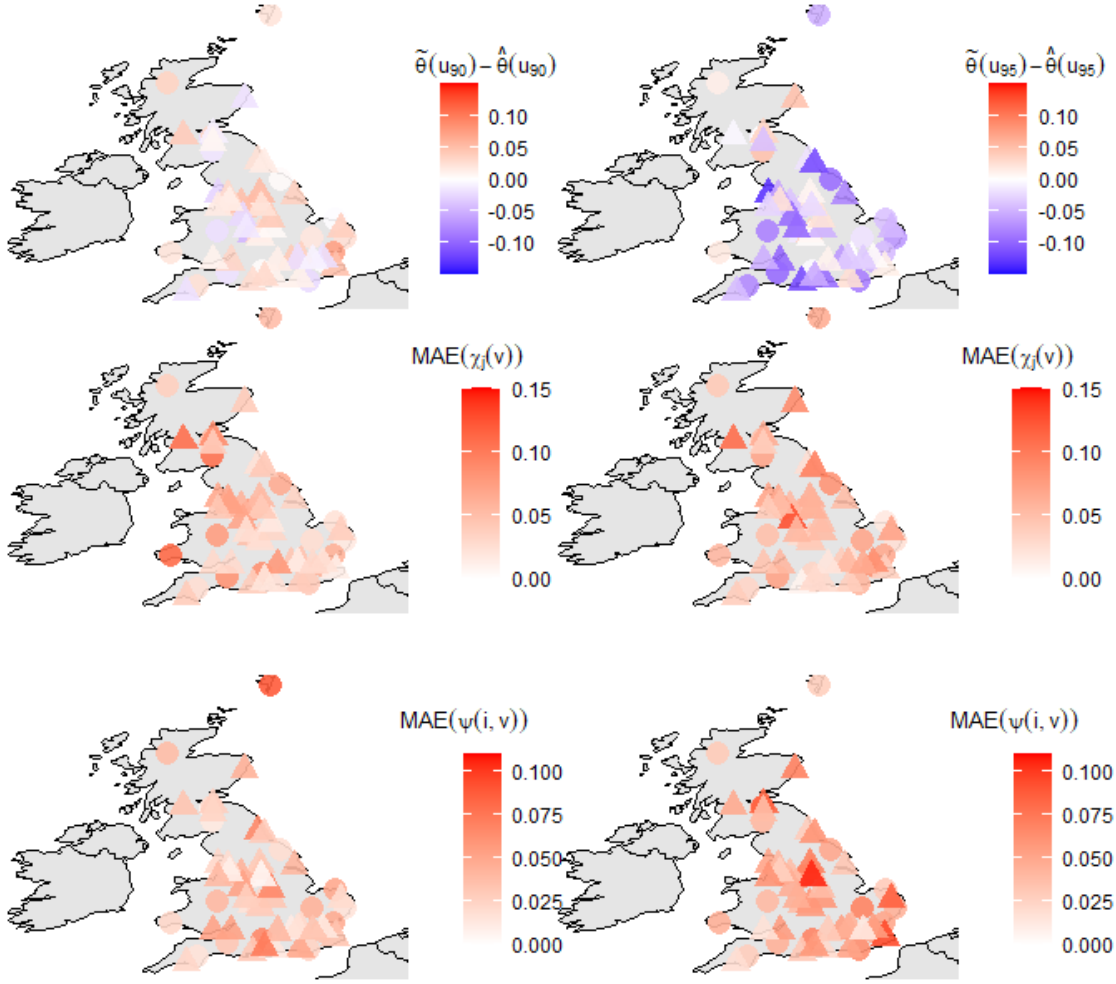


Figure 5.9: Difference between the empirical estimate and the model-based estimate of $\theta(v)$ and the mean absolute error of $\chi_j(v)$ and $\psi(i,v)$ respectively by row. We average over the first five lags for χ_j(v) and $\psi(i,v)$. The cluster functionals evaluated for $v = u_{90,L}$ and $v = u_{95,L}$ respectively by column. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

Figure 5.9 shows the difference between the empirical estimate and model-based estimate of the cluster functionals across all sites. For $\theta(u_{90,L})$ the difference ranges from -0.047 to 0.073 with a median difference of just 0.012 and 31.6% (38.5%)

of rural (urban) background sites underestimate. This increase to 71% and 84.2% respectively when $v = u_{95,L}$. The magnitude of underestimating increases with a difference ranging between -0.130 to 0.047 and a median value of -0.037.

There is little difference between the ranges and medians of the mean absolute error of $\chi_j(v)$, however the spatial distribution of these change. For $v = u_{90,L}$ the MAEs are larger on the west-coast with the smallest in East Anglia. This changes to the south-coast having the smallest MAEs and North England and Scotland having the largest for $v = u_{95,L}$. For $\psi(i, v)$ there appears to be no obvious spatial trend in the MAEs but a noticeable difference in the magnitude between the two quantiles. The range goes from (0.006,0.088) to (0.009, 0.126) and the median increases by 63% from 0.029 to 0.046. Note all the cluster functionals take values between zero and one. These estimates and their corresponding standard errors computed by the modified block-bootstrap approach, demonstrated in Table 5.2 for our case study studies, indicate little difference between the empirical estimates and the model-based estimates meaning that the model captures the extremal structure of the trained data well.

Code	$\theta(v)$	$\chi_j(v)$	$\psi(i, v)$
BAR3	0.013 (0.017)	0.059 (0.010)	0.046 (0.006)
CANT	0.048 (0.009)	0.014 (0.002)	0.034 (0.010)
COAL	0.029 (0.018)	0.032 (0.005)	0.023 (0.011)
GLKP	0.040 (0.016)	0.108 (0.011)	0.033 (0.003)
HIL	0.042 (0.015)	0.045 (0.007)	0.012 (0.017)
WIG5	0.013 (0.024)	0.074 (0.013)	0.009 (0.008)
MACK	-0.019 (0.018)	0.082 (0.007)	0.032 (0.017)
HM	0.004 (0.011)	0.064 (0.007)	0.019 (0.009)
SIB	0.024 (0.012)	0.017 (0.003)	0.016 (0.008)

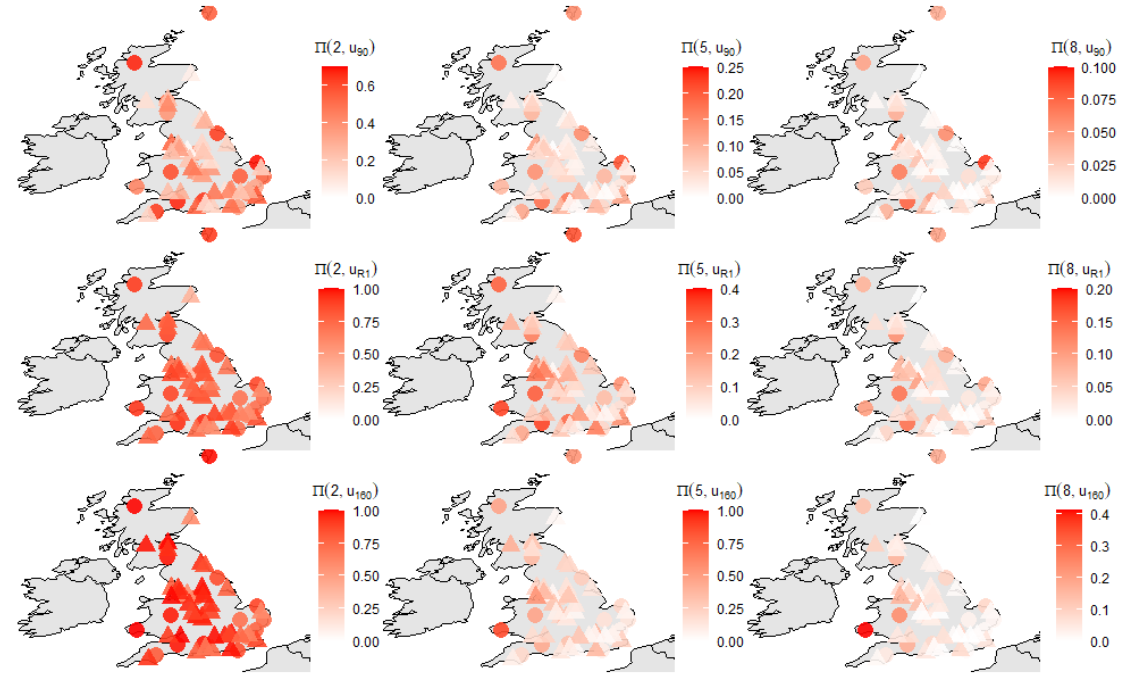
Table 5.2: Difference between the empirical estimate and the model-based estimate of $\theta(v)$ and mean absolute error of $\chi_j(v)$ and $\psi(i, v)$ at the nine case study sites with $v = u_{90,L}$. Mean is taken over the first five lags. Standard errors are given in parentheses computed by the modified block-bootstrap approach.

5.5 Forecasting

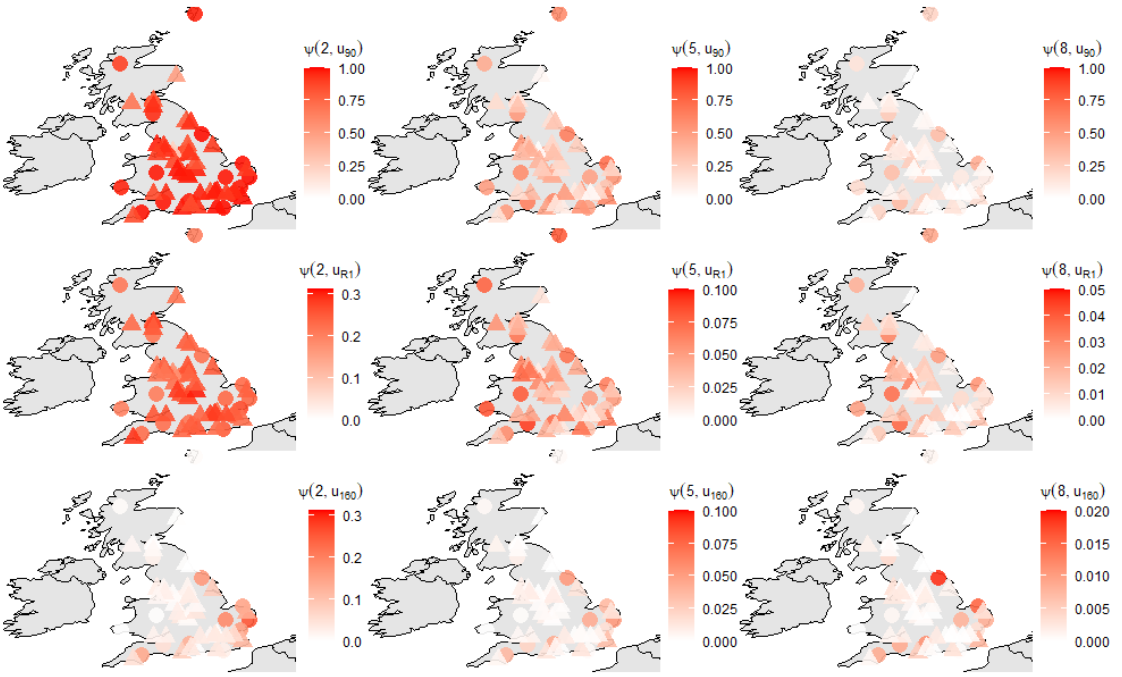
Scenario simulation

Scenario forecasting aids in warning system development and decisions of policy makers by providing useful information into what could be expected during an extreme event. Such useful information includes the cluster functionals. Here, we provide cluster functional estimates for three scenarios where the initial exceedance is greater than: (a) the marginal 90th quantile (b) the site-specific one-year return level, and (c) 160 $\mu g/m^3$ (minimum for DAQI of 7, high health risk). These scenarios are selected based on modelling quantities and public health risk levels. To obtain the cluster functional estimates we simulate 5000 extremal chains by following the algorithm in Section 5.3.2 but using an alternate cluster definition - where a cluster is said to have ended once the concentration falls below 100 $\mu g/m^3$. This is done since a concentration below 100 $\mu g/m^3$ is said to pose a low risk to public health according to the DAQI. We present cluster functional estimates for $i = 2, 5$, and 8 to demonstrate the short-, medium- and long-range dependence structure. Throughout, across site standard deviations are given in parentheses.

Figure 5.10a shows the estimated cluster functionals $\Pi(i, v)$ for each scenario. Increasing X_0 increases the probability of observing a longer event. Further sites, in particular rural background sites, on the coast tend to have the largest probability of witnessing longer extreme events. On average the probability that after the first exceedance the concentration is above 100 $\mu g/m^3$ are 0.481 (0.155) and 0.267 (0.151) across the rural and urban background sites, respectively. This is approximately 6-fold higher than the empirical estimates 0.086 (0.033) and 0.048 (0.025), respectively. There is little difference in scenario (b) and (c)'s probabilities of witnessing an event at least two days long at rural background sites and urban background sites. However, rural background sites have on average more than twice the chance of seeing an event longer than five and eight days than urban background sites and almost three times the probability in the case of scenario (a).



(a) Estimates of $\Pi(i, v)$



(b) Estimates of $\psi(i, v)$

Figure 5.10: Cluster functional estimates for each scenario (rows) and for $i = 2, 5, 8$ day of the simulated event (columns). Note the changing of scales throughout the maps. ($\circ$) points denote rural background sites and ($\triangle$) points denote urban background sites.

Figure 5.10b gives the probability of observing an extreme event in an ozone season.

There is little difference in the estimate between rural background and urban background sites for at least two-day long events for all three scenarios. Under scenario (a) the probability we observe an event of at least five days long above the moderate health risk is on average 0.477 (0.145) and 0.273 (0.153) at rural and urban background sites respectively. This decreases to 0.005 (0.006) and 0.001 (0.001) for scenario (c). Scenario (c) indicates that the east-coast and south-coast sites have the largest probability of witnessing a longer event in an ozone season.

Short-lead-time forecasting

As short-lead-time forecasting plays a fundamental role in providing the public with health warnings, we wish to demonstrate the forecasting skill of the kth-order Markov extremes model by forecasting out-of-sample events. More specifically, we extract all extreme events that occurred in 2020, an out-of-sample year, at each of our sampled monitoring sites and use the fitted kth-order Markov extremes model to see how well the model does at *forecasting* these extreme events given the initial exceedance. This is conducted by, for each sampled monitoring site and extreme event, simulating 5000 extremal chains using the simulation algorithm in Section 5.3.2 with initialising step of setting v to be the initial observed exceedance. These simulations are then used to determine the evolution of the event. Note three sites did not witness an extreme event during 2020 and are omitted here.

Figure 5.11 depicts the distribution of ozone concentration on the next i days after the initial exceedance for two events at Aston Hill and Wirral Tranmere. The points indicate the observed concentration and evolution of the event. Such figures can be used to determine the probability of persistence and provide probabilities of remaining in different risk levels. For example, the Figure 5.11(b) indicates that for days 1, 2 and 3 after the initial exceedance the probability of being at least moderate risk ($>100\,\mu g/m^3$) is 0.843, 0.611, 0.467 respectively. These figures (across all sites) indicate that on average the median of the chains on each day captures the correct classification, marginal threshold exceedance or non-exceedance.

