Maynooth University
National University of Ireland Maynooth

Non-stationary spatio-temporal extremal processes;

an analysis of extreme temperature events in Ireland

A thesis submitted in fulfillment of the requirements of the degree of Doctor of Philosophy by

Dáire Healy

Under the supervision of

Prof. Andrew Parnell,
Prof. Jonathan A. Tawn,
Prof. Peter Thorne

Hamilton Institute – Maynooth University
Institiúid Hamilton – Ollscoil Mhá Nuad

August 2023

Dedicated to my parents, thanks for everything!

Acknowledgements

Firstly, I owe a huge thanks to my academic supervisors Andrew, Jon, and Peter whose combined experience and wisdom is nothing less than extreme. Your advice, continuous support, and patience during my PhD has been truly appreciated. You have helped me get through many obstacles along the way and were always generous with your time and knowledge. Thank you for teaching me how to be a researcher.

I am very grateful to have been part of the CRT and Hamilton Institute. I feel incredibly lucky to have been surrounded by such good people over the last few years.

Special thanks go to Janet, Kate, Rosemary, Joanna, Ken, Janice, and David for running the show and providing endless hours of help through technical and administrative struggles.

Finally, I want to express my gratitude to my family and friends for their support and encouragement throughout the PhD process. Without their support, I certainly would not have made it to the end.

Declaration

I, Dáire Healy, declare that I have produced this thesis, titled, "Non-stationary spatio-temporal extremal processes; an analysis of extreme temperature events in Ireland", without the prohibited assistance of any third parties and without making use of aids other than those specified.

The thesis work was conducted from September 2019 to August 2023 under the supervision of Professor Andrew Parnell in the Hamilton Institute, National University of Ireland Maynooth, Professor Jonathan A. Tawn in the Mathematics and Statistics Department, Lancaster University, and Professor Peter Thorne in ICARUS, Maynooth University.

Dáire Healy,
Maynooth, Ireland.
August 2023

Sponsor

This work was supported by a Science Foundation Ireland grant number 18/CRT/6049.

Publications

Of the three chapters in this thesis which cover novel contributions, Chapter 4 is under review by the Journal of the Royal Statistical Society Series C (Applied Statistics). Chapter 3 and Chapter 5 are in preparation to be submitted for publication.

Submitted articles (under review):

- Healy, D., Tawn, J. A., Thorne, P. & Parnell, A. (2023), 'Inference for extreme spatial temperature events in a changing climate with application to Ireland', *arXiv preprint arXiv:2111.08616*.

Abstract

Climate change has resulted in extreme events becoming more frequent, intense, and destructive. It is essential to understand the behaviour of extreme weather processes for societal preparedness. This thesis focuses on the development of spatio-temporal statistical models for extremely hot summer and cold winter temperature events in Ireland, in the context of climate change. To provide reliable estimates of extreme temperatures of unobserved magnitudes, we rely on asymptotically justified extrapolations provided by extreme value theory.

Modelling observational weather records is problematic for several reasons. Firstly, these data tend to have low spatial resolution and contain spatial bias. This thesis presents a framework to unify information about spatial extreme events using both historical temperature observations and output from climate models to enrich the statistical model's topographic information while addressing spatial bias. Secondly, observational weather records are incomplete, with sites having different record lengths as well as missing values. To deal with this issue we develop a spatial extreme value model which allows for an estimation of the full dependence structure without imputing or removing spatial observations that are not fully observed.

We identify a lack of a unified measure of extremity in the spatio-temporal context. Such tools are critical for practitioners to summarise and present risks associated with non-stationary spatio-temporal extremal processes. To address this gap in the literature, we develop several novel summary tools to synthesise and visually communicate non-stationarity in the extreme values of spatio-temporal processes.

We apply our methodology to analyse the extremal behaviour of both summer

maximum and winter minimum daily temperatures in Ireland. We characterise changes on a marginal level (i.e., the behaviour at any location in Ireland) as well as on a spatial level. For the marginal model, we use spatial covariates derived from climate model outputs and explore several potential temporal covariates. To model the spatial extremal dependence, we use an r-Pareto process with Hüsler-Reiss margins in order to incorporate incomplete spatial observations and apply our measures of spatio-temporal non-stationarity to simulations from our fitted model of summer and winter extreme temperatures.

Contents

List of Figures

List of Tables

1

Introduction

In this chapter we define, in general terms, what are the extreme events we are interested in modelling. We highlight the impacts of extreme weather events and the motivation behind the development of extreme value theory to describe them. We discuss the motivation behind the research in this PhD and our research questions.

1.1 Extreme climate events

Extreme climate events are observed with very low probability and although infrequent, can be incredibly impactful. While there is no unified definition of extremes, they commonly comprise unusually hot summers, cold winters, strong windstorms, or heavy rains. Unusual weather events can be devastating, impacting many facets of our environment. It is imperative to understand the mechanisms driving extreme climate activity. In this thesis, we focus primarily on understanding the spatial and temporal changes in extreme temperature events over the island of Ireland.

1.1.1 Trends in global extreme temperatures

Humans have caused an increased concentration of greenhouse gases in our atmosphere, such as carbon dioxide (CO_2), a long-lived gas that absorbs and radiates heat (Friedlingstein et al. 2022). This has resulted in the warming of our planet over the last century (see Figure 1.1 left panel, with temperature anomalies derived from HadCRUT5, Morice et al. 2021). Feedback loops exacerbate this increase in warming, most notably polar amplification. Polar amplification is the tendency for the earth's poles to warm faster than the global average (England et al. 2021). As Arctic ice melts, it reflects less heat away from the planet, which increases warming, and so on. Increasing temperatures cause permafrost (ground that remains frozen for two or more years) to thaw, resulting in increased bio-activity, and contributing to atmospheric greenhouse gasses. The Arctic has warmed almost four times as fast as the global average, leading to changes in temperature gradients over the globe (Rantanen et al. 2022). The changing temperature gradient affects air pressure, resulting in "wavier" jet streams (EASAC 2018), allowing for the formation of extreme weather systems such as persistent rains and increased frequency and magnitude of extreme heat events (Dosio et al. 2018, Moon et al. 2022) and cold winters (Francis & Vavrus 2015, Hallam et al. 2022).

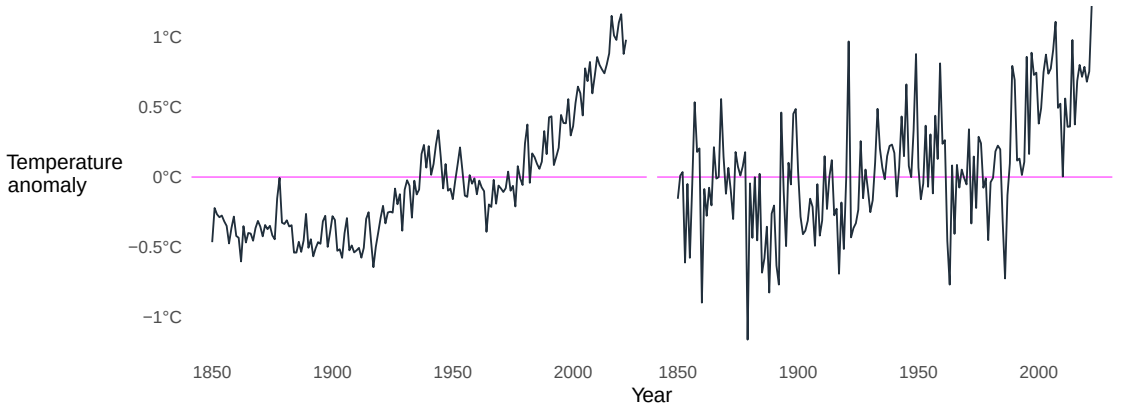

Figure 1.1: Global (left) and Irish (right) mean temperature anomalies from 1850–2023 derived from HadCRUT5.

The Intergovernmental Panel on Climate Change reports an observable change in extreme weather and climate events since about 1950 (IPCC 2021). It is generally accepted that the occurrence of heatwaves has increased in almost all regions around the globe (Brown et al. 2008, Orlowsky & Seneviratne 2012, The Royal Society 2014) and that this increase can be attributed to human activities (Fischer & Knutti 2015, Otto et al. 2017, Gan et al. 2023). Similarly, there are fewer cold days observed around the globe, and an increase in deadly extreme temperatures. Heatwaves kill thousands of people annually, accounting for some of the highest-casualty disasters of recent years (IFRC 2022). Some of the highest temperatures in recorded history have happened very recently. In 2022 alone, record-breaking high temperatures were observed in Europe, South America, South Asia (White et al. 2023), Antarctica, the North and Southwestern Pacific Ocean, the Atlantic, and the Southeastern Pacific Oceans (Cheng et al. 2023). Without a significant reduction of atmospheric greenhouse gas concentrations (which currently continue to rise), the Earth's capacity to support large populations will be significantly reduced in many areas (Lyon et al. 2022) and some parts of the Earth will potentially become uninhabitable (IPCC 2023).

1.1.2 Trends in Irish extreme temperatures

The increase in mean annual temperature in Ireland reflects global trends (Sweeney et al. 2002, McElwain & Sweeney 2003, Walsh 2017, Mateus & Coonan 2022), which

have increased by around 0.9°C over the period 1900-2020 (García et al. 2022), with a rapid increase observed from the 1990s (McElwain & Sweeney 2003). Ireland's mean annual temperature has been higher than average over the past 12 years (2010-2022) with 2022 seeing the warmest mean annual temperature since 1990 (NOAA 2023). This increase is shown in the right-hand panel of Figure 1.1. The strong collinear relationship between global and Irish warming is shown in Figure 1.2. The increase in mean temperatures results in more frequent extreme observations (van der Wiel & Bintanja 2021). McElwain & Sweeney (2007) note that from 1961 to 2005 an "alteration of the temperature distribution" in Ireland has occurred, resulting in a warming of both maximum and minimum temperatures for all sites they considered. García et al. (2022) report an increase in the number of "warm spell" days and a decrease in the number of cold spells from 1961–2018. McElwain & Sweeney (2003) suggest, through exploratory analysis, that both winter and summer temperatures are increasing, with the most significant warming occurring in the winter months. There was a significant decrease in the number of frost days and an increasing number of "hot days" (where the maximum daily temperature exceeds 18°C). O'Sullivan et al. (2020) have shown that the frequency of extreme temperature events over Dublin has increased over the period 1981–2010. Recent summers have been the driest on record (Sweeney 2020). The summer of 2022 saw the hottest temperature recorded in Irish history; a temperature of 33°C was observed at Phoenix Park on 18 July 2022[a]. The evidence of changing temperatures over Ireland is unanimous. It is vital to understand trends and their consequences of unprecedented extreme temperatures heralded by a changing climate.

[a]Note that some historic records dispute this record, stating that the hottest temperature recorded in Ireland is 33.3°C at Kilkenny Castle, 26 June 1887; however, this record has recently been shown to be theoretically implausible (Dooley et al. 2023).

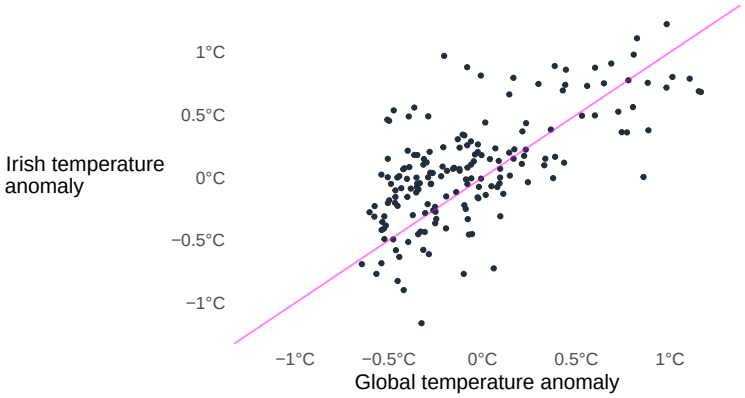

Figure 1.2: Irish against global mean temperature anomalies, derived from HadCRUT5, over the period 1950–2022. The diagonal line is the line of equality. There is a strong relationship between the variables with a correlation coefficient of 0.66.

1.2 Impacts of extreme climate events

"Extreme" is a contextual adjective that can only be interpretable in the context of the climate of a given region. Extreme temperatures in Ireland can be very destructive when considering the infrastructure's inability to deal with them. Ireland has experienced extreme flooding, drought, and snowfall, though the scale of these events is much less than those experienced elsewhere around the world (Government of Ireland 2019). Changes in our environment impact all aspects of society. The immediate and most devastating results of climate change are caused by extreme weather events (Otto et al. 2017). Here we cover a selection of facets of society negatively impacted by extreme climate events caused by a changing climate.

1.2.1 Increased mortality

Extreme weather can, and does, lead directly to increased mortality (The Royal Society 2014). High temperatures accelerate respiratory and cardiovascular diseases, while cold temperatures adversely affect the respiratory and cardiovascular system, increasing heart rate and blood pressure (Schneider & Breitner 2016). The elderly population is at increased risk for health complications caused by extreme weather. Furthermore, an ageing population (Sheehan & O'Sullivan 2020) means that the

proportion of the population at risk is increasing (Cullen 2007). Ireland was found to be the country with the highest rate of excess winter mortality in Europe between 1984 and 2007, with 2,800 excess deaths each winter (Zeka et al. 2014). An excess of 294 deaths was attributed to heatwaves in Ireland during the years 1983, 1984, 1995, and 2006 (Pascal et al. 2013). The risk associated with extremely hot weather in Ireland is moderate, but not insignificant, and is expected to increase (Sweeney et al. 2002).

1.2.2 Disproportionate effects of extreme weather

The negative effects of extreme weather are not equally distributed and densely populated areas are at a disproportionately higher risk (Monteiro et al. 2022). The population of Ireland has increased to 5.6 million (an increase of 7.6%) since 2016 (CSO 2022), recording the highest population since 1841. Urban centres are becoming more densely populated, For example, the population of Saggart, County Dublin has increased by 261% from 2006-2016 (CSO 2017). Rapid urbanisation places resources under increasing pressure (e.g., insufficient Irish health services, lack of housing, water demands; Irish Medical Organisation 2020). Globally, as regions are made uninhabitable by extreme weather events and climate disasters, people move to more opportunistic areas, which are most typically urban centres (Adger et al. 2020). Urbanisation is expected to intensify as the climate continues to change, driving more extreme weather events, and forcing people to find livable areas.

1.2.3 Loss of biodiversity

Changes in weather lead to changes in biodiversity (Sweeney 2020). Species richness is largely driven by climatic factors (Harrison 2014). Warming in Irish spring temperatures has advanced phenological phases (Wingler et al. 2022) for many organisms, potentially disturbing symbioses (Gleeson et al. 2013). Globally, several species have been observed migrating towards the poles over the last century due to increasing temperatures (Hawkins et al. 2009). Increasing extreme weather events, including extreme temperatures, will negatively affect the abundance of Irish species, as well as their spatial distribution, most dramatically in coastal and upland areas (Gorman et al. 2023). These vulnerable Irish habitats are placed under

threat by increasing extreme temperatures (NPWS 2019), placing the inhabiting species at risk. For example, extreme temperature can have a negative impact on species habiting intertidal zones, resulting in premature death (Hiscock et al. 2004). Sabater et al. (2023) show that extreme weather events dramatically affect the biodiversity of rivers, substantially reducing species. Garrabou et al. (2009) link a heat wave in 2003 resulting in the highest sea temperature in recorded history with the mass mortality of several sea organisms.

1.2.4 Econometric effects

The global estimated cost of extreme weather events between 1980 and 2004 is 1.3 trillion euro (The Royal Society 2014). The size of this figure underscores the unpreparedness of our infrastructure to face extreme weather. Heatwaves resulting in drought have a major influence on water supply. For example, in 2018 and 2020, extreme temperatures saw water conservation efforts calling for restrictions on irrigation, which caused widespread crop failure (Antwi et al. 2022). Annual losses from drought in the EU and the UK are estimated to be around 9 billion euro (Cammalleri et al. 2020). Ireland imports the majority of its energy (67% in 2018; Ó Cléirigh 2020) and so is vulnerable to supply interruptions caused by extreme weather globally. For example, recently, due to increased energy costs and extreme cold temperatures in Spain, Ireland was affected by an interrupted supply of fresh produce (Kelleher 2023).

1.3 Extreme value theory

Extreme value theory concerns itself with extrapolation, i.e., assigning probabilities to events that are outside the range of observations. It is incredibly valuable to make predictions about the behaviour of observations that are more extreme than previously observed. Classical methods fall short here since they are concerned with describing the body of the data; thus, a model of the entire process will be biased toward the centre of the data (Davison et al. 2013). The use of these classical methods can lead to underestimating and under-predicting the severity of extreme events. Empirical methods are also insufficient since they limit the analysis to the data range. We can achieve extrapolation through modelling extreme observations

with theoretically justified asymptotic limits described by extreme value theory.

1.3.1 Univariate extreme value theory

Kotz & Nadarajah (2000) and David & Edwards (2001) have traced the origin of extreme value theory to 1709, when Bernoulli considered the longest lifespan of a group of individuals as the expected value of the maximum in a sample. The theory was not of much practical use until a formal extreme value theory framework was developed in the early 20th century by Fréchet (1927), Fisher & Tippett (1928), Von Mises (1936), and later by Gnedenko (1943) and Gumbel (1958). This rapid succession of publications quickly led to the development of the foundational Fisher–Tippett–Gnedenko theorem, often referred to as the *Extremal Types theorem* (covered in Chapter 2). This theorem eventually led to the block maxima modelling technique. Todorovic & Zelenhasic (1970), Todorovic & Rousselle (1971), and Natural Environment Research Council (1975) were concerned with the loss of information when modelling hydrological processes in the block maxima regime. This led to Balkema & de Haan (1974), Pickands (1975), and Leadbetter (1991) describing the asymptotic behaviour of a random variable exceeding some high threshold. Statistical inference procedures were then set out by Davison (1984), Smith (1984), and Davison & Smith (1990), leading to the development of the threshold exceedance modelling framework.

The approximate behaviour of a random process can be assumed to follow a family of known distributions; namely, the generalised extreme value (GEV) distribution in the block maxima setting or the generalised Pareto (GP) distribution in the threshold exceedance setting. Both results were critical in providing a theoretical framework to understand and describe the stochastic patterns governing extremal processes.

1.3.2 Multivariate extreme value theory

It is often useful to consider extreme events in more than one dimension. Extreme value theory is widely applied to environmental processes which are inherently multivariate and/or spatial processes. For example, extreme sea level and sea surge at one location (Coles & Tawn 1994) or river flooding at different locations (Gumbel 1958); both early examples in the multivariate literature. Modelling these processes

in a multivariate way gives a more robust, holistic, and informative analysis. Certainly, modelling a collection of covariates and their relationship with a predictor gives a more complete view of the process.

Properties of bi-variate extreme value processes were initially characterised by Finkelshteyn (1953), de Oliveira (1958), Geffroy (1958), Gumbel (1960), Sibuya (1960), Berman (1961), and de Oliveira (1962). It was known from early on that, unlike the univariate case, there is no given parametric family of distributions to model multivariate extreme observations. Specifically, there is no ubiquitous distribution for the dependence structure between extremal variables. We must choose an appropriate model for the dependence between the tails of the variables, with potentially "infinitely many types of limit distributions" (Sibuya 1960). Initially, empirical and theoretical properties of extremal bi-variate processes were explored. Eventually, parametric models for bi-variate extremal dependence structures were proposed (Tawn 1988). Seminal results by de Haan & Resnick (1977) made modelling multivariate extremal processes (beyond bi-variate) possible. Multivariate extreme value theory subsequently allowed for the analysis of spatial extreme value processes.

The infinite-dimensional analogue to the univariate block maxima approach is the max-stable process (de Haan 1984, Davison et al. 2012). Construction of a max-stable process involves analysis of the limiting behaviour of the multivariate distribution of component-wise maxima over a set of observation sites (when considering a spatial setting). The max-stable process provided a framework to analyse multivariate extreme value processes, however one disadvantage of these methods is their computational cost, with even moderate dimensions being infeasible (Wadsworth & Tawn 2022). Furthermore, the construction of component-wise maxima can disrupt temporal information in the data. Entire spatial observations are conflated and an amalgam of observations make up a single "component-wise event".

Similar to the univariate setting, a multivariate threshold exceedance approach was developed to address concerns of loss of data in the component-wise maxima setting. As a step towards multivariate threshold exceedance modelling, Coles & Tawn (1994) summarise and project multivariate observations onto a univariate process. However, this approach discards spatial information. Analogous to the univariate case, we can consider the distribution of observations that exceed some high thresh-

old. The multivariate generalised Pareto distribution results as the distribution of multivariate observations exceeding a threshold in at least one margin (Falk et al. 2010, Tajvidi 1996, Rootzén & Tajvidi 2006). This approach was extended to the functional setting, with early exploration by Buishand et al. (2008) and later, theoretical development by Ferreira & de Haan (2014) who defined the generalised Pareto process. The authors show that multivariate observations, whose supremum exceeds a threshold, converge to a multivariate generalised Pareto distribution. This was generalised further by Dombry & Ribatet (2015) and de Fondeville & Davison (2018), replacing the supremum with a prescribed function, allowing for a range of functions to define the magnitude of multivariate observations. Choosing a function to measure the magnitude of an event provides greater flexibility in model specification and allows the practitioner to isolate processes of interest. Recently, Palacios-Rodríguez et al. (2020) illustrate a space-time modelling framework, while de Fondeville et al. (2021) and de Fondeville & Davison (2022) have extended further the class of functionals available to measure events while also alleviating some restrictions of the Pareto process.

1.3.3 Recent advances

Extreme value theory is a rapidly developing field. One growing area of research focuses on high dimensional extreme value analysis. Multivariate extreme value models involve complex likelihoods. The computation of these likelihoods can become infeasible as dimensions increase. Methods to reduce computation complexity through approximations have proved very powerful for applying extreme value models to high dimensional problems (Opitz et al. 2018, Simpson et al. 2023). When modelling multivariate extreme events, accurately capturing the dependence between the variables is challenging. Recent developments allow models to capture a wider range of dependence regimes. These include asymptotic independence (Wadsworth & Tawn 2022) and sub-asymptotic models (transitioning from one dependence structure to another; Lugrin et al. 2021). Capturing non-stationarities in extremal data allows us to model the underlying process, which is particularly important in the analysis of extreme weather in a changing climate (Rohmer et al. 2021). More recent developments allow for the extremal dependence structures to vary over time and space (Zhong, Huser & Opitz 2022). An individual variable in and of itself need not be extreme, however the compounding effect of variables

can result in an extreme event (Leonard et al. 2014). The challenging paradigm of concurrent and compound extreme events has received increased attention in recent years (Vignotto et al. 2021, Wazneh et al. 2020).

1.3.4 Thesis overview

The overarching aim of this thesis is to better understand the behaviour of extreme temperatures in Ireland. To date, extreme temperature events in Ireland have been modelled at the marginal level. There is limited knowledge about the behaviour of spatial extreme temperature events in the Irish context. Environmental processes are spatial by nature and so much of their complexity is lost when we consider them in the univariate setting. This thesis presents a novel modelling methodology which captures the temporal and spatial nature of extreme events. We apply this methodology to describe recent changes in both hot summer temperatures and cold winter temperatures in Ireland. Contributions to the literature to achieve this analysis are highlighted below.

1.4 Thesis outline and contributions

Here we provide a brief summary of the contents of each chapter in this thesis, with key contributions highlighted.

In Chapter 2, *"Extreme value theory"*, we give an overview of the necessary literature for the work presented in this thesis. We cover existing modelling methodologies starting from univariate theory, and extending to multivariate spatial analysis. We discuss some limitations of extreme value theory in practice. We derive some foundational results in order to motivate extensions presented later in the thesis.

In Chapter 3, *"Extremal vectors of unequal dimensions with the Hüsler-Reiss distribution"*, we explore properties of the finite-marginal distributions of the Brown-Resnick process, which are Hüsler-Reiss distributed (HRD). The Brown-Resnick process is the basis of a popular parametric model for fitting r-Pareto processes in spatial extreme value theory. This HRD describes the asymptotic behaviour of suitably normalised Gaussian vectors. We provide some theoretical background and illustrate its adaptability to describe the asymptotic behaviour of vectors which are

not necessarily of equal dimension. We are motivated by the temporal inconsistencies in environmental synoptic observations. Typically, modelling a multivariate process across a number of sites requires observations mutual to each site at each time point. Such restrictions result in the practitioner disregarding time spans of data which are not present at all sites. Since extremal observations are typically few, omitting data is clearly not desirable. Our method enables us to develop a parametric r-Pareto process which can make use of all available extremal data.

In Chapter 4, *"Inference for extreme spatial temperature events in a changing climate with application to Ireland"*, we perform an analysis of the extreme max daily summer temperature in Ireland. We investigate the changing nature of the frequency, magnitude and spatial extent of extreme temperatures in Ireland from 1942 to 2020. We develop an extreme value model that captures spatial and temporal non-stationarity in extreme daily maximum temperature data. We use weather station observations for modelling extreme events data while leveraging climate model data to overcome issues linked to the sparse and biased sampling of the observations. Our analysis identifies a temporal change in the marginal behaviour of extreme temperature events over the study domain, which is much larger than the change in mean temperature levels over this time window. We illustrate how these characteristics result in increased spatial coverage of events that exceed critical temperatures.

In Chapter 5, *"Unusual winter events in Ireland"*, we extend the methodology developed in Chapter 4 to analyse extreme daily minimum temperatures in winter months. We model the marginal distributions of extreme winter minima using a generalised Pareto distribution (GPD), capturing temporal and spatial non-stationarities in the parameters of the GPD. We investigate two independent temporal non-stationarities in extreme winter minima. We model the long-term trend in magnitude of extreme winter minima as well as the short-term, large fluctuations in magnitude caused by anomalous behaviour of the jet stream. We measure magnitudes of spatial events with a carefully chosen risk function and fit an r-Pareto process to extreme events exceeding a high-risk threshold. Our analysis is based on synoptic data observations courtesy of Met Éireann and MIDAS. We show that the frequency of extreme cold winter events is decreasing over the study period. The magnitude of extreme winter events is also decreasing, indicating that winters are warming, and apparently warming at a faster rate than maximum daily summer

temperatures. We find that a climate model output which was informative as a covariate for modelling extremely warm summer temperatures is less informative as a covariate for extremely cold winter temperatures. However, we show that the climate model output is useful for informing our non-extreme temperature model.

Finally, in Chapter 6, *"Discussion and Conclusion"*, we review and summarise the work and contributions of this thesis. We highlight some areas of further work and recommend possible extensions to our research.

1.5 Software

All analysis in the thesis was done throughout the `R` programming language. The software presented in Chapter 3 is available at github.com/dairer/hrd. Code associated with analysis in Chapter 4 is available at github.com/dairer/Extreme-Irish-Summer-Temperatures. Code associated with the analysis carried out in Chapter 5 is available at github.com/dairer/Extreme-Irish-Winter-Temperatures.

2

Extreme value theory

In this chapter we set out the relevant and necessary theoretical results for this thesis. In Section 2.1 we define and summarise results on extreme value theory in the univariate setting. In Section 2.2 we cover multivariate results followed by recent developments in the literature We provide an overview of standard multivariate modelling approaches to date, highlighting cases of their appropriate use, and giving examples and critiques throughout. In Section 2.3 we provide a guide to extending multivariate extreme value models to the spatial setting.

2.1 Univariate extremes

There are two widely applied definitions of univariate extremes with both resulting in different analysis procedures. We will cover both methods and give outlines of their derivations here. The block maxima approach is covered in Section 2.1.1 and the threshold exceedance approach in Section 2.1.2.

2.1.1 Block maxima

Consider n independent and identically distributed (*iid*) continuous random variables from a distribution function F, $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$. We are interested in the behaviour of the sample maxima, which is defined as

$$M_n = \max\{X_1, X_2, \ldots, X_n\}.$$

We can straightforwardly derive the distribution of sample maxima as

$$\begin{aligned}\Pr(M_n \leqslant x) &= \Pr(X_1 \leqslant x,\ X_2 \leqslant x,\ \ldots,\ X_n \leqslant x) \\ &= \Pr(X_1 \leqslant x) \times \Pr(X_2 \leqslant x) \times \cdots \times \Pr(X_n \leqslant x) \\ &= F^n(x).\end{aligned}$$

As we are interested in the distribution of the sample maxima as we collect more extreme observations, we consider the distribution of M_n as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \Pr(M_n \leqslant x) = \lim_{n\to\infty} F^n(x).$$

Since we are taking the sample maxima of increasingly larger samples, the distribution of M_n will have its mass entirely on the upper-end point of F as $n \to \infty$. Furthermore, if $F(x) < 1$, then $\lim_{n\to\infty} \Pr(M_n \leqslant x) = 0$ and if $F(x) = 1$, then $\lim_{n\to\infty} \Pr(M_n \leqslant x) = 1$.

To avoid the distribution of the sample maxima collapsing upon a single point, we apply a general linear scaling that adjusts the sample maxima with respect to the

sample size. Specifically, we consider

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leqslant x\right) &= \lim_{n\to\infty} \Pr\left(M_n \leqslant a_n x + b_n\right) \\ &= \lim_{n\to\infty} F^n(a_n x + b_n).\end{aligned} \tag{2.1}$$

The Fisher–Tippett–Gnedenko theorem (Fisher & Tippett 1928, Gnedenko 1943), a foundational result in extreme value statistics, tells us that if there exist sequences $a_n > 0,\ b_n \in \mathbb{R}$, such that (2.1) has a non-degenerate limiting distribution function,

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant x\right) \to G(x) \text{ as } n \to \infty, \tag{2.2}$$

then G is one of only three possibilities. The three possible limiting distributions are the Gumbel

$$G(x) = \exp\left\{-\exp(-x)\right\}, \text{ for } x \in (-\infty, \infty),$$

Fréchet

$$G(x) = \begin{cases} 0, & \text{for } x \leq 0, \\ \exp\left(-x^{-\alpha}\right), & \text{for } x > 0, \end{cases}$$

and reversed Weibull

$$G(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\}, & \text{for } x < 0, \\ 1, & \text{for } x \geq 0, \end{cases}$$

distributions, where $\alpha > 0$. This is quite a remarkable and surprising result. Essentially, without knowing anything about a distribution function F, we can know the distribution of its sample maxima. Each distribution describes three potential extremal behaviours of the sample maxima, with three potential rates of upper tail decay. The Fréchet distribution has a very heavy tail with polynomial decay. The Gumbel distribution has a lighter, exponentially decaying tail while the reversed Weibull distribution has an upper bound. Figure 2.1 shows an example of each of the three distributions.

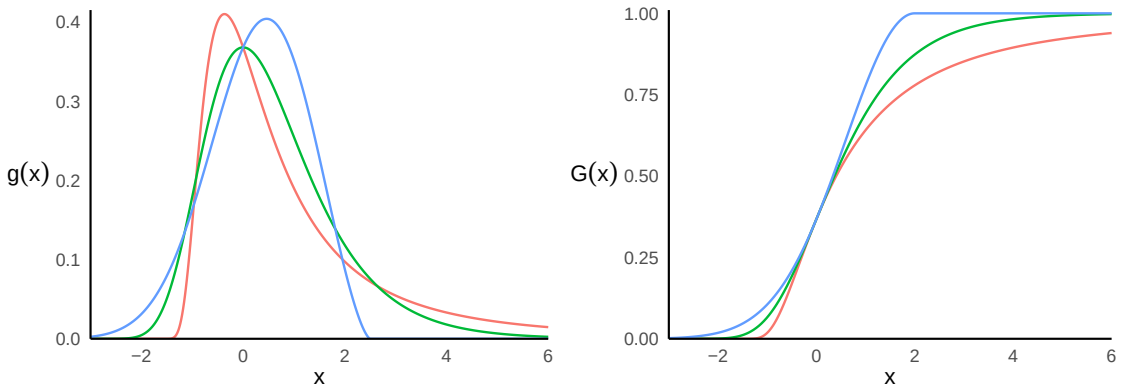

Figure 2.1: Density function (left) and distribution function (right) of the Generalised Extreme Value distribution (2.3) with a location parameter of 0 ($\mu = 0$), a scale parameter of 1 ($\sigma = 1$) and all three possible distribution classes determined by the shape parameter: the Gumbel distribution (green), with a shape parameter of 0 ($\xi = 0$) and exponential tail decay; the reversed Weibull distribution (blue) with a negative shape parameter ($\xi = -0.4$) and a finite upper endpoint; the Fréchet distribution (orange), with a positive shape parameter ($\xi = 0.4$) and polynomial tail decay.

The Gumbel, Fréchet, and reversed Weibull distributions can be generalised and expressed as one distribution function, the General Extreme Value (GEV) distribution

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{2.3}$$

parameterised by the location parameter μ, the scale parameter $\sigma > 0$, and the shape parameter, ξ. The three distributions are recovered by specifying ξ, the shape parameter, i.e., $\xi = 0$ gives the Gumbel distribution, $\xi > 0$ gives the Fréchet distribution and $\xi < 0$ gives the reversed Weibull distribution. This parameterisation alleviates the onus on the practitioner to decide which extreme value distribution best describes their data set as the fit will provide evidence for ξ being positive, negative or zero.

2.1.1.1 The GEV in practice

We say that an arbitrary distribution function F is in the max-domain of attraction of a particular extreme value distribution G, if G is the asymptotically limiting distribution of the scaled sample maxima of F. If a collection of random variables

$X_1, X_2, \ldots, X_n$ are in the max-domain of attraction of an extreme value distribution G, we know that limit (2.2) holds so, in practice, it is reasonable to assume that for large n

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant x\right) = G(x),$$

which implies that,

$$\Pr\left(M_n \leqslant x\right) = G\left(\frac{x - b_n}{a_n}\right) = G^*(x),$$

where G^* is also a GEV distribution with different location and scale parameter values than G. The scaling series a_n and b_n are absorbed into the parameters of the GEV distribution G^*. This means that we can fit a GEV distribution directly to a set of sample maxima and estimate G^*. Construction of sample maxima can be achieved by dividing a series of observations into blocks of equal size and taking each block maximum. We interpret the constructed block maxima as samples from a GEV distribution. Figure 2.2 gives an example block maxima analysis using simulated *iid* standard Gaussian data, in the left plot we partition the data into blocks of equal size, take the maximum observation in each block, and model the distribution of their magnitudes using the GEV. The right figure shows the QQ plot of the data's sample maxima against the quantiles of the fitted GEV distribution.

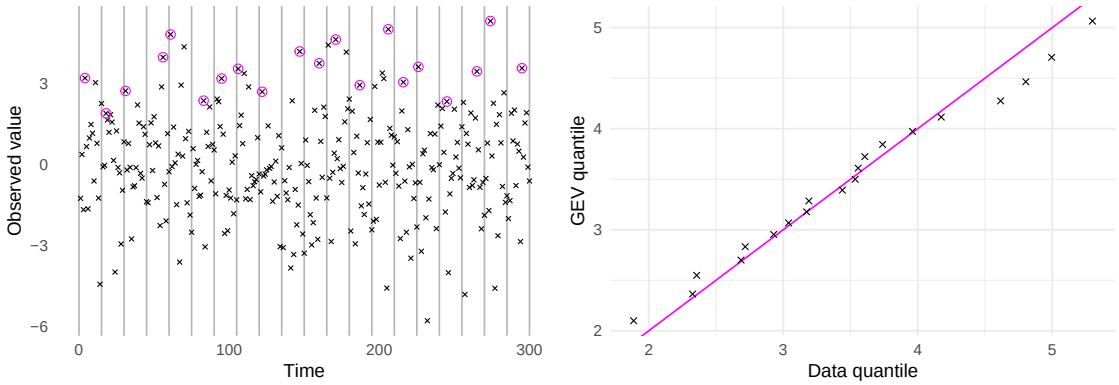

Figure 2.2: Simulated time series, constructed with *iid* samples from a Gaussian distribution (left). Vertical lines show the block partitions, where the maxima in each block (circled in magenta) are assumed to follow a GEV distribution. QQ-plot of the maxima in each block against the quantiles of the fitted GEV distribution (right), the diagonal magenta line shows the line of equality.

In many applications, blocking the data can be intuitive and instinctive, for example, taking the yearly maxima of some observed climate variable. However, from a statistical perspective, it is challenging to specify block length. Simply taking natural time blocks can lead to poor use of data. Taking only one value per block can lead to inefficient usage of extreme observations. For example, looking at the simulated time series in Figure 2.2, we can see several extreme data points that are not classified as extreme since they are not the largest in their respective block. In general, the block maxima approach is wasteful of extremal data. We will now cover the threshold exceedance approach, which tends to be more frugal with extreme observations.

2.1.2 Threshold exceedances

An alternative definition of univariate extremes with its own asymptotically limiting distribution was developed by Balkema & de Haan (1974), Pickands (1975), and Leadbetter (1991). Rather than taking only the largest value in a time span, the authors considered the properties of all observations that exceeded some high threshold. The following section presents work originally from Smith (1984), for a textbook treatment see Coles (2001).

Consider a set of n, *iid* variables, $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$. We define a point process, denoted $\mathcal{Q}_n \in \mathbb{R}^2$, as

$$\mathcal{Q}_n = \left\{ \frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right\}, \text{ for } i = 1, \ldots, n,$$

for some $a_n, b_n \in \mathbb{R}$ and $a_n > 0$. The first coordinate of $\mathcal{Q}_n$ is scaled by $n+1$ to ensure it lies within the range $(0, 1)$. The second coordinate is linearly scaled in a similar manner as described in Section 2.1.1 to avoid the distribution degenerating upon a single mass, i.e., we assume there exists some appropriate linear scaling of sample maxima such that $\Pr\{(M_n - b_n)/a_n \leqslant x\} \to G(x)$ as $n \to \infty$ where G is a non-degenerate. The point process $\mathcal{Q}_n$ converges to a non-homogeneous Poisson point process (NHPP) $\mathcal{Q}$ on any set $\mathcal{A} \subseteq [0, 1] \times (x_G, \infty)$ as $n \to \infty$, where $x_G = \lim_{n\to\infty}(x_F - b_n)/a_n$ and x_F and is the lower end point of the distribution F. The point process Q has intensity measure on $\mathcal{A}$, defined as $\Lambda(\mathcal{A})$, which we will next derive.

For point process Q_n in $\mathbb{R}^2$, we define the number of points in any subset of that space $\mathcal{A} \subset \mathbb{R}^2$, as $N_n(\mathcal{A})$, and the expected number of these points as $n \to \infty$ is $\Lambda(\mathcal{A}) := \lim_{n\to\infty} \mathbb{E}\{N_n(\mathcal{A})\}$. We restrict the subset $\mathcal{A}$ to be a subset of the Cartesian product $\mathcal{A} = [t_1, t_2] \times \left[u, x^G\right) \subset \mathbb{R}^2$, where $0 \leqslant t_1 < t_2 \leqslant 1$ and $x_G < u < x^G$, where $x^G = \lim_{n\to\infty}(x^F - b_n)/a_n$ and x^F is the upper end point of F. Figure 2.3 illustrates the set $\mathcal{A}$; intuitively, u will act as a threshold, with points exceeding it considered as extreme.

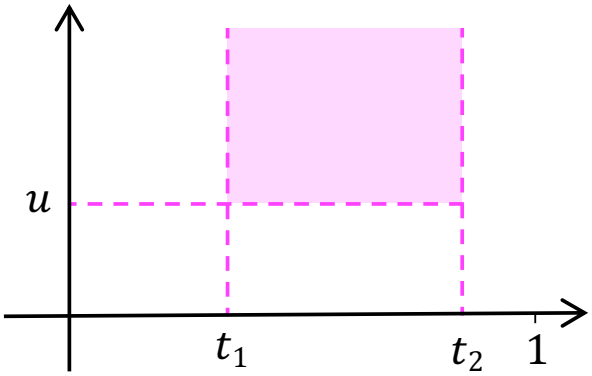

Figure 2.3: The set $\mathcal{A} = [t_1, t_2] \times \left[u, x^G\right) \subseteq \mathbb{R}^2$ is shaded.

The probability of observing N_n points of $\mathcal{Q}_n$ exceeding a level u between times nt_1 and nt_2 can be described by a Binomial distribution, i.e.,

$$N_n([t_1, t_2] \times \left[u, x^G\right)) \sim \text{Bin}\left(\lfloor (t_2 - t_1)\, n \rfloor, \Pr\left\{(X_i - b_n)/a_n > u\right\}\right), \tag{2.4}$$

where $\lfloor \cdot \rfloor$ is the floor operator. The number of observations in (2.4) is calculated as $\lfloor (t_2 - t_1)\, n \rfloor$ since data are independent and uniformly distributed on $[0, 1]$ and so the expected number of points in $[t_1, t_2]$ is $t_2 - t_1$.

As the number of observations grows, a binomial distribution can be approximated by a Poisson, with the rate parameter of the Poisson, or the expected value, being equal to the number of observations times the probability of success in the binomial distribution (Blitzstein & Hwang 2014, Theorem 4.8.3). So, as $n \to \infty$,

$$N_n([t_1, t_2] \times \left[u, x^G\right)) \sim \text{Poisson}\left\{(t_2 - t_1)\, n \Pr\left\{(X_i - b_n)/a_n > u\right\}\right\},$$

with the expected number of points exceeding u being

$$\begin{aligned}\Lambda\left([t_1,t_2]\times\left[u,x^G\right)\right) &= \lim_{n\to\infty}\left(t_2-t_1\right)n\Pr\left\{(X_i-b_n)/a_n>u\right\}\\ &= \left(t_2-t_1\right)\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}.\end{aligned}$$

Therefore, we have that for $\mathcal{A}=[t_1,t_2]\times\left[x,x^G\right)$

$$\Lambda(\mathcal{A})=\left(t_2-t_1\right)\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi},$$

Assuming there exists an intensity function $\lambda:\mathbb{R}^2\to\mathbb{R}_+$ whose integral over the set $\mathcal{A}$ describes the expected number of points in $\mathcal{A}$, we can define the intensity measure of the process $\mathcal{Q}$ as

$$\Lambda(\mathcal{A})=\int_{\mathcal{A}}\lambda(\boldsymbol{x})d\boldsymbol{x}. \tag{2.5}$$

Specifically, if we define $\mathcal{A}=[0,t]\times\left[x,x^G\right)$, the intensity function is calculated as

$$\lambda(t,x)=-\frac{\partial^2\Lambda\left(\mathcal{A}\right)}{\partial t\partial x}=\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi},$$

for $t<1$ and $u<x<x^G$.

We are interested in calculating the conditional distribution of observations that exceed a high threshold. Taking any $x>u$, we have that

$$\begin{aligned}\Pr\left(\frac{X_i-b_n}{a_n}>x\;\middle|\;\frac{X_i-b_n}{a_n}>u\right) &= \frac{\Lambda\left([t_1,t_2]\times\left[x,x^G\right)\right)}{\Lambda\left([t_1,t_2]\times[u,x^G)\right)}\\ &= \frac{\left(t_2-t_1\right)\left[1+\xi(x-\mu)/\sigma\right]^{-1/\xi}}{\left(t_2-t_1\right)\left[1+\xi(u-\mu)/\sigma\right]^{-1/\xi}}\\ &= \left[1+\frac{\xi(x-u)}{\sigma^*}\right]_+^{-1/\xi},\end{aligned} \tag{2.6}$$

where $\sigma^*=\sigma+\xi(u-\mu)$. So, we have the generalised Pareto distribution (GPD),

which we will denote as H,

$$H(x) = 1 - [1 + \xi x/\sigma^*]_+^{-1/\xi}, \tag{2.7}$$

where $x \geqslant 0$ and $(1 + \xi x/\sigma^*) > 0$.

The shape parameters of the GEV and the GPD are equivalent. For $\xi = 0$, (as $\xi \to 0$) we have the exponential distribution. If $\xi < 0$, the GPD is bounded at $u - \sigma^*/\xi$. If $\xi > 0$, the distribution has a heavy tail, polynomial rate of decay. An example of each distribution class can be seen in Figure 2.4.

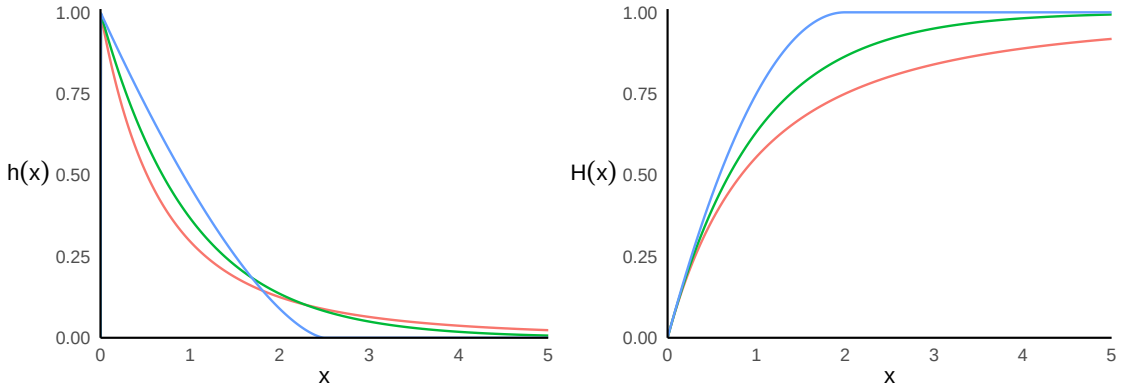

Figure 2.4: Density function (left) and distribution function (right) of the generalised Pareto distribution (2.7) with a location parameter of 0 ($\mu = 0$), a scale parameter of 1 ($\sigma = 1$) and all three possible distribution classes determined by the shape parameter. A shape parameter of 0 ($\xi = 0$) with exponential tail decay (green). A negative shape parameter ($\xi = -0.4$) with a finite upper endpoint (blue). A positive shape parameter ($\xi = 0.4$) with polynomial tail decay (orange).

2.1.2.1 The GPD in practice

Similar to Section 2.1.1, we do not need to know the scaling series a_n and b_n in practice. This is because the parameters of the GPD can absorb the scaling constants, so the distribution of the standardised maxima can itself be approximated by the GPD. Let $u_n = a_n v + b_n$, for some v, then for $x > 0$

$$\Pr\left(\frac{X_i - b_n}{a_n} \leqslant x + v \;\middle|\; \frac{X_i - b_n}{a_n} > v\right) = \Pr\left(X_i \leqslant a_n x + u_n \mid X_i > u_n\right).$$

As $n \to \infty$, applying (2.6), we can see that

$$\Pr\left(X_i \leqslant a_n x + u_n \mid X_i > u_n\right) \to 1 - \left[1 + \frac{\xi x}{\sigma^*}\right]_+^{-1/\xi}. \tag{2.8}$$

For large n, we consider (2.8) as an equality letting $u = u_n$, that is, for $y > 0$,

$$\Pr\left(X \leqslant y + u \mid X > u\right) = 1 - \left[1 + \frac{\xi y}{\tilde{\sigma}^*}\right]^{-1/\xi} = H(y),$$

where $\tilde{\sigma}^* = a_n \sigma^*$. The unconditional probability for $x > u$ is then

$$\begin{aligned}\Pr(X \leqslant x) &= (1 - \lambda_u) + \lambda_u H(x - u) \\ &= 1 - \lambda_u \left[1 + \frac{\xi(x-u)}{\tilde{\sigma}^*}\right]_+^{-1/\xi}\end{aligned}$$

where

$$\lambda_u = \Pr(X > u) \tag{2.9}$$

is the threshold exceedance probability.

In practice, we pick u to be a high quantile of the data since it arises as $u = u_n = a_n v + b_n$, so as $n \to \infty$ this term would tend to x_F. With careful checks on the selection of u, observations above that threshold can be assumed to follow a generalised Pareto distribution, H.

Figure 2.5 presents an example analysis of simulated *iid* standard Gaussian data. In the left-hand plot of this figure, we have the identical simulated time series as presented in Figure 2.2. Here, the horizontal line corresponds to the threshold, u, (chosen as the 95th percentile of the data). The data exceeding it are classified as extreme and the distribution of their magnitudes is assumed to follow the GPD. The right-hand plot shows a QQ-plot of the extremal data quantiles against the fitted GPD.

The GPD satisfies a threshold stability property (Davison & Smith 1990). Once a sufficiently high threshold is chosen, say u_a, the distribution of exceedances of that threshold are assumed to be a GPD. A consequence of this assumption is that the distribution of excesses of any higher threshold, $u_b > u_a$ will also be described by a GPD, with a deterministic shift in the scale parameter. For example, suppose that

$(X - u_a | X > u_a) \sim \text{GPD}(\sigma_a^*, \xi)$. Choosing a higher threshold, $u_b > u_a$, we have $(X - u_b | X > u_b) \sim \text{GPD}(\sigma_b^*, \xi)$ where $\sigma_b^* = \sigma_a^* + \xi(u_b - u_a)$ with the same shape parameter, ξ.

This raises the issue of how to choose an appropriate threshold, u. There is extensive literature on the issue of threshold selection and validation with a review given by Scarrott & MacDonald (2012). Of course, the higher the value for the threshold, the more accurate the assumption of the GPD, however the higher the threshold, the fewer exceedances. There is ultimately an application-specific, bias–variance tradeoff which requires consideration by the practitioner when choosing a threshold. Standard procedures to assess the fit of the GPD help determine if the threshold is appropriate, for example, parameter stability and mean residual life plots (Coles 2001).

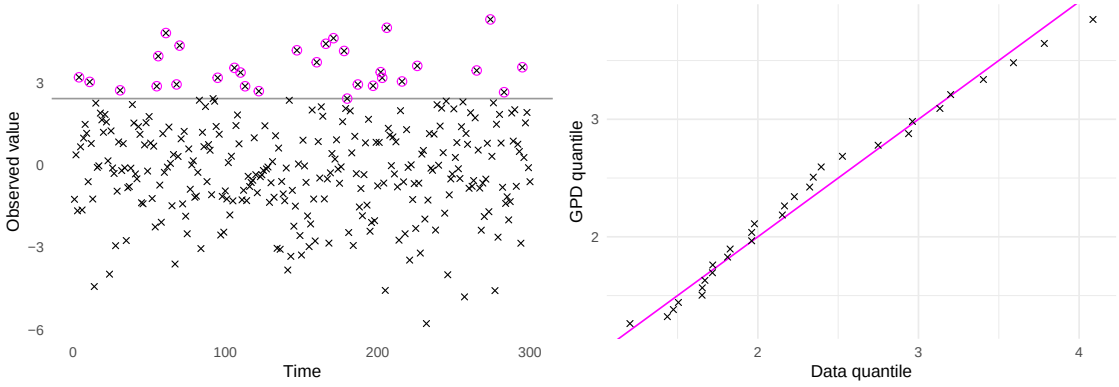

Figure 2.5: Simulated time series, constructed with *iid* samples from a Gaussian distribution (left). The horizontal line shows the threshold, u, and points exceeding u are assumed to follow a GPD distribution. QQ-plot of the threshold exceedances against the quantiles of the fitted GPD distribution (right), the diagonal magenta line shows the line of equality.

2.1.3 Typical modelling challenges

2.1.3.1 Extreme minima

Modelling extreme minima is readily achievable through modelling the maxima of the negated data. Given $X_1, X_2, \ldots, X_n \sim F$, we are interested in modelling the extreme minima of these variables. This is equivalent to analysing the extreme

maxima of $-X_1, -X_2, \ldots, -X_n$.

$$\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n).$$

In this way, all theoretical results in Sections 2.1.1 and 2.1.2 hold for extreme minima.

2.1.3.2 Model validation

As in any statistical modelling, assessment of a model's performance is a critical task. We can generate a QQ plot of data quantiles against model quantiles for a visual indicator of a model's performance. Given data $x_1, x_2, \ldots, x_n$, we wish to check the quality of the fit of a distribution F with estimated parameters $\widehat{\boldsymbol{\theta}}$. The $\{i/(n+1)\}$-th empirical quantile is, $x_{(i)}$ where $x_{(1)} < \cdots < x_{(i)} < x_{(i+1)} < \cdots < x_{(n)}$ are the ordered values of $x_1, x_2, \ldots, x_n$. The predicted values of the fitted distribution at those quantiles are

$$\hat{x}_{(i)} = F^{-1}\left(\frac{i}{n+1}; \widehat{\boldsymbol{\theta}}\right). \tag{2.10}$$

Plotting the model's quantiles $\hat{x}_{(i)}$ against the data $x_{(i)}$, for $i = 1, \ldots, n$, gives a visual diagnostic QQ-plot of the accuracy of the fitted distribution. Ideally, all points would lie along the diagonal and $x_{(i)} = \hat{x}_{(i)}$, for each $i = 1, \ldots, n$. The closer the points to the diagonal the more accurate the fitted distribution.

In reality, observed values x_i for $i = 1, 2, \ldots, n$, are generated from some true, unknown, high-dimensional process, F^*. In natural processes, there are typically complex and high-dimensional dependencies contained within the observations from F^* and so the data are not independently distributed. The true likelihood function of the entire process is often difficult or infeasible to specify. However, the lower dimensional, univariate marginal distributions can be assumed and modelled (Cox & Reid 2004). An approximation of the likelihood can be achieved by intentionally, falsely assuming independence in the data. Then, the resulting so-called pseudo-

likelihood is given by,

$$\mathcal{L}_p(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}),$$

where f is the density function of the assumed univariate marginal distributions, parameterised by $\boldsymbol{\theta}$.

When using a pseudo-likelihoods, classic likelihood comparisons such as AIC or BIC between models are no longer accurate (Claeskens & Hjort 2008). If the true model, F^*, which captures the full dependence structure, is considered, the information criterion will select it as the best model. However, if the model assumes independence, dependence in the data causes a steeper gradient of the likelihood, biasing the information criteria and potentially choosing an over-parameterised model. Note that the block maxima approach described in Section 2.1.1 disrupts the temporal and spatial information in the data, removing dependence and alleviating this issue to some degree. Similar alleviation can be achieved within the threshold exceedance paradigm by temporally declustering threshold exceedances.

We generally see strong spatio-temporal dependence within climatological data which can bias information criteria, potentially choosing an over-parameterised model, hence alternative methods of model selection are preferable. There exist adjustments of the likelihood to account for this, e.g., CLIC (composite likelihood information criterion) which adjusts likelihood based on the gradient (Davis et al. 2013), however these are asymptotically motivated adjustments so may not be ideal in the extreme value context given small samples.

If sufficient data are available, out-of-sample model validation such as k-fold cross-validation can potentially be used as a non-biased model comparator. Empirical comparisons can then be performed between the out-of-sample data and the model's predictions in terms of, for example, root mean squared error (RMSE). For RMSE, the error is taken as the difference between out-of-sample observations and the predicted values at the corresponding quantiles of the model, as in (2.10). To calculate RMSE, we must define the quantiles of our observations. Alternatively, we can use metrics such as continuous ranked probability score (CRPS) which avoid specifying the quantile of the data (Zamo & Naveau 2018). The CRPS measures the difference between the distribution function F and a step distribution function

at the observed value y,

$$\text{CRPS}(F, y) = \int \left(F(x) - \mathbf{1}_{\{x \geq y\}}\right)^2 dx,$$

where $\mathbf{1}_{\{.\}}$ denotes the indicator function. The CRPS can be calculated iteratively for each out-of-sample data point. Finally, an average score over all data points can be used as a model performance metric.

2.1.4 Return levels and return periods

Return periods and return levels are often reported in the extreme value literature as they are readily interpretable statistics to a non-statistical audience. For example, say X is observed daily, let x^t be the return level with return period t of the random variable X. We expect to observe a level x^t, on average, once every t days. That is,

$$\Pr(X > x^t) = 1/t.$$

In other words, the average time between events of magnitude x^t is t days.

Clearly, we can derive a return period/level from any distribution function. However, if we are interested in making inferences about return levels beyond the time span of historical records, or estimating return periods associated with magnitudes rarely or not yet observed, it is most sensible to derive them from distributions which accurately describe the upper-most tail of the distribution. That is, from the GEV or GP distributions. Given a distribution function F, a return level with return period t is calculated by solving $F(x^t) = 1 - 1/t$ for x^t, i.e., $x^t = F^{-1}(1 - 1/t)$, where F^{-1} is the inverse of the distribution function. The return value associated with return period t of the GEV is

$$x^t = \begin{cases} \mu - \sigma/\xi \left\{1 - \left[-\log\left(1 - 1/t\right)\right]^{-\xi}\right\}, & \text{for } \xi \neq 0, \\ \mu - \sigma \log\left[-\log\left(1 - 1/t\right)\right], & \text{otherwise.} \end{cases}$$

The return value associated with return period t of the GPD with threshold ex-

ceedance probability λ_u, as defined in (2.9), is

$$x^t = \begin{cases} u + \sigma/\xi \left[(t\lambda_u)^\xi - 1\right], & \text{for } \xi \neq 0, \\ u + \sigma \log(t\lambda_u), & \text{otherwise.} \end{cases}$$

Calculating return levels assumes that the probability of occurrence of extreme events is constant over time, so reporting return levels in a non-stationary setting is clearly not ideal. However, a powerful insight into the level of non-stationarity is the change in return level over time. For example, in climate change analysis we can estimate the 100-year return level of a process. A 100-year event corresponds to an event with a 0.01 probability of occurrence in a given year. We can report the change in return level over time, and highlight, in an interpretable way, the rate of change of the process.

2.2 Multivariate extremes

In the multivariate setting, unlike in the univariate case, there is no ubiquitous distribution to rely on once you have defined your extreme observations. Careful consideration of the dependence between the tails of the marginal variables is required and determines which modelling techniques are appropriate. This has resulted in a challenging modelling framework. However, this also affords quite a lot of modelling approaches and has resulted in some very interesting theory. In this section, we will cover some of the main theoretical frameworks used in multivariate extreme value theory and which are appropriate in different dependence settings.

2.2.1 Multivariate ordering

The first challenge a practitioner is faced with when performing multivariate extreme value analysis is the issue of ordering multivariate observations. How do we decide that one observation is more extreme than another? There are several possible orderings, each with different theoretical consequences (Barnett 1976). We will focus on two classes of orderings which have found wide applicability in the literature, and which are analogous to those used in the univariate theory described in Section 2.1 (see Figure 2.6). Namely, a block maxima approach, resulting

in max-stable distributions, and a threshold exceedance approach resulting in the multivariate Pareto distribution. Respectively, these two approaches lead to the max-stable process and the generalised Pareto process in the continuous spatial context.

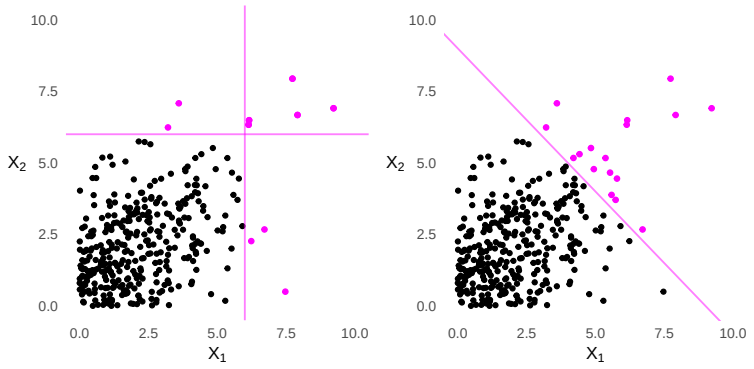

Figure 2.6: Alternative definitions of multivariate extremes, a component-wise maxima approach (left), and a threshold exceedance approach (right). Extreme observations are highlighted in both cases.

2.2.2 Marginal standardisation

To meaningfully discuss the dependence between two variables, it is convenient to assume they have equivalent marginal distributions. This is not an unreasonable assumption since it is easily constructed. For example, suppose $X \sim F$ with X a continuous random variable; applying probability integral transform we can map this random variable to a uniform distribution, i.e., $F(X)$ is uniformly distributed (Angus 1994). Following this, we can transform the variable to have any distribution with cumulative distribution function F_*, via the transformation: $F_*^{-1}\{F(X)\}$. For example, to transform variable X with distribution F to be Fréchet or Pareto distributed we have

$$X^{\text{Fréchet}} = -1/\log F(X), \qquad X^{\text{Pareto}} = 1/(1 - F(X)) \tag{2.11}$$

respectively, both of which are commonly used standardisations in the multivariate extreme value literature. The choice of standardisation is mostly arbitrary, however some distributions have attractive features which make them useful in studying multivariate extremes. For example, consider the Fréchet distribution; it has the majority of its mass near 0 with heavy polynomial upper tails. This accentuates

the data in the very tails of the distribution. Transforming data to be Fréchet distributed can highlight and focus attention on extremal patterns in the data. The same can be said of any heavy-tailed distribution, such as the Pareto distribution.

2.2.3 Extremal dependence

With early characterisation by Sibuya (1960), there are two possible dependence regimes in the multivariate setting; asymptotic dependence and asymptotic independence. The class of extremal dependence exhibited by the data determines the appropriate modelling strategies a practitioner should use. Standardising data to have a common marginal distribution allows us to focus solely on the dependence between the random variables.

Consider a pair of random variables $X_1 \sim F_1$ and $X_2 \sim F_2$. We are interested in characterising the upper dependence of the tails of these random variables; that is, when X_1 is extreme, how likely is it for X_2 to be similarly extreme, and *vice versa*. The upper tail dependence coefficient describes this asymptotic dependence between two random vectors and is defined as

$$\chi(X_1, X_2) = \lim_{z \to 1^-} \Pr\left[F_1(X_1) > z \mid F_2(X_2) > z\right].$$

Note that the lower tail dependence can be characterised by the upper tail dependence of the negated data, as in Section 2.1.3.1. In practice, to get an estimate of χ, we pick $z = u$ to be a high quantile and denote the conditional probability as,

$$\chi_u(X_1, X_2) = \Pr\left[F_1(X_1) > u \mid F_2(X_2) > u\right],$$

which can be estimated empirically for $0 < u < 1$. Calculating χ_u over a sequence of high quantiles gives the practitioner an indication of the extremal dependence regime. If $\chi(X_1, X_2) > 0$, or in practice, if $\chi_u(X_1, X_2) \to c$ as $u \to 1$ and $c > 0$, the variables are judged to be asymptotically dependent. If $\chi(X_1, X_2) = 0$, or in practice, if $\chi_u(X_1, X_2) \to 0$ as $u \to 1^-$, the variables are interpreted as asymptotically independent (Coles et al. 1999). Note that $\chi(X_1, X_2) = \chi(X_2, X_1)$ as the margins of X_1 and X_2 are equal.

Historically, models describing asymptotically-dependent data have received disproportionate attention and development in the literature. However, in reality,

extremal data is quite often asymptotically independent. See Coles et al. (1999) for a discussion of the measure $\bar{\chi}$, an analogous measure to χ, which instead describes the degree of asymptotic independence between random variables. Incorrectly assuming asymptotic dependence can lead to overestimation of joint extreme events (Coles et al. 1999). The data analysed in this thesis exhibit asymptotic dependence so we won't cover a great deal of asymptotic independence theory. Suffice to say, numerous models have been proposed that can capture a wide range of dependence regimes. Models describing asymptotically independent spatial data (Wadsworth & Tawn 2022) and sub-asymptotic data (transitioning from one dependence structure to another at further distances; Lugrin et al. 2021, Wadsworth & Tawn 2012, Huser & Wadsworth 2022) reflect many natural processes. Huser & Wadsworth (2022) give a thorough review of recent advances in modelling a wide range of dependence regimes in the spatial context.

2.2.4 Component-wise maxima

The component-wise maxima approach is the multivariate analogue to the univariate block maxima approach as described in Section 2.1.1. Consider a set of n, independent d-dimensional random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from a distribution function F, where $\boldsymbol{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,d})$ for $i = 1, 2, \ldots, n$. Define the vector of component-wise maxima of these vectors as

$$\boldsymbol{M}_n = \{M_{n,1}, M_{n,2}, \ldots, M_{n,d}\},$$

where $M_{n,k}$ is the maximum observation in the k-th position across the n vectors. We can derive the d-dimensional distribution of component-wise maxima $\boldsymbol{M}_n$ as

$$\begin{aligned}
\Pr(\boldsymbol{M}_n \leqslant \boldsymbol{x}) &= \Pr(M_{n,1} \leqslant x_1, M_{n,2} \leqslant x_2, \ldots, M_{n,d} \leqslant x_d) \\
&= \Pr\left(\max_{1 \leqslant i \leqslant n} X_{i,1} \leqslant x_1, \max_{1 \leqslant i \leqslant n} X_{i,2} \leqslant x_2, \ldots, \max_{1 \leqslant i \leqslant n} X_{i,d} \leqslant x_d\right) \\
&= \Pr\left(X_{1,1} \leqslant x_1, \ldots, X_{n,1} \leqslant x_1, \ldots, X_{1,d} \leqslant x_d, \ldots, X_{n,d} \leqslant x_d\right),
\end{aligned}$$

where here and subsequently we use component-wise vector algebra. Reordering, we see that

$$\begin{aligned}\Pr\left(\boldsymbol{M}_n \leqslant \boldsymbol{x}\right) &= \Pr\left(X_{1,1} \leqslant x_1, \ldots, X_{1,d} \leqslant x_d, \ldots, X_{n,1} \leqslant x_1, \ldots, X_{n,d} \leqslant x_d\right) \\ &= \Pr\left(\boldsymbol{X}_1 \leqslant \boldsymbol{x}, \ldots, \boldsymbol{X}_n \leqslant \boldsymbol{x}\right).\end{aligned}$$

Recall that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \boldsymbol{X}_n \overset{\text{iid}}{\sim} F$, so we have

$$\begin{aligned}\Pr\left(\boldsymbol{M}_n \leqslant \boldsymbol{x}\right) &= \Pr\left(\boldsymbol{X}_1 \leqslant \boldsymbol{x}, \ldots, \boldsymbol{X}_n \leqslant \boldsymbol{x}\right) \\ &= \Pr\left(\boldsymbol{X}_1 \leqslant \boldsymbol{x}\right) \times \cdots \times \Pr\left(\boldsymbol{X}_n \leqslant \boldsymbol{x}\right) \\ &= F^n(\boldsymbol{x}).\end{aligned}$$

Since we are interested in the extremal behaviour of the component-wise maxima $\boldsymbol{M}_n$ as $n \to \infty$, we consider

$$\lim_{n\to\infty} \Pr(\boldsymbol{M}_n \leqslant \boldsymbol{x}) = \lim_{n\to\infty} F^n(\boldsymbol{x}). \tag{2.12}$$

Due to similar arguments presented in the univariate case (Section 2.1.1), it is clear that the R.H.S. of the expression (2.12) will collapse upon a single mass as $n \to \infty$. To avoid this, we apply a general linear scaling to each element of $\boldsymbol{M}_n$. So, we have

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leqslant \boldsymbol{x}\right) &= \lim_{n\to\infty} \Pr\left(\boldsymbol{M}_n \leqslant \boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n\right) \\ &= \lim_{n\to\infty} F^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n).\end{aligned} \tag{2.13}$$

If there exists a sequence of linear scaling constants $\boldsymbol{a}_n > \boldsymbol{0},\ \boldsymbol{b}_n \in \mathbb{R}^d$, such that (2.13) has a limiting distribution function

$$F^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}),\ \text{as } n \to \infty, \tag{2.14}$$

which is non-degenerate in each margin, then G is called a multivariate extreme value distribution function. Furthermore, since the $\boldsymbol{X}_i$ are independently distributed, univariate extreme value results presented in Section 2.1.1 apply to each element of $\boldsymbol{M}_n$ in turn and so the margins of G must belong to the GEV (2.3) distribution family. We say that the distribution F in (2.14) is in the maximum domain of attraction of G.

Without loss of generality, we can assume that the marginal distributions of F are standard Fréchet distributed (if they belong to a different distribution within the GEV they can be transformed to Fréchet through probability integral transform as in Section 2.2.2). The appropriate linear scaling of a Fréchet random variable such that its sample maxima are non-degenerate as $n \to \infty$ is found by setting $\boldsymbol{a}_n = n\mathbf{1} \in \mathbb{R}^d$ and $\boldsymbol{b}_n = \mathbf{0} \in \mathbb{R}^d$ (see example 3.2, Coles 2001). Assuming standard Fréchet margins gives a convenient expression of (2.14), that is,

$$\Pr\left(n^{-1}\boldsymbol{M}_n \leqslant \boldsymbol{x}\right) = F^n(n\boldsymbol{x}) \to G(\boldsymbol{x}), \text{ as } n \to \infty,$$

and the multivariate extreme value distribution G has the form

$$G(\boldsymbol{x}) = \exp\{-V(\boldsymbol{x})\}, \tag{2.15}$$

where V is typically referred to as the exponent measure of G (a derivation of this expression will be given in Section 2.2.6). Since marginal distributions are standard Fréchet distributed, for all $i = 1, \ldots, d$ we have,

$$G(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty) = \exp\left(-\frac{1}{x_i}\right), \tag{2.16}$$

so $\Lambda(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty) = x_i^{-1}$. Furthermore, G is max-stable by its derivation, i.e., $G^m(mx) = G(x)$ for all $x \in \mathbb{R}_+^d$ and all $m > 0$. Max-stability of G implies that V is homogeneous of order -1, that means, for any $m \in \mathbb{R}$, with $m > 0$,

$$V(m\boldsymbol{x}) = m^{-1}V(\boldsymbol{x}),$$

for all $x \in \mathbb{R}_+^d$. A distribution is max-stable if it is in its own max-domain of attraction. That is, if G is max-stable, the appropriately scaled sample maxima from G also have distribution G with a deterministic linear adjustment.