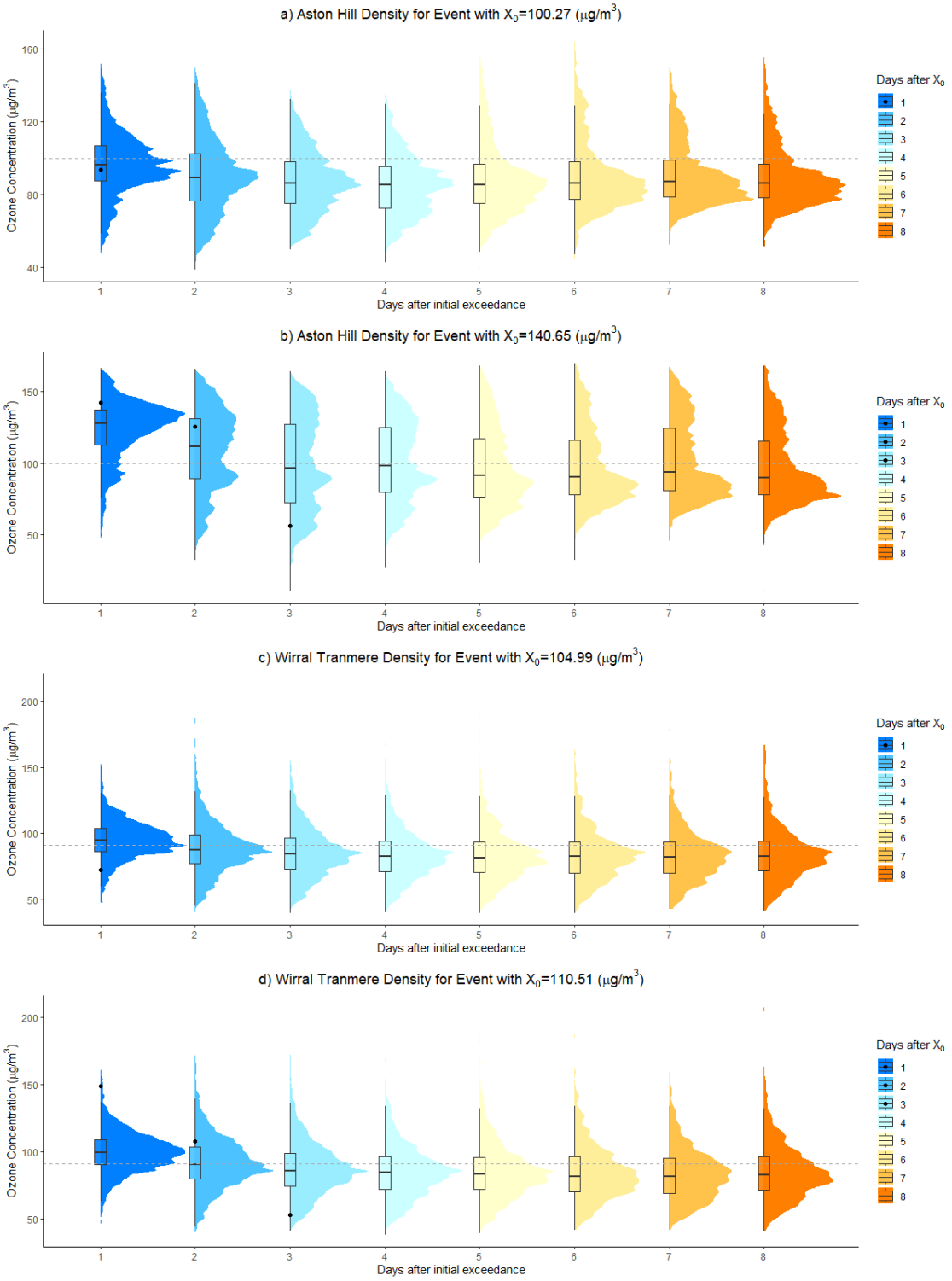


Figure 5.11: Density plots of ozone concentration on the ith day after the initial exceedance for two events at (a, b) Aston Hill and (c, d) Wirral Tranmere, with the marginal threshold given by the grey dashed line. The black points indicate the observed concentration on each day of the event (the threshold exceedances and the first non-exceedance).

To numerically quantify the forecasting performance of the model we compute the sub-asymptotic extremal index $\theta(v)$ and the average integrated concentration discrepancy across the extreme events with the median estimating the concentration on each day. On average, the $\theta(v)$ estimates indicate an approximate underestimation of the length of extreme event by one day, however 35 sites have larger empirical $\theta(v)$ estimates for 2020 than $2011 - 2019$ which indicates longer extreme events were observed. Figure 5.12 shows the average integrated difference using all the simulated chains or chains of observed event length only, respectively. Using all the simulated chains the kth-order Markov extremes model on average overestimates the ozone concentration by approximately $9.471\,\mu g/m^3$ per day, whereas the chains of observed event length indicate an average overestimation of $3.368\,\mu g/m^3$ at 45 sites and underestimates at 12 sites by approximately $1.358\,\mu g/m^3$ on average per day. The integrated difference is largest around London and south-east England indicating a potential improvement is required.

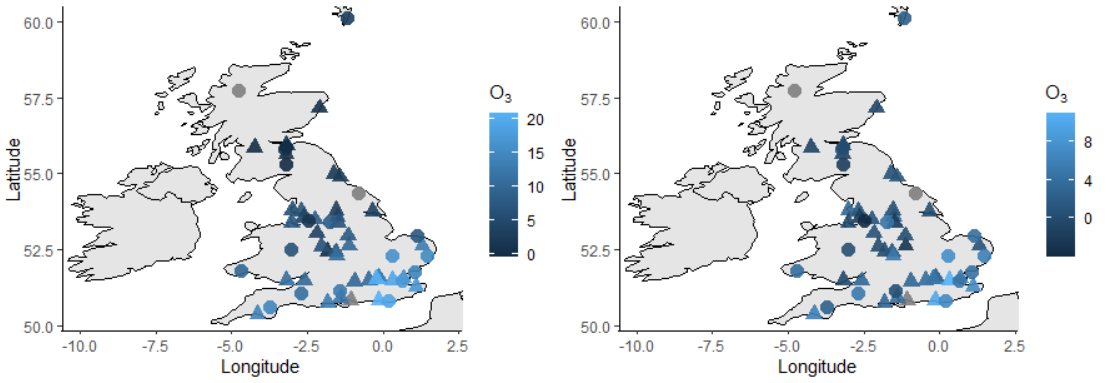


Figure 5.12: Maps showing the average integrated difference between the observed event and model-based median estimates using (left) all the simulated chains (right) chains of observed event length. Grey coloured sites are the sites where no extreme events occurred. (○) points denote rural background sites and (△) points denote urban background sites.

5.6 Conclusions

Extreme levels of DM8 ozone often occur in clusters exhibiting temporal dependence of the form asymptotic dependence and asymptotic independence. To capture this

structure, we used the kth-order Markov extremes model of Winter and Tawn (2017). To identify the optimal model order, we introduced a new statistic that adapts work of Knutti et al. (2017) by computing weights for each model order based on cluster functional estimates across a range of lags and quantiles. For our sampled AURN sites we found that on average urban background sites required a more complex model compared to rural background sites and the selected model order provides suitable approximations of the dependence structure.

In Section 5.5 we introduced two novel applications of the model, simulating important scenarios and short-lead-time forecasts of an out-of-sample event. We forecast all 2020 extreme events at all sites and found that the kth-order Markov extremes model provides good forecasts by capturing the duration of the events with an average overestimation of ozone concentration less than $3\,\mu g/m^3$ on each day of the event. The model struggles to capture the correct magnitude of ozone around London. To improve these forecasts, one could use more sites within the region, using the London Air Quality Network, and average over them to provide a more representative estimate. A bias correction technique, similar to the one described in Section 2.4.2 and used by the UK Met Office, may also be appropriate here as one could correct for the simulation error throughout an event or pool information across sites which have similar temporal dependence structure.

Although we have only presented for DM8 ozone concentrations this model can easily be implemented on the hourly scale given an appropriate condition to end an extreme event, for example a run length of at least twelve hours to account for the diurnal cycle. Further extensions include modelling the marginal parameter estimates or the threshold as a function of time in the form of a Fourier series to account for all seasonality. Another adaptation of the model would be to incorporate covariates such as temperature, pressure and precursor concentrations in both the marginal and conditional parameters. Tendijck et al. (2021, 2023) has started extensions by modelling multivariate time series of oceanographic data with a multivariate kth-order Markov extremes model. In operational terms, we would dynamically update

the dependence parameters and model order as more measurements are taken and extending the time series to make sure the model provides efficient cluster functional estimates and forecasts. Further, we would want to be using backward simulation processes from each exceedance in the chain to obtain the simulated chains that have captured the previous estimates accurately and then simulate these chains forward to obtain cluster functional estimates.

Chapter 6

A spatial Bayesian hierarchical model for extreme ozone

6.1 Introduction

Ozone (O_3) is a natural constituent of the atmosphere and is present in both the stratosphere and the troposphere. Tropospheric ozone is a secondary, trans-boundary air pollutant that is formed through photochemical reactions between anthropogenic nitrogen oxides and volatile organic compounds. Unlike stratospheric ozone which protects life on earth, tropospheric ozone can cause adverse effects when present in high concentrations. Short-term exposure can lead to difficulty breathing and aggravation to pre-existing respiratory issues like asthma, emphysema and chronic bronchitis; whereas long-term exposure can lead to development of such respiratory illnesses and potential death (Bell et al., 2004; Wilson et al., 2014; WHO, 2013).

Statistical modelling of high (extreme) concentrations of ozone is crucial in providing efficient forecasts and warnings to the public to mitigate morbidity. Extreme value theory (EVT) provides a method to characterise the tail behaviour of any distribution and the fundamentals can be found in for example: Pickands (1975); Leadbetter et al. (1983); Resnick (2007); Beirlant et al. (2004); Coles (2001). In

recent times, modelling spatial extremes has gained traction and three distinct methods have been developed.

The first is based on *max-stable processes* which are the natural generalization of the generalised extreme value (GEV) distribution to the infinite-dimensional setting. Max-stable processes arise by looking at the limit process of the renormalised pointwise maxima of the spatial stochastic process, see e.g. Davison et al. (2012), Smith (1990b) or de Haan and Ferreira (2007) for details of the approach. These models have been used in a wide variety of environmental applications e.g. Vettori et al. (2019) to model maxima of air pollution concentration and temperatures in Los Angeles and Oesting et al. (2016) uses a bivariate Brown-Resnick process to model extreme wind gusts over Germany. However, inference for max-stable models in high-dimensions is difficult due to the complicated form of the likelihood function (Padoan et al., 2010; Ribatet et al., 2012). Further, the block maxima approach relies on creating artificial spatial block maxima; that is if the chosen block-resolution is too large or too small then one can be imposing unobserved spatial structure when obtaining the block maxima values. To model the observed spatial processes directly one can use the analogue of max-stable processes, namely the *generalised Pareto processes* where one uses the threshold exceedance approach in a spatial context (Ferreira and de Haan, 2014). These models also benefit from usually having simpler likelihood functions. However, the issue with max-stable and generalised Pareto processes is the limited dependence structures they can capture; full independence and asymptotic dependence.

The extremal dependence class of a stochastic process is determined by the conditional exceedance probability between two sites. More precisely, let $\mathbf{X} = (X_1, \ldots, X_m)$ be the observed concentration of ozone at m sites, often denoted as belonging to the set of all sites $\mathcal{S}$ such that $|\mathcal{S}| = m$, then a stochastic process $X(\mathbf{s})$ over $\mathcal{S} \subset \mathbb{R}^m$, is said to be asymptotically dependent if for any two sites $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$

the conditional exceedance probability

$$\chi_u(\mathbf{s}_1, \mathbf{s}_2) = \Pr\left(X(\mathbf{s}_1) > F_1^{-1}(u) | X(\mathbf{s}_2) > F_2^{-1}(u)\right), \tag{6.1.1}$$

has a positive limit as $u \to 1$ (Coles et al., 1999), where F_1, F_2 are the random variables associated with $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ respectively and $F_{\cdot}^{-1}$ are the corresponding inverses. The stochastic process is asymptotically independent if the limit of equation (6.1.1) is zero. Often in environmental data the strength of asymptotic dependence weakens with quantiles and distance i.e. more extreme events occur at the same time at locations closer together, while extremes at further away locations rarely occur at the same time. As such, misclassification of the dependence structure can lead to overestimation of risk measure estimates.

The second method for modelling spatial extremes is the *spatial conditional extremes model* which alleviates this issue as it allows for a broader range of dependence structures (Wadsworth and Tawn, 2022; Richards et al., 2023; Shooter et al., 2021). This model allows for both asymptotic dependence and asymptotic independence by conditioning on a single site being extreme in place of assuming all components of the spatial process have a positive probability of being jointly extreme simultaneously. Wadsworth and Tawn (2022) proposed the use of composite likelihoods, multiplying the likelihoods obtained by conditioning on each site, to combine information and overcome the limitation of having no natural conditioning site. Richards et al. (2023) extends the model by allowing a mixture of two components with different marginal and dependence models with an application to precipitation across southern UK. A further advantage of this model over the max-stable and generalised Pareto processes is the scalability to larger spatial domains due to the semi-parametric method and use of composite likelihoods. However, a limitation of the conditional approach is it is more heavily parametrised and difficult to interpret unconditionally.

Although the conditional approach to modelling spatial extremes allows for a wider range of dependence structures, can be applied in reasonably high dimensions and provides simple conditional simulation at unobserved locations (Wadsworth and

Tawn, 2022), the marginal distribution at unobserved locations and risk measures such as return levels are difficult to estimate. This is due to the use of a copula formulation, where the marginal models are fitted separately for each site before the dependence component is estimated. Consequently, the interpretation of the model is difficult. For a extensive review of the above models see Huser and Wadsworth (2020).

The final method uses *Bayesian hierarchical models* with a latent spatial structure to describe the spatial variation of the marginal distribution parameters. The advantage of these models is the capability of obtaining marginal parameter estimates and uncertainties at locations with little or no observations. This is done through obtaining samples from the posterior and predictive distributions using Markov chain Monte Carlo (MCMC) algorithms and assuming that the data model parameters vary smoothly over space. However, due to the inference procedure, for increasing number of locations m the procedure becomes increasingly computationally intensive as the inverse of a $m \times m$ covariance matrix is computed at each MCMC step. A solution to this dimension limitation is proposed in Banerjee et al. (2008). However, hierarchical models cannot be used to quantify spatial events due to the site to site independence assumption.

The first instance of latent processes in spatial extremes was presented by Coles and Casson (1998) and Casson and Coles (1999) who modelled wind speed data of hurricanes on the U.S. Gulf Coast. Cooley et al. (2007) developed a hierarchical spatial model for the generalised Pareto distribution (GPD) parameters and constructed maps of extreme precipitation return levels in Colorado. Other instances of using hierarchical models include: Sang and Gelfand (2009) who employed a spatio-temporal model to extreme precipitation by assuming conditional independence given the spatially correlated parameters; Clancy et al. (2016) used a spatial Bayesian hierarchical model to capture the spatial variation of extreme sea states in the west coast of Ireland; Sharkey and Winter (2019) proposed the use of an adjusted likelihood to account for spatial and temporal dependence by imposing

a condition of spatial similarity on the Bayesian hierarchical model parameters with an application to precipitation across UK.

Many studies have explored air pollutants and air quality, for example: Kitabo (2020) explored ozone using max-stable and kriging methods across South Korea, Menezes et al. (2016) studied NO_2 in Portugal using geostatistical tools from a spatio-temporal approach and Russell et al. (2016) uses a hierarchical approach to spatially model the effects of meteorological drivers of extreme ozone. Further studies that explored all levels of ozone in a Bayesian framework include, Huerta et al. (2004) who developed a spatio-temporal model to describe the ozone levels across Mexico City and Mukhopadhyay and Sahu (2018) who estimated long-term exposure to outdoor air pollution across the UK using a Bayesian spatio-temporal model. However, to our best knowledge, no studies have analysed just the extreme levels of ozone with a spatial Bayesian hierarchical model. Thus, this work focuses on the spatial analysis of ozone threshold exceedances across Great Britain using the Bayesian framework and adapts the work of Cooley et al. (2007) for our application. The goal of this study is to propose a hierarchical model with a spatial structure in the GPD parameters which is generalised by a latent spatial process. Further we explore how including different geographical covariates as regression coefficients in the latent process effects the model fit and the risk measures, here we look at return levels.

The chapter is structured as follows. Section 6.2 describes the ozone data used in the study. Section 6.3 provides the details of the model used by first outlining the generalised Pareto distribution and then each layer of the three-layer Bayesian hierarchical model used to model the GPD parameters. Section 6.3.3 describes the ordinary kriging process to create the threshold surface, followed by the Markov chain Monte Carlo inference and the interpolation procedure in Section 6.3.4 and Section 6.3.5 respectively. The results and model validation are presented in Section 6.4 with conclusions and further work explored in Section 6.5.

6.2 Data

This study is performed using the daily maximum of the 8-hour running mean (DM8) ozone concentration observations from 57 monitoring sites sampled from the Automatic Urban and Rural Network (AURN) across Great Britain for the period $2011 - 2019$, as described in Section 2.4.1, and for the months of March to August inclusively. Figure 6.1 provides the locations of the monitoring sites and surfaces of two geographical covariates used here, distance-to-coast (Dist) and elevation (Elev). The covariate surfaces are found through R packages, **rnaturalearth** and **elevatr** respectively. Further, Table 6.1 provides details of the nine monitoring sites we withhold from the model fitting process, which are used to cross-validate our model.

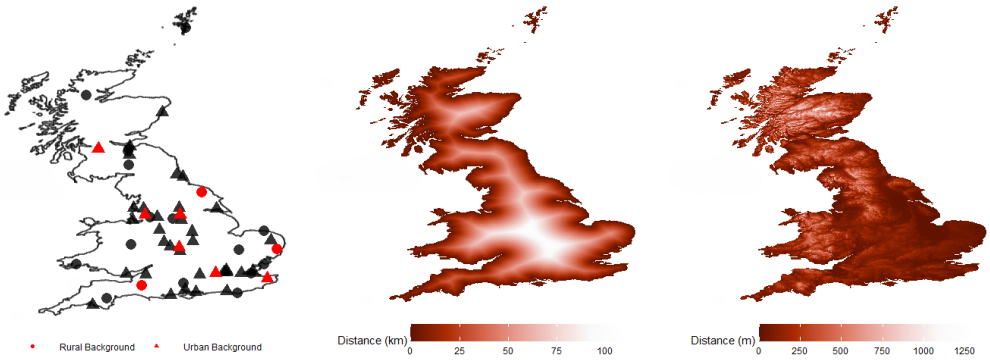


Figure 6.1: Maps of (left) the AURN sites used where ▲ (●) denotes urban (rural) background sites and those in red are the randomly selected sites used as a validation set (middle) the distance-to-coast covariate surface and (right) the elevation covariate surface.

Code	Long.	Lat.	Min.	1st qu.	Median	3rd qu.	u_{90}	Max.
BAR3	-1.51044	53.56292	8.82	52.08	62.04	72.59	83.29	154.81
CANT	1.09806	51.27399	6.88	67.12	78.02	88.83	102.74	176.18
COAL	-1.56023	52.41156	23.83	59.10	69.58	80.74	92.31	146.23
GLKP	-4.24363	55.86578	21.22	51.74	62.35	73.97	84.43	139.44
HIL	-0.46086	51.49633	2.64	38.58	54.39	69.84	82.92	154.67
WIG5	-2.63814	53.54914	6.34	60.61	71.05	82.62	92.00	175.44
MACK	-2.68345	51.05625	10.26	67.91	80.08	91.77	102.59	165.03
HM	-0.80855	54.33494	27.16	67.46	79.52	90.63	101.38	182.69
SIB	1.46350	52.29404	13.45	67.76	77.76	87.65	100.30	206.03

Table 6.1: Information of the nine case study monitoring sites: site-code, latitude and longitude, minimum, maximum and quantiles of the observed DM8 ozone data.

6.3 The model

6.3.1 Extreme value theory

Extreme value theory (EVT) mainly deals with modelling the tail of a distribution. The most popular approach in tail behaviour is to define an extreme value as an exceedance of a sufficiently high threshold u, such that the threshold exceedances should approximately follow a generalised Pareto distribution (GPD), as given in equation (3.2.5). The tail of the distribution is either bounded ($\xi < 0$), light ($\xi \to 0$) or heavy ($\xi > 0$). Further, EVT provides a method to extrapolate beyond the range of the data and estimate quantities such as return levels. T-year return levels can be estimated by using the equation (3.2.10).

6.3.2 Spatial hierarchical model for threshold exceedances

We are interested in producing return level maps of DM8 ozone across Great Britain and thus must first construct maps for the threshold exceedances and the threshold. To do this we employ Bayesian methods to estimate spatial hierarchical models as they allow for simpler computation of parameter estimates compared to maximum likelihood estimation. Here, we first introduce the Bayesian framework and then build the structure of our spatial hierarchical model used.

Bayesian framework

Let $\mathbf{X}$ denote a vector of observations from some distribution depending on unknown parameter $\boldsymbol{\theta}$ and some covariates $\mathbf{Z}$. The joint probability distribution for $\mathbf{X}$, $\mathbf{Z}$ and $\boldsymbol{\theta}$ is given by,

$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) = p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z}), \tag{6.3.1}$$

where $p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta})$ is the data likelihood and $p(\boldsymbol{\theta}|\mathbf{Z})$ is the prior distribution. Conditioning on the observed data $\mathbf{X}$, we obtain an expression for the posterior

density of the parameters

$$\pi(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z}) = \frac{p(\mathbf{X}|\mathbf{Z},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z})}{p(\mathbf{X},\mathbf{Z})}, \tag{6.3.2}$$

where $p(\mathbf{X},\mathbf{Z}) = \int p(\mathbf{X}|\mathbf{Z},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z})\, d\boldsymbol{\theta}$ is the marginal distribution of the data. We focus on only the numerator since $p(\mathbf{X},\mathbf{Z})$ is often intractable, as such equation (6.3.2) can be expressed as

$$\pi(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z}) \propto p(\mathbf{X}|\mathbf{Z},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z}). \tag{6.3.3}$$

Equation (6.3.3) forms the basis for Bayesian statistics and provides the building blocks for hierarchical models (Gelman et al., 2003).

Bayesian hierarchical model

Hierarchical models provide a method to model a complex process and the relationship to observations without the issue of over-fitting that non-hierarchical models can face (Gelman et al., 2003). From equation (6.3.3) a Bayesian hierarchical model can be built. Following Cooley et al. (2007) a three layer hierarchy can be constructed where the first layer of the hierarchy models the data at each monitoring site, the second layer models the latent process that drives the data and the third defines the prior distributions for the latent parameters given in layer two. Let $\mathbf{X}(\mathbf{s}) = (\mathbf{X}(\mathbf{s}_1),\ldots,\mathbf{X}(\mathbf{s}_n))$ be the matrix of observations where $\mathbf{X}(\mathbf{s}_i)$ is the observations at location $\mathbf{s}_i$. Then Bayes' rule provides the inference for the parameters in our models $\boldsymbol{\theta}$ given the data at each location $\mathbf{s}$;

$$\pi(\boldsymbol{\theta}|\mathbf{X}(\mathbf{s})) \propto p(\mathbf{X}(\mathbf{s})|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{6.3.4}$$

where π denotes a *posterior* distribution, $p(\mathbf{X}(\mathbf{s})|\boldsymbol{\theta})$ is the data likelihood and $p(\boldsymbol{\theta})$ is the prior distribution. We can express equation (6.3.4) in terms of the conditional

distributions of our hierarchical model,

$$\pi(\boldsymbol{\theta}|\mathbf{X}(\mathbf{s})) \propto p_1(\mathbf{X}(\mathbf{s})|\boldsymbol{\theta}_1)p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p_3(\boldsymbol{\theta}_2), \tag{6.3.5}$$

where p_1 is the data likelihood given the latent process, p_2 is the prior density of the latent process given the latent parameters and p_3 is the prior density for the latent parameters. Each layer of our hierarchical model is described in the remainder of this subsection. Section 6.3.3 provides the model and method to generate the threshold surface. Section 6.3.4 describes the parameter estimation techniques that are employed and Section 6.3.5 provides the interpolation procedure.

Data model

In the first layer, we assume that the ozone exceedances of the marginal 90th quantile at each monitoring site follow a generalised Pareto distribution with separate scale parameters (σ_i) and separate shape parameters (ξ_i). The exceedance rate φ is fixed to 0.1, resulting in the threshold varying spatially, see Section 6.3.3 for details.
Following Cooley et al. (2007), we reparametrise the scale parameter, $\phi_i = \log \sigma_i$, allowing ϕ_i to take on both positive and negative values, although we note the orthogonal parametrisation of Chavez-Demoulin and Davison (2005) could be a useful alternative. We do this as we are assuming the underlying process is Gaussian. For each monitoring site a separate GPD is fitted, resulting in the likelihood function given by

$$p_1(\mathbf{X}(\mathbf{s}) \mid \boldsymbol{\theta}_1) = \prod_{i=1}^{m}\prod_{k=1}^{n_i} \frac{1}{\exp(\phi(\mathbf{s}_i))}\left[1 + \xi(\mathbf{s}_i)\left(\frac{x_k(\mathbf{s}_i) - u(\mathbf{s}_i)}{\exp(\phi(\mathbf{s}_i))}\right)\right]^{-1/\xi(\mathbf{s}_i)-1}, \tag{6.3.6}$$

where $\boldsymbol{\theta}_1 = [\boldsymbol{\phi}, \boldsymbol{\xi}]^T$, $\mathbf{s}_i$ are the spatial coordinates of the ith site, m is the number of sites, n_i is the number of observations at the ith site and $x_k(\mathbf{s}_i)$ is the kth observation at the ith site.

Process model

In the second layer of our hierarchy, the parameters of the GPD are assumed to follow a spatial model with two separate spatial features: one describing spatial trends in the GPD parameters and the other describing the dependence that cannot be attributed to co-dependence on a geographical covariate. As such, using standard geostatistical methods (Matheron, 1963; Cressie, 1993), the parameter $\phi(\mathbf{s})$ is specified through a Gaussian process with $\mathbb{E}[\phi(\mathbf{s})] = \mu_\phi(\mathbf{s})$ and $\text{Cov}(\phi(\mathbf{s}), \phi(\mathbf{s}')) = C_\phi(\mathbf{s}, \mathbf{s}')$. The mean $\mu_\phi(\mathbf{s})$ is a function of parameters $\boldsymbol{\alpha}_\phi$ and the covariates $\mathbf{Z}(\mathbf{s})$:

$$\mu_\phi(\mathbf{s}) = f_\phi(\boldsymbol{\alpha}_\phi, \mathbf{Z}(\mathbf{s})). \tag{6.3.7}$$

The covariance is a function of the distance between monitoring sites and parameters $\boldsymbol{\beta}_\phi$ and κ_ϕ,

$$C_\phi(\mathbf{s}, \mathbf{s}') = \beta_{\phi,0} \exp\left(-\left(\frac{||\mathbf{s} - \mathbf{s}'||}{\beta_{\phi,1}}\right)^{\kappa_\phi}\right), \tag{6.3.8}$$

which corresponds to a κ-exponential covariance function. The parameters $\beta_{\phi,0}$ and $\beta_{\phi,1}$ are often called the *sill* and the *range*, respectively. The parameter $\beta_{\phi,1}$ controls how quickly the spatial dependence decays as a function of the distance apart, large $\beta_{\phi,1}$ represents long-range correlations and small $\beta_{\phi,1}$ means that points close together are nearly independent. This model assumes the process is stationary and isotropic; invariant to translation and rotation. As such the second piece of equation (6.3.5) is given by,

$$p_2(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2) = \frac{1}{\sqrt{(2\pi)^m |\Sigma_\phi|}} \exp\left(-\frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\mu}_\phi)^T \Sigma_\phi^{-1} (\boldsymbol{\phi} - \boldsymbol{\mu}_\phi)\right) p_\xi(\boldsymbol{\xi}|\boldsymbol{\theta}_\xi), \tag{6.3.9}$$

where $\boldsymbol{\mu}_\phi$ is defined by equation (6.3.7) evaluated at the covariates of the monitoring site locations $\boldsymbol{s}_i$, Σ_ϕ is the covariance matrix produced by equation (6.3.8) at the monitoring site locations, and the density function p_ξ is defined as the prior distribution for the shape parameter $\boldsymbol{\xi}$ with parameters $\boldsymbol{\theta}_\xi = (\boldsymbol{\alpha}_\xi, \boldsymbol{\beta}_\xi)$ and

$\boldsymbol{\theta}_2 = [\boldsymbol{\alpha}_\phi, \boldsymbol{\beta}_\phi, \boldsymbol{\theta}_\xi]^T$, with $\boldsymbol{\theta}_2$ often termed the *hyperparameters*. The form of p_ξ is determined by the model fitted to the shape parameter, which also determines the form of $\boldsymbol{\theta}_\xi$. If the shape parameter is modelled spatially, we use a spatial Gaussian process with similar structure as that of the scale parameter, i.e. where $\boldsymbol{\alpha}_\xi$ and $\boldsymbol{\beta}_\xi$ exist and require prior distributions. If the shape parameter is assumed to be constant over the study space, a uniform prior is used. Consequently, the density function p_ξ is absorbed into the proportionality constant of equation (6.3.5).

Prior model

In the third and final layer of our hierarchy, we define the priors for the hyperparameters. Throughout, independence is assumed in all stages of the model including the prior stage,

$$p_3(\boldsymbol{\theta_2}) = p_{\boldsymbol{\alpha}_\phi}(\boldsymbol{\alpha}_\phi) p_{\boldsymbol{\beta}_\phi}(\boldsymbol{\beta}_\phi) p_{\boldsymbol{\alpha}_\xi}(\boldsymbol{\alpha}_\xi) p_{\boldsymbol{\beta}_\xi}(\boldsymbol{\beta}_\xi). \tag{6.3.10}$$

We have no prior information on how the GPD parameters ϕ and ξ relate to the covariates, therefore uninformative priors for $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\alpha}_\xi$ are chosen. For all models we set $\alpha_{\phi,i} \sim \text{Unif}(-\infty, \infty)$ and for all models that involve $\boldsymbol{\alpha}_\xi$ we set $\alpha_{\xi,i} \sim \text{Unif}(-\infty, \infty)$, which still provides proper posterior distributions, i.e. integrates to one over the support (Banerjee et al., 2004). Berger et al. (2001) and Banerjee et al. (2004) provide explanations to why improper priors for the sill and range parameters often results in improper posteriors. As a result informative priors are chosen to ensure proper posterior distributions are obtained.

As previously mentioned, a GPD is fitted independently at each monitoring site to obtain a maximum likelihood estimate for σ and thus ϕ. Following Cooley et al. (2007) an empirical variogram is fitted to $\hat{\boldsymbol{\phi}}$. By definition $\beta_{\phi,0}$ controls the sill of the variogram model given in equation (6.3.8). Figure 6.2 (left) demonstrates such a SSE-minimised variogram (solid) and the variogram associated with the prior chosen for $\beta_{\phi,0}$ (dashed). The prior chosen is $\beta_{\phi,0} \sim \text{Unif}(0.001, 0.15)$ as it provides a wide

envelope of possible variograms and can be used across all models tested.

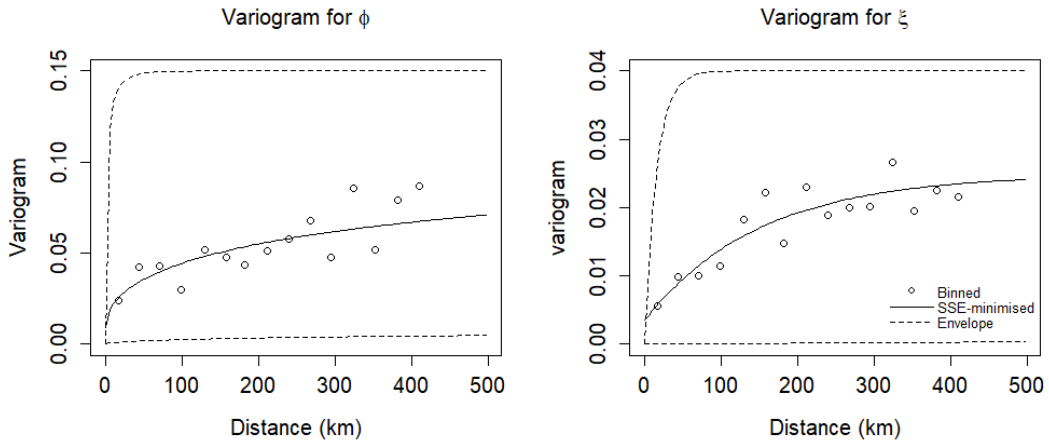


Figure 6.2: Binned variogram estimates (∘) and the SEE-minimising variogram (—) are plotted for the MLE-estimated ϕ and ξ parameters respectively. The dashed lines denote the envelope of possible variograms given the priors for $\beta_{\cdot,0}$ (sill) and $\beta_{\cdot,1}$ (range).