The exponent measure V has the form

$$V(\boldsymbol{y}) = d\int_{S_d}\left\{\max_{i=1,\ldots,d}\left(\frac{\omega_i}{y_i}\right)\right\} dH(\omega_1, \ldots, \omega_d) \tag{2.17}$$

for $\boldsymbol{y} = (y_1, \ldots, y_d) \in \mathbb{R}_+^d$, where H is called the spectral measure and is a distribu-

tion function defined on the $(d-1)$-dimensional simplex S_d,

$$S_d = \left\{ [w_1, w_2, \ldots, w_d] \in [0,1]^d \ : \ \sum_{i=1}^{d} w_i = 1 \right\}.$$

Illustrations of S_d are given in Figure 2.7.

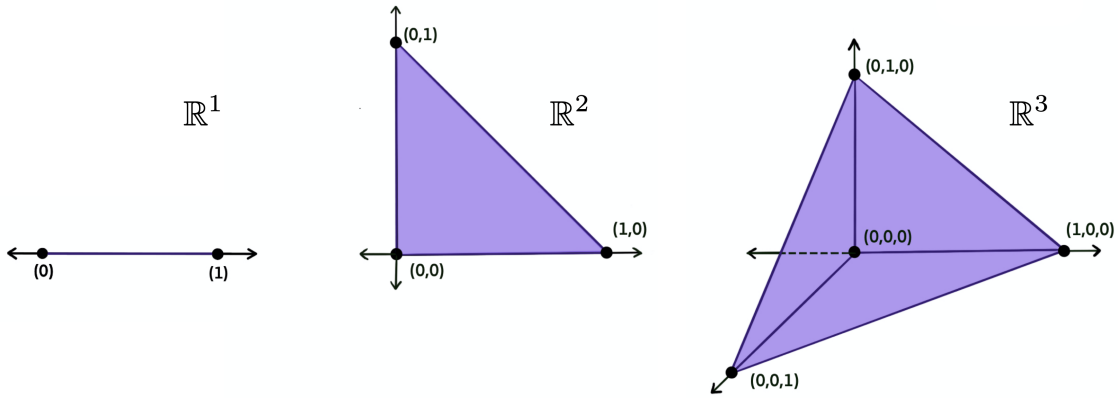

Figure 2.7: Diagrams of the simplexes S_2, S_3, and S_4

To ensure that H is a valid distribution function, it must satisfy

$$\int_{S_d} dH\left(\boldsymbol{\omega}\right) = 1 \tag{2.18}$$

and from the marginal condition (2.16) it follows that

$$\int_{S_d} \omega_i dH\left(\boldsymbol{\omega}\right) = 1/d, \tag{2.19}$$

for $i = 1, 2, \ldots, d$ (Pickands 1981). Consequently, (2.19) implies that the univariate marginals of H all have a mean measure of $1/d$. We use the notation dH to include the cases where H is not differentiable, i.e., in the case of complete dependence or independence. In the case of complete independence, H will have all its mass on the nodes of the simplex, in the case of complete dependence all mass will be placed on the centre of the simplex and V is a discrete measure. If the variables are strongly dependent then H will have a lot of mass on the interior of the simplex. If the variables are weakly dependent H will have a lot of mass in the boundaries of the simplex (Coles & Tawn 1991). If H is differentiable with density h, then (2.17),

becomes

$$V(\boldsymbol{y}) = d \int_{S_d} \left\{ \max_{i=1,\ldots,d} \left(\frac{\omega_i}{y_i} \right) \right\} h(\omega_1, \ldots, \omega_d)\, d\omega_1, \ldots, d\omega_d.$$

In the following sections, we will discuss derivations and the development of parametric forms of the exponent measure V and subsequently the spectral density h.

2.2.5 Parametric forms for G

There are uncountably infinite possible distribution functions which satisfy (2.18) and (2.19). This means, as mentioned above, there is no prescribed set of distributions as is the case in univariate extreme value modelling. However, there is a collection of popular extreme value distribution functions. Coles & Tawn (1991) give further links between the density H and exponent measure V, allowing for the development of parametric models. Specifically, they show that

$$\frac{\partial^d V(\boldsymbol{x})}{\partial x_1, \ldots, \partial x_d} = -\frac{d}{\|\boldsymbol{x}\|^{d+1}} h\left(\boldsymbol{x}/\|\boldsymbol{x}\|\right), \tag{2.20}$$

where $\|\boldsymbol{x}\| = \sum_{i=1}^{d} x_i$ for $\boldsymbol{x} \in \mathbb{R}_+^d$. Using (2.20), and through the specification of h, satisfying (2.18) and (2.19), we can derive parametric models for V and in turn, multivariate extreme value models G.

An early example of a parametric bivariate extreme value distribution is the logistic model (Gumbel 1960), which has, for a parameter $0 < \alpha \leqslant 1$, the form

$$G(x_1, x_2) = \exp\left\{ -\left(x_1^{-1/\alpha} + x_2^{-1/\alpha} \right)^{\alpha} \right\}$$

with the associated spectral density function recovered using (2.20),

$$h(\omega) = \frac{1}{2}\left(\alpha^{-1} - 1\right) \{\omega(1-\omega)\}^{-1-1/\alpha} \left\{ \omega^{-1/\alpha} + (1-\omega)^{-1/\alpha} \right\}^{\alpha-2}.$$

where $0 \leqslant \omega \leqslant 1$. The logistic model is flexible in that, complete dependence and independence are special cases as $\alpha \to 0^+$ and $\alpha \to 1^-$ respectively (Tawn 1988).

Another example of a popular bivariate extreme value model is given by Hüsler &

Reiss (1989) which arises as the maxima of linearly normalised Gaussian vectors where the correlation in pair n tends to 1 as $n \to \infty$ at a suitable rate to avoid independence or perfect dependence. The bivariate distribution function with parameter $\lambda \in [0, \infty)$ is given by

$$G(x_1, x_2) = \exp\left\{-\frac{1}{x_1}\Phi\left(\lambda + \frac{1}{2\lambda}\log\frac{x_2}{x_1}\right) - \frac{1}{x_2}\Phi\left(\lambda + \frac{1}{2\lambda}\log\frac{x_1}{x_2}\right)\right\}, \quad (2.21)$$

for $(x_1, x_2) \in \mathbb{R}^2_+$, with spectral density function

$$h(\omega) = \frac{\lambda}{2\omega(1-\omega)^2}\phi\left(\frac{1}{\lambda} + \frac{\lambda}{2}\ln\left[\frac{\omega}{1-\omega}\right]\right),$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard Gaussian distribution and density functions respectively. The dependence structure of the bivariate Hüsler-Reiss distribution is determined entirely by the parameter λ. Complete dependence corresponds to the case where $\lambda \to 0$, and complete independence where $\lambda \to \infty$.

Note that, in the copula literature, it is conventional to use an alternative parameterisation to the original distribution described by Hüsler & Reiss (1989) in (2.21) which instead takes the inverse of the dependence parameter (Joe 1994). Adopting this convention, we subsequently use $\lambda = 1/\lambda$ hereafter. In this case, complete dependence corresponds to the case where $\lambda \to \infty$, and complete independence where $\lambda \to 0$. Figure 2.8 shows a sample of points from a bivariate Hüsler-Reiss distribution at increasing levels of dependence from left to right, with the associated spectral density function below. We provide further discussion of the Hüsler-Reiss distribution, including treatment of the higher dimensional cases, in Chapter 3.

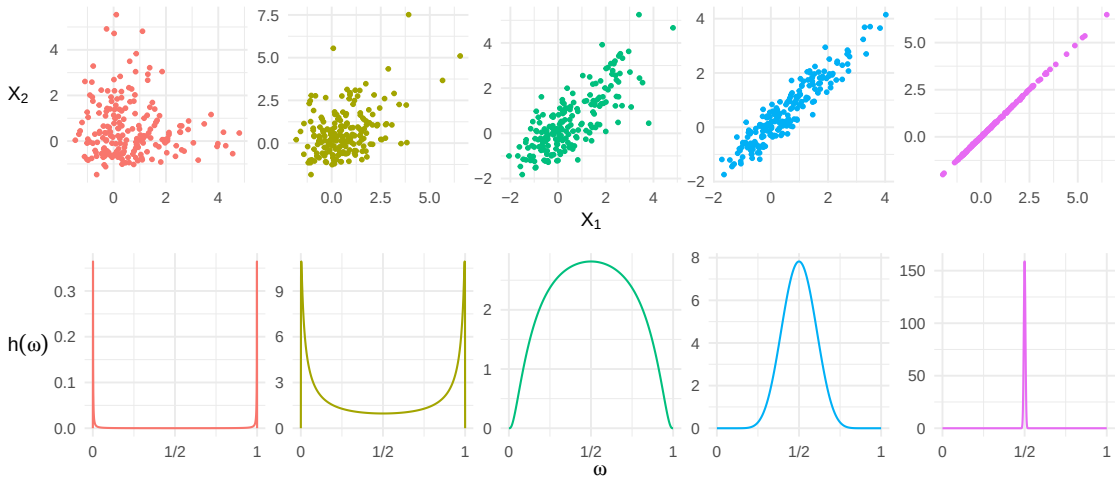

Figure 2.8: Sample of points from a bivariate Hüsler-Reiss extreme value distribution (top row) at increasing levels of extremal dependence from left to right with $\lambda = 0.25, 1, 2, 5, 100$. The marginal distributions are GEV(0,1,0). Spectral density of the Hüsler-Reiss distribution (bottom row) from which the points in the respective plots above of the same colour are sampled.

2.2.6 Point process representation

To see the expression of the exponent measure, V in (2.17), we first define the pseudo-polar decomposition of a vector $\boldsymbol{x} = (x_1, \ldots, x_d)$ using the function $T : \mathbb{R}_+^d \to (0, \infty) \times S_d$,

$$T(x_1, \ldots, x_d) = \left\{ \sum_{j=1}^{d} x_j, \left(\frac{x_1}{\sum_{j=1}^{d} x_j}, \ldots, \frac{x_{d-1}}{\sum_{j=1}^{d} x_j} \right) \right\}. \tag{2.22}$$

Suppose that $\boldsymbol{X}$ has standard Fréchet or Pareto(1) margins, we take $T(\boldsymbol{X}) = (\|\boldsymbol{X}\|, \boldsymbol{X}/\|\boldsymbol{X}\|) = (R, \boldsymbol{W})$ where the variable R describes the radial measure or magnitude of an observation and W represents the angular or dependence component of an event. This decomposition is illustrated in the bivariate case in the left-hand panel of Figure 2.9.

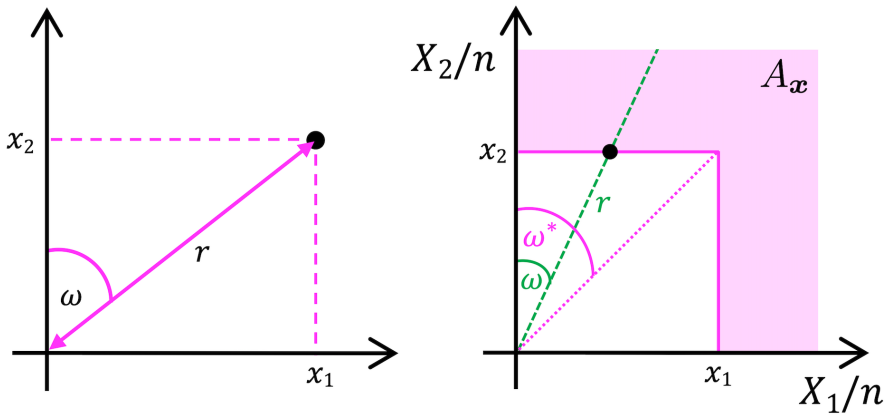

Figure 2.9: Pseudo-polar decomposition (left) of a point $(x_1, x_2) \in \mathbb{R}^2$ into radial components $(r, \omega) = T(x_1, x_2)$ where T is defined in expression (2.22). Illustration of the set $A_{\boldsymbol{x}}$ (right).

As in the univariate case, the point process setting allows for a convenient and intuitive derivation of multivariate extreme value models. Without loss of generality assume that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} F$, where F has standard Fréchet marginal distributions. Consider the d-dimensional point process

$$\mathcal{P}_n = \left\{ \left(\frac{X_{i,1}}{n}, \ldots, \frac{X_{i,d}}{n} \right), i = 1, \ldots, n \right\}.$$

The point process $\mathcal{P}_n$ converges to a non-homogeneous Poisson process (NHPP) $\mathcal{P}$ on $\mathbb{R}_+^d \backslash \{\mathbf{0}\}$ as $n \to \infty$, with intensity measure $\lambda(dr \times d\boldsymbol{\omega}) = dr^{-2} dr dH(\boldsymbol{\omega})$ (Coles & Tawn 1991).

Reusing the notation set out in Section 2.1.2, define the number of points in any subset $\mathcal{A} \subset \mathbb{R}^d$, as $N_n(\mathcal{A})$, the expected number of these points in $\mathcal{A}$ as $\mathbb{E}\{N_n(\mathcal{A})\}$, and $\Lambda(\mathcal{A}) := \lim_{n\to\infty} \mathbb{E}\{N_n(\mathcal{A})\}$. The probability of observing j points in $\mathcal{A}$ is

$$\lim_{n\to\infty} \Pr\{N_n(\mathcal{A}) = j\} = \frac{\{\Lambda(\mathcal{A})\}^j}{j!} e^{-\Lambda(\mathcal{A})}. \tag{2.23}$$

Now consider restricting the set $\mathcal{A}$ to be the positive orthant bounded away from the origin, i.e., consider the set $A_{\boldsymbol{x}} = (0, \infty)^d \backslash \{(0, x_1) \times (0, x_2) \times \cdots \times (0, x_d)\}$, where $x_i > 0$ for $i = 1, \ldots, d$. See the right-hand panel in Figure 2.9 for an illustrated bivariate example of the set $A_{\boldsymbol{x}}$. We can consider the number of points in this set, $N_n(A_{\boldsymbol{x}})$, as $n \to \infty$. As the sample size gets larger, the values of $x_1, x_2, \ldots, x_d$

correspond to increasing values of $X_1, \ldots, X_d$ as each is only exceeded if $X_i/n > x_i$. That means

$$\begin{aligned} \Pr(n^{-1}\boldsymbol{M}_n \leqslant \boldsymbol{x}) &= \Pr(n^{-1}M_{n,1} \leqslant x_1, n^{-1}M_{n,2} \leqslant x_2, \ldots, n^{-1}M_{n,d} \leqslant x_d) \\ &= \Pr\left\{N_n(A_{\boldsymbol{x}}) = 0\right\}, \end{aligned} \tag{2.24}$$

where $\boldsymbol{M}_n$ is the vector of component-wise maxima $(M_{n,k}; k = 1, \ldots, d)$. Combining (2.23) and (2.24) we see that, as $n \to \infty$

$$\Pr\left\{N_n(A_{\boldsymbol{x}}) = 0\right\} = \frac{\{\Lambda(A_{\boldsymbol{x}})\}^0}{0!} e^{-\Lambda(A_{\boldsymbol{x}})} = \exp\{-\Lambda(A_{\boldsymbol{x}})\} = G(\boldsymbol{x}).$$

To arrive at our earlier definition of the multivariate extreme value distributions in (2.15) we now need to derive the form of $\Lambda(A_{\boldsymbol{x}})$. Taking the integrated intensity function of the NHPP (derived in Appendix A.1)

$$\lambda(r, \boldsymbol{\omega}) dr d\boldsymbol{\omega} = d\frac{dr}{r^2} dH(\boldsymbol{\omega})$$

and applying relation (2.5), we see that for an arbitrary set $\mathcal{A} \in \mathbb{R}_+^d \backslash \{0\}$, the integrated intensity over $\mathcal{A}$ is given by

$$\Lambda(\mathcal{A}) = d \int_{\mathcal{A}} \frac{dr}{r^2} dH(\boldsymbol{\omega}). \tag{2.25}$$

So, in our bivariate illustration, $d = 2$ and $\boldsymbol{\omega} = \omega$.

Integrating over the set $A_{\boldsymbol{x}}$ can be achieved by integrating over the dashed green line for r, where it intersects the set $A_{\boldsymbol{x}}$ for all angles $\omega \in [0, 1]$. We have that $r = x_1 + x_2$ and so $x_1 = r\omega$ and $x_2 = r(1 - \omega)$. For $\omega \leqslant \omega^*$, we have that $r = x_2/(1 - \omega)$ and for $\omega > \omega^*$, $r = x_1/\omega$. The appropriate form is easily found by taking $\min\{x_2/(1 - \omega), x_1/\omega\}$. Formally, this means,

$$\begin{aligned}\Lambda(A_{\boldsymbol{x}}) =& 2\int_0^1 \int_{\min\left(\frac{x_2}{1-\omega}, \frac{x_1}{\omega}\right)}^{\infty} r^{-2} dr dH(\omega) \\ =& 2\int_0^1 \frac{1}{\min\left(\frac{x_2}{1-\omega}, \frac{x_1}{\omega}\right)} dH(\omega) \\ =& 2\int_0^1 \max\left(\frac{1-\omega}{x_2}, \frac{\omega}{x_1}\right) dH(\omega).\end{aligned}$$

Extended to the d-dimensional setting, this becomes

$$\Lambda(A_{\boldsymbol{x}}) = d\int_{S_d} \left\{\max_{i=1,\ldots,d}\left(\frac{\omega_i}{x_i}\right)\right\} dH(\boldsymbol{\omega}),$$

which gives the exponent measure $V(\boldsymbol{x})$ as defined in (2.17).

2.2.7 Multivariate regular variation

From Resnick (1987), given a measurable subset of the unit simplex, $B \subseteq S_d$, and spectral measure H defined on S_d (as in expression (2.17)), multivariate regular variation for random variables with unit Fréchet marginals tells us that for $r \geqslant 1$ and $u > 0$,

$$\lim_{n\to\infty} \Pr(n^{-1}R > ur, \boldsymbol{W} \in B \mid n^{-1}R > u) = r^{-1}H(B). \tag{2.26}$$

If $\boldsymbol{X}$ is regularly varying, then $\boldsymbol{X}$ is in the maximum domain of attraction of a multivariate extreme value distribution (de Haan 1984). Limit (2.26) tells us that R (the magnitude of an observation $\boldsymbol{X}$) and $\boldsymbol{W}$ (the extremal dependence structure of the observation $\boldsymbol{X}$) are independent as the radial variable tends to infinity, i.e., as $n \to \infty$. Notice how this parallels the factorisation of λ in Section 2.2.6. It is reasonable to assume that, for the most extreme events ($R/n > u$), the magnitude of the event and the extremal dependence structure are independent if u is taken to be large enough. In practice, this allows the practitioner to separate the statistical modelling of the dependence structure ($\boldsymbol{W}$) and the magnitude (R) of extremal observations, and most critically, justifiably extrapolate to higher magnitudes.

To see how the limit (2.26) arises, let A_u be the set bounded away from the origin

by the hyperplane $n^{-1}R = \|\boldsymbol{X}\| = u$. The set A_u in the bivariate setting is illustrated in the left-hand panel of Figure 2.10 and corresponds to the conditioning set of expression (2.26) (i.e., $\{n^{-1}R > u\}$). Let $A_{ur,B}$ be the set bounded away from the origin by the hyperplane $n^{-1}R = \|\boldsymbol{X}\| = ru$, where $r > 1$, and with $B = \{\boldsymbol{X} : \boldsymbol{X}/\|\boldsymbol{X}\| \leqslant \omega^*\}$ for $0 < \omega^* < 1$ as illustrated in the right-hand sketch in Figure 2.10 for the bivariate setting, i.e., B is the set of $W \in [0, \omega^*]$. We can see that the set $A_{ur,B}$ corresponds to the conditioned set in expression (2.26), (i.e., $\{n^{-1}R > ur, \boldsymbol{W} \in B\}$). We then have, as $n \to \infty$,

$$\begin{aligned} \Pr(n^{-1}R > ur, \boldsymbol{W} \in B \mid n^{-1}R > u) &= \Pr(n^{-1}\boldsymbol{X} \in A_{ur,B} | n^{-1}\boldsymbol{X} \in A_u) \\ &\to \Lambda(A_{ur,B})/\Lambda(A_u). \end{aligned} \tag{2.27}$$

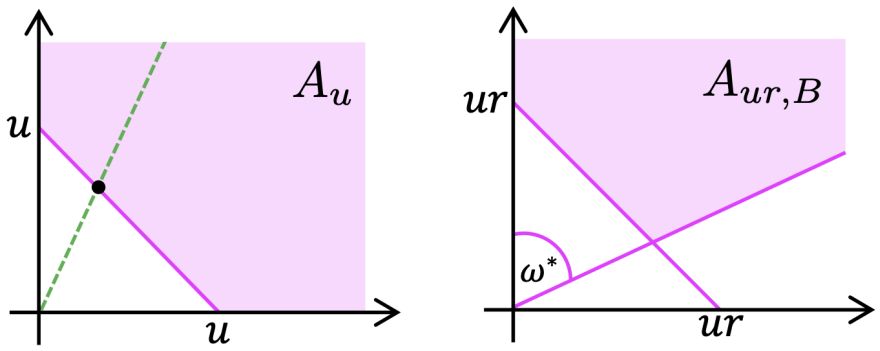

Figure 2.10: Sketch of extremal set A_u shaded (left) and extremal set $A_{ur,B}$ shaded (right).

We calculate both $\Lambda(A_u)$ and $\Lambda(A_{ur,B})$ using expression (2.25), while again making use of the pseudo-polar decomposition described in expression (2.22). Similar to the derivation of the exponent measure $V(\boldsymbol{x}) = \Lambda(A_{\boldsymbol{x}})$, in Section 2.2.6, we achieve integration over the sets A_u and $A_{ur,B}$ by integrating over the radial r (shown as the dashed green line in Figure 2.10), where it intersects the extremal sets for all $\omega \in [0, 1]$. We therefore have,

$$\Lambda(A_u) = 2\int_0^1 \int_u^\infty \frac{dr}{r^2} dH(\omega) = 2\int_0^1 \frac{1}{u} dH(\omega) = \frac{2}{u}, \tag{2.28}$$

and

$$\Lambda(A_{ur,B}) = 2\int_0^{\omega^*}\int_{ur}^{\infty}\frac{dr}{r^2}dH(\omega) = 2\int_0^{\omega^*}\frac{1}{ur}dH(\omega) = \frac{2}{ur}H(B), \qquad (2.29)$$

since $B = [0, \omega^*]$. Finally, plugging (2.28) and (2.29) into expression (2.27), we get

$$\lim_{n\to\infty}\Pr(R/n > ur, W \in B \mid R/n > u) = \lim_{u\to\infty}\frac{2(ur)^{-1}H(B)}{2u^{-1}} = r^{-1}H(B),$$

arriving at the expression for multivariate regular variation stated in (2.26). Finally notice that as the limit does not depend on u, we can combine n and u as $t = nu$ and get

$$\lim_{t\to\infty}\Pr(R > tr, \boldsymbol{W} \in B \mid R > t) = r^{-1}H(B),$$

the usual form of how multivariate regular variation is expressed.

Figure 2.11 demonstrates the behaviour of the angular density or spectral density h of H, of bivariate extreme observations when transformed to Fréchet margins at different levels of extremal dependence. The top row shows three illustrative data sets and the middle row shows the same data transformed to Fréchet margins. Considering each column, from left to right we see increasing extremal dependence in the data. Notice that after standardising the margins to a heavy-tailed distribution, only the extremal observations "survive" away from the origin (for illustrative purposes in the figures presented here, we take the empirical 90-th percentile of L_1-norms of observations as the threshold for an observation being extreme (i.e., $R > r_{0.9}$), a theoretical justification of choosing such a threshold is later given in Section 2.2.7). Only points that are not extreme in both dimensions are pulled towards the axis. The estimated spectral density (bottom row), given as a kernel density estimate, on S_1 of the observations of $W \mid R > r_{0.9}$ has increasing mass in the centre of the simplex at (1/2) and less at the boundaries as we move to stronger dependence cases.

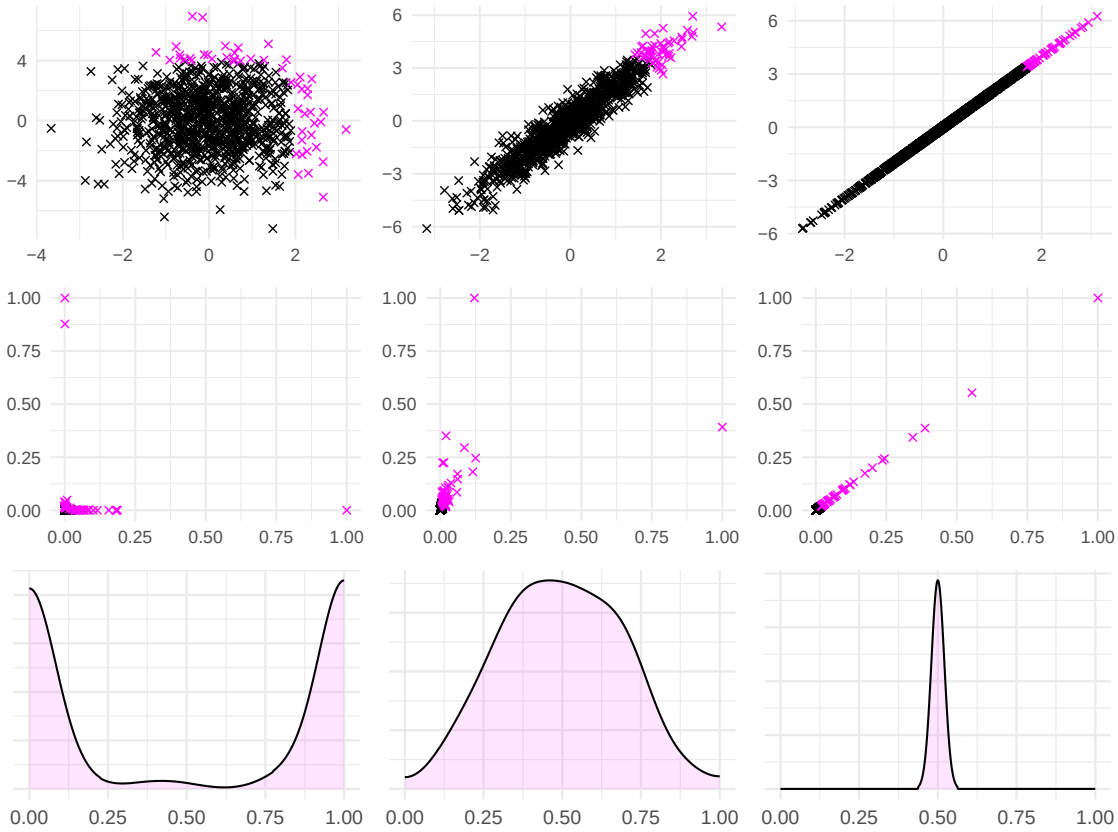

Figure 2.11: The three columns in the figure from left to right showcase three different dependence levels; independence ($\rho = 0$), strong dependence ($\rho = 0.95$) and perfect dependence ($\rho = 1$). Simulations from a bivariate Gaussian distribution (top row). Events with extreme magnitude (above the empirical 90-th percentile of L_1-norms) are highlighted. Data from the respective plots above transformed to have (scaled) standard Fréchet margins (middle row). Kernel density estimate of the angular density of the extremal points in the plot above (bottom row).

An empirical, illustrative example of the spectral density extended to tri-variate random variables (X_1, X_2, X_3) is given in Figure 2.12. This figure can be interpreted as the equivalent of the density in the bottom row of Figure 2.11 where now we take the angular density from a 3-dimensional distribution. The distribution on the simplex is representative of the pairwise extremal dependence of three random variables, X_1, X_2, and X_3. Mass at the lower left node indicates extremal behaviour of variable X_1; mass at the upper left and lower right nodes indicate the equivalent for X_2 and X_3, respectively. Mass between two nodes indicates extremal dependence between those nodes. The first plot in this figure represents data with low extremal

dependence, notice that spectral density has most mass at the nodes of the simplex and very little between nodes. The second plot shows high extremal dependence among all three random variables, resulting in most density being at the centre of the simplex. The third plot shows an asymmetric spectral density depicting unequal extremal pairwise dependence among the random variable, with strong extremal dependence between variables X_1 and X_3 and low extremal dependence between X_2 and both X_1 and X_3. Notice that, in general, the observations of $\boldsymbol{W} \mid R > r_{0.9}$ have increasing mass towards the centre of the simplex at $(1/d, \ldots, 1/d)$ where $d = 3$ and less at the boundaries as we move to stronger dependence cases. In general, if two variables are asymptotically independent all mass in the simplex between their respective nodes will be on those nodes. If there is any mass between the two nodes, then the variables are asymptotically dependent.

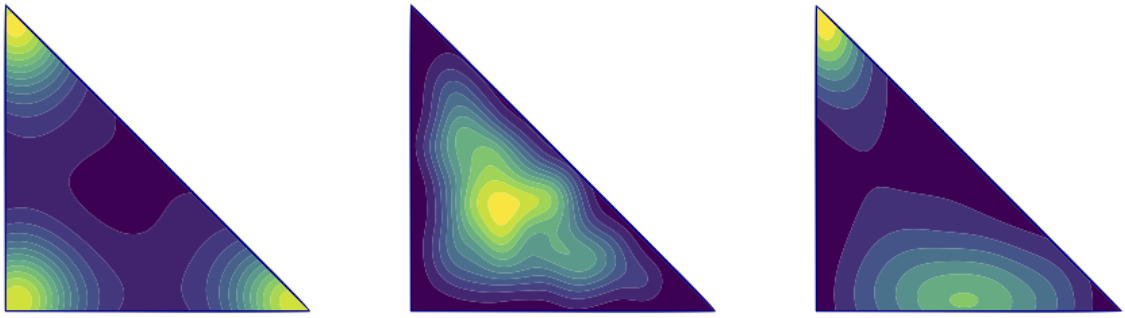

Figure 2.12: Spectral density of simulated random variables X_1, X_2 and X_3 from 3-dimensional Hüsler-Reiss distribution. In each simplex, the lower left corner corresponds to X_1, the upper left corner corresponds to X_2 and the lower right, X_3. Lighter colours correspond to higher density.

2.2.8 Copulae

A copula is simply a multivariate distribution function with uniform marginal distributions on $[0, 1]$. The copula function, $C : [0, 1]^d \to [0, 1]$, of the variables $(X_1, X_2, \ldots, X_d)$ is the multivariate distribution of these variables after being transformed to a uniform distribution,

$$C(u_1, u_2, \ldots, u_d) = \Pr\{F_1(X_1) \leqslant u_1, F_2(X_2) \leqslant u_2, \ldots, F_n(X_d) \leqslant u_d\} \tag{2.30}$$

where F_i is the distribution function of X_i. From Sklar (1959), if the random variables $(X_1, X_2, \ldots, X_d)$ have a joint distribution F, with continuous marginal

distributions F_i, then the copula function in (2.30) is unique. Since the copula function is unique, it is invariant to marginal transformations. That means we can transform the margins to any convenient distribution without disrupting the dependence structure between them. For example, using (2.11) we can transform the copula to have standard Fréchet margins,

$$C\left[-1/\log F_1(x_1), \ldots, -1/\log F_d(x_d)\right]$$

to focus attention on the dependence in the tails of the distribution.

Copulas offer a convenient way to describe the relationship between any set of random variables whose marginal distributions are known since any random variable can be transformed to have uniform distribution through the probability integral transform. Any copula function that satisfies max-stability is a valid multivariate extreme value distribution (Gudendorf & Segers 2010). Copulas are a very popular modelling tool in multivariate extreme value theory as they allow the practitioner to deal with the modelling of the marginal and multivariate components of a collection of variables separately.

Joint parameter estimation of the marginal distributions and dependence structure is also possible. Suppose we have a collection of *iid* bivariate continuous random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ from a distribution F and copula function C. Suppose the margins have distribution functions F_1 and F_2. The bivariate density f of F can then be expressed as

$$f(x_1, x_2) = c\left[F_1(x_1), F_2(x_2)\right] \times f_1(x_1) \times f_2(x_2)$$

where c is the density function of the copula C and f_1 and f_2 are the densities of the marginal distributions F_1 and F_2 respectively. Following this we can derive the likelihood function of any parametric model which can be used for parameter estimation.

2.3 Spatial extremes

Extreme value processes are widely applied to spatial environmental applications. In this section, we discuss the extension of multivariate extreme value theory to an infinite-dimensional process setting.

2.3.1 Model-based geostatistics

The field of model-based geostatistics is concerned with modelling random variables that have some inherent spatial information. We define a spatial domain $\mathcal{S} \subset \mathbb{R}^d$, where for the majority of environmental applications $d = 2$. The domain $\mathcal{S}$ is made up of a set of locations, $\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots \in \mathcal{S}$, where typically $\boldsymbol{s}_i$ describes the spatial coordinate of a point, i.e., $\boldsymbol{s}_i = (\text{long}_i, \text{lat}_i)$. We define a spatial process $Y : \mathcal{S} \to \mathbb{R}$, which maps the spatial domain to some variable we are interested in understanding. For example, $Y(\boldsymbol{s}_i)$ could describe the altitude at location $\boldsymbol{s}_i = (\text{long}_i, \text{lat}_i)$. We define a spatio-temporal process as $Y : \mathbb{R} \times \mathcal{S} \to \mathbb{R}$, where Y in the spatial domain $\mathcal{S}$ takes different values over time $t \in \mathbb{R}$. Let $t_1, t_2, \cdots \in \mathbb{R}$ be a set of time points at which observations on Y are made. Then $Y(t_i, \boldsymbol{s}_j)$ describes the process at location $\boldsymbol{s}_j$ and time t_i. For example, $Y(t_i, \boldsymbol{s}_j)$ could describe the temperature at location $\boldsymbol{s}_j$ and time t_i. If the process $Y(t, \boldsymbol{s})$ is stationary in space (Diggle et al. 1998) then its statistical properties (e.g., expected value and variance) are identical at all locations over the spatial domain and the inter-site covariance depends only on the separation between the two locations. Furthermore, if the process is stationary in time, the statistical properties such as the variance and expected value of the process are equal at every time point t. A spatial process is isotropic if the relationship between the process at any two sites is a function of only the distance between them, i.e., not determined by this angle or the direction between them (Diggle et al. 1998).

2.3.1.1 Variogram

To reflect natural processes, we typically assume that the relationship between two locations is a function of the separation between them. A classic statistical tool for describing this feature of spatial processes is the variogram, $\gamma : \mathbb{R}^2 \to \mathbb{R}$. A variogram describes the correlation of two sites given their separation. For simplicity, we will assume isotropy throughout the remainder of this review, and so,

the variogram becomes a function of the distance between two sites. Given a pair of sites $\boldsymbol{s}_i, \boldsymbol{s}_j \in \mathcal{S}$, we denote the distance between them as $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$. The variogram for stationary isotropic processes is defined as

$$2\gamma(h) = \mathbb{E}\left[\{Y(\boldsymbol{s}_2) - Y(\boldsymbol{s}_1)\}^2\right]$$

or,

$$\gamma(h) = \frac{1}{2}\text{var}\left\{Y(\boldsymbol{s}_2) - Y(\boldsymbol{s}_1)\right\}, \tag{2.31}$$

where $\|\boldsymbol{s}_2 - \boldsymbol{s}_1\| = h$ (Diggle et al. 1998). Calculating $\gamma(h)$ for a range of distances provides valuable information on the level of spatial auto-correlation of the process Y over the domain.

There are many parametric forms of (2.31). One such example is the Matérn variogram,

$$\gamma(h) = \alpha\left\{1 - (2\sqrt{\nu}h/\phi)^{\nu}2^{1-\nu}\Gamma(\nu)^{-1}\text{K}_{\nu}(2\sqrt{\nu}h/\phi)\right\}, \tag{2.32}$$

where K_ν is a modified Bessel function of the second kind and the parameters $\alpha > 0$, $\phi > 0$, and $\nu > 0$ determine the sill/variance, scale/range, and smoothness, respectively (Banerjee et al. 2014). The sill, α, represents the maximum variability of the process and is approached asymptotically by the variogram as $h \to \infty$. The range, ϕ, characterises the distance at which the spatial auto-correlation becomes negligible or the distance at which the variance levels out to α. The smoothness parameter, ν, controls the rate of decay of spatial auto-correlation as the distance between locations increases. Different combinations of parameter values allow the Mátern variogram to capture both short-range and long-range spatial dependence. Figure 2.13 shows a Matérn variogram with $\alpha = 0.75$, $\phi = 1$, and increasing values of the smoothness parameter $\nu = \{0.1, 1, 2\}$ corresponding to slower convergence to α resulting in smoother spatial processes.

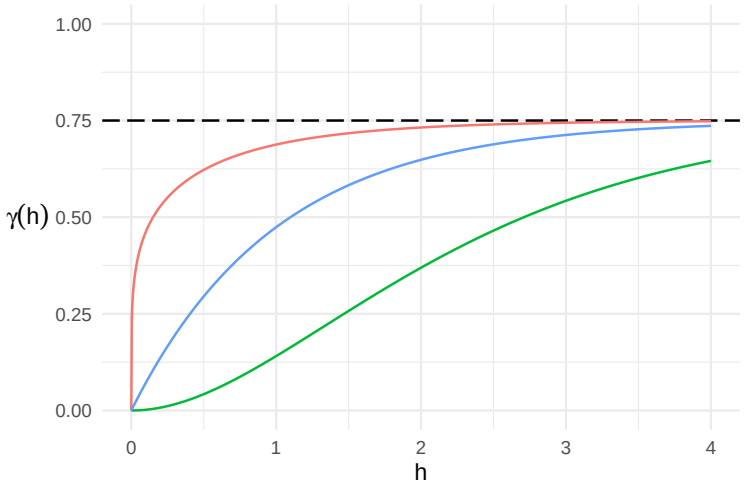

Figure 2.13: The Matérn variogram with $\alpha = 0.75$ (horizontal dashed line), $\phi = 1$, and different values of $\nu = \{0.1, 0.5, 2\}$ the smoothness parameter in red, blue, and green, respectively, corresponding to slower convergence to α, resulting in smoother spatial processes.

2.3.2 Max-stable process

A max-stable process extends the multivariate extreme value distributions to the spatial setting. Let $X(t, \boldsymbol{s})$ be *iid* observations from a stochastic process X, at discrete times $t = 1, 2, \ldots, n$ and over locations $\boldsymbol{s} \in \mathcal{S}$, where X is continuous in space and time. We take the component-wise maxima of the process, scaled to avoid margins collapsing, giving

$$Z_n(\boldsymbol{s}) = \max_{t=1,\ldots,n} \frac{X(t, \boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})}, \ \boldsymbol{s} \in \mathcal{S}, \tag{2.33}$$

where $a_n(\boldsymbol{s}) > 0$ and $b_n(\boldsymbol{s}) \in \mathbb{R}$. The spatio-temporal component-wise maximum corresponds to the maximum observed value over time at each site. In Figure 2.14 we give a sample of simulated events, with the process $Z_n(\boldsymbol{s})$ (consisting of component-wise maxima) highlighted. Notice that the component-wise maxima consist of more than one spatial observation.

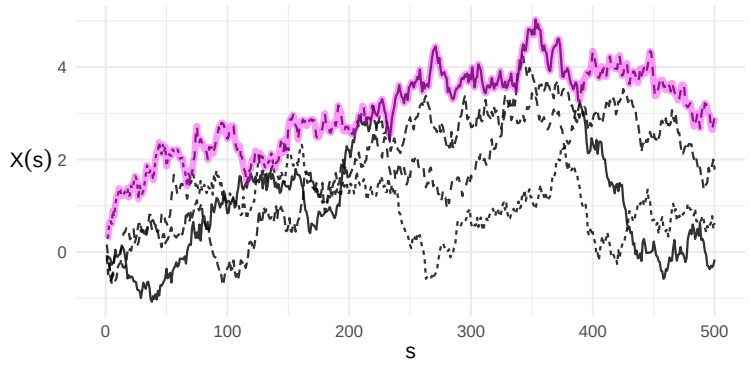

Figure 2.14: Illustrative simulated observations $\boldsymbol{X}_i(\boldsymbol{s})$ for $\boldsymbol{s} \in [0, 500]$ from a 1-dimensional spatial process at 500 locations. The component-wise maxima are highlighted.

To examine the extremal behaviour of this process we define,

$$Z(\boldsymbol{s}) = \lim_{n \to \infty} Z_n(\boldsymbol{s}).$$

If sequences $b_n(\boldsymbol{s})$ and $a_n(\boldsymbol{s})$ exist, such that $Z(\boldsymbol{s})$ has non-degenerate marginal distributions for all $\boldsymbol{s} \in \mathcal{S}$, then $Z(\boldsymbol{s})$ is called a max-stable process (Smith 1990). The marginal distributions of $Z(\boldsymbol{s})$ are max-stable and must belong to the GEV distribution family. Using equivalent arguments, as in Section 2.2.4, assuming that $X(t, \boldsymbol{s})$ has standard Fréchet marginal distributions, we take $b_n(\boldsymbol{s}) = 0$ and $a_n(\boldsymbol{s}) = n$. It follows that, given a collection of sites $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k \in \mathcal{S}$, we can describe the finite-dimensional joint distribution of $Z(\boldsymbol{s})$ at those sites as

$$\Pr\left\{Z\left(s_1\right) \leqslant z_1, \ldots, Z\left(s_k\right) \leqslant z_k\right\} = G(\boldsymbol{z}),$$

where G is a multivariate extreme value distribution, describing the extremal dependence between the process at sites $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k$.

Parametric models for max-stable processes are typically derived through the point process representation as described in Section 2.2.6. Let $\{R_i\}_{i \geq 1}$ be the points of a Poisson process on $[0, \infty)$ with intensity $r^{-2}dr$ and $\{W_i(\cdot)\}_i$ be *iid* replicates of a process on $\mathcal{S}$ with truncated mean 1 (i.e., $\mathbb{E}[\max\{0, W(\boldsymbol{s})\}] = 1$). Then

$$Z(\boldsymbol{s}) = \sup_{i \geq 1} \left\{R_i W_i(\boldsymbol{s})\right\}, \ \ \boldsymbol{s} \in \mathcal{S}$$

is a stationary max-stable process with unit Fréchet margins (Smith 1990). As discussed previously, the point process provides a physically interpretable representation of extreme environmental processes. Recall that R corresponds to the magnitude of events and W describes the spatial structure of the extremal dependence. Furthermore, $\max\{0, W(\boldsymbol{s})\}$ describes the contribution of site $\boldsymbol{s}$ to the extreme event. Through the appropriate specification of the random process W, we can derive parametric max-stable processes. For example, taking a probability density function f as any non-negative function such that $\int_{\mathbb{R}^d} f(\boldsymbol{s})d\boldsymbol{s} = 1$ and $\{\boldsymbol{Y}_i\}_{i\geq 1}$ to be the points of a homogeneous Poisson process in $\mathbb{R}^d$, we get the max-stable process, $\{Z(S) : s \in \mathbb{R}^d\}$, where

$$Z(\boldsymbol{s}) = \max_{i\geq 1} R_i f(\boldsymbol{s} - \boldsymbol{Y}_i).$$

This construction allows for a more meaningful interpretation where now f describes the spatial distribution or the profile of the extreme event and $\boldsymbol{Y}_i$ is the spatial "location" of the extreme event. As an illustrative example, suppose $Z(\boldsymbol{s})$ describes rain storms or heatwaves centred at $\boldsymbol{Y}_i$ and the function f captures the spatial extent of the event and how it falls off with increasing distance, while R describes the intensity of the storm or heatwave. Smith (1990) chooses f to be a multivariate normal distribution (MVN). The resulting model is called the Smith max-stable model or the "storm profile model", and has a bivariate distribution function for the process at sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ as

$$G(z_1, z_2) = \exp\left\{-\frac{1}{z_1}\Phi\left(\frac{\lambda}{2} + \frac{1}{\lambda}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{\lambda}{2} + \frac{1}{\lambda}\log\frac{z_1}{z_2}\right)\right\}, \quad (2.34)$$

where $\Phi(\cdot)$ is the standard normal distribution function. The Smith max-stable model is parameterised by $\lambda = \sqrt{h^T\Sigma^{-1}h}$, where h is the vector separation from site $\boldsymbol{s}_1$ to $\boldsymbol{s}_2$ and Σ is the covariance matrix of the Gaussian distribution controlling the level of extremal dependence between the two sites.

The assumption that extreme events are Gaussian (i.e., elliptical and smooth) is often restrictive and unrepresentative of natural processes. There have been several alternative model parameterisations to alleviate the determinism of the shape of the extreme event (e.g. Schlather 2002). One such popular model is the Brown-Resnick max-stable process. Taking $W(\boldsymbol{s}) = \exp(\epsilon(\boldsymbol{s}) - \gamma(\boldsymbol{s}))$ for $\boldsymbol{s} \in \mathcal{S}$, where ϵ is

a stationary Gaussian process and the variogram, γ has the form

$$\gamma(h) = \frac{1}{2}\text{var}\left\{W(\boldsymbol{s}) - W(\mathbf{0})\right\}.$$

where here, $h = \|\boldsymbol{s} - \mathbf{0}\|$. Then W is called a Brown-Resnick process, as described by Brown & Resnick (1977), which has a bivariate distribution function identical to the joint distribution in expression (2.34), where now $\lambda = \sqrt{2\gamma(h)}$. Engelke & Hitz (2020) show that the marginal distributions of the Brown-Resnick process are Hüsler-Reiss distributed, which can be seen in expression (2.21).

Figures 2.15 and 2.16 compare the general characteristics between the Smith and the Brown-Resnick max-stable processes. The left-hand panel in both figures give simulations in $\mathbb{R}^1$ (Fig. 2.15) and $\mathbb{R}^2$ (Fig. 2.16) respectively of the Smith max-stable process with characteristic Gaussian features. The right-hand panel of both figures shows the equivalent from the Brown-Resnick process, which has much more intricate, and less prescriptive features.

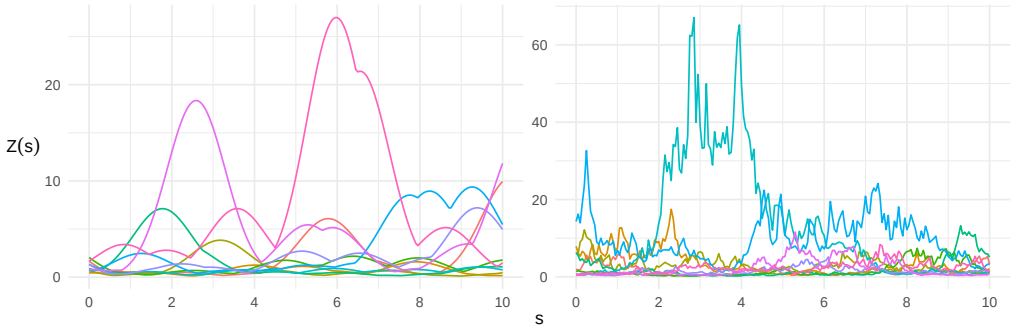

Figure 2.15: Collection of 10 simulations from a 1-dimensional Smith max-stable process with Σ having 1 on the diagonal and 0.5 on the off-diagonals (left), and simulations from a Brown-Resnick process with variogram $\gamma(h) = 0.5h^{1.5}$ (right).

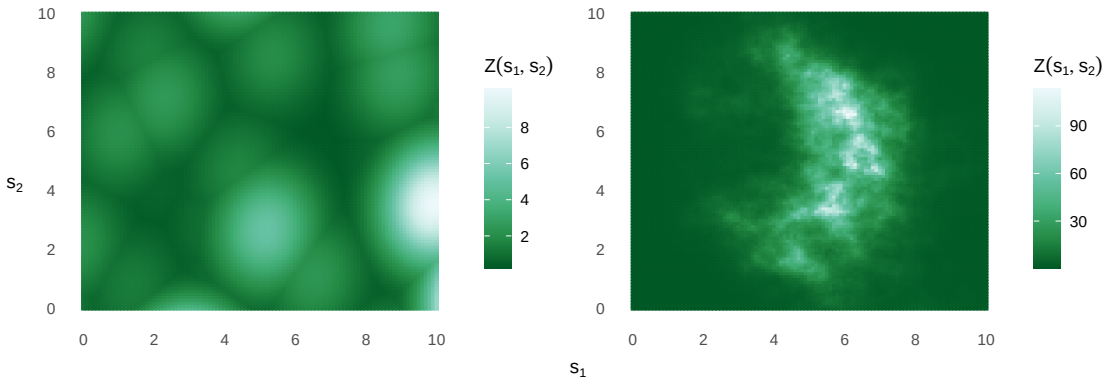

Figure 2.16: Simulations from a 2-dimensional max-stable process on a grid of $10,000$ locations over $[0, 10]^2$ with Σ having $[1, 2]$ on the diagonal and 0.1 on the off-diagonals (left), simulations from a Brown-Resnick process on the same grid with variogram, $\gamma(h) = 0.5h^{1.5}$ (right).

Analytical likelihood functions for max-stable processes are practically infeasible to write down for processes observed on even a moderate number of spatial locations. To deal with this, a composite likelihood consisting of bivariate pairwise likelihood functions is used (Padoan et al. 2010). Even so, calculating the composite likelihood is computationally expensive and quantifying parameter uncertainty is difficult.

2.3.3 Functional exceedances

Analogous issues arise between the univariate block maxima and the max-stable process modelling procedures (see Section 2.1.1.1). Namely, nuances of the extremal process being analysed are missed when the component-wise maxima is an amalgam of several independent extreme events. Summarising spatio-temporal data through block maxima is clearly not ideal as it can disrupt valuable physical information embedded in the data, as is often the case in environmental data. de Fondeville & Davison (2018) provide the example of modelling flood events resulting from different precipitation processes. Taking site-wise maxima may result in only modelling local and intense rain storms while disregarding widespread, cyclonic storms that appear less intense at a single site but of critical importance for widespread flood risk.

An alternative definition of extreme multivariate events was proposed by Coles & Tawn (1994), wherein d-dimensional observations are summarised and projected

onto a one-dimensional variable via a risk functional $r : \mathbb{R}^d \to \mathbb{R}$. The function r can be chosen to capture the combined effects of several variables that lead to an extreme event of interest, for example, the sum of rainfall over a collection of sites, $r(X) = \sum_{\boldsymbol{s}} X(\boldsymbol{s})$, where a large enough accumulation may lead to flooding. Projecting a series of multivariate observations onto a so-called "structure variable" immediately establishes a meaningful ordering. This approach transforms data to the univariate setting where well-established extreme value modelling techniques are available. For example, taking the risk of events that exceed a high enough threshold, i.e., the events $X \mid r(X) > v$, then $r(X) - v$, for large enough v, can be assumed to follow a generalised Pareto distribution (see Section 2.1.2). An appealing feature of this classification of events is the flexibility afforded to the practitioner of defining what extremal behaviours are of interest, i.e., to be able to choose the function r.

2.3.4 Generalised Pareto process

The method of Coles & Tawn (1994) described above does not tell us how individual variables of a process contribute to extreme events. To achieve this, Buishand et al. (2008) and Ferreira & de Haan (2014) consider the behaviour of the original process X conditioning on $\sup_{\boldsymbol{s}} X(\boldsymbol{s})$ being large. The authors show that, for $a_n(\boldsymbol{s})$ and $b_n(\boldsymbol{s})$ as in (2.33), the process

$$\left[1 + \xi(\boldsymbol{s}) \left\{ \frac{X(\boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})} \right\} \right]_+^{1/\xi(\boldsymbol{s})} \Bigg| \left(\sup_{\boldsymbol{s} \in \mathcal{S}} \left\{ \frac{X(\boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})} \right\} > 0 \right)$$

converges to a generalised Pareto process as $n \to \infty$, where $\xi(\boldsymbol{s})$ is the scale parameter of the process at site $\boldsymbol{s}$, describing the tail decay at that site. Transforming $X(\boldsymbol{s})$ to standardised margins $X^*(\boldsymbol{s})$, specifically Pareto(1), we see that

$$\left\{ u^{-1} X^*(\boldsymbol{s}) \,\middle|\, \sup_{\boldsymbol{s} \in \mathcal{S}} X^*(\boldsymbol{s}) > u \right\} \to Y^*(\boldsymbol{s}) \text{ as } u \to \infty, \tag{2.35}$$

where Y^* is a generalised Pareto process with standardised margins. Recall that, by construction of the max-stable process we have that the margins are GEV distributed. Since now we are taking the entire random field, and not component-wise maxima, the margins of the Pareto process do not necessarily belong to an extreme value distribution. However, we know that exceedances above a suitably

high threshold at each site $\boldsymbol{s}$ will follow the GPD.

Figure 2.17 shows a sample of simulated events over 1-dimensional space. In the left-hand panel, the highlighted simulation has the largest value over the observed domain, whereas the right-hand panel highlights the simulation with the largest mean value over the domain. This figure can be examined in contrast to Figure 2.14 to see how the generalised Pareto process retains the physical interpretation of an extreme event. By modelling functional exceedances we capture spatial patterns that one could expect to observe from the original process. Such a feature is incredibly valuable when working with climate data as a lot of information is embedded in the spatial behaviour of the phenomenon being studied. Since simulations from the resulting model are potentially physically realisable events they are immediately interpretable and can be used to inform impact models to detect possible damage to infrastructure (de Fondeville & Davison 2022).

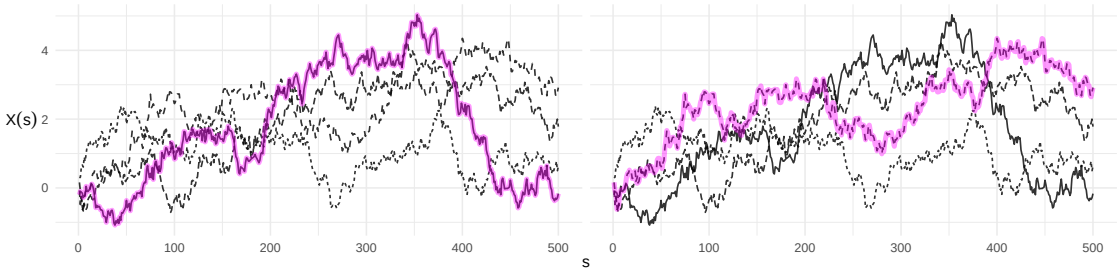

Figure 2.17: Simulation of one-dimensional spatial event, extreme observation highlighted. In the left-hand plot, the event with the largest observation is highlighted. In the right-hand plot, the event with the largest mean value is highlighted.

2.3.5 r-Pareto Processes

One disadvantage of conditioning on the level of the supremum of the process is that it requires the process to be observed over the entirety of the domain $\mathcal{S}$. Clearly, this is not always possible in certain applications, for example, weather observation stations which are located non-uniformly and not continuously throughout space and time. A relaxation of this restriction was introduced via the ℓ-Pareto processes (Dombry & Ribatet 2015), later called the r-Pareto processes (de Fondeville & Davison 2018); we will adopt the latter notation here. Akin to the approach of Coles & Tawn (1994) described above, the r-Pareto processes allow specification of a risk

function through which events are ordered. Effectively providing a generalisation of (2.35), allowing for a more flexible conditioning event by allowing a wider range of functions in place of the $\sup(\cdot)$ risk function, giving

$$\left\{ u^{-1}X^* \;\middle|\; r\left(X^*\right) > u \right\} \to \{Y^*(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\} \text{ as } u \to \infty, \tag{2.36}$$

where $X^* = \{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$. The limiting process is called the generalised r-Pareto process. Dombry & Ribatet (2015) show, under weak conditions on the process X^*, that $\{Y^*(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ is marginally non-degenerate in all margins. The level of the extremity of the stochastic process X^* is determined by a risk function $r(X^*) \in [0, \infty)$, where the only other constraint on r is that it is homogeneous of order 1, i.e., for any constant $c > 0$, then $r(c\boldsymbol{x}) = cr(\boldsymbol{x})$ for any $\boldsymbol{x}$ with non-negative components. Specification of the risk function r allows modelling tailored to the specific spatial nature of the process under consideration. Some useful functions which satisfy this restriction are the spatial maximum (or supremum), sum, mean, minimum over $\mathcal{S}$, or the cost at a particular site of interest. For example, in Figure 2.17, the right-hand side highlights the extreme event with the largest mean value over the observed events.

As in Section 2.2.7, $Y^* = \{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ can be decomposed into two independent stochastic components as follows:

$$Y(\boldsymbol{s}) = RW(\boldsymbol{s}) \; \forall \; \boldsymbol{s} \in \mathcal{S},$$

where $R \sim$ Pareto(1) and, $r(W) = 1$ where $W = \{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$. From (2.36), given an extremal set $\mathcal{A}$ which is a subset of $\{X_i^* : r(X^*) > 1\}$,

$$\Pr\left\{u^{-1}X^* \in \mathcal{A} \mid r(X^* > u)\right\} \to \Pr\{Y^* \in \mathcal{A}\}$$

as $u \to \infty$, where Y^* is an r-Pareto process. This result is analogous to the threshold stability property of the generalised Pareto distribution. This feature of threshold stability officiates extrapolation to yet unobserved extreme events.

If the limit provided in equation (2.36) is a good approximation for large u, then those spatial events with a risk function exceeding u will be well-approximated by an r-Pareto process. Similar to the construction of max-stable models in Section 2.3.2,

the choice of process W leads to many possible models. A popular choice is to take W to be a Brown-Resnick process, where inter-site extremal dependence is characterised entirely by a variogram $\gamma(h)$ (Engelke et al. 2015).

A

Appendix to Chapter 2

A.1 Deriving of intensity measure λ

We recall the fundamental theorem of calculus. Define $f(x)$ to be continuous function on the interval $[x_1, x_2]$ and $F(X)$ to be the anti-derivative of f, i.e., $F'(x) = f(x)$, then,

$$\int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1).$$

We have the integrated intensity function $r^{-1}H(\omega)$ and we wish to find $\lambda(r, \omega)$. So, we know that,

$$r^{-1}H(w) = \int_r^{\infty} \int_0^w \lambda(r', w')dw'dr'.$$

First, take the derivative of the left-hand side

$$\frac{\partial r^{-1}H(\omega)}{\partial r \partial \omega} = \frac{\partial r^{-1}h(\omega)}{\partial r} = -r^{-2}h(\omega).$$

Then, take the derivative of the right-hand side

$$\frac{\partial}{\partial r \partial \omega} \int_r^{\infty} \int_0^{\omega} \lambda(r', \omega')dw'dr' = \frac{\partial}{\partial r} \int_r^{\infty} \lambda(r', \omega)dr' = -\lambda(r, \omega).$$

Now, equating the R.H.S. and L.H.S. we get

$$\lambda(r, \omega) = r^{-2}h(\omega).$$

We express h as dH to highlight the fact that the derivative does not always exist for all $0 \leqslant \omega \leqslant 1$ in special cases. For example, in the case of complete dependence ($dH(1/2) = 1$), the intensity function is

$$\lambda(r, \omega) = r^{-2} dH(\omega),$$

or equivalently,

$$\lambda(r, \omega) dr d\omega = \frac{dr}{r^2} dH(\omega).$$

3

Extremal vectors of unequal dimensions with the Hüsler-Reiss distribution

In this chapter we explore properties of the finite-marginal distributions of the Brown-Resnick process, which which follow the Hüsler-Reiss distribution (HRD). The Brown-Resnick process is the basis of a popular parametric model for fitting r-Pareto processes in spatial extreme value theory. The HRD describes the asymptotic behaviour of suitably normalised normal vectors. We provide some theoretical background and illustrate its adaptability to describe the asymptotic behaviour of vectors which are not necessarily of equal dimension. We are motivated by the temporal inconsistencis in environmental synoptic observations. Typically, modelling a multivariate process across a number of sites requires observations mutual to each site at each time point. Such restrictions result in the practitioner disregarding time spans of data which are not present at all sites. Since extremal observations are typically few, omitting data is clearly not desirable. Our method enables us to develop a parametric r-Pareto process which can make use of all available extremal data.

3.1 Introduction

The default approach in standard spatial modelling is to standardise the marginal distributions of a process to be Gaussian and assume it is a realisation of a Gaussian process (Cressie 1993). This approach is not applicable in many extremal settings since the Gaussian process is asymptotically independent at spatial lags (Hüsler & Reiss 1989). The data analysed in this thesis exhibit a strong form of extremal dependence, called asymptotic dependence (see Section 2.2.3) which cannot be captured by Gaussian processes. Davison et al. (2012) provide thorough reviews of available approaches for the statistical modelling of spatial extreme events. Most notably, these include the copula (see Section 2.2.8) and max-stable process (see Section 2.3.2) modelling approaches. However, we are interested in modelling spatial threshold exceedances, which is made possible via the r-Pareto process modelling approach.

The r-Pareto process over the spatial domain $\mathcal{S} \subset \mathbb{R}^2$ is the limiting process of suitably normalised processes, $X^* = \{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$, with Pareto distributed marginal variables, whose risk (calculated by a suitable homogeneous order 1 function r) exceeds a high threshold, u. Specifically,

$$\left\{ u^{-1} X^* \;\middle|\; r\left(X^*\right) > u \right\} \to \{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\} \text{ as } u \to \infty,$$

where Y is the limiting r-Pareto process (Dombry & Ribatet 2015). In practice, spatial events whose risk, r, exceeds a sufficiently high threshold, u, will be well-approximated by an r-Pareto process. A restriction of this modelling approach is that it assumes asymptotic dependence over the entire domain, however for many environmental processes over a small region, this is often a sound assumption. A full treatment of r-Pareto theory will be provided in Chapter 4. Suffice to say, Y can be expressed as two stochastically independent components; that is $Y(\boldsymbol{s}) = RW(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathcal{S}$. The component R is simply a univariate heavy-tailed random variable (e.g., Pareto). However, $W^* \coloneqq \{W(\boldsymbol{s}); \boldsymbol{s} \in \mathcal{S}\}$ is a stochastic process which describes the extremal dependence between sites, or the contribution of site $\boldsymbol{s}$ to an extreme event for all $\boldsymbol{s} \in \mathcal{S}$. A popular parametric model for W^* is the Brown-Resnick process (Brown & Resnick 1977), whose joint finite-dimensional marginal distributions for component-wise maxima of X^* are known to follow the HRD (Huser & Davison

2013).

The HRD has been extensively studied and applied within the extreme value literature and can be viewed as the extremal analogue of the multivariate normal distribution, characterising the asymptotic behaviour of suitably normalised normal vectors (Hüsler & Reiss 1989), in particular, the multivariate joint distribution is entirely determined by the complete set of bivariate distributions. The distribution was originally considered in the context of component-wise maxima and so is a max-stable distribution by derivation. The spatial extension to the HRD led to the popular Brown-Resnick max-stable process which has found wide applicability in environmental extreme value statistics (Engelke et al. 2011). More recently, the Brown-Resnick process has found application outside of the max-stable process and can be used to model the spatial component of the r-Pareto process. In this chapter, we examine in detail the features of the HRD, highlighting its versatility in adapting to typical practical issues of observational environmental data, and in turn, extending the applicability of the r-Pareto process.

The motivation behind this investigation arises from the temporal inconsistencies observed in environmental observations. In extreme spatio-temporal statistics, missing data poses a significant challenge. A time series at any site will likely have missing observations within its observed period for numerous possible reasons, and it is reasonable to assume they are missing at random. In many multivariate or spatial statistical analyses, gap-free data sets are a fundamental prerequisite. No existing spatial extremes approach deals with the issue of missingness, all implicitly assume no data are missing. Missing data imputation is often a necessary procedure when deletion is the only alternative. Many methods of missing data imputation are available, see (Enders 2022). However, the task of imputation is highly non-trivial, especially in data with a large proportion of missingness. As noted by Mishra & Khare (2014) the proportion of missingness significantly impacts the reliability of imputation methods. Yozgatligil et al. (2013) caution that imputation methods become increasingly unreliable and many algorithms, e.g., the EM-MCMC method (or the popular MICE algorithm Wu et al. 2022) become unsuitable beyond 50% missingness.

Furthermore, imputation methods that do not account for the uncertainty in the imputed values will introduce a level of artificial reduction in uncertainties in model

estimates (Glasbey 1995, Griffith & Liau 2021), and incorrect imputations may introduce bias, with both giving misleading results. This issue is again exacerbated by the proportion of missingness. In our analysis of maximum daily summer temperature (Chapter 4), our observed time series comprises approximately 33% of the complete data set. Imputing such a huge proportion of the data set would require significant effort which is beyond the scope of our research goals.

The use of multiple imputation (MI) is suggested when dealing with extensive missingness (Madley-Dowda et al. 2019). MI generates a range of plausible values for each missing observation. This alleviates false precision associated with single imputation methods. The increase in computational cost is manifold as the practitioner is required to perform the statistical analysis on each imputed data set. Our analysis of extreme temperatures is already computationally burdensome and so MI is not practical. Furthermore, the success of MI in the case of large missingness is owed to the use of information borrowed from auxiliary variables. In our analysis, there are no appropriate covariates available, making MI an infeasible solution.

Finally, imputing missing values in spatio-temporal data can be challenging due to the intricate spatial and temporal dependencies, particularly in the extremal setting. In general, the imputation of extreme values is difficult (Clavel et al. 2014, Wu et al. 2022, Chapon et al. 2023). The goal of imputation is to heavily rely on the observed data and minimally on theoretical assumptions (Jean-Baptiste & Jean-Luc 2014, Gao et al. 2018). However, accurate imputation of missing extreme observations must rely on an extreme value theory-based model, and so is typically not helpful. We are therefore motivated to describe a spatio-temporal extreme model that can handle the inconsistent dimensionality of spatial observations caused by missing values without the need for imputation. Our simulation study in Section 3.7 illustrates that our parameter estimation method, which avoids data imputation, gives more accurate parameter estimates of an r-Pareto process than a model fit to only complete spatial observations, which is the standard approach.

In environmental statistics, a typical analysis involves modelling the spatial behaviour of a process given a set of observations at different locations over time. Consider the usual case in weather observation data where a small subset of locations have long-term historical records extending back several decades with a larger subset having observations over a more recent time period. We wish to fit a full

spatio-temporal model which describes the dependence between all sites over time. To include the longer-term historical records in this model, the practitioner can choose to model only those sites with the longest records, likely leading to a spatially sparse data set. Alternatively, the practitioner could disregard events that are not observed at every site in the analysis. This practice of large data omittance leads to several issues including biased and unrepresentative inference as well as weaker inference with more uncertainty (Graham 2009). Disregarding available data from historical time periods removes a lot of information about the long-term behaviour of that process. We risk overlooking important historical contexts and potentially losing valuable information about wider spatio-temporal domains. In spatial extreme value theory, as observations of rare events are of utmost importance, omitting data can lead to the exclusion of crucial extreme observations. Leaving out extreme observations, which are already scarce and valuable for understanding rare events can greatly impact inference and prediction of extreme events potentially leading to underestimation of their associated risks. We propose a method that enables us to utilise all available data, allowing practitioners to make informed inferences based on the complete data set, leading to a more robust and accurate spatio-temporal description of extreme events.

In this chapter, we explore the properties and characteristics of the Hüsler-Reiss distribution, exploring its adaptability in describing the asymptotic behaviour of vectors with unequal dimensions. Throughout, we aim to provide a comprehensive understanding of the theoretical background of the HRD, ensuring that researchers and practitioners can effectively apply such a model. The layout of this chapter is as follows. In Section 3.2 we describe the definition of the HRD along with an outline of its derivation. In Section 3.3 we derive the spectral density of the HRD and prove that it is closed under marginalisation in Section 3.4. In Section 3.5, we derive estimators for the HRD in an r-Pareto process setting, accounting for missing data. In Section 3.6 we present software that can be used to readily apply the HRD in different scenarios followed by a simulation study comparing an r-Pareto process model (parameterised by the spatial HRD, i.e., the Brown-Resnick process) that accounts for missingness with a model that only deals with complete observations in Section 3.7.

3.2 The Hüsler-Reiss distribution

Given a collection of *iid* d-dimensional random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, where $\boldsymbol{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,d})$ for $i = 1, 2, \ldots, n$ which are standard Fréchet distributed (achievable through probability integral transformation), we define the vector of suitably normalised component-wise maxima as

$$\boldsymbol{M}_n = \left\{ \max_{i=1,\ldots,n} X_{i,1}, \max_{i=1,\ldots,n} X_{i,2}, \ldots, \max_{i=1,\ldots,n} X_{i,d} \right\}.$$

We know that if

$$\Pr\left(n^{-1}\boldsymbol{M}_n \leqslant \boldsymbol{x}\right) \to G(\boldsymbol{x}), \text{ as } n \to \infty,$$

where G is non-degenerate in each margin, then G is a multivariate extreme value distribution. Furthermore, G has the form

$$G(\boldsymbol{x}) = \exp\left\{-V(\boldsymbol{x})\right\}, \tag{3.1}$$

where V is called the exponent measure, and has the form

$$V(\boldsymbol{x}) = d \int_{S_d} \left\{ \max_{i=1,\ldots,d} \left(\frac{\omega_i}{x_i} \right) \right\} dH(\omega_1, \ldots, \omega_d) \tag{3.2}$$

for $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}_+^d$, where H is called the spectral measure and is a distribution function defined on the $(d-1)$-dimensional simplex S_d,

$$S_d = \left\{ [w_1, w_2, \ldots, w_d] \in [0,1]^d \ : \ \sum_{i=1}^{d} w_i = 1 \right\},$$

with H satisfying an equal marginal moment constraint. A detailed introduction and derivation of the multivariate extreme value distribution are discussed in detail in Chapter 2 (Section 2.2).

For modelling extreme dependence between a collection of Fréchet marginal random variables, it is useful and indeed typically necessary to specify a parametric form for the multivariate extreme value distribution, G. The HRD is one such valid parametric form of G and has enjoyed significant application across diverse domains

such as finance (Tian et al. 2023), engineering (Fu & Sayed 2022) and environmental science (Bargaoui & Bardossy 2015). Hüsler & Reiss (1989) derived the HRD through consideration of the asymptotic behaviour of component-wise multivariate sample maxima of normal random vectors.

If $\boldsymbol{X}_i$ are *iid* with a multivariate normal copula (with no correlation parameter equal to 1) then it is well-known that the associated multivariate extreme value distribution G as defined in (3.1) is a mutually independent distribution. To achieve asymptotic dependence, Hüsler & Reiss (1989) circumvented this issue by considering a triangular array where the n-th row contains n samples from a multivariate normal distribution with correlation matrix being a function of n, with these correlations tending to 1 as $n \to \infty$. So, they remove the identically distributed assumption of the above.

For simplicity, firstly consider $\boldsymbol{X}_i$ with bivariate normal copula whose correlation coefficient, $\rho(n)$, is a function of the sample size n. Hüsler & Reiss (1989) show that if $\rho(n) = \rho < 1$ for all n then the component-wise maxima are independent in the limit as $n \to \infty$. In contrast, if $\rho(n) = \rho = 1$ for all n then the component-wise maxima are completely dependent in the limit as $n \to \infty$. However, if $\rho(n) \to 1$ as $n \to \infty$ such that

$$[1 - \rho(n)] \log n \to \lambda^2 \text{ as } n \to \infty, \tag{3.3}$$

and $\lambda \in (0, \infty)$, then the limiting component-wise maxima are neither independent nor completely dependent but are asymptotically dependent and non-degenerate. Since we can normalise the margins through probability integral transform, in practice data does not need to have standard normal margins or the limiting Gumbel or Fréchet forms. The bivariate Hüsler-Reiss copula is given by

$$C(u, v; \lambda) = \exp\left\{-\tilde{u}\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log\frac{\tilde{u}}{\tilde{v}}\right) - \tilde{v}\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log\frac{\tilde{v}}{\tilde{u}}\right)\right\}, \tag{3.4}$$

for $0 \leqslant u, v \leqslant 1$, where $\tilde{u} = -\log u$, $\tilde{v} = -\log v$, and $\Phi(\cdot)$ is the standard normal distribution (Joe 1997). Notice that if $\lambda \to 0$, we would have that

$$\lim_{\lambda \to 0} C(u, v; \lambda) = \exp\{-\tilde{u} - \tilde{v}\} = \exp\{\log u + \log v\} = \exp\{\log(uv)\} = uv,$$

which shows independence between the marginals. If $u > v$, and so $\log(\tilde{u}/\tilde{v}) < 0$, then we have $\lim_{\lambda\to\infty} C(u, v; \lambda) = \exp\{-\tilde{v}\} = \exp\{\log v\} = v$. Conversely, if $u < v$, and so $\log(\tilde{v}/\tilde{u}) < 0$, we have $\lim_{\lambda\to\infty} C(u, v; \lambda) = u$. So, in general, if $\lambda \to \infty$ we have

$$\lim_{\lambda\to\infty} C(u, v; \lambda) = \min(u, v),$$

which is the copula of a completely dependent distribution.

The Hüsler-Reiss copula can readily be transformed to have any marginal distribution through probability integral transform. The top row of Figure 3.1 shows samples from a Hülser-Reiss copula, with increasing values of dependence parameter, λ, from left to right. The bottom row of the same figure shows the data from the respective copula above, transformed to have standard Gumbel margins i.e., GEV(0,1,0), so the bottom row shows samples from a bivariate HRD with Gumbel margins. We can see with larger values of the dependence parameter λ we capture stronger dependence.

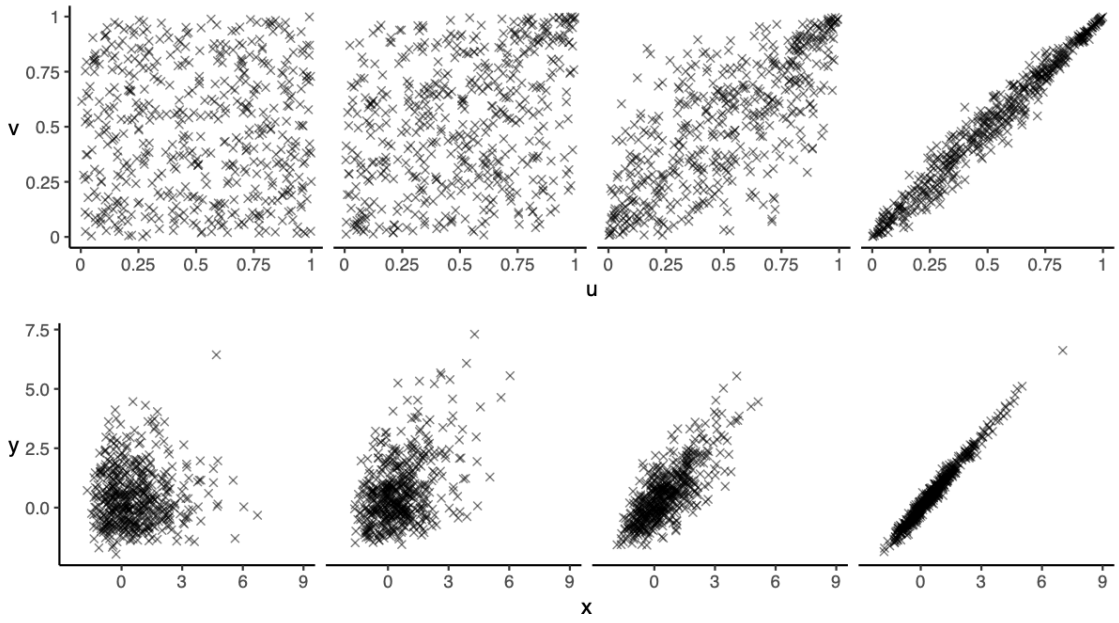

Figure 3.1: Simulations from a bivariate Hüsler-Reiss copula with increasing dependence parameter $\lambda = 0.1, 1, 2, 10$ from left to right (top row). Data from the respective plot above, transformed to have GEV margins with parameters $\mu = 0, \sigma = 1$ and $\xi = 0$ (bottom row).

Now we outline the d-dimensional case. Consider the triangular array Q_n, where row n consists of n-samples from a d-dimensional multivariate normal distribution,

$$Q_n = \begin{bmatrix} \boldsymbol{X}_{1,1} & & & \\ \boldsymbol{X}_{2,1} & \boldsymbol{X}_{2,2} & & \\ \boldsymbol{X}_{3,1} & \boldsymbol{X}_{3,2} & \boldsymbol{X}_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \\ \boldsymbol{X}_{n,1} & \boldsymbol{X}_{n,2} & \boldsymbol{X}_{n,3} & \ldots\ldots \boldsymbol{X}_{n,n} \end{bmatrix}.$$

All multivariate variables in the triangular array are independent. The n-th row of Q_n follows a MVN distribution with correlation matrix $P_d(n) \in \mathbb{R}^{d\times d}$ given by

$$P_d(n) = \begin{bmatrix} \rho_{1,1}(n) & \rho_{1,1}(n) \ldots\ldots \rho_{1,d}(n) \\ \rho_{2,1}(n) & \rho_{2,2}(n) \ldots\ldots \rho_{2,d}(n) \\ \vdots & \vdots \quad \ddots \quad \vdots \\ \rho_{d,1}(n) & \rho_{d,2}(n) \ldots\ldots \rho_{d,d}(n) \end{bmatrix},$$

which is a function of the sample size, n, with correlation coefficient elements $\rho_{i,j}(n)$ where $1 \leqslant i < j \leqslant d$ and with $\rho_{ij}(n) = \rho_{ji}(n)$ when $j < i$. The condition (3.3) extends naturally to the $d > 2$ dimensional case. If $P_d(n)$ satisfies

$$\boldsymbol{\Lambda} = \lim_{n\to\infty} \log(n). \begin{bmatrix} 1-\rho_{1,1}(n) & 1-\rho_{1,1}(n) \ldots\ldots 1-\rho_{1,d}(n) \\ 1-\rho_{2,1}(n) & 1-\rho_{2,2}(n) \ldots\ldots 1-\rho_{2,d}(n) \\ \vdots & \vdots \quad \ddots \quad \vdots \\ 1-\rho_{d,1}(n) & 1-\rho_{d,2}(n) \ldots\ldots 1-\rho_{d,d}(n) \end{bmatrix},$$

where $\boldsymbol{\Lambda} \in (0,\infty)^{d\times d}$ belongs to the set of symmetric, strictly conditionally negative definite matrices, the normalised row maxima of Q_n converge to the d-dimensional HRD, which is completely characterised by the matrix

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{1,1}^2 & \lambda_{1,1}^2 \ldots\ldots \lambda_{1,i}^2 \\ \lambda_{2,1}^2 & \lambda_{2,2}^2 \ldots\ldots \lambda_{2,i}^2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ \lambda_{i,1}^2 & \lambda_{i,2}^2 \ldots\ldots \lambda_{i,i}^2 \end{bmatrix},$$

where $\lambda_{i,j}^2 \in [0,\infty)$, with $\lambda_{i,j}^2 = \lim_{n\to\infty}(\log n)[1-\rho_{i,j}(n)]$.

Explicit derivation of any higher dimensional ($d > 2$) HRD is complex and requires the following notation setup. Given the motivating example of incomplete weather observations, we adopt environmental statistics nomenclature. Suppose we have observations $\boldsymbol{x} = (x_0, x_1, \ldots, x_k) \in \mathbb{R}^{k+1}$ at a set of sites $\mathcal{S} \in \mathbb{R}^{k+1}$, where $\mathcal{S} = \{0, \ldots, k\}$ denotes the indices of all observable sites in the region $\mathcal{S}$. We are ultimately interested in describing the $d = k + 1$ variate HRD of an extremal process over these sites. We let $\mathcal{M}^r$ denote all ordered subsets of $\mathcal{S}$ of size r. That is, $\mathcal{M}^r = \{m_0, \ldots, m_{r-1}\} \subseteq \mathcal{S}$ such that $|\mathcal{M}^r| = r$ and $0 = m_0 < m_1 < \cdots < m_{r-1} \leqslant k$. Following the definition of the HRD as described in the original article by Hüsler & Reiss (1989), for $\boldsymbol{\Lambda} = \left(\lambda^2_{i,j}\right)_{0 \leqslant i,j \leqslant k}$ and $\boldsymbol{m} \in \mathcal{M}^{\ell+1}$, we define the matrix

$$\boldsymbol{\Psi}_{\ell,\boldsymbol{m}}(\boldsymbol{\Lambda}) = 2\left(\lambda^2_{m_i,m_0} + \lambda^2_{m_j,m_0} - \lambda^2_{m_i,m_j}\right)_{1 \leqslant i,j \leqslant \ell} \in \mathbb{R}^{\ell \times \ell},$$

where $\ell = 1, \ldots, k$, where for ease of notation, if $\ell = k$ we denote $\boldsymbol{\Psi}_{k,\mathcal{M}^{k+1}}(\boldsymbol{\Lambda})$ as $\Sigma \in \mathbb{R}^{k \times k}$. The matrix $\boldsymbol{\Psi}$ is simply a function of a sub-matrix of $\boldsymbol{\Lambda}$ and site m_0. Then, we denote the ℓ-dimensional survivor function of a normal random vector with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Psi}$ as $\varsigma_\ell(\,\cdot\,;\boldsymbol{\Psi})$, i.e., $\varsigma_\ell(\,\cdot\,;\boldsymbol{\Psi}) = \int_\cdot^\infty \cdots \int_\cdot^\infty \phi_\ell(\boldsymbol{u}_\ell\,;\boldsymbol{\Psi}) du_1 du_2 \ldots du_l$, where $\boldsymbol{u}_\ell = (u_1, u_2, \ldots, u_\ell) \in \mathbb{R}^\ell$ and $\phi_\ell(\cdot)$ is the ℓ-dimensional normal density function. Then we have that the $(k+1)$-dimensional HRD with standard Gumbel margins, for $\boldsymbol{x} = (x_0, x_1, \ldots, x_k) \in \mathbb{R}^{k+1}$ is

$$H_{\boldsymbol{\Lambda}}(\boldsymbol{x}) = \exp\left(\sum_{\ell=0}^{k}(-1)^{\ell+1} \sum_{\boldsymbol{m} \in \mathcal{M}^{\ell+1}} f_{\ell,\boldsymbol{m},\boldsymbol{\Lambda}}(x_{m_0}, x_{m_1}, \ldots, x_{m_\ell})\right),$$

where

$$f_{\ell,\boldsymbol{m},\boldsymbol{\Lambda}}(y_0, \ldots, y_\ell) = \int_{y_0}^{\infty} \varsigma_\ell\left(\left(y_i - z + 2\lambda^2_{m_i,m_0}\right)_{i=1,\ldots,\ell}\,;\boldsymbol{\Psi}_{\ell,\boldsymbol{m}}(\boldsymbol{\Lambda})\right) e^{-z} dz,$$

where $\boldsymbol{y} = (y_0, y_1, \ldots, y_l) \in \mathbb{R}^{\ell+1}$, $1 \leqslant \ell \leqslant k$, and $f_{0,\boldsymbol{m},\boldsymbol{\Lambda}}(y_{m_0}) = e^{-y_{m_0}}$ when $\boldsymbol{m} = m_0$. Hence, for $\boldsymbol{m} \in \mathcal{M}^{\ell+1}$,

$$\varsigma_\ell\left(\left(y_i - z + 2\lambda^2_{m_i,m_0}\right)_{i=1,\ldots,\ell}\,;\boldsymbol{\Psi}_{l,\boldsymbol{m}}(\boldsymbol{\Lambda})\right) = \int_{y_1 - z + 2\lambda^2_{m_1,m_0}}^{\infty} \cdots \int_{y_l - z + 2\lambda^2_{m_l,m_0}}^{\infty} \phi_l(\boldsymbol{u}_\ell\,;\boldsymbol{\Psi}_{\ell,\boldsymbol{m}}(\boldsymbol{\Lambda})) du_\ell \ldots du_1.$$

An explicit expression for the $d > 2$ dimensional HRD can be derived. However, given the complexity of the distribution it quickly becomes very difficult to write down. An explicit expression for the 3-dimensional HRD can be found in Genton et al. (2011). Note that the HRD is a limiting distribution of normalised component-wise maxima and so is a max-stable distribution. In the following section, we give a detailed treatment of the spectral density of the HRD which we later use to derive an estimator for the distribution's parameters.

3.3 Spectral Density of HRD

In practice, the estimation of the parameters of G is most effectively achieved by estimating the distribution's associated spectral density (Ballani & Schlather 2011). For the purpose of deriving an expression for the spectral density function $h_{\boldsymbol{\Lambda}}$ of the HRD, it is more convenient to consider standard Fréchet margins, so we define $G_{\boldsymbol{\Lambda}}$ as the HRD parameterised by dependence matrix $\boldsymbol{\Lambda}$, with standard Fréchet margins $G_{\boldsymbol{\Lambda}}(\boldsymbol{x}) = H_{\boldsymbol{\Lambda}}(\log \boldsymbol{x})$,

$$G_{\boldsymbol{\Lambda}}(\boldsymbol{x}) = \exp\left(\sum_{\ell=0}^{k}(-1)^{\ell+1}\sum_{\boldsymbol{m}\in\mathcal{M}^{\ell+1}} f_{l,\boldsymbol{m},\boldsymbol{\Lambda}}(\log x_{m_0}, \log x_{m_1}, \ldots, \log x_{m_\ell})\right).$$

Using (3.2) and (3.2) for $\boldsymbol{m} \in \mathcal{M}^{\ell+1}$ we can rewrite $f_{\ell,\boldsymbol{m},\boldsymbol{\Lambda}}$ as

$$f_{\ell,\boldsymbol{m},\boldsymbol{\Lambda}}(\log y_0, \ldots, \log y_\ell) = \int_{\log y_0}^{\infty}\int_{\log y_1 - z + 2\lambda^2_{m_1,m_0}}^{\infty}\cdots\int_{\log y_\ell - z + 2\lambda^2_{m_l,m_0}}^{\infty} \phi_l(\boldsymbol{u}_\ell\,; \boldsymbol{\Psi}_{\ell,\boldsymbol{m}}(\boldsymbol{\Lambda}))du_\ell \ldots du_1.e^{-z}dz.$$

Let $V_{\mathcal{S},\boldsymbol{\Lambda}}(\boldsymbol{x})$ denote the exponent measure and $h_{\mathcal{S},\boldsymbol{\Lambda}}(\boldsymbol{x})$ the spectral density at $\boldsymbol{x}$, both parameterised by the matrix $\boldsymbol{\Lambda}$. From Coles & Tawn (1991, Theorem 1) we can calculate the spectral density $h_{\mathcal{S},\boldsymbol{\Lambda}}$ by taking partial derivatives of the exponent measure $V_{\mathcal{S},\boldsymbol{\Lambda}}$. That is,

$$h_{\mathcal{S},\boldsymbol{\Lambda}}(\boldsymbol{x}) = -\frac{\partial V_{\mathcal{S},\boldsymbol{\Lambda}}(\boldsymbol{x})}{\partial x_0 \partial x_1 \ldots \partial x_k}. \tag{3.5}$$

Hereafter, we drop the subscript $\boldsymbol{\Lambda}$ from the function f, the spectral density h, the

exponent measure V, and the HRD G. However, they remain parameterised by the matrix $\mathbf{\Lambda}$ unless stated otherwise. Furthermore, for ease of notation, we define $\alpha_i = \log x_i - z + 2\lambda_{i,0}^2$ for $i = 1, \ldots, k$. From expression (3.1), the exponent measure $V_{\mathcal{S}}(\boldsymbol{x})$ is defined as $-\log G(\boldsymbol{x})$. We get a more convenient expression for $G(\boldsymbol{x})$ by substituting the expression for $f_{\ell,\boldsymbol{m}}$ in (3.3) into (3.3) which gives us

$$G(\boldsymbol{x}) = \exp\left(\sum_{\ell=0}^{k}(-1)^{\ell+1}\sum_{\boldsymbol{m}\in\mathcal{M}^{\ell+1}}\int_{\log x_{m_0}}^{\infty}\int_{\alpha_1}^{\infty}\cdots\int_{\alpha_\ell}^{\infty}\phi_\ell(\boldsymbol{u}_\ell\,;\boldsymbol{\Psi}_{\ell,\boldsymbol{m}}(\mathbf{\Lambda}))du_\ell\ldots du_1e^{-z}dz\right),$$

with exponent measure,

$$V_{\mathcal{S}}(\boldsymbol{x}) = \sum_{\ell=0}^{k}(-1)^{\ell}\sum_{\boldsymbol{m}\in\mathcal{M}^{\ell+1}}\int_{\log x_{m_0}}^{\infty}\int_{\alpha_1}^{\infty}\cdots\int_{\alpha_\ell}^{\infty}\phi_\ell(\boldsymbol{u}_\ell\,;\boldsymbol{\Psi}_{\ell,\boldsymbol{m}}(\mathbf{\Lambda}))du_\ell\ldots du_1e^{-z}dz.$$

Expanding this, we have,

$$\begin{aligned}V_{\mathcal{S}}(\boldsymbol{x}) =&(-1)^0\sum_{\boldsymbol{m}\in\mathcal{M}^1}f_{0,\boldsymbol{m}}(x_{m_0}) + (-1)^1\sum_{\boldsymbol{m}\in\mathcal{M}^2}f_{1,\boldsymbol{m}}(x_{m_0},x_{m_1})\\ &+(-1)^2\sum_{\boldsymbol{m}\in\mathcal{M}^3}f_{2,\boldsymbol{m}}(x_{m_0},x_{m_1},x_{m_2})+\cdots+(-1)^k f_{k,\mathcal{M}^{k+1}}(x_{m_0},\ldots x_{m_k}),\end{aligned}$$

where we drop the summation in the last term since $\mathcal{M}^{k+1} = \mathcal{S}$, which simplifies the expression further to,

$$\begin{aligned}V_{\mathcal{S}}(\boldsymbol{x}) =&\sum_{i=0}^{k}e^{-x_i} - \sum_{\boldsymbol{m}\in\mathcal{M}^2}f_{1,\boldsymbol{m}}(x_{m_0},x_{m_1}) + \sum_{\boldsymbol{m}\in\mathcal{M}^3}f_{2,\boldsymbol{m}}(x_{m_0},x_{m_1},x_{m_2})\\ &+\cdots+(-1)^k f_{k,\mathcal{M}^{k+1}}(x_0,\ldots x_k).\end{aligned}\tag{3.6}$$

Considering the fact that we will be using (3.5) to derive an expression from the spectral density, we see that all terms in (3.6) will differentiate to 0 (when we jointly differentiate with respect to $x_0, \ldots, x_k$) except for the final term. So, we only consider this term, and by (3.5) we have that

$$h_{\mathcal{S}}(\boldsymbol{x}) = \frac{\partial}{\partial x_0\partial x_1\ldots\partial x_k}(-1)^{k+1}\int_{\log x_0}^{\infty}\int_{\alpha_1}^{\infty}\cdots\int_{\alpha_k}^{\infty}\phi_k(\boldsymbol{u}_k\,;\Sigma)du_k\ldots du_1e^{-z}dz.$$

In general (by the chain rule), when $F(x) \coloneqq \int_a^x f(t)df$, we have that

$$\frac{d}{dx}\int_a^{g(x)} f(t)dt = [F(g(x))]' = F'(g(x))g'(x) = f(g(x)) \cdot g'(x).$$

Applying this to each partial derivative/integral pair in (3.3) we get

$$\begin{aligned} h_{\mathcal{S}}(\boldsymbol{x}) &= \frac{\partial}{\partial x_0}(-1)^{k+1}\int_{\log x_0}^{\infty} \frac{(-1)^k}{x_1 \dots x_k}\phi_k(\tilde{\boldsymbol{u}}_k\,;\Sigma)e^{-z}dz \\ &= \frac{-1}{x_1 \dots x_k}\frac{\partial}{\partial x_0}\int_{\log x_0}^{\infty} \phi_k(\tilde{\boldsymbol{u}}_k\,;\Sigma)e^{-z}dz, \end{aligned}$$

where $\tilde{\mathbf{u}}_k = (\tilde{u}_1, \dots \tilde{u}_k)$, with i^{th} component $\tilde{u}_i = \log x_i - z + 2\lambda_{i,0}^2$ for $i = 1, \dots, k$, so

$$\begin{aligned} \frac{-1}{x_1 \dots x_k}\frac{\partial}{\partial x_0}\int_{\log x_0}^{\infty} \phi_k(\tilde{\boldsymbol{u}}_k\,;\Sigma)e^{-z}dz &= \frac{1}{x_1 \dots x_k}\phi_k(\tilde{\boldsymbol{u}}_k\,;\Sigma)e^{-\log x_0}\frac{1}{x_0} \\ &= \frac{1}{x_0^2 x_1 \dots x_k}\phi_k(\tilde{\boldsymbol{u}}_k\,;\Sigma), \end{aligned}$$

where now, we can see that $\tilde{\mathbf{u}}_k = (\tilde{u}_1, \dots \tilde{u}_k)$ with i^{th} component $\tilde{u}_i = \log(x_i/x_0) + 2\lambda_{i,0}^2$, $i = 1, \dots, k$. Finally, we have that,

$$h_{\mathcal{S}}(x_0, x_1, \dots, x_k) = \frac{1}{x_0^2 x_1 \dots x_k}\phi_k(\tilde{\boldsymbol{u}}_k\,;\Sigma),$$

which we can write as

$$h_{\mathcal{S}}(x_0, x_1, \dots, x_k) = \frac{1}{x_0^2 x_1 \cdots x_k (2\pi)^{k/2}|\det\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{x}}^\top \Sigma^{-1}\tilde{\boldsymbol{x}}\right),$$

where $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \dots, \tilde{x}_k)$ with i^{th} component $\tilde{x}_i = \log(x_i/x_0) + 2\lambda_{i,0}^2$, $i = 1, \dots, k$ and recall that $\Sigma = \boldsymbol{\Psi}_{k,\{0,\dots,k\}}(\boldsymbol{\Lambda}) \in \mathbb{R}^{k\times k}$. Methods for estimating the dependence matrix $\boldsymbol{\Lambda}$ based on the spectral density $h_{\mathcal{S}}$ are given later in Section 3.5.

3.3.1 Spectral density on the simplex

It is often useful to transform data to pseudo-polar coordinates to show how we can define the spectral density on the unit simplex. Let $x_i = rw_i$ where $r = x_0 + \cdots + x_k$ for $i = 1 \dots, k$ and $x_0 = r\,[1 - (w_1 + \cdots + w_k)]$. Rearranging, we get that $w_i = x_i/r$

for $i = 0, \ldots, k$, where $0 \leqslant w_1 + \cdots + w_k \leqslant 1$ and $w_0 = 1 - (w_1 + \cdots + w_k)$. From this transformation, we can express the spectral density as,

$$h_{\mathcal{S}}(\boldsymbol{w}) = \frac{1}{w_0^2 w_1 \cdots w_k (2\pi)^{k/2} |\det \Sigma|^{1/2}} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{w}}^\top \Sigma^{-1} \tilde{\boldsymbol{w}}\right),$$

with $\boldsymbol{w} = (w_0, w_1, \ldots, w_k) \in S_{k+1}$ and $\tilde{\boldsymbol{w}} = (\tilde{w}_1, \ldots, \tilde{w}_k)$ with i^{th} component $\tilde{w}_i = \log(w_i/w_0) + 2\lambda_{i,0}^2$, $i = 1, \ldots, k$, and $\Sigma = \boldsymbol{\Psi}_{k,\{0,\ldots,k\}}(\boldsymbol{\Lambda}) \in \mathbb{R}^{k \times k}$. This highlights the connection of the HRD to expression (3.2).

3.4 Marginalising the HRD

In this section, we illustrate that the HRD is closed under marginalisation. We do this in two ways. Firstly, in Section 3.4.1, considering that the HRD is parameterised in terms of the exponent measure which describes the probability of being in an extremal set bounded away from the origin, we show that marginalisation can be achieved by taking the limit of the exponent measure *wrt* the missing elements. Secondly, in Section 3.4.2, through a change of variables, we express the HRD as a MVN distribution. This allows for a mapping of properties of the MVN distribution (which is trivially closed under marginalisation) to the HRD setting. This transformation is useful as it provides a framework for any feature of the MVN distribution to be mapped back to the HRD.

We begin by setting up some notation to allow us to describe spatially incomplete observations. Recall that $\mathcal{S} = \{0, \ldots, k\}$ denotes the indices of all sites. Suppose we have a collection of observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, with $\boldsymbol{x} = (x_0, x_1, \ldots, x_k)$, where x_s, for $s \in \mathcal{S}$, corresponds to an observation at site s. Ideally, in each $\boldsymbol{x}_i$, for $i = 1, \ldots, n$, we observe the entire spatial domain, $\mathcal{S}$ (i.e., $|\boldsymbol{x}_i| = |\mathcal{S}|$). We are interested in the case where this is not true. Let $\mathcal{D} \subset \mathbb{N}$ denote the set of indices of sites that are missing for a given spatial event and let $\mathcal{E} = (\mathcal{S} \backslash \mathcal{D})$. The observation over sites $\mathcal{E}$ is denoted as $\boldsymbol{x}_{\mathcal{E}}$, i.e., $\boldsymbol{x}_{\mathcal{E}} = \{x_j\}_{j \in \mathcal{E}}$ and let $\boldsymbol{x}_{\mathcal{D}}$ denote the missing observations in the given spatial event. We denote the spectral density in this case as $h_{\mathcal{E}}$.

3.4.1 Marginalising HRD with limit argument

The exponent measure of the HRD is given by

$$V_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}}) = (-1)^0 \sum_{i=0}^{k} e^{-x_i} + (-1)^1 \sum_{\boldsymbol{m}\in\mathcal{M}^2} f_{1,\boldsymbol{m}}(x_0, x_1) + \cdots + (-1)^k f_{k,\mathcal{M}^{k+1}}(\boldsymbol{x}_{\mathcal{S}}),$$

which we can expand to

$$\begin{aligned} V_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}}) = &\sum_{i=0}^{k} e^{-x_i} - \sum_{\boldsymbol{m}\in\mathcal{M}^2} \int_{\log x_{m_0}}^{\infty} \int_{\alpha_{m_1}}^{\infty} \phi_1(\boldsymbol{u}_1; \boldsymbol{\Psi}_{1,\boldsymbol{m}}(\boldsymbol{\Lambda})) du_1 e^{-z} dz \\ &+ \cdots + (-1)^k \int_{\log x_{m_0}}^{\infty} \int_{\alpha_{m_1}}^{\infty} \cdots \int_{\alpha_{m_k}}^{\infty} \phi_k(\boldsymbol{u}_k; \boldsymbol{\Psi}_{k,\mathcal{M}^{k+1}}(\boldsymbol{\Lambda})) du_k \ldots du_1 e^{-z} dz. \end{aligned}$$

Recall that $G_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}}) = \exp\{-V_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}})\}$, and so to marginalise out the variables $\boldsymbol{x}_{\mathcal{D}} = \{x_i\}_{i\in\mathcal{D}}$ we consider $G_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) = \lim_{\boldsymbol{x}_{\mathcal{D}}\to\boldsymbol{\infty}} G_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}}) = \lim_{\boldsymbol{x}_{\mathcal{D}}\to\boldsymbol{\infty}} \exp\{-V_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}})\} = \exp\{-\lim_{\boldsymbol{x}_{\mathcal{D}}\to\boldsymbol{\infty}} V_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}})\}$. To get an expression for $V_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}})$, the exponent measure where only sites indexed by $\mathcal{E}$ are observed, we take the limit

$$V_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) = \lim_{\boldsymbol{x}_{\mathcal{D}}\to\boldsymbol{\infty}} V_{\mathcal{S}}(\boldsymbol{x}_{\mathcal{S}}).$$

So,

$$\begin{aligned} V_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) = \lim_{\boldsymbol{x}_{\mathcal{D}}\to\boldsymbol{\infty}} \Bigg(&\sum_{i=0}^{k} e^{-x_i} + \sum_{\boldsymbol{m}\in\mathcal{M}^2} \int_{\log x_{m_0}}^{\infty} \int_{\alpha_{m_1}}^{\infty} \phi_1(\boldsymbol{u}_1; \boldsymbol{\Psi}_{1,\boldsymbol{m}}(\boldsymbol{\Lambda})) du_1 e^{-z} dz \\ &+ \cdots + (-1)^k \int_{\log x_{m_0}}^{\infty} \int_{\alpha_{m_1}}^{\infty} \cdots \int_{\alpha_{m_k}}^{\infty} \phi_k(\boldsymbol{u}_k; \boldsymbol{\Psi}_{k,\mathcal{M}^{k+1}}(\boldsymbol{\Lambda})) du_k \ldots du_1 . e^{-z} dz \Bigg). \end{aligned} \tag{3.7}$$

Consider the limit of integrals in (3.7) which contain a missing observation, i.e., which have α_j in the lower limit, where $j \in \mathcal{D}$, that is, expressions of the form

$$\lim_{\substack{x_j\to\infty \\ j\in\mathcal{D}}} \int_{\log x_{m_0}}^{\infty} \int_{\alpha_{m_1}}^{\infty} \cdots \int_{\alpha_j}^{\infty} \cdots \int_{\alpha_{m_\ell}}^{\infty} \phi_\ell(\boldsymbol{u}_\ell; \boldsymbol{\Psi}_{\ell,\{m_0,\ldots,j,\ldots,m_\ell\}}(\boldsymbol{\Lambda})) du_\ell \ldots du_1 e^{-z} dz, \tag{3.8}$$

where $\ell \leqslant k$. In order to make the integrals in (3.8) independent of z in the limit, we take the substitution $q_i = u_i + z$ for $i = 1, \ldots, \ell$, then the limits of the integral,

α_{m_i} for $i = 1, \ldots, \ell$ become $\alpha_{m_i} + z = \log(x_{m_i}) - z + 2\lambda_{i,0}^2 + z = \log(x_{m_i}) + 2\lambda_{i,0}^2$. Now, the limit in expression (3.7) equals expression (3.8) where all lower limits $\alpha_j + z$ become $\lim_{x_j \to \infty}(\alpha_j + z) = \lim_{x_j \to \infty}(\log(x_j) + 2\lambda_{j,0}^2) = \log(\infty) + 2\lambda_{j,0}^2 = \infty$ for $j \in \mathcal{D}$. That means, the limit in (3.8) becomes

$$\int_{\log x_{m_0}}^{\infty} \cdots \int_{\infty}^{\infty} \cdots \int_{\alpha_{m_\ell}+z}^{\infty} \phi_\ell \left(\boldsymbol{q} - \mathbf{1}z; \boldsymbol{\Psi}_{\ell,\{m_0,\ldots,j,\ldots,m_\ell\}}(\boldsymbol{\Lambda}) \right) dq_\ell \ldots dq_1 e^{-z} dz = 0.$$

So, after undoing the substitution, expression (3.7) becomes

$$\begin{aligned} V_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) = &\sum_{i=0}^{k} e^{-x_i} - \sum_{\boldsymbol{m} \in \mathcal{M}_{\mathcal{E}}^2} \int_{\log x_{m_0}}^{\infty} \int_{\alpha_{m_1}}^{\infty} \phi_1(\boldsymbol{u}_1; \boldsymbol{\Psi}_{1,\boldsymbol{m}}(\boldsymbol{\Lambda})) du_1 e^{-z} dz \\ &+ \cdots + (-1)^{(j)} \int_{\log x_{m_0}}^{\infty} \cdots \int_{\alpha_{m_j}}^{\infty} \phi_{|\mathcal{E}|}(\boldsymbol{u}_{|\mathcal{E}|}; \boldsymbol{\Psi}_{|\mathcal{E}|-1,\mathcal{M}_{\mathcal{E}}^{|\mathcal{E}|}}(\boldsymbol{\Lambda})) du_j \ldots du_i e^{-z} dz, \end{aligned}$$

where $\mathcal{M}_{\mathcal{E}}^r$ denote all ordered subsets of $\mathcal{E}$ of size r and $\boldsymbol{u}_{|\mathcal{E}|} = (u_0, u_i, \ldots, u_j)$ where $0 < i \leqslant j \leqslant k$, $(i, j) \in \mathcal{E}$.

From expression (3.5), we get an expression of the spectral density $h_{\mathcal{E}}$ by taking the partial derivative of the corresponding exponent measure, i.e., we have that

$$\begin{aligned} h_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) = &- \frac{\partial_{|\mathcal{E}|}}{\partial x_0 \partial x_i \ldots \partial x_j} V_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) \\ = &\frac{(-1)^{(j+1)} \partial^{|\mathcal{E}|}}{\partial x_0 \partial x_i \ldots \partial x_j} \int_{\log x_0}^{\infty} \int_{\alpha_i}^{\infty} \cdots \int_{\alpha_j}^{\infty} \phi_{|\mathcal{E}|}(\tilde{\boldsymbol{u}}_{|\mathcal{E}|}; \boldsymbol{\Psi}_{|\mathcal{E}|-1,\mathcal{M}_{\mathcal{E}}^{|\mathcal{E}|}}(\boldsymbol{\Lambda})) du_j \ldots du_i e^{-z} dz \\ = &\frac{1}{x_0^2 x_i \ldots x_j} \phi_{|\mathcal{E}|}(\tilde{\boldsymbol{u}}_{|\mathcal{E}|}; \boldsymbol{\Psi}_{|\mathcal{E}|-1,\mathcal{M}_{\mathcal{E}}^{|\mathcal{E}|}}(\boldsymbol{\Lambda})), \end{aligned} \tag{3.9}$$

where $\tilde{\boldsymbol{u}}_{|\mathcal{E}|} = (\tilde{u}_i, \ldots, \tilde{u}_j)$ and $\tilde{u}_i = \log(x_i/x_0) + 2\lambda_{i,0}^2$ for $0 < i \leqslant j \leqslant k$, $(i, j) \in \mathcal{E}$. We can see that the last line of (3.9) is exactly the spectral density of the HRD as seen in (3.3), without margins associated with $\mathcal{D}$. Therefore, the HRD is closed under marginalisation.

3.4.2 Marginalisation of the HRD using symmetry

In this section, we provide an alternative and more fruitful method to arrive at the result above. From (3.3), we have the spectral density h in terms of $\boldsymbol{x} =$

$(x_0, x_1, \ldots, x_k) \in \mathbb{R}_+^{k+1}$,

$$h^X(\boldsymbol{x}) = x_0^{-2} \left(\prod_{i=1}^{k} x_i^{-1} \right) \phi_k \left(\tilde{\boldsymbol{x}}_k; \Sigma \right)$$

and where $\tilde{\boldsymbol{x}}_k = (\tilde{x}_1, \ldots, \tilde{x}_k)$ with component $\tilde{x}_i = \log(x_i/x_0) + 2\lambda_{i,0}^2$, for $i = 1, \ldots, k$. Notice that this expression of the spectral density misleadingly appears asymmetric and conditional on the element x_0. We perform a change of variables to illustrate the symmetry of the components of $\boldsymbol{x}$ in spectral density h. Let $\boldsymbol{y} = (y_0, y_1, \ldots, y_k)$, where $y_i = \log x_i - \log x_0 + 2\lambda_{i,0}^2$, for $i = 1, \ldots, k$ and $y_0 = 1/x_0$. We express $h^Y(\boldsymbol{y})$ as $h^X(\boldsymbol{x})/|\det \mathrm{J}(\boldsymbol{x})|$ with $\boldsymbol{x}$ expressed in terms of $\boldsymbol{y}$ and $J(\boldsymbol{x})$ is the Jacobian matrix of the transformation from $\boldsymbol{x}$ to $\boldsymbol{y}$. The Jacobian matrix has (i, j)-th element

$$\{\mathrm{J}(\boldsymbol{x})\}_{i,j} = \frac{\partial y_j}{\partial x_i}, \text{ for } i, j = 0, \ldots, k.$$

So,

$$\mathrm{J}(\boldsymbol{x}) = \begin{pmatrix} \partial y_0/\partial x_0 & \partial y_0/\partial x_1 \ldots\ldots \partial y_0/\partial x_k \\ \partial y_1/\partial x_0 & \partial y_1/\partial x_1 \ldots\ldots \partial y_1/\partial x_k \\ \vdots & \vdots \quad \ddots \quad \vdots \\ \partial y_k/\partial x_0 & \partial y_k/\partial x_1 \ldots\ldots \partial y_k/\partial x_k \end{pmatrix} = \begin{pmatrix} -x_0^{-2} & 0 & \ldots\ldots & 0 \\ x_0^{-1} & x_1^{-1} & & \vdots \\ x_0^{-1} & 0 & x_2^{-1} & \vdots \\ \vdots & \vdots & \ddots & 0 \\ x_0^{-1} & 0 & \ldots\ldots 0 & x_k^{-1} \end{pmatrix}$$

$$\det \mathrm{J}(\boldsymbol{x}) = \frac{-1}{x_0^2} . \det \begin{pmatrix} x_1^{-1} & 0 & \ldots\ldots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots\ldots & 0 & x_k^{-1} \end{pmatrix} = \frac{-1}{x_0^2 x_1} . \det \begin{pmatrix} x_2^{-1} & 0 & \ldots\ldots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots\ldots & 0 & x_k^{-1} \end{pmatrix}$$
$$= \cdots = \frac{-1}{x_0^2 x_1 \ldots x_k}.$$

Then, $|\det \mathrm{J}(\boldsymbol{x})| = 1/(x_0^2 x_1 \dots x_k)$. So,

$$h^Y(\boldsymbol{y}) = \frac{h^X(\boldsymbol{x})}{|\det \mathrm{J}(\boldsymbol{x})|} = x_0^2 x_1 \dots x_k h^X(\boldsymbol{x}), \tag{3.10}$$

Expanding the RHS of (3.10) we have

$$h^Y(\boldsymbol{y}) = \cancel{x_0^2 x_1 \dots x_k} \cdot \frac{1}{\cancel{x_0^2 x_1 \dots x_k}(2\pi)^{k/2}|\det \Sigma|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^\top \Sigma^{-1}\boldsymbol{y}\right),$$

and finally,

$$h^Y(\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}\boldsymbol{y}^\top \Sigma^{-1}\boldsymbol{y}\right) = \phi_k(\boldsymbol{y}; \Sigma)\,.$$

Hence, $\boldsymbol{Y} \sim \mathrm{MVN}(\boldsymbol{0}, \Sigma)$.

Since $h^Y_{\mathcal{S}}(\boldsymbol{y})$ is multivariate normal density, it is closed under marginalisation. This means

$$h^Y_{\mathcal{E}}(\boldsymbol{y}_{\mathcal{E}}) = \int h^Y_{\mathcal{S}}(\boldsymbol{y}) d\boldsymbol{y}_{\mathcal{D}} = \phi_{|\mathcal{E}|}\left(\boldsymbol{y}_{\mathcal{E}}; \Sigma_{\mathcal{E}}\right),$$

where $\Sigma_{\mathcal{E}} = \boldsymbol{\Psi}_{|\mathcal{E}|-1, \mathcal{M}_{\mathcal{E}}^{|\mathcal{E}|}}(\boldsymbol{\Lambda})$, i.e., the matrix of Σ, after removing all rows and columns that correspond with sites in $\mathcal{D}$. Recall the change of variables $h^Y_{\mathcal{S}}(\boldsymbol{y}) = h^X_{\mathcal{S}}(\boldsymbol{x})/|\det \mathrm{J}(\boldsymbol{x})|$. Then,

$$h^X_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) = h^Y_{\mathcal{E}}(\boldsymbol{y}_{\mathcal{E}})|\det \mathrm{J}(\boldsymbol{x}_{\mathcal{E}})|$$

$$\det \mathrm{J}(\boldsymbol{x}_{\mathcal{E}}) = \det \begin{pmatrix} -x_0^{-2} & 0 & \dots & \dots & 0 \\ x_0^{-1} & x_{s_1}^{-1} & \ddots & & \vdots \\ x_0^{-1} & 0 & x_{s_2}^{-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ x_0^{-1} & 0 & \dots & 0 & x_{s_{|\mathcal{E}|}}^{-1} \end{pmatrix} = -x_0^{-2} \prod_{i \in \mathcal{E}} x_i^{-1},$$

where $s_1, s_2, \dots, s_{|\mathcal{E}|} \in \mathcal{E}$. Therefore,

$$\begin{aligned} h^X_{\mathcal{E}}(\boldsymbol{x}_{\mathcal{E}}) &= x_0^{-2}\left(\prod_{i \in \mathcal{E}} x_i^{-1}\right) \phi_{|\mathcal{E}|}\left(\boldsymbol{y}_{\mathcal{E}}; \Sigma_{\mathcal{E}}\right) \\ &= x_0^{-2}\left(\prod_{i \in \mathcal{E}} x_i^{-1}\right) \phi_{|\mathcal{E}|}\left(\tilde{\boldsymbol{x}}_{\mathcal{E}}; \Sigma_{\mathcal{E}}\right), \end{aligned} \tag{3.11}$$

where $\tilde{\boldsymbol{x}}_{\mathcal{E}} = \{\tilde{x}_i\}_{i\in\mathcal{E}}$ and $\tilde{x}_i = \log(x_i/x_0) + 2\lambda_{i,0}^2$. Notice that (3.11) is similar to expression (3.9), which is exactly the spectral density of the HRD without marginal distributions associated with missing components in $\mathcal{D}$, which again shows that the HRD is closed under marginalisation.

3.5 Inference

In this section, we provide detail about how we can estimate the parameter matrix of the HRD for the underlying spatial isotropic Brown-Resnick process. Recall that

$$\Sigma_{\mathcal{S}} = \boldsymbol{\Psi}_{k,\mathcal{M}^{k+1}}(\boldsymbol{\Lambda}) = 2\left(\lambda_{i,0}^2 + \lambda_{j,0}^2 - \lambda_{i,j}^2\right)_{1\leqslant i,j\leqslant k} \in \mathbb{R}^{k\times k}$$

and that $\boldsymbol{\Lambda} = \left(\lambda_{i,j}^2\right)_{0\leqslant i,j\leqslant k}$. Now, take the variogram of the normal process component to be γ_Θ, with parameter set Θ, and re-parameterise the matrix $\boldsymbol{\Lambda}$ as

$$\boldsymbol{\Lambda} = \{\gamma(|s_i - s_j|;\Theta)/2\}_{0\leqslant i,j\leqslant k}\,.$$

Then,

$$\Sigma_{\mathcal{S}} = \Big(\gamma_\Theta(|s_i - s_0|) + \gamma_\Theta(|s_j - s_0|) - \gamma_\Theta(|s_i - s_j|)\Big)_{1\leqslant i,j\in\mathcal{S}}.$$

Let $t \in \mathcal{T} = \{1, 2, \ldots, n\}$ denote the indices of observed events, i.e., $\boldsymbol{x}_t$ is observed at time t from the random variable $\boldsymbol{X}$. Suppose each event is observed at each measurement site in the spatial domain, i.e., $|\boldsymbol{x}_t| = |\mathcal{S}| = k + 1$ for each event t. Let $x_{t,i}$ denote the i^{th} element in event t of random variable $\boldsymbol{X}$. Let $\mathcal{L}(\Theta)$ be the likelihood of *iid* observations $\boldsymbol{x}_t$ for $t = 1, 2, \ldots, n$ over sites $\mathcal{S}$ given the parameter matrix $\Sigma_{\mathcal{S}}$ parameterised via a variogram function,

$$\mathcal{L}(\Theta) = \prod_{t=1}^{n} h_{\mathcal{S}}(\boldsymbol{x}_n\,;\Sigma_{\mathcal{S}}).$$

Let $\log \mathcal{L}(\Theta)$ denote the corresponding log-likelihood, which gives,

$$\begin{aligned}
\log \mathcal{L}(\Theta) &= \log \prod_{t=1}^{n} h_{\mathcal{S}}(\boldsymbol{x}_t \,;\Sigma_{\mathcal{S}}) = \sum_{t=1}^{n} \log h_{\mathcal{S}}(\boldsymbol{x}_t \,;\Sigma_{\mathcal{S}}) \\
&= \sum_{t=1}^{n} \log \left\{ x_{t,0}^{-2} \left(\prod_{i=1}^{k} x_{t,i}^{-1} \right) \phi_k \left(\tilde{\boldsymbol{x}}_t \,;\Sigma_{\mathcal{S}} \right) \right\} \\
&= \sum_{t=1}^{n} \log \left\{ x_{t,0}^{-2} \left(\prod_{i=1}^{k} x_{t,i}^{-1} \right) \frac{1}{(2\pi)^{k/2} |\det \Sigma_{\mathcal{S}}|^{1/2}} \exp\left(-\frac{1}{2} \tilde{\boldsymbol{x}}_t^{\top} \Sigma_{\mathcal{S}}^{-1} \tilde{\boldsymbol{x}}_t \right) \right\} \\
&= \sum_{t=1}^{n} \left[\log \left\{ x_{t,0}^{-2} \left(\prod_{i=1}^{k} x_{t,i}^{-1} \right) \right\} - \log \left\{ (2\pi)^{k/2} |\det \Sigma_{\mathcal{S}}|^{1/2} \right\} - \frac{1}{2} \tilde{\boldsymbol{x}}_t^{\top} \Sigma_{\mathcal{S}}^{-1} \tilde{\boldsymbol{x}}_t \right] \\
&= \sum_{t=1}^{n} \left[\log \left\{ x_{t,0}^{-2} \left(\prod_{i=1}^{k} x_{t,i}^{-1} \right) \right\} - \frac{k}{2} \log \{2\pi\} - \frac{1}{2} \log \{ |\det \Sigma_{\mathcal{S}}| \} - \frac{1}{2} \tilde{\boldsymbol{x}}_t^{\top} \Sigma_{\mathcal{S}}^{-1} \tilde{\boldsymbol{x}}_t \right],
\end{aligned}$$

where $\tilde{\boldsymbol{x}}_t = \log(x_{t,i}/x_{t,0}) + 2\lambda_{i,0}^2$. Since we are interested in determining $\operatorname{argmax}_{\Theta} h(\boldsymbol{X} \mid \Sigma_{\mathcal{S}})$ which corresponds to maximising $\log \mathcal{L}(\Theta)$, we can disregard the terms which are constant with respect to $\Sigma_{\mathcal{S}}$. So, we estimate $\widehat{\Sigma}_{\mathcal{S}}$ as

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \left\{ -\frac{n}{2} \log \left(|\det \Sigma_{\mathcal{S}}| \right) - \frac{1}{2} \sum_{t=1}^{n} \left(\tilde{\boldsymbol{x}}_t^{\top} \Sigma_{\mathcal{S}}^{-1} \tilde{\boldsymbol{x}}_t \right) \right\}, \tag{3.12}$$

which we optimise numerically to estimate the parameters Θ of the variogram γ.

3.5.1 The case of missing data

Now suppose that events are not necessarily observed over the entire spatial domain, $\mathcal{S}$, and that instead, event t is missing observations over sites $\mathcal{D}_t$ and has observations over sites $\mathcal{E}_t$, i.e., $|\boldsymbol{x}_t| = |\mathcal{E}_t| \leqslant k+1$, where $\mathcal{E}_t = \mathcal{S} \backslash \mathcal{D}_t$. Let $\mathcal{L}(\Theta)$ be the likelihood of *iid* observations $\boldsymbol{x}_t$ for $t = 1, 2, \ldots, n$. The log-likelihood of the

parameters Θ is,

$$\begin{aligned}\log \mathcal{L}(\Theta) =& \log \prod_{t=1}^{n} h_{\mathcal{E}_t}(\boldsymbol{x}_t \,; \Sigma_{\mathcal{E}_t}) = \sum_{t=1}^{n} \log h_{\mathcal{E}_t}(\boldsymbol{x}_t \,; \Sigma_{\mathcal{E}_t}) \\ =& \sum_{t=1}^{n} \log \left\{ x_{t,0}^{-2} \left(\prod_{i \in \mathcal{E}_t} x_{t,i}^{-1} \right) \frac{1}{(2\pi)^{|\mathcal{E}_t|/2} |\det \Sigma_{\mathcal{E}_t}|^{1/2}} \exp\left(-\frac{1}{2} \tilde{\boldsymbol{x}}_t^\top \Sigma_{\mathcal{E}_t}^{-1} \tilde{\boldsymbol{x}}_t \right) \right\} \\ =& \sum_{t=1}^{n} \left[\log \left\{ x_{t,0}^{-2} \left(\prod_{i \in \mathcal{E}_t} x_{t,i}^{-1} \right) \right\} - \frac{|\mathcal{E}_t|}{2} \log\{2\pi\} - \frac{1}{2} \log\{|\det \Sigma_{\mathcal{E}_t}|\} - \frac{1}{2} \tilde{\boldsymbol{x}}_t^\top \Sigma_{\mathcal{E}_t}^{-1} \tilde{\boldsymbol{x}}_t \right].\end{aligned}$$

We can derive an estimator of the parameters of the variogram as,

$$\begin{aligned}\widehat{\Theta} =& \underset{\Theta}{\operatorname{argmax}} \left[\sum_{t=1}^{n} \left\{ -\frac{|\mathcal{E}_t|}{2} \log(2\pi) - \frac{1}{2} \log(|\det \Sigma_{\mathcal{E}_t}|) - \frac{1}{2} \tilde{\boldsymbol{x}}_t^\top \Sigma_{\mathcal{E}_t}^{-1} \tilde{\boldsymbol{x}}_t \right\} \right] \\ =& \underset{\Theta}{\operatorname{argmax}} \left(-\frac{1}{2} \sum_{t=1}^{n} \log |\det \Sigma_{\mathcal{E}_t}| - \frac{1}{2} \sum_{t=1}^{t} \tilde{\boldsymbol{x}}_t^\top \Sigma_{\mathcal{E}_t}^{-1} \tilde{\boldsymbol{x}}_t \right)\end{aligned}$$

which we estimate through numerical optimisation.

3.6 hrd software package in R

In this section, we present a new software package, named `hrd`, implemented in R to allow practitioners to readily implement several Hüsler-Reiss based models as described above. Firstly, we provide some functions for exploratory analysis as well as functions to fit the bivariate Hüsler-Reiss copula in two scenarios; where the margins are known *a priori* (or estimated separately), and a framework to estimate the margins and copula jointly in the component-wise maxima setting. Following this, we provide an extension to the software in `mvPot` (de Fondeville & Davison 2018) where we estimate the r-Pareto process parameterised by the Brown-Resnick process in the case of incomplete data records. The package along with installation instructions can be found at `github.com/dairer/hrd`.

To illustrate the functions provided in the `hrd` package, we look at daily maximum reanalysis sea surge data from the ERA5 data set (Hersbach et al. 2020). We take 23 sites across Ireland and the UK as shown in Figure 3.2, where we have a time series of maximum daily sea surge values from 1979 to 2019 for each site. This data

set is complete with no missing observations.

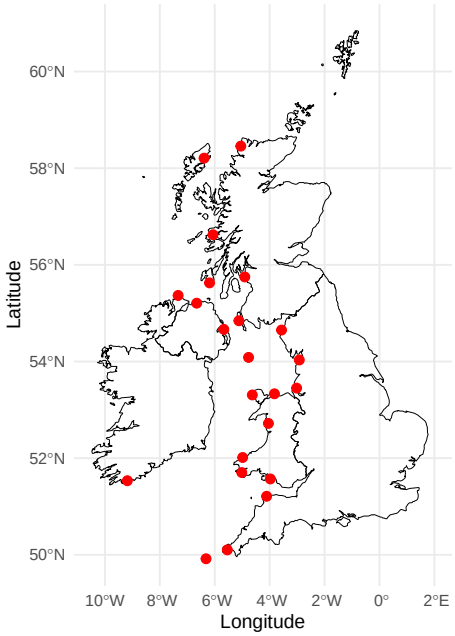

Figure 3.2: ERA5 data locations with maximum daily storm surge data from 1979–2019.

In the following sections, we demonstrate how a practitioner may employ the software provided in `hrd` to easily assess and analyse the extremal dependence of these spatio-temporal data.

3.6.1 Exploratory analysis

We provide exploratory functions to gauge the structure of pairwise asymptotic dependence between of a collection of variables. The function `explore_lambda()` plots an estimate of the dependence matrix $\mathbf{\Lambda}$. By default the function plots the lower diagonal matrix, however the full matrix can be visualised by using the argument `lower_diag = FALSE`. The variables are plotted in alphabetical order by default, however, a specific order can be supplied to the argument `ord`. Alternatively, by setting the argument `use_seriate = TRUE`, the function uses the R package `seriation` to order the variables, using the method specified by the `method`, argument. The default ordering method when `use_seriate = TRUE` is `method="Spectral"` (see Hahsler et al. 2008 for details on the `seriation` package and ordering methods).

The matrix $\mathbf{\Lambda}$ is estimated by standardising the marginal variables to a uniform distribution using an empirical ranking followed by calculating the dependence parameter of the Hüsler-Reiss copula fit to every pair of variables. The copula is fit using the provided function `fit_hrd()` which we discuss separately later. The estimates of the dependence parameters can be translated into the extremal dependence coefficient χ using the relation $\chi = 2\,(1 - \Phi(1/\lambda))$ where $\Phi(\cdot)$ is the standard normal density function. The matrix of estimates of the pairwise dependence coefficient can be calculated and plotted using the function `explore_chi()`. The output of both `explore_lambda()` and `explore_chi()` can be seen in the left and right-hand plot of Figure 3.3 respectively. The variables are plotted in order of longitude, i.e., from most northerly to most southerly.

Estimates of parametrically derived χ from the estimated dependence matrix of the HRD can be compared to empirical estimates of extremal dependence coefficients corresponding to a high quantile u, which we denote as χ_u. Estimates of χ_u serve as an empirical estimator of χ. We can estimate χ_u for two random vectors X and Y as

$$\widehat{\chi}_u(X, Y) = \frac{|\{x \in X : x > X^{(u)}\} \cap \{y \in Y : y > Y^{(u)}\}|}{|\{x \in X : x > X^{(u)}\}|}$$

where $X^{(u)}$ and $Y^{(u)}$ denotes the u-th quantile of X and Y respectively. The function `explore_emp_chi()` plots $\widehat{\chi}_u$. For each of the three exploratory functions, the practitioner is required to supply a matrix or data frame with d columns corresponding to the d sites being analysed, where each row corresponds to an observation at each site. The matrix supplied to `explore_lambda()` and `explore_chi()` should contain bivariate component-wise maxima in each row whereas the matrix supplied to `explore_emp_chi()` should contain all observations. The data can have any marginal distribution. However, for the function `explore_emp_chi()`, the practitioner must also specify the quantile level `u` for which to estimate extremal dependence. The matrix of $\widehat{\chi}_u$ where $u = 0.8$ (the default value) is shown in Figure 3.4, where again the variables are plotted in order of longitude. We see a similar pattern between empirical estimates from the HRD (right-hand plot of Figure 3.3) and empirical estimates (Figure 3.4). However, a comparison of these plots suggests that the HRD is overestimating the extremal dependence between some sites. This is evidenced by the much lower level of extremal dependence estimated empirically as

compared to parametrically in many sites.

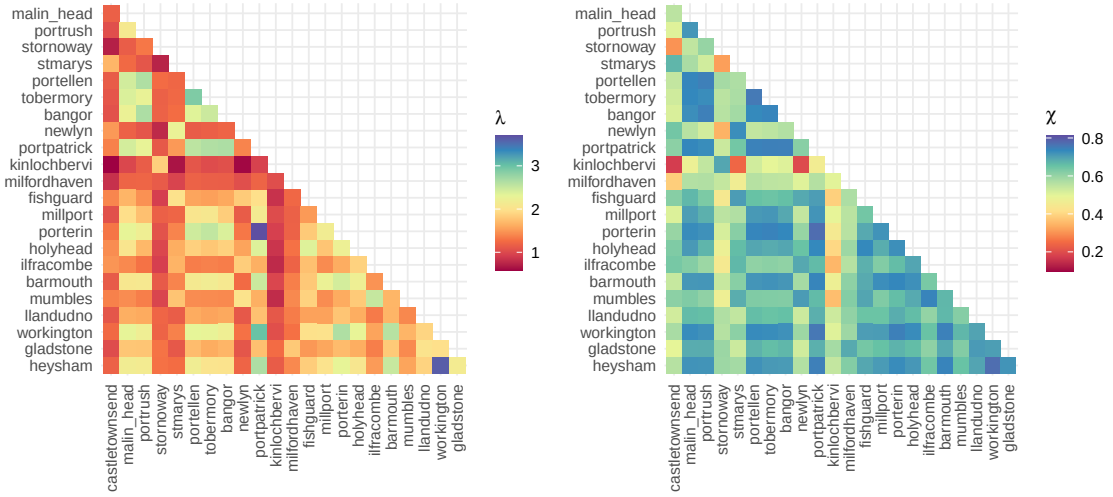

Figure 3.3: Output of the function `explore_lambda()` which gives the estimated lower dependence matrix of the HRD (left). Output of the function `explore_chi()` which gives χ derived from respective estimates in the left-hand plot (right).

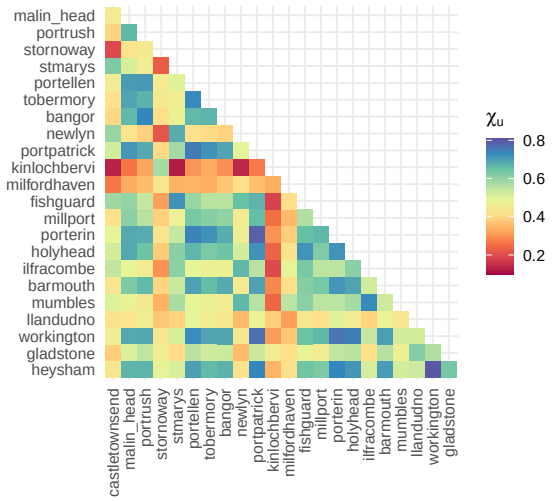

Figure 3.4: Lower diagonal matrix of pairwise empirical estimates of χ_u for $u = 0.8$.

We focus on three sites along the north coast of Ireland, to illustrate how a practitioner could carry out a more rigorous pairwise analysis. These sites are shown in the left-hand plot of Figure 3.5. A subsection of the time series from one of the

sites (Malin Head) is shown in the right-hand plot of Figure 3.5. To assess the class of extremal dependence and to ascertain if the pairwise combinations of these sites exhibit asymptotic dependence, we provide a simple function, `emp_chi()` to empirically calculate the pairwise extremal dependence coefficient χ over a range of quantiles $u \in \mathbb{R}$.

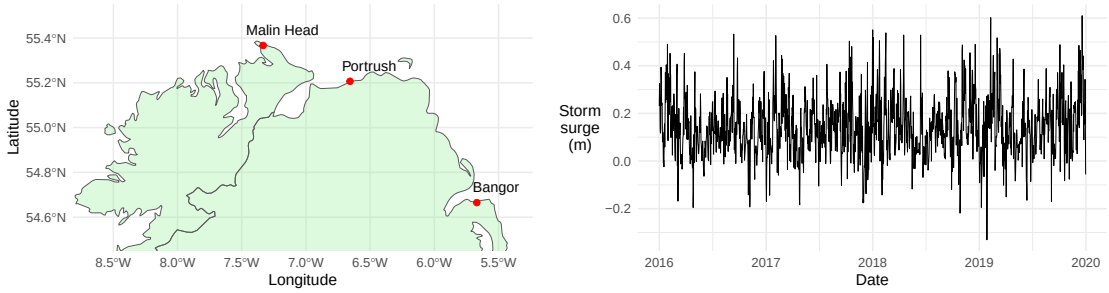

Figure 3.5: Three sites along the north coast of Ireland with maximum daily storm surge data from 1979 – 2019 (left). Time series of storm surge from Malin Head from 2016 – 2019 (right).

The practitioner again supplies a matrix or data frame with two columns corresponding to two sites being analysed, where each row corresponds to an observation at both sites. However, here the user also supplies a vector of quantiles over which to estimate χ. Plotting estimates of χ_u over a range of u provides evidence to determine whether the data exhibit asymptotic dependence or not and hence whether an asymptotic dependent model such as the Hüsler-Reiss distribution is appropriate. Plotting χ_u against u for a range of high quantiles for each pair of sites along the north of Ireland we can see that it is reasonable to assume that each are asymptotically dependent. This is seen in Figure 3.6, since $\hat{\chi}_u$ doesn't asymptotically decay to 0 as we move to higher quantiles. This means that the practitioner may be interested in applying a Hüsler-Reiss based model which we illustrate next.

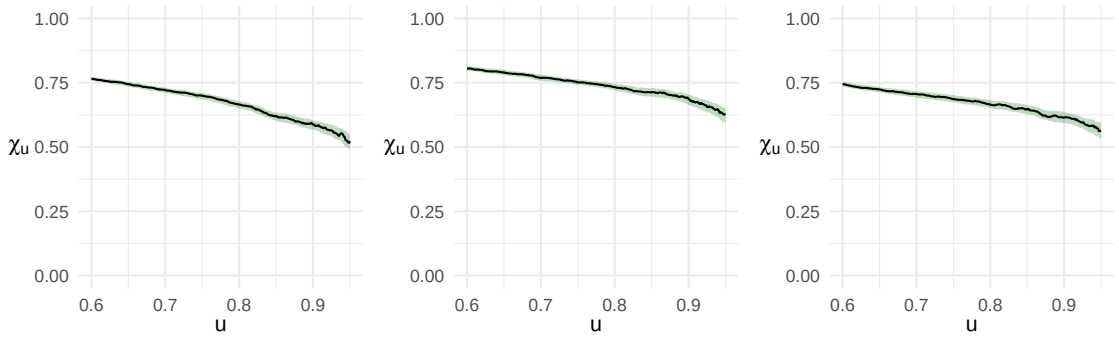

Figure 3.6: Extremal dependence coefficient of three pairs of sites over a range of high quantiles. Plots left-right correspond to pairs of sites Portrush - Malin Head, Bangor - Portrush and Bangor - Malin Head. The shaded green areas show the 95% confidence intervals of the extremal dependence coefficient estimates.

3.6.2 Bivariate modelling

Suppose initially we are interested in applying a component-wise maxima-based analysis of these three sites in Figure 3.2. The functions `fit_hrc()` and `fit_hrc_gev()` can be used to fit a bivariate Hüsler-Reiss copula using maximum likelihood estimation. The first function, `fit_hrc()`, takes data that have been transformed to uniform margins, i.e., the practitioner knows the marginal distributions *a priori* and uses probability integral transformation to standardise both variables to be uniformly distributed. The second function, `fit_hrc_gev()`, assumes that the data are block maxima and on the original data scale. The function estimates the marginal distributions which are assumed to follow a GEV distribution jointly with the dependence parameter of the Hüsler-Reiss copula. For maximum likelihood estimation, the function `fit_hrc()` uses the Brent optimisation algorithm (Brent 1973) implemented in the function `optim()` provided by base `R`. The function `fit_hrc_gev()` uses the Nelder-Mead algorithm (Nelder & Mead 1965), which is better suited to multivariate optimisation and is also implemented in `optim()`. Uncertainties in the parameters are estimated using the estimated Hessian matrix.

For our data, we construct component-wise maxima taking monthly maxima sea surge values at each site. In Figure 3.7, for each pair of sites we plot the component-wise maxima which are assumed to follow a max-stable distribution (e.g., the HRD) with margins following generalised extreme value distributions.

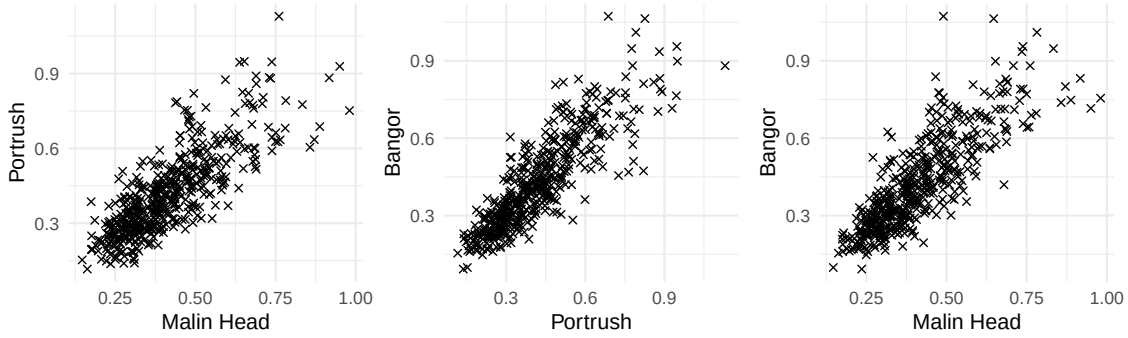

Figure 3.7: Component-wise monthly block maxima of sea surge data for all pairs of three sites along the north coast of Ireland.

We denote any two vectors of block maxima for two asymptotically dependent random variables as X and Y. To estimate the margins and copula separately for X and Y, the practitioner first uses the provided function `fit_gevd()` to estimate the parameters of the GEV of both vectors. The maximum likelihood estimator of the GEV is

$$\hat{\boldsymbol{\Theta}}_X = \underset{\boldsymbol{\Theta}_X}{\operatorname{argmax}} \sum_{i=1}^{n} \log f_{GEV}(x_i; \boldsymbol{\Theta}_X),$$

where f_{GEV} is the density function of the GEV and the vector $\boldsymbol{\Theta}_X$ are the parameters of the GEV distribution for variables X. Using the estimated parameters of the GEV, we can standardise X and Y to have uniform margins as $X^{\text{unif}} = F_{GEV}(X; \boldsymbol{\Theta}_X)$ and $Y^{\text{unif}} = F_{GEV}(Y; \boldsymbol{\Theta}_Y)$, where F^X_{GEV} and F^Y_{GEV} are the distribution functions of the GEV. The maximum likelihood estimator of the HR copula in the function `fit_hrd()` is then

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log c\left(x_i^{\text{unif}}, y_i^{\text{unif}}; \lambda\right),$$

where c is the density function of the bivariate Hüsler-Reiss copula, given as

$$c(u, v\ ; \lambda) = \frac{1}{uv} C(u, v\ ; \lambda) \left[\Phi\left(\frac{1}{\lambda} + \frac{1}{2}\lambda \log z^{-1}\right) \Phi\left(\frac{1}{\lambda} + \frac{1}{2}\lambda \log z\right) + \frac{1}{2}\lambda v^{-1} \phi\left(\frac{1}{\lambda} + \frac{1}{2}\lambda \log z\right) \right],$$

where C is the corresponding distribution function, as in expression (3.4).