Cooley et al. (2007) used knowledge of the study space to determine the prior for $\beta_{\phi,1}$, the distance at which the correlation between the scale parameters is less than 0.05. For exponential variograms, this distance is approximately $3/\beta_{\phi,1}$. Therefore, the limits of a suitable uniform prior are approximately $3/\min(d)$ and $3/\max(d)$, where d is the distances associated with the binned variogram estimates as given in Figure 6.2. As we are using a κ-exponential variogram model to describe the covariance of $\phi(\mathbf{s})$ a different approach is required. Exploratory analysis of the variogram cloud and the behaviour of the posterior distributions guide our prior choice. The prior chosen is $\beta_{\phi,1} \sim \text{Gamma}(3,1)$ as we can control where the mass of the distribution lies and the probability of accepting unreasonable proposals. Further we reject the proposed values of $\beta_{\phi,1}$ and $\beta_{\phi,0}$ if the proposed value goes outside the range of observed distance between monitoring sites as our modelling assumptions may not hold beyond the support.

For the shape parameter $\xi(\mathbf{s})$, only when modelled as a Gaussian process are priors for $\boldsymbol{\alpha}_\xi$ and $\boldsymbol{\beta}_\xi$ required. Similarly to $\boldsymbol{\phi}$, following Cooley et al. (2007), empirical information is used to determine the prior for $\beta_{\xi,0}$. We choose $\beta_{\xi,0} \sim$

Unif$(0.001, 0.04)$ to provide a wide envelope of possible variograms, as seen in Figure 6.2 (right). Although for $\xi(\mathbf{s})$ we use the exponential variogram model, as suggested by the empirical variogram in Figure 6.2 (right), we still follow similar exploratory analysis approach taken for $\phi(\mathbf{s})$ and chose the prior $\beta_{\xi,1} \sim \text{Gamma}(4, 0.5)$.
Although we use a κ-exponential variogram model to describe the covariance of $\phi(\boldsymbol{s})$, we do not require a prior for κ_ϕ. We obtain the value for κ_ϕ by fitting a range of variograms with varying κ_ϕ values and take the value which corresponds to the SSE-minimised variograms. This results in $\kappa_\phi = 0.5$ and using a similar approach produces $\kappa_\xi = 1$. This decreases the number of parameters required to be estimated and therefore reduced the uncertainty in parameter estimation.

6.3.3 Model for threshold

To produce surfaces of return level estimates we require surfaces for the scale, shape and threshold. As defined in Section 6.3.2 we work with only the largest 10% of observations at each monitoring site, as such the threshold will vary spatially. To model the spatial variation we employ ordinary kriging (Matheron, 1963; Zimmerman et al., 1999). Therefore, the estimate of threshold u at an arbitrary location $(\mathbf{s}_0)$ is defined as

$$\hat{u}(\mathbf{s}_0) = \sum_{i=1}^{m} \lambda_i u(\mathbf{s}_i) \tag{6.3.11}$$

that minimises the mean squared prediction error. The kriging weights λ_i are derived from the estimated spatial structure of the empirical 90th quantiles from sites with data, in this case from the empirically fitted covariance function to these quantiles as given in Figure 6.3 (left). This corresponds to Stein's parametrisation of the Matérn covariance function given by,

$$\text{Cov}(||\mathbf{s}-\mathbf{s}'||) = \frac{\sqrt{\pi}\rho}{2^{\nu-1}\Gamma(\nu+\frac{1}{2})\gamma^{2\nu}}(\gamma||\mathbf{s}-\mathbf{s}'||)^{\nu}\mathcal{K}_\nu(\gamma||\mathbf{s}-\mathbf{s}'||) + \mathbb{I}_{||\mathbf{s}-\mathbf{s}'||=0}\cdot\tau^2, \tag{6.3.12}$$

where $\nu > 0$ represents the smoothness parameter, γ partially functions like an inverse range parameter and affects low-frequency behaviours, ρ is the overall scale

parameter that is seen in the high-frequency behaviour of the spatial process, τ^2 represents the nugget effect and $\mathcal{K}_\nu(\cdot)$ denotes the modified Bessel function of the second kind with order ν, (Matérn, 1960; Handcock and Stein, 1993; Stein, 1999; Loh, 2005).

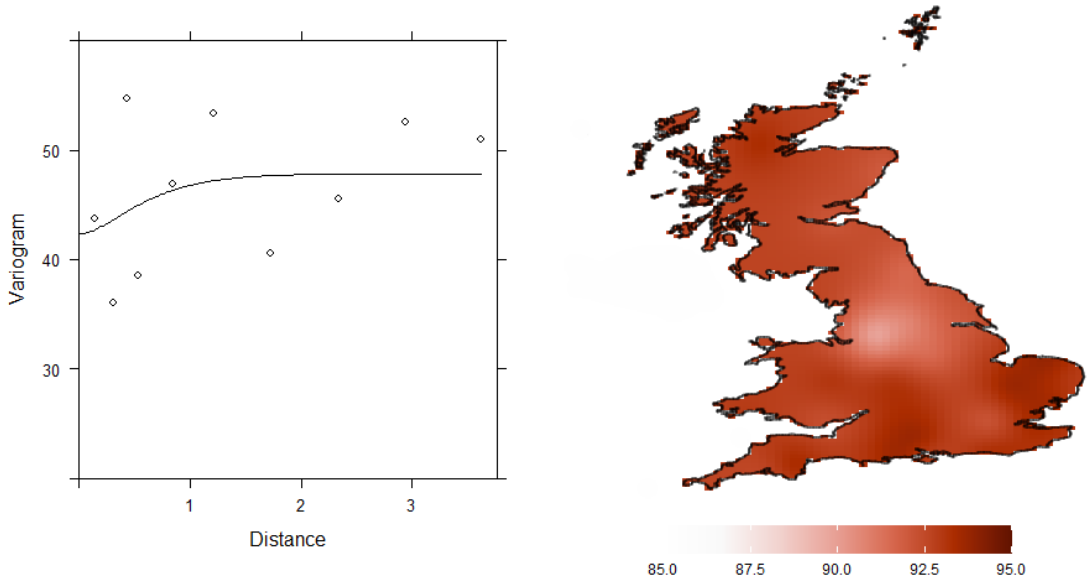


Figure 6.3: (left) The binned variogram estimates (∘) and the empirically estimated Stein's parametrisation of the Matérn covariance function for threshold. (right) The ordinary kriged threshold surface across Great Britain estimated by the variogram. Note: distances are in degrees.

The ordinary kriged surface of threshold, Figure 6.3 (right) indicates a relatively flat surface across Great Britain. The lowest thresholds are across North England, spanning from the urban background sites around Liverpool and Manchester to York. The highest thresholds are observed in East Anglia which is reflected in the surface. Due to the nature of the model, the thresholds vary around an unknown mean and thus the empirical thresholds at the monitoring sites do vary from the surface. The inverse distance weighted interpolation method was explored but the generated surface created too artificial circular neighbourhoods around the monitoring sites.

6.3.4 MCMC structure

The dimension of the model increases with the number of sites and covariates. An analytical solution for the parameters becomes impractical and intractable as there is no closed form for the normalising constant in the posterior. However, Markov chain Monte Carlo (MCMC) algorithms can be used to obtain approximate draws from the posterior distribution (Robert and Casella, 1999; Gelman et al., 2003). We employ Metropolis-Hastings (MH) steps within a Gibbs sampler to update each parameter of the model. This involves drawing a potential value from an appropriate candidate distribution and accepting or rejecting it according to the Hastings ratio (Geyer, 2011). The MH candidate densities for all parameters are implemented as random walks. The appropriate steps of the random walks are selected to approximately achieve the optimal rate of acceptance of 0.434 and 0.23 for parameters of dimension 1 to 4 or greater than 4 respectively (Gelman et al., 1996; Roberts and Rosenthal, 1998, 2001).

Three parallel chains were run for each model to determine if the true stationary posterior distribution is obtained. Each simulation consisted of $20,000$ iterations, of which 2000 were considered as burn-in. To reduce the dependence across the samples only every 5th was kept as indicated by the partial auto-correlation function (PACF). Convergence of the *thinned* chains for each parameter was checked using the MCMC convergence criterion,

$$\hat{R} = \frac{\frac{L-1}{L}W + \frac{1}{L}B}{W}, \tag{6.3.13}$$

where L is the number of of iterations in each chain after the burn-in period, W is the averaged variances of the individual chains across all chains and B is the variance of the means of the chains (Gelman and Rubin, 1992). All parameter chains indicate convergence as $\hat{R}$ is below the suggested criterion of 1.2 (Gelman, 1996).

6.3.5 Spatial interpolation

Once we have fitted the model described in Section 6.3.2, we are able to predict quantities of interest at any location, such as the T-year return level. We shall demonstrate the method to predict the scale parameter $\boldsymbol{\phi}$ at a location, say location s^*. Assuming that ϕ_0 is the scale parameter at s^*, the joint distribution of the modelled sites and s^* is given by

$$\begin{pmatrix} \phi_0 \\ \boldsymbol{\phi} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_0 \\ \boldsymbol{\mu} \end{pmatrix}, \begin{bmatrix} \sigma_0 & \Sigma_{0m} \\ \Sigma_{m0} & \Sigma_\phi \end{bmatrix} \right) \tag{6.3.14}$$

where μ_0 and $\boldsymbol{\mu}$ are the means of the distributions at s^* and modelled sites respectively, σ_0 is the variance at s^*, Σ_ϕ is the covariance matrix across the modelled sites, and Σ_{0m} and Σ_{m0} are the covariance matrices between the modelled sites and s^*. Multivariate normal distribution theory provides the result that the conditional distribution of ϕ_0 given $\boldsymbol{\phi}$ is normal with mean

$$\mu_1 = \mu_0 + \Sigma_{0m}\Sigma_\phi^{-1}(\boldsymbol{\phi} - \boldsymbol{\mu}) \tag{6.3.15}$$

and variance

$$\sigma_1 = \sigma_0 - \Sigma_{0m}\Sigma_\phi^{-1}\Sigma_{m0}, \tag{6.3.16}$$

where μ_0 and $\boldsymbol{\mu}$ are given by equation (6.3.7) evaluated at s^* and the modelled sites respectively, Σ_{0m} and Σ_{m0} are given by equation (6.3.8) evaluated between s^* and the modelled sites and Σ_ϕ is given by equation (6.3.8) evaluated across the modelled sites. Once the mean μ_1 and variance σ_1 have been estimated, samples of the scale parameter at s^* can be drawn from the distribution,

$$\phi_0 \mid \boldsymbol{\phi} \sim \text{N}(\mu_1, \sigma_1). \tag{6.3.17}$$

When the shape parameter is modelled by a Gaussian process a similar distribution is formed that can be used to draw samples from. We will use the interpolation method

to obtain parameter and return level estimates at the nine withheld monitoring sites to cross-validate the modelling approach, see Section 6.4.4.

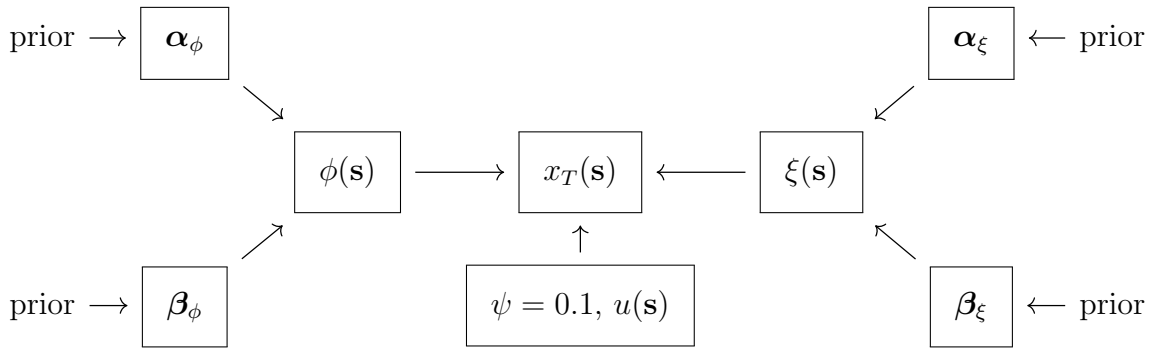


Figure 6.4: Schematic of the model used to estimate the T-year return level map $x_T(\mathbf{s})$, adaptation of Cooley et al. (2007).

Further, to generate an estimate for the return level map across the study space we divide the region into a regular 10km grid and estimate the threshold, scale parameter and shape parameter at each grid location. The distributions of the scale and shape parameters at these grid locations are based on the estimated latent parameters and the covariates. Considering each latent parameter taken from its posterior distribution and the covariate values for each grid location, the mean and covariance are obtained. We take 1000 samples from the conditional distributions, equation (6.3.17), and the average is taken to represent one *posterior draw*. This procedure is repeated for each step of the MCMC chain to obtain estimates of posterior distributions for the scale and shape parameters at the grid location. We obtain the *posterior pointwise mean* of each parameter by taking the average of the posterior draws and the 95% credible interval is obtained by taking the 0.025 quantile and the 0.975 quantile of the posterior draws. To obtain the return level maps and consequently the credible interval of the return level maps we follow the schematic given in Figure 6.4.

6.4 Results

6.4.1 Model selection

In this study, several models are generated based on different covariates, including elevation (Elev) and distance-to-coast (Dist), and are compared using the deviance information criterion (DIC) (Spiegelhalter et al., 2002). The DIC is a simple and robust criterion that is used widely in model selection e.g. Reich (2012) and Sang and Gelfand (2009). The DIC value is the sum of the model fit measure $\bar{D}$ and the model complexity measure p_D. The model fit measure $\bar{D}$ is the posterior expectation of the deviance, where the posterior distribution of the deviance statistic is given by

$$D(\mathbf{X}(\mathbf{s}) \mid \boldsymbol{\theta_1}) = -2 \sum_i \sum_k \log p_1(\mathbf{X}(\mathbf{s}) \mid \boldsymbol{\theta_1}). \tag{6.4.1}$$

The model complexity measure p_D is evaluated as $\bar{D} - D(\mathbf{X}(\mathbf{s}) \mid \bar{\boldsymbol{\theta}}_1)$. As we are running three chains we produce DIC values for each and take the average. This is done to help minimise the randomness in the DIC values since they are a result of posterior samples generated from an MCMC run. A model is considered better if it has a smaller DIC value. Further we do not solely rely on the DIC values to determine the best model, we also consider the return level maps and their uncertainties as well as how they capture the behaviour of ozone at the case study monitoring sites. Table 6.2 shows a selection of the models tested increasing in complexity from the base model, where a common scale parameter and common shape parameter are assumed over the study space. We start by adding in more complex structure to the scale parameter such that $\phi(\mathbf{s})$ is modelled as in Section 6.3.2. This improves the model considerably over the base model. We then model the shape parameter $\xi(\mathbf{s})$ with the structure in Section 6.3.2 (Model 3) and an improved model performance was indicated by the DIC value. We then allow the mean of the scale parameter to be a linear function of covariates: latitude (Lat), longitude (Long), elevation (Elev) and/or distance-to-coast (Dist) (Models $4-10$). An improvement to the DIC value is

observed for when latitude or distance-to-coast are included in the mean of the scale parameter. However, a combination of covariates did not indicate an improvement in model performance. Adding complexity to the mean of the shape parameter did not improve the model. Model 6 is chosen to be the most appropriate model tested based on its DIC value and also the return level maps.