To estimate the margins and copula jointly for X and Y, the practitioner can use the function `fit_hrd_gev()`, which has maximum likelihood estimator

$$\left\{\widehat{\lambda}, \widehat{\boldsymbol{\Theta}}_X, \widehat{\boldsymbol{\Theta}}_Y\right\} = \underset{\{\lambda, \boldsymbol{\Theta}_X, \boldsymbol{\Theta}_Y\}}{\operatorname{argmax}} \sum_{i=1}^{n} \log[f^X_{GEV}(x_i; \boldsymbol{\Theta}_X) f^Y_{GEV}(y_i; \boldsymbol{\Theta}_Y) \\ c\left\{F^X_{GEV}(x_i; \boldsymbol{\Theta}_X), F^Y_{GEV}(y_i; \boldsymbol{\Theta}_Y); \lambda\right\}], \tag{3.13}$$

where c is the density function of the bivariate Hüsler-Reiss copula, F^X_{GEV} and F^Y_{GEV} are the estimated distribution functions of the GEV for X and Y respectively and f^X_{GEV} and f^Y_{GEV} are the density function of the GEV for X and Y respectively. The vectors $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Y$ are the parameters of the GEV distribution for variables X and Y respectively.

To compare both methods described above, we estimate the margins and copula of monthly maxima sea surge data at Malin Head and Portrush. In Figure 3.8, we plot estimates for $\boldsymbol{\Theta}_X = (\mu_1, \sigma_1, \xi_1)$ and $\boldsymbol{\Theta}_Y = (\mu_2, \sigma_2, \xi_2)$, which are the location, scale and shape parameters of the two marginal distributions Malin Head and Portrush respectively as well as the dependence parameter λ. The black points represent estimates from the two-step, marginal (using `fit_gevd()`) followed by dependence estimate (using `fit_hrd()`) procedure. The magenta points are estimates from the joint marginal and dependence estimate (using `fit_hrd_gev()`). As can be seen in Figure 3.8, the two fitting procedures give comparable results. However, we note that the likelihood in `fit_hrd_gev()`, as in expression (3.13) is trickier to optimise. Specifically, the estimator is sensitive to initial conditions. This joint marginal and dependence modelling procedure seems to give a more certain estimate of λ, the dependence parameter. However, the estimates of the shape parameter of the marginal distributions, in particular, are more uncertain. Where computationally feasible, the joint fitting procedure provided in `fit_hrd_gev()` may be a more desirable method, giving a more realistic model, irrespective of the uncertainty values.

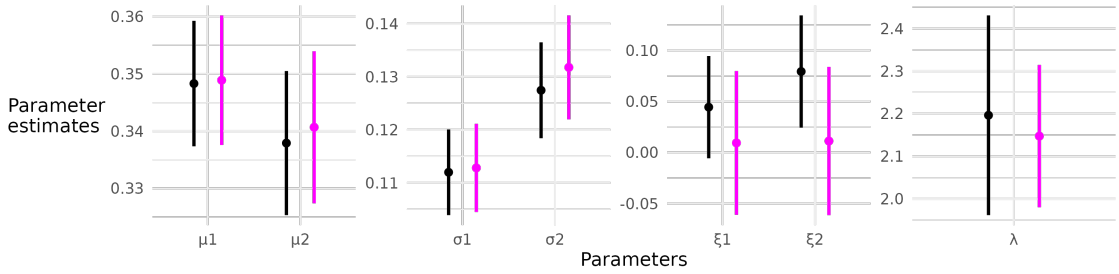

Figure 3.8: Comparison of GEV and bivariate HRD parameter estimates from two-step modelling procedure, consisting of marginal estimates followed by copula estimate (black) and full joint marginal and copula estimate (magenta). In each case, the vertical line segments show 95% confidence interval of their respective point estimates.

We note that due to closure to marginalisation of the HRD, all information about the dependence matrix of higher dimensional HRD ($d > 2$) is contained in the set of all bivariate pairs. That means higher dimensional analysis can be achieved by estimating the bivariate HRD between all pairs. However, in the case of missing data, the construction of matrix $\mathbf{\Lambda}$ from pairwise λ estimates may not give a strictly conditionally negative definite matrix and so care is needed to satisfy this constraint. We estimate the bivariate HRD between all pairs of sites along the north of Ireland as shown in Figure 3.7 using the two-step estimation method. Using the function `rhr()`, we can generate bivariate simulations from the HRD on uniform margins. The practitioner supplies the number of desired simulations `n` along with the estimated dependence parameter `lambda`. The simulations can then be transformed to a data scale using the inverse of the marginal distribution function. We implement the inverse of the generalised extreme value distribution in the function `qgev()`. The simulations can be used to answer probabilistic questions about extreme events of the process. We plot a collection of 100 simulations in red over the component-wise maxima in Figure 3.9.

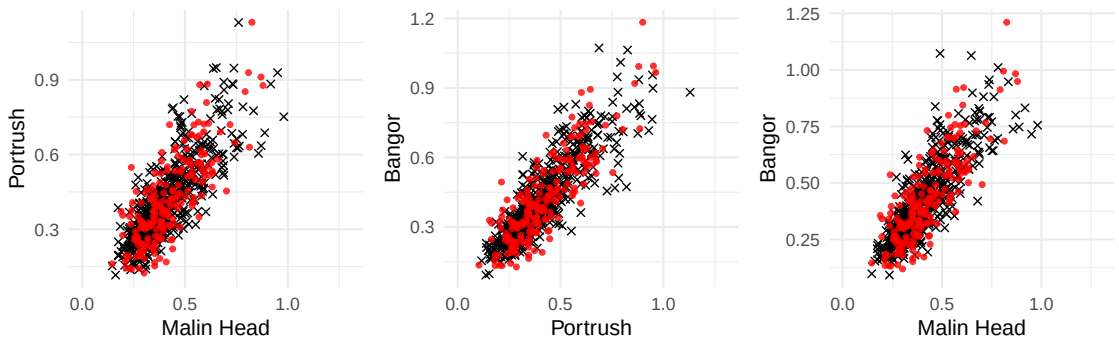

Figure 3.9: Simulations of component-wise monthly block maxima from the HRD transformed to data scale margins through the inverse of GEV distribution fitted to each series of block maxima.

We provide code for the bivariate copula as it proves very useful in many settings. For example, we can model the dependence between two variables at a single site. For example, we could model the dependence between max daily temperature and min daily temperature on extremely hot days. This is an important topic since heat relief is a critical covariate in heat-related mortality. We highlight this as future work and give a more detailed discussion of this topic in Chapter 6.

3.6.3 Fitting a Brown-Resnick process

The `mvPot` package developed by de Fondeville & Davison (2018) provides software to estimate an r-Pareto process parameterised by a Brown-Resnick process following the method of Engelke et al. (2015). We extend our software to allow for changing dimensions. We implement the estimator as described in Section 3.5.1 in the function `fit_rpareto_br()`.

To fit this model, the user must specify a variogram function. The variogram function must accept two arguments, firstly a numeric value $h \in \mathbb{R}_+$ which is the distance between two sites (calculated within the function `fit_rpareto_br()`) and secondly, a numeric vector which stores the parameters of the variogram being estimated. For example, to use the power law variogram the practitioner can define the function as follows

```
> my_vario <- function(dist, pars) pars[1] * dist^pars[2]
```

which will be used internally by the function `fit_rpareto_br()`.

The practitioner supplies two lists `dt` and `locs` in addition to the variogram function, `vr` to the function `fit_rpareto_br()`, each with a length equal to the number of observed spatial events. The first list, `dt`, should contain vectors of observed values on standardised margins. The second list, `locs`, should contain data frames with two columns corresponding to longitude and latitude, and each row describing the location of the observation in the respective position of the vector in the corresponding element of the list `dt`. Note that if the function is passed a data frame instead of a list for `locs`, the function assumes there are no missing data and uses the estimator described in expression (3.12). Simulations can be generated using the `mvPot` package.

3.7 Simulation study of Brown-Resnick process parameter estimation with missing data

We present a simulation study to highlight the potential benefits of incorporating all available data to model incomplete spatio-temporal records rather than only modelling fully observed spatial observations. The left-hand plot of Figure 3.10 shows the simulated site locations with point size and colour indicating the record length at each site. Our data simulation procedure is as follows. We sample 10 random locations uniformly in $[0, 1] \times [0, 1]$. We sample the number of records (or number of days observed) at each site from an exponential distribution with a rate parameter of $1/5000$ which gives an average of $5,000$ simulated 'days' or approximately 13.7 'years' per site. We increment each duration by 365×5 to have at least 5 'years' of data at each site. The right-hand plot of Figure 3.10 further highlights and compares the duration of simulated data at each site.

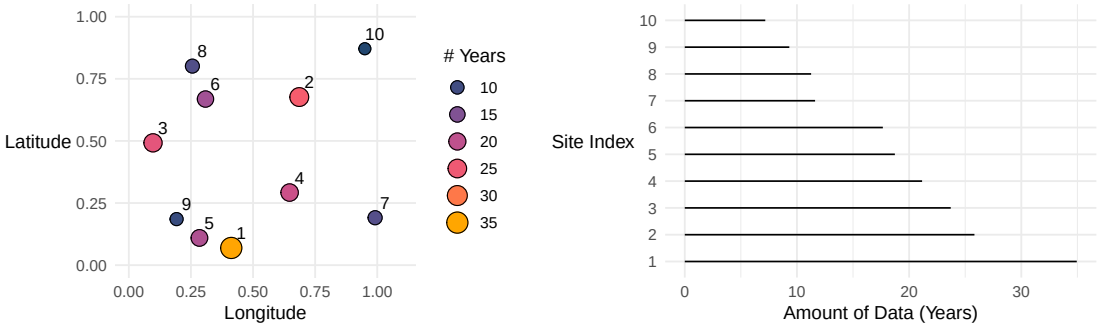

Figure 3.10: Uniformly random sampled locations in $[0,1] \times [0,1] \in \mathbb{R}^2$, with size and colour of points indicating the amount of data observed at that site (left). Time span of records at each site, presented in years (right).

Using the `R` package `mvPot`, we simulate from the multivariate Pareto distribution associated to the Brown-Resnick process over the sites shown in the left-hand plot of Figure 3.10 using a power law variogram $\gamma(h) = \alpha|h|^\beta$ with scale parameter $\alpha = 0.8$ and range parameter $\beta = 1$. We simulate $12,754$ spatial events, corresponding to the length of the longest simulated record (in site indexed 1). We remove values from those sites marked as not observed in accordance with the missingness pattern shown in the right-hand plot of Figure 3.10. Since the simulations are *iid* we do not have to account for the "temporal" ordering of events when removing observations.

We now compare estimates of the variogram parameters using only complete spatial observations to estimates using all spatial observations with at least two observed sites. We have $2,620$ fully spatially observed events each and $9,425$ events with at least two sites observed. Recall that, spatial events whose risk, r, exceeds a sufficiently high threshold, u, will be well-approximated by an r-Pareto process. We denote by $u^{(\tau)}$ the τ-th quantile of risks for all simulated spatial events being modelled. We take risk functional r to be the mean of observations in an event. For both modelling approaches, we take extreme events as those whose risk (calculated as the average value in an event) is in the top 20-th, 10-th and 5-th percentile, i.e., we take $\tau = 0.8, 0.9, 0.95$ and estimate the spatial/angular component (W) of the r-Pareto process ($Y = RW$) using the Brown-Resnick process. Table 3.7 presents the number of extreme spatial events and total marginal level observations we are left with in each case.

	Fully observed events		**All records** ($d > 2$)	
τ	# extreme events	# observations	# extreme events	# observations
0.8	524	5,240	1,885	12,721
0.9	262	2,620	943	6,382
0.95	131	1,310	472	3,150

Table 3.1: Number of extreme spatial observations and marginal level observations above the $v = 0.8, 0.9$ and 0.95 quantiles when considering only fully observed records and when considering all records with at least two sites observed.

The estimates of the parameters of the variogram from both modelling approaches can be seen in Figure 3.11. Both methods give reliable estimates of the true values for extreme events about the 20-th and 10-th quantiles of risk, with the true value within the 95% uncertainty bounds. However, unsurprisingly, incorporating more data gives less uncertain estimates throughout. As we move to the highest threshold of $u^{(0.95)}$, we see that using only the complete records is poor as the true values lie outside the parameter's 95% uncertainty estimates. This is in contrast to our proposed method which avails of much more data, achieving more reliable and robust estimates estimation of parameters. This is evident in the estimated value's proximity to the true values as well as the low uncertainty as compared to the complete cases method, which gives weaker inference with more uncertainty even at lower threshold levels.

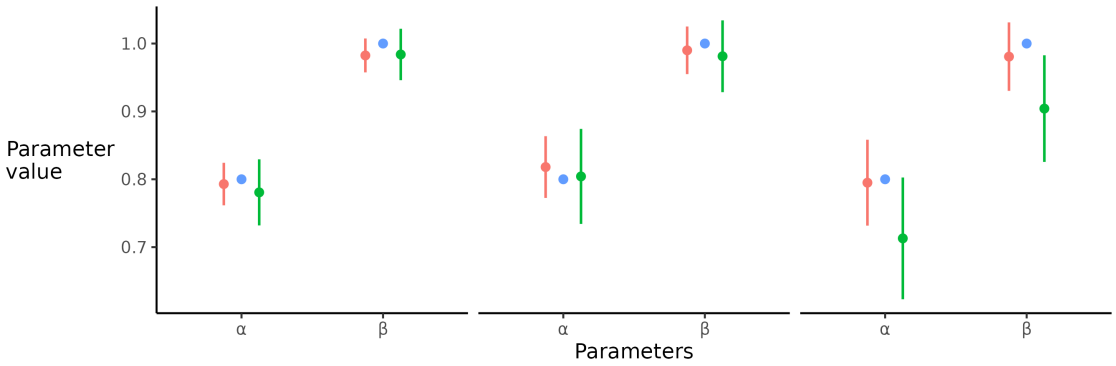

Figure 3.11: Variogram parameter estimates for r-Pareto process simulation in the top 20-th (left), 10-th (centre) and 5-th (right) percentiles. The blue centre point in each plot represents the true parameter values, scale $\alpha = 0.8$ and range $\beta = 1$. The points and vertical segments represent maximum likelihood estimates of parameters along with 95% uncertainties. The orange points represent estimates using all spatial events that have at least two sites observed while the green points represent estimates using only complete spatial records.

3.8 Conclusion

In this chapter, we have presented a modelling framework that allows us to fit an r-Pareto process, parameterised by a Brown-Resnick process, to spatial observations with incomplete and inconsistent dimensionality. Our framework avoids the need for data imputation as well as the omittance of events that are not fully observed over sites being analysed. Our approach allows for the inclusion of much more data in typical observational environmental data settings where data records have differing lengths over sites. Our simulation study reveals that the presented methodology allows for more accurate and precise parameter estimation as compared to estimates based on the subset of fully observed events. We also provide a software package, `hrd` in `R` which is publicly available on GitHub to allow applied scientists to avail of this modelling methodology. The method is applied to observational temperature records over Ireland in the following Chapters 4 & 5.

4

Inference for extreme spatial temperature events in a changing climate with application to Ireland

We investigate the changing nature of the frequency, magnitude, and spatial extent of extreme temperatures in Ireland from 1942 to 2020. We develop an extreme value model that captures spatial and temporal non-stationarity in extreme daily maximum temperature data. We model the tails of the marginal variables using the generalised Pareto distribution and the spatial dependence of extreme events by a semi-parametric Brown-Resnick r-Pareto process, with parameters of each model allowed to change over time. We use weather station observations for modelling extreme events since data from climate models (not conditioned on observational data) can over-smooth these events and have trends determined by the specific climate model configuration. However, climate models do provide valuable information about the detailed physiography over Ireland and the associated climate response. We propose novel methods which exploit the climate model data to overcome issues linked to the sparse and biased sampling of the observations. Our analysis identifies a temporal change in the marginal behaviour of extreme temperature events over the study domain, which is much larger than the change in mean temperature levels over this time window. We illustrate how these characteristics result in increased spatial coverage of the events that exceed critical temperatures.

4.1 Introduction

The Intergovernmental Panel on Climate Change (IPCC 2021, Chapter 11) reports an observable change in extreme weather and climate events since around 1950. Characterisation of extreme temperature events is crucial for societal development, for estimating risks, and to enable the mitigation of their effects for many sectors, e.g., healthcare, economic growth, agricultural disruption, and infrastructure. Brown et al. (2008) observed a warming of both maximum and minimum temperatures since 1950 for most regions indicating an increasing number of warm days, longer heatwaves, and fewer cold extremes.

In Ireland, changing extreme temperature behaviour has also been observed. McElwain & Sweeney (2007) found that a warming of both maximum and minimum temperature observations occurred for all sites over 1961–2005. O'Sullivan et al. (2020) showed that the frequency of extreme temperature events for County Dublin has increased over the period 1981–2010. Both these approaches considered only the marginal behaviour of extremes. To the best of our knowledge, the only modelling of spatial extreme temperature events in Ireland is by Huser & Wadsworth (2022). They used a gridded Irish temperature data set, which has the potential to be over-smooth relative to the observed process, and fitted their model to these data under the assumption of stationarity over time and space. Under similar stationarity assumptions, Fuentes et al. (2013) and Cebrián et al. (2022) present analyses of extreme spatial temperatures for other locations.

We are interested in developing a model which captures the temporal evolution of spatial extreme temperature events over Ireland. This involves modelling how the marginal distributions vary over space, accounting for spatial dependence within extreme events, and modelling how these two elements vary over time. Our focus is on modelling extreme-value data. However for a spatial process, an extreme event can consist of abnormally high values in part of the region and typical values elsewhere (Davison et al. 2012).

Observational extreme event data are sparse and so they need to be used efficiently. The traditional statistical approach is to model these data with powerful probabilistic characterisations from extreme value theory. This theory provides a parsimonious asymptotic justification for extrapolation which enables us to describe

the properties and behaviour of events which are more extreme than those previously observed. However, the theory by itself will not provide information on how to spatially interpolate over heterogeneous geography or how to account for when the characteristics of complex spatial events change over time. Here we propose a novel approach to address these issues which exploits the physical knowledge of the climate processes from information given by fine-scale climate model data. We review existing extreme value methods for spatial and temporal processes and outline our strategies for using climate model data.

The theory of univariate extreme values for stationary processes (Leadbetter et al. 1983), and the associated statistical models (Coles 2001), fully determine a simple parametric distributional family, namely the generalised Pareto distribution (GPD), as the non-degenerate limit distribution for the normalised excesses of a threshold, as that threshold tends to the upper endpoint of the marginal distribution. The GPD has been very widely used in diverse applications since the exposition of Davison & Smith (1990). To deal with non-stationarity, the GPD parameters have been allowed to change smoothly with covariates, initially using fully parametric regression models and more recently with a range of different non-parametric smoothing methods (Chavez-Demoulin & Davison 2005, Youngman 2022). We model the upper tails of the marginal distribution of the temperature process using the GPD, with covariates selected from space, time, information from climate models, and established measures/causes of climate change.

The most established approach to spatial extreme modelling uses max-stable process models (de Haan & Ferreira 2006). These processes are the class of non-degenerate limiting distributions of linearly normalised site-wise maxima, typically fitted to annual maxima data observed at each site in a set of locations over years. They are a natural extension of univariate block maximum limit theory, and so have generalised extreme value distributions for their margins. Brown & Resnick (1977) introduced a widely used subclass of these models, derived from Gaussian random fields, known as Brown-Resnick processes, with Davis et al. (2013) applying this model to spatio-temporal data.

The major problem with max-stable models is that they do not model, and so cannot capture, spatial patterns of observed extreme events. Inference using these models can lead to biased estimation of dependence (Huser & Wadsworth 2022). A recent

development in the modelling of spatial threshold exceedances is the generalised r-Pareto process (Thibaud & Opitz 2015, de Fondeville & Davison 2018, 2022). Generalised r-Pareto processes, like max-stable processes, exhibit a strong form of dependence, known as asymptotic dependence (defined in Section 4.4.2) between all sites. This implies that for an event which is extreme at any location in space, there is a positive probability that this event will be extreme everywhere else in the spatial domain. For processes over spatial domains that are large relative to the scale of the spatial dependence of the process, this is an unrealistic assumption. More flexible spatial models, building from those in Wadsworth & Tawn (2022) are discussed in Appendix B.7.1.

In Section 4.4 we define these extremal dependence properties precisely and provide evidence that generalised r-Pareto processes are suitable for daily maximum temperatures over Ireland. We identify extreme spatial fields, based on observations at d sites, as those which exceed a sufficiently high threshold for a risk function $r : \mathbb{R}^d \to \mathbb{R}_+$. We model these fields as realisations of a Brown-Resnick Pareto process, which is closed under marginalisation (Engelke & Hitz 2020), an important property given the time-varying level of missing temperature data in our application.

We have observational daily maximum data for a network of 182 Irish temperature stations, with only 38% of these having more than 30 years of data due to differential operational periods and quality controls, which is further compounded by spatial selection bias in the station locations. We also have a rich spatio-temporally complete data set generated from a climate model, giving daily maximum temperatures over 56 years on a fine grid over the island of Ireland. These climate model data are not conditioned on the observed weather, so their values on any given day have no correlation to the observed data, but they have similar probability distributions to the observed data at the associated sites and time of the year. The climate model data have no missing values or location biases; they are on a dense regular grid and incorporate the impact of known geophysical structures on the temperature process.

Although it may be tempting to analyse the simpler climate model data than the observed station data, climate models involve some abstraction of the physical processes they model and so tend to under-predict extreme events in magnitude and to

over-estimate dependence owing to the climate model's smoothness over space and time, see Section 4.3 and Section 4.4. So direct analysis of the climate data is not ideal but clearly, they offer vital additional information to the observational data. Various attempts have been made to downscale the climate model data to produce a proxy for the observed data which gives a spatial and temporally complete data set, e.g., Maraun et al. (2017), with the focus to date being on assessing marginal features. We prefer to let the observational data stand for themselves, particularly in relation to the information they provide about temporal non-stationarity as the climate model data have trends determined by the climate model configuration that is an imperfect representation of the real-world processes.

The novelty of our method is achieved through the use of state-of-the-art extreme value methods for marginal distributions, spatial dependence, and temporal non-stationarity which collectively exploit knowledge from climate science and through the use of appropriate metrics for describing changes in spatial extreme events. Our use of climate science relies heavily on how our inference for the observational temperature data leverages core information from the climate model data, i.e., parameter estimates (within sample quantiles and GPD parameters) over space, and through our careful assessment of, and sensitivity to, the effects of the inclusion of various climate-based covariates.

The chapter is organised as follows. Section 4.2 details the observational and climate model data used. Section 4.3 and Section 4.4 describe the marginal and dependence modelling of the process respectively, in each case accounting for their changing behaviour over time. In Section 4.5 we use the model to explore how the properties of spatial extreme events have changed over time. Conclusions and a broader discussion are given in Section 4.6 and Section 4.7. All our code and instructions on how to access the data are available on GitHub[a].

4.2 Data

We start with a note on nomenclature since we use multiple data sets which differ in their structure and use. We use the term 'station data' to refer to data taken

[a] https://github.com/dairer/Extreme-Irish-Summer-Temperatures.

directly from weather stations. These are irregularly located and suffer from missing values. The term 'climate model data' refers to physics-based simulations of the weather system which are run on high-resolution grids and do not aim to match individual weather events, rather they model the spatio-temporal dynamics of the weather system. Finally, 'observation-based data products' are gridded data sets which arise from some form of statistical or physical interpolation of station data.

4.2.1 Station data

Our daily maximum temperature data comprise 182 Irish temperature stations compiled from two sources, the locations of which are shown in Figure 4.1. For the Republic of Ireland, data for 151 stations came from Met Éireann's archive[b]. Data for 31 Northern Ireland sites were obtained through the CEDA archive (Met Office 2012). Collectively, these data have many missing values, with the availability of data decreasing further back in time. We have more than twice the data from the 1950s than the 1940s, with all stations pre-1950 (except one) being coastal. We have 56% of daily values observed in the last 30 years and only 0.53% observed before 1942. We present and contrast results hereafter for the years 1942 and 2020 and focus our discussion on this period. No single day has data for every station. The average span of data for each station is about 30 years, with observations ranging from 1931 to 2022. The sites with the most data tend to be located near the coast reflecting historical and present-day observational priorities.

Our interest is in extreme warm temperatures in Ireland, so we restrict our analysis to data from the summer months June, July, and August (JJA). This choice is supported by the finding that 93% of all the days with temperatures above the 99% site-wise marginal quantile occur in summer, with this proportion increasing with the level of marginal quantile. Furthermore, we assume that within each summer the process is temporally stationary, with exceedances being reasonably spread across the summer. Exploratory analysis supporting both of these choices is reported in Appendix B.2.3.

[b]Copyright Met Éireann. Source: https://met.ie/climate/available-data/historical-data.

4.2.2 Climate model data

Climate models are mathematical representations of the physical processes driving weather and climate and represent our best understanding of these natural phenomena (Giorgi 2019). Climate models are broadly run on two scales; large-scale global climate models (GCMs) and finer-scale regional climate models (RCMs). An RCM is informed by the GCM at its boundary. Crucially, climate model data do not correlate with the observed time-evolution of weather, rather they have probabilistic structures that reflect plausible weather sequences which could occur. They are typically designed to investigate the effect of potential external forcing on the climate system by, e.g., increases in greenhouse gas concentrations arising from anthropogenic activities. We identify and exploit physical and topographical features in the output of these models and use them to adjust for spatial and temporal bias in the observed data set.

We use RCMs for their detailed topographical information and their physical description of temperature processes. When relying only on climate models to understand extreme weather it is common to consider several GCM/RCM combinations, each with different initial conditions and future climate scenarios. This choice will have limited impact for us as we use station data to describe the magnitude and frequency of temperature events, and the climate model data only to inform non-temporal features. We use data from the CLMcom-CLM-CCLM4-8-17 RCM combined with the ICHEC-EC-EARTH GCM. Specifically, we have daily maximum temperatures over a 56-year period (created using the atmospheric climate drivers from 1950 to 2005) on a regular grid of 558 points over Ireland (corresponding to a 0.11^2 degree resolution). Figure 4.1 (right) shows the values for the day with the largest average temperature over Ireland in this data set. This plot illustrates two features which we exploit in Section 4.3 and Section 4.4 respectively. Firstly, the RCM provides much greater spatial coverage in the interior of Ireland than the stations in Figure 4.1 (left). Secondly, extreme temperature events can be very widely spread across Ireland, since even the sites with the lowest values on this day have temperatures in their marginal distributions' upper tails.

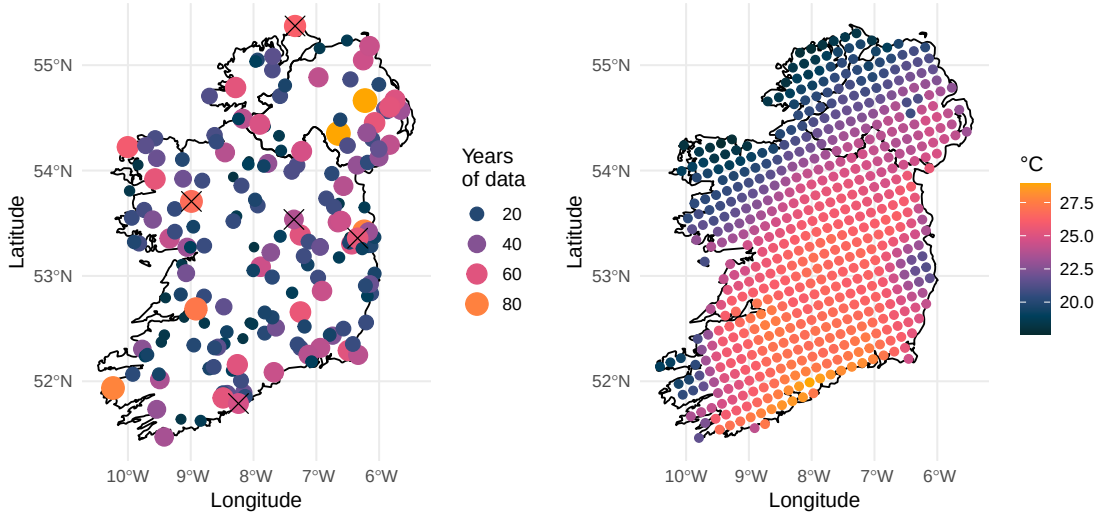

Figure 4.1: Ireland data locations: station data sites, with the amount of data indicated by colour and size (left). Sites marked with an '×' correspond to Malin Head (North), Roches Point (South), Phoenix Park (East), Claremorris (West) and Mullingar (Centre); climate model data from MOHC-HadREM3-GA7-05 showing a generated extreme temperature event (right).

4.2.3 Covariates: observation-based data products

To model temporal non-stationarity of extreme temperature data it is common to use time as the sole covariate, although this will have severe limitations outside the range of the data as potential emission scenarios diverge. Instead, we use the time-varying covariates that climate scientists believe best represent changes in observed mean temperatures. These are predictable into the future under different emission scenarios. We use two covariates; smoothed monthly average temperature anomalies for the global average $M^G(t)$ and for the grid box over Ireland $M^I(t)$, from the observation-based data product HadCRUT5, over time t. See Appendix B.2.1 for details and plots of the covariates. Over 1942–2022, both covariates increase by $\approx 1°$C, with the change accelerating.

Our exploratory analysis, using spline-based models, identified that the shortest distance to the coast, for each site, was a potential descriptor of the change of temporal trends. We define this covariate by $C(\boldsymbol{s})$, for each site $\boldsymbol{s}$. To estimate the coastal distance $C(\boldsymbol{s})$, we use the `st_distance()` function from the `R` package `sf` (Pebesma et al. 2023), to calculating the shortest euclidean distance between

each synoptic weather station and the Irish coastline. Data for the Irish coastline was downloaded using the `R` package `rnaturalearth` (Massicotte et al. 2023). We consider the covariate of the annual CO_2 emissions ($CO_2(t)$) for Ireland; see plot and details in Appendix B.2.2. Since there is strong collinearity present in the collective covariates $\boldsymbol{z}_t := \{t, M^G(t), M^I(t), \mathrm{CO}_2(t), C(\boldsymbol{s})\}$ we only use one of these at a time in each model.

4.3 Marginal models

4.3.1 Overview and strategy

Let $X_{\mathrm{o}}(t, \boldsymbol{s})$ denote the observed station data comprising summer maximum daily temperature at time t and site $\boldsymbol{s}$, and let $X_{\mathrm{c}}(t, \boldsymbol{s})$ be the equivalent process from the climate model data. We assume temporal stationarity within each year for each site and each process. Here $t \in \mathbb{N}$ indexes summer days within and across years and $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$, where $\mathcal{S}$ denotes Ireland, with $\boldsymbol{s}$ corresponding to the vector of latitude and longitude. We have data on the two processes at $\mathcal{S}_o \subset \mathcal{S}$ and $\mathcal{S}_c \subset \mathcal{S}$ and at times $\mathcal{T}_o$ and $\mathcal{T}_c$ respectively. For $\mathcal{T}_o$ we also have missing data for some of the stations as discussed in Section 4.2. We use the subscripts to identify the type of process throughout, though the indexing is dropped when discussing methods which apply similarly to both processes.

In Section 4.3.2, we propose a spatial and temporal quantile regression model for the data to derive an estimate of the distribution function of $X(t, \boldsymbol{s})$. As the tails of this distribution are particularly important to model well, we introduce a threshold $u(\boldsymbol{s})$ which is fixed over time but varies in space, above which we replace the quantile model with the generalised Pareto distribution (GPD) parametric model with temporal and spatial covariates. The justification for our choice of a constant threshold over time is discussed in Section 4.7.2. Novelty in our approach comes from using estimates from the X_{c} process to infer features of the X_{o} process, which is appealing as $|\mathcal{S}_o| \ll |\mathcal{S}_c|$ and $|\mathcal{T}_o| \ll |\mathcal{T}_c|$ for most sites. Given all these considerations, we need to estimate thresholds, the temporally varying marginal distributions over $\mathcal{S}$ for below the thresholds, the GPD parameters for above the thresholds, and to do this for both the X_{c} and X_{o} processes.

When analysing the climate model data we need to account for the following issues. First, our use of these data is to improve our spatial mapping and to overcome issues of missing data in the analysis of the observational data. Second, climate model data can show different time dynamics from that of the observed process since they are based on incomplete physics and forcing detail. We want our analysis to be robust to temporal non-stationary aspects of the climate model data, so we assume that $X_c(t, \boldsymbol{s})$ are temporally stationary in our analyses. As the trend in the climate model data is 4% of the variation in the data at each site, this is not too restrictive an assumption.

4.3.2 Modelling the body of the distribution

Given the issues raised in Section 4.3.1 about $X_c(t, \boldsymbol{s})$, we take the following simple approach for the inference of its distribution. For site $\boldsymbol{s}$ we estimate the τth quantile of X_c by using the empirical sample quantile for the climate model data at that site alone; we denote this estimator by $q_c^{(\tau)}(\boldsymbol{s})$. We use this approach for all τ over the range $[0.01, 0.99]$. This estimator is reliable as we have sufficient data (5152 days with none missing) and, due to the climate model data being numerical model output, their spatial variation is very smooth, so statistical spatial smoothing methods are more likely to induce bias than to enhance the analysis through information sharing.

For the analysis of $X_{\mathrm{o}}(t, \boldsymbol{s})$, the issues of spatial sparsity of stations, limited data, varying periods of records of stations, and the need to account for temporal variations, lead to a different approach than for $X_{\mathrm{c}}(t, \boldsymbol{s})$. We follow the approach of Yu & Moyeed (2001) and that of the R package `evgam` (Youngman 2022) by using the asymmetric Laplacian distribution (ALD) for quantile regression to estimate a range of spatially and temporally varying τth quantiles, $q_o^{(\tau)}(t, \boldsymbol{s})$, for a grid of $\tau \in [0.01, 0.99]$, for all $t \in \mathcal{T}$ and $\boldsymbol{s} \in \mathcal{S}$. The density function of the ALD_τ is

$$f_{\mathrm{ALD}_\tau}(y; q, \psi) = \tau(1-\tau)\psi^{-1} \exp\left\{-\rho_\tau(y-q)\,\psi^{-1}\right\},\ y \in \mathbb{R}, \tag{4.1}$$

where $\rho_\tau(z) = \{\tau - I(z < 0)\}z$ is the check function, $q \in \mathbb{R}$ is a location parameter, corresponding to the τth quantile of interest, and $\psi > 0$ is a scale parameter. We assume that q and ψ vary smoothly over $\mathcal{S}_o$ and $\mathcal{T}_o$.

For estimating $q_o^{(\tau)}(t, \boldsymbol{s})$ and $\log\{\psi_o^{(\tau)}\}$ we consider not just t and $\boldsymbol{s}$ as covariates but also incorporate as potential covariates the associated quantile from the climate model data $q_c^{(\tau)}(\boldsymbol{s})$ and each of the climate-based covariates of Section 4.2.3. The former provides richer spatial information that is not captured in the observational data set, and the latter gives a causal set of time-varying covariates. Details of the analysis using these models are presented in Section 4.3.5.

To provide estimates for all τ, we fit this model separately for a grid of τ values and use a cubic interpolation spline for each $\boldsymbol{s}$ to give a continuous estimate over $0.01 \leq \tau \leq 0.99$ for $\hat{q}_o^{(\tau)}(t, \boldsymbol{s})$. We keep the grid of τ values relatively coarse to avoid issues of quantile estimates crossing. This gives us an estimate of the distribution function of $X_o(t, \boldsymbol{s})$ as

$$F_{o,t,\boldsymbol{s}}(\hat{q}_o^{(\tau)}(t, \boldsymbol{s})) := \Pr\{X_o(t, \boldsymbol{s}) < \hat{q}_o^{(\tau)}(t, \boldsymbol{s})\} \approx \tau \,\forall\, \tau, \,\forall\, \boldsymbol{s} \in \mathcal{S}. \tag{4.2}$$

This model provides estimates for all quantiles for any $\boldsymbol{s} \in \mathcal{S}_c$, not just $\mathcal{S}_o$, and at all times where we have the covariates, e.g., not just for $t \in \mathcal{T}_o$. At each site $\boldsymbol{s}$, below the threshold $u(\boldsymbol{s})$ (defined in Section 4.3.3) we use this distributional model $F_{o,t,\boldsymbol{s}}$.

4.3.3 Modelling the tails of the distribution

It is well known that quantile regression, and hence the ALD model, is unreliable for estimating quantiles in the tails of the distribution and provides no means to extrapolate beyond the observed data. As the upper extremes of the distribution of $X_o(t, \boldsymbol{s})$ are important to us, we chose to use a different model based on extreme value methods. This enabled us to produce a model that is continuous over all t and $\boldsymbol{s}$, with the extreme value model being used above a high threshold, and the ALD model describing data below.

One option is to have a threshold, $u(t, \boldsymbol{s})$, that varies over time and space computed for a given quantile, e.g., $u(t, \boldsymbol{s}) = \hat{q}^{(\tau)}(t, \boldsymbol{s})$, for a choice of τ. However, in extreme value inference, it is well-known that it is difficult to objectively select a threshold and/or to account for the uncertainty in that choice (Northrop et al. 2017). Here it is the temporal change in extreme events which is of most interest, and this trend is small relative to other sources of variations in the data. We do not select a time-

varying threshold using information from the body of the distribution, as this may bias results for the extremes. Instead, we choose the threshold to be constant over time but varying over space, i.e., $u(\boldsymbol{s})$. Our choice is discussed further in Section 4.7.

To reduce subjectivity, for each site and for both X_c and X_o processes, we use a common exceedance probability for the fixed-over-time threshold. Based on the use of standard extreme value threshold selection methods for stationary processes (Coles 2001) which we applied at each site/process separately, we identified that the 90% quantile was suitable. For the reasons discussed in Section 4.3.1 and Section 4.3.2 we use the site-specific 90% empirical sample quantile for $u_c(\boldsymbol{s})$ but a model-based estimate for $u_o(\boldsymbol{s})$. Specifically, we fit the model for density (4.1) with $\tau = 0.9$ with the location parameter structured as $u_o(\boldsymbol{s}) := q_o^{(0.9)}(\boldsymbol{s}) = \beta_0 + \beta_1 u_c(\boldsymbol{s})$, with (β_0, β_1) parameters. Thus the climate model data provides a means by which the spatially varying threshold u_o for the observed data X_o can be estimated. This routine aims to overcome the data quality limitations and to provide estimates for all $\boldsymbol{s} \in \mathcal{S}_c$.

For a given threshold there are two remaining elements required to model the extremes, namely the threshold exceedance probability $\lambda_o(t, \boldsymbol{s})$ and the distribution H of the excesses of the threshold (Chavez-Demoulin & Davison 2005). We consider these in turn. We estimate $\lambda_o(t, \boldsymbol{s})$ from the model for the body of the distribution, using the set of estimated distribution functions (4.2). Specifically,

$$\lambda_o(t, \boldsymbol{s}) = 1 - \tau_{u_o}(t, \boldsymbol{s}),$$

where $\tau_{u_o}(t, \boldsymbol{s})$ is the value of τ, at time t, which makes $\hat{q}_o^{(\tau)}(t, \boldsymbol{s}) = u_o(\boldsymbol{s})$. If there is no temporal non-stationarity in $X_o(t, \boldsymbol{s})$ then by construction of the threshold $u_o(\boldsymbol{s})$, we would have $\lambda_o(t, \boldsymbol{s}) = 1 - \tau = 0.1$ across $t \in \mathcal{T}_o$ and $\boldsymbol{s} \in \mathcal{S}_o$.

For each site $\boldsymbol{s}$ we assume that excesses of the threshold $u(\boldsymbol{s})$ follow a generalised Pareto distribution (GPD); see Pickands (1975) and Davison & Smith (1990) for the probabilistic justification and properties. The GPD(σ, ξ) has distribution function

$$H(y; \sigma, \xi) = 1 - (1 + \xi y/\sigma)_+^{-1/\xi}$$

for $y > 0$, with a shape parameter $\xi \in \mathbb{R}$ and a scale parameter $\sigma > 0$, with the notation $x_+ = \max(x, 0)$, and $\xi = 0$ is obtained by taking the limit as $\xi \to 0$.

When $X(t, \boldsymbol{s}) > u(\boldsymbol{s})$ the threshold excess, $Y(t, \boldsymbol{s}) = X(t, \boldsymbol{s}) - u(\boldsymbol{s})$, is taken to be distributed as

$$Y(t, \boldsymbol{s}) \sim \text{GPD}(\sigma(t, \boldsymbol{s}), \xi),$$

where we discuss the choice of models for $\sigma(t, \boldsymbol{s})$ below, and where the shape parameter is taken to be constant over time and space. This choice of homogeneity for the shape parameter for both X_o and X_c (i.e., values ξ_o and ξ_c respectively) is supported by exploratory analysis in Appendix B.3.2, but it is typical in GPD modelling as there is limited evidence against this in almost all applications, and for the pragmatic reason that even a homogeneous value is difficult to estimate well. Furthermore, this choice reduces the risk of parameter identifiability problems (Davison et al. 2012).

Combining $\lambda_o(t, \boldsymbol{s})$ with the model for H gives our overall marginal distributional model $F_{o;t,\boldsymbol{s}}$ for the upper tail of $X_o(t, \boldsymbol{s})$. Specifically for $y > u_o(\boldsymbol{s})$ we have

$$F_{o;t,\boldsymbol{s}}(y) = 1 - \lambda_o(t, \boldsymbol{s})[1 - H(y - u_o(\boldsymbol{s}); \sigma_o(t, \boldsymbol{s}), \xi_o)]. \tag{4.3}$$

As with the estimation of the quantiles below $u_o(\boldsymbol{s})$ for X_o, we use information from X_c to provide a spatial covariate for $\sigma_o(t, \boldsymbol{s})$. We aim to learn about temporal non-stationarity exclusively from the observational data, so only information about the spatial variation of the marginal tail distribution is taken from X_c. We fit a model of the form

$$Y_{\text{c}}(t, \boldsymbol{s}) \sim \text{GPD}\{\sigma_{\text{c}}(\boldsymbol{s}), \xi_{\text{c}}\}, \tag{4.4}$$

for the excesses of $u_c(\boldsymbol{s}) = q_c^{(0.9)}(\boldsymbol{s})$. When modelling the climate model data we believe we have sufficient observations and spatial consistency, from their generation, to treat $\sigma_{\text{c}}(\boldsymbol{s})$ as site-specific, i.e., not imposing any spatial smoothness on the GPD scale parameters over $\boldsymbol{s} \in \mathcal{S}$. Clearly, it would be wrong to smooth well-estimated parameters spatially if we want to capture the relevant geophysical features of the climate system. As discussed in Section 4.3.1 we do not allow the temporal variation in the climate model to be informative about the observational data so we keep σ_{c} constant over t.

Full likelihood inference is not possible as any realistic model for the station data is likely to be highly complex, requiring spatial and temporal dependence of the data

to be modelled. Instead, we use a pseudo-log-likelihood

$$p\ell_c\left(\sigma_c(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}_c; \xi_c\right) = \sum_{\boldsymbol{s}\in\mathcal{S}_c}\left[\sum_{t\in\mathcal{T}_c} \log h(y_{c,t,\boldsymbol{s}}; \sigma_c(\boldsymbol{s}), \xi_c)\right], \tag{4.5}$$

constructed under the false assumption of spatial and temporal independence, with h being the density function of the GPD. This inference approach is commonly used, e.g., by Davison et al. (2012). The maximisation of this function can be broken down into a series of 1-dimensional optimisations by alternating the maximisation over ξ_c with the scale parameters fixed, and then exploiting the partition of the function $p\ell_c$ with respect to $\boldsymbol{s}$ when maximising over each $\sigma_c(\boldsymbol{s})$ in turn whilst treating ξ_c as constant. Iterating in this way until convergence is achieved gives estimated values $\{\hat{\sigma}_c(\boldsymbol{s}\}; \boldsymbol{s} \in \mathcal{S}_c\}$ and $\hat{\xi}_c$.

Next, we model the extreme observational data excess above the threshold, $u_o(\boldsymbol{s})$, denoted by $Y_o(t, \boldsymbol{s})$. The generic form of each of the models we consider is

$$Y_o(t, \boldsymbol{s}) \sim \text{GPD}(\sigma_o\{\boldsymbol{z}_t, \hat{\sigma}_c(\boldsymbol{s})\}, \xi_o),$$

where we model $\log \sigma_o\{\boldsymbol{z}_t, \hat{\sigma}_c(\boldsymbol{s})\}$ as either a parametric linear model of $\log \hat{\sigma}_c(\boldsymbol{s})$ and the covariates $\boldsymbol{z}_t$ (defined in Section 4.2.3) or via a generalized additive model (GAM) formulation. We denote the parameters of σ_o by $\boldsymbol{\theta}$. As with the inference for the climate model data we have to use a pseudo-log-likelihood, constructed under the false assumption of spatial and temporal independence. For the fully parametric model, the pseudo-log-likelihood is

$$p\ell_o\left(\boldsymbol{\theta}; \xi_c\right) = \sum_{\boldsymbol{s}\in\mathcal{S}_o}\left[\sum_{t\in\mathcal{T}_o} \log h(y_{o,t,\boldsymbol{s}}; \sigma_o, \xi_o)\right],$$

whereas in the GAM setting $p\ell_o$ is adapted by incorporating an additive spline smoothing penalty term (Wood 2006). Given the use of a pseudo-penalised likelihood, we cannot use standard methods for the evaluation of parameter uncertainty and model selection. Instead, the approaches we use are discussed in Section 4.3.4, with our marginal tail inference for the data being presented in Section 4.3.5.

4.3.4 Model uncertainty quantification and selection

In cases where a pseudo-likelihood is used, as in Section 4.3.3, the most widely adopted method for model selection is to adapt standard information criteria to account for model/likelihood mis-specification to greater penalise complexity relative to a better pseudo-likelihood fit. For spatial extremes, the composite likelihood information criterion (CLIC, Davison et al. 2012) is sometimes used, which includes a first-order asymptotically motivated additive adjustment factor. However, despite being used in many pseudo-likelihood approaches, we have chosen not to use CLIC for model selection. This is because the likelihoods for extremes are far from the asymptotic elliptical forms around the mode, yet CLIC relies on such asymptotic theory; CLIC measures only goodness of fit in the sample yet we have rich enough data to exploit out-of-sample model assessment; and for determining the parameter uncertainty we are not relying on asymptotic theory. Below we outline the alternative approaches we use.

4.3.4.1 Bootstrap methods

For both model selection and parameter uncertainty evaluation we generate bootstrapped samples X_o^* of $\{X_{\text{o}}(t, \boldsymbol{s}) : t \in \mathcal{T}_o, \boldsymbol{s} \in \mathcal{S}_o\}$ for a given marginal distribution model. These bootstrap samples need to preserve all spatial dependencies, short-range temporal dependence consistent with the passage of weather systems, missing data patterns of the observational data, and to exhibit the temporal non-stationarity of the fitted model.

For a given marginal model, the bootstrap takes the set of transformed observed data $\{x_{\text{o}}^U(t, \boldsymbol{s}) = F_{\text{o},t,\boldsymbol{s}}(x_{\text{o}}(t, \boldsymbol{s})) : t \in \mathcal{T}_o, \boldsymbol{s} \in \mathcal{S}_o\}$ where $F_{\text{o},t,\boldsymbol{s}}$ is given by the two model components of Section 4.3.2 and Section 4.3.3. The $x_{\text{o}}^U(t, \boldsymbol{s})$ values are realisations of Uniform$(0, 1)$ random variables that are identically distributed over time for each $\boldsymbol{s} \in \mathcal{S}_o$, but with the temporal and spatial dependence structure of the $X_o(t, \boldsymbol{s})$ process retained. To these data, we apply a vector temporal block bootstrap, with details of block structure and adaptions to account for the missing data described in Appendix B.3.4. For each bootstrapped data set $X^{U*} = \{X_{\text{o}}^{U*}(t, \boldsymbol{s}) : t \in \mathcal{T}_o, \boldsymbol{s} \in \mathcal{S}_o\}$ we use the inverse of the distribution function $F_{o,t,\boldsymbol{s}}$ to create the bootstrapped sample X_o^* with elements

$$X_o^*(t, \boldsymbol{s}) = F_{o,t,\boldsymbol{s}}^{-1}\{X_o^{U*}(t, \boldsymbol{s})\}. \tag{4.6}$$

Applying this raw bootstrap method induces bias in parameter estimates, and hence in sampling distribution estimates. The bias stems from ties in the extreme bootstrapped data that this method produces. As the very largest observations in a data set are known to be the most influential on the GPD model fit (Davison & Smith 1990) this is particularly problematic. There is negative bias in the estimate of the shape parameter of the GPD. Since the shape and scale parameters are negatively correlated, there is also a positive bias in the scale parameter estimator. To adjust for these biases we use a bootstrap error correction, as described in Appendix B.3.5. This is a two-step procedure with a location adjustment to the bootstrapped shape parameter estimate, then the scale parameter is re-estimated after fixing the adjusted shape parameter.

4.3.4.2 Cross-validation

For model fit diagnostics, we use two types of cross-validation (CV) to evaluate the performance of our models on out-of-sample data (Hastie et al. 2001, Ch 7.). We use standard 15-fold CV (15-CV) so that the data are divided into 15 groups (folds), where each fold is removed in turn and the model is fitted to the remaining folds. Since standard CV can perform poorly when the data have spatial or temporal correlation (Roberts et al. 2017), we also use a spatio-temporal CV (ST-CV) with 15 folds, corresponding to 5 spatial clusters of station data (i.e., divided spatially into 5 contiguous groups) and 3 temporal folds. Each temporal fold consists of every third week in the summer months, preserving long-term temporal non-stationarity. We choose 5 spatial partitions of our 182 sites as being low enough to help account for, and reduce, bias introduced via spatial auto-correlations as well as being sufficiently large that it reduces variance in our performance metrics across folds (Schratz et al. 2019). We define the 15 ST-CV folds as all combinations of spatial and temporal clusters, taking the intersection as a fold. We also investigated higher number of spatial clusters (up to 30) and higher number of random folds (up to 90) and found equivalent model performance rankings as those presented here.

For each left out fold, we compute two different goodness-of-fit measures to evaluate out-of-sample performance, the root mean square error (RMSE) and the continuous ranked probability score (CRPS, Gneiting & Katzfuss 2014). The RMSE evaluates the general closeness between the empirically estimated and predicted quantiles,

whilst the CRPS aims to match both the calibration and the sharpness of these extremes quantiles (Zamo & Naveau 2018). Here the empirical quantile, $x_o^{(\tau)}(t, \boldsymbol{s})$ is evaluated using the ordered data at site $\boldsymbol{s}$ and the year which contains time t, whereas the predicted quantiles are estimated as $\hat{x}_o^{(\tau)}(t, \boldsymbol{s}) = F^{-1}_{o,t,\boldsymbol{s}}(\tau)$ for quantile τ from the appropriate model. The comparisons between $x_o^{(\tau)}$ and $\hat{x}_o^{(\tau)}$, for the same t, $\boldsymbol{s}$, and τ, are averaged across the folds. Lower values of RMSE and CPRS generally indicate a superior fit.

4.3.5 Marginal data analysis

4.3.5.1 Body of distribution

Following exploratory analysis, we identified three potential models for the body of the distribution, which we present in Table 4.1 along with their CV RMSE. The first model serves as a base, in which the location parameter is constant over space and time for each quantile τ. In the second model we allow the quantile regression to vary spatially by using the corresponding climate model data quantiles $q_c^{(\tau)}(\boldsymbol{s})$ as a covariate. The third model also includes the temporal Irish mean temperature covariate $M^I(t)$. The inclusion of the climate model covariate reduces the RMSE for both types of CV, whereas $M^I(t)$ improves the CV scores further, though not as much. We fitted a number of other covariate combinations for $\boldsymbol{z}_t$, as well as using the principal components of $\boldsymbol{z}_t$ to avoid issues of collinearity. Overall, we found the third model provides the best balance of simplicity and fit, so use this for subsequent analysis.

Table 4.1: Cross-validation (RMSE) on the quantile regression analysis for the body of the distribution.

Model structure for $\hat{q}_o^{(\tau)}(t, \boldsymbol{s})$	**ST-CV**	**15-CV**
$\beta_0^{(\tau)}$	1.442	1.455
$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\boldsymbol{s})$	1.350	1.347
$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\boldsymbol{s}) + \beta_2^{(\tau)} M^I(t)$	**1.322**	**1.319**

Appendix B.3.1 provides estimates of $\beta_2^{(\tau)}$, which show a slight decrease with τ although the confidence intervals widen. For all τ, $\beta_2^{(\tau)} = 1$ appears consistent with the data, indicating that mean summer temperatures in Ireland are a good representation of the temporal change for all the body of the distribution. The estimates

of $\beta_1^{(\tau)}$ (not plotted) decrease, approximately linearly, from 0.75 to around 0.65 with $0 < \tau < 1$, showing that the climate model is not giving identical descriptions to the station data, as the estimates differ from 1 significantly and change with τ.

4.3.5.2 Tails of the distribution

For selecting the threshold $u_o(\boldsymbol{s})$, we use the second model in Table 4.1 with $\tau = 0.9$, providing a threshold that varies in space but not time. Figure 4.2 (left) shows the threshold $u_o(\boldsymbol{s})$ over Ireland, with cooler temperature values on the west of Ireland and coastal regions on the south and north coasts. For this $u_o(\boldsymbol{s})$ we estimate the threshold exceedance probability $\lambda_o(t, \boldsymbol{s})$ and its spatial average $\lambda_o(t) = \int_{\boldsymbol{s}\in\mathcal{S}_c} \lambda_o(t, \boldsymbol{s}) d\boldsymbol{s}/|\mathcal{S}_c|$. Estimates of the latter are shown in Appendix B.3.1. The $\lambda_o(t)$ estimates show an increasing exceedance rate, with the average rate over time of 0.1 reflecting the choice of the threshold, increasing by around 35% with 95% confidence interval 28-44% from 1942–2020. We see the same features at individual sites, but with wider confidence intervals.

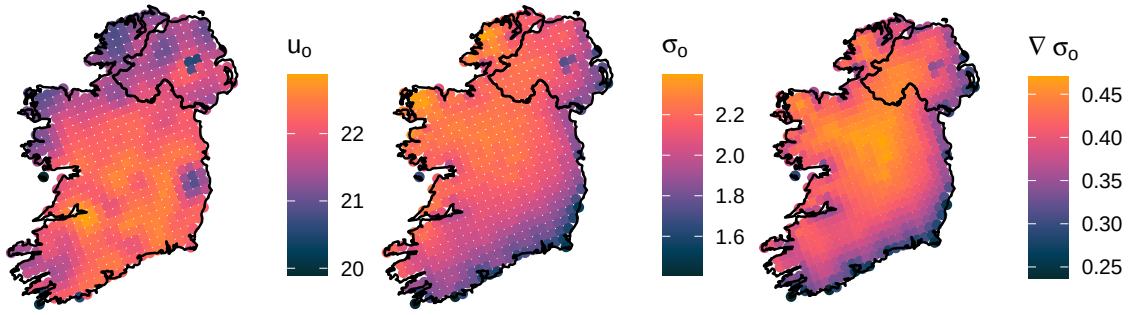

Figure 4.2: Estimated values of threshold $u_o(\boldsymbol{s})$ (left), GPD scale parameter according to M_2 in 2020, (centre), and the estimated change in the scale parameter since 1942, $\nabla\sigma_o(\boldsymbol{s})$ (right).

Table 4.2 presents a subset of the models that we explored for the GPD scale parameter: incorporating climate model data via σ_c, defined by expression (4.4) and constant over time in model M_0; allowing for temporal non-stationarity via $M^I(t)$ in M_1; allowing also temporal non-stationarity via $M^I(t)$ in M_2 as both a constant rate of change and as an interaction with coastal distance $C(\boldsymbol{s})$. Other models were attempted with differing covariates and spline structures included but these failed to improve over models M_0–M_2. However, over a range of spline models

we noticed that they were consistently suggesting evidence for different temporal trends on the coast relative to inland, hence our introduction of the $C(\boldsymbol{s})$ covariate.

Table 4.2: Models M_0–M_2 for GPD log-scale parameter, $\log \sigma_o(t, \boldsymbol{s})$, along with cross-validation results and estimated shape parameter, ξ_o, with bootstrapped 95% confidence intervals. Numbers in bold font show the lowest CV values.

		ST-CV		**15-CV**		ξ_o
		RMSE	CRPS	RMSE	CRPS	
M_0	$\beta_0 + \beta_1 \log \sigma_c(\boldsymbol{s})$	0.945	0.895	0.928	0.882	$-0.152(-0.237, -0.092)$
M_1	$\beta_0 + \beta_1 \log \sigma_c(\boldsymbol{s}) + \beta_2 \mathrm{M}^I(t)$	0.938	0.894	0.918	0.880	$-0.156(-0.204, -0.110)$
M_2	$\beta_0+\beta_1 \log \sigma_c(\boldsymbol{s})+\beta_2 \log \mathrm{C}(\boldsymbol{s})+$ $\beta_3 \mathrm{M}^I(t) + \beta_4 \log \mathrm{C}(\boldsymbol{s}) \mathrm{M}^I(t)$	**0.934**	**0.893**	**0.908**	**0.878**	$-0.158(-0.194, -0.110)$

Table 4.2 presents our model selection diagnostics based on CV metrics (CRPS and RMSE). All four approaches favour model M_2, with M_0 and M_1 having similar, slightly inferior performance, and we find that M_0 is too simplistic. Models M_0–M_2 estimate the coefficient of $\log \sigma_c$ as close to 1 in all cases, showing that the climate model is providing very helpful information as a spatial covariate. The estimates of the shape parameter ξ_o are also given in Table 4.2. As the GPD scale parameter model is made increasingly flexible (from model M_0 to M_2), the value of ξ_o decreases, lightening the tail decay, indicating that each model is progressively reducing sources of variation in the tail. Since there is some uncertainty in the marginal model choice, we take M_0, M_1, and M_2 through the spatial dependence analysis to assess the sensitivity of the risk measures, with details for model M_2 reported here and for M_0 and M_1 in Appendix B.

Model M_2 shows that the most variable excess distribution is on the west coast (see centre plot in Figure 4.2), with a decay in values from west to east, so almost the opposite of the behaviour of $u_o(\boldsymbol{s})$. We also investigated the estimated change in the scale parameter over the observation period, denoted $\nabla \sigma_o(\boldsymbol{s}) = \sigma_o(2020, \boldsymbol{s}) - \sigma_o(1942, \boldsymbol{s})$, see Figure 4.2 (right), and found it to be largest in the centre of Ireland, with the change there being close to double that on the coast. The scale parameter is increasing over time everywhere, leading to warmer extreme temperatures.

The model selection diagnostics in Table 4.2 show primarily the relative quality of the three model fits. To assess the absolute quality of the fitted model M_2 we use

pooled QQ plots in Figure 4.3, pooling over all sites and years. Due to the spatio-temporal non-stationarity of the marginal model, we transform the data through our fitted model into a common uniform scale and to a common exponential scale (for the conditional distribution of threshold excesses). The choice of scales helps identify key departures of fit in the body and tails of the distribution respectively. We see evidence of an exceptionally strong fit in both components of the distribution, with values near the lines of equality, and in the far upper tail, all values falling within the pointwise tolerance bounds which were derived assuming independence of time and space (so are much narrower than necessary). Examples of similar site-specific QQ plots are shown in the appendix, Figure B.13 for the five stations identified in Figure 4.1 and for five other randomly selected stations. These show a slightly more varied quality of fit, with the least good fits occurring on the coastlines, e.g., Malin Head, but with very good fits at most stations.

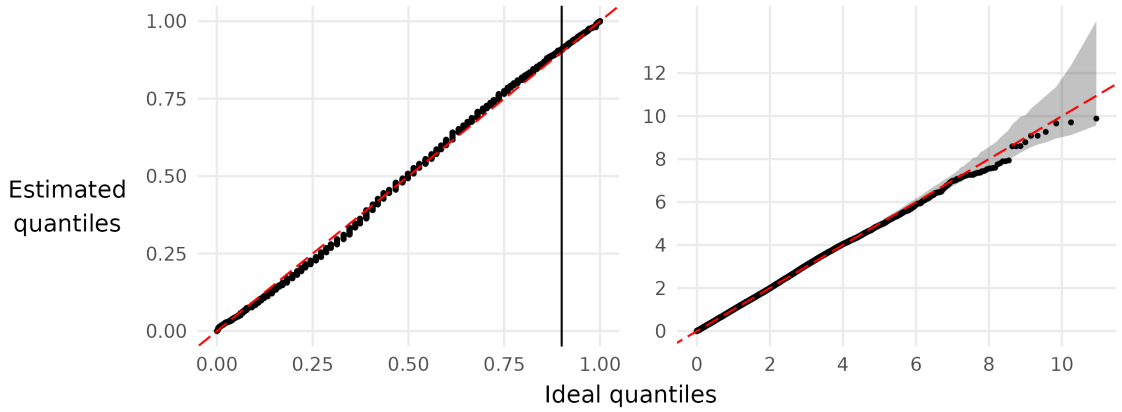

Figure 4.3: Spatially and temporally pooled QQ-plots for model M_2: all data on uniform margins, threshold shown as vertical line (left); tail model (GPD) on exponential margins. The shaded region shows pointwise 95% tolerance intervals (right). The lines of equality are in red.

4.4 Spatial models

4.4.1 Standardising data

When modelling dependence between variables with differing marginal distributions and covariates, it is common to first standardise the marginal variables so that they have an identical distribution over variables and covariates (Coles 2001, de Haan

& Ferreira 2006). Here we transform the data to (unit) Pareto distributions, X^P, X_c^P, and X_o^P, using the same subscript notation as in Section 4.3. The choice of Pareto marginal scale is ideal for studying asymptotically dependent variables, a property defined in Section 4.4.2, such as r-Pareto processes (de Haan & Ferreira 2006), but less ideal for asymptotically independent variables, where shorter-tailed Exponential or Laplace distributions are favoured (Wadsworth & Tawn 2022). We make use of the probability integral transform, i.e.,

$$X^P(t, \boldsymbol{s}) = 1/\left[1 - F_{t,\boldsymbol{s}}\{X(t, \boldsymbol{s})\}\right], \text{ for all } \boldsymbol{s} \in \mathcal{S} \text{ and all } t, \tag{4.7}$$

where the marginal distribution function $F_{t,\boldsymbol{s}}$ takes a different estimated form below and above $u(\boldsymbol{s})$, see Section 4.3.2 and Section 4.3.3. Thus, if the marginal model is perfectly estimated, we have $\Pr(X^P(t, \boldsymbol{s}) > y) = y^{-1}$ for all $y > 1$, t, and $\boldsymbol{s}$. In our uncertainty assessment in the subsequent inference, the marginal model uncertainty is accounted for through our bootstrap procedures. To transform from standard Pareto margins to the original distribution at $\boldsymbol{s}$ and t, the inverse of the transformation (4.7) is used.

4.4.2 Classification of extremal dependence type

We now explore the nature of the extremal spatial dependence structure in the processes X_c^P and X_o^P. To simplify notation we omit the temporal dimension of these spatial processes but always consider the process on the same day at different locations. Following Coles et al. (1999), we estimate the pairwise coefficient of asymptotic dependence, χ. Specifically, for the process X^P at sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, $\chi = \chi^P(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is defined by

$$\chi^P(\boldsymbol{s}_i, \boldsymbol{s}_j) = \lim_{v \to \infty} \Pr(X^P(\boldsymbol{s}_j) > v \big| X^P(\boldsymbol{s}_i) > v).$$

If $\chi^P(\boldsymbol{s}_i, \boldsymbol{s}_j) > 0$ (or equals 0) then X^P is said to be asymptotically dependent (or asymptotically independent) respectively for these sites. The larger the value of χ ($0 < \chi \leq 1$) the stronger the asymptotic dependence.

The selection of the appropriate extremal dependence model for the data depends on whether or not the process is better approximated as being asymptotically dependent for all $\boldsymbol{s}_i, \boldsymbol{s}_j \in \mathcal{S}$ or not. The base quantity that is typically used to identify

asymptotic dependence for a pair of sites is $\chi^P(\cdot,\cdot;p)$, where

$$\chi^P(\boldsymbol{s}_i,\boldsymbol{s}_j;p) = \Pr(X^P(\boldsymbol{s}_j) > v_p \big| X^P(\boldsymbol{s}_i) > v_p), \tag{4.8}$$

with $v_p = 1/(1-p)$ being the pth marginal quantile of X^P. An empirical estimate of $\chi^P(\boldsymbol{s}_i,\boldsymbol{s}_j;p)$ exploits the replication over t by assuming spatial dependence does not change with t. We denote this estimator by $\tilde{\chi}^P(\boldsymbol{s}_i,\boldsymbol{s}_j;p)$. We expect approximate spatial stationarity and isotropy of the spatial extreme process. Plotting (not shown) the cloud of $\tilde{\chi}^P(\boldsymbol{s}_i,\boldsymbol{s}_j;p)$ against the Euclidean distance between the sites ($h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$), reveals a decay with distance that is somewhat hidden by the sampling variation of the points, with the variation depending on the overlap in time of samples at the pairs of sites. A better empirical estimate of $\chi^P(h;p)$, the pairwise extremal dependence at separation distance h, exploits the property that it changes smoothly over h and that we can obtain the sampling distribution of $\tilde{\chi}^P(\boldsymbol{s}_i,\boldsymbol{s}_j;p)$ through the bootstrap. Together these enable us to construct a weighted estimate $\tilde{\chi}^P(h;p)$ from the cloud of points (using pairs with h_{ij} close to h) and obtain its sampling distribution. We used 300 bootstraps and 30 binned distances, each with an equal number of pairs of sites.

Figure 4.4 shows the behaviour of $\tilde{\chi}^P(h;p)$, for both X_c^P and X_o^P processes. It shows estimates and intervals that account for 95% of the marginal estimation uncertainty, which for X_o we used model M_2. These estimates are shown for $p \in (0.8, 0.85, 0.9)$, the latter corresponding to only 9 days per summer. Despite the climate model having a much richer set of pairs of sites and longer simultaneous data, both processes provide very similar qualitative findings. Naturally, $\chi^P(h;p)$ decreases with distance but it is far from zero, in all cases, even at the longest distances of pairs of sites from $\mathcal{S}$. For short distances, the estimates of $\chi_c(h;p)$, exhibit less variation than in the estimated values for $\chi_o(h;p)$. We have that $\hat{\chi}_o^P(h;p) \leq \hat{\chi}_c^P(h;p)$ for all distances, suggesting that the X_c data are overestimating the extremal dependence in X_o. This difference is important when looking at extreme events spatially, as it suggests using the climate model data alone (or when down-scaled) will lead to an overestimate of the risk of widespread heatwaves in Ireland.

Most critical for our modelling of the observed process is to assess whether, as p increases to 1, the $\hat{\chi}_o^P(h;p)$ values decay to zero or stabilise at a non-zero limit indicating asymptotic independence and asymptotic dependence respectively. There

is a small decline, at all distances, however even when $p = 0.9$ these estimates are far from zero for both X_c^P and X_o^P. So we conclude that it seems reasonable that maximum daily temperature data are consistent with asymptotic dependence over Ireland.

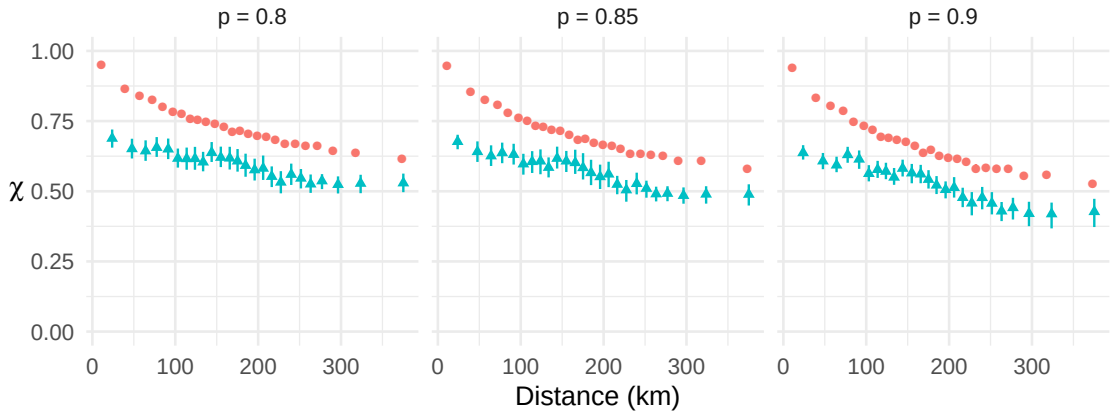

Figure 4.4: Empirical estimates $\tilde{\chi}^P(h; p)$ plotted against inter-site distance h for the climate model (χ_c^P) (orange) and station data (χ_o^P) (blue). Plots use the marginal model M_2 for $p = 0.8, 0.85$ and 0.9 (left to right) with 95% confidence intervals shown as vertical lines.