Baseline Model		$\bar{D}$	p_D	DIC
Model 1:	$\boldsymbol{\phi} = \phi$ $\boldsymbol{\xi} = \xi$	49087.02	2.00	49089.02
Spatial Models		$\bar{D}$	p_D	DIC
Model 2:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \epsilon_\phi$ $\boldsymbol{\xi} = \xi$	48866.79	33.25	48900.05
Model 3:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48815.75	48.42	48864.17
Spatial Covariate Models		$\bar{D}$	p_D	DIC
Model 4:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Long}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48817.48	47.38	48864.86
Model 5:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Lat}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48822.75	40.88	48863.64
Model 6:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Dist}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	**48811.84**	**45.06**	**48856.89**
Model 7:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Elev}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48818.43	49.31	48867.74
Model 8:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Dist}) + \alpha_{\phi,2}(\text{Elev}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48811.65	46.56	48858.21
Model 9:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Dist}) + \alpha_{\phi,2}(\text{Lat}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48817.85	39.80	48857.65
Model 10:	$\boldsymbol{\phi} = \alpha_{\phi,0} + \alpha_{\phi,1}(\text{Dist}) + \alpha_{\phi,2}(\text{Long}) + \epsilon_\phi$ $\boldsymbol{\xi} = \alpha_{\xi,0} + \epsilon_\xi$	48817.86	46.14	48864.00

Table 6.2: GPD hierarchical models tested and their corresponding DIC scores. Note: $\epsilon_\cdot \sim \text{MVN}(0, \Sigma_\cdot)$, where $[\Sigma_\cdot]_{i,j} = \beta_{\cdot,0} \exp(-(||\mathbf{s}_i - \mathbf{s}_j||/\beta_{\cdot,1})^{\kappa_\cdot})$.

6.4.2 Modelling parameters

Let us now consider Model 6 in greater detail. Figure 6.5 shows the posterior densities for the parameters $\alpha_{\phi,0}$ and $\alpha_{\phi,1}$. The right-hand panel shows that the

distance-to-coast coefficient is positive with a mean of 0.003 and 95% credible interval of $(0.001, 0.006)$. This indicates the further away from the coast the larger the scale parameter. Although this value is marginally above zero, remember $\phi(\mathbf{s})$ is on the logarithmic scale, once transformed onto the original margins the effects of the covariate distance-to-coast is more pronounced. To provide a demonstration, Aston Hill, located approximately 62km away from the coast, has a scale parameter estimate of 2.720 (2.591, 2.843) whereas Wirral Tranmere, located approximately 3km away from the coast, has an estimate of 2.511 (2.362, 2.672), see Figure 6.6 (top left) for their posterior densities. These are comparable to the maximum likelihood estimates from the individually fitted GPD model, 2.892 (2.652, 3.086) and 2.490 (2.161, 2.737) respectively, and to the individually fitted Bayesian GPD model, 2.887 $(2, 681,$ 3.067) and 2.531 (2.211, 2.750) respectively. Notice the narrower intervals for the estimates obtained by the Bayesian hierarchical model.

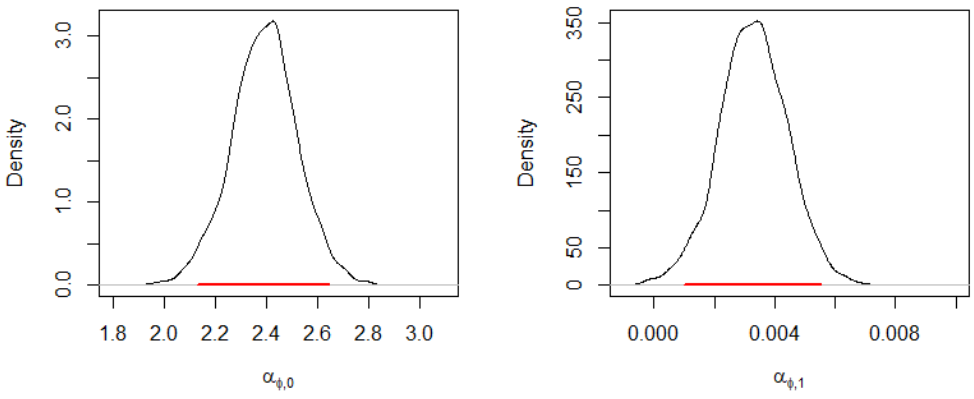


Figure 6.5: The posterior densities for $\alpha_{\phi,0}$ and $\alpha_{\phi,1}$ of Model 6. The red lines shows the 95% credible interval for each parameter.

The posterior densities for the remaining hyperparameters are given in Figure 6.6 with their modelling priors. Using uniform priors for the sill parameters $\beta_{\cdot,0}$ provides a pointwise mean of 0.041 (0.020, 0.075) and 0.013 (0.005, 0.029) with the 95% credible interval in parentheses. The posterior densities for the range parameters $\beta_{\cdot,1}$ have a pointwise mean of 4.541 (1.658, 8.563) and 2.435 (0.946, 4.732) respectively.

Note these values are in degrees where a difference of one degree is approximately 111km. The $\beta_{\cdot,1}$ estimates have a larger credible interval for the scale parameter than the shape parameter. Also the *effective range*, the distance at which 95% of the sill is exceeded, is larger for the scale parameter indicating a stronger dependence relationship with distance between monitoring sites.

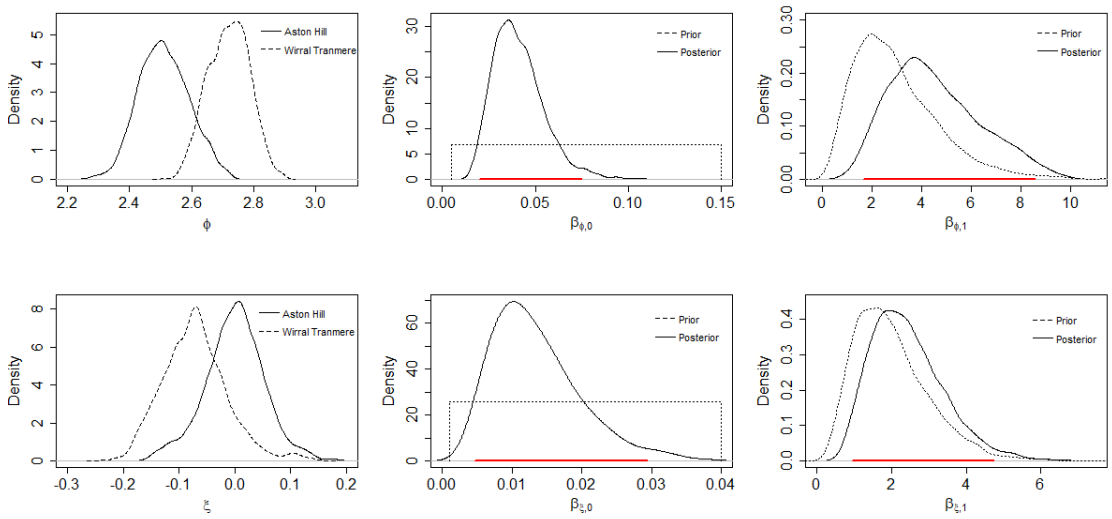


Figure 6.6: Posterior densities for Model 6 parameters. First column shows ϕ and ξ parameters from two sampled sites, Aston Hill $(\cdots)$ and Wirral Tranmere (—) respectively. The middle column shows the posterior density for the sill of the variogram, $\beta_{\cdot,0}$ (—) and its prior $(\cdots)$ and the last column shows the range parameter of the variogram, $\beta_{\cdot,1}$ (—) and its prior $(\cdots)$. The red line shows the 95% credible interval of the posterior distributions.

With the posterior distributions for the latent parameters we can estimate the scale and shape parameters at ungauged locations, as described in Section 6.3.5. Figure 6.7 shows the pointwise mean maps for the scale (σ) and shape (ξ) parameters generated over the study space using a regular 10km grid and the covariate surfaces given in Figure 6.1. The scale parameter map shows the importance distance-to-coast is a spatial covariate since the spatial pattern seen is largely similar to the covariate map, Figure 6.1 (middle). The largest scale estimates are around London and Southeast England and the smallest being in North England and Scotland.

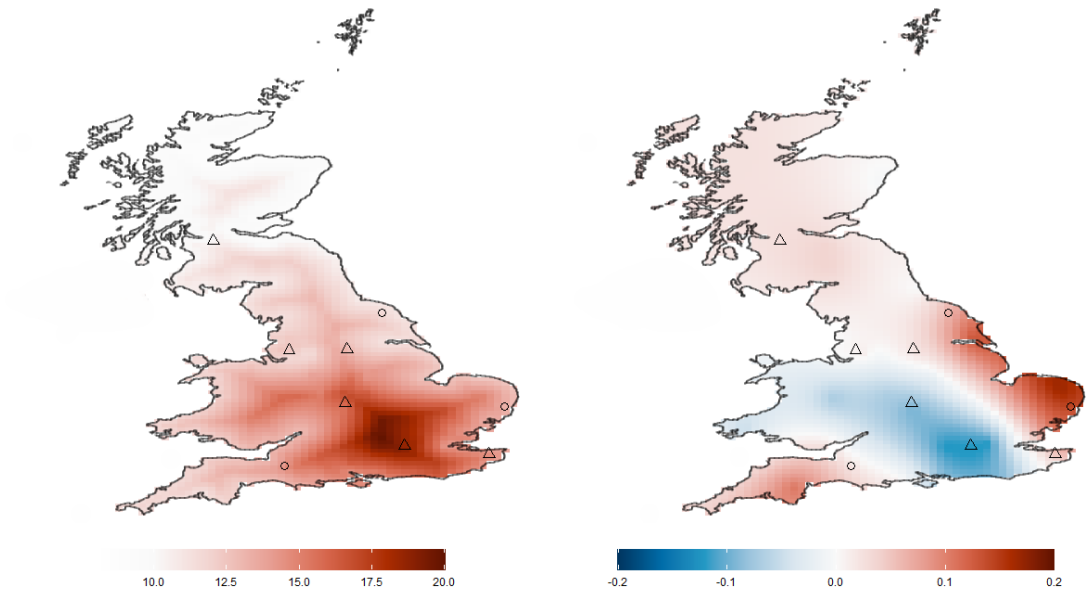


Figure 6.7: Maps showing the pointwise mean estimates for GPD scale parameter σ and GPD shape parameter ξ from Model 6 respectively. The highlighted locations are the validation monitoring sites.

The shape parameter map highlights a clear spatial structure across Great Britain, Figure 6.7 (right). The study region splits into approximately three bands: South-West England having positive shape estimates; Wales, southern parts of the North West, West Midlands, Greater London and Southeast England having negative shape estimates; Scotland, northern parts of the North West, North East, Yorkshire and Humber, East Midlands and East Anglia having positive shape estimates.

6.4.3 Return level maps

We now turn to the T-year return levels of DM8 ozone concentration. After obtaining the posterior distributions for the scale and shape parameters, we can follow the schematic in Figure 6.4 to obtain the one- and ten-year return levels across Great Britain. Figure 6.8(b,f) provide the pointwise mean estimate of the one- and ten-year return level maps with the 0.025 and 0.975 quantiles either side respectively. The map shows the expected one time exceedance in one- and ten-years is lower in Scotland and northern England than the south of England. Figure

6.8(d,h) illustrate the width of the credible interval of the estimate of the one- and ten-year return levels.

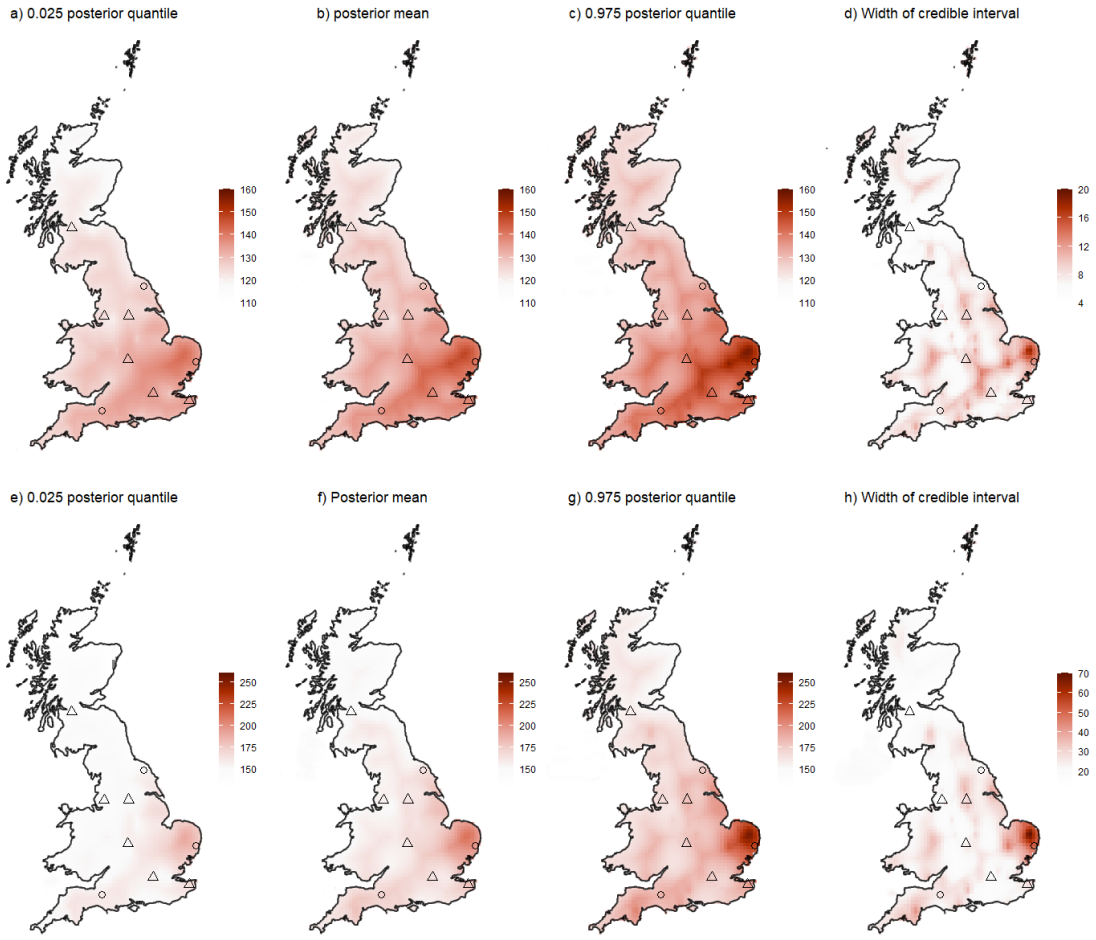


Figure 6.8: Estimate of the (a-d) one-year return levels and the (e-h) ten-year return levels across Great Britain, where (b,f) show the posterior mean estimates, and the corresponding credible intervals are given in (a,c) and (e,g), with the width of the interval given in (d,h). The points on the map indicate the withheld validation monitoring sites and their site type, where (○) points denote rural background sites and (△) points denote urban background sites.

The largest return levels are observed and predicted in East Anglia, with largest concentration from the Bayesian hierarchical modelling being 147.80 $\mu g/m^3$ (141.85, 156.99) compared to the individually fit GPD return levels of 155.67 $\mu g/m^3$ (143.95, 167.40), where the values in parentheses are the 95% credible (confidence) intervals respectively. The lower maximum estimate can be attributed to the smoothed

threshold surface. Similar to the scale parameter map, a vein from East Anglia to approximately Plymouth is observed, which highlights the importance of distance-to-coast as a spatial covariate. The ten-year return levels show similar spatial patterns but indicates a larger difference between East Anglia and the vein. The maximum ten-year return level estimate from the Bayesian hierarchical model is $211.28\,\mu g/m^3$ (190.31, 246.77) compared to the individually fit GPD model of $229.66\,\mu g/m^3$ (144.65, 314.68). Notice the narrower intervals for estimates obtained by the Bayesian hierarchical model. This is due to the sharing of data across the monitoring sites in the data layer of the model. Further, the Bayesian hierarchical model on average indicates smaller uncertainty in the estimates across Scotland and northern England which can be attributed to fewer large extreme events causing a smoothing effect in the estimates, although one may expect greater differences at ungauged sites since there are fewer monitoring sites to interpolate from. The greatest uncertainty is seen in East Anglia, the area where highest extreme levels of ozone are observed and have highest return level estimates, and across the vein.