4.4.3 r-Pareto Processes

We now model the extreme values of the process $X^P(t, \boldsymbol{s})$ over $\boldsymbol{s} \in \mathcal{S}$, with unit Pareto distributed marginal variables. We look at spatial fields separately for each t and simplify notation by dropping the argument t. First, we define what we mean by a spatially extreme event as there is no natural ordering of multivariate or spatial processes. Here the level of extremity of the stochastic process $X^P := \{X^P(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ is determined by a risk function $r(X^P) \in [0, \infty)$, where the only constraint on r is that it is homogeneous of order 1, i.e., $r(c\boldsymbol{x}) = cr(\boldsymbol{x})$ for any constant $c > 0$ and with $\min(\boldsymbol{x}) > 0$. de Fondeville & Davison (2018, 2022) suggest taking r as the magnitude at one particular site of interest, or the spatial mean, median, maximum, or minimum over $\mathcal{S}$.

Under weak conditions on X^P, de Fondeville & Davison (2018) report that

$$\Pr\left\{\left[v^{-1}X^P(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\right] \in \cdot \mid r(X^P) > v\right\} \to \Pr\left\{\left[Y_r^P(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\right] \in \cdot\right\} \tag{4.9}$$

as $v \to \infty$, where $\{Y_r^P(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\} := Y_r^P$ is marginally non-degenerate for all $\boldsymbol{s} \in \mathcal{S}$, with Y_r^P a r-Pareto process. Limit (4.9) implies that scaled events of the X^P process with risk exceeding a threshold of v are increasingly well-approximated by an r-Pareto process, as the risk threshold increases to infinity. The limit (4.9) is used for statistical modelling by taking it as an equality for a suitably large value for v, denoted by v_r with $v_r > 1$, then those spatial events with a risk function exceeding v_r are treated as realisations from an r-Pareto process. Specifically, taking an extreme set $B \subseteq \{\boldsymbol{x} : r(\boldsymbol{x}) > 1\}$ leads to the modelling assumption that

$$\Pr\left\{X^P \in v_r B \mid r(X^P) > v_r\right\} = \Pr\left(Y_r^P \in B\right). \tag{4.10}$$

Hence, defining the set $A = v_r B$ and un-doing the conditioning on the left-hand side of equality (4.10), for any $A \subseteq \mathcal{A}_r := \{\boldsymbol{x} : r(\boldsymbol{x}) > v_r\}$ we obtain that

$$\Pr\left\{X^P \in A\right\} = \Pr\{r(X^P) > v_r\} \Pr\left(v_r Y_r^P \in A\right). \tag{4.11}$$

The r-Pareto process exhibits properties which can be exploited for efficient evaluation of $\Pr\left(X^P \in A\right)$. Specifically, Y_r^P decomposes into two independent components:

$$Y_r^P(\boldsymbol{s}) = RW_r(\boldsymbol{s}) \text{ for all } \boldsymbol{s} \in \mathcal{S}, \tag{4.12}$$

where R is unit Pareto distributed and is interpreted as the risk of the process, and $W_r := \{W_r(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ is a stochastic process which describes the spatial profile of the extreme event, i.e., the proportion of the risk function r contributed by each site. The limiting dependence structure of X^P is entirely determined by the stochastic properties of W_r. By construction $R = r(Y_r^P)$ and $r(W_r) = 1$. A consequence of the limiting approximation (4.10) holding above v_r and R having a Pareto distribution is that expression (4.11) simplifies as we have $\Pr\{r(X^P) > v_r\} = c_r v_r^{-1}$ where, in general, $c_r \in (0, 1]$ depends only on the choice of risk functional and the dependence structure of Y. However, for our choice of r, given by expression (4.15), $c_r = 1$ always, see Coles & Tawn (1994).

The characterisation (4.12) is powerful for the extrapolation to larger events than those observed due to R having a known distribution and the independence property ensuring that the spatial profiles of larger events have exactly the same stochastic properties for any event with a risk greater than 1. For any r-Pareto process and

a set $A \subseteq \mathcal{A}_r$, there exists a constant $b_A \in [1, \infty)$ such that for any $b \in [1, b_A]$ we have that

$$\Pr\left(v_r Y_r^P \in A\right) = b^{-1} \Pr\left(b v_r Y_r^P \in A\right). \tag{4.13}$$

Although the two sides of this expression are equal, the two probabilities are not, with Opitz et al. (2021) noting that the latter is much more efficient to estimate using Monte Carlo methods. Taking $b > b_A$ will give bias, as some smaller outcomes in A will be missed by simulations of $bv_r Y_r^P$ while $b < b_A$ leads to unnecessary variability in the empirical estimator. So we look to scale by b_A in Section 4.4.4, where we discuss how to obtain b_A and illustrate its usage in estimating the right-hand side of expression (4.11).

The above shows that inference for any extreme events is relatively straightforward once we have a model for the process W_r. We follow Engelke et al. (2015) and de Fondeville & Davison (2018) by modelling W_r as a spatial stationary isotropic log-Gaussian stochastic process which is determined solely by a variogram $\gamma(h)$, for inter-site distance $h \geq 0$. We use the Matérn variogram family

$$\gamma_{\text{mat}}(h; t) = \alpha \left\{1 - (2\sqrt{\nu} h/\phi)^{\nu} 2^{1-\nu} \Gamma(\nu)^{-1} \mathrm{K}_{\nu}(2\sqrt{\nu} h/\phi)\right\}, \tag{4.14}$$

where K_ν is a modified Bessel function of the second kind and the positive parameters $(\alpha = \alpha_t, \phi = \phi_t, \nu = \nu_t)$ determine the variance, range, and smoothness respectively at time t. Our choice of a bounded variogram was based on the evidence from Figure 4.4 which suggested that the summer temperature process is asymptotically dependent, even at the longest distances in Ireland. See Appendix B.7 for additional discussion on the choice of variogram function including the support for isotropy.

The issue of missing data in spatial extremes applications seems to be rarely discussed. A possible reason for this is that with composite likelihood fitting methods for max-stable processes (Padoan et al. 2010) the implications are restricted as only pairwise joint likelihood contributions are used so the impact of missing data is limited. This is not the case for r-Pareto processes, which model jointly across all sites, and the issue of missing data in this context does not seem to have been discussed. When encountering missing data it is tempting to remove all observations at that time point from all sites in the network. However, we observed in Section 4.2 that this tactic would leave us with no data. Fortunately, thanks to the properties

of the log-Gaussian process, it is possible to show that the model is closed under marginalisation (Engelke & Hitz 2020).

With missing data, we need to be careful in selecting a suitable risk function r. The choice of risk function needs to be invariant to the changing dimension of partially observed events, whatever their missing patterns. Hence, in our statistical inference at time t, for all $t \in \mathcal{T}_o$, we take the risk function r_t to be the average of the standardised variables over the stations which were observed at time t, i.e.,

$$r_t(X_o^P(t, \boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}) = \sum X_o^P(t, \boldsymbol{s}_i) I_o(t, \boldsymbol{s}_i) \Big/ \sum I_o(t, \boldsymbol{s}_i), \tag{4.15}$$

where $I_o(t, \boldsymbol{s}_i)$ is the indicator variable for whether $X_o(t, \boldsymbol{s}_i)$ is observed or not, and the sums are from $i = 1, \ldots, |\mathcal{S}_o|$. For evaluating r_t we would have liked to use a subset of stations that are observed for all t and reasonably evenly spread across Ireland, but this was not possible. However the data from Aldergrove (north) are used as this site has very little missing data over the period 1942–2020.

We set the risk threshold v_r, used in expression (4.10) to define extreme spatial events, at the 80% sample quantile of the risk values calculated from all observed events, i.e., we use the empirical estimate $\widetilde{\Pr}\{r(X_r^P) > v_r\} = 0.2$. We explored different threshold choices and selected the lowest level we could whilst making the usual bias/variance trade-off for tail selection. Figure B.16 and Figure B.17 in the appendix show that the parametric estimate of χ_o^P, derived from the variogram, agrees well with empirical estimates for each marginal model M_0, M_1, and M_2.

We explored the effect of a time-changing dependence model. We allowed for the variance and range parameters of the Matérn variogram (5.7) to vary over time t, while keeping ν constant. We considered a range of constant and log-linear models using each of the marginal models M_0 to M_2. For M_0 we find some evidence, at the 5% level, for α_t increasing with $M^I(t)$ (weakening dependence over time), but not for the improved marginal models M_1 or M_2. Evidence for a change in extremal dependence was not statistically significant and so we keep a temporally stationary r-Pareto process.

4.4.4 Simulation and efficient inference for spatial extreme events

We simulate spatial extreme events on the observational scale in year t by first simulating an event from the r-Pareto process and then map this pointwise to the data scale using the inverse of transform (4.7) for the required t. For each step of this process we use the selected statistical model and the simulated values are generated using the fitted parameters of that model, or for assessing uncertainty in the point estimates, using the bootstrapped realisations of these parameters. As the estimated r-Pareto process in our application is found to be well-approximated by a stationary process over time, we can generate identically distributed events of the r-Pareto process to transform for each location and time t using the time-varying marginal model. The r-Pareto process simulations are generated using the R package `mvPot` (de Fondeville et al. 2021). We denote these simulations by $\boldsymbol{y}_1^P, \boldsymbol{y}_2^P, \ldots, \boldsymbol{y}_m^P$, for m simulations, with the ith simulation consisting of the spatial realisation $\boldsymbol{y}_i^P = \{y_i^P(\boldsymbol{s}) : \ \boldsymbol{s} \in \mathcal{S}\}$. For the ith realisation of the r-Pareto process, $\boldsymbol{y}_i^P$, we define , $r_i = r(\boldsymbol{y}_i^P) > 1$ as the risk, $\boldsymbol{w}_i = \boldsymbol{y}_i^P / r_i$ as the spatial profile, and $w_i(\boldsymbol{s}) = y_i^P(\boldsymbol{s})/r_i$ as the value of the r-Pareto process at site $\boldsymbol{s} \in \mathcal{S}$.

Appendix B.5, Figure B.21 shows five simulated extreme events, transformed to observational scale under 2020 conditions and the exact same events in 1942 conditions (presented as a difference in temperatures at each site, for each event). A positive difference shows the equivalent event in the two years to be hotter in 2020 than 1942, with that difference found to be largest for the hotter events. As the r-Pareto process realisations can have marginal values in the range $(0, 1)$ at some sites, i.e., outside the domain of the Pareto variable, we follow de Fondeville & Davison (2018) and use Fréchet marginals rather than Pareto for transformation to the observational space.

Although expression (4.11) provides a basis for inference for the probability of occurrence in any extreme event $A \subseteq \mathcal{A}_r$ by the process X^P, we are most interested in making inference about spatial events of the observational process which exceed a critical temperature of T°C somewhere over Ireland at time t. We denote these events by

$$A_{t,\mathcal{S}}(T) = \{X_o(t, \boldsymbol{s}), \boldsymbol{s} \in \mathcal{S} : \exists \ \boldsymbol{s}_0 \in \mathcal{S} \text{ with } X_o(t, \boldsymbol{s}_0) > T\}. \tag{4.16}$$

After the marginal transformation to Pareto margins, this event is equal to

$$A_{t,\mathcal{S}}(T) = \{X_o^P(t,\boldsymbol{s}), \boldsymbol{s} \in \mathcal{S} : \exists\ \boldsymbol{s}_0 \in \mathcal{S} \text{ with } X_o^P(t,\boldsymbol{s}_0) > T^P(t,\boldsymbol{s}_0)\},$$

where $T^P(t,\boldsymbol{s})$ is the mapping of T through the transformation (4.7) at time t and for site $\boldsymbol{s}$. To use the r-Pareto approximation, all elements of $A_{t,\mathcal{S}}(T)$ must have a risk exceeding v_r; this imposes a lower bound $T > 20.6°\text{C}$ across all t. We also focus on marginal extreme events, which further restricts $T \geq \max_{i\in\mathcal{S}} u_i = 22.9°\text{C}$. As all the results we present in Section 4.5 are for $T \geq 26°C$ this lower bound is not restrictive for our purposes.

To estimate $\Pr\{A_{t,\mathcal{S}}(T)\}$, there have been a set of possibilities proposed, see Appendix B.6.1. We focus on the most efficient of these estimators, which exploits the independence property (4.12), and the scaling property (4.13). Specifically, $\boldsymbol{w}_i$ and r_j are independent realisations for all i, j, and there is no reason to restrict ourselves to the observed r_j as we know they are unit Pareto realisations. So we supplement the information to have $\{r_j^P; j = 1,\ldots,L\}$, which are *iid* realisations of a unit Pareto variable, where L is taken as large as possible to improve computational efficiency. To find the optimal scaling factor $b_{T(t)}$ we first define component-wise maxima of the simulated Pareto processes scaled to have unit cost, i.e., $\omega_{(m)}(\boldsymbol{s}) = \max_{i=1,\ldots,m} w_i(\boldsymbol{s})$, for each $\boldsymbol{s} \in \mathcal{S}$. At time t, we want to scale these component-wise maxima by as much as possible without producing a scaled event with an exceedance of $T^P(t,\boldsymbol{s})$ for some $\boldsymbol{s} \in \mathcal{S}$. The appropriate scaling is then $b_{T(t)} = \min_{\boldsymbol{s}\in\mathcal{S}} \left\{T^P(t,\boldsymbol{s})/\omega_{(m)}(\boldsymbol{s})\right\}$. Here $b_{T(t)} = v_r b_A$ in expression (4.13). Combined together, these give a form of importance sampling estimator

$$\widehat{\Pr}_{imp}\{A_{t,\mathcal{S}}(T)\} = \frac{1}{b_{T(t)} mL} \sum_{i=1}^{m}\sum_{j=1}^{L} I\left\{\exists \boldsymbol{s_0} \in \mathcal{S} : r_j^P b_{T(t)} \frac{y_i^P(\boldsymbol{s_0})}{r_i} > T^P(t,\boldsymbol{s_0})\right\}, \tag{4.17}$$

see Appendix B.5, Algorithm 2. With this scaling choice, and the extrapolation from the $r_j^P > \max(r_1,\ldots r_m)$, we are guaranteed to have at least L out of the mL simulated fields which achieve at least temperature $T°C$ somewhere in $\mathcal{S}$ in year t.

4.5 Temporal changes in spatial extreme events

We present a range of summaries detailing how spatial daily maximum temperature extreme events in Ireland are changing over the period 1942–2020. First, we look at the changes in the marginal quantiles. Figure 4.5 shows estimates of the level exceeded by daily values with probability 1/9200, i.e., that of a 100-year return level if the process was stationary in time. For simplicity, we refer to these as the 100-year levels changing over time. For model M_2, we show these estimates for 2020 and also present the estimated difference between them for 2020 and 1942. In the latter, a positive value represents a warming of temperatures. The 100-year return level has increased between $1.2 - 2.2$°C across Ireland, with the larger increases away from the south and east coasts. These changes in extreme temperatures over the observed record are substantially larger than the 1°C change of M^I and M^G over this period, illustrating that climate change is more radically affecting summer temperature extreme events than mean levels.

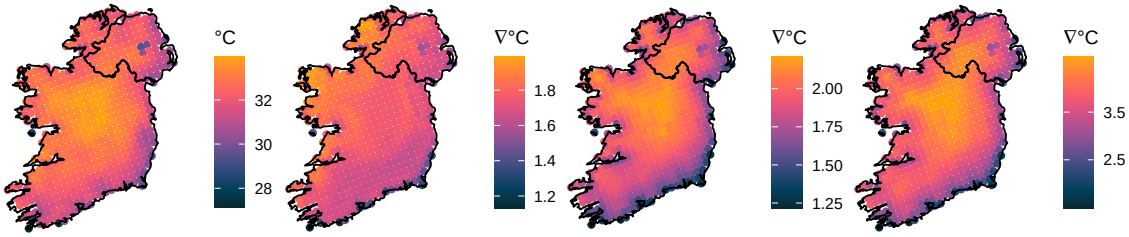

Figure 4.5: Estimated 100-year marginal return level for the year 2020 (1st plot); estimated change in 100-year marginal return level from 1942 to 2020 (3rd plot); lower and upper 95% CI limits respectively for the change in 100-year return level from 1942 to 2020 (2nd and 4th plots respectively).

In Figure 4.5 we also report the change in the upper and lower 2.5% quantile of change in return level from 1942 - 2020 across all bootstrap samples. In both cases, and at all sites, these changes are positive, with the rate of change in these features being greater than that of the point estimates. Although we present the results for the 100-year return levels, similar results hold for all high quantiles and for the finite upper endpoints of the marginal distribution; the latter as the GPD shape parameter being negative (see Table 4.2). Figures B.14 and B.15 give equivalent figures for M_0 and M_1.

Next, we consider summaries that also reflect the dependence structure of extreme events. There are no established analytical closed-form expressions of such changes. Instead, we revert to using simulated fields of extreme events and presenting risk measures based on empirical summaries using large samples of these fields. The simulation strategy set out in Section 4.4.4 gives replicated independent spatial fields. In particular, we focus on the occurrence of events $A_{t,\mathcal{S}}(T)$, i.e., an extreme temperature of at least T°C somewhere in Ireland (determined by the set of locations as required), and then summaries of the properties of such events. We estimate $\Pr(A_{t,\mathcal{S}}(T))$ using the estimator $\widehat{\Pr}_{imp}$, with $m = 25,000$ and $L = 300$. Figure 4.6 shows this estimated probability (expressed as a return period) for a range of temperatures $T \in [26, 36]$ for years 1942 and 2020 separately for $\mathcal{S}_o$ and $\mathcal{S}_c$. For $\mathcal{S}_o$, the plot reveals a marked change with estimated return periods being shorter in 2020 compared to 1942 for the same T. To illustrate this, consider the event with the hottest temperature observed anywhere at the station network, a temperature of 33°C at Phoenix Park, Dublin, July 2022. The spatial event $A_{t,\mathcal{S}}(33)$ changes from being a 1 in 182-year event in 1942 to a 1 in 8.7-year event in 2020. Furthermore, the model estimates that a temperature in excess of 34°C, i.e., a value not yet recorded in Ireland, changes from a 1 in 1,588-year event to a 1 in 27.5 year event over this time window.

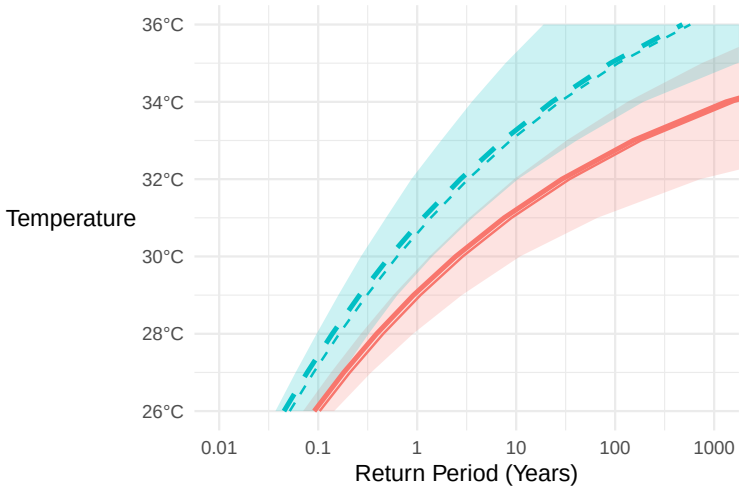

Figure 4.6: Return period of the event $A_{t,\mathcal{S}}(T)$ where an extreme temperature exceeding T°C occurs somewhere on the Irish station network, $\mathcal{S}_o$. Blue dashed (solid orange) lines correspond to $t = 2020$ (1942). Shaded regions show pointwise 95% confidence intervals for the return periods. The higher bold curves show the corresponding point estimates for the climate model grid $\mathcal{S}_c$.

Figure 4.6 shows that, for a given return period, hotter temperatures are expected somewhere in $\mathcal{S}_c$ than on $\mathcal{S}_o$, as the former is denser and with better coverage than the station network. The difference between the results for the two collections of sites is very small. This slight change shows that the station network, when all gauges are working, has the ability to fully capture all extreme temperature events over Ireland. Such information has not been available previously given the complexity of addressing spatial dependence, marginal non-stationary, and missing data in the station network.

We now propose risk metrics to summarise the features of events satisfying $A_{t,\mathcal{S}}(T)$. First consider a measure of pairwise dependence, which extends the idea behind χ^P but applied to the data scale, so it combines the effects of changes over time in the marginal distributions and the estimated extremal dependence structure. Specifically, we define

$$\chi_o(h; A_{t,\mathcal{S}}(T)) = \Pr\{X_o(t, \boldsymbol{s}^h) > T \mid A_{t,\mathcal{S}}(T)\},$$

where $\boldsymbol{s}^h$ is a randomly selected site in $\mathcal{S}$ with $\left\|\boldsymbol{s}^h - \boldsymbol{s}_0\right\| = h$, i.e., the conditional probability of the observational process exceeding temperature T on day t at a site which is a distance h away from a site $\boldsymbol{s}_0$ that has a temperature exceeding T on that day. We also investigate the associated unconditional risk measure

$$\chi_o(h; T, t) = \Pr(\exists\ \boldsymbol{s}_0 \in \mathcal{S} : \min[X_o(t, \boldsymbol{s}_0), X_o(t, \boldsymbol{s}^h)] > T).$$

Figure 4.7 presents estimates of each of these two risk measures for a range of h and T, between 1942 and 2020. On the observed data scale, we find that extremal spatial dependence decreases with distance as would be expected, but beyond this, there are quite different findings from the two measures. Risk measure $\chi_o(h; A_{t,\mathcal{S}}(T))$ is broadly stable over the presented range of T and t whilst the unconditional $\chi_o(h; T, t)$ is substantially different. The former is perhaps not too surprising given that the model is asymptotically dependent (the dependence structure is invariant to any change in the extremity of an event) and the extremal dependence structure is estimated to be stationary over time. However, as the event is on the marginal scale and the marginal distributions are changing with time, this finding was not anticipated. For $\chi_o(h; T, t)$ we do see that the joint probability of temperature being

above T at sites h apart changes notably with time, e.g., taking $h = 100$ km, we find that $\chi_o(h; T, t)$ has increased by a factor of 2.8, 3.5, and 4.7 for $T = 28, 29$, and 30°C respectively.

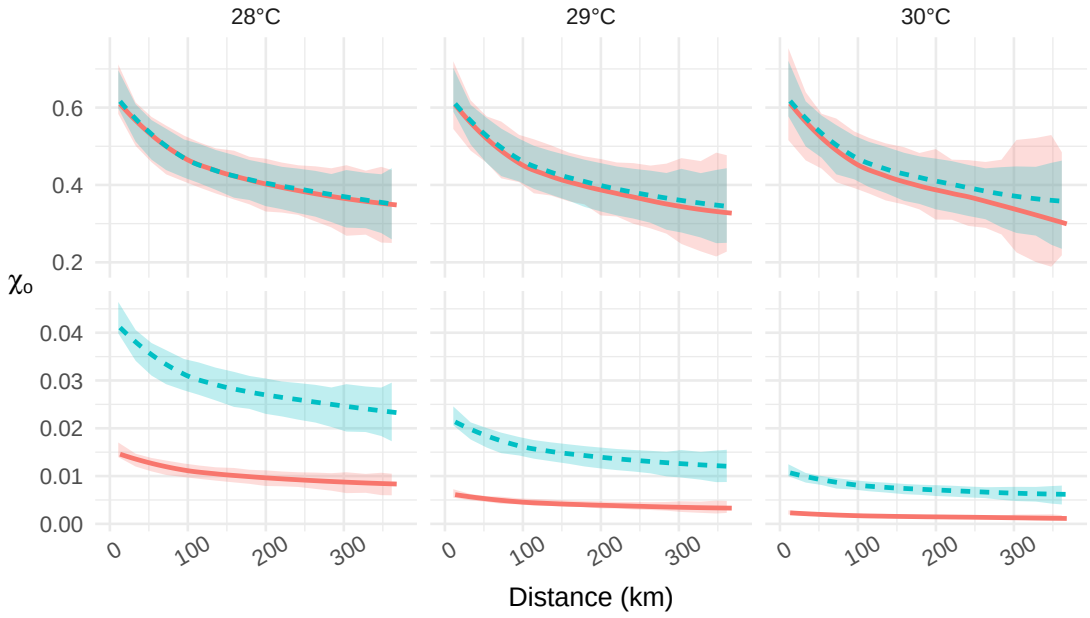

Figure 4.7: Estimates of $\chi_o(h; A_{t,\mathcal{S}}(T))$ (top row) and $\chi_o(h; T, t)$ (bottom row) against h (in km) for $T = 28, 29$, and 30°C for 1942 (solid, orange line) and for 2020 (dashed, blue line) for model M_2. Confidence intervals are based on 10,000 simulations for each 300 bootstrap sample data sets.

Finally, we look at a spatial risk measure based on the proportion, C, of a spatial field over the network that exceeds T°C at time t. Specifically, we consider the expected value of C, denoted $E_o(C; t, T)$. We also consider the conditional expected value of C, given by $E_o\left(C \mid A_{t,\mathcal{S}}(T)\right)$, i.e., given that we have observed a temperature somewhere on the station network. This conditional expectation is closely related to a functional used in characterising heatwave events (Cebrián et al. 2022). Specifically, these functionals and their relationships are given as follows:

$$\begin{aligned} E_o(C; t, T) &= \mathrm{E}\left(\frac{1}{|\mathcal{S}|}\int_{\mathcal{S}} I\left(X_o(t, \boldsymbol{s}) > T\right) d\boldsymbol{s}\right) \\ &= \mathrm{E}\left(\frac{1}{|\mathcal{S}|}\int_{\mathcal{S}} I\left(X_o(t, \boldsymbol{s}) > T\right) d\boldsymbol{s} \;\middle|\; A_{t,\mathcal{S}}(T)\right) \times \Pr\left(A_{t,\mathcal{S}}(T)\right) \\ &= E_o\left(C \mid A_{t,\mathcal{S}}(T)\right) \times \Pr\left(A_{t,\mathcal{S}}(T)\right), \end{aligned}$$

where $I(B)$ is the indicator function of event B.

Figure 4.8 shows that estimates of both of these measures for the station network over Ireland have increased from 1942 to 2020. The changes are highly significant, a factor of 90 larger for $T = 34°C$ when considering the unconditional expectation $E_o(C; t, T)$. However, when conditioning on the event $A_{t,\mathcal{S}}(T)$ this expected coverage proportion exhibits more limited changes with the largest difference being a doubling of the expected area affected when $T = 34°C$. In this latter case, the estimated change is small by comparison with its associated uncertainties.

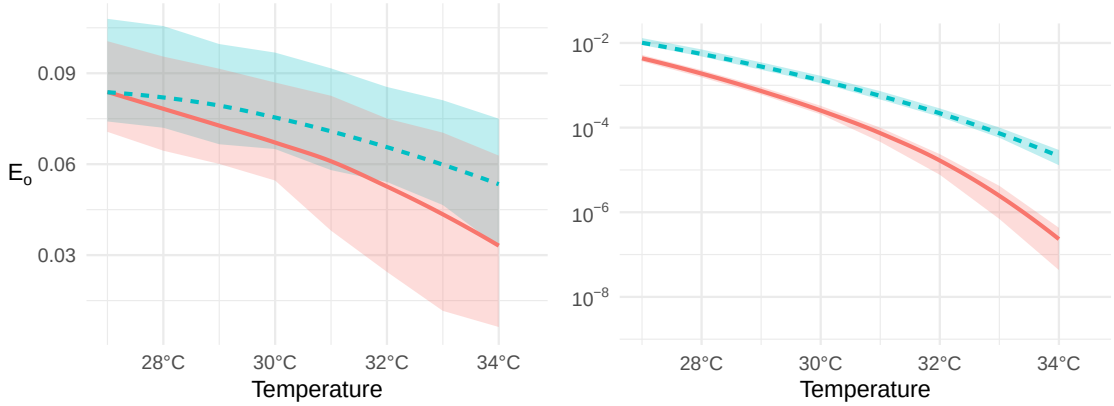

Figure 4.8: Expected proportion, $E_o\left(C \mid A_{t,\mathcal{S}}(T)\right)$, of Ireland that exceeds a temperature of T°C in an extreme event given that at least one site in Ireland (at the station network) exceeds T°C according to M_2 (left). The equivalent unconditioned estimates, i.e., estimates of $E_o(C; t, T)$ (right). Estimates are plotted against T for 1942 (solid, orange line) and for 2020 (dashed, blue line). The shaded regions give associated pointwise 95% confidence intervals, based on $10,000$ simulated fields for each 300 bootstrap sample data sets.

We finish by reporting on our investigation of the sensitivity of our risk measure analyses to our marginal modelling choices for temporal non-stationarity. The results for the models M_0 and M_1, i.e., the less well-fitting models, are given in Appendix B.6.2, with these being given for the same features shown for model M_2 in Section 4.5. Unsurprisingly, the inclusion of temporal non-stationarity in the tail model gives markedly different conclusions for all risk measures compared to those derived from the stationary model M_0. The inclusion of a coastal proximity covariate in model M_2 leads to larger scale parameter estimates inland and lower estimates in coastal regions than M_1. See Appendix B.3.3, Figure B.7. This is

reflected in the estimates of $\Pr(A_{t,\mathcal{S}}(T))$ and $E_o(C; t, T)$, with M_0 being lower than both M_1 and M_2, and in M_1 giving slightly higher estimates than M_2 on the station network. This change results in the probability of observing 33°C somewhere in $\mathcal{S}_o$ increasing by a factor of 1.5, 25, and 21 between 1942 and 2020 for models M_0, M_1, and M_2 respectively, showing that the difference in key conclusions is not too large between models M_1 and M_2.

4.6 Conclusions

We have presented some novel candidate approaches to merge information from spatially and temporally complete climate models into the spatial extreme value analysis of sparse and temporally incomplete observed temperatures from available meteorological stations. New methodological features include using outputs from an extreme value analysis of the climate model data to provide a covariate for the equivalent analysis of observational data, and dealing with r-Pareto processes in a missing data framework. We also presented novel metrics, combining both marginal and dependence features, to describe changes in spatial risk over time.

Our analysis was for daily maximum summer temperatures over Ireland. We found that the climate model data were more helpful for marginal modelling of the observational data than for dependence modelling, as they have the potential to overestimate extremal spatial dependence relative to the observational data. We pooled data from across stations to fit our model and found evidence that the Irish summer temperature anomalies were the best-fitting covariate, appearing to mostly affect marginal behaviour with minimal effect on spatial extremal dependence, see Appedix B.2 and Appendix B.7.2. We found that from 1942 to 2020 the occurrence rates of high threshold exceedances have increased by 35%, with 95% confidence interval 28-44%, and extreme quantiles have increased by 1.2 – 2.2°C, the latter ≈ 1°C greater than the change of mean summer temperature anomalies for Ireland and globally. Finally, we found that spatial heatwave events over thresholds that are critical for society have become much larger, having at least doubled in spatial extent for 28°C, with this change increasing at more extreme temperatures.

4.7 Discussion

4.7.1 Use of climate models

In the analysis of climate extremes, practitioners tend to make use of some combination of observations and climate models, but without necessarily recognising the inherent trade-offs between them and the respective limitations in the data sources. Perhaps more critically we believe further thought could be given to the synergies which might enable a better set of tools for practitioners. Downscaling from climate models to match observational properties is a major area in statistical climate science, but it risks introducing biases in spatial dependence from climate models given its focus on the marginal agreement. Here we have illustrated what appears to be an effective new approach aiming instead to enhance the observational analysis by exploiting the strong information about the physical properties of the climate system and the greater spatial coverage of information embedded in the climate model data.

As a proof of concept, we have restricted the information we extract from the climate models to that arising from one climate model output. There exists a broad ensemble of global and regional models that could be used. Each combination of such models encapsulates different modelling assumptions and therefore would provide a distinct estimate of the behaviour of maximum temperatures over Ireland. To fully quantify the uncertainty in our estimates would require an adequate sampling strategy to select global-regional model combinations from the available ensemble.

Our analysis looks only at the change in extreme temperature events over the historical record. We do not have observations of the future. Climate models are our only reliable tool for predicting future temporally non-stationary extremal behaviour under different scenarios, so being able to link the temporal non-stationarity of observational and climate data is essential to understanding model strengths and weaknesses. This is particularly important for the consideration of highly non-linear change linked to instabilities such as the possible effects of any change in the strength of the Atlantic meridional overturning circulation (AMOC) which has a profound modulating effect on Irish climate. Most global models suggest some weakening of the AMOC through to 2100, and a complete shutdown cannot be ruled out (IPCC 2021). Some work in this area has been carried out by Zhong,

Brunner, Opitz & Huser (2022), where the authors examine future time evolution of precipitation extremes using r-Pareto processes applied to future climate model simulations.

Finally, although we focus on temperature, many other variables (e.g., wind, rainfall) are important for assessing climate change for extreme meteorological events marginally, jointly, and integrated over different time windows. Climate model data are likely to provide improvements in extremal inference of such joint distributions under the assumption that they better capture the physical interactions between processes. This can be used to enhance the equivalent empirical information from the observational data.

4.7.2 Choice of threshold in non-stationary analysis

Threshold choice for the GPD and other tail models for identically distributed univariate extremes has been a major area of research for much of the last 40 years. Therefore, it is not surprising that there are a number of different perspectives for picking a systematic threshold selection criteria in our temporally non-stationary spatial context.

For univariate temporally non-stationary problems Eastoe & Tawn (2009) propose pre-processing the data using models fitted to the body of the distribution before modelling the extremes of the residuals with a constant threshold. Another approach is to use a conditional quantile (Northrop & Jonathan 2011). As noted in Section 4.3.3 it is difficult to account for the uncertainty in the threshold selection, so incorporating a temporal trend into the threshold undermines our ability to account for the uncertainty in estimating the temporal change in extreme data, which is our primary focus. We are pleased to see that even with our constant threshold at each station we found a simple model for how the GPD scale parameter changes over time and that the GPD is a good fit globally. More generally, the signal-to-noise ratio is critical in determining whether non-stationarity is accounted for in selecting the set of "extreme data" to analyse.

We had an additional threshold to select for the extreme spatial dependence modelling via the risk function r. We had to face issues of missing data, with our approach presenting the first methods, we are aware of, for this. Climate mod-

els may help here either through exploring the sensitivity of different missing data patterns or through the use of reanalyses (weather forecast models conditioned on the observed values) to replace missing data, as these will help identify the largest events over space correctly.

4.7.3 Choice of spatial extremes dependence model

The scale of Ireland relative to the physical systems that drive temperature extremes has also played a key role in our choice of extreme value approach for modelling spatial dependence. This enabled us to take a simple model which is asymptotically dependent at the largest required spatial separation, which we achieved via an r-Pareto process coupled to a log-Gaussian latent process with a bounded variogram. We do not believe our approach would be applicable at much broader scales.

Even over the scale of Ireland, the asymptotic dependence property will not necessarily hold for other climatic variables, e.g., precipitation, which are manifest on smaller scales and with higher variability. In such cases, the modelling approaches need to incorporate asymptotic independence and to address issues about the scale over which asymptotic dependence holds (Wadsworth & Tawn 2022, Zhong, Huser & Opitz 2022) or even whether the spatial process is stationary over different mixture type events, e.g., convective or frontal precipitation (Richards et al. 2023). Any application of this method to different regions or processes should certainly involve an assessment of the evidence for asymptotic dependence, as our impression is that this assumption is made too readily. We find that, as a good approximation, we can assume spatial stationarity and isotropy. Ireland has, relatively speaking, simple topography with few ranges of hills of significant altitude. It does not follow that the method would be readily applicable to more complex alpine regions without, at a minimum, considerable additional validation.

4.7.4 Choice of metrics for assessing change

Section 4.5 illustrates the challenge of finding effective metrics to illustrate temporal change when considering extremes of spatial fields. Metrics for marginal variables, e.g., high quantiles, are well-established and parsimonious. We see the development of spatial risk measures which enable the simple assessment of changing risk over time as an important avenue for further research. Spatial extreme value model

inference also lacks well-established diagnostic methods for assessing the fit with observed events. Pairwise, and potentially higher order versions of the measure χ, used in Section 4.4.2 and Section 4.5, can be helpful but these may not be sufficient in practice. Winter et al. (2016) used severity-area frequency curves as a basis of comparison, but these focused only on assessing the performance of the model for the dependence structure. Picking metrics that directly link to risk assessment from heatwaves, such as health factors (Winter & Tawn 2016) or crop failures or forest fires (Zhang et al. 2022), is likely to be valuable for planners.

We focused on the spatial properties of the extreme events. Models for spatio-temporal extremal dependence of the process are needed to capture the evolution over time of spatial extreme events. This is an area where greater focus is required. For processes that are asymptotically dependent in space and time, some methods have been developed (Davis et al. 2013, Huser & Davison 2014), and it is pleasing to see recent extensions to incorporate asymptotic independence (Simpson & Wadsworth 2021).

B

Supplementary material to Chapter 4

B.1 Introduction

In this supplementary material, we provide extra details that support our main analysis in Chapter 4. We considered three marginal models for the data, named M_0, M_1, and M_2, each of which has a structure as described in Table 4.2 in Chapter 4. The latter model (M_2) was selected as the superior model and forms the basis for our spatial analysis. However, we carry M_0 and M_1 throughout the entire analysis procedure to explore the sensitivity of our model choices. In each case, we provide additional comparisons to the model M_2 for completeness. We begin in Section B.2 with details on the covariates used in our analysis, we detail the use of different temporal covariates in the marginal models and give justification for our choice of using only temperatures from summer months for our primary analysis. In Section B.3, we present supplementary results from our modelling of the body of the distribution. We explore the effect of our model choice on the GPD parameters in our marginal models, as well as estimate uncertainties through our bias-corrected bootstrap resampling, along with a simulation study to justify this method. We also present diagnostic plots of marginal model fits. In Section B.4, we provide additional information about the spatial results derived from models M_0 and M_1. In Section B.5, we describe, in algorithmic terms, our method to simulate extreme spatial events. In Section B.6, we present alternative methods to calculate the probability of observing extreme events from a set of simulations from the r-Pareto process along with summaries of simulations of extreme events based on M_0

and M_1. Finally, in Section B.7 we discuss possible modelling choices where the assumption of asymptotic dependence is not appropriate and we also give a possible extension to our distance metric.

B.2 Additional detail on data

B.2.1 Covariates

As discussed in Section 4.2.3, we use smoothed temperature anomalies for the global average $M^G(t)$, and for the grid box over Ireland $M^I(t)$ (see right panel of Figure B.1) to model temporal non-stationarity. These covariates are derived from the observation-based data product HadCRUT5 (Morice et al. 2021). These data are a set of gridded temperature anomalies, aiming to represent observed data, produced as a combination of a land temperature anomaly data set (CRUTEM5; Osborn et al. 2021) merged with a sea-surface temperature anomaly data set (HadSST4; Kennedy et al. 2019). The merging of these data sets is achieved via weighting as a function of land area fraction (Morice et al. 2021).

The HadCRUT5 data set is provided at a monthly resolution on a global 5^2 degree grid. It is produced by taking 200 samples from the observed anomalies (with added modelled noise) and subsequently taking the average at each time point. An attractive feature of this procedure is that it accounts for, e.g., the state of the North Atlantic Oscillation (NAO), a major influence on European climate. We perform LOESS temporal smoothing on the summer monthly average temperature anomalies for both the global average $M^G(t)$, and for the grid box over Ireland $M^I(t)$, to decrease bias (Clarke & Richardson 2021) and use these as covariates. Using the `R` language implementation, LOESS smoothing was performed with the default settings of a span of 0.75 and a degree of 2 (R Core Team 2023). Figure B.1 (left) shows the LOESS-smoothed $M^G(t)$ and $M^I(t)$, with respective 95% confidence intervals, while Figure B.1 (right) shows the HadCRUT5 grid over Europe. Over the period of our observations, both Ireland and global mean temperatures are increasing by approximately 1°C.

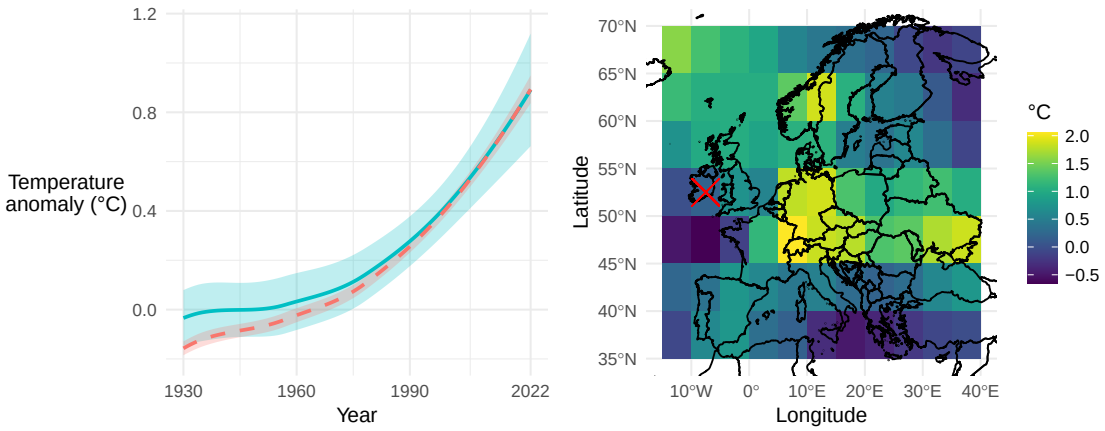

Figure B.1: LOESS smoothed Irish (global) summer temperature anomalies shown by solid-blue (dashed-orange) lines respectively, with shaded regions showing their 95% confidence intervals (left); an event from the HadCRUT5 grid over Europe in July 1991 (right). The grid box for Ireland is marked by a red ×.

B.2.2 Alternative covariates

We explore the use of several alternative temporal covariates to those used in Chapter 4. The alternative covariates included time, global temperature anomalies, and both local and global CO_2 emissions. We consider the covariate of the annual CO_2 emissions ($CO_2(t)$) for Ireland (Ritchie et al. 2020) since carbon dioxide is well mixed and national emissions continue to be a small overall fraction of global emissions, the national emissions can be seen as a proxy for local climatic effects of atmospheric particulates co-emitted with fossil fuel combustion and which crucially have lifetimes of days to weeks. Local CO_2 emissions are plotted in Figure B.2.

In an earlier version of the study, the best-performing model used time, local CO_2 emissions, and global mean temperature anomalies as covariates. Reviewers of this manuscript rightly highlighted issues with collinearity between the temporal covariates included. Model structures with collinear covariates are certainly not ideal both for inference and interpretation reasons. To address this, we performed an eigen-decomposition of these covariates. The newer model addressed the issue of collinearity while preserving model performance, however it greatly restricted our interpretation of the model. Finally, as presented in Chapter 4, we decided to use the HadCRUT5 mean temperature anomalies over Ireland as a temporal

covariate in lieu of others discussed here. The parsimony, performance and ease of interpretation of this covariate in the model made it an effective formulation.

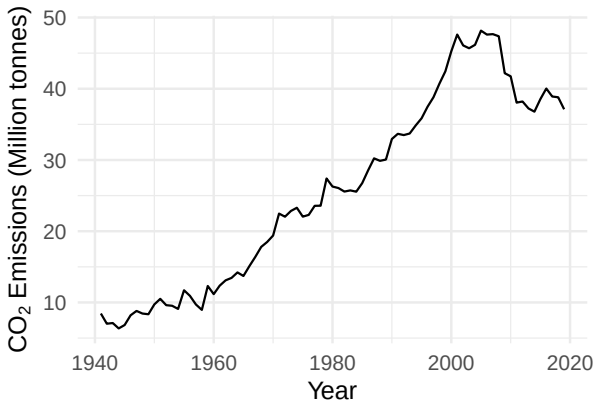

Figure B.2: CO_2 emissions in Ireland from 1940 to 2022, explored as an alternative temporal covariate in our model.

B.2.3 Seasonality

We justify our restriction of using only the summer months (June – August inclusive) in our selected data set. We start by investigating the period of the year in which the hottest temperatures arise. Figure B.3 shows that taking thresholds at the 95th and 99th quantiles across the year separately at each site, 85% and 93% respectively of threshold exceedances occur in these summer months. Furthermore, the domination of these months in containing extreme temperatures continues to increase with higher threshold magnitudes. The high temperatures are reasonably spread across the three summer months, with most extremes in July. Thus, considering only these months in the analysis largely avoids the need for modelling seasonality within a year, although there may be some benefit in modelling within-summer non-stationarity. We leave this as a task for future research.

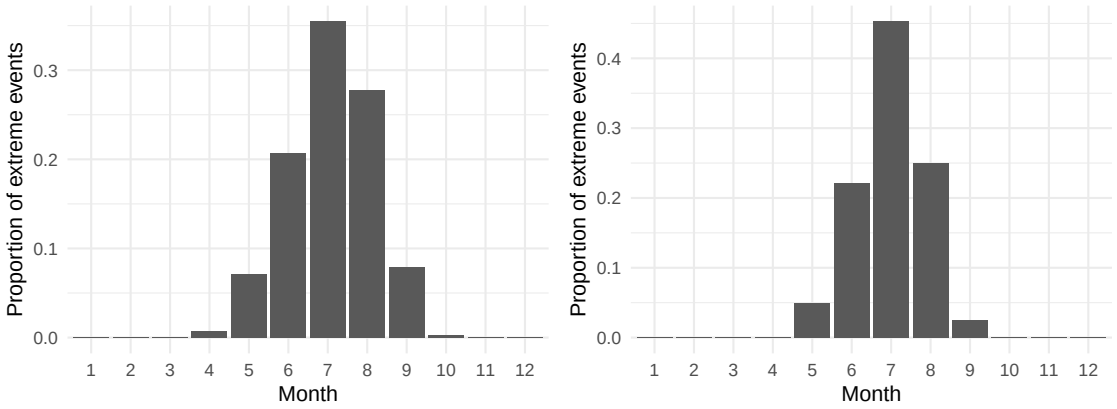

Figure B.3: Proportion of threshold exceedances per month where the threshold corresponds to 95th (left) and 99th (right) site-wise quantiles. Months labelled 1–12 correspond to January–December, respectively.

B.3 Marginal models

B.3.1 Body of distribution model

In this section, we present supplementary figures for the body of the distribution model fit. Figure B.4 (left) shows the estimated $\beta_2^{(\tau)}$ values against τ for our chosen model, M_2. Figure B.4 (right) shows the spatial average threshold exceedance probability (defined in Section 4.3.5.2) for model M_2.

B.3.2 Justification for constant shape parameter over space

Here we expand on our decision to keep the shape parameter, ξ, constant over space in the GPD model while allowing the scale parameter to vary over space, for both the climate model output and the observed data. Take the null hypothesis, M_N, as the simpler model, with the shape parameter held constant over space. The alternative model, M_A, then allows for a different shape parameter for the available sites, $\boldsymbol{s}$, for the respective data sets. We perform a log-likelihood ratio test at each site to determine whether allowing the shape parameter to vary over space provides a significant improvement in the model fit.

To estimate constant shape parameters ξ_o and ξ_c over Ireland (for the purpose of the log-likelihood ratio test), we use the same approach set out in Chapter 4

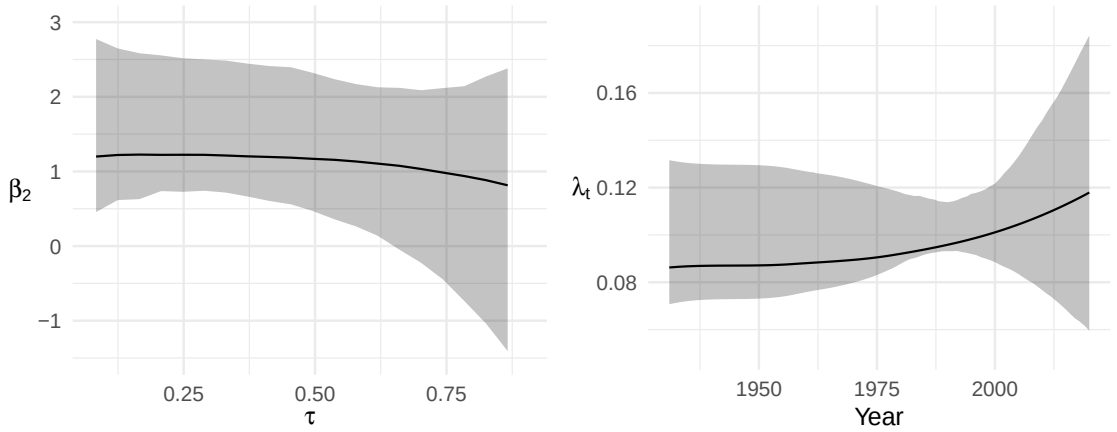

Figure B.4: Estimates of coefficient $\beta_2^{(\tau)}$ of $M^I(t)$ over a range of τ (left) and the threshold exceedance probability $\lambda_o(t)$ for t from 1942 to 2020 (right). In each case, the shaded region indicates bootstrap-based pointwise 95% confidence intervals.

for estimating ξ_c, i.e., using the pseudo-log-likelihood (4.5) with a separate scale parameter at each site, which is constant over time, but here for the observational and climate model data respectively. Under the null model, M_N, the estimated parameter values[a] were $\hat{\xi}_o = -0.174$ and $\hat{\xi}_c = -0.189$. For the alternative model, M_A, we fit a GPD distribution at each site independently, estimating both the shape and scale parameters specific to that site with the scale parameter constant over time.

For each site $\boldsymbol{s}$, for both the climate model output and observational data, we find the maximised pseudo-likelihood of both null and alternative models giving $p\mathcal{L}\{M_N(\mathbf{s})\}$ and $p\mathcal{L}\{M_A(\mathbf{s})\}$ respectively. At each site, the pseudo-likelihood ratio test statistic is $2\ln[p\mathcal{L}\{M_A(\mathbf{s})\}/p\mathcal{L}\{M_N(\mathbf{s})\}]$, which is plotted in Figure B.5 for both data sets. All the values of the test statistics are very close to zero[b] for all sites. Thus, there is insufficient evidence to reject the null hypothesis, so we subsequently fix the respective shape parameter estimators of ξ_o and ξ_c to be constant over space.

[a]This estimate for ξ_o is not the final estimate used in Chapter 4, as that comes from a model with spatially and temporally varying scale parameters.

[b]If this was a standard likelihood test, the 95% critical value would be 3.84, and all values are an order of magnitude smaller than that.

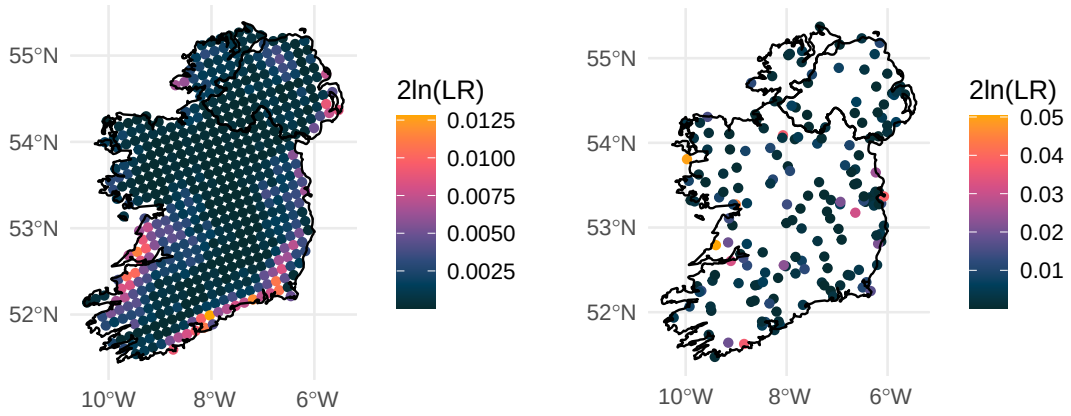

Figure B.5: Log-Likelihood ratio test statistic for constant shape parameter in the GPD model on climate output (left) and observational data (right) over Ireland.

B.3.3 Scale parameter estimates

Here we detail the parameter estimates of the marginal models, M_0 and M_1, defined in Table 4.2. Since the shape parameter is taken to be constant over space and time, all marginal non-stationarity is captured by the scale parameter, σ, in the generalised Pareto distribution part of the marginal models. In Figure B.6 we compare the scale parameter estimated using models M_0 and M_1. The marginal models M_0 and M_1 have identical spatial structures since they each use the same spatial covariate derived from climate model outputs (σ_c). The first two subplots in Figure B.6 show the estimate of the scale parameter according to model M_0 and model M_1 (in 2020) respectively. Model M_1 estimates a lower scale parameter than M_0. Model M_0 estimates the scale parameter as constant in time (which gives a poorer description of the data than a temporally varying model). This difference could possibly be the cause of the increase in the overall variance in the model fit, inflating the point estimate of the scale parameter overall. It is also important to note that these models have different shape parameters (ξ) associated with them and these parameters tend to be negatively correlated with the associated scale parameter estimates in practice. The shape parameter of models M_0 and M_1 are -0.152 and -0.156 respectively, so this difference will have a limited effect on the scale parameters.

The third plot in Figure B.6 shows that the estimate for the scale parameter for model M_1 increases by around 0.4 in most regions of the island from 1942 to 2020. This increase is largely comparable in magnitude to that estimated using model M_2 (see Figure 4.2), although the inclusion of the coastal proximity covariate in model M_2 results in a different spatial structure from both models M_0 and M_1. The difference in these estimated scale parameters is shown in Figure B.7 here. In both comparisons, model M_2 estimates a higher scale parameter inland and decreasing estimates towards coastal regions with lower estimates than model M_1 on the coast.

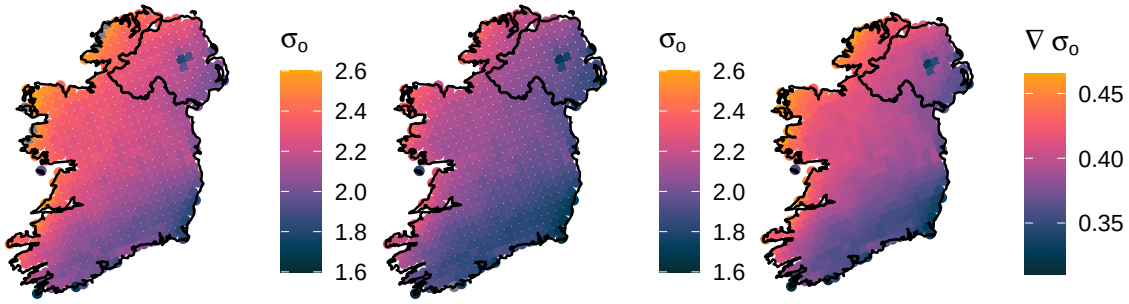

Figure B.6: Estimated values of scale parameter according to M_0 (left), scale parameter according to M_1 in 2020 (middle), and change in scale parameter in the period 1942–2020 according to M_1 (right).

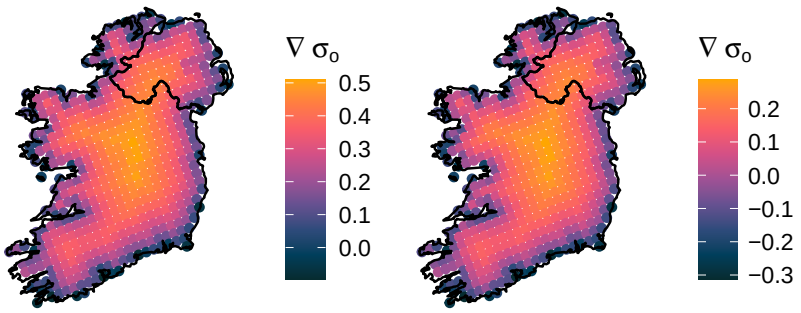

Figure B.7: Estimates of scale parameter according to model M_0 subtracted from estimates according to model M_2 in 2020 (Left). Estimates of scale parameter according to model M_1 subtracted from estimates according to model M_2 in 2020 (right). A positive value indicates a higher scale parameter estimated in M_2.

B.3.4 Additional details of the bootstrap methods

As stated in Section 4.3.4.1, we apply a standard vector temporal block bootstrap to $x_{\mathrm{o}}^{U}(t, \boldsymbol{s})$, the observed data transformed to uniform margins, with the temporal and spatial dependence structure of the process $(X_o(t, \boldsymbol{s}))$ retained. Here, we describe the bootstrap procedure. We apply a standard vector temporal block bootstrap to data (Politis & Romano 1994). Here temporal blocks of the spatial vectors are sampled throughout the observation period, with block lengths randomly generated from a geometric distribution. This fully retains the empirical spatial dependence whilst recognising that temporally the data exhibit short-term dependence only. We selected the geometric distribution mean to be 5 as the empirically estimated probability of an extreme temperature at a site on day $t + 5$ given that a site experienced an extreme temperature on day t is 0.2 (where extreme here refers to exceeding the 0.9 marginal quantile), so the majority of extreme events finish within 5 days. Finally, by retaining the covariates $\boldsymbol{z}_t$ in this selected block we also retain the links to the covariates and their joint distribution.

If there were no missing data, we can simply repeat this block selection at random until the size of the bootstrapped data set matches the actual data. However, for the station data, the bootstrap sample needs to exhibit the same missingness pattern as the actual data. Hence, when sampling to find a bootstrapped block of data of a specified length, to correspond to the station data missingness pattern at that time, we restrict the search to include only those blocks of the required length with patterns of recorded values which match or exceed the actual data. Thus, if the station series has any pair $(t, \boldsymbol{s})$ missing, and the associated bootstrap block has that pair as recorded, then these data values are discarded. We found this to work well, although as there are less data in the early years containing the observation pattern of recent years we obtain bootstrapped samples with greater across-sample variation in the early years than the later years. We then transform the bootstrapped data set $X^{U*} = \{X_{\mathrm{o}}^{U*}(t, \boldsymbol{s}) : t \in \mathcal{T}_o, \boldsymbol{s} \in \mathcal{S}_o\}$ to the data scale using the inverse of the distribution function $F_{o,t,\boldsymbol{s}}$ to create the bootstrapped sample X_o^* as shown in expression (4.6).

B.3.5 Details of bias correction of marginal model parameter estimates

To calculate uncertainties in the parameters of any given marginal model we repeatedly fit the same model to 300 spatio-temporal bootstrap samples. An issue encountered with bootstrapping extreme values is the occurrence of ties in the most extreme observations in the bootstrapped data set. In extreme value theory, the very largest observations in a data set are known to be the most influential on the GPD model fit (Davison & Smith 1990), this issue is exacerbated with bootstrapping since cannot sample values larger than those observed and risk sampling the largest few observations several times in a bootstrap sample. When this repeated sampling occurs it can lead to bias in the estimate of the shape parameter of the GPD causing it to be smaller than what would be expected. The bias arises because the ties in the largest values indicate it is very unlikely to obtain much larger values than that already observed. Since the shape and scale parameters are negatively correlated, this bias in the shape parameter estimator is inherited by the scale parameter estimator, with a positive bias. To solve the problem we use a bootstrap error correction factor which we have found works well in simulations. Specifically, we apply Algorithm 1 to each bootstrap, X_o^*, for each GPD model, M_k, where $k = \{0, 1, 2\}$. This procedure involves fitting the GPD model M_k to all bootstrap data sets X_o^*, finding the average shape parameter value from these fits, and comparing this with the actual point estimate from the model fitted to the observed data. The difference between the mean of the bootstrapped shape parameter estimates and the actual estimate acts as our bias correction term. To correct for the bias in any bootstrap fit, we take the estimated shape parameter, add the bias correction term, Δ_ξ, and fix this parameter while re-estimating the scale parameter for that bootstrap sample. We repeat this for all bootstraps to get a corrected uncertainty estimate (for the corrected bootstrap sampling distribution) for the GPD parameters.

Algorithm 1 Correcting bias in GPD fits to bootstrap samples for model M_k

Let b be the number of bootstrap samples.

Let $\widehat{\xi}$ be the estimate of the shape parameter from fitting M_k to all observed data.

for all bootstrap samples $X_o^{*(i)}$, where $i = \{1, 2, \ldots, b\}$

Estimate $\xi^{*(i)}$ by fitting M_k to $X_o^{*(i)}$.

end for

Calculate bias correction term for shape parameter, $\Delta_\xi = (\sum_i^b \xi^{*(i)})/b - \widehat{\xi}$.

for all bootstrap samples $X_o^{*(i)}$, where $i = \{1, 2, \ldots, b\}$

Re-fit M_k to X_o^* while fixing shape parameter to be $\xi^{*(i)} + \Delta_\xi$.

end for

To illustrate the performance of Algorithm 1 in correcting for bootstrap bias, Figure B.8, Figure B.9, and Figure B.10 show plots of the sampling density of each parameter of models M_0 and M_1 fitted to 200 bootstrap samples and M_2 fitted to 300 bootstrap samples. The top row of each plot corresponds to the estimates with shape parameters $\widehat{\xi}$ using the raw bootstrap samples. The sampling densities are estimated using the default settings of the `density()` function in `R` (R Core Team 2023). This row of plots (in each figure) shows a clear negative mean bias in the sampling density of ξ relative to $\widehat{\xi}$, and the resultant clear positive bias in the scale parameter through the sampling density for β_0. The bottom rows show the parameter estimates after applying Algorithm 1. In each model, we can see the bias-corrected bootstrap has corrected for bias in the shape parameter (ξ) by construction. This correction has reduced bias in the intercept term, β_0 of the scale parameter, and slightly reduced bias in the bootstrapped uncertainty estimate of other terms.

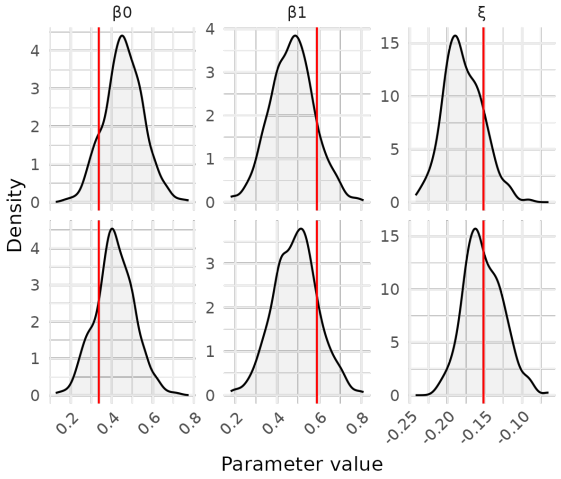

Figure B.8: Sampling density of raw (top) and bias-corrected (bottom) bootstrapped parameter estimates of model M_0 with a value estimated from the original data shown as a vertical line.

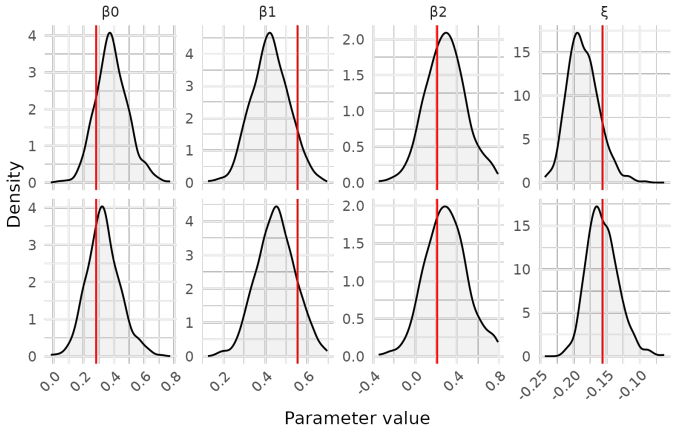

Figure B.9: Sampling density of raw (top) and bias-corrected (bottom) bootstrapped parameter estimates of model M_1 with a value estimated from the original data shown as a vertical line.

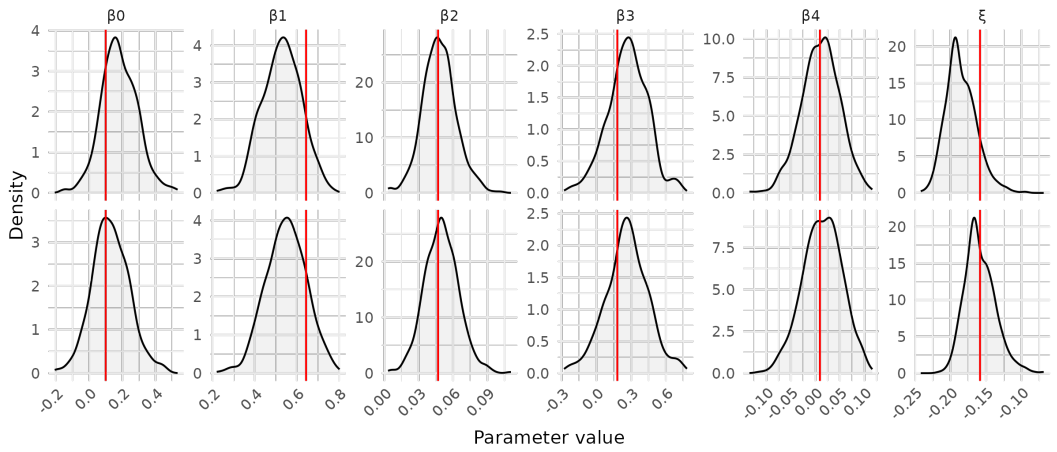

Figure B.10: Sampling density of raw (top) and bias-corrected (bottom) bootstrapped parameter estimates of model M_2 with a value estimated from the original data shown as a vertical line.

B.3.6 Simulation study to assess the performance of the bias correction method

In Section B.3.5, we showed that the bias correction in the bootstrapping better centres the resulting sampling distribution on the point estimates and in Section 3.4.1 of the main paper we suggested that the bias arose due to the ties in the data. In this section, we perform simple investigations of both these aspects. Specifically, in Section B.3.6.1 we show that rounding GPD data to induce ties does in fact produce a bias similar to that found for our data and we show that correcting for the bias in the shape parameter, similarly to the bias-correction method of Section B.3.5, helps account for, and reduce, the associated bias in the scale parameter. Furthermore, in Section B.3.6.2 we illustrate that for the key parameters, the bias correction process in the bootstrap produces confidence intervals with notably better coverage (i.e., closer to the nominal coverage) than the raw bootstrap method.

B.3.6.1 The impact of rounded data on bootstrap sampling distribution

Rather than using the full generality of the spatial likelihood, here we illustrate the effect of ties within the data on the bootstrap procedures of Section B.3.5 simply through a data analysis of univariate GPD simulated data. Specifically,

we have GPD scale parameter $\sigma = 1.5$ and shape parameter $\xi = -0.15$, with these parameters of a similar order of magnitude to our threshold exceedance temperature data. We generate a random sample $\boldsymbol{x}$ of size 500 from this distribution. We then round the data in $\boldsymbol{x}$ to give data $\boldsymbol{x}^r$, inducing ties in the data, including potentially also for the largest observed values. We choose one decimal place for the rounding as this is the resolution of our temperature data in the main paper.

We study the sampling distributions of the maximum likelihood estimators of the GPD scale and shape parameters using 1000 independent replicated data sets of $\boldsymbol{x}^r$, as generated above. In these fits, we do not account for ties within the data. First, we fit the GPD for each of these 1000 data sets, giving 1000 values, denoted $(\hat{\sigma}_i^{raw}, \hat{\xi}_i^{raw})$ for $i = 1, \ldots, 1000$, from the sampling distribution of the scale and shape parameters. We term these the raw maximum likelihood estimates and use these replicates to derive an estimate of the associated sampling distributions for $\hat{\sigma}^{raw}$ and $\hat{\xi}^{raw}$.

We then attempt to capture the key feature of the bias-correction bootstrap methods of Section B.3.5, to improve these raw sampling distributions. Specifically, as here we know the true shape parameter ξ_{true} for each sample, we correct the ith raw maximum likelihood estimate $\hat{\xi}_i^{raw}$ to a bias-corrected value $\hat{\xi}_i^{corr}$ via

$$\hat{\xi}_i^{corr} = \hat{\xi}_i^{raw} + \left(\frac{1}{1000} \sum_{j=1}^{1000} \hat{\xi}_j^{raw} - \xi_{true} \right) \text{ for } i = 1, \ldots, 1000.$$

Thus, the sampling distribution of the $\{\hat{\xi}_i^{corr}\}$ values has mean at ξ_{true}, so $\hat{\xi}^{corr}$ is unbiased. As with the methods in Section B.3.5, for each of the data samples, we then fix the shape parameter value at the corrected values, i.e., for the ith sample at $\hat{\xi}_i^{corr}$, and estimate the GPD scale parameter using the associated maximum likelihood estimator for this constrained likelihood. We denote the resulting estimates by $\{\hat{\sigma}_i^{corr}\}$.