6.4.4 Model validation

In this study, we first assess the performance of the model by estimating return levels at the nine randomly selected monitoring sites given in Table 6.1 which were withheld during the parameter estimation. The remaining monitoring sites are used to determine the distributions of the latent parameters through the MCMC process given in Section 6.3.4. The 95% confidence intervals for GPD parameters and the one-year return levels are compared to the 95% credible intervals obtained from the Bayesian hierarchical model, Model 6. As seen in Figure 6.9, using the Bayesian hierarchical model dramatically reduces the uncertainty in the estimates since data from different monitoring sites are combined in the data model stage of the modelling procedure. Note the overestimate of the scale parameter at site HIL can be explained by its location, Greater London the region with the highest scale estimates.

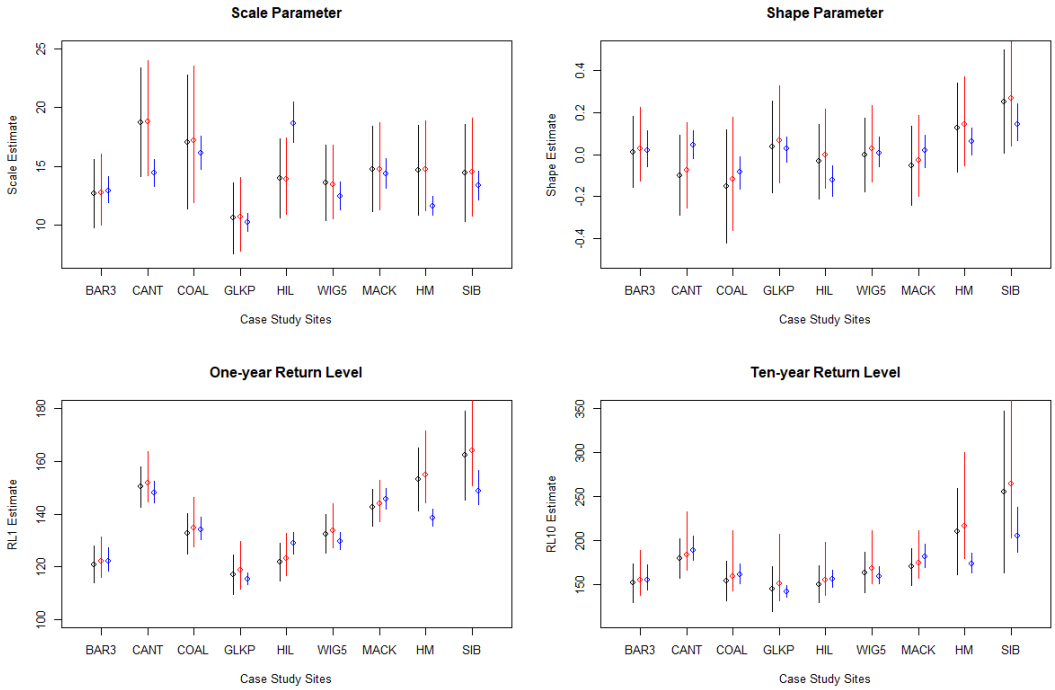


Figure 6.9: Model checking for the latent variable model. Comparing the point estimates of σ, ξ, the one- and ten-year return levels obtained from fitting (black) the GPD distribution and numerically optimising the likelihood, (red) the GPD distribution and obtaining Bayesians estimates and (blue) the Bayesian hierarchical model, each with their corresponding 95% confidence intervals and 95% credible intervals, respectively.

We further validate the Bayesian hierarchical model by using the *posterior predictive distribution* to obtain predictions of ozone concentrations at the nine withheld monitoring sites (Gelman et al., 1996). We do the following at each monitoring site. For a given site i, we observe n_i threshold exceedances. We then randomly draw n_i values for the scale and shape parameters from their posterior distributions. Each scale and shape parameter draw is considered a separate generalised Pareto distribution, which we generated one value from. This produces a vector of predictive threshold exceedances for the site that is the same length as the observed threshold exceedances. After ordering, we plot the results in a scatter plot to demonstrate the predictive skill of the model. In general, we see a good match between the observed and the predictive exceedances. For some of the larger exceedances an overestimation is observed, however this could be due to the random

selection process of a large scale estimate and overly positive shape parameter estimate. The correlation coefficient is 0.961 (0.928, 0.979) and the average difference between the ordered observed exceedances and the ordered predicted exceedances is $-0.812\,\mu g/m^3$ $(-1.643,\ -0.001)$ where the 95% bootstrapped confidence intervals are given in parentheses. This difference indicates a slight underestimation across the nine withheld monitoring sites. We note that Figure 6.10 indicates a lack of fit in the upper portion of the tail, which could be due to the random sampling of the shape parameter from the corresponding posterior distribution.

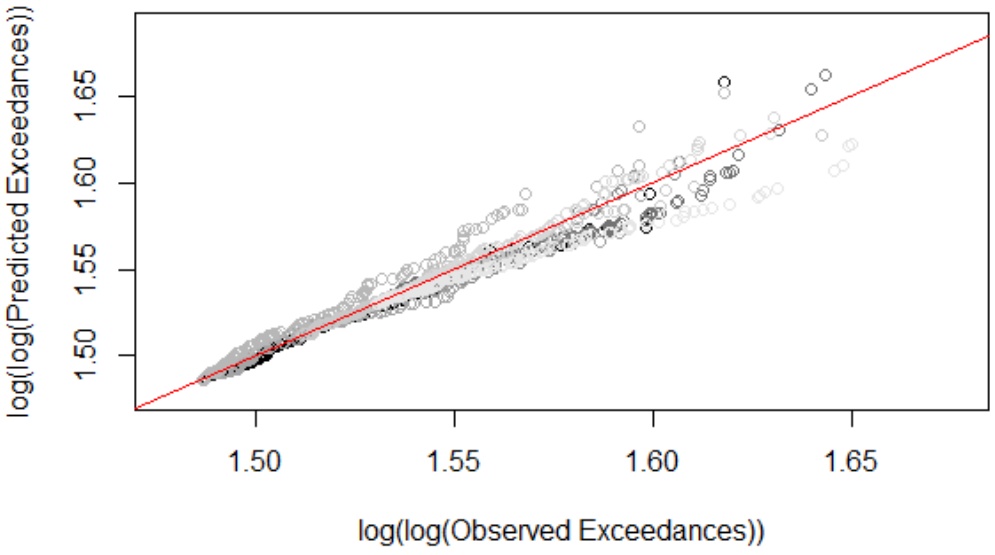


Figure 6.10: Validating the Bayesian hierarchical model by comparing observed exceedances with those predicted by the model at the nine withheld monitoring sites (different grey tone for each site).

6.4.5 Probability maps

To characterise the risk, the probability of exceeding $100\,\mu g/m^3$ (moderate health risk, DAQI 4) of the air quality index employed across the UK is presented in Figure 6.11. Based on the Bayesian hierarchical model fit to the 2011-2019 data, the probability of DM8 ozone exceeding $100\,\mu g/m^3$ on a given day during March to August ranges between 4.05% and 6.93%. The lowest probabilities

occur in Northwest England around Manchester and Liverpool and the north-east coast of the Scottish Lowlands. The largest probabilities occur across East and Southeast England and follow similar spatial pattern to the distance-to-coast covariate. Further, a distinct divide is evident in the probabilities with those above the latitude of approximately 53° having less chance to observe an exceedance of $100\,\mu g/m^3$ than those below, as indicated by the dotted line in Figure 6.11. The areas above have on average a probability of 4.93% whereas those below have an average probability of 6.13%.

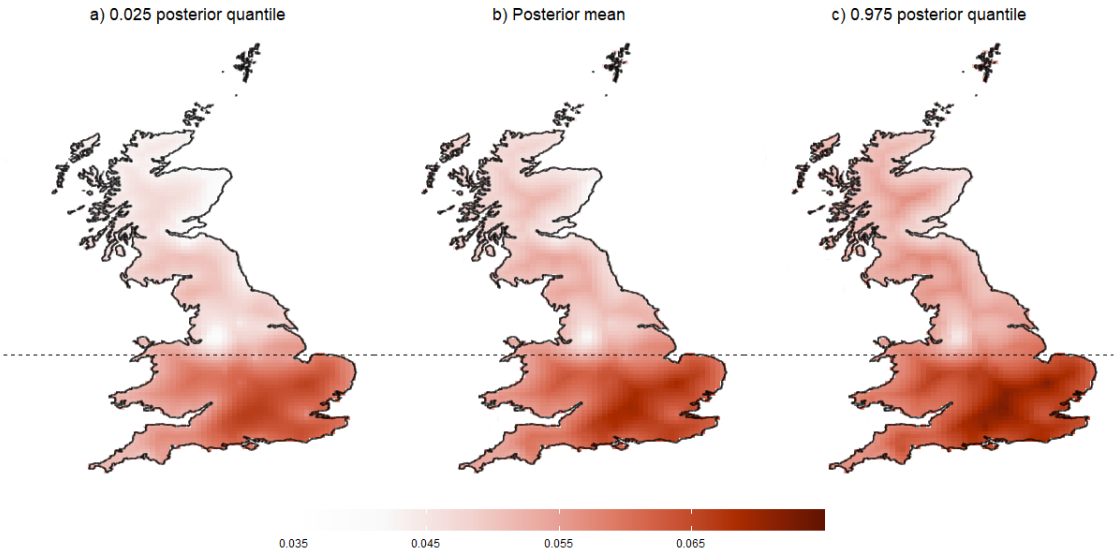


Figure 6.11: Maps showing the probability of exceeding $100\,\mu g/m^3$ on any given day during an ozone period, (a) 0.025 quantile (b) posterior pointwise mean (c) 0.975 quantile. Dotted line indicates latitude of 53°.

6.5 Conclusions

In this study, we have proposed a spatial model of extreme values to evaluate the risk of extreme levels of ozone, an adaptation of the model detailed in Cooley et al. (2007). We assumed that the threshold exceedances of the marginal 90th quantile are independent and identically distributed, allowing us to fix the rate parameter and model the threshold spatially by ordinary kriging. Return level maps are obtained for the generalised Pareto distribution. The estimation of parameters was obtained

through a Bayesian approach allowing us to include a covariance function and model the mean of the parameters as a function of covariates. We tested an array of models and found the most effective model, given the covariates at hand, was Model 6 where both the GPD scale parameter and shape parameter are modelled by a Gaussian process with the mean of the scale parameter being a function of distance-to-coast. The scale parameter and thus the return level is positively correlated with distance-to-coast as indicated by a positive regression coefficient. The map of the estimated shape parameter clearly splits the study space into three distinct regions, see Figure 6.7. The largest one-year return level obtained by Bayesian hierarchical model is approximately $8\,\mu g/m^3$ less than the one-year return level obtained by individually fitted GPD models, however within the 95% confidence intervals.

Model verification is assessed based on evaluating the return level at withheld monitoring sites and drawing from the posterior predictive distribution to obtain predicted threshold exceedances. Using the Bayesian hierarchical model dramatically reduces the uncertainty in the return levels compared to fitting the GPD *at each site individually* due to sharing of observations in the data model stage. Drawing from the posterior predictive distribution generates samples of predicted threshold exceedances that, when ordered, correlate with the ordered observed threshold exceedances. Across the nine withheld monitoring sites the correlation coefficient is 0.961 (0.928, 0.979) and the average forecast error is $-0.812\,\mu g/m^3$ (-1.643, -0.001) where the 95% bootstrapped confidence intervals are given in parentheses.

To provide context to public health we produced maps showing the estimated probability of exceeding $100\,\mu g/m^3$ on a given day during March-August. This probability ranges between 4.05% and 6.93% with the East and Southeast regions having the largest probabilities. Similar spatial structure can be seen in Gouldsbrough et al. (2024), however are less smooth due to the higher resolution and the larger number of covariates used. In their paper, distance-to-coast was found to be the 11th most important feature but is found to be the second most important spatial feature,

second to distance from major roads, (see Figure 3 in Gouldsbrough et al. 2024). As such, an initial extension could involve including more spatial covariates, such as distance to major roads. A further extension would be to expand the model to a spatio-temporal paradigm similar to Garcia et al. (2023). Another option is to implement a non-stationary GPD by including covariates into the marginal scale and shape parameters and modelling each regression coefficient as a spatial Gaussian process. Finally, as DM8 ozone possesses temporal extremal dependence one could extend the modelling approach proposed by Winter and Tawn (2017) by assuming some spatial structure exists on the conditional parameters and the joint and conditional distributions.

Chapter 7

Conclusions and further work

In this concluding chapter, we summarise our contributions to the area of extreme value statistics that result from Chapters 4-6 of this thesis. The research presented advances existing ideas and methodologies in evaluating extremes from process-based forecasts and modelling the extremal behaviour of ozone. We summarise the content of each chapter in Section 7.1, before proposing potential avenues for further research in Section 7.2.

7.1 Summary of contributions

In Chapter 4 we demonstrated several novel applications of extreme value theory in validating numerical process-based forecasts. These applications involved seeing if the tail behaviour of the ozone observations and the tail behaviour of the AQUM forecasts come from the same distribution, through modelling the tails using the generalised Pareto distribution (GPD) and constructing likelihood ratio tests. Our results showed the marginal distributions differed at 69% of the studied sites, indicating improvements are required in forecasting extreme levels of ozone. Further, we modelled the joint distribution of AURN and AQUM using the bivariate extremes value distribution with a range of spectral measures. Selecting the model that minimised the Akaike information criterion (AIC) resulted in 1 asymmetric

logistic model, 7 negative logistic models, 14 Dirichlet distribution models, 19 logistic models and 20 Hüsler-Reiss distribution models, which indicated the joint behaviour favoured the use of symmetric models. These models were then used to probabilistically estimate conditional probabilities of correctly forecasting an exceedance given an exceedance has been observed. Our results indicated that when ozone events become more extreme and pose greater risk to public health, AQUM struggles to accurately capture the magnitude particularly for observed events at least $120\,\mu g/m^3$ (moderate health risk, DAQI 5). The methodology discussed in Chapter 4 could apply to other pollutants and environmental fields where validation of the tail behaviour in forecasts, real-analysis or ensembles is required.

Chapter 5 presented a framework to investigate the stochastic behaviour of the DM8 ozone observations as prolonged episodes of extreme levels of ozone can cause damage to public health through development of respiratory illnesses and potential death. Consequently, understanding the stochastic behaviour of extreme ozone and accounting for such temporal structure is of upmost importance. We used the kth-order Markov extremes model of Winter and Tawn (2017) with a modified kernel to describe the temporal dependence structure. We introduced a new statistic that adapts work of Knutti et al. (2017) by computing weights for each model order based on cluster functional estimates across a range of lags and quantiles. Our results showed that urban background sites on average required more complex models than rural background sites, which coincides with the more varied atmospheric conditions (microclimates by the urban heat island effect) experienced in urban areas (Yang et al., 2016). After obtaining the optimal order model, we then used them in two novel applications: simulating scenarios and short-lead-time forecasts of an out-of-sample event. Our results from the applications showed the models provide accurate forecasts for the year of 2020, however could be improved around the more densely populated areas, like London. Such simulations of scenarios could be used to guide changes in legislation by providing a deeper understanding of the extremal dependence structure of ozone and how extreme ozone events change with time. The

ideas presented in Chapter 5 could be applied to more pollutants, but also to other environmental variables which may exhibit stochastic behaviour.

In Chapter 6 we proposed a spatial Bayesian hierarchical model for threshold exceedances using a generalised Pareto distribution in the data layer, applied to ozone data across Great Britain. These models aim to borrow information across monitoring sites to improve estimation of marginal return levels, therefore reducing uncertainty of these estimates. Minimising the DIC resulted in the model where both the GPD scale and shape parameters are modelled by a Gaussian process with the mean of the scale parameter being a function of distance-to-coast. We assumed that the threshold exceedances of the marginal 90th quantile are independent and identically distributed, allowing us to fix the rate parameter and model the threshold spatially by ordinary kriging. Our resulting surface for the shape parameter showed a clear splitting of the study space, from positive to negative to positive as you move Southwest to Northeast. The resulting return level surfaces highlighted the importance of distance-to-coast as a geographical covariate and showed that East Anglia is the region expected to observe the highest concentrations of ozone. Cross-validation across the nine withheld monitoring sites showed shrinkage in the credible intervals of one- and ten-year return level estimates and the posterior pointwise mean estimates were comparable to the estimates from the individually fitted generalised Pareto distribution.