Figure B.11 shows the estimated sampling distributions in terms of density estimates. The raw estimates (top row) produce densities with the modes not aligned to the true values for either parameter. We conclude that the presence of ties in the data risks artificially re-enforcing the upper-end point of the data, creating a negative bias in the shape parameter of the GPD. Since the shape and scale parameters

of the GPD are negatively correlated, we see this bias reflected in a positive bias in the scale parameter. In contrast, the bias-corrected sampling densities method (bottom row) are centred on the truth, with the bias correction for the shape parameter in turn reducing the bias in the scale parameter estimate.

In summary, we believe this small study shows that the rounding/ties in the observed data is a likely cause for bias in the marginal data analysis in the main paper. Furthermore, it shows that our bias reduction approaches, used in the corrected bootstrap in Section B.3.5, appears to suitably address this bias and produce viable unbiased estimates.

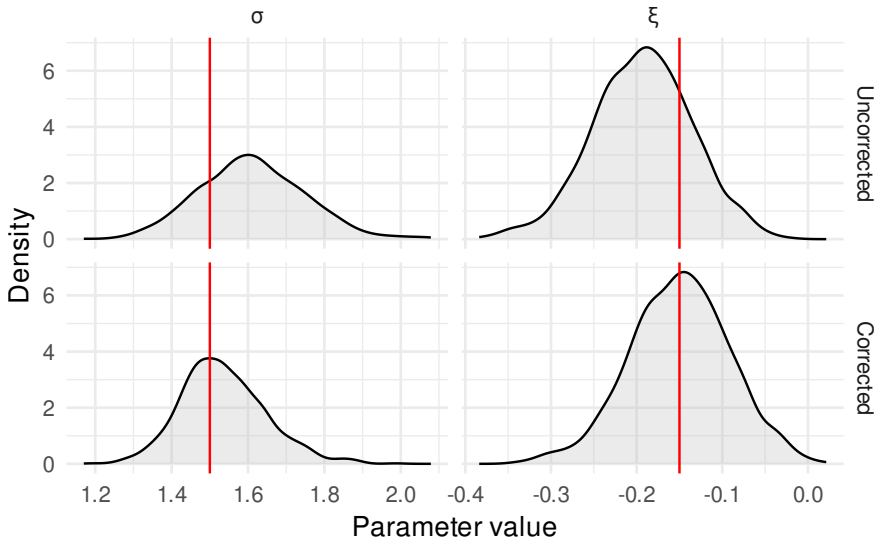

Figure B.11: Inference for rounded GPD data: sampling density of raw (top) and bias-corrected (bottom) bootstrapped GPD parameter estimates with known values shown as vertical lines.

B.3.6.2 Assessment of coverage for bootstrap confidence intervals

To illustrate that the proposed bias-corrected bootstrap method of Section B.3.5 gives improved confidence intervals, we show that the coverage properties of these intervals are both close to their nominal values and much better than the raw bootstrap-based methods. Specifically, we generate a number of replicate data sets, where the truth is known, with the same spatial-temporal varying marginal

distributional structure as in the fitted model M_2 and the same extremal dependence structure as for the fitted r-Pareto process as in the main paper. We also imposed that the simulated data are recorded at an identical network of 182 stations as for the observational data, and that they have the same amount of data over time, i.e., $8,464$ days, with an identical missing pattern as for the observed station data. Details of this simulation model are given in Section B.3.6.3.

We generated 200 replicated data sets from this joint spatial-temporal model. We apply both the raw and the bias-corrected bootstrap methods for each simulated data set, each giving nominal 95% confidence intervals for the GPD parameters and associated quantiles (return levels) for each site. From these estimated intervals the coverage probabilities, produced by the two proposed bootstrap methods, are evaluated as the proportion of the replicated intervals which contain the truth.

For nominal 95% confidence intervals, the coverage percentage for the shape parameter, ξ, increased substantially from a coverage of 45% to 81% from the raw bootstrap to the bias-corrected bootstrap, whereas for the intercept term of the log-scale parameter, i.e., β_0, it is decreased slightly from 68% to 65%. Given the strong trade-off between the estimates of the GPD parameters, we also assess the coverage for return level confidence intervals across return periods, sites, and time. Here we find much-improved coverage with the use of the bias-corrected bootstrap in all cases.

In Table B.1, we present these coverage probabilities for 10-, 100-, and $1,000$-year return levels at points in space and time corresponding to the mean spatial location in Ireland, as determined by the spatial covariates in model M_2, and for the median and both the lower and upper quartiles of the temporal covariate, denoted by $M^I_{(0.5)}$, $M^I_{(0.25)}$, and $M^I_{(0.75)}$ respectively. Here we see a very similar improvement in coverage to that found for the shape parameter, which is not too unexpected given the importance of the shape parameter for high quantile estimation.

Table B.1: Coverage probabilities (reported as rounded percentages) from raw and bias-corrected bootstrapped nominal 95% confidence interval estimates of return levels (RL). Results are for the marginal model M_2 and are based on 200 simulated data sets.

	10-year RL		**100-year RL**		**1,000-year RL**	
Temporal covariate	Raw	Corrected	Raw	Corrected	Raw	Corrected
$M^I_{(0.25)}$	54	77	50	79	50	81
$M^I_{(0.5)}$	47	71	46	74	46	75
$M^I_{(0.75)}$	44	65	44	68	40	71

So, although the coverage of the bias-corrected bootstrap is still underestimating the nominal value, it is a substantial improvement over the raw bootstrap method which is the current standard approach in this type of context in spatial extreme value modelling.

B.3.6.3 Simulation model for confidence interval coverage investigation

For our simulation model, we generate spatially dependent data that is temporally independent over days. As we are only interested in realisations on the observational scale that are above the marginal threshold, corresponding to the $1 - \tau$ quantile, we can simplify the generation process without making any critical differences from the station data characteristics.

For a simulation for a given day, we firstly generate data from our fitted r-Pareto process, X^P, for a risk function, corresponding to the average realisation over sites, exceeding a chosen cost of $c = 0.01$. We select this value of c so that this reproduces values marginally in the observational data after transformation of the margins. We remove components of X^P corresponding to the observational missingness pattern for that day in the station data. The resulting marginal distributions of this process are only Pareto when conditional on exceeding the level $u^P_m = 182c$, with lower marginal values following a more complex distribution depending on the r-Pareto dependence structure.

Secondly, we transform, marginally, the simulated r-Pareto process to a set of de-

pendent uniform random variables $\{X_i^U : i = 1, \ldots, 182\}$, via a componentwise probability integral transform. Specifically, for the ith marginal of X^P we use the empirical rank transformation of the i component of the X^P data below u_m^P, whereas above u_m^P we use the Pareto tail model.

Thirdly, we transform the uniform variables to the required GPD tail model M_2 for that day for each site with observations on that day. In practice, we only need to transform values with $X_i^U > 1 - \tau$ as the likelihood we use for inference of model M_2 only uses threshold exceedances. Finally, we round the simulated GPD data to one decimal place to replicate the nature of the network station data.

B.3.7 Marginal model fit diagnostic plots

We now present a broader assessment of model fit than that given in Chapter 4. Figure 4.3 in Chapter 4 presented a QQ-plot, pooled over all sites, showing that the model fits very well over Ireland when assessed globally over all sites. Here we look in more detail at the fits for each site individually, which we illustrate for 10 selected stations: the five named sites in Figure 4.1, and five other sites selected to be away from the coast. At each site, we assess the fit slightly differently for the body and tail models. To give the modelled quantile temperature, we use a different formulation for below and above the threshold exceedance probability: the inverse of expressions (4.2) and (4.3) for below and above the threshold respectively, with the numbering of these expressions being for Chapter 4.

Figure B.12 shows the estimated quantiles for the model for the body of the distribution. As we have sufficient data to assess the fit on a year-by-year basis below the threshold, we present a QQ curve for each year of observed data. Generally, there is a strong similarity in the curves over years, indicating that in the body of the distribution, there is limited change over the 1942–2020 time window. The model appears to perform well at most sites, possibly least well in the very lower tail, but also rather poorly at two of these sites, namely Malin Head and Roches Point. These two sites are located on the coast, at the very north and south respectively (see Figure 4.1) and so it is not too surprising that they are less well modelled, as temperatures change markedly at the land-sea interface.

For assessing the model fit above the threshold, we do not have sufficient data to

follow the strategy used in Figure B.12 of looking at the fit for each year separately. Instead, we pool data over years at each site above the threshold. As the modelled distribution changes over time, we standardise the excess of threshold data at each site to be identically exponentially distributed using the probability integral transform, and we also show the pointwise tolerance bounds in Figure B.13 to give bounds on natural variation in such a plot. Similar findings are obtained in the tails as in the body of the distribution; again some caution is required for estimates for some coastal locations but otherwise, the model fits well. As noted above, the model fit is very strong globally, so there will be only a very few coastal stations with limited quality fits.

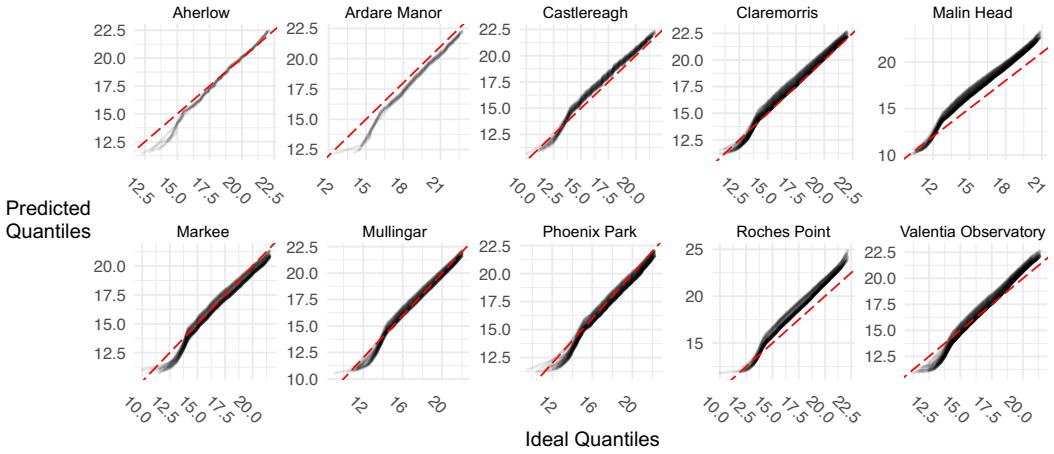

Figure B.12: QQ-plot of bulk model fits at a selection of sites. In each plot, lines join predictions within a year at their respective site. The lines of equality are in red.

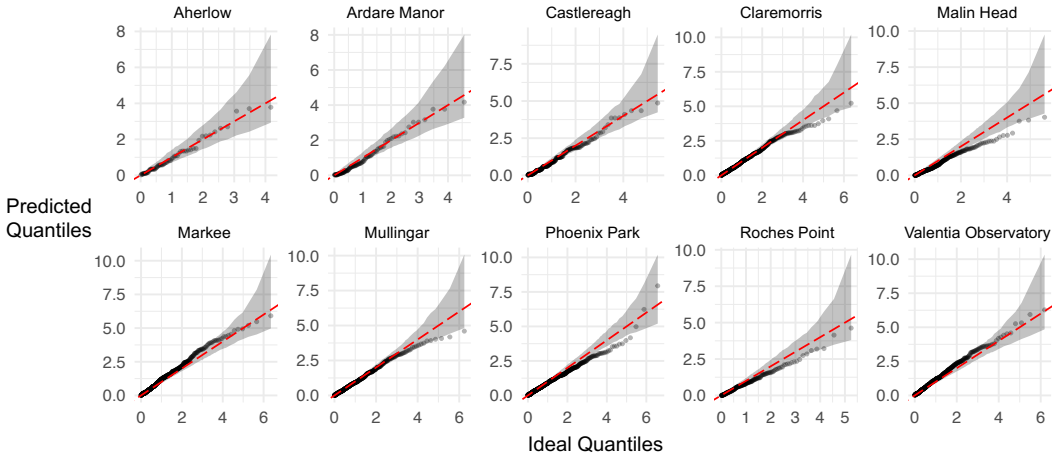

Figure B.13: QQ-plot of GPD model fits at a selection of sites on exponential margins. The shaded region shows pointwise 95% tolerance intervals. The lines of equality are in red.

B.3.8 Return level estimates

Figures B.14 and B.15 contain the return level estimates derived from models M_0 and M_1 respectively, with these providing similar plots to Figure 5 in Section 4.5 for model M_2. The left-hand panel in each figure corresponds to the 100-year return levels in 2020. The estimate according to model M_1 is slightly higher than that of model M_0. The estimate of the 100-year return level according to both M_0 and M_1 is comparable to that of model M_2 presented in Figure 4.5 with a magnitude of around 32°C over the island. The third plot in both figures shows the change of the 100-year return level from 1942 to 2020. The change in the return level estimated by model M_0 is very small since the only non-stationarity in this model comes from a time-varying threshold exceedance rate. The estimated change in model M_1 is quite similar to that estimated by model M_2 presented in Chapter 4 with respect to magnitude although quite dissimilar with respect to spatial structure. This is due to the inclusion of a coastal proximity covariate in M_2. Model M_2 suggests that the change in scale parameter from 1942–2020 is lowest in coastal regions and highest inland (see Figure 4.5).

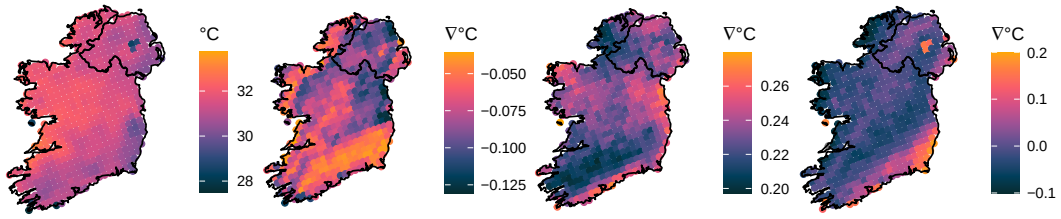

Figure B.14: Features of the 100-year return level obtained using model M_0: estimated 100-year return level for the year 2020 (1st plot); change in lower 95% CI for change in 100-year return level from 1942 to 2020 (2nd plot); estimated difference between the 100-year return level estimated for the years 2020 and 1942 (3rd plot); (4th plot) change in upper 95% CI for change in 100-year return level in the period 1942–2020 (4th plot).

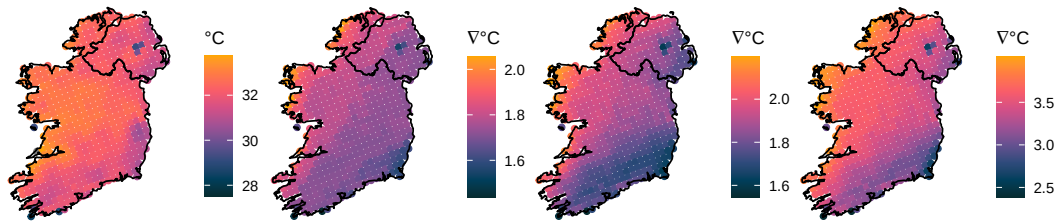

Figure B.15: Features of the 100-year return level obtained using model M_1: estimated 100-year return level for the year 2020 (1st plot); change in lower 95% CI for change in 100-year return level from 1942 to 2020 (2nd plot); estimated difference between the 100-year return level estimated for the years 2020 and 1942 (3rd plot); change in upper 95% CI for change in 100-year return level in the period 1942–2020 (4th plot).

B.4 Spatial models

B.4.1 Parameter estimates of spatial models

Here we present details on the fitting of the r-Pareto process to data standardised the marginal models. For computational efficiency, we opt for a faster and less memory-intensive algorithm on fewer bootstrap samples to illustrate uncertainties for the r-Pareto model fits for models not presented in Chapter 4. To estimate the uncertainty associated with r-Pareto fits for models M_0 and M_1, we use the limited-memory modification of the BFGS quasi-Newton method (Byrd et al. 1995) optimisation algorithm on 200 spatiotemporal bootstrap samples. This is unlike the 300 bootstrap samples modelled with the Nelder–Mead optimisation algorithm which we used in Chapter 4, and which we found to be more reliable (but with the cost of greater computational effort, approximately twice as slow). Figure B.16 shows the estimated χ_o^P with uncertainty for models M_0 and M_1. The equivalent for model M_2 can be seen in Figure B.17. No major discernible features can be distinguished between the models based on this visualisation, suggesting that the choice between M_0, M_1, and M_2 has a very limited effect on the estimated dependence structure, a feature identified more broadly in copula modelling by Genest et al. (1995).

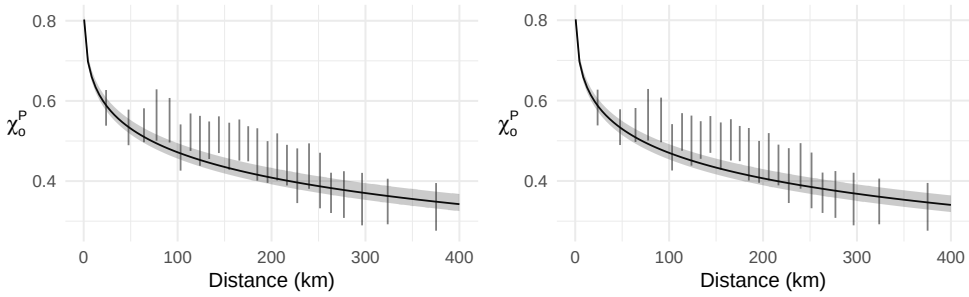

Figure B.16: Estimates of χ_o^P for model M_0 (left) and M_1 (right) against inter-site distance h (in km): The vertical line segments show the pairwise empirical estimator described expression (4.8), Section 4.4.2, with $p = 0.95$; the smooth lines display limiting values for χ_o^P estimated using the r-Pareto process with Matérn variogram when fitted above a risk threshold of 80%. Confidence intervals are estimated based on 200 bootstrapped r-Pareto fits using the limited-memory BFGS optimisation algorithm.

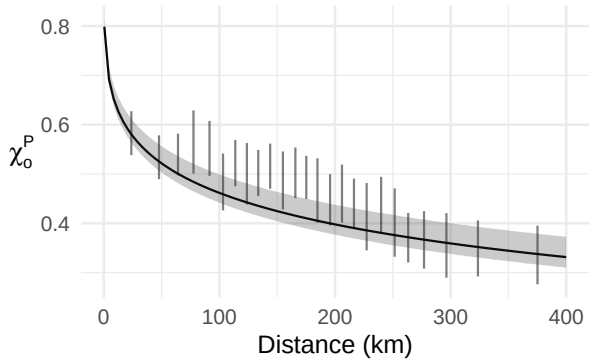

Figure B.17: Estimates of χ_o^P against inter-site distance h for model M_2: the pairwise empirical estimator with $p = 0.95$ shown as vertical line segments reflecting the bootstrap uncertainty in the marginal distribution estimates; the smooth lines display the limiting value for χ_o^P, estimated using the r-Pareto process with a Matérn variogram when fitted above an 80% risk threshold, with 95% confidence intervals, capturing both marginal and extremal dependence bootstrap uncertainties.

B.5 Scaling simulations

In Section 4.4.4, we lay out an importance sampling procedure to re-scale a set of simulated events to reduce Monte Carlo noise on estimated probabilities of extreme events, and to generate a sample of spatial events which exceed the critical temperature level T°C somewhere on the simulated field. Here we present the method in detail, in Algorithm 2.

Figure B.18 presents a bivariate illustrative example of our scaling methodology, with $A_{t,\mathcal{S}}(T)$ as in Equation (4.16). Here, for illustrative purposes we take $\mathcal{S} = \{1, 2\}$ and $T = 30$°C. As defined in Chapter 4, $\boldsymbol{y}_i^P$ is a simulation on the standardised scale, the values of these are shown in Figure B.18 (i) for $m = 10,000$. The associated sample from the r-Pareto process' spatial profile W_r is denoted by $\boldsymbol{\omega}_i$, for $i = 1, \ldots, m$. These are equivalent to simulations with cost 1, i.e., $\boldsymbol{\omega}_i = \boldsymbol{y}_i^P / r(\boldsymbol{y}_i^P)$, as shown in Figure B.18 (ii). Figure B.18 (iii) illustrates this set of $\boldsymbol{\omega}_i$ values being scaled to have a cost of at most temperature $T^P(t, \boldsymbol{s})$, i.e., $b_{T(t)}\boldsymbol{y}_i^P / r(\boldsymbol{y}_i^P)$, according to two different time periods, 1942 (red) and 2020 (blue), with $b_{T(t)}$ is defined in Chapter 4. Finally, Figure B.18 (iv) shows the scaled fields with costs sampled from a standard Pareto distribution, which gives us $r_j^P b_{T(t)}\boldsymbol{y}_i^P / r(\boldsymbol{y}_i^P)$, used in Equa-

tion (4.17), where $r_j^P \sim$ Pareto(1). The scaling of the points gives multiple points corresponding to a realisation of the event $A_{t,\mathcal{S}}(T)$, from which the importance sampling estimator $\widehat{\Pr}_{imp}$ derives its estimate. Again, different coloured points show samples corresponding to different time periods.

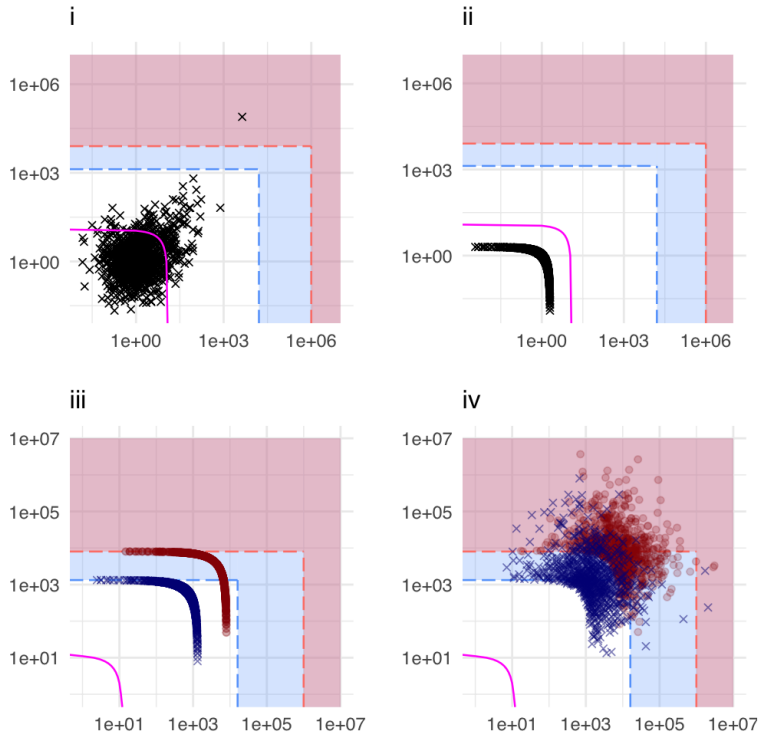

Figure B.18: Illustration of the simulation scaling procedure for a bivariate process. In each figure, the magenta line shows the threshold for extreme costs, the shaded red region is an extreme event of interest in 1942 and the blue shaded region is an equivalent event in 2020. The extreme event is taken to be 30°C at either site: (i) scatter plot of 10,000 simulated events from r-Pareto process of two sites, plotted on a log scale; (ii) events standardised to have a cost of 1; (iii) events scaled for each year; (iv) new cost sampled.

The top rows of Figure B.19 and Figure B.20 show a sample of spatial simulations, over $\mathcal{S}_o$, transformed to data scale for the marginal models M_0 and M_1 respectively. The bottom rows in these figures correspond to the change in magnitude of those actual simulations from 1942 to 2020 (a positive value corresponding to an increase over time). The equivalent plots, according to model M_2, are shown in Figure B.21.

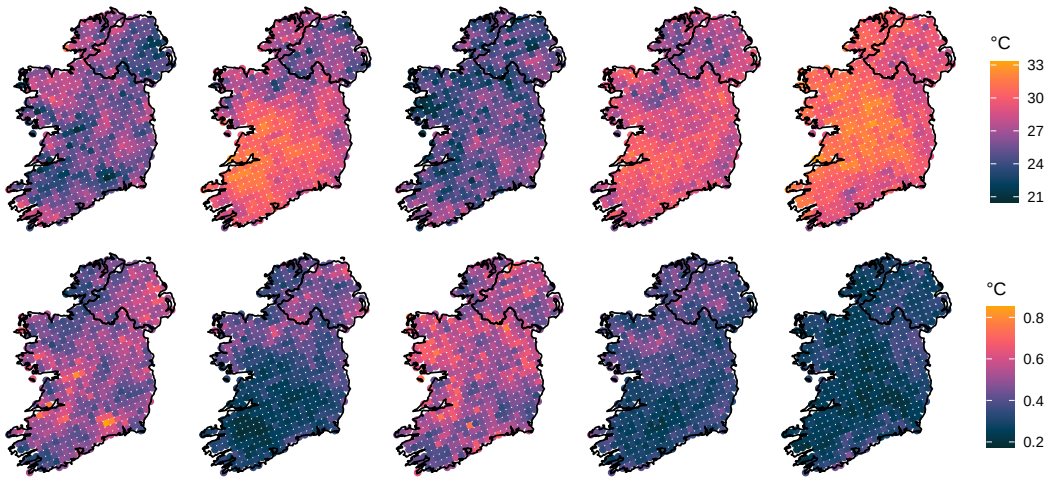

Figure B.19: Extreme events simulated from M_0 on the data scale according to 2020 (top row) and the difference in magnitude from the same simulations in the context of 1942 (bottom row).

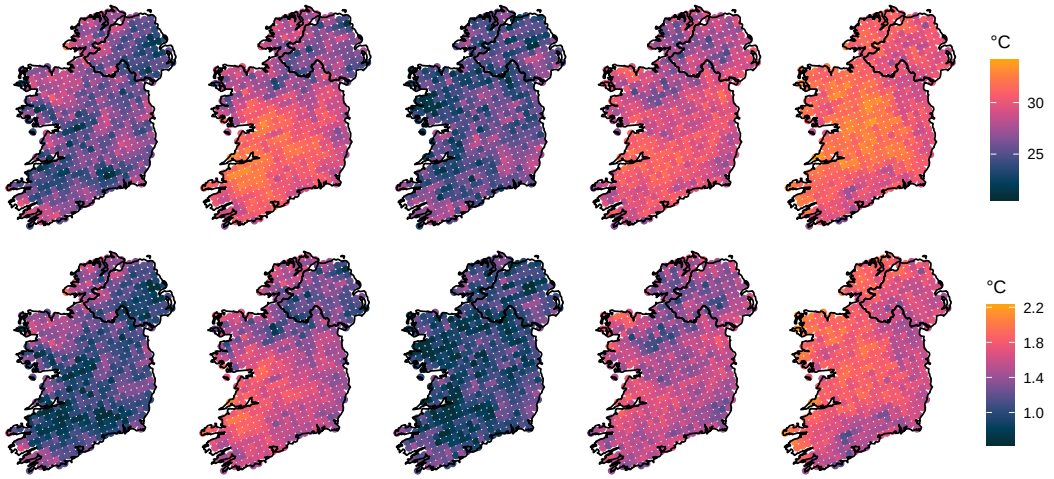

Figure B.20: Extreme events simulated from M_1 on the data scale according to 2020 (top row) and the difference in magnitude from the same simulations in the context of 1942 (bottom row).

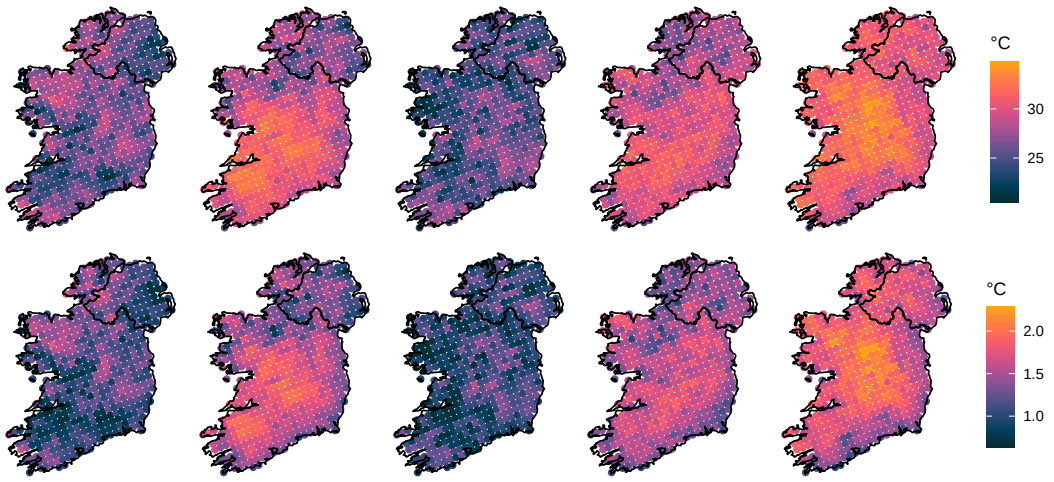

Figure B.21: Simulated extreme events from M_2 on the data scale according to 2020 (top row) and the difference in magnitude from the same simulations in the context of 1942 (bottom row).

Algorithm 2 Calculating the probability of observing an extreme temperature on $\mathcal{S}$

Input: Fitted angular model, W_r.

Choose $m \in \mathbb{R}$ // Number of simulations
Choose t // Year of interest
Choose T // Extreme temperature of interest in year t
Choose L // Number of importance samples
Calculate $T^P(t, \boldsymbol{s})$ // T on standardised scale, e.g. Pareto or Fréchet

for `i = 1,2, ..., m` **do**
    Sample $\boldsymbol{w}_i \sim \widehat{W}$ // $\boldsymbol{w}_i \in \mathbb{R}^{|\mathcal{S}|\times 1}$, $r(\boldsymbol{w}_i) = 1$
end for

for each site $\boldsymbol{s} = \{s_1, s_2, \dots\} \in \mathcal{S}$ **do**
    Calculate $\omega_{(m)}(\boldsymbol{s}) = \max_{i=1,\dots,m} \boldsymbol{w}_i(\boldsymbol{s})$ // Largest value simulated at location $\boldsymbol{s}$
    Calculate $\pi(\boldsymbol{s}) = T^P(\boldsymbol{s}, t)/\omega_{(m)}(\boldsymbol{s})$ // Scaling factor at location $\boldsymbol{s}$
end for

Calculate $b_{T(t)} = \min\{\pi(s_1), \pi(s_2), \dots\}$ // Minimum scaling factor

Initialise `exSimCount` $= 0$ // Counts number of extreme simulations

for `j = 1,2, ..., L` **do**
    Sample $r_j \sim$ Pareto(1) // Sample a new magnitude
end for

for `i = 1,2, ..., m` **do**
    for `j = 1,2, ..., L` **do**
        Let $\boldsymbol{y}^P_{i,j} = r_j \times b_{T(t)} \times \boldsymbol{w}_i$ // Scale magnitude
        Transform to data scale, $\boldsymbol{y}^D_{i,j}$
    end for

    if $\boldsymbol{y}^D_{i,j} > T$ somewhere **then**
        Increment `exSimCount`
    end if
end for

Return `exSimCount`$/\left(b_{T(t)} \times m \times L\right)$

B.6 Summary of spatial extremes

B.6.1 Various estimators of $\Pr\left(A_{t,\mathcal{S}}(T)\right)$

To estimate $\Pr\{A_{t,\mathcal{S}}(T)\}$, defined in Section 4.4.4, there are a set of possibilities. The empirical estimate from the observed data is the most basic estimator, i.e.,

$$\widehat{\Pr}_{emp}\{A_{t,\mathcal{S}}(T)\} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\left\{\exists \boldsymbol{s_0} \in \mathcal{S} : x_i^P(\boldsymbol{s_0}) > T^P(t, \boldsymbol{s_0})\right\},$$

where $x_i^P(\boldsymbol{s})$ is a realisation of $X_o^P(t_i, \boldsymbol{s})$ for observed field at time t_i and where n is the number of observed fields. For small T, the estimator $\widehat{\Pr}_{emp}$ can give useful values for testing the model fit, but for T larger than any observed event the estimate will be 0. This is where the r-Pareto process is valuable. Specifically, based on limiting approximation (Equation (4.11)), with the latter term estimated using Monte Carlo integration based on the simulated sample $\boldsymbol{y}_1^P, \boldsymbol{y}_2^P, \ldots, \boldsymbol{y}_m^P$, our raw estimator is

$$\widehat{\Pr}_{raw}\{A_{t,\mathcal{S}}(T)\} = \frac{1}{v_r m}\sum_{i=1}^{m} \mathrm{I}\left\{\exists \boldsymbol{s_0} \in \mathcal{S} : v_r y_i^P(\boldsymbol{s_0}) > T^P(t, \boldsymbol{s_0})\right\}.$$

Although $\widehat{\Pr}_{raw}$ improves on $\widehat{\Pr}_{emp}$ for moderate T, we find that $\widehat{\Pr}_{raw}$ is quite variable for large T, even for big m, and will be zero when no simulated fields exceed T°C.

Figure B.22 compares the different estimators of $\Pr\left(A_{t,\mathcal{S}}(T)\right)$ for $t = 2020$, for the station network, based on estimates $\widehat{\Pr}_{emp}$, $\widehat{\Pr}_{raw}$, and $\widehat{\Pr}_{imp}$ ($\widehat{\Pr}_{imp}$ is described in Section 4.4.4). Estimates based on $\widehat{\Pr}_{emp}$ do not extend beyond the range of the data. Estimates based on $\widehat{\Pr}_{raw}$ and $\widehat{\Pr}_{imp}$ extend to much higher temperatures. However, $\widehat{\Pr}_{imp}$ gives much less variable estimates as we extrapolate further from the data. Only the estimates from $\widehat{\Pr}_{imp}$ are presented in Chapter 4.

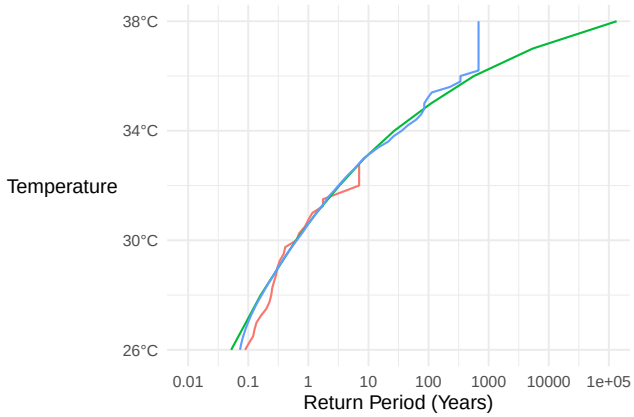

Figure B.22: Estimates of $\Pr\left(A_{t,\mathcal{S}}(T)\right)$ for $t = 2020$, for the station network, based on $\widehat{\Pr}_{emp}$ (red), $\widehat{\Pr}_{raw}$ (green), and $\widehat{\Pr}_{imp}$ (blue).

B.6.2 Summary of Spatial Extremes for model M_0 and M_1

Figure B.23 shows the return period of the event where an extreme temperature can occur anywhere on the observational network in both 2020 (dashed blue line) and 1942 (solid orange line) for model M_0 and M_1. The equivalent plot for model M_2 is presented in Figure 4.6. We can see a very slight increase in return level from 1942 to 2020 according to M_0. However, this increase is not statistically significant relative to other inference uncertainties and comes entirely from the time-changing threshold exceedance probabilities. Model M_1 estimates a substantial and much greater increase in return level from 1942 to 2020 as compared to M_0. Furthermore, return levels are much higher in 2020 according to model M_1 than M_0. These increases in return level are due to the GPD scale parameter in M_1 being allowed to vary with time. However, model M_1 suggests a longer estimated return period for each temperature level compared to M_2. This is due to M_2 having a larger scale parameter than M_1 and thus higher probabilities of associated extreme observations in the period 1942–2020.

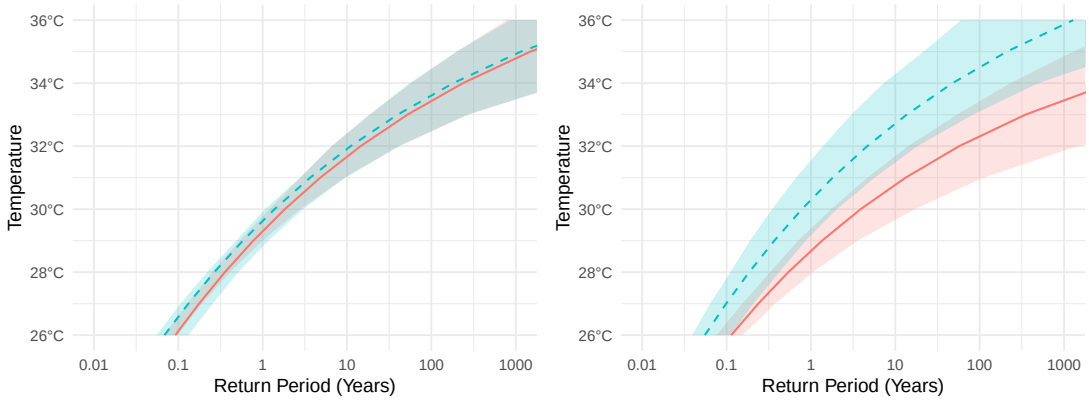

Figure B.23: Return period of an extreme temperature anywhere on the Irish observational grid. Blue dashed line corresponds to 2020 and solid orange, 1942 according to M_0 (left) and M_1 (right). Darker shaded regions correspond to overlapping uncertainties.

Figures B.24 and B.25 show the expected proportion of sites in Ireland which exceed an extreme temperature, given that one site exceeds that temperature according to model M_0 and M_1 on the observational network, the equivalent analysis for M_2 can be seen in Figure 4.8. Estimates in the left-hand panel of each plot are based on simulations under the respective model being at least T, for $T \in [26, 36]$°C, somewhere on the observational network; the right-hand panel shows the unconditioned equivalent. Model M_0 shows small levels of temporal non-stationarity (again inherited from non-constant threshold exceedance rates) while model M_1 shows levels of non-stationarity which are quite similar in magnitude to those of model M_2.

Figure B.26 and Figure B.27 show the data scale pairwise extremal coefficient, χ_o (as described in Section 4.5) for models M_0 and M_1 respectively at a selection of critical temperatures. Since model M_0 only has temporal non-stationarity represented in its marginal threshold exceedance rate and not in the actual model of the extreme temperatures, there is quite a small difference between extremal temperature dependence in 1942 and 2020. Model M_1 has very similar extremal dependence structures on the data scale to that of M_2.

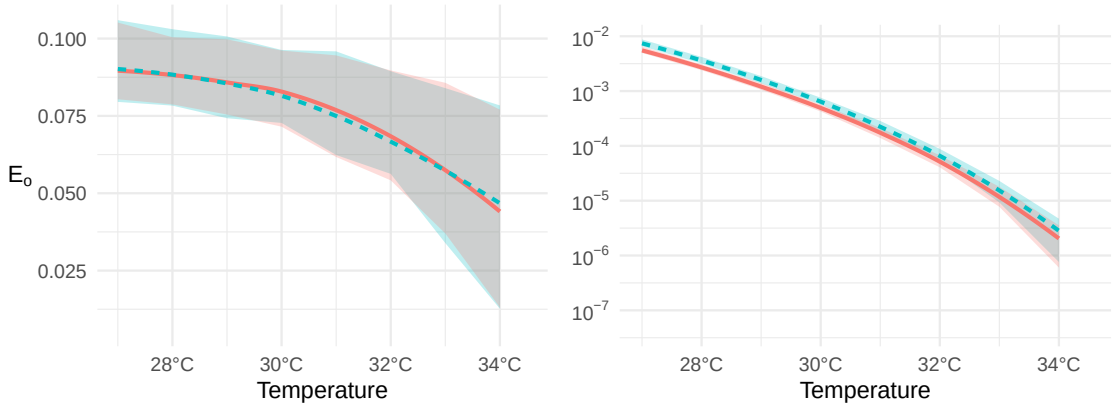

Figure B.24: Expected proportion $E_o(T;t)$ of Ireland that exceeds a temperature of T°C in an extreme event given that at least one site in Ireland exceeds T°C according to M_0. Estimates are plotted against T for 1942 (solid, orange line) and for 2020 (dashed, blue line). The left panel corresponds to estimates based on simulations scaled to be at least T°C, right panel corresponds to the unconditioned estimates.

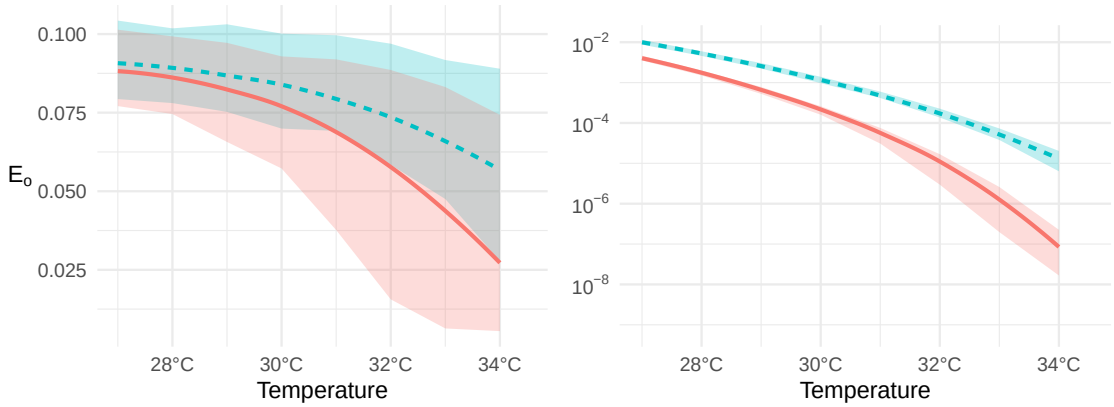

Figure B.25: Expected proportion $E_o(T;t)$ of Ireland that exceeds a temperature of T°C in an extreme event given that at least one site in Ireland exceeds T°C according to M_1. Estimates are plotted against T for 1942 (solid, orange line) and for 2020 (dashed, blue line). The left panel corresponds to estimates based on simulations scaled to be at least T°C, right panel corresponds to the unconditional estimates.

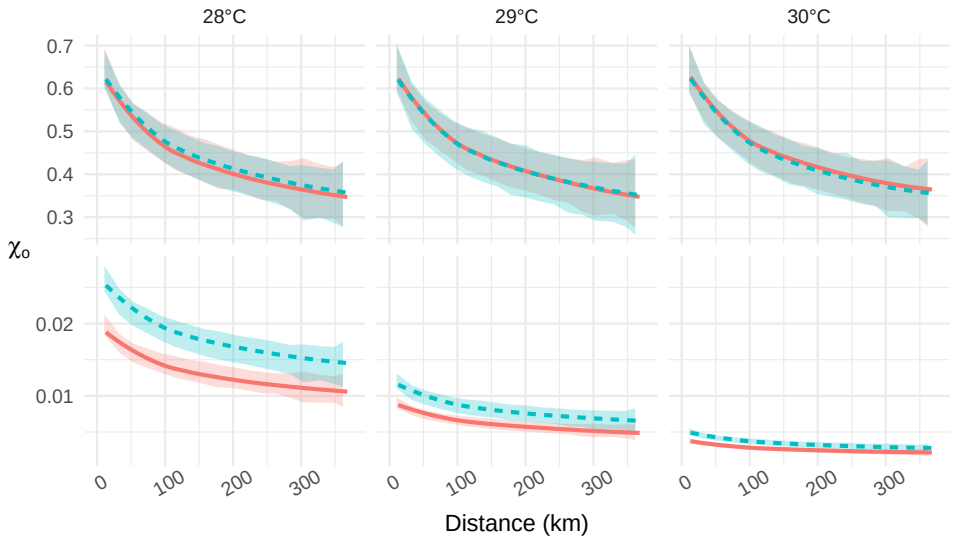

Figure B.26: Estimates of $\chi_o(h;T,t)$ against h (in km) for $T = 28, 29$, and 30°C for 1942 (solid, orange line) and for 2020 (dashed, blue line) for model M_0. Confidence intervals are based on 10,000 simulations for each of 200 bootstrapped data sets. The top row corresponds to estimates conditioned on simulations being at least 27°C, bottom row corresponds to unconditioned estimates.

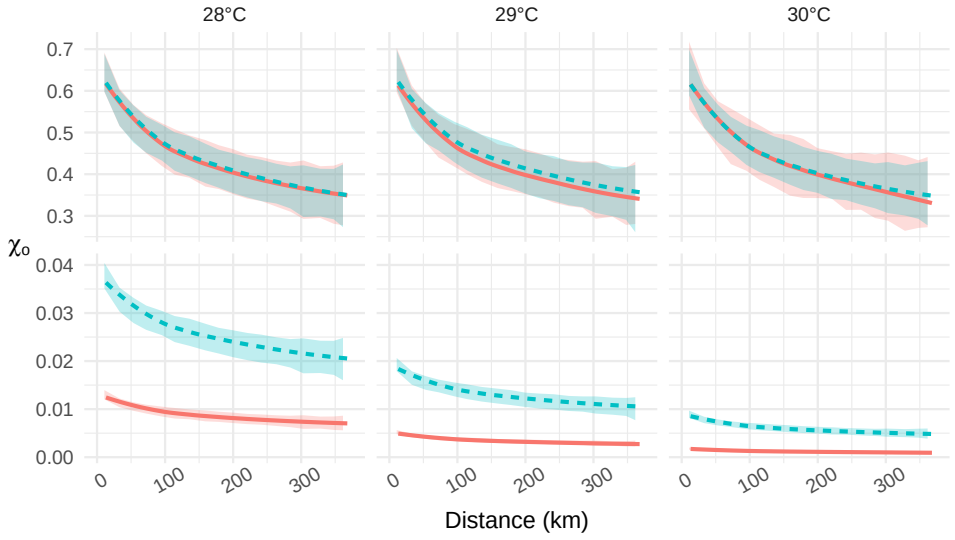

Figure B.27: Estimates of $\chi_o(h;T,t)$ against h (in km) for $T = 28, 29$, and 30°C for 1942 (solid, orange line) and for 2020 (dashed, blue line) for model M_1. Confidence intervals are based on 10,000 simulations for each 200 bootstrapped data set. The top row corresponds to estimates conditioned on simulations being at least T°C, bottom row corresponds to unconditioned estimates.

B.7 Models for alternative dependence structures

B.7.1 Asymptotic independence

Max-stable processes and r-Pareto processes exhibit a strong form of dependence, known as asymptotic dependence. Wadsworth & Tawn (2012, 2022) were the first to develop pairwise and conditional extreme value dependence models respectively which overcome this restriction, using models that also capture processes known as asymptotic independence, which for spatial processes leads to extreme events (when marginal distributions are identical) becoming more localised as they increase in marginal magnitude, see also max-infinite divisible processes which possess some similar features (Zhong, Huser & Opitz 2022). Shooter et al. (2021) and Richards et al. (2022) provide illustrations of conditional extreme value methods for inference for extreme waves at short- and long-ranges and extremes of spatially aggregated rainfall over different spatial scales respectively. Simpson & Wadsworth (2021) provides such an approach for spatio-temporal modelling of sea surface temperatures. The above examples focus on processes with a dependence structure which only depends on distance, with Huser & Genton (2016) and Richards & Wadsworth (2021) presenting extensions for when this is an unrealistic assumption.

We appreciate that for other regions and different environmental processes, this may not be a valid choice, as this variogram model cannot capture full independence between sites. More critically, the r-Pareto process fails to describe adequately any process exhibiting asymptotic independence at any distance (other than distances where the process is fully independent). For such processes, different modelling approaches such as conditional extremal (Wadsworth & Tawn 2022) or max-infinitely divisible (Bopp et al. 2021) processes need to be considered.

B.7.2 Spatially non-stationary dependence modelling

Our analysis has shown that the change in spatial extreme temperature events over Ireland in 1942–2020 has been in the marginal variables and not in the dependence structure. We have explored several potential covariates for this change, finding Ireland's mean temperature anomalies as the best descriptor, but the statistical attribution for a different rate of change in extreme temperatures remains unclear.

Climate models, run under different scenarios, using a designed experiment, could provide insights. Also, exploration of temporally lagged extremal dependence could reveal a more complex dependence structure.

We assume spatial stationarity and isotropy. Our use of the climate model data has helped identify that this may be too simplistic, with the distance of sites from the coast being a potentially useful descriptor of dependence. In exploratory work we have had reasonable success with the distance metric $\tilde{h}_{ij}$ between $\boldsymbol{s}_i, \boldsymbol{s}_j \in \mathcal{S}$ of

$$\tilde{h}_{ij} = h_{ij} + (\delta_0 + \delta_1 h_{ij}) \exp\{-\delta_2 \min(C(\boldsymbol{s}_i), C(\boldsymbol{s}_j))\}, \tag{B.1}$$

where h_{ij} is the Euclidean distance between the sites, $C(\boldsymbol{s})$ is the shortest Euclidean distances to the coastline from $\boldsymbol{s}$, and δ_0, δ_1, and δ_2 are non-negative parameters. However, any measure of distance can be used, such as the Haversine distance, which is more appropriate when analysing processes over larger spatial domains where the Euclidean distance is no longer accurate. Future work should incorporate this separation metric, with the likely consequence being the creation of extreme events which are more localised along the coast. Additionally, alternative topographical descriptors, such as altitude, could be explored in place of coastal distance in Equation (B.1).

5

Unusually cold winter events in Ireland

In this chapter we extend the methodology developed in Chapter 4 to analyse extreme daily minimum temperatures in winter months. We model the marginal distributions of extreme winter minima using a generalised Pareto distribution (GPD), capturing temporal and spatial non-stationarities in the parameters of the GPD. We investigate two independent temporal non-stationarities in extreme winter minima. We model the long-term trend in magnitude of extreme winter minima as well as the short-term, large fluctuations in magnitude caused by anomalous behaviour of the jet stream. We measure magnitudes of spatial events with a carefully chosen risk function and fit an r-Pareto process to extreme events exceeding a high-risk threshold. Our analysis is based on synoptic data observations courtesy of Met Éireann and MIDAS. We show that the frequency of extreme cold winter events is decreasing over the study period. The magnitude of extreme winter events is also decreasing, indicating that winters are warming, and apparently warming at a faster rate than maximum daily summer temperatures. We find that a climate model output which was informative as a covariate for modelling extremely warm summer temperatures is less informative as a covariate for extremely cold winter temperatures. However, we show that the climate model output is useful for informing a non-extreme temperature model.

5.1 Introduction

The frequency and intensity of extremely cold temperatures have decreased globally since 1950 with this trend expected to continue as global mean temperatures rise (IPCC 2021, Chapter 11). Winter temperatures have been particularly warming in the northern mid-latitudes (Matthes et al. 2015, Van Oldenborgh et al. 2019). In Ireland, significant warming of minimum air temperatures has been observed with decreasing frequency of cold events over the period 1885–2018 and are in line with global trends (Nolan & Flanagan 2020, Mateus & Coonan 2022, García et al. 2022). Extreme minimum temperatures are increasing faster than maxima, both globally (Dunn et al. 2020) and in Ireland, where the rate of change of temperature minima is twice that of maxima over the period 1950–2003 (McElwain & Sweeney 2003).

Extremely high temperatures receive a substantially disproportionate amount of analysis in the literature given their direct and immediate link to loss of life (Ballester et al. 2023), crop failure (He et al. 2022), etc. Furthermore, the intensity, duration and extent of heatwaves are expected to increase with global temperatures (see Chapter 4), exacerbating impacts and demanding immediate attention. While the warming of winter temperatures is less immediately disruptive, it nonetheless harbours many potentially devastating consequences. Changes in winter extremes are having an effect on many facets of our environment and society, from mortality and morbidity rates (Conlon et al. 2011), to agricultural activity (Bindi & Olesen 2011, Hooker et al. 2008) and ecosystems (Osland et al. 2021). For example, many regions rely on cold winters to control pathogens and pest populations (Skendžić et al. 2021). It has been argued that the consequences of warming winters have been understudied (Boucek et al. 2016).

Irish winters are generally less severe compared to many countries at similar latitudes due to the regulating effect of the North Atlantic Ocean and the Irish Sea surrounding the island. The winter of 2019 was the warmest on record for Ireland at 0.9°C above the 1961–1990 average winter temperature (Met Éireann 2019). However, in northern Europe, warming trends are contrasted with occasional, extreme cold events associated with variability in climatic oscillations, i.e., Arctic Oscillation (AO) and the North Atlantic Oscillation (NAO) (Vihma et al. 2020). The coldest temperatures in some parts of Ireland were observed as recently as 2010, for

example, a record low of -17.5°C was observed in Co. Mayo, Straide (Christiansen et al. 2018). The extremely cold winter of 2009/10 was caused by an anomalously persistent negative phase of the North-Atlantic Oscillation (Cattiaux et al. 2010).

Figure 5.1 shows that globally (left-hand plot), the winter months of December, January, and February (DJF) of 2010 were anomalously warm, despite large variations locally, for example in Ireland (right-hand plot). Disentangling these juxtaposing trends in winter extremes is important to fully ascertain the overall trend of winter extreme temperatures. Christiansen et al. (2018) show that the cold winter of 2009/10 was in fact warmer than expected, in the context of the climate conditions at the time. The authors estimate that the occurrence probability of extremely cold winter temperatures such as those seen in 2009/10 has reduced by a factor of 2 due to anthropogenic-induced climate change, and so were in line with global warming winter trends.

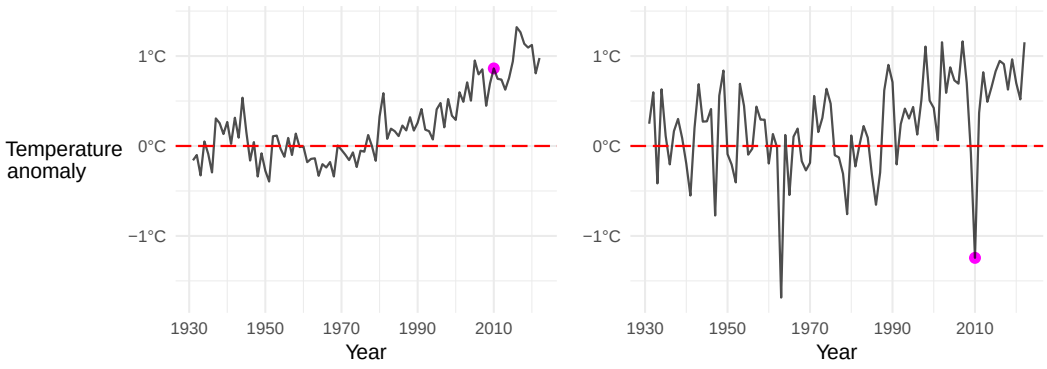

Figure 5.1: Global (left) and Irish (right) mean winter (DJF) temperature anomalies with 2010 highlighted. Calculated from HadCRUT5 (Morice et al. 2021).

Our analysis aims to test for and identify non-stationarities in winter temperature extremes to give a more thorough picture of the impact of climate change on winter temperatures in Ireland. To avoid underestimating the severity of extreme weather events, we rely on asymptotically justified statistical models provided by extreme value theory. Advances in the extreme value literature enable us to model the spatial dependence of extreme observations, which allows the simulation of spatially realistic winter temperature events. We use the generalised r-Pareto process (Dombry & Ribatet 2015) to model extremal dependence since this approach involves fitting the model to full spatial observations. The r-Pareto process more

accurately captures realistic spatial patterns of extreme events as opposed to the max-stable process, which models constructed componentwise maxima, which typically do not correspond to the actual observed spatial events. We generate high dimensional, spatially realistic extreme temperature events over Ireland and make inferences about future spatial events, which have higher magnitudes than previously observed. It is generally accepted that winter temperatures are warming but to our knowledge, no analysis has been done on spatio-temporal non-stationary winter extremes in Ireland. An analysis of trends in unusually cold winter temperatures is vital for characterising expected changes in the future. Furthermore, a local analysis allows for the development of bespoke policy and the circumvention of potential impacts.

In our analysis of extreme minimum winter temperatures, we largely follow the modelling methodology developed in Chapter 4. In order to focus on the extensions and novelty of the method here, the reader is referred to the corresponding subsections in Chapter 4 where a detailed discussion of the modelling procedure is given and further technical detail is provided. This chapter is organised as follows. Section 5.2 details the observational and climate model data used as well as additional covariates explored. Section 5.3 and Section 5.4 describe the marginal and dependence modelling of the process respectively. In Section 5.5, we use the model to explore how the properties of spatial extreme events have changed over time. Conclusions and a broader discussion are given in Section 5.6. All our code and instructions on how to access the data are available on GitHub[a].

5.2 Data

Modelling extreme minima of a series of *iid* observations is practically equivalent to modelling extreme maxima (see Section 2.1.3.1). This is achieved through an inverse transformation of the data, that is,

$$\min(X) = -\max(-X). \tag{5.1}$$

[a] https://github.com/dairer/Extreme-Irish-Winter-Temperatures.

For our analysis, we apply transformation (5.1) to our series of minimum daily temperatures and analyse them as maxima. Finally, we report all results on the original scale, after applying the inverse of (5.1). For ease of interpretation, we discuss temperatures on the original scale throughout. We use several data sources to achieve a full spatial and temporally non-stationary model of extremely cold winter temperatures. These data sources are discussed below.

5.2.1 Station data

For the purpose of this study, we focus on modelling the extremal behaviour of observational temperature data. We consider temperature data to avoid pre-processed data products which may influence or smooth out extremal behaviour (Donat et al. 2014). This can include omittance of extreme values incorrectly classified as outliers and smoothing of spatial patterns which can have a great impact on the magnitude of unusually large/small events.

Our daily minimum temperature data comprise Irish temperature stations compiled from two sources. The data sources are equivalent to those described in Section 4.2, with observational data for the Republic of Ireland provided by Met Éireann's data archive[b], while Northern Ireland sites were obtained through the CEDA archive (Met Office 2012). The data exhibits the same missingness complications as discussed in Section 4.2.1, with data availability greatly decreasing as we go back in time. Although we have some data prior to 1950, for the purpose of comparing different climatic covariates (discussed in Section 5.2.3.1), which are only available mutually since 1950, we further restrict our analysis to data from the period 1950–2022. The data set contains some outliers that were not physically realistic, though we took a conservative approach to remove these values to avoid discarding any true, extreme observations. Thus, we only removed any temperatures that were greater than 4 (empirically estimated) standard deviations away from the mean minimum temperature on a given day. This procedure removed 93 unrealistic observations from the data set. Keeping sites that had at least 5 years of data, we have a collection of 125 sites with a total number of approximately $330,000$ daily minimum winter (DJF) temperature records over Ireland.

Our interest is in extreme cold temperatures in Ireland, so we restrict our analysis

[b] https://www.met.ie/climate/available-data/historical-data.

to data from the winter months (DJF). In Figure 5.2, we look at the number of exceedances per month when we take high site-wise quantiles as thresholds. The majority of extreme cold events occur in the winter months December, January, and February. Specifically, we find that more than 80% of all the days with temperatures below the 1% site-wise marginal quantile occur in winter (DJF). However, there is a non-negligible number of extreme events in March (12% of events below the site-wise 1% quantile). We initially modelled the winter months (DJF) including March in our analysis, however we found that the weather processes governing extremely low temperatures in the month of March were not captured well by our modelling procedure. The inclusion of March requires careful modelling of seasonality which is beyond the scope of this work and so we leave this as future research. We thus model only the winter months (DJF) which contain the majority of temperature extremes.

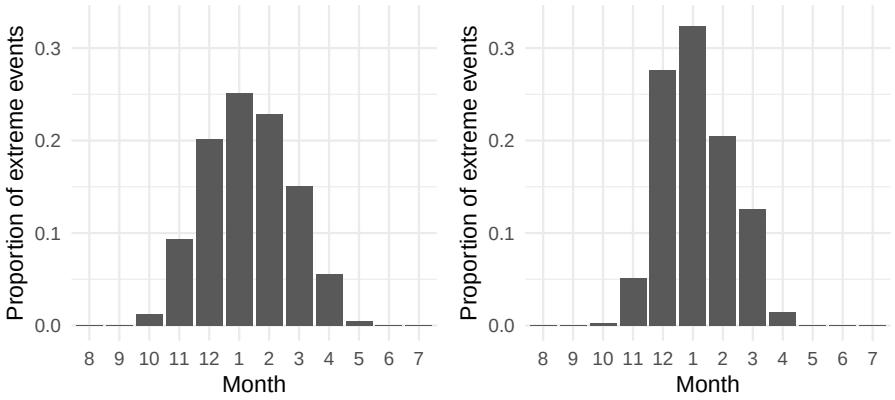

Figure 5.2: Proportion of threshold exceedances per month where the threshold corresponds to 5th (left) and 1st (right) site-wise percentile. Months labelled 1–12 correspond to January–December, respectively.

5.2.2 Spatial covariates

5.2.2.1 Climate model output

We explore the efficacy of exploiting physical information from climate model outputs as a spatial covariate. Climate models are mathematical representations of the physical processes driving weather and climate and represent our best understanding of these natural phenomena (Giorgi 2019). They are computationally expensive, and so to model the climate on a fine scale requires limiting ourselves to

specific regions. Climate models are broadly run on two scales, global climate models (GCMs) and regional climate models (RCMs). We use the output from RCMs to incorporate more detailed descriptions of the physics and spatial characteristics into the modelling of temperature extremes. We take data from the CLMcom-CLM-CCLM4-8-17 RCM (see Figure 5.3) combined with the ICHEC-EC-EARTH GCM (Copernicus Climate Change Service 2019), which was used in Chapter 4 and proved useful for informing extreme summer temperature models. From the models we have daily minimum temperature values over a 56-year period, on a regular grid of 558 points over Ireland (corresponding to a 0.11^2 degree resolution). For a given ensemble we are required to select a so-called "experimental configuration". Historical experiments are climate simulations for a period in which observational data exists. They reflect observed changes in climate. For our chosen ensemble this period covers 1951–2005. To simulate data beyond this period it is necessary to make some assumptions about climate change. However, our analysis does require climate model output for informing temporal non-stationarity so we rely solely on the historical experiment.

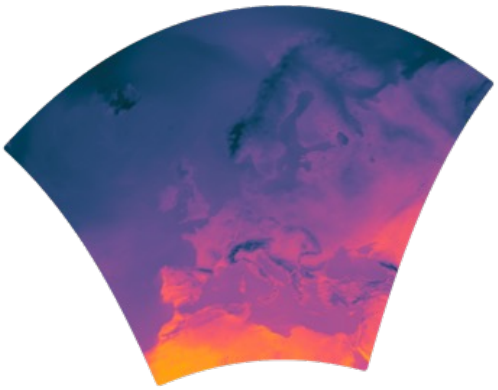

Figure 5.3: Full spatial extent of the RCM CLMcom-CLM-CCLM4-8-17, showing a generated temperature event for a randomly chosen day.

5.2.2.2 Coastal distance

Our analysis in Chapter 4 found coastal distance to be a useful covariate for describing the spatial variation in extremely hot summer temperatures. Furthermore, there is an evident difference in temperature levels between coastal and inland areas in Ireland, especially so in winter (Mateus & Coonan 2022). This spatial effect is due to the strong influence of the Irish Sea and the Atlantic Ocean on Irish air

temperatures. We are thus motivated to employ coastal proximity as a covariate to examine its efficacy in explaining the spatial distribution of cold extremes. We denote coastal proximity at site $\boldsymbol{s}$ as $C(\boldsymbol{s})$.

5.2.3 Long-term temporal trend covariate

For modelling the long-term temporal trend in winter extremes, we use temperature anomalies as a temporal covariate in order to ascertain how extreme cold temperatures are changing with respect to mean temperature levels. We perform LOESS smoothing on temperature anomalies during winter months, over the grid point containing Ireland from the HadCRUT5 data set (Morice et al. 2021). We use the R language implementation of LOESS smoothing, taking the default span of 0.75 and degree of 2 (R Core Team 2023). We denote this smoothed temporal covariate as $M^I(t)$ and use it to account for the large-scale temporal trend in the data. The covariate $M^I(t)$ is plotted in Figure 5.4, illustrating an increase of approximately 0.9°C in mean winter temperature anomalies over the period 1950–2022.

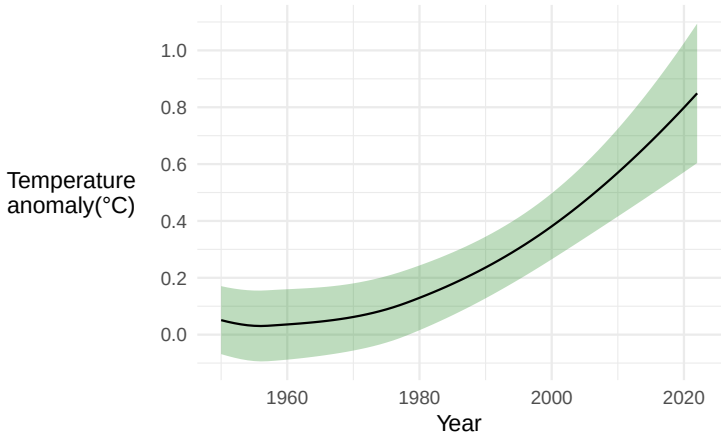

Figure 5.4: LOESS smoothed Irish winter temperature anomalies from HadCRUT5 with shaded regions showing 95% confidence intervals.

5.2.3.1 Climatic oscillation covariates

In our analysis, we aim to disentangle the general, long-term trend of extremely cold winter temperatures from the large variations caused by extremal climatic processes. To this end, we explore numerous climatic variables to help explain

extreme, sudden cold snaps, which can misleadingly be interpreted as contradicting warming winters.

It is well understood that the behaviour of the jet stream can lead to shifts in weather patterns, including the occurrence of extreme events such as cold spells (Stendel et al. 2021). The jet stream is a high-altitude, fast-flowing air current that circumvents the northern hemisphere in the upper troposphere and lower stratosphere. The jet stream plays a crucial role in shaping weather patterns and controlling the movement of storms and weather systems (Hoskins & James 2014). The jet stream meanders in a wave-like pattern as it circumnavigates the globe. These waves form and change in response to several climatic conditions and are characterised by northward or southward oscillations. When these waves become amplified, we see a greater north-south displacement of air masses, resulting in the transport of air masses from different latitudes. When a deep southward dip forms in the jet stream, it can allow cold air to spill southward from the polar regions into lower latitudes causing extremely cold temperatures (Thompson & Wallace 2001, Francis & Vavrus 2012).

There is a strong link between the behaviour of the jet stream and extratropical teleconnections such as the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) (Hall et al. 2015). Both the NAO and AO have the greatest variability in winter months (Sen Gupta & McNeil 2012) which is when they have a substantial influence on weather patterns in the northern hemisphere (Thompson & Wallace 1998). Furthermore, persistent and strong negative phases of the AO and the NAO have been linked with extremely low minimum air temperatures in Ireland (Mateus & Coonan 2022). Thus, we are motivated to consider the efficacy of using NAO or AO as a covariate for anomalous jet stream behaviour and subsequently extreme cold winter temperatures in Ireland.

The NAO is a climatic phenomenon that describes the state of the atmospheric pressure difference between the Icelandic Low and the Azores High in the North Atlantic. The NAO is known to influence the behaviour and position of the jet stream (Gerber & Vallis 2009). In the positive phase, the pressure difference between the Icelandic Low and the Azores High is stronger than average. The stronger pressure difference confines the jet stream to higher altitudes, with a steady eastward flow. Conversely, during the negative phase of the NAO, the pressure difference weakens,

allowing for a more meandering and meridional path of the jet stream. This allows the jet stream to dip southwards, pouring cold air masses onto northern Europe, leading to extremely cold temperatures. As a result, the jet stream may meander or become more amplified. A wavier jet stream generally leads to slower progression of weather systems, and subsequently more persistent weather regimes, such as persistent cold air in a cold spell (Sousa et al. 2018). The NAO can be interpreted as the North Atlantic sector manifestation of the AO (IPCC 2021, Annex IV: Modes of Variability) and so, the two variables are highly correlated (with a correlation coefficient of approximately 0.77) and can be seen in Figure 5.5.

The AO characterises changes in atmospheric pressure over the Arctic region (Thompson & Wallace 1998). Analogous to the NAO, during the positive phase of the AO, the pressure difference between the Arctic and mid-latitudes is weaker and the jet stream is confined to higher latitudes. This means that weather systems and air masses tend to move more smoothly and quickly across the mid-latitudes (Deser 2000). Conversely, during the negative phase of the AO, the pressure difference between the Arctic and mid-latitudes increases with elevated air pressure over the Arctic region and lower air pressure over the northern Pacific and Atlantic Oceans. Lower air pressure over the northern Pacific and Atlantic Oceans allows for an amplified meandering of the jet stream with a higher probability of extremely cold, polar air outbreaks in the mid-latitudes during winters.

We retrieve data for the NAO and AO from the National Weather Service, Climate Prediction Center[c], taking monthly and yearly mean AO and NAO values from 1950 onwards. The climatic oscillation covariates are shown in Figure 5.5. Notice that in both plots, we see a large deviation from the mean around the year 2009/10, which appears more pronounced in the AO series.

[c]Available at: https://www.cpc.ncep.noaa.gov.

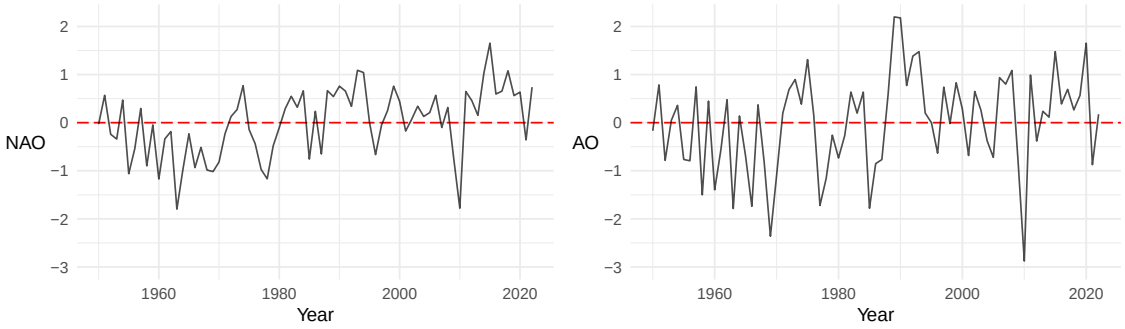

Figure 5.5: Yearly North Atlantic Oscillation (NAO) index over the period 1950–2022 (left). Yearly Arctic Oscillation (AO) index over the period 1950–2022 (right).

NAO and AO indices are not a perfect proxy for the behaviour of the jet stream. Other atmospheric and oceanic patterns, such as high-pressure systems and sea surface temperatures, can also interact with the jet stream and influence the occurrence of cold weather events (Screen & Simmonds 2014). Overall, the interaction between the jet stream, atmospheric, and oceanic patterns is a complex and dynamic process (Hall et al. 2015). The interplay of these factors can lead to changes in the jet stream's behaviour, which in turn affects weather patterns, including the occurrence of cold spells (Kidston et al. 2015).

5.2.4 Short-term temporal trend covariate

The HadCRUT5 data set which we use to derive our long-term temporal covariate (see Section 5.2.3) seems an attractive candidate to describe processes governing cold temperatures since HadCRUT5 (Morice et al. 2021) is an observation-based data product of temperatures produced as a combination of a land temperature anomaly data set (CRUTEM5; Osborn et al. 2021) merged with a sea-surface temperature anomaly data set (HadSST4; Kennedy et al. 2019). An appealing feature of this procedure is that it accounts for and encapsulates the effects of all climate processes that affect temperature, e.g., including the NAO and AO, and therefore the behaviour of the jet stream. We wish to capture large temperature deviations which are encoded in the HadCRUT5 data set. To achieve this, we use the residuals from the LOESS smoothing (see Section 5.2.3) as a covariate to explain the effect of large-scale climatic oscillatory patterns on winter extremes. In this way, we con-

struct two independent covariates from the HadCRUT5 data set. We explore the efficacy of taking residuals from the LOESS smoothing of temperature anomalies during winter months ($M^I(t)$) on a monthly (denoted $M^I_{r,m}(t)$) and yearly averaged (denoted $M^I_r(t)$) basis, both of which can be seen in Figure 5.6. Notice in these time series, the large positive deviation around the year 2009/10, as the LOESS smoothing under-represents these extremely cold winters. This deviation corresponds to a large, negative temperature anomaly over Ireland described in the HadCRUT5 data set.

Furthermore, we explore the efficacy of using temperature residuals from grid boxes surrounding Ireland, as well as the grid box directly above Ireland, after removing the smoothed temperature $M^I(t)$ trend. Both time series as well as their associated HadCRUT5 grid box(es) can be seen in Figure 5.7. In both time series, we can clearly see large deviations in mean temperatures, for example in the year 2009/10. Since we are taking the residual temperature anomalies after removing the smoothed temperature trend $M^I(t)$, positive values correspond to anomalously cold temperatures not captured by $M^I(t)$.

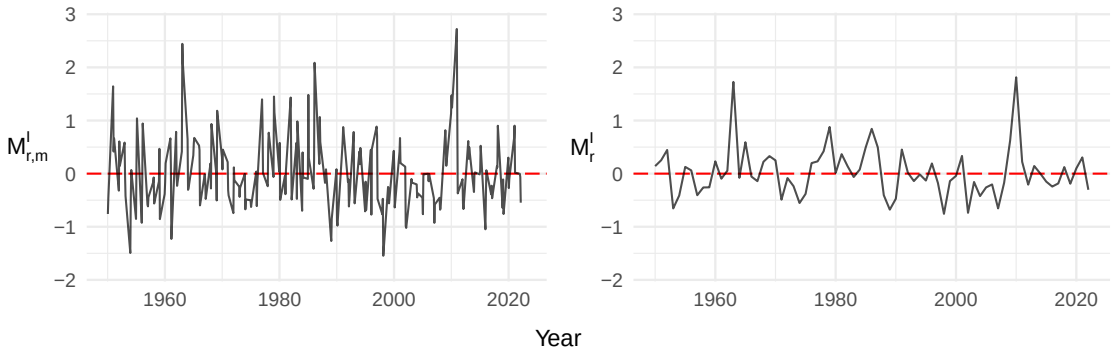

Figure 5.6: Residuals from HadCRUT5 temperature values over Ireland after LOESS smoothing (left). Average yearly residuals from HadCRUT5 winter temperature values over Ireland after LOESS smoothing (right).

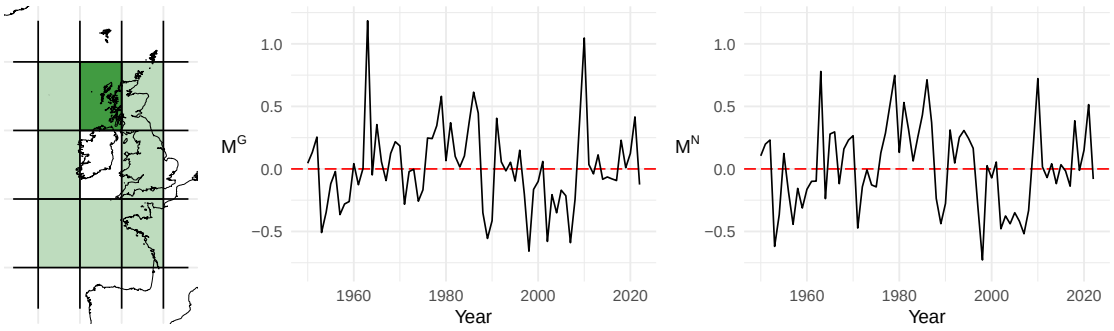

Figure 5.7: Illustration of gridding of the HadCRUT5 data set over northwest Europe centred on Ireland, with grid boxes surrounding Ireland lightly shaded and the grid box north of Ireland heavily shaded (left). Time series of residual yearly temperature anomalies averaged from all shaded grid boxes surrounding Ireland after removing the smoothed temperature $M^I(t)$ trend (centre). Time series of yearly residual temperature anomalies from grid box north of Ireland after removing the smoothed temperature $M^I(t)$ trend (right).

5.3 Marginal models

5.3.1 Overview and strategy

Extreme minimum daily temperature data presented some unique challenges which required adaptations and extensions to the methodology set out in Chapter 4. Namely, our methodology is adapted here to account for high levels of autocorrelation within cold temperature extremes, and high variability of extreme temperatures governed by external large-scale climatic patterns. We carry the notation set out by Chapter 4, where instead now let $X_{\mathrm{o}}(t, \boldsymbol{s})$ denote the station data comprising winter minimum daily temperature at time t and site $\boldsymbol{s}$, and let $X_{\mathrm{c}}(t, \boldsymbol{s})$ be the equivalent process from the climate model data. The subscripts o and c are used throughout, referring to observation and climate model processes respectively, the subscripts are dropped when discussing methods mutual to both. Here $t \in \mathbb{N}$ indexes winter days within and across years and $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$, where $\mathcal{S}$ denotes Ireland, with $\boldsymbol{s}$ corresponding to the vector of latitude and longitude. We have data on the two processes at $\mathcal{S}_o \subset \mathcal{S}$ and $\mathcal{S}_c \subset \mathcal{S}$ and at times $\mathcal{T}_o$ and $\mathcal{T}_c$ respectively. Throughout the analysis we investigate the effect of monthly covariates; these will be indicated by a subscript m. Covariates without subscript m can be assumed to

be yearly.

For modelling $X_o(t, \boldsymbol{s})$, we account separately for the behaviour below and above a threshold $u_o(\boldsymbol{s})$, which is fixed over time. In Section 5.3.3.1, we use interpolated quantiles estimated via spatial quantile regression for the data below the threshold. We also derive our estimate of the threshold and threshold exceedance parameter in Section 5.3.3.3, allowing for a smooth and continuous transition from the bulk to the tail model. In Section 5.3.3.4, we present the parameterisations of the generalised Pareto distribution we explored. We discuss the efficacy of using the climate model output to inform the spatial behaviour of extreme winter temperatures. In Section 5.3.3.6, we highlight our bootstrapping procedure for uncertainty estimations. We detail a novel, relaxed bootstrapping algorithm that deals with highly auto-correlated extremal data with increased missing data.

5.3.2 Cross-validation

For characterising marginal model fit performance and model selection, we use cross-validation (Hastie et al. 2001, Ch 7.). We use both standard 45-fold CV (45-CV) and spatio-temporal CV (ST-CV) with 45 folds. For 45-CV the data are divided into 45 groups (folds), where each fold is removed in turn and the model is fitted to the remaining folds. For ST-CV, observational sites are divided spatially into 15 contiguous groups and 3 temporal folds. Each temporal fold consists of every third week in the winter months in order to preserve long-term temporal non-stationarity. We define the 45 ST-CV folds as all combinations of spatial and temporal clusters, taking the intersection as a fold.

We iteratively excluded one fold from the data set and fit the model to the remaining data. We test the model's predictive ability on the out-of-sample data using the root mean square error (RMSE) and the continuous ranked probability score (CRPS, Gneiting & Katzfuss 2014). The empirical quantile $x_o^{(\tau)}(t, \boldsymbol{s})$ is evaluated using the ordered data at site $\boldsymbol{s}$ and the year which contains time t, whereas the predicted quantiles are estimated as $\hat{x}_o^{(\tau)}(t, \boldsymbol{s}) = F^{-1}_{o,t,\boldsymbol{s}}(\tau)$ for quantile τ from the appropriate model F. This process is repeated for each fold, with final performance metrics reported as the average across folds.

5.3.3 Marginal data analysis

5.3.3.1 A model for the body of the distribution

Following the method described in Section 4.3.2, to estimate the distribution function of $X_o(t, \boldsymbol{s})$, we use the asymmetric Laplacian distribution (ALD) for quantile regression. We estimate a range of spatially and temporally varying τth quantiles of $X_o(t, \boldsymbol{s})$ (denoted as $q_o^{(\tau)}(t, \boldsymbol{s})$) for a grid of $\tau \in \{0.01, 0.02, \ldots, 0.99\}$, for all $t \in \mathcal{T}$ and $\boldsymbol{s} \in \mathcal{S}$. Following this, we use a cubic interpolation spline for each t and $\boldsymbol{s}$ to give a continuous estimate over $\tau \in [0.01, 0.99]$.

We explored several potential parameterisations for the ALD, which we present in Table 5.1, along with their cross-validation metrics (calculated as in Section 5.3.2). The first four models are similar to those explored in Section 4.3.5.2. The first model in Table 5.1, *a.*, can be considered as the base model where a constant quantile is estimated over space and time for each τ. The second model, *b.*, includes the smoothed temperature anomalies over Ireland, $M^I(t)$, as a covariate to allow for temporal non-stationarity in the model of the body of the distribution. The third model, *c.*, allows the quantile estimates to vary spatially using the corresponding climate model output quantiles $q_c^{(\tau)}(\boldsymbol{s})$ as a covariate. The fourth model *d.* combines the coefficients of models *b.* and *c.*, allowing for both spatial and temporal non-stationarity.

Model *d.* was the best-performing model for describing maximum daily summer temperatures found in Chapter 4. However, for modelling minimum winter temperatures, we found the performance of the bulk model greatly improved when incorporating a coastal proximity covariate $C(\boldsymbol{s})$. This is shown in the reduction of RMSE values achieved by model *e.* over *d.* in Table 5.1. We found a further substantial improvement in RMSE with the inclusion of a covariate to account for unusually cold weather events caused by large-scale climate oscillation. We chose the covariate that was best performing in the tail model (discussed in detail in Section 5.3.3.4) for the sake of parsimony, continuity, and interpretability from the bulk-to-tail model. That is, we use the residuals from the LOESS smoothing of temperature anomalies during winter months, $M^I_{r,m}(t)$ as a covariate in the bulk model. We use bulk model *f.* for the remainder of the analysis.

Table 5.1: Cross-validation (RMSE) on the quantile regression analysis for the body of the distribution.

	Model structure for $\hat{q}_o^{(\tau)}(t, \boldsymbol{s})$	**ST-CV**	**90-CV**
a.	$\beta_0^{(\tau)}$	2.019	2.028
b.	$\beta_0^{(\tau)} + \beta_1^{(\tau)} M^{\mathrm{I}}(t)$	1.991	2.028
c.	$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\boldsymbol{s})$	1.898	1.912
d.	$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\boldsymbol{s}) + \beta_2^{(\tau)} M^{\mathrm{I}}(t)$	1.864	1.891
e.	$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\boldsymbol{s}) + \beta_2^{(\tau)} M^{\mathrm{I}}(t) + \beta_3^{(\tau)} C(\boldsymbol{s})$	1.764	1.756
f.	$\beta_0^{(\tau)} + \beta_1^{(\tau)} q_c^{(\tau)}(\boldsymbol{s}) + \beta_2^{(\tau)} M^{\mathrm{I}}(t) + \beta_3^{(\tau)} C(\boldsymbol{s}) + \beta_4^{(\tau)} M_{r,m}^{\mathrm{I}}(t)$	**1.218**	**1.221**

The coefficients of model *f.* are shown in Figure 5.8 along with 95% uncertainty intervals based on 200 spatio-temporal bootstrap samples, described later in Section 5.3.3.6. Since we are modelling negated temperatures, higher quantiles correspond to colder temperatures. The estimates of $\beta_1^{(\tau)}$ show that the climate model does not provide a perfect description of the station data, as the estimates differ from 1 significantly and change with τ. This suggests that the chosen climate model output is not sufficient in and of itself in representing daily winter minima. However, the climate model is still informative here, especially at quantiles near the median of the data ($0.4 < \tau < 0.6$), where it is significantly different from zero. However, the climate model covariate appears to become less informative towards the extremes of the process as we see the effect of the q_c^{τ} falling off at both tails of the distribution ($\tau < 0.2$ and $\tau > 0.8$). For the majority of the body of the distribution, we see that $\beta_2^{(\tau)} \approx -1$. This indicates that mean winter temperatures in Ireland are a good representation of the temporal change for the majority of the body of the distribution. However, we see the effect of $M^I(t)$ increasing at higher quantiles ($\tau > 0.75$), suggesting extremely cold daily winter temperatures events are changing faster than mean minimum daily winter temperatures. We see that the effect of coastal proximity increases almost linearly with quantiles τ, giving a sufficient spatial description of the process not provided by the climate model output. Finally, $\beta_4^{(\tau)}$ suggests, unsurprisingly, that large-scale oscillations are useful for explaining the most extreme cold events, for example, seen as the increase in the uppermost quantiles ($\tau > 0.75$).

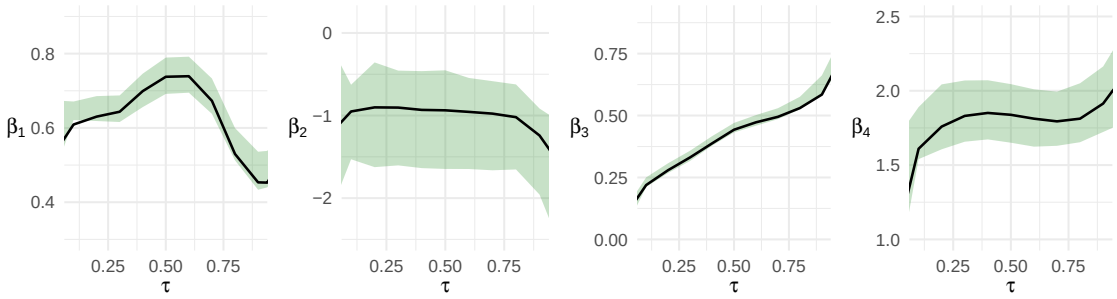

Figure 5.8: Estimates of bulk model *f.* coefficients $\beta_1^{(\tau)}$, $\beta_2^{(\tau)}$, $\beta_3^{(\tau)}$, and $\beta_4^{(\tau)}$ of $q_c^{(\tau)}$ (corresponding climate output quantile), $M^I(t)$ (smoothed temperature anomalies over Ireland), $C(\boldsymbol{s})$ (coastal distance), and $M^I_{r,m}(t)$ (monthly residuals of $M^I(t)$) respectively over a range of quantiles τ. In each case, the shaded region indicates bootstrap-based pointwise 95% confidence intervals. Higher quantiles correspond to colder temperatures.

5.3.3.2 Phases of short-term climate variability (SCV)

Since we use a highly-variable covariate, $M^I_{r,m}(t)$, in the bulk model parameterisation *f.* (and in the tail model), every result derived using this covariate will be similarly highly variable and not smooth. For the sake of interpretability, we report results at three different quantile levels of the variable $M^I_{r,m}(t)$. That is, we present results associated with the 0.1, 0.5 and 0.9 quantiles of the covariate $M^I_{r,m}(t)$, and denote these levels as $M^{(\tau),I}_{r,m}(t)$ where $\tau \in \{0.1, 0.5, 0.9\}$. Throughout this analysis, we will refer to these as low, median, and high phases of short-term climate variability (SCV) respectively. Reporting results associated with quantiles of the covariate alleviates any findings associated with the variability of any specific covariate. Instead, this method provides a general interpretation of different levels of extremity in the context of different phases of large-scale climatic processes. Note that the high phase of SCV in $M^I_{r,m}(t)$, (i.e., $M^{(0.9),I}_{r,m}(t)$) corresponds to a large negative phase of NAO and AO and subsequently a pronounced meandering of the jet stream which we know is linked to more extreme cold spells in Northern Europe, as well as Ireland (as discussed in Section 5.2.3.1)

5.3.3.3 Threshold selection

For threshold selection, we examine the stability of the parameters of the generalised Pareto distribution over a range of thresholds as shown in Figure 5.9. For a sequence

of 200 quantiles, τ, between 0.8 and 0.999 we empirically estimate the temperature at that quantile for each site ($\boldsymbol{s}$) separately. We treat this quantile as a threshold and fit a generalised Pareto distribution to the excesses above each threshold, pooled over space. The estimates of the generalised Pareto distribution parameters $\ln \sigma$ and ξ are plotted, along with associated 95% confidence intervals for each quantile. Since the generalised Pareto has a threshold stability property (Davison & Smith 1990), we know that once a sufficiently high threshold is chosen, the exceedance of any higher threshold will be similarly generalised Pareto distributed with the same shape parameter and a deterministic shift in the scale parameter. In Figure 5.9, we can see this appears to happen around the 0.96 quantile. We are conscious of the fact that this is quite a high threshold (given only three months of data per year) and so to ascertain the sensitivity of this choice, we perform analysis based on the three quantiles marked as vertical lines in the figure, i.e., we use the $0.95, 0.96$, and 0.97 quantiles as thresholds. Our primary analysis is performed using the 0.96 quantile, however we calculate results at all three quantiles (where computationally feasible) to establish threshold selection sensitivity.

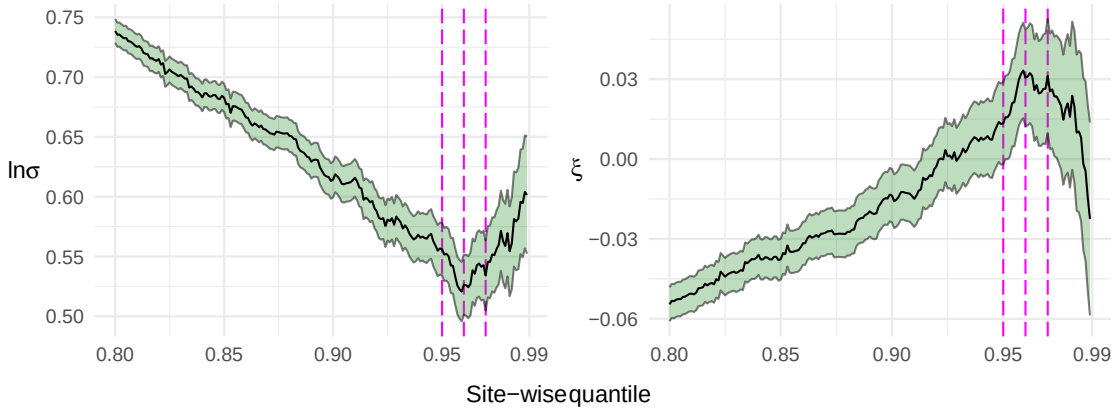

Figure 5.9: Parameter stability plot of the GPD parameters, σ and ξ over a range of quantiles used to calculate a site-wise threshold. Three vertical dashed lines represent the 0.95, 0.96, and 0.97 quantiles. The shape parameter, ξ, appears to stabilise around the 0.96 quantile.

The winter of 2009/10 represents a disproportionately large percentage of all extreme events. We have over 73 years of data, with almost 17% of extreme hourly observations occurring in the winter of 2009/10, while, on average, each year has 1.36% of all extreme events. It is clear the behaviour of extremes are different

in 2009/10 from those in other years. For example, the temperature at the 0.9 quantile of minimum daily temperatures in 2009/10 is approximately twice as cold (in the Celsius scale) as the average yearly temperature at the 0.9 quantile. This anomalous period certainly has a substantial impact on the tail of the entire data distribution. Despite this, extreme observations above the 0.96 quantile of temperatures are evenly spread across space and time, with the exclusion of the winter 2009/10. Even at this unusually high threshold, we believe we have sufficient extreme observations to achieve our modelling goal of allowing the scale parameter to capture the long-term non-stationarity in a parsimonious manner.

To alleviate the impact of the winter of 2009/10 on the estimation of the threshold, we could use a different threshold estimate during 2009/10 from other years. However, a step function at the threshold greatly reduces the interpretability and parsimony of our model. Instead, we rely on more physically interpretable covariates (described in Section 5.3.3.2, such as $M^I_{r,m}(t)$) which capture SCV, and directly reflect the unusual behaviour of the especially extreme behaviour 2009/10. We could incorporate this covariate into our threshold estimation. It is a non-trivial choice as to whether to deal with this complex non-stationarity in the threshold or GPD parameters (see Section 4.7.2 for a discussion of this issue). Since we have data evenly spread across all periods at our high threshold and to preserve the parsimony of the model, we choose to keep a temporally constant threshold, and account for the highly variable behaviour in the temperature process by incorporating the SCV covariate into the scale parameter of the GPD.

An alternative approach to be explored in future work, is to use an extended generalised Pareto distribution (eGPD) based models (Papastathopoulos & Tawn 2013). The eGPD allows for a lower threshold selection, through the inclusion of an additional shape parameter. In this way, we can include more extreme observations in our analysis. Furthermore, recent developments by Naveau et al. (2016) extend the methodology of (Papastathopoulos & Tawn 2013), to eliminate the specification of a threshold, by allowing a smooth transition between the bulk and tail. These methods maintain the generalised Pareto distribution in the tail of the distribution.

Regardless, we are conscious of the fact that this is quite a high threshold (given only three months of data per year) and so to ascertain the sensitivity of this choice, we perform analysis based on the three quantiles marked as vertical lines in Figure 5.9,

i.e., we use the $0.95, 0.96$, and 0.97 quantiles as thresholds.

Initially, following our best-performing analysis procedure for modelling maximum daily summer temperatures found in Chapter 4, we chose model *c.* in Table 5.1 as the threshold model. Estimates of the spatial threshold, $u_o^{(\tau)}(\boldsymbol{s})$, for each $\tau \in \{0.95, 0.96, 0.97\}$ can be seen in the top row of Figure 5.10, where columns left to right correspond to the increasing values of τ. However, we found that model *c.* generally over-estimated lower quantiles and under-estimated higher quantiles. Through exploratory analysis, we found this was largely due to coastal bias. The inclusion of a coastal proximity covariate helped correct this bias. Finally, we chose model *c.* with the inclusion of coastal proximity covariate $C(\boldsymbol{s})$ as our threshold model. Note that this is equivalent to model *e.* and *f.*, with no temporal non-stationarity. That is, the estimate of the spatial threshold is calculated as

$$u_o^{(\tau)}(\boldsymbol{s}) = \beta_0^{(\tau)} + \beta_1^{(\tau)} q_c(\boldsymbol{s}) + \beta_2^{(\tau)} C(\boldsymbol{s}), \tag{5.2}$$

where $\tau = 0.95, 0.96$, or 0.97. Estimates of $u_o^{(\tau)}(\boldsymbol{s})$ using (5.2) for each $\tau \in \{0.95, 0.96, 0.97\}$ can be seen in the bottom row of Figure 5.10, where again, columns left to right correspond to increasing values of τ. Figure 5.10 shows that inland areas of Ireland experience cooler winter temperatures.

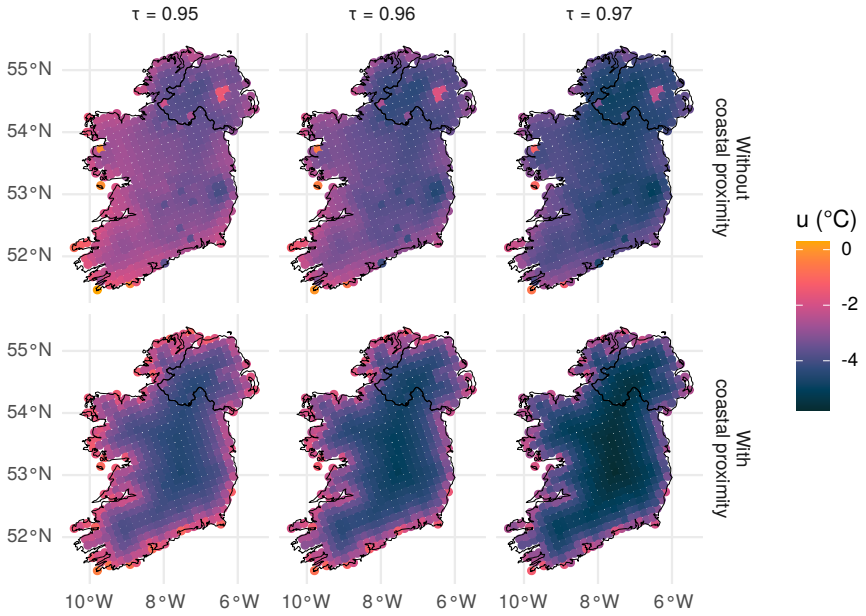

Figure 5.10: Estimates of threshold $u_o^{(\tau)}(\boldsymbol{s})$ without accounting for coastal proximity with values of $\tau = 0.95, 0.96, 0.97$ from left to right (top row). Equivalent to the corresponding figures above, with the inclusion of coastal proximity as a covariate (bottom row).

In Figure 5.11, we calculate estimates of the threshold exceedance probability, λ, associated with low, median, and high phases of SCV and present its spatial average $\lambda_o(t) = \int_{\boldsymbol{s}\in\mathcal{S}_o} \lambda_o(t,\boldsymbol{s})d\boldsymbol{s}/|\mathcal{S}_o|$. The presented estimates of $\lambda_o(t)$ show a decreasing exceedance rate over the period 1950–2022, for each phase of SCV (low to high from left to right). For the median phase of SCV, i.e., $M_{r,m}^{(0.5),I}(t)$, in the centre plot, the threshold exceedance rate over time of $0.04 - 0.015$ reflects the choice of the threshold as 0.96. During the low phase of SCV, i.e., $M_{r,m}^{(0.1),I}(t)$, in the left-hand plot, we have much smaller threshold exceedance probabilities, indicating a much lower occurrence rate of extremely cold winter temperatures. In stark contrast, during the high phase of SCV, i.e., $M_{r,m}^{(0.9),I}(t)$, in the right-hand plot, we see a much higher occurrence of extremely cold winter temperatures. During these high phases, we experience winters such as those in 2009/10.

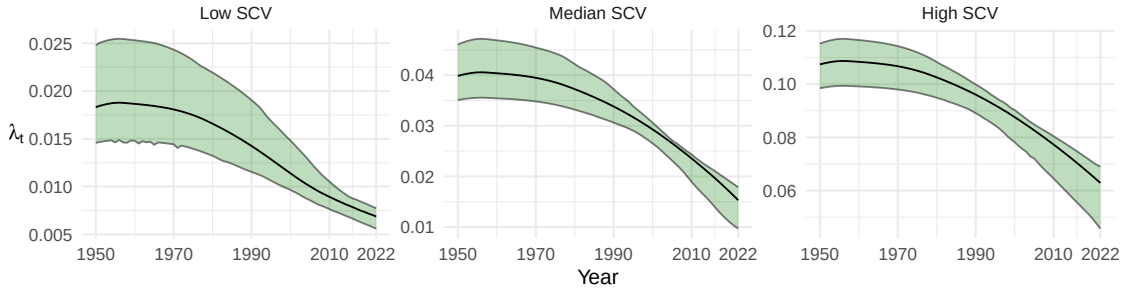

Figure 5.11: Threshold exceedance probability $\lambda_o(t)$ for t from 1950 to 2022 for model E2 averaged over $\mathcal{S}$, corresponding to low (left), median (centre) and high phases (right) of SCV. In each case, the shaded region indicates bootstrap-based pointwise 96% confidence intervals.

5.3.3.4 Tail model

For each site $\boldsymbol{s}$, we assume that excesses of the threshold $u(\boldsymbol{s})$ follow a generalised Pareto distribution (GPD), with distribution function

$$H(y;\sigma,\xi) = 1 - (1+\xi y/\sigma)_+^{-1/\xi}$$

for $y > 0$, with a shape parameter $\xi \in \mathbb{R}$ and a scale parameter $\sigma > 0$, with the notation $x_+ = \max(x, 0)$, and $\xi = 0$ is obtained by taking the limit as $\xi \to 0$. When $X(t, \boldsymbol{s}) > u(\boldsymbol{s})$ the threshold excess, $Y(t, \boldsymbol{s}) = X(t, \boldsymbol{s}) - u(\boldsymbol{s})$, is taken to be distributed as

$$Y(t, \boldsymbol{s}) \sim \text{GPD}(\sigma(t, \boldsymbol{s}), \xi).$$

We investigated numerous choices of models for $\sigma(t, \boldsymbol{s})$, broadly grouped into seven as presented in Table 5.2. In all models, we take the shape parameter to be constant over space and time for each model. We perform a log-likelihood ratio test at each site to determine whether allowing the shape parameter to vary over space provides a significant improvement in the model fit. We take M_N, the null hypothesis, to be the model with constant shape parameter and M_A, the alternative, where each site has a different shape parameter estimate. For each site we compare the maximised pseudo-likelihood of M_N (i.e., $p\mathcal{L}\{M_N(\mathbf{s})\}$) and M_A, (i.e., $p\mathcal{L}\{M_A(\mathbf{s})\}$) and calculate the pseudo-likelihood ratio test statistic as $2\ln[p\mathcal{L}\{M_A(\mathbf{s})\}/p\mathcal{L}\{M_N(\mathbf{s})\}]$. We found that over 91% of sites had a likelihood ratio below 0.05, with 95% of sites

below 0.06. The total proportion of data in the sites with a test statistic greater than 0.05 is about 2.5% of extremal observations substantially below the standard likelihood ratio test statistic of 3.84. Furthermore, there was no clear spatial pattern in those sites with a test statistic higher than 0.05. Thus, we choose to keep a constant shape parameter.

As discussed in Section 5.1, extreme Irish winter minima are highly variable and strongly associated with large-scale climatic oscillations. We explore a set of potential covariates to explain the variability unaccounted for by $M^I(t)$. This extra covariate is denoted as $z(t)$ in Table 5.2.

Group	**Model structure for** $\ln \sigma_o$
A	$\beta_0 + \beta_1 M^I(t) + \beta_2 z(t)$
B	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 z(t)$
C	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 C(\boldsymbol{s}) + \beta_4 z(t)$
D	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 z(t)$
E	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 z(t)$
F	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_3 z(t)$
G	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 z(t)$
H	$\{\beta_0,\ \beta_0 + \beta_1 \sigma_c(\boldsymbol{s}),\ \beta_0 + \beta_1 C(\boldsymbol{s})\}$,

Table 5.2: GPD model groups.

The covariates $z(t)$ capture short-term climatic variability (SCV) and are detailed in Section 5.2.3.1. In each group in Table 5.2, $z(t)$ can also be the empty-set $\varnothing$, so in each model group, we excluded the covariate $z(t)$ and explore the assumption that there is no climate variability unaccounted for by a global trend. The set of covariates we present cross-validation results for is $z(t) = \{0,\ M^I_{r,m}(t),\ M^I_r(t),\ \mathrm{NAO}(t),\ \mathrm{NAO}_r(t),\ \mathrm{NAO}_m(t),\ \mathrm{NAO}_{r,m}(t),\ \mathrm{AO}(t),\ \mathrm{AO}_r(t),\ \mathrm{AO}_m(t),\ \mathrm{AO}_{r,m}(t),\ M^G_r(t),\ M^N_r(t)\}$, totalling 94 models which can be seen in Table 5.3. The covariates $\mathrm{NAO}(t)$ and $\mathrm{AO}(t)$, refer to the North Atlantic and Arctic oscillation values, respectively; $M^G_r(t)$ refers to the average HadCRUT5 values over the grid points surrounding Ireland, excluding the grid point over Ireland, and $M^N_r(t)$ refers to the average HadCRUT5 values over the grid points above Ireland. In each case, the subscript r refers to the residuals of that covariate after the trend of $M^I(t)$ is removed. A subscript m indicates monthly values, whereas no subscript m indicates that the average value over the winter months was taken for each year. We also investigated monthly values of $M^N_r(t)$ and $M^G_r(t)$. However, as compared to $M^I_r(t)$, their performance was

generally worse in each case, so we do not include them for brevity. The spatial covariate $\sigma_c(\boldsymbol{s})$ represents the scale parameter of the GPD, fitted to $X_c(t, s)$, the climate model from the grid point closest to site $\boldsymbol{s}$. The covariate $\sigma_c(\boldsymbol{s})$ for each threshold $u_c^{(0.95)}(\boldsymbol{s})$,$u_c^{(0.96)}(\boldsymbol{s})$, and $u_c^{(0.97)}(\boldsymbol{s})$ estimated using climate model data is shown in Figure 5.12.

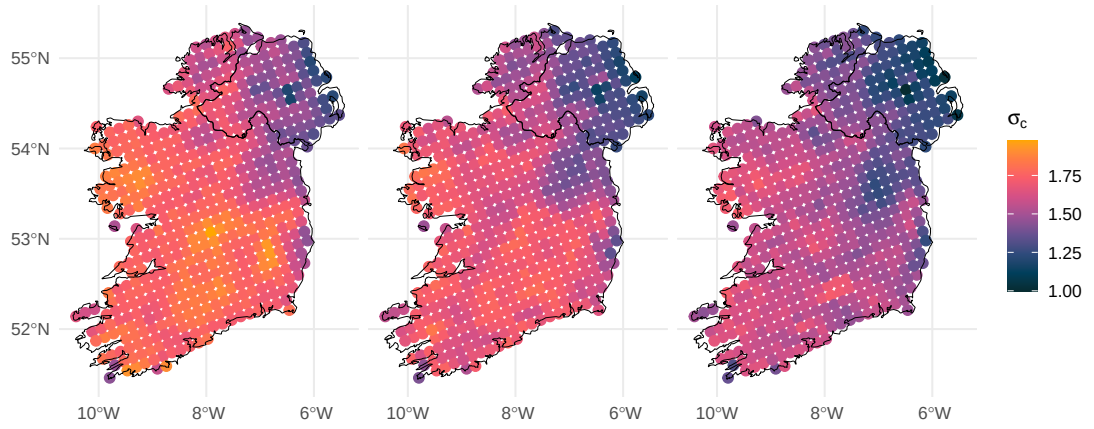

Figure 5.12: Climate scale covariate $\sigma_c(\boldsymbol{s})$ associated thresholds $u_c^{(0.95)}(\boldsymbol{s})$ (left), $u_c^{(0.96)}(\boldsymbol{s})$ (centre) and $u_c^{(0.97)}(\boldsymbol{s})$ (right) each with associated shape parameters $\xi = -0.12, -0.11$, and -0.09 respectively.

5.3.3.5 Model selection

In Table 5.2, Group A assumes spatial stationarity in the scale parameter. Groups B and C borrow spatial information from a climate model output similar to the procedure used in Chapter 4, where C also explores a coastal proximity effect. Group D is equivalent to C but explores the interaction between the coast and long-term temporal trends. Groups E and F do not borrow information from the climate model output, instead relying on coastal proximity to describe all spatial non-stationarity. Group G has no main spatial term, instead, it explores an interaction in space and time. We also consider some miscellaneous "base" models, which we will call group H. Group H consists of three models, model H1 has no spatial or temporal non-stationarity. Model H2 incorporates $\sigma_c(\boldsymbol{s})$ as a spatial covariate. Surprisingly, this yields no improvement over the spatially stationary model H1, suggesting the inefficacy of $\sigma_c(\boldsymbol{s})$ for informing the spatial nature of extreme cold temperatures.

However, the incorporation of the coastal proximity covariate $C(\boldsymbol{s})$, in model H3, achieves an improved cross-validation score over the spatially stationary model H1. Surprisingly, the chosen climate model output is not as effective as a spatial covariate for modelling extremely cold winter events as we found it to be for extremely hot summer temperatures in Chapter 4.

For the cross-validation metrics presented in Table 5.3, empirical quantiles used for calculating RMSE were estimated on yearly blocks of data for each site (in order to have sufficient data for estimating reasonable empirical quantiles, as well as preserving temporal non-stationarity). Therefore we do not report RMSE associated with models that have monthly non-stationary as their interpretation is not helpful. The continuous ranked probability score (CRPS, Gneiting & Katzfuss 2014) does not require the specification of an observation quantile to assess the model performance. We thus use RMSE and CRPS jointly to identify the best-performing covariate. Following this we use CRPS to choose whether to take a given covariate on a monthly or yearly scale.

The best models (highlighted in Table 5.3) all favour using HadCRUT5 residuals, $M_r^I(t)$, as a covariate for explaining SCV. This is perhaps unsurprising as all climatic forcings are encoded within HadCRUT5, and most informatively, their direct impact on temperature processes. The next most successful and informative covariates $z(t)$ were $\{\text{AO}_r, \text{NAO}_r$ and, $M_r^G\}$. Furthermore, each of the best-performing models used coastal proximity as a covariate.

In each pair of best-performing models, within each model group, CRPS prefers the monthly covariate $M_{r,m}^I(t)$ (highlighted in red in Table 5.3) rather than the yearly averaged $M_r^I(t)$ (highlighted in blue). This narrows the selection down to models C2, E2, and F2. Models C2 and F2 are equivalent, only for their main spatial covariate, with C2 using $\sigma_c(\boldsymbol{s})$ and F2 using $C(\boldsymbol{s})$. Model F2 outperforms model C2 in n-fold CV, while model C2 outperforms model F2 in spatio-temporal CV by the same very small margin. In both models F2 and C2, the coefficient β_3 of the spatio-temporal interaction between $C(\boldsymbol{s})$ and $M^I(t)$ was not statistically significant with 95% confidence bounds containing 0. In this regard, we prefer model E2 over F2, since E2 is a nested model of F2, simply without the spatio-temporal interaction. We note that model F2 gives a very slight improvement in n-fold CRPS CV yet no improvement in spatio-temporal CV. Comparing models C2

and E2, both with very similar CV metrics and considering the earlier discussion of base model H3 outperforming H2, we prefer model E2. Finally, we choose model E2 for its parsimony, predictive and modelling performance, and ease of interpretation.

To summarise, our chosen marginal GPD model E2 uses coastal proximity, $C(\boldsymbol{s})$, as a spatial covariate, smoothed temperature anomalies over Ireland, $M^I(t)$, as a long-term temporal covariate, and monthly residuals from smoothed covariate, $M^I_{r,m}(t)$, as a SVC covariate, capturing unusually cold winters. To assess the absolute quality of the chosen model E2, we create a pooled QQ-plot in the right-hand panel of Figure 5.13, pooling over all sites and years. We transform the data through our fitted model into a common uniform scale and to a common exponential scale. The exponential scale accentuates the upper tail of the data and highlights the model's performance in capturing the extreme values. We see evidence of a good fit, with values near the lines of equality, and in the far upper tail, all values falling within the pointwise tolerance bounds. The left-hand plot of the same figure presents the equivalent plot, but using model E1 (equivalent to model E2 but with no covariate $z(t)$ to capture SCV). This clearly highlights how not accounting for climate variability in extreme winter temperatures leads to a poor representation of the data in the upper tail. This leads to model E1 having a heavier tail ($\hat{\xi} = 0.011$) than our preferred model E2 ($\hat{\xi} = -0.079$) and thus, overestimates the heaviness of the tail of the distribution.

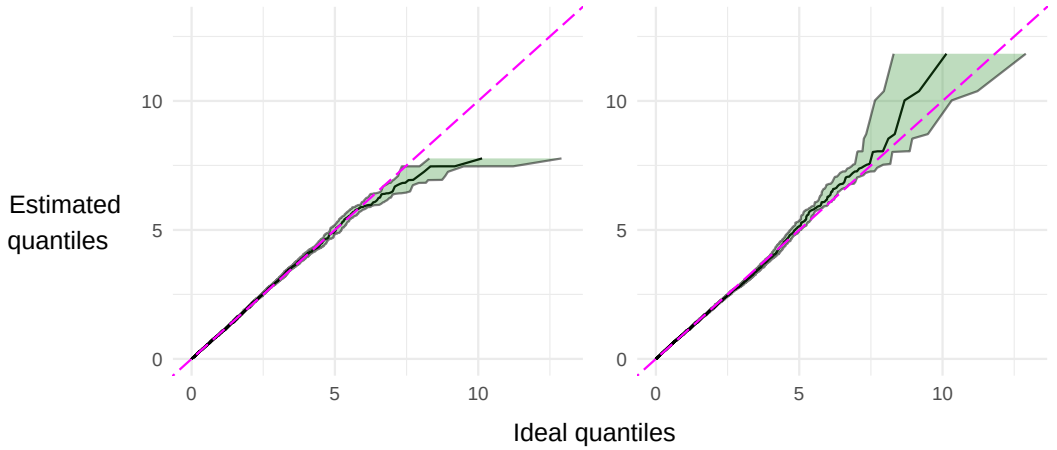

Figure 5.13: QQ-plot of two models on exponential margins to accentuate the behaviour of the tail. Model E1 without climate oscillation covariate (left), final chosen model E2 with climate oscillation covariate (right)

Figure 5.14 plots the scale parameter, σ, estimated via model E2. This figure shows that the most variable excess distribution is inland, with a decay in values moving towards the coast, highlighting the ocean's regulatory effect. Furthermore, we see the largest estimated scale parameter at higher levels of SCV. The estimated change in the scale parameter over the observation period, denoted $\nabla\sigma_o(\boldsymbol{s})$, during low, median, and high phases of SCV, is plotted on the bottom row of the same figure. The scale parameter has decreased overall levels of SCV with the largest decrease during phases of high SCV. Recall that we are modelling the maximum of negative minimum temperatures (see Equation (5.1)), so the decreasing scale parameter suggests a warming of extreme minimum winter temperatures, with the greatest warming observed during high levels of SCV.

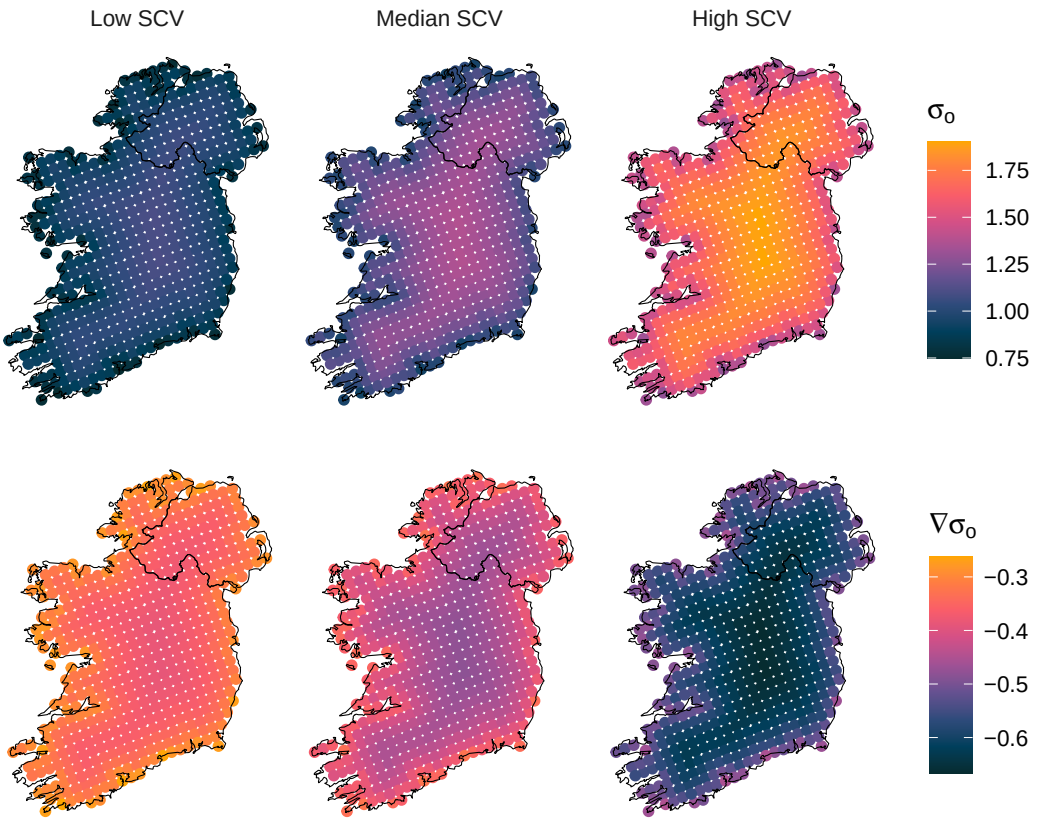

Figure 5.14: Estimated GPD scale parameter according to model E2 in 2022 during low, median, and high phases of SCV (top row, left to right respectively). Estimated change in the scale parameter since 1950, $\nabla\sigma_o(\boldsymbol{s})$ with respect to scale estimates during low, median, and high phases of SCV as plotted above (bottom row, left to right respectively).

Table 5.3: Cross-validation metrics for each GPD model.

Model		Parameterisation of $\ln \sigma_o$	ST-CV		n-CV	
			RMSE	CRPS	RMSE	CRPS
A	1	$\beta_0 + \beta_1 M^I(t)$	1.223	0.886	1.350	0.946
	2	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I_{r,m}(t)$	-	0.824	-	0.880
	3	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I_r(t)$	1.106	0.836	1.198	0.895
	4	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^G_r(t)$	1.122	0.842	1.221	0.901
	5	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^G_r(t)$	1.128	0.848	1.224	0.906
	6	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{AO}(t)$	1.112	0.842	1.208	0.901
	7	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{AO}_r(t)$	1.112	0.842	1.208	0.901
	8	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{AO}_m(t)$	-	0.862	-	0.918
	9	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{AO}_{r,m}(t)$	-	0.862	-	0.918
	10	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{NAO}(t)$	1.121	0.843	1.220	0.904
	11	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{NAO}_r(t)$	1.121	0.843	1.220	0.904
	12	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{NAO}_m(t)$	-	0.845	-	0.902
	13	$\beta_0 + \beta_1 M^I(t) + \beta_2 \mathrm{NAO}_{r,m}(t)$	-	0.845	-	0.902
B	1	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t)$	1.225	0.887	1.349	0.946
	2	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I_{r,m}(t)$	-	0.823	-	0.880
	3	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I_r(t)$	1.110	0.836	1.199	0.895
	4	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^G_r(t)$	1.126	0.843	1.222	0.901
	5	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^G_r(t)$	1.130	0.848	1.224	0.906
	6	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{AO}(t)$	1.116	0.842	1.208	0.901
	7	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{AO}_r(t)$	1.115	0.842	1.208	0.901
	8	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{AO}_m(t)$	-	0.862	-	0.918
	9	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{AO}_{r,m}(t)$	-	0.862	-	0.918
	10	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{NAO}(t)$	1.125	0.844	1.220	0.903
	11	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{NAO}_r(t)$	1.125	0.843	1.220	0.903
	12	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{NAO}_m(t)$	-	0.845	-	0.902
	13	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \mathrm{NAO}_{r,m}(t)$	-	0.845	-	0.902
C	1	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s})$	1.189	0.879	1.320	0.939
	2	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 M^I_{r,m}(t)$	-	0.815	-	0.874
	3	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 M^I_r(t)$	1.064	0.828	1.166	0.889
	4	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 M^G_r(t)$	1.083	0.835	1.188	0.895
	5	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 M^G_r(t)$	1.087	0.840	1.191	0.899
	6	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{AO}(t)$	1.075	0.834	1.176	0.895
	7	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{AO}_r(t)$	1.073	0.834	1.177	0.896
	8	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{AO}_m(t)$	-	0.855	-	0.911
	9	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{AO}_{r,m}(t)$	-	0.855	-	0.911
	10	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{NAO}(t)$	1.085	0.838	1.186	0.898
	11	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{NAO}_r(t)$	1.077	0.835	1.187	0.898
	12	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t) C(\boldsymbol{s}) + \beta_4 \mathrm{NAO}_m(t)$	-	0.838	-	0.896

	13	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_{r,m}(t)$	-	0.838	-	0.896
D	1	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s})$	1.225	0.887	1.349	0.945
	2	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^I_{r,m}(t)$	-	0.824	-	0.880
	3	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^I_r(t)$	1.112	0.837	1.200	0.895
	4	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^G_r(t)$	1.125	0.844	1.222	0.902
	5	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^G_r(t)$	1.131	0.849	1.224	0.906
	6	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}(t)$	1.117	0.842	1.208	0.901
	7	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}_r(t)$	1.116	0.842	1.208	0.901
	8	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}_m(t)$	-	0.863	-	0.918
	9	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}_{r,m}(t)$	-	0.863	-	0.918
	10	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}(t)$	1.123	0.843	1.219	0.904
	11	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_r(t)$	1.122	0.844	1.222	0.904
	12	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_m(t)$	-	0.845	-	0.902
	13	$\beta_0 + \beta_1\sigma_c(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_{r,m}(t)$	-	0.844	-	0.902
E	1	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t)$	1.189	0.878	1.320	0.939
	2	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I_{r,m}(t)$	-	0.816	-	0.874
	3	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I_r(t)$	1.065	0.829	1.163	0.888
	4	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^G_r(t)$	1.080	0.834	1.187	0.895
	5	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^G_r(t)$	1.086	0.840	1.190	0.899
	6	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{AO}(t)$	1.073	0.834	1.176	0.895
	7	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{AO}_r(t)$	1.073	0.834	1.176	0.895
	8	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{AO}_m(t)$	-	0.855	-	0.912
	9	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{AO}_{r,m}(t)$	-	0.855	-	0.911
	10	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{NAO}(t)$	1.081	0.836	1.187	0.898
	11	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{NAO}_r(t)$	1.081	0.836	1.187	0.898
	12	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{NAO}_m(t)$	-	0.837	-	0.896
	13	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 \text{NAO}_{r,m}(t)$	-	0.837	-	0.896
F	1	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s})$	1.187	0.878	1.316	0.938
	2	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^I_{r,m}(t)$	-	0.816	-	0.873
	3	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^I_r(t)$	1.063	0.828	1.159	0.888
	4	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^G_r(t)$	1.079	0.835	1.181	0.894
	5	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 M^G_r(t)$	1.083	0.840	1.183	0.898
	6	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}(t)$	1.072	0.835	1.172	0.895
	7	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}_r(t)$	1.069	0.833	1.171	0.895
	8	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}_m(t)$	-	0.854	-	0.911
	9	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{AO}_{r,m}(t)$	-	0.854	-	0.911
	10	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}(t)$	1.081	0.836	1.185	0.897
	11	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_r(t)$	1.082	0.836	1.184	0.897
	12	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_m(t)$	-	0.837	-	0.895
	13	$\beta_0 + \beta_1 C(\boldsymbol{s}) + \beta_2 M^I(t) + \beta_3 M^I(t)C(\boldsymbol{s}) + \beta_4 \text{NAO}_{r,m}(t)$	-	0.837	-	0.895
G	1	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t)C(\boldsymbol{s})$	1.224	0.886	1.349	0.945
	2	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t)C(\boldsymbol{s}) + \beta_4 M^I_{r,m}(t)$	-	0.824	-	0.880

	3	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 M_r^I(t)$	1.107	0.836	1.198	0.895
	4	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 M_r^G(t)$	1.122	0.843	1.221	0.901
	5	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 M_r^G(t)$	1.129	0.849	1.224	0.906
	6	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{AO}(t)$	1.113	0.842	1.208	0.901
	7	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{AO}_r(t)$	1.113	0.842	1.208	0.901
	8	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{AO}_m(t)$	-	0.862	-	0.918
	9	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{AO}_{r,m}(t)$	-	0.862	-	0.918
	10	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{NAO}(t)$	1.124	0.844	1.220	0.904
	11	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{NAO}_r(t)$	1.122	0.844	1.221	0.904
	12	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{NAO}_m(t)$	-	0.845	-	0.902
	13	$\beta_0 + \beta_1 M^I(t) + \beta_2 M^I(t) C(\boldsymbol{s}) + \beta_4 \text{NAO}_{r,m}(t)$	-	0.845	-	0.902
H	1	β_0	1.224	0.885	1.354	0.945
	2	$\beta_0 + \beta_1 \sigma_c(\boldsymbol{s})$	1.225	0.885	1.354	0.945
	3	$\beta_0 + \beta_1 C(\boldsymbol{s})$	1.189	0.878	1.322	0.939

5.3.3.6 Bootstrapping procedure

For our analysis, we rely on bootstrapping to characterise and carry uncertainty in our modelling assumptions throughout the entire analysis, from the marginal bulk and tail model to the r-Pareto simulations. In Section 4.3.4.1 we present a bootstrapping algorithm designed to preserve spatial and short-term temporal dependence in maximum daily summer temperature data, while also matching missingness patterns in the observation data. This was achieved through vector temporal block bootstrapping of the observed data after being transformed to be uniformly distributed, $X_o^U(t, \boldsymbol{s}) = F\{X_o(t, \boldsymbol{s})\}$. The resampled data are then transformed back to the data scale, giving a bootstrapped sample data set, $X_o^*(t, \boldsymbol{s})$, generated under the chosen marginal model. This bootstrapping procedure was not immediately applicable to our winter data set for two reasons. Firstly, extremely cold winter events have substantially stronger temporal dependence than hot summer temperatures. Secondly, we found that temporal trends in cold winter extremes were much more variable and more complex in terms of dependence on climate indicators than in hot summer extremes. Here we describe modifications to our bootstrapping algorithm to account for these features and capture them in our bootstrap samples.

Extreme cold winter events have much stronger temporal dependence compared to hot summer events. Solar insolation is lowest in Ireland during winter months and so factors such as clouds tend to have much lower local impact leading to more spatially

and temporally smooth structures (Liou 2002). Through exploratory analysis, we found that the expected duration of an extreme cold event at any site in our data set is on average approximately 20 days, see Figure 5.15. Temporally de-clustering extreme events at a single site would create independent observations and remove the need to perform block bootstrapping. However, in a spatio-temporal setting, de-clustering each site separately would greatly increase the permutations of observed sites across each day, greatly reducing the number of possible spatial matches for resampling.

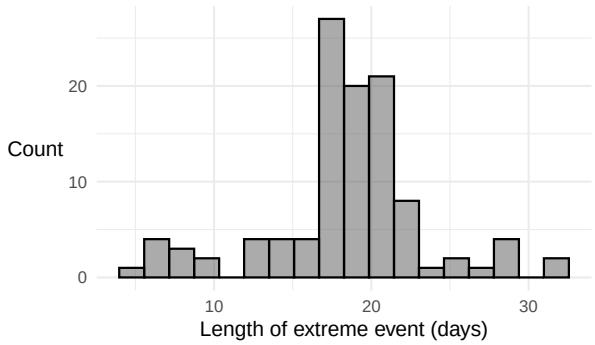

Figure 5.15: The mean across all sites of the length of extreme events in days. The duration of a single extreme event is the number of consecutive extreme observations that occur within a seven-day period of another at an individual site.

Instead, we summarise all days in the data set by the minimum temperature observed across sites on that day, creating a time series of $\min_{\boldsymbol{s}\in\mathcal{S}}\{X(t,s)\}$, as shown in Figure 5.16. We considered two approaches using this series. By temporally de-clustering this time series, taking local minima within a ± 10 day range, and keeping all data within the de-clustered dates gives approximately independent observations without disrupting the spatial dimension. However, this risks under-representing areas that experience less severe extremes while also being wasteful of data. Instead, we create spatio-temporal blocks on the original data, centred on the de-clustered local minima of $\min_{\boldsymbol{s}\in\mathcal{S}}\{X(t,s)\}$, extending to the midpoint between the previous and subsequent local minima. These blocks are assumed to be independent and capture the short-range temporal dependence in the data. We found the best results when we extended the range for declustering to a ± 15 day range, allowing us to capture the majority of the events with very strong autocorrelation while being small enough to allow us to define a suitably large number of blocks for resampling.

The de-clustered local minima are highlighted in red in Figure 5.16.

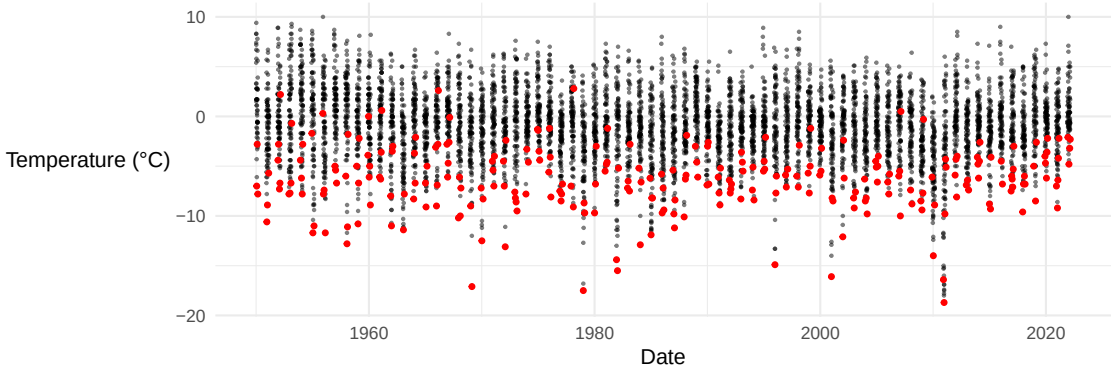

Figure 5.16: Time series of $\min_{\boldsymbol{s}\in\mathcal{S}}\{X(t,s)\}$ over the period 1950–2022 with de-clustered local minima (in $\pm$15 days) highlighted in red.

A key issue with the bootstrapping procedure presented in Chapter 4 is that it assumes the model is correct and that $F\{X_o(t,\boldsymbol{s})\}$ is uniformly distributed. Assuming that the model captures all temporal non-stationarity within the data is certainly unrealistic when the non-stationarity is complex, highly variable, and the choice of the best-performing model is not obvious. Our aim is to capture and accurately describe the overarching trend of winter extremes in Ireland, disentangling the long-term trend from complex climatic variations. Furthermore, we found that the most extreme events (such as winter 2009/10, as discussed in Section 5.1) greatly affected extremal inference on that bootstrapped data set. Furthermore, as is shown in Figure 5.2, the winter of 2009/2010 constitutes a disproportionately large portion of extremal observations and hence is more likely to be resampled. With the aim of preserving the global trend of the data, in the context of numerous and highly variable models, while accurately representing expected large variation, we restrict resampling within windows of $\pm$10 years. That is, we accept a resample if it is within a 10-year window of the data being replaced.

When resampling, we initially only accepted resampled blocks if they had the exact spatio-temporal pattern as the block being replaced as a subset of its data. However, given our set of fixed and predefined blocks, this led to relatively few suitable choices available to resample which in turn led to little variation in the bootstrapped data sets. To deal with this, we experimented with a "relaxed" resampling algorithm for spatial pattern matching. We accepted a resampled block if it had a high

proportion of the sites in the block it was replacing. We found that we had sufficient variation when accepting blocks that had 90% of the sites being replaced. To preserve missingness patterns in the data we discarded observations from sites in the resampled block that are not in the block that was being replaced. If the resampled block was of a longer duration than the block being replaced we took a subset of the resampled block, randomly.

5.4 Multivariate dependence models

5.4.1 Overview and strategy

We investigate the level of extremal dependence of extreme minimum daily winter temperature data, following the process as described in Section 4.4.1. We first standardise the marginal variables so that they have an identical distribution over variables and covariates such that we can fairly investigate extremal dependence. We first transform the data to uniform margins using the probability integral transform, and following this, we transform to standard Pareto using

$$X^P(t, \boldsymbol{s}) = \frac{1}{1 - F_{t,\boldsymbol{s}}\{X(t, \boldsymbol{s})\}}, \text{ for all } \boldsymbol{s} \in \mathcal{S} \text{ and all } t, \tag{5.3}$$

where F is the estimated distribution function of $X(t, \boldsymbol{s})$ with quantile regression bulk model below $u(\boldsymbol{s})$ and GPD tail model above. Following this, we fit a generalised r-Pareto process to extreme spatial observations as decided via a cost function to model the extremal dependence. Asymptotic dependence is a prerequisite for fitting the r-Pareto model and so we test if the data satisfies this requirement. A discussion of the extremal dependence of the data is shown in Section 5.4.2. After fitting the r-Pareto process (covered in Section 5.4.3) we generate spatial simulations under different phases of SCV.

5.4.2 Evidence for asymptotic dependence

The pairwise coefficient of asymptotic dependence, χ, measures the probability of a process at two sites being jointly extreme (Coles et al. 1999). Recall, for the process

X^P at sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, $\chi = \chi^P(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is defined by

$$\chi^P(\boldsymbol{s}_i, \boldsymbol{s}_j) = \lim_{v \to \infty} \Pr(X^P(\boldsymbol{s}_j) > v \big| X^P(\boldsymbol{s}_i) > v)$$

If $\chi^P(\boldsymbol{s}_i, \boldsymbol{s}_j) > 0$ (or equals 0) then process X^P is asymptotically dependent (or asymptotically independent) at this pair of sites respectively. The value of χ ($0 < \chi \leq 1$) determines the degree of asymptotic dependence, with χ increasing as this dependence strengthens.

We define an empirical estimator of the tail coefficient of processes X_1 and X_2 at quantile u as

$$\chi_u(X_1, X_2) = \frac{P[F_{X_1}(X_1) > u, F_{X_2}(X_2) > u]}{P[F_{X_2}(X_2) > u]}, \tag{5.4}$$

where F_{X_i} denotes the distribution function of process X_i. Empirical estimates of χ are plotted against the distance between sites for all pairs of sites for both the observational data, X_o^P, and climate model data X_c^P in Figure 5.17. This figure suggests that the climate model output overestimates extremal dependence in observational winter temperatures. The empirically estimated coefficient of extremal dependence for both processes remains non-zero, even at considerable distances and increasing quantiles with the magnitude of χ changes minimally as quantiles increase. Our analysis suggests that the assumption of asymptotic dependence in the minimum daily temperature in Ireland is reasonable.

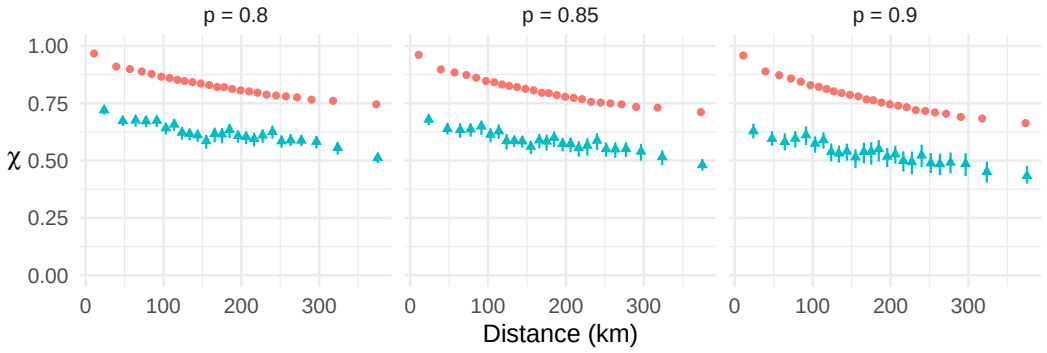

Figure 5.17: Estimates of χ plotted against inter-site distance for observational data (blue), standardised using marginal model E2, and climate model output (orange) for $u = 0.8, 0.85$, and 0.9 with 95% confidence intervals plotted as vertical lines.

5.4.3 r-Pareto process

After standardising the data to unit Pareto margins, we model the extreme dependence of the process $X^P(t, \boldsymbol{s})$ over $\boldsymbol{s} \in \mathcal{S}$ with an r-Pareto process. For a detailed treatment of r-Pareto process theory see Section 4.4.3. To briefly recall, we define a risk function to provide a measure and ordering to levels of extremity of spatial observations, $r(X^P) := r\{X^P(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\} \in [0, \infty)$, where r must be homogeneous of order 1. Dombry & Ribatet (2015) show that under weak conditions on the process X^P that

$$\Pr\left\{v^{-1}X^P(\boldsymbol{s} : \boldsymbol{s} \in \mathcal{S}) \in \cdot \mid r(X^P(\boldsymbol{s} : \boldsymbol{s} \in \mathcal{S})) > v\right\} \to \Pr\{Y_r^P(\boldsymbol{s} : \boldsymbol{s} \in \mathcal{S}) \in \cdot\}, \tag{5.5}$$

as $v \to \infty$, where $\{Y_r^P(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ is marginally non-degenerate in any margins. Here, Y_r^P is called the r-Pareto process. If the limit provided in Equation (5.5) is a good approximation for large v, then those spatial events with a risk function exceeding v will be well-approximated by an r-Pareto process. Crucially, Y_r^P can be decomposed into two independent stochastic components as follows:

$$Y_r^P(\boldsymbol{s}) = RW(\boldsymbol{s}) \text{ for all } \boldsymbol{s} \in \mathcal{S} \tag{5.6}$$

where R is Pareto distributed and can be interpreted as the risk of the process, and $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ is a stochastic process which describes the spatial profile of the extreme event, i.e., the proportion of the risk at each site. By construction, we have that $R = r(Y_r^P(\boldsymbol{s} : \boldsymbol{s} \in \mathcal{S}))$ and $r(W(\boldsymbol{s} : \boldsymbol{s} \in \mathcal{S})) = 1$, where $R \sim \text{Pareto}(1)$ can be interpreted as the magnitude or cost of the process, i.e., $R = r(Y)$. The spatial process of Y is then W. This characterisation is powerful as it allows extrapolation to events larger than those previously observed.

Similar to the analysis in Chapter 4, we take the cost function to be

$$r_t(X_o^P(t, \boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}) = \frac{\sum_{i=1}^d X_o^P(t, \boldsymbol{s}_i) I_o(t, \boldsymbol{s}_i)}{\sum_{i=1}^d I_o(t, \boldsymbol{s}_i)},$$

where $I_o(t, \boldsymbol{s}_i)$ is the indicator variable for whether $X_o(t, \boldsymbol{s}_i)$ is observed or not, d is the number of sampling locations selected for the risk function evaluation using sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$. Thus, the risk function r_t is the average of standardised variables over the d sites which were observed at time t, and so is invariant to the changing

dimension of the partially observed event.

Expression (5.6) allows us to model the spatial profile of extremal dependence separately from the magnitude. Furthermore, the choice of a parametric spatial process W also affords great flexibility. Following de Fondeville & Davison (2018) and Palacios-Rodríguez et al. (2020), we choose to model W using the log-Gaussian stochastic process (also referred to as the Brown–Resnick process in the max-stable literature). Modelling the angular component W of the r-Pareto process via log-Gaussian processes affords us the ability to make use of classical geostatistical methods. Engelke et al. (2015) note that the finite-dimensional marginal distribution of the log-Gaussian process at sites $s_0, s_1, \ldots, s_k$ is the Hüsler-Reiss distribution (Hüsler & Reiss 1989) with spectral density h and parameter matrix $\Lambda = (\lambda^2_{i,j})_{0\leq i,j\leq k} = (\gamma(s_i - s_j)/2)_{0\leq i,j\leq k}$. So, we can express h in terms of this variogram γ. We use the Matérn variogram family

$$\gamma_{\text{mat}}(h;t) = \alpha\left\{1 - (2\sqrt{\nu}h/\phi)^{\nu}2^{1-\nu}\Gamma(\nu)^{-1}\mathrm{K}_{\nu}(2\sqrt{\nu}h/\phi)\right\}, \tag{5.7}$$

for inter-site distance $h \geq 0$, K_ν a modified Bessel function of the second kind, and the positive parameters (α, ϕ, ν) which determine the variance, range, and smoothness, respectively (Banerjee et al. 2014).