7.2 Further work

In the following sections we discuss potential interesting avenues for further research which would extend the ideas and methodologies explored in this thesis. In Section 7.2.1 we present ideas that further develops our novel approach for evaluating and valdating the extreme forecasts from process-based models, which was introduced in Chapter 4. Then in Section 7.2.2 we discuss ways to extend our work from Chapter 6 regarding temporal dependence modelling using the kth-order Markov model.

Lastly, in Section 7.2.3 we suggest ways to extend the spatial Bayesian hierarchical model to account for further covariate effects and the temporal dependence structure captured in Chapter 5.

7.2.1 Evaluation of ozone forecasts in Chapter 4

A key part of our methodologies in Chapter 4 was validating the AQUM forecasts by investigating the marginal and joint tail behaviours of the AURN observations and the AQUM forecasts. Although we aggregated to the daily maximum of the 8-hour running mean (DM8) scale, slight temporal dependence existed. It would be insightful to see how accounting for the temporal dependence affects the confidence intervals of the GPD parameter estimates and the resulting return levels. To account for any temporal dependence, the non-stationary GPD of Smith (1989) and Davison and Smith (1990) could be utilised. The non-stationary GPD allows the constant parameters to be parametrised as functions of covariates, where seasonal variability is often accounted for by using harmonics (Coles et al., 1994). It would be interesting to see how the inclusion of temperature, time of day, time of year, wind speed and wind direction in the GPD parameters would alter the modelling parameter estimates and the return levels.

Further, using the bivariate extreme value distribution to model the joint behaviour of extreme ozone assumes asymptotic dependence, where this might not be the case at some monitoring sites or for other pollutants/environmental applications. The conditional extremes model of Heffernan and Tawn (2004) can capture both asymptotic dependence and asymptotic independence. As such, an interesting avenue would be to fit the conditional extremes model to the pair (AURN, AQUM) to examine the dependence structure more deeply. This model could also be used to test whether the temporal dependence of the observations is captured in the forecasts by fitting the model to the pair (AURN, AURN_T) and (AQUM, AQUM_T), where subscript T denotes the T-lagged version of the data, and contrast the resulting parameter estimates.

An impactful extension of the validation process presented in Chapter 4 would be to turn this offline validation tool into an on-line operational validation tool. This could allow for the existing post-processing technique, as described in Chapter 2, to be updated to have two separate regimes. This could be where the first regime continues to correct non-exceedances with the current SPPO method, and when an exceedance has been observed or forecasted by AQUM then the second regime would kick in where the joint extremes model would be used to validate the forecasts of the extreme event. The method of modelling of the joint behaviour allows for probabilistic estimates to be obtained, which could be used to provide probabilities of observing an exceedance on a given day. This information would be invaluable to present to the public as it provides more information about the upcoming air quality, rather than just the Daily Air Quality Index.

7.2.2 Temporal behaviour modelling in Chapter 5

Our simulation procedure for the clusters of extreme ozone using the kth-order Markov extremes model, in Chapter 6, demonstrated promising results for replicating the temporal dependence structure of the observations. While our simulations of out-of-sample extreme events matched the observations well (see Section 5.5 of Chapter 5), further work could investigate if improvements could be made. One option is to pool information across sites that share similar site type and similar temporal dependence characteristics. Another option could be, rather than aggregating to the DM8 to account for the diurnal cycle, we could have explored the hourly observations and redefine when a cluster ends i.e. using a run length of at least 12 hours to account for the diurnal cycle. Further, before using the probability integral transformation to transform onto Laplace margins, we could have parametrised the GPD parameters to be functions of time using Fourier Series or inclusion of other covariates. Tendijck et al. (2021, 2023) has started the extension by modelling multivariate time series with a multivariate kth-order Markov extremal model, although this is on an oceanographic dataset.

We presented the use of forward simulation to generate the chains with kth-order dependence structure from an initial exceedance. However, extreme concentrations of ozone may have a building up period due to certain atmospheric conditions. Therefore, it would be interesting to see if using backward simulation from the first exceedance to generate a couple observations and then simulate forward these chains improves the cluster functionals estimates, given the backward simulation generates estimates approximately the same as the observations.

It could be interesting to apply the methodology in Chapter 5 to the AQUM data explored in Chapter 4 and combine the findings. This would add another element to the validation process of extremes from a process-based forecast model. The combined findings could allow for extensions into an operational setting, where it would be used to generate forecasts for the remainder of an extreme event. We would need to dynamically update the dependence parameters and model order as more observations become available. This could further improve the quality of information about the air quality for future days available to the public, allowing individuals who are more susceptible to high concentrations of air pollution to take the appropriate action.

7.2.3 Bayesian hierarchical modelling in Chapter 6

Our motivation driving the research presented in Chapter 6 was the apparent spatial patterns in the GPD parameter estimates, return levels, dependence parameter estimates and the model order of the kth-order extremal Markov models we observed in Chapter 4 and Chapter 5. As such, a spatial Bayesian hierarchical model for threshold exceedances is presented in Chapter 6 to allow for interpolation to ungauged locations. We fixed the rate of exceedance and modelled the threshold using the ordinary kriging method, which created a very smooth surface. An improvement to our proposed model would be to fix the threshold to say $100\,\mu g/m^3$ (DAQI of 4, moderate health risk) and model the rate of exceedance spatially.

MCMC techniques were used to interpolate our spatial Bayesian hierarchical model

to ungauged locations and generate estimates for return levels and probabilities of observing a moderate health risk or worse on a given day in an ozone period. Our estimated surfaces show similar spatial structure as that of Gouldsbrough et al. (2024), however are less smooth due to the higher resolution and the larger number of covariates used. In their paper, distance-to-coast was found to be the 11th most important feature but is found to be the second most important spatial feature, second to distance from major roads, (see Figure 3 in Gouldsbrough et al. 2024). Therefore, it would be interesting to include distance to major roads as a covariate and contrast the results.

As seen in Chapter 5, ozone possesses temporal dependence and the conditional extremes model captures this structure well. A potential avenue of further work would be to assume the parameters of the conditional extremes model has some underlying spatial structure. This could be done by following the work of Wadsworth and Tawn (2019) where the model is constructed by conditioning on threshold exceedances at a single location and composite likelihood approach used in the inference, allowing for contributions from different conditioning sites. An extension could be to use Simpson and Wadsworth (2021), which is an spatio-temporal extension of the work by Wadsworth and Tawn (2019), or the spatio-temporal framework of Garcia et al. (2023).

Appendix A

AURN site information

	Site Codes	Type	Latitude	Longitude	OS Start	OS End
1	AH	RB	52.50385	-3.03418	March	July
2	ACTH	RB	55.79216	-3.24290	March	July
3	BUSH	RB	55.86228	-3.20578	March	May
4	MACK	RB	51.05625	-2.68345	March	July
5	CHBO	RB	51.14962	-1.43823	April	August
6	ESK	RB	55.31531	-3.20611	March	July
7	GLAZ	RB	53.46008	-2.47206	March	July
8	HM	RB	54.33494	-0.80855	March	August
9	LB	RB	53.40337	-1.75201	March	July
10	LERW	RB	60.13922	-1.18532	March	July
11	LH	RB	50.79370	0.18125	March	September
12	PEMB	RB	51.78178	-4.69146	March	July
13	ROCH	RB	51.45617	0.63489	April	August
14	SIB	RB	52.29440	1.46350	March	August
15	OSY	RB	51.77798	1.04903	March	August
16	SV	RB	57.73446	-4.77658	March	May
17	WEYB	RB	52.95049	1.12202	March	August
18	WFEN	RB	52.29850	0.29092	April	August

19	YW	RB	50.5976	-3.71651	March	July
20	ABD	UB	57.15736	-2.09428	March	June
21	BAR3	UB	53.56292	-1.51044	March	July
22	AGRN	UB	52.43716	-1.83000	March	August
23	BLC2	UB	53.80489	-3.00718	March	July
24	BORN	UB	50.73957	-1.82674	April	August
25	BRT3	UB	50.84084	-0.14757	April	September
26	BRS8	UB	51.46284	-2.58448	March	July
27	CANT	UB	51.27399	1.09806	March	August
28	CARD	UB	51.48178	-3.17625	April	July
29	COAL	UB	52.41156	-1.56023	April	September
30	ED3	UB	55.94559	-3.18219	March	May
31	GLKP	UB	55.86578	-4.24363	March	June
32	HUL2	UB	53.74878	-0.34122	March	August
33	LEAM	UB	52.28881	-1.53312	March	August
34	LEED	UB	53.80378	-1.54647	March	July
35	LECU	UB	52.61982	-1.12731	April	August
36	CLL2	UB	51.52229	-0.12589	April	August
37	HG4	UB	51.58413	-0.12525	March	August
38	HIL	UB	51.49633	-0.46086	March	August
39	KC1	UB	51.52105	-0.21349	April	August
40	MAN3	UB	53.48152	-2.23788	March	July
41	NEWC	UB	54.97825	-1.61053	March	July
42	NO12	UB	52.61419	1.30198	March	August
43	NOTT	UB	52.95473	-1.14645	March	August
44	PEEB	UB	55.65747	-3.19653	March	June
45	PLYM	UB	50.37167	-4.14236	March	July
46	PMTH	UB	50.82881	-1.06858	March	September
47	PRES	UB	53.76559	-2.68035	March	July

48	REA1	UB	51.45309	-0.94407	April	August
49	SHDG	UB	53.37862	-1.47810	March	August
50	SOUT	UB	50.90814	-1.39578	March	August
51	SEND	UB	51.54421	0.67841	April	August
52	STOK	UB	53.02821	-2.17513	March	July
53	SUN2	UB	54.88361	-1.40688	March	July
54	THUR	UB	51.47707	0.31797	April	August
55	WAL4	UB	52.60562	-2.03052	March	July
56	WIG5	UB	53.54914	-2.63814	March	July
57	TRAN	UB	53.37287	-3.02272	March	July

Table A.1: AURN site information for sites used in the chapter analysis. OS denotes *ozone season*. The site types are split into urban background (UB), rural background (RB).

Appendix B

Case study sites

Code	Lat.	Long.	u_{90}	$\hat{\sigma}_{u_{90}}$	$\hat{\xi}$
BAR3	53.56292	-1.51044	84.927	12.612 (1.608)	-0.001 (0.092)
CANT	51.27399	1.09806	102.737	18.729 (2.369)	-0.098 (0.097)
COAL	52.41156	-1.56023	92.727	15.815 (2.765)	-0.106 (0.142)
GLKP	55.86578	-4.24363	87.041	8.187 (1.668)	0.188 (0.168)
HIL	51.49633	-0.46086	82.921	13.953 (1.720)	-0.032 (0.090)
WIG5	53.54914	-2.63814	94.528	14.719 (1.863)	-0.040 (0.089)
MACK	51.05625	-2.68345	103.652	16.069 (2.196)	-0.093 (0.102)
HM	54.33494	-0.80855	101.376	14.649 (1.968)	0.127 (0.108)
SIB	52.29440	1.46350	100.299	14.421 (2.112)	0.250 (0.125)

Table B.1: Information of the nine case study sites: site code, the latitude and longitude, marginal threshold and generalised Pareto distribution parameter estimates. Values in parentheses are the associated standard errors.

Bibliography

Andreae, M. O., Jones, C. D., and Cox, P. M. (2005). Strong present-day aerosol cooling implies a hot future. *Nature*, 435(7046):1187–1190.

Arneth, A., Harrison, S. P., Zaehle, S., Tsigaridis, K., Menon, S., Bartlein, P. J., Feichter, J., Korhola, A., Kulmala, M., O'Donnell, D., Schurgers, G., Sorvari, S., and Vesala, T. (2010). Terrestrial biogeochemical feedbacks in the climate system. *Nature Geoscience*, 3:525–532.

Arneth, A., Unger, N., Kulmala, M., and Andreae, M. O. (2009). Perspectives: Clean the air, heat the planet? *Science*, 326:672–673.

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC Press.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Methodological)*, 70(4):825–848.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–355.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugals, J. (2004). *Statistics of extremes: Theory and Applications*. John Wiley and Sons Ltd, Chichester, UK.

Bell, M. L., Dominici, F., and Samet, J. M. (2005). A meta-analysis of time-series

studies of ozone and mortality with comparison to the national morbidity, mortality and air pollution study. *Epidemiology*, 16:436–445.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M., and Dominici, F. (2004). Ozone and short-term mortality in 95 US urban communities 1987 − 2000. *Journal of the American Medical Association (JAMA)*, 292(19):2372–2378.

Berger, J., DeOiveira, J., and Sanso, B. (2001). Objective Bayesian Analysis of Spatially Correlated Data. *Journal of the American Statistical Association*, 96:1361–1374.

Borrego, C., Monteiro, A., Ferreira, J., Miranda, A. I., Costa, A. M., Carvalho, A. C., and Lopes, M. (2008). Procedures for estimation of modelling uncertainty in air quality assessment. *Environment International*, 34:613–620.

Brimblecombe, P. and Grossi, C. M. (2007). Damage to Buildings from Future Climate and Pollution. *The Journal of Preservation Technology*, 38(2/3):13–18.

Casson, E. and Coles, S. G. (1999). Spatial regression models for extremes. *Extremes*, 1(4):449–468.

CE Delft (2018). Health impacts and costs of diesel emissions in the EU.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*. CRC Press.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):207–222.

Clancy, C., O'Sullivan, J., Sweeney, C., Dias, F., and Parnell, A. (2016). Spatial Bayesian hierarchical modelling of extreme sea states. *Ocean Modelling*, 107:1–13.

Cohen, A. J., Burnett, M. B. R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska,

L., Pope 3rd, C. A., Shin, H., Straif, K., Shaddick, G., van Dingenen, M. T. R., van Donkelaar, A., Vos, T., Murray, C. J. L., and Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*, 389(10082):1907–1918.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Coles, S. G. and Casson, E. (1998). Extreme value modelling of hurricane wind speeds. *Structural Safety*, 20(3):283–296.

Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical Methods for multivariate Extremes: an Application to Structural Design. *Applied Statistics*, 43:1–48.

Coles, S. G., Tawn, J. A., and Smith, R. L. (1994). A seasonal Markov model for extremely low temperatures. *Environmentrics*, 5(3):221–239.

COMEAP (2011). Review of the UK Air Quality Index - A report by the Committee on the Medical Effects of Air Pollutants. Technical report, COMEAP. `https://webarchive.nationalarchives.gov.uk/ukgwa/20140714074352/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317137023144`.

Connolly, E., Fuller, G., Baker, T., and Willis, P. (2013). Update of Implementation of the Daily Air Quality Index: Information for Data Providers and Publishers. `http://uk-air.defra.gov.uk/reports/cat14/1304251155_Update_on_Implementation_of_the_DAQI_April_2013_Final.pdf`.

Contzen, J., Dickhaus, T., and Lohmann, G. (2022). Variability and extremes: statistical validation of the Alfred Wegener Institute Earth System Model (AWI-ESM). *Geoscientific Model Development*, 15:1803–1820.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cooper, O. R., Parrish, D. D., Ziemke, J., Balashov, N. V., Cupeiro, M., Galbally, I. E., Gilge, S., Horowitz, L., Jensen, N. R., Lamarque, J.-F., Naik, V., Oltmans, S. J., Schwab, J., Shindell, D. T., Thompson, A. M., Thouret, V., Wang, Y., and Zbinden, R. M. (2014). Global distribution and trends of tropospheric ozone: An observation-based review. *Elementa: Science of the Anthropocene*, 2.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

Davison, A. C. and Smith, R. L. (1990). Models for Exceedances over High Thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 52:393–442.

de Haan, L. and Ferreira, A. (2007). *Extreme value theory : an introduction*. Springer Science & Business Media.

Defra (accessed 2020). `http://www.airquality.co.uk/archive/data_and_statistics.php`.

DEFRA and UK-AIR (accessed 2022). Daily Air Quality Index. online. `https://uk-air.defra.gov.uk/air-pollution/daqi?view=more-info`.

Denby, B., Horalek, J., Walker, S. E., Eben, K., and Fiala, J. (2005). Interpolation and Assimilation Methods for European Scale Air Quality Assessment and

Mapping. https://acm.eionet.europa.eu/docs/ETCACC_TechnPaper_2005_7_SpatAQ_Interpol_Part_I.pdf Accessed 22/09/2022.

Devlin, R. B., Raub, J. A., and Folinsbee, L. J. (1997). Health effects of ozone. *SciMedicine*, pages 8–17.

Dingenen, R. V., Dentener, F. J., Raes, F., Krol, M. C., Emberson, L., and Cofala, J. (2009). The global impact of ozone on agricultural crop yields under current and future air quality legislation. *Atmospheric Environment*, 43(3):604–618.

Doob, J. L. (1935). The Limiting Distributions of Certain Statistics. *Annals of Mathematical Statistics*, 6(3):160–169.

Düsterhus, A. and Hense, A. (2012). Advanced information criterion for environmental data quality assurance. *Advances in Science and Research*, 8:99–104.

EEA and WHO (2007). Air pollution in Europe 1990-2004, EEA Report No 2/2007.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26.

Eyring, V., Arblaster, J. M., Cionni, I., Sedláček, J., Perlwitz, J., Young, P. J., Bekki, S., Bergmann, D., Cameron-Smith, P., Collins, W. J., Faluvegi, G., Gottschaldt, K.-D., Horowitz, L. W., Kinnison, D. E., Lamarque, J.-F., Marsh, D. R., Saint-Martin, D., Shindell, D. T., Sudo, K., Szopa, S., and Watanabe, S. (2013). Long-term ozone changes and associated climate impacts in CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10):5029–5060.

Fawcett, L. and Walshaw, D. (2006). Markov chain models for extreme wind speeds. *Envionmetrics*, 17(8):795–809.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, R., Prinn, R., Raga, G., Schulz, M., and van Dorland, R. (2007). *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Fougères, A.-L. (2004). *Extreme values in finance, telecommunications and the environment*, chapter Multivariate Extremes, pages 373–388. Chapmen & Hall/CRC, Boca Raton.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70:674–680.

Garcia, J. A., Acero, F. J., and Portero, J. (2023). A Bayesian hierarchical spatio-temporal model for extreme temperatures in Extremadura (Spain) simulated by a Regional Climate Model. *Climate Dynamics*, 61:1489–1503.

GCNTO (accessed 2022). Global Challenge Network on Tropospheric Ozone - Ecosystem effects on ozone. `http://www.ozone-net.org.uk/sites/ozone-net.org.uk/files/documents/filedepot/4/GNC_OzoneFactSheets_EcosystemEffects.pdf`.

Gelman, A. (1996). *Inference and Monitoring Convergence*. London: Chapman & Hall.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis Jumping Rules. *Bayesian Statistics*, 5:599–607.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.

Geyer, C. J. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press.

Gouldsbrough, L., Hossaini, R., Eastoe, E., Young, P. J., and Vieno, M. (2024). A machine learning approach to downscale EMEP4UK: analysis of UK ozone variability and trends. *Atmospheric Chemistry and Physics*, 24:3163–3196.

Guicherit, R. and van Dop, H. (1977). Photochemical production in western Europe (1971-1975) and its relation to meteorology. *Atmospheric Environment*, 11:145–156.

Gumbel, E. J. (1958). *Statistics of Extremes*. New York Chichester, West Sussex: Columbia University Press.

Gumbel, E. J. (1960). Distributions des valeurs extremes en plusieurs dimensions. *Publications de l'Institut de statistique de l'Université de Paris*, 9:171–173.

Handcock, M. S. and Stein, M. L. (1993). A Bayesian Analysis of Kriging. *Technometrics*, 35(4):403–410.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society: Series B (Methodological)*, 66(3):497–546.

Holland, M., Kinghorn, S., Emberson, L., Cinderby, S., Ashmore, M., Mills, G., and Harmens, H. (2006). Development of a framework for a probabilistic assessment of the economic losses caused by ozone damage to crops in Europe. Centre for Ecology and Hydrology, Bangor.

Hou, P. and Wu, S. (2016). Long-term Changes in Extreme Air Pollution Meteorology and the Implications for Air Quality. *Scientific Reports*, 6(23792).

Huangfu, P. and Atkinson, P. (2020). Long-term exposure to NO_2 and O_3 and all-cause and respiratory mortality: A systemic review and meta-analysis. *Environment International*, 144(105998).

Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatiotemporal model for Mexico City ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):231–248.

Huser, R. and Wadsworth, J. (2020). Advances in statistical modeling of spatial extremes. *WIREs Computational Statistics*, 14(e1537).

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters.*, 9:283–286.

Ji, M., Cohan, D. S., and Bell, M. L. (2011). Meta-analysis of the association between short-term exposure to ambient ozone and respiratory hospital admissions. *Environmental Research Letters*, 6(024006).

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, 9:75–81.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall/CRC.

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A. M., Helber, R., and Arnone, R. A. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, 76(1-2):64–82.

Kang, D., Eder, B. K., Stein, A. F., Grell, G. A., Peckham, S. E., and McHenry, J. (2005). The new england air quality forecasting pilot program: Development of an evaluation protocol and performance benchmark. *Journal of the Air & Waste Management Association*, 55:1782–1796.

Kang, D., Mathur, R., Rao, S. T., and Shaocai, Y. (2008). Bias adjustment techniques for improving ozone air quality forecasts. *Journal of Geophysical Research: Atmospheres*, 113(D23).

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Kitabo, C. A. (2020). Bayesian Spatial and Trend Analysis on Extreme Data in South Korea: 1991-2015. *Advances in Meteorology*, 2020(8839455).

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, 44(4):1909–1918.

Kukkonen, J., Olsson, T., Schultz, D. M., Baklanov, A., Klein, T., Miranda, A. I., Monteiro, A., Hirtl, M., Tarvainen, V., Boy, M., Peuch, V.-H., Poupkou, A., Kioutsioukis, I., Finardi, S., Sofiev, M., Sokhi, R., Lehtinen, K. E. J., Karatzas, K., San José, R., Astitha, M., Kallos, G., Schaap, M., Reimer, E., Jakobs, H., and Eben, K. (2012). A review of operational, regional-scale, chemical weather forecasting models in europe. *Atmospheric Chemistry and Physics*, 12(1):1–87.

LAEI (accessed 2024). `https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2019`.

Lavigne, E., Gasparrini, A., Wang, X., Chen, H., Yagouti, A., Fleury, M. D., and Cakmak, S. (2014). Extreme ambient temperatures and cardiorespiratory emergency room visits: assessing risk by comorbid health conditions in a time series study. *Environmental Health*, 13(5).

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer-Verlag, New-York.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 59(2):475–499.

Ledford, A. W. and Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 65(2):521–543.

Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., and Münzel, T. (2019). Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *European Heart Journal*, 40(20):1590–1596.

Lin, C., Heal, M. R., Vieno, M., MacKenzie, I. A., Armstrong, B. G., Butland, B. K., Milojevic, A., Chalabi, Z., Atkinson, R. W., Stevenson, D. S., Doherty, R. M., and Wilkinson, P. (2017). Spatiotemporal evaluation of EMEP4UK-WRF v4.3 atmospheric chemistry transport simulations of health-related metrics for NO_2, O_3, PM_{10}, and $PM_{2.5}$ for 2001–2010. *Geoscientific Model Development*, 10:1767–1787.

Liu, J. and West, M. (2001). *Combined Parameter and State Estimation in Simulation-Based Filtering*, pages 197–223. Springer New York.

Loh, W. (2005). Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *The annals of Statistics*, 33(5):2344–2394.

Matérn, B. (1960). *Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surverys and Other Sampling Investigations*, volume 49. Medd. Statens Skogsforsknings institut, Stockholm.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8):1246–1266.

Menezes, R., Piairo, H., Garcia-Soidan, P., and Sousa, I. (2016). Spatial-temporal modellization of the NO_2 concentration data through geostatistical tools. *Statistical Methods and Applications*, 25:107–124.

Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., and Cox, P. M. (2009). Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, 458(7241):1014–1017.

Mukhopadhyay, S. and Sahu, S. K. (2018). A Bayesian spatiotemporal model to estimate long-term exposure to outdoor air pollution at coarser administrative geographies in England and Wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(2):465–486.

Murray, C. J. L., A. Y. Aravkin, P. Z., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., et al. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet*, 396(10258):1223–1249.

Neal, L. S., Agnew, P., Moseley, S., Ordonez, C., Savage, N. H., and Tilbee, M. (2014). Application of a statistical post-processing technique to a gridded, operational, air quality forecast. *Atmospheric Environment*, 98:385–393.

Oehlert, G. W. (1992). A Note on the Delta Method. *The American Statistician*, 46(1):27–29.

Oesting, M., Schlather, M., and Friederichs, P. (2016). Statistical post-processing of forecasts for extremes using bivariate brown-resnick processes with an application to wind gusts. *Extremes*, 20:309–332.

Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Papastathopoulos, I., Strokorb, K., Tawn, J. A., and Butler, A. (2017). Extreme events of Markov chains. *Advances in Applied Probability*, 49(1):134–161.

Papastathopoulos, I. and Tawn, J. A. (2013). Graphical structures in extreme multivariate events. In *Proceedings of 25th Panhellenic Statistics Conference*, pages 315–323.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131.

Pickands, J. (1981). Multivariate extreme value distribution. In *Proceedings 43th, Session of International Statistical Institution*, pages 859–878.

Pleijel, H., Danielsson, H., Simpson, D., and Mills, G. (2014). Have ozone effects on carbon sequestration been overestimated? A new biomass response function for wheat. *Biogeosciences*, 11:4521–4528.

Raes, F., Liao, H., Chen, W. T., and Seinfeld, J. H. (2010). Atmospheric chemistry-climate feedbacks. *Journal of Geophysical Research*, 115(D12).

Reich, B. J. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal statistical Society: Series C (Applied Statistics)*, 61(4):535–553.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Science & Business Media, Berlin, Germany.

Ribatet, M., Cooley, D. S., and Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22:813–845.

Ribatet, M., Ouarda, T., Sauquet, E., and Gresillon, J.-M. (2009). Modeling all exceedances above a threshold using an extremal dependence structure: Inferences on several flood characteristics. *Water Resources Research*, 45(3).

Richards, J., Tawn, J. A., and Brown, S. (2023). Joint estimation of extreme spatially aggregated precipitation at different scales through mixture modelling. *Spatial Statistics*, 53(100725).

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

Roberts, G. O. and Rosenthal, J. S. (1998). Markov-Chain Monte Carlo: Some Practical Implications of Theoretical Results. *The Canadian Journal of Statistics*, 26(1):5–20.

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings Algorithms. *Statistical Science*, 16(4):351–367.

Rootzén, H. (1988). Maxima and exceedances of stationary markov chains. *Advances in Applied Probability*, 20:371–390.

Russell, B. T., Cooley, D. S., Porter, W. C., and Heald, C. L. (2016). Modeling the spatial behaviour of the meteorological drivers' effects on extreme ozone. *Environmetrics*, 27:334–344.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.

SAQD (accessed 2024). `https://www.scottishairquality.scot/data`.

Savage, N. H., Agnew, P., Davis, L. S., Ordonez, C., Thorpe, R., Johnson, C. E., O'Connor, F. M., and Dalvi, M. (2013). Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation. *Geoscientific Model Development*, 6:353–372.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal*, 10(1):33–60.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley.

Sharkey, P. and Winter, H. C. (2019). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *Environmetrics*, 30(1).

Shooter, R., Tawn, J. A., Ross, E., and Jonathan, P. (2021). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24:241–265.

Simpson, D., Arneth, A., Mills, G., Solberg, S., and Uddling, J. (2014). Ozone - the persistent menace; interactions with the N-cycle and climate change. *Current Operations in Environmental Sustainability*, 9-10:9–19.

Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for red sea surface temperatures. *Spatial Statistics*, 41:100482.

Sitch, S., Cox, P. M., Collins, W. J., and Huntingford, C. (2007). Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature*, 448(7155):791–794.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, pages 7267–7292.

Smith, R. L. (1989). Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. *Statistical Science*, 4(4):367–377.

Smith, R. L. (1990a). *Extreme Value Theory. Handbook of Applicable Mathematics.* W. Ledermann. John Wiley, Chichester.

Smith, R. L. (1990b). *Max-stable Processes and Spatial Extremes.* University of Surrey, Guildford, UK.

Smith, R. L. (1992). The extremal index for a markov chain. *Journal of Applied Probability*, 29(1):37–45.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):515–528.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Methodological)*, 64:583–639.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer Science & Business Media.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–147.

Tancredi, A., Anderson, C. W., and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9:87–106.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 20:397–415.

Tawn, J. A. (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika*, 77(2):245–253.

Tendijck, S., Eastoe, E., Tawn, J. A., Randell, D., and Jonathan, P. (2021). Modeling the Extremes of Bivariate Mixture Distributions With Application

to Oceanographic Data. *Journal of the American Statistical Association*, 118(542):1373–1384.

Tendijck, S., Jonathan, P., Randell, D., and Tawn, J. A. (2023). Temporal evolution of the extreme excursions of multivariate kth order markov processes with application to oceanographic data. *Environmetrics*, 35(3):e2834.

The Royal Society (2008). Ground-level ozone in the 21st century: future trends, impacts and policy implications.

Thunis, P., Pederzoli, A., and Pernigotti, D. (2012). Performance criteria to evaluate air quality modeling applications. *Atmospheric Environment*, 59:476–482.

Thunis P, Pernigotti D, G. M. (2013). Model quality objectives based on measurement uncertainty. Part I: Ozone. *Atmospheric Environment*, 79(JRC77727):861–868.

Tippett, L. H. C. and Fisher, R. A. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Vettori, S., Huser, R., and Genton, M. G. (2019). Bayesian modelling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, 75(3):831–841.

Vohra, K., Vodonos, A., Schwartz, J., Marais, E. A., Sulprizio, M. P., and Mickley, L. J. (2021). Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem. *Environmental Research*, 195.

Vukovich, F. M., Bach, W. D., Crissman, B. W., and King, W. J. (1977). On the relationship between high ozone in the rural surface layer and high pressure systems. *Atmospheric Environment*, 11:967–983.

Wadsworth, J. and Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51(100677).

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.

Wadsworth, J. L. and Tawn, J. A. (2019). Higher-dimensional spatial extremes via single-site conditioning. arXiv:1912.06560.

Weller, G. B., Cooley, D. S., and Sain, S. R. (2012). An investigation of the pineapple express phenomenon via bivariate extreme value theory. *Environmetrics*, 23(5):420–439.

Whitfield, R. G., Biller, W. F., Jusko, M. J., and Keisler, J. M. (1996). A Probabilistic Assessment of Health Risks Associated with Short-term Exposure to Tropospheric Ozone. Report prepared for U.S. EPA, OAQPS. Argonne National Laboratory; Argonne, IL.

WHO (2003). Health Aspects of Air Pollution with Particulate matter, Ozone and Nitrogen Dioxide.

WHO (2013). Review of evidence on health aspects of air pollution - REVIHAAP Project: Technical Report. *Copenhagen: WHO Regional Office for Europe*.

WHO (2021). World Health Organisation air quality guidelines: Particulate matter, ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. `https://www.who.int/publications/i/item/9789240034228`.

WHO (2022). Billions of people still breathe unhealthy air: new WHO data. `https://www.who.int/news/item/04-04-2022-billions-of-people-still-breathe-unhealthy-air-new-who-data`.

Wilson, A., Rappold, A. G., Neas, L. M., and Reich, B. J. (2014). Modeling the effect of temperature on ozone-related mortality. *Annals of Applied Statistics*, 8(3):1728–1749.

Winter, H. C. and Tawn, J. A. (2017). kth-order markov extremal models for assessing heatwave risks. *Extremes*, 20:393–415.

Yang, L., Qian, F., Song, D.-X., and Zheng, K.-J. (2016). Research on urban heat-island effect. *Procedia Engineering*, 169:11–18. Fourth International Conference on Countermeasures to Urban Heat Island, 30-31 May and 1 June 2016.

Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M. P. (1999). An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Mathematical Geology*, 31:375–390.