In general, the estimation of the smoothness parameter, ν, is difficult in practice (Bai et al. 2012). A popular approach to alleviate this issue is to estimate ν via a grid search (H. L. Ip & W. K. 2017). We fit the r-Pareto process over a grid of values for ν and found that $\nu = 0.1$ maximised the likelihood function. We found the risk threshold we used to model extreme extremal dependence for summer temperature data provided an accurate choice for modelling winter minima. That is, we choose events whose risk is above the 80% quantile of the risks calculated for all observed events. Figure 5.18 shows the estimated χ^P_o derived from the fitted r-Pareto process with associated 95% uncertainty bounds based on 200 spatio-temporal bootstrap samples, along with the pairwise empirical estimator as in expression (5.4), estimated for a high quantile ($u = 0.92$). The figure shows a good fit to the data, suggesting the r-Pareto process is capturing the level of asymptotic dependence well. We tested for linear non-stationarity in both the scale parameter of the Mátern variogram, α, which controls the maximum variability of the process and is approached asymptotically by the variogram. We use year, scaled

between 0 and 1, as a temporal covariate. However, similar to extreme maximum daily temperatures Chapter 4, we do not find any statistically significant evidence for change in the variogram's parameters.

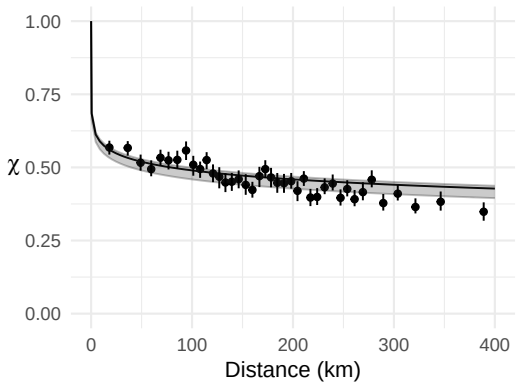

Figure 5.18: Estimates of χ_o^P against inter-site distance h for model E2: the pairwise empirical estimator of χ_o^P for $u = 0.92$, calculated as in expression (5.4) is shown as vertical line segments reflecting the bootstrap uncertainty in the marginal distribution estimates; the smooth lines display the limiting value for χ_o^P, estimated using the r-Pareto process with a Matérn variogram when fitted above an 80% risk threshold, with 95% confidence intervals, capturing both marginal and extremal dependence bootstrap uncertainties.

5.5 Results

In this section, we present a range of summaries of how daily minimum winter temperature extreme events in Ireland are changing over the period 1950–2022. Firstly, we characterise how winter temperature extremes are changing on a marginal level in Section 5.5.1. Following this, in Section 5.5.2, we summarise temporal non-stationarities in the spatial context. We present a range of extremal simulation summary statistics detailing how spatial daily minimum temperature extreme events in Ireland are changing over the period 1950–2022, parallel to those presented in Section 4.5. Throughout this section, we present results in the context of low, median, and high SCV.

5.5.1 Marginal return level results

Figure 5.19 presents estimates of the 100-year return level in the context of the year 2022 for model E2. The 100-year level corresponds to an occurrence probability of $1/(100\times90.25)$. In this figure, we see the return level associated with different phases of SCV, i.e., $M_{r,m}^{(0.1,I)}(t)$, $M_{r,m}^{(0.5,I)}(t)$, and $M_{r,m}^{(0.9,I)}(t)$, from left to right. The figure clearly shows that, in lower phases of SCV, we see much less extreme temperatures. We estimate a return level of -8.9°C to -2.7°C over Ireland associated with the covariate level $M_{r,m}^{(0.1,I)}(t)$ corresponding to the low phase of SCV. At increasing levels of SCV, we see more intense return levels, i.e., we estimate a return level of -10.9°C to -3.9°C and -14.6°C to -6.8°C associated to the levels $M_{r,m}^{(0.5,I)}(t)$ and $M_{r,m}^{(0.9,I)}(t)$, respectively. The different return levels in each case have equivalent occurrence probabilities in the context of their respective phases of climatic oscillations.

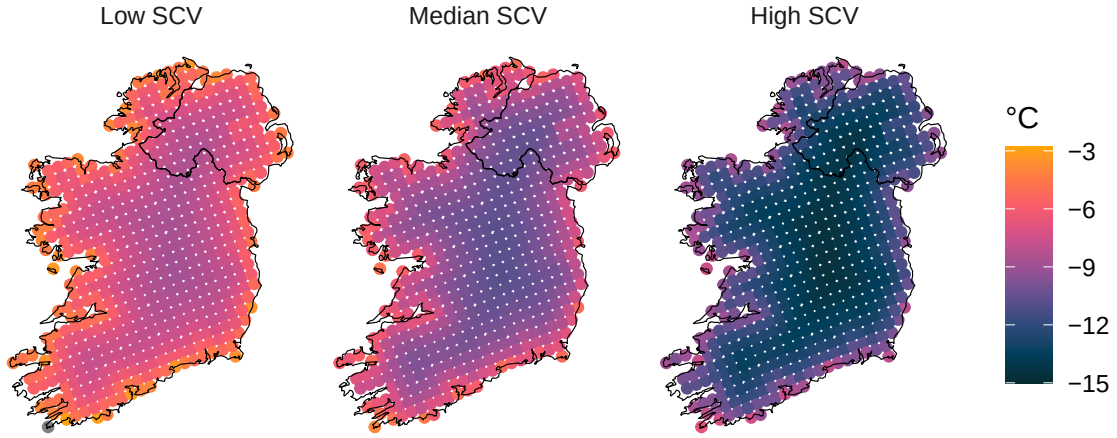

Figure 5.19: Estimated 100-year marginal return level in 2022 for (left-right) low, median, and high phases of SCV derived from model E2.

Figure 5.20 shows half the width of the 95% confidence interval for 100-year return levels presented in the respective plots in Figure 5.19 based on 200 bootstrapped data sets. For each phase of SCV, we can see the most uncertainty inland, where we have the shortest records of data. Furthermore, among the three levels of SCV, most uncertainty occurs during high levels, which correspond to unusually cold events such as that of winter 2009/10.

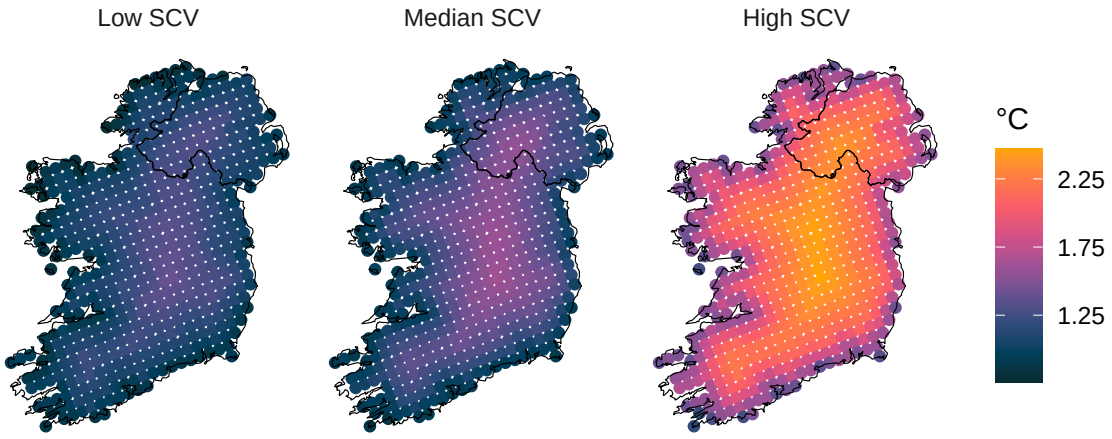

Figure 5.20: Half of the width of the 95% confidence interval of 100-year return level in 2022 for low, median, and high phases (left-right) of SCV derived from model E2 derived from 200 spatio-temporal bootstraps.

Figure 5.21 shows how the estimated 100-year level derived from model E2 has changed since 1950 during different levels of SCV. In each case, we see that the largest increases are inland. The 100-year return level during levels of median SCV (centre plot; Figure 5.21) has increased between 2.1–3°C. This increase in extreme temperatures over the observed record is substantially larger than the approx 0.8–1°C change of $M^I(t)$ over the same period (see Figure 5.4). This result reiterates that climate change is more radically affecting extreme temperatures than mean temperatures. Furthermore, an increase of 2.1–3°C of the 100-year level is in contrast to the respectively lower increase of 1.2–2.2°C estimated for maximum summer 100-year levels over a similar period (1942–2020) in the analysis carried out in Chapter 4. This indicates that winter extreme minimum temperatures are warming faster than summer extreme maximum temperatures, corroborating Ireland's reflection of global trends. We see the most substantial increase in 100-year return levels is in the context of high SCV (right-hand plot; Figure 5.21) with an estimated increase of 2.9–4°C over the island. This suggests that climatic conditions that would typically result in very cold temperatures in Ireland are now resulting in much warmer temperatures. This aligns with the findings of Christiansen et al. (2018), which suggest that the winter of 2009/10 was much milder than expected given the climatic conditions at the time. We estimate the lowest levels of warming during the low phase of SCV (left-hand plot; Figure 5.21) with an increase 1.9–2.7°C over the island. Figure 5.22 reports the 95% confidence interval of changes

in 100-year return levels calculated using 200 bootstrapped data sets. The columns in Figure 5.22 correspond to the respective column of Figure 5.21. The top row corresponds to the lower bound while the bottom row corresponds to the upper bound. At all three levels of SCV, and at all sites, the changes are positive.

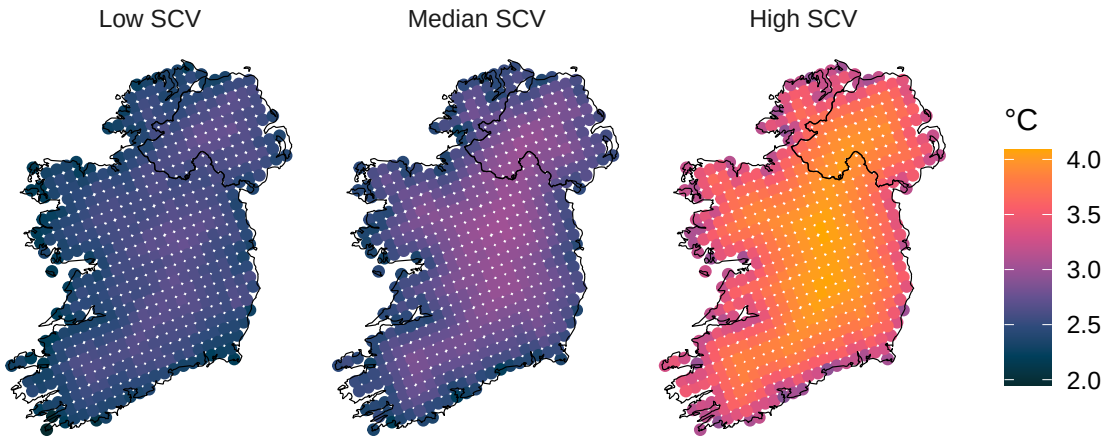

Figure 5.21: Estimated change in 100-year marginal return level over the period 1950–2022 for low, median, and high phases (left-right) of SCV derived from model E2.

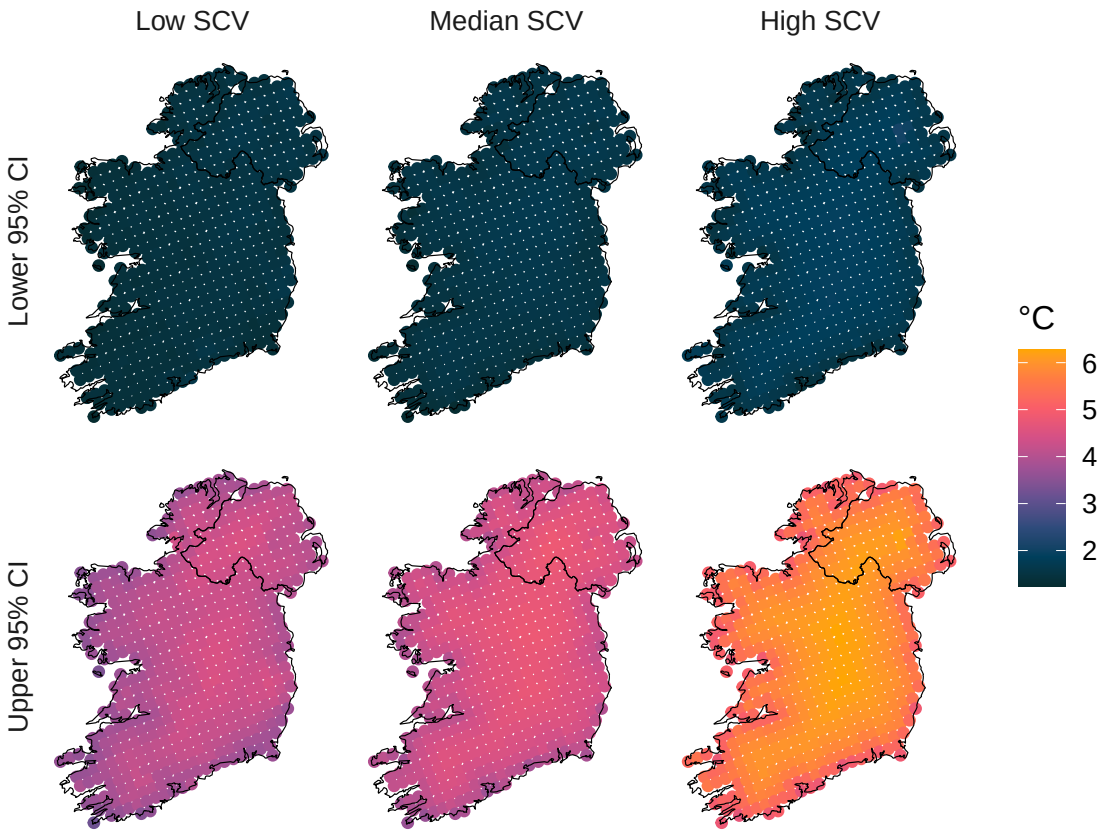

Figure 5.22: The lower bound of 95% confidence interval of change in 100-year return level according to model E2 for low, median, and high phases of SCV (top row, left to right respectively). The upper bound of 95% confidence interval of the corresponding plot in the top row (bottom row). Confidence intervals are based on 200 bootstrapped data sets.

We investigate the sensitivity of our threshold choice as discussed in Section 5.3.3.3. Figure 5.23 reports the difference between the 100-year level derived from our chosen model, E2, using our chosen threshold, $u^{(0.96)}(\boldsymbol{s})$, with 100-year levels estimated using a lower threshold (left-hand plot, with $\tau = 0.95$) and higher threshold (right-hand plot, with $\tau = 0.97$). The differences between the return level estimates are well within the 95% uncertainty interval of our chosen estimate. This suggests that while there is a small difference in return levels estimated with slightly different thresholds, these differences are captured by our bootstrapped uncertainties.

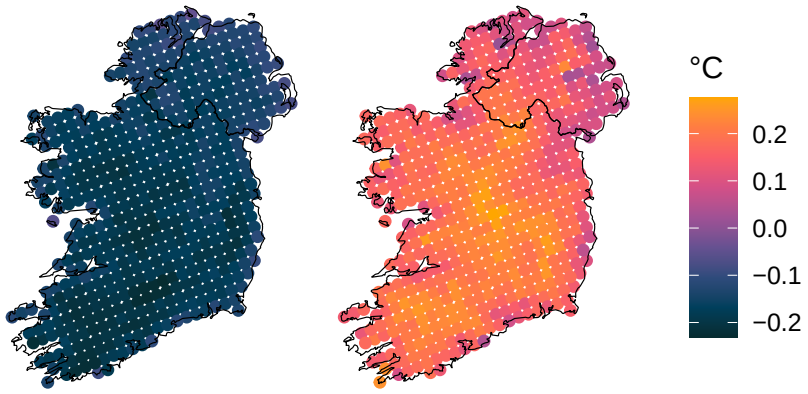

Figure 5.23: The difference in the 100-year level during the median phase of SCV in 2022 between model E2 fit to exceedances above the chosen threshold of $u^{(0.96)}(\boldsymbol{s})$ and to exceedances above both a lower threshold of $u^{(0.95)}(\boldsymbol{s})$ (left) and higher threshold of $u^{(0.97)}(\boldsymbol{s})$ (right).

5.5.2 Spatial results

To derive summary statistics of spatial extreme events, we first generate a database of spatial extreme simulations and present empirically estimated risk measures using these simulations. Our approach to simulation here is described in Section 4.4.4. To recall, we first simulate an event from the r-Pareto process, on Pareto margins, and then back transform this to the data scale using the inverse of expression (5.3). As the r-Pareto process is stationary over time, we can generate identical events in Pareto margins to transform to any specific time and any phase of SCV using the time-varying model E2. The r-Pareto process simulations are generated using the R package `mvPot` (de Fondeville et al. 2021), which exploits the decomposition in (5.6). Point estimates of the spatial risk summaries are based on $100,000$ simulations, for each phase of the SCV. Uncertainties are based on $10,000$ simulations for each r-Pareto process fitted to each of the 200 spatio-temporal bootstrap data sets, totalling an additional $200,000$ simulations for each phase of SCV.

We are most interested in making inferences about spatial events of the observational process that exceed a critical temperature of T°C somewhere over Ireland at time t. We denote these events by

$$A_{t,\mathcal{S}}(T) = \{X_o^P(t,\boldsymbol{s}), \boldsymbol{s} \in \mathcal{S} : \exists\ \boldsymbol{s}_0 \in \mathcal{S} \text{ with } X_o^P(t,\boldsymbol{s}_0) < T^P(t,\boldsymbol{s}_0)\}, \quad (5.8)$$

where $T^P(t, \boldsymbol{s})$ is the mapping of T through the transformation (5.3) at time t and for site $\boldsymbol{s}$. Recall that we are modelling negated minimum daily temperatures, so the set $A_{t,\mathcal{S}}(T)$ in (5.8), corresponds to extremely cold temperatures. To estimate $\Pr\{A_{t,\mathcal{S}}(T)\}$ we use the importance sampling procedure proposed in Equation (4.17), Section 4.4.4. Estimates of $\Pr\{A_{t,\mathcal{S}}(T)\}$ for a range of extremal temperatures and for different phases of SCV are given in Figure 5.24. Figure 5.24 shows this estimated probability (expressed as a return period) for a range of temperatures $T \in [-20, -5]$ for years 1950 and 2022 separately, for $\boldsymbol{s} \in \mathcal{S}_o$ (due to computational infeasibility we do not present estimated based on $\mathcal{S}_c$, however based on the analysis presented in Section 4.5, we can assume that estimates based on $\mathcal{S}_c$ will be only slightly different and within uncertainty estimates based on $\mathcal{S}_o$).

Figure 5.24: Return period of the event $A_{t,\mathcal{S}}(T)$ where an extreme temperature exceeding T°C occurs somewhere on the Irish station network, $\mathcal{S}_o$. Blue dashed (solid orange) lines correspond to $t = 2022$ (1950). Shaded regions show pointwise 95% confidence intervals for the return periods at low, median, and high SCV from left-right.

Rohan (1986) notes that a temperature of −10°C is an extremely cold temperature in Ireland and is in the coldest 1 percentile of daily temperature minima. Our model finds that the spatial event $A_{t,\mathcal{S}}(-10)$ has increased from a 1 in 0.2-year to a 1 in 0.7-year event in high phases of SCV, a 1 in 1.4-year to a 1 in 1-12 year event in median phases of SCV and a 1 in 5-year to a 1 in 70-year event in low phases of SCV. If we consider the coldest temperature in the republic of Ireland, observed in the winter of 2009/10, which was −17.5°C on the 25th of December 2010 in Co. Mayo, the spatial event $A_{t,\mathcal{S}}(-17.5)$ changed from approximately a 1 in 10-year event in 1950 to a 1 in 370-year event in 2022 in the context of a high phase of SCV. Regarding the coldest temperature in recorded history on the island

of Ireland of −19.4°C recorded at Omagh, Co. Tyrone on the 23rd of January 1881 (Hickey 2011), the spatial event $A_{t,\mathcal{S}}(-19.4)$ changes from being a 1 in 38-year event in 1950 to a 1 in 3,700-year event in 2022, in the context of a high phase of SCV. However, note that a temperature of this extremity is not in our data set so estimates are very uncertain.

The simulation strategy set out in Section 4.4.4 allows us to generate replicated independent spatial fields which exceed T°C for at least one $\boldsymbol{s} \in \mathcal{S}$. This is essentially achieved by exploiting the decomposition in (5.6) to scale simulations by multiplying R with a known change in probability of occurrence, i.e., $\Pr\{A_{t,\mathcal{S}}(T)\}$ as estimated above. We scale simulations to be below temperature T°C for at least one site. Estimates of the following risk measures are calculated on the scaled simulations, following this the estimates are rescaled to interpretable levels. Firstly, we investigate the behaviour of the "data scale χ", which combines the effect of changes in the marginal distributions over time with the estimated extremal dependence structure,

$$\chi_o(h; T, t) = \Pr(\exists\ \boldsymbol{s}_0 \in \mathcal{S} : \min[X_o(t, \boldsymbol{s}_0), X_o(t, \boldsymbol{s}^h)] > T),$$

where $\boldsymbol{s}^h$ is a randomly selected site in $\mathcal{S}$ with $\|\boldsymbol{s}^h - \boldsymbol{s}_0\| = h$. Since we are working with negated temperatures, this corresponds to two sites, distance h apart both being below a temperature T. Figure 5.25 presents estimates of this risk measure for a range of h and $T = \{-8, -9, -10\}$, between 1950 and 2022. Our analysis indicates a consistent pattern across the three phases of SCV. We observe a higher likelihood of joint cold extreme temperatures occurring at greater distances during earlier periods. As temperatures become increasingly extreme, there is a tendency for them to become more localised. Additionally, the probability of joint extreme temperature events is greater during periods of high phases of SCV, even at the furthest distances. We note that the uncertainties associated with these findings exhibit a high degree of noise. This is primarily due to the substantial computational resources required for generating and storing simulations for calculating this statistic. The computational burden is increased as we are interested in making inferences at different levels of climatic variability. We report uncertainty as the 95% quantiles of $\chi_o(h; T, t)$ in bins of 25km. For $\chi_o(h; T, t)$ we see that the joint probability of temperature being below T at sites h apart changes notably with

time, e.g., taking $h = 100$ km, we find that $\chi_o(h; T, t)$ has increased by a factor of 7.4, 13.7, and 24.2 for $T = -8, -9$, and -10°C, respectively, during the low phase of SCV. The respective equivalent estimates for median and high phases of SCV are 4.7, 6.9, 11.9 and 2.8, 3.3, 3.9. Notice that we see the smallest increases during the highest phase of SCV, this is due to the fact that extremely low temperatures are much more likely to occur in these periods, and so these temperatures are much farther into the tail of the GPD in phases of low SCV.

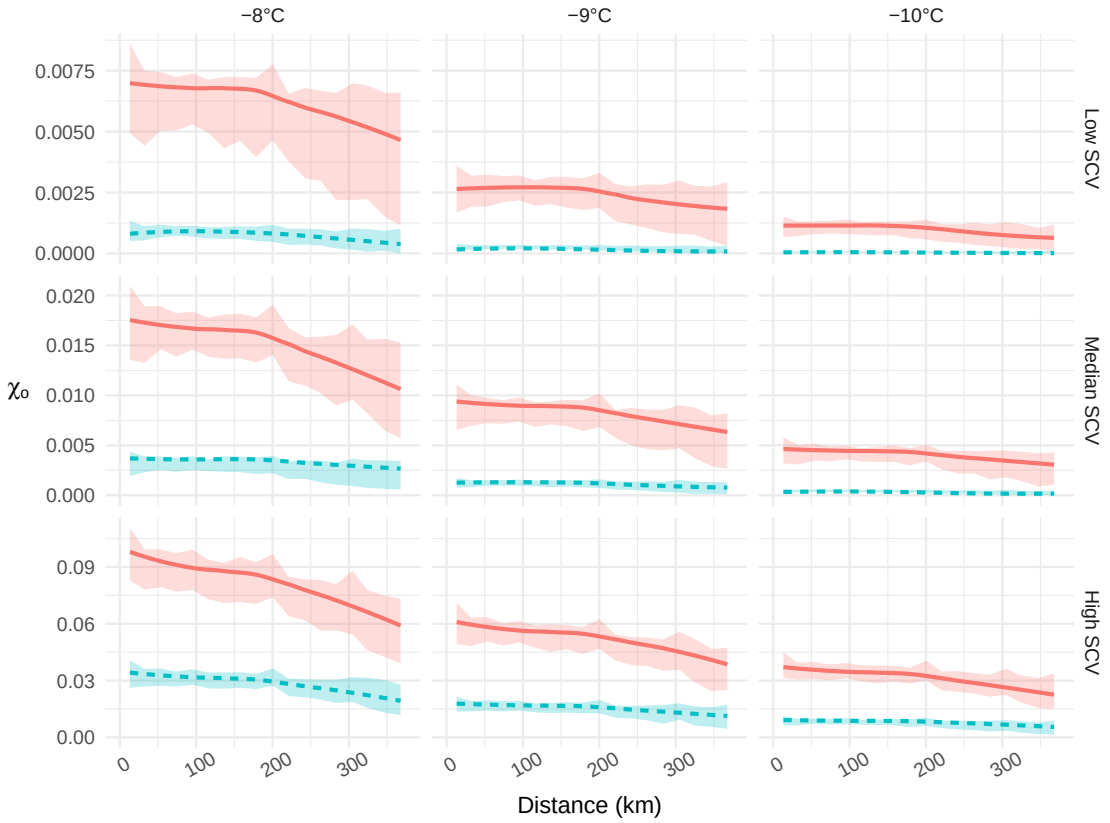

Figure 5.25: Estimates of $\chi_o(h; T, t)$ against h (in km) for $T = -8, -9$, and -10°C (left-right respectively) for 1950 (solid, orange line) and for 2022 (dashed, blue line) for model E2 at low, median, and high levels of SCV (top to bottom row respectively). Confidence intervals are based on 10,000 simulations for each 200 bootstrap sample data set and each level of SCV.

Finally, in Figure 5.26 we look at a fully spatial risk measure on the station network over Ireland from 1950 to 2022. This figure plots the expected proportion of a temperature field on $\mathcal{S}$, which exceeds T°C, i.e., it plots $E_o^D(T; t)$ against T, defined

as:

$$E_o(C;t,T) = \mathrm{E}\left(\frac{1}{|\mathcal{S}|}\int_{\mathcal{S}} I\{X_o(t,\boldsymbol{s}) > T\}d\boldsymbol{s}\right)$$

where $I(B)$ is the indicator function of event B. Again, since data is negated, $E_o(C;t,T)$ corresponds to the expected proportion of sites that are colder than temperature T. The figure shows that the relative change in spatial extent over this period decreases with more extreme cold temperatures T. Figure 5.26 shows that for $T = -10°\mathrm{C}$, estimates of $E_o(C;t,T)$ have decreased by a factor of 17.2, 10.2, and 4.68 in phases of low, median, and high SCV. This indicates that events are becoming more localised at all phases of short-term climate variability.

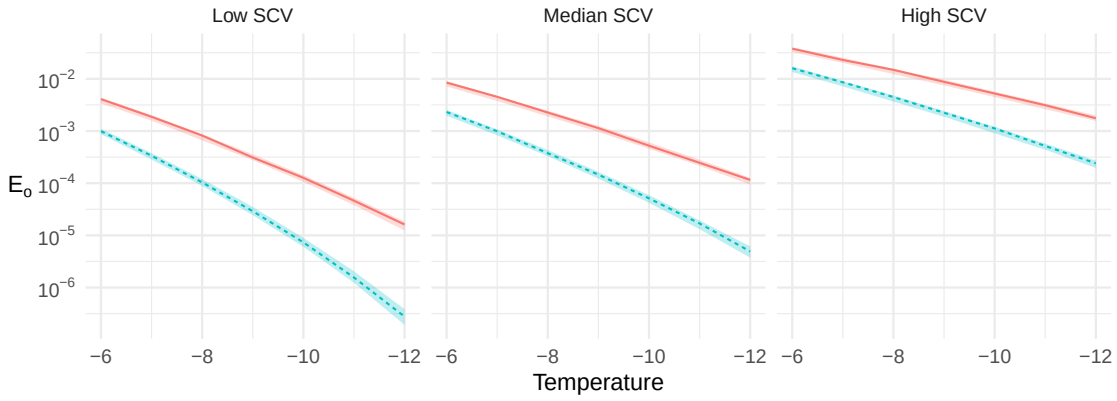

Figure 5.26: Expected proportion, $E_o(C;t,T)$, of Ireland that exceeds a temperature of $T°\mathrm{C}$ in an extreme event given that at least one site in Ireland (at the station network) exceeds $T°\mathrm{C}$ according to E2 for low, median, and high levels of SCV (left to right, respectively) in 1950 (solid, orange line) and 2022 (dashed, blue line). Confidence intervals are based on 10,000 simulations for each 200 bootstrap sample data set and each level of SCV.

5.6 Conclusion & discussion

The aim of this research was to identify and characterise non-stationarities of extremely cold daily winter temperatures in Ireland. Given the complex and variable nature of extreme minimum winter temperatures, we emphasise the importance of considering the climatic context (i.e., the atmospheric and oceanic patterns) in which they occur. Specifically, we discuss how shifts in the jet stream's behaviour can lead to the increased occurrence of extremely cold winter temperatures. We have presented some novel candidate approaches to account for large variations in cold extremes during different phases of large-scale climatic patterns. We explore using NAO and AO indices as covariates to help explain the behaviour of the jet stream and its impact on weather patterns. However, we find that using the HadCRUT5 data set, which combines land and sea surface temperature anomalies, best captures and represents all climatic processes driving extremely cold temperatures. Furthermore, we show that not accounting for this climatic variability conflates unusually cold winter events with the contrasting general trend of warming winters, and underestimates long-term warming trends.

We have presented a method of characterising and visualising climate risk associated with different phases of what we define as "short-term climatic variability" (SCV). We have developed a novel bootstrapping approach (used to account for uncertainty throughout) to accommodate the stronger temporal dependence and complex variability observed in extremely cold winter events, as compared to hot summer events. We found that the climate model data were less helpful as a covariate for the GPD as compared to the case for maximum daily summer temperatures. However, for non-extreme minimum daily summer temperatures, the climate model output proved informative and helpful.

Our study reveals that, across Ireland, there has been a decrease in the frequency and intensity of extremely cold temperatures since 1950. Furthermore, the rate of warming of extreme minimum temperatures is substantially greater than that of maximum summer temperatures as found in Chapter 4. Overall, the analysis reveals there has been a decrease in the frequency and intensity of extremely cold temperatures since 1950 in Ireland, with the largest changes observed in regions further inland, during all phases of SCV considered.

We found that from 1950 to 2022 the occurrence rates of high threshold exceedances have decreased for each phase of SCV. We found that extreme quantiles have increased by 2–4°C, over different phases of SCV. We describe changes in spatial risk over time, combining both marginal and dependence features, to describe changes in spatial risk over time. We found that spatial cold events have become more localised, with this change increasing at more extreme temperatures, and greatest for low phases of SCV. Overall, spatial cold extreme events are becoming less frequent and warmer over time.

6

Discussion & Conclusion

In this chapter, we review and summarise the work and contributions of this thesis. We highlight some areas of further work and recommend possible extensions to our research.

6.1 Thesis Summary

The aim of this thesis is to present novel theoretical and methodological results for modelling extreme spatio-temporal events, with the motivation of modelling extreme temperatures over Ireland in the context of climate change. We briefly re-examine the work presented in previous chapters, examining some model advantages and limitations, providing final remarks, and discussing further work and possible extensions in each case.

Chapter 3

In Chapter 3 we explore the theoretical and practical properties of the Hüsler-Reiss multivariate extreme value distribution, which constitutes the finite-dimensional joint distributions of the popular spatial max stable process (the Brown-Resnick process) or the log-Gaussian process. We provide a derivation of the spectral density of the Hüsler-Reiss distribution, which leads us to two natural proofs that the distribution is closed under marginalisation. We discuss how this allows a practitioner to account for, and deal with, a common issue in spatio-temporal environmental statistics; that is, time-series records of differing lengths as well as missing data in general. We provide open-source software as well as a demonstration of that software, to make this modelling procedure widely accessible to applied scientists.

Chapter 4

In Chapter 4 we examine spatio-temporal non-stationarities in extremely hot daily summer temperatures in Ireland. We present novel candidate approaches to merge information from spatially and temporally complete climate models into the spatial extreme value analysis of sparse and temporally incomplete observed temperatures from available meteorological stations. Methods introduced in this chapter allow us to use climate model outputs to incorporate detailed topographical features on the spatial extremal modelling of spatially-coarse meteorological observation stations. The ability to account for detailed local spatial features in extremal inference gives a more physically realistic model of the process and can help highlight areas of particular risk or impact. We use the theoretical results presented in Chapter 3 to apply r-Pareto processes modelling in a missing data framework. We found

a substantial and statistically significant increase in the occurrence rates of high threshold exceedances as well as an increase in the 100-year return level from 1942-2020. We found that in general, extremely hot summer temperatures in Ireland have increased faster than the change in mean Ireland and global summer temperature anomalies over the study period. In this chapter we present novel simulation summarising metrics, combining marginal and spatial dependence features, to describe changes in spatial risk over time. Using visualisations of these metrics, we show that spatial heatwave events over thresholds that are critical for society have become much larger, with these changes increasing at more extreme temperatures. In the appendix to this chapter, we consider alternative marginal modelling choices compared to those taken in the main analysis. We use this as a means to ascertain the sensitivity of our model choices.

Chapter 5

In Chapter 5 we identify and characterise spatio-temporal non-stationarities of extremely cold winter temperatures in Ireland. Given the complex nature of extreme minimum winter temperatures, we emphasise the importance of considering the climatic context (i.e., the atmospheric and oceanic patterns) in which they occur, particularly the shifts in the jet stream's behaviour, which has a strong influence on the occurrence of extremely cold winters in northern Europe. We present some novel candidate approaches to account for large variations in cold extremes during different phases of large-scale climatic patterns. We examine winter extremes at three phases of large-scale oscillations and build upon visualisation methods in Chapter 4, introducing a method to characterise and visualise climate risk associated with different phases of what we define as "short-term climatic variability" (SCV). We found that the climate model data were less helpful as a covariate for modelling the spatial pattern of extremely cold winter temperatures as compared to the case for maximum daily summer temperatures in Chapter 4. However, we still incorporate this physiographic information to inform the non-extreme model as well as the spatial pattern of the threshold, used to define what temperatures are extreme. Our analysis suggests coastal proximity is a sufficient spatial covariate for extremely cold winter temperatures. Spatial cold extreme events are becoming less frequent and warmer from 1950–2022. In Ireland, the occurrence rates of threshold exceedances for extremely cold temperatures have decreased over the study period,

with the most significant changes observed in inland regions. We found the occurrence rates of threshold exceedances have decreased, at a faster rate than hot summer extremes have increased. We find that the rate of warming of extreme minimum temperatures is substantially greater than that of the level of warming in maximum summer temperatures analysed in Chapter 4.

6.2 Limitations of our methodology

In this section, we will highlight some drawbacks of our developed methodology as well as give suggestions for improvements and extensions.

Our analysis relies on one climate model output. Any climate model will capture different modelling assumptions and therefore would provide an alternate estimate of the extremal process being modelled. An extension to our approach could involve the comparison or combination of several climate model outputs as spatial covariates of extreme temperatures over Ireland.

Since the r-Pareto process describes asymptotic dependent data, a limitation of this modelling framework is its potential lack of adaptability to other, larger spatial domains or regions of complex topography or different more localised extremal processes (e.g., precipitation). In other applications, it is likely that the practitioners will need to account for asymptotic independence and the decay of asymptotic independence at further distances. This requires an alternative modelling procedure of extremal dependence that can capture these features, examples of such models are discussed below in Section 6.3.2.

A major advantage of the r-Pareto process modelling procedure is the flexibility afforded in defining the risk functional r, which can pull out extreme events of particular interest. The only restriction of the risk function is that it must be homogenous of order 1. In our methodology, an additional restriction is imposed, that being that the risk function must be as close as possible to being invariant to the changing dimension of partially observed events, whatever their missing patterns. Future research could involve the investigation and demonstration of alternate risk functions that satisfy these restrictions.

In both major analyses presented in this thesis, we focus on a specific season in

order to capture the most impactful extreme events, i.e., summer for daily maxima and winter for daily minima. In future work, to develop a year-round model, we would include a non-stationary threshold, such that levels of extremity reflect the context of the season in which an event occurs. This could initially be achieved by taking a high/low quantile in each month as a threshold.

A limitation in extending our methodology to alternate climatic processes is the necessary process of model selection. Given a set of potentially useful covariates, the choice of which best represents the underlying phenomenon being modelled, as well as how best to parameterise the model is non-trivial, tedious, and time-consuming. In future, we suggest the exploration of non-parametric machine-learning methods, such as BART (Chipman et al. 2010), which remove this onus from the practitioner, especially if the model space is extensive. Specifically, an extension of the work developed by Murray (2017) (which uses data augmentation to fit BART models to non-Gaussian data) to allow the fitting of BART to GPD or GEV distributed data could be explored.

Communicating risk in an interpretable way is essential to inspire action against human-caused climate warming. However involved and theoretically complex any such multivariate extreme analysis is, the results must be presented in an interpretable way. The purpose of analyses such as those in Chapter 4 and Chapter 5 is to help our understanding of the impact of climate change. Climate change is of global public concern and so our best and current understanding of climate change should be globally and publicly understandable. Although we present some potential methods for summarising our analyses here, additional work is required to develop tools to effectively demonstrate and communicate risks associated with processes that are inherently spatial and temporal.

6.3 Further work

In this section, we discuss several potential avenues for future work. In Section 6.3.1, we discuss complications involved in extrapolating inference beyond the observed time period, while in Section 6.3.2 we propose possible approaches for expanding the analysis over larger spatial domains. In Section 6.3.3, we highlight an important and impactful extremal process that is under-studied, requiring novel methodologies. In

Section 6.3.4, we present a stepping stone towards a fully Bayesian model for the spatial extremal analyses we have tackled via likelihood-based inference.

6.3.1 Temporal extrapolation

A clear extension to our research would be the quantification of future extreme Irish temperature events, e.g., what will a 100-year event look like at a given point in the future? One way to achieve this is to model data from simulations of future climate. Potential future climate scenarios are categorised by a representative concentration pathway (RCP) which is indicative of the level of radiative forcing caused by possible concentrations of greenhouse gasses. Radiative forcing is a metric used to classify potential future climate scenarios, and measures the "net change in the energy balance of the Earth system due to some imposed perturbation" (Huang et al. 2013). In this case, uncertainty quantification is essential and it is achieved by modelling outputs from several climate model ensembles. A multi-model ensemble is made up of a collection of regional and global climate model pairs (Giorgi 2019). A major challenge in climate model ensemble analysis is the computational cost of storing and processing numerous climate model outputs. Furthermore, climate models may provide a poor representation of extreme weather events (Li et al. 2018, Shin et al. 2019) and so it is important to consider the inherent limitations of climate models when assessing extreme weather events.

In both analyses in Chapter 4 and Chapter 5, we use global mean temperature as a covariate for temporal non-stationarity where we estimate the coefficient of this covariate using observational data. An alternative approach to temporal extrapolations may involve deriving the equivalent mean temperature covariates from several different future simulations of climate allowing a projection of the estimated GPD into future scenarios. Extrapolating covariates is less computationally expensive and gives a more interpretable descriptor of climate change. Here we rely on climate models' description of mean temperature behaviour which is much more successfully simulated.

6.3.1.1 Oceanic influences

Considering all potentialities of future climate change is practically impossible and it is typically only feasible to analyse the most likely and impactful scenarios. One

major potential climate event we wish to highlight here and consider in future work is the overturning of Atlantic Meridional Overturning Circulation (AMOC). In recent decades, the majority of the Earth's surface has warmed except for a so-called "cold blob" over the North Atlantic, south of Greenland and to the west of Ireland. This cooling is caused by a slowing down of the AMOC (Rahmstorf et al. 2015, Caesar et al. 2018) which is currently at its weakest in the last millennium (Caesar et al. 2021). The slowing of the AMOC has been linked to anthropogenic forcings (Qasmi 2023) and under the current scenario of future emissions is expected to halt by 2050 (Ditlevsen & Ditlevsen 2023). The AMOC (sometimes called the "global ocean conveyor belt") transports warm water from the Gulf Stream along the surface of the Atlantic Ocean northward, where it eventually cools (warming the atmosphere), sinks, and is pushed southward to the Antarctic. The cold water travels along the bottom of the ocean to the Antarctic, rises due to local climatic forces (e.g., upwelling), and gets pushed back into the Gulf Steam. The AMOC is central to the Earth's climate and has a major influence on weather systems in the mid-latitude of the Northern Hemisphere. However, an influx of freshwater (melting icecaps, increased rainfall) has lowered the salinity of surface water in the northern Atlantic, making it less dense, and less likely to sink, thus weakening the AMOC.

It is important to consider the behaviour and influence of a slowing AMOC in the projection and modelling of non-stationarities in extreme weather events over Ireland (and indeed globally) as a potentially important driver of extreme weather events. For example, Yin & Zhao (2021) show that the AMOC moderates extreme cold events in the United States, and a reduction in heat transfer by the AMOC could cause more extremely cold winter temperatures. The climate works to stabilise and compensate for heat differences between land and ocean (Yang et al. 2018), so a collapse of the AMOC could result in a significant cooling over the northern hemisphere (Jackson et al. 2015). There is a huge amount of uncertainty in estimates associated with the behaviour of the AMOC given the relatively course and short observational records. Research in this area heavily relies upon proxy records and climate models. Given recent estimates of a near-future collapse of the AMOC, urgent research efforts and collaboration between statisticians or climate scientists are required to understand what consequences this will have on extreme weather events. An initial analysis could involve investigating the relationship between measures of the AMOC (including proxy records) and trends in extreme

temperatures over Ireland, historically. Since records are few, climate model simulations of different scenarios will likely be the most viable source of data.

Concurrently, we have recently seen the most extreme heatwaves over the Atlantic Ocean from 1940 to present (the duration of ERA5 records, Hersbach et al. 2020). June 2023 saw the highest average sea surface temperatures (SST) over the north-eastern Atlantic with an average temperature of 1.36°C above average for the entire month (calculated using a 30-year reference period from 1991–2020, Copernicus 2023). The average temperature of the northern Atlantic for the same month was around 0.5°C higher than the previous record for June, set in 2010. NOAA's Marine Heatwave Watch classified the heatwave to the west of Ireland in June 2023 as "Beyond Extreme", with max temperatures about 4–5°C above average (Hobday et al. 2018). Further analysis is required to better understand the influence of extreme SSTs and their effects on land surface temperatures. A possible avenue of future work is the investigation and modelling of the extremal dependence between SST and land temperature in order to accurately account for the influence of extreme oceanic heatwaves or cooling.

6.3.2 Spatial extrapolation

A very valuable potential extension of our research would be to model extreme temperatures over the European continent. By conducting a broader analysis of temperature extremes over Europe, we can ascertain local levels of climate change within the context of the entire continent and potentially discover trends and patterns that are not observable at a local level. Specifically, a wider assessment of extreme temperature events enables a more contextualised analysis of climate impacts in Ireland, within the framework of global climatic variations. However, such an analysis presents major methodological challenges, from collating and pre-processing observational data to managing the computational cost required to model such a large data set. Moreover, the need to account for the likely low levels of extremal dependence persisting over such a vast region raises questions about whether an r-Pareto process approach is viable, and if not, how instead the phenomenon is best modelled.

To ease the burden of data cleaning and management, we might rely on a gridded observational-based data set, for example, the E-OBS data set illustrated in

Figure 6.1. However, the computational complexity and dependence limitations persist. Given Europe's diverse climates, incorporating relevant covariates, such as land type, coastal distance, and elevation, becomes vital for this analysis. While our marginal modelling procedure is applicable to data sets over Europe with the use of such covariates, implementing an extremal dependence model requires careful consideration of extremal dependence. The assumption of asymptotic dependence (a prerequisite for r-Pareto processes) over such a large domain is very likely to no longer be valid.

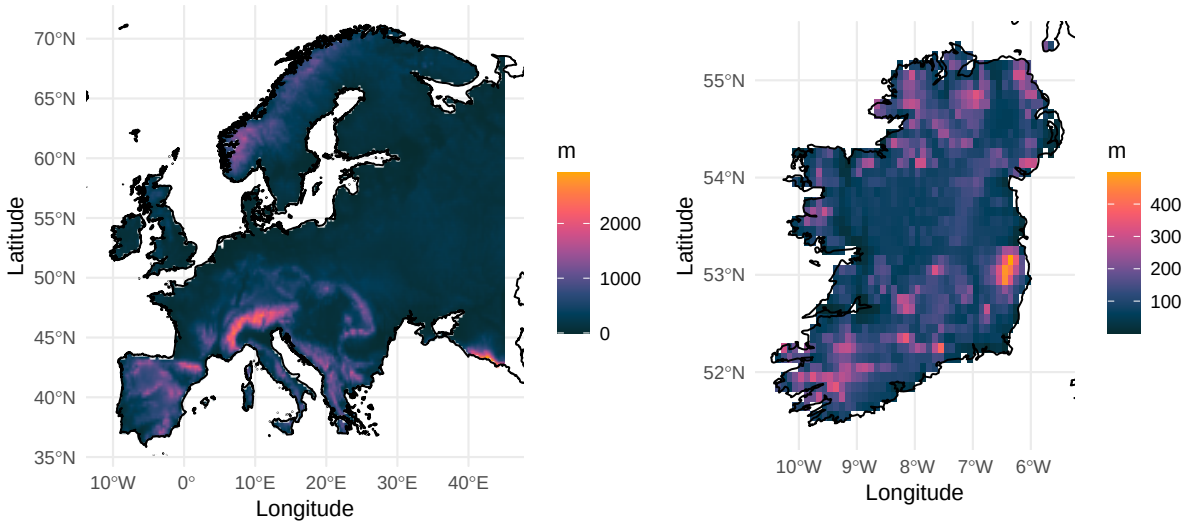

Figure 6.1: Average elevation (in metres) in each grid point of the E-OBS data set over Europe (left) and Ireland (right).

There are some existing methodologies dealing with multiple dependence classes, such as the "Gaussian scale mixtures" models proposed by Huser et al. (2017) and Wadsworth & Tawn (2022). The authors develop a model which allows a smooth transition between dependence classes. However, these methods are limited to moderately sized applications as the computational cost required to fit them remains a significant constraint. Zhang et al. (2022) alleviates the computation burden of these approaches to circumvent the most onerous computations involved in fitting the model while still maintaining the theoretical properties of the model. However, they cannot capture complete independence even at infinite distance. Hazra et al. (2023) develops a model based on the Gaussian scale mixture model, which can capture short-range asymptotic dependence, mid-range asymptotic independence, and

long-range exact independence. The authors achieve this by replacing the univariate random scale (representing the extremity of the event) with a spatial random process, allowing local scaling of events rather than global over the domain. However, this model has been applied to the rainfall data over the country of Bangladesh (a data set of 195 grid cells), which is a small spatial domain compared to the continent of Europe.

An alternative approach to handle weakly dependent extreme processes at large distances was proposed by Wadsworth & Tawn (2022). They extend the multivariate conditional extremal approach modelling approach, using a spatial conditional extremes model combined with latent Gaussian models. Simpson et al. (2023) showed that the modelling high dimensional data within the conditional extremal modelling framework can be made possible with efficient estimation using INLA. A major difficulty involved with this methodology is the choice of a conditioning site. The authors note that a composite likelihood can be constructed using all locations in the data set as the conditioning site. However, this is computationally burdensome. While analysis of extreme temperature events over Europe is crucial for climate change understanding, the challenges involved are non-trivial. Existing methodologies offer potential solutions for an initial analysis and areas of exciting research.

6.3.3 Multi-process extreme events

Globally, much research attention has been placed on the trends of unusually high maximum daily temperature events with less attention placed on extremely warm minimum night-time temperatures and even less (to our knowledge, none) on the extremal dependence between daily max and min temperatures during hot events. We wish to highlight this area as an important and impactful avenue of future research. The combination of a low diurnal temperature range with an extremely high maximum temperature can pose considerable risks to human health. Without adequate relief from extreme heat, individuals are susceptible to potentially fatal health consequences (Basu 2009, Murage et al. 2017, Kim et al. 2023). Extreme heat during the summer of 2022 caused an estimated 61,000 deaths in Europe alone (Ballester et al. 2023). The IPCC (2021) defines a compound extreme event as "the combination of multiple drivers and/or hazards that contribute to societal or

environmental risk". We suggest that the diurnal features of temperature processes can be considered as a compound extreme event.

Diurnal temperature ranges have generally decreased since the 1950s (Thorne et al. 2016, Sun et al. 2019), with daily minimum temperatures increasing at a rate nearly double that of maximum temperatures around the same period (IPCC 2021). Analysis by Vose et al. (2017) supports that nights are warming faster than days on average across the USA, demonstrating the asymmetry of warming between minimum and maximum daily temperatures. In light of this, we propose a separate treatment of upper and lower extremes, followed by an analysis of dependence when the processes are jointly extreme during hot events (i.e., extremely high maximum temperature and extremely low diurnal range). A popular approach for examining compound extreme events is that of copula-based approaches. For example, Wazneh et al. (2020) look at the interdependence of extreme precipitation and temperature in the south of Ontario, Canada using copula models. Critically, they found an underestimation of concurrent extreme temperature and precipitation when they ignored the relationship between the two variables. Similarly, Zhou & Liu (2018) describe a method to estimate the likelihood of concurrent precipitation and temperature extremes in China using copula-based models to estimate the occurrence probability. The authors highlight the need to consider both spatial and temporal non-stationarity in concurrent extreme modelling.

The challenging paradigm of concurrent and compound extreme events has received increased attention in recent years. However, methods so far have lacked a natural spatial interpretation. More recently, Vignotto et al. (2021) provide a spatial clustering of compound events, where they cluster the U.K. and Ireland into regions of similar extremal dependence structures between of compound precipitation and wind extremes. It is unlikely that asymptotic dependence would persist over any sizeable areas between two variables, however by clustering the domain into mutually asymptotically dependent sub-regions we could model them separately, at the cost of a larger scale interpretation. An area of potential future research involves the adaption of the r-Pareto process in order to capture the physical features of compounding extreme events. In this case, the individual processes need to be asymptotically dependent over the entire study region. An initial analysis may involve summarising the bivariate process (max and min temperature) at each site to a single variable at each site (e.g., diurnal range), where extremal dependence

modelling methodologies are well developed.

6.3.4 Bayesian modelling in extreme value theory

One of the main advantages of Bayesian models is their ability to quantify uncertainty. Instead of providing a single-point estimate, Bayesian models produce a joint probability distribution over all the model parameters. This would alleviate the onerous computational cost of creating full spatio-temporal bootstrapped data sets for uncertainty estimation as we have done in Chapter 4 and Chapter 5.

As a step towards a fully Bayesian framework, we explored a hierarchical Bayesian approach to estimate the parameters of the GPD. The use of Bayesian models for univariate extreme value models is a popular and widely applied method in the literature (Clancy et al. 2016, Sharkey & Winter 2019, Jiménez-Hernández et al. 2020). This framework is a potentially useful tool to address the spurious negative correlation often found between the scale and shape parameters of the GPD (Cooley & Sain 2010). Additionally, this framework is potentially useful for incorporating information from climate model outputs in extreme value inference. We present here an initial modelling framework with which we shall explore in future work. Given a process X observed over the spatial domain $\mathcal{S}$, i.e., $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$, our Bayesian generalised Pareto model is defined as,

$$\{X(\boldsymbol{s}) - u(\boldsymbol{s}) \;\mid\; X(\boldsymbol{s}) > u(\boldsymbol{s})\} \;\sim\; \text{GPD}\left(\sigma(\boldsymbol{s}), \xi(\boldsymbol{s})\right),$$

where,

$$\begin{bmatrix} \sigma(\boldsymbol{s}) \\ \xi(\boldsymbol{s}) \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu_{\sigma(\boldsymbol{s})} \\ \mu_{\xi(\boldsymbol{s})} \end{bmatrix}, \Sigma \right),$$

and,

$$\Sigma = \begin{pmatrix} \psi^2_{\sigma(\boldsymbol{s})} & \rho\psi_{\sigma(\boldsymbol{s})}\psi_{\xi(\boldsymbol{s})} \\ \rho\psi_{\sigma(\boldsymbol{s})}\psi_{\xi(\boldsymbol{s})} & \psi^2_{\xi(\boldsymbol{s})} \end{pmatrix}$$

with prior distributions, on $\mu_{\sigma(\boldsymbol{s})}$, $\mu_{\xi(\boldsymbol{s})}$, $\psi_{\sigma(\boldsymbol{s})}$, $\psi_{\xi(\boldsymbol{s})}$, and ρ. In Bayesian analysis, prior distributions reflect our prior knowledge of credible physical values while allowing them to be flexible enough to allow the data to inform the posterior distributions. Incorporation of this information about credible parameter values can be incredibly beneficial when working with extreme data given that the observations

are typically few. For example, we could place Gaussian priors on the mean vector of the MVN for each site, corresponding to the mean σ parameter and mean ξ parameter at that site. One potential is to centre the priors on parameter estimates from an MLE fitting of the GPD fitted to the climate model output. For restricting the standard deviation away from very large values while allowing flexibility near 0, Gelman (2006) recommends using a distribution from the half-t family. The Cauchy distribution is a special case of the t-distribution with degree of freedom equaling 1 (Ghosh et al. 2018). Therefore, we suggest initial exploration using the truncated Cauchy distribution for each of the of the standard deviations in the model. This prior is well known to be a weakly informative prior due to it having much lower tail decay as compared to the normal distribution. Finally, for the correlation coefficient ρ, we suggest either an uninformative uniform prior over $[-1, 1]$, or a more informative prior with most mass on zero, and very little mass on large values of ρ (either negative or positive), for example a beta distribution with large shape parameter values, re-scaled around zero.

To address the problematic correlation between scale and shape estimates we treat the scale and shape parameters at each site as multivariate normally (MVN) distributed with correlation ρ. The mean vector of the MVN for each site corresponds to the mean scale and shape parameter of the GPD at that site. In this way, we can make direct inferences about the correlation between scale and shape and incorporate prior knowledge about what we believe to be a realistic correlation or to restrict it.

Furthermore, we can incorporate hierarchy into the model by placing hyper-priors on the hyper-parameters of the model's prior distributions. Making this model hierarchical allows the parameters at each site to borrow strength from each other which allows for a more realistic estimation of the parameters.

We investigated an alternative implementation of Bayesian inference through the use of the computationally efficient `R-INLA` package (Lindgren & Håvard 2015). However, a restriction of the current implementation of the GPD in INLA is that the shape parameter must be positive. This restriction is not appropriate for temperature extremes in Ireland. We note that the model proposed above could be made temporally non-stationary in the same way as our GPD parameterisations in

Chapter 4 and Chapter 5, that is, for example, instead of modelling $\mu_{\sigma(\boldsymbol{s})}$, let

$$\mu_{\sigma(t,\boldsymbol{s})} = \beta_0 + \beta_1 \sigma_c(\boldsymbol{s}) + \beta_2 M^I(t)$$

describe the mean scale parameter at site $\boldsymbol{s}$ and time t where, we place prior distributions on the parameters β_0, β_1, and β_2, where the covariates σ_c and $M^I(t)$ are those as described in Chapter 4 and Chapter 5.

6.3.5 Final remarks

In conclusion, there are many possible directions for future work based on the research presented in this thesis. Fruitful avenues of further research range from the modelling of alternate extremal processes in Ireland using improved techniques, to extrapolation of inference into future climate scenarios and the modelling of extremes on larger spatial domains.

Bibliography

Adger, W. N., Crépin, A. S., Folke, C., Ospina, D., Chapin, F. S., Segerson, K., Seto, K. C., Anderies, J. M., Barrett, S., Bennett, E. M., Daily, G., Elmqvist, T., Fischer, J., Kautsky, N., Levin, S. A., Shogren, J. F., van den Bergh, J., Walker, B. & Wilen, J. (2020), 'Urbanization, migration, and adaptation to climate change', *One Earth* **3**(4), 396–399.

Angus, J. E. (1994), 'The probability integral transform and related results', *SIAM Review* **36**(4), 652–654.

Antwi, S. H., Rolston, A., Linnane, S. & Getty, D. (2022), 'Communicating water availability to improve awareness and implementation of water conservation: a study of the 2018 and 2020 drought events in the Republic of Ireland', *Science of The Total Environment* **807**, 150865.

Bai, Y., X.-K. Song, P. & Raghunathan, T. E. (2012), 'Joint composite estimating functions in spatiotemporal models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**(5), 799–824.

Balkema, A. A. & de Haan, L. (1974), 'Residual life time at great age', *The Annals of Probability* **2**(5), 792–804.

Ballani, F. & Schlather, M. (2011), 'A construction principle for multivariate extreme value distributions', *Biometrika* **98**(3), 633–645.

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R. F., Pegenaute, F., Herrmann, F. R., Robine, J. M., Basagaña, X., Tonne, C., Antó, J. M. & Achebak, H. (2023), 'Heat-related mortality in Europe during the summer of 2022', *Nature Medicine* **29**(7), 1857–1866.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, New York.

Bargaoui, Z. K. & Bardossy, A. (2015), 'Modeling short duration extreme precipitation patterns using copula and generalized maximum pseudo-likelihood estimation with censoring', *Advances in Water Resources* **84**, 1–13.

Barnett, V. (1976), 'The ordering of multivariate data', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **139**(3), 318–344.

Basu, R. (2009), 'High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008', *Environmental Health* **8**(40), 1–13.

Berman, S. M. (1961), 'Convergence to bivariate limiting extreme value distributions', *Annals of the Institute of Statistical Mathematic* **13**, 217–223.

Bindi, M. & Olesen, J. E. (2011), 'The responses of agriculture in Europe to climate change', *Regional Environmental Change* **11**(SUPPL. 1), 151–158.

Blitzstein, J. K. & Hwang, J. (2014), *Introduction to Probability*, Chapman & Hall/CRC Texts in Statistical Science, CRC Press/Taylor & Francis Group , USA.

Bopp, G. P., Shaby, B. A. & Huser, R. (2021), 'A hierarchical max-infinitely divisible spatial model for extreme precipitation', *Journal of the American Statistical Association* **116**(533), 93–106.

Boucek, R. E., Gaiser, E. E., Liu, H. & Rehage, J. S. (2016), 'A review of subtropical community resistance and resilience to extreme cold spells', *Ecosphere* **7**(10).

Brent, R. P. (1973), *Algorithms for Minimization Without Derivatives*, Dover Publications, Inc., New York.

Brown, B. M. & Resnick, S. I. (1977), 'Extreme values of independent stochastic processes', *Journal of Applied Probability* **14**(4), 732–739.

Brown, S. J., Caesar, J. & Ferro, C. A. (2008), 'Global changes in extreme daily temperature since 1950', *Journal of Geophysical Research Atmospheres* **113**(5), D05115.

Buishand, T. A., de Haan, L. & Zhou, C. (2008), 'On spatial extremes: with application to a rainfall problem', *The Annals of Applied Statistics* **2**(2), 624–642.

Byrd, R., Lu, P., Nocedal, J. & Zhu, C. (1995), 'A limited memory algorithm for bound constrained optimisation', *Journal of Scientific Computing* **16**(5), 1190–1208.

Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N. & Rahmstorf, S. (2021), 'Current Atlantic meridional overturning circulation weakest in last millennium', *Nature Geoscience* **14**(3), 118–120.

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G. & Saba, V. (2018), 'Observed fingerprint of a weakening Atlantic Ocean overturning circulation', *Nature* **556**(7700), 191–196.

Cammalleri, C., Naumann, G., Mentaschi, L., Formetta, G., Forzieri, G., Gosling, S., Bisselink, B., De Roo, A. & Feyen, L. (2020), Global warming and drought impacts in the EU, Technical report, European Commission, Ispra, Italy.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V. & Codron, F. (2010), 'Winter 2010 in Europe: a cold extreme in a warming climate', *Geophysical Research Letters* **37**(20).

Cebrián, A. C., Asín, J., Gelfand, A. E., Schliep, E. M., Castillo-Mateo, J., Beamonte, M. A. & Abaurrea, J. (2022), 'Spatio-temporal analysis of the extent of an extreme heat event', *Stochastic Environmental Research and Risk Assessment* **36**(9), 2737–2751.

Chapon, A., Ouarda, T. B. M. J. & Hamdi, Y. (2023), 'Imputation of missing values in environmental time series by D-vine copulas', *Weather and Climate Extremes* **41**(1), 100591.

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

Cheng, L., Abraham, J., Trenberth, K. E., Fasullo, J., Boyer, T., Mann, M. E., Zhu, J., Wang, F., Locarnini, R., Li, Y., Zhang, B., Yu, F., Wan, L., Chen, X., Feng, L., Song, X., Liu, Y., Reseghetti, F., Simoncelli, S., Gouretski, V., Chen, G., Mishonov, A., Reagan, J. & Li, G. (2023), 'Another year of record heat for the oceans', *Advances in Atmospheric Sciences* **40**(6), 963–974.

Chipman, H. A., George, E. I. & McCulloch, R. E. (2010), 'BART: Bayesian additive regression trees', *The Annals of Applied Statistic* **4**(1), 266–298.

Christiansen, B. O., Alvarez-Castro, C., Christidis, N., Ciavarella, A., Colfescu, I., Cowan, T., Eden, J., Hauser, M., Hempelmann, N., Klehmet, K., Lott, F., Nangini, C., Jan Van Oldenborgh, G., Orth, R., Stott, P., Tett, S., Vautard, R., Wilcox, L. & Yiou, P. (2018), 'Was the cold European winter of 2009/10 modified by anthropogenic climate change? An attribution study', *Journal of Climate* **31**(9), 3387–3410.

Claeskens, G. & Hjort, N. L. (2008), *Model Selection and Model Averaging*, Cambridge University Press, New York.

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

Clarke, D. C. & Richardson, M. (2021), 'The benefits of continuous local regression for quantifying global warming', *Earth and Space Science* **8**(5), 1–21.

Clavel, J., Merceron, G. & Escarguel, G. (2014), 'Missing data estimation in morphometrics: how much is too much?', *Systematic Biology* **63**(2), 203–218.

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

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

Coles, S. G. & Tawn, J. A. (1991), 'Modelling extreme multivariate events', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**(2), 377–392.

Coles, S. G. & Tawn, J. A. (1994), 'Statistical methods for multivariate extremes: an application to structural design (with discussion)', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **43**(1), 1–48.

Conlon, K. C., Rajkovich, N. B., White-Newsome, J. L., Larsen, L. & O'Neill, M. S. (2011), 'Preventing cold-related morbidity and mortality in a changing climate'.

Cooley, D. & Sain, S. R. (2010), 'Spatial hierarchical modeling of precipitation extremes from a regional climate model', *Journal of Agricultural, Biological, and Environmental Statistics* **15**(3), 381–402.

Copernicus (2023), 'Record-breaking North Atlantic Ocean temperatures contribute to extreme marine heatwaves'.
URL: *https://climate.copernicus.eu/record-breaking-north-atlantic-ocean-temperatures-contribute-extreme-marine-heatwaves*

Copernicus Climate Change Service (2019), 'CORDEX regional climate model data on single levels'.
URL: *https://cds.climate.copernicus.eu*

Cox, D. R. & Reid, N. (2004), 'A note on pseudolikelihood constructed from marginal densities', *Biometrika* **91**(3), 729–737.

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

CSO (2017), Census of Population 2016 - Profile 2 Population Distribution and Movements, Technical report, Central Statistics Office Ireland, Cork, Ireland.

CSO (2022), Census of Population 2022 - Preliminary Results, Technical report, Central Statistics Office Ireland, Cork, Ireland.

Cullen, E. (2007), Climate change and health in Ireland: A national vulnerability assessment, PhD thesis, Maynooth University.

David, H. A. & Edwards, A. W. F. (2001), Beginnings of extreme-value theory, *in* 'Annotated Readings in the History of Statistics', Springer, New York, pp. 145–150.

Davis, R. A., Kluppelberg, C. & Steinkohl, C. (2013), 'Statistical inference for max-stable processes in space and time', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(5), 791–819.

Davison, A. C. (1984), Modelling excesses over high thresholds, with an application, *in* J. T. de Oliveira, ed., 'Statistical Extremes and Applications', Vol. 131, Springer, Dordrecht, pp. 461–482.

Davison, A. C., Huser, R. & Thibaud, E. (2013), 'Geostatistics of dependent and asymptotically independent extremes', *Mathematical Geosciences* **45**(5).

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

Davison, A. C. & Smith, R. L. (1990), 'Models for exceedances over high thresholds (with discussion)', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **52**(3), 393–442.

de Fondeville, R., Belzile, L. & Thibaud, E. (2021), 'mvPot: multivariate peaks-over-threshold modelling for spatial extreme events'.

de Fondeville, R. & Davison, A. C. (2018), 'High-dimensional peaks-over-threshold inference', *Biometrika* **105**(3), 575–592.

de Fondeville, R. & Davison, A. C. (2022), 'Functional peaks-over-threshold analysis', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **84**(4), 1392–1422.

de Haan, L. (1984), 'A spectral representation for max-stable processes', *The Annals of Probability* **2**(4), 1194–1204.

de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer, New York.

de Haan, L. & Resnick, S. I. (1977), 'Limit theory for multivariate sample extremes', *Probability Theory and Related Fields* **40**(4), 317–337.

de Oliveira, T. J. (1958), 'Extremal distributions', *Rev. Fac. Ciências Lisboa* **8**.

de Oliveira, T. J. (1962), 'Structure theory of bivariate extremes: extensions', *Estudos de Matematica, Estatistica, e Economicos* **7**, 165–195.

Deser, C. (2000), 'On the teleconnectivity of the Arctic Oscillation', *Geophysical Research Letters* **27**(6), 779–782.

Diggle, P. J., Tawn, J. A. & Moyeed, R. A. (1998), 'Model-Based geostatistics', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**(3), 299–350.

Ditlevsen, P. & Ditlevsen, S. (2023), 'Warning of a forthcoming collapse of the Atlantic meridional overturning circulation', *Nature Communications* **14**(1), 4254.

Dombry, C. & Ribatet, M. (2015), 'Functional regular variations, Pareto processes and peaks over threshold', *Statistics and its Interface* **8**(1), 9–17.

Donat, M. G., Sillmann, J., Wild, S., Alexander, L. V., Lippmann, T. & Zwiers, F. W. (2014), 'Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets', *Journal of Climate* **27**(13), 5019–5035.

Dooley, K., Kelly, C., Seifert, N., Myslinski, T., O'Kelly, S., Siraj, R., Crosby, C., Dunne, J. K., McCauley, K., Donoghue, J., Gaddren, E., Conway, D., Cooney, J., McCarthy, N., Cullen, E., Noone, S., Murphy, C. & Thorne, P. (2023), 'Reassessing long-standing meteorological records: an example using the national hottest day in Ireland', *Climate of the Past* **19**(1), 1–22.

Dosio, A., Mentaschi, L., Fischer, E. M. & Wyse, r. K. (2018), 'Extreme heat waves under 1.5°C and 2°C global warming', *Environmental Research Letters* **13**(1), 054006.

Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A. A., Brunet, M., Caesar, J., Chagnaud, G., Cheng, V., Cinco, T., Durre, I., Guzman, R., Htay, T. M., Wan Ibadullah, W. M., Bin Ibrahim, M. K. I., Khoshkam, M., Kruger, A., Kubota, H., Leng, T. W., Lim, G., Li-Sha, L., Marengo, J., Mbatha, S., McGree, S., Menne, M., Milagros Skansi, M., Ngwenya, S., Nkrumah, F., Oonariya, C., Pabon-Caicedo, J. D., Panthou, G., Pham, C., Rahimzadeh, F., Ramos, A., Salgado, E., Salinger, J., Sané, Y., Sopaheluwakan, A., Srivastava, A., Sun, Y., Timbal, B., Trachow, N., Trewin, B., Schrier, G., Vazquez-Aguirre, J., Vasquez, R., Villarroel, C., Vincent, L., Vischel, T., Vose, R. & Bin Hj Yussof, M. N. (2020), 'Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3', *Journal of Geophysical Research: Atmospheres* **125**(16), e2019JD032263.

EASAC (2018), Extreme weather events in Europe, Technical Report March, European Academies' Science Advisory Council, Brussels.

Eastoe, E. F. & Tawn, J. A. (2009), 'Modelling non-stationary extremes with application to surface level ozone', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **58**(1), 25–45.

Enders, C. K. (2022), *Applied Missing Data Analysis*, second edn, The Guilford Press, New York.

Engelke, S. & Hitz, A. S. (2020), 'Graphical models for extremes (with discussion)', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**(4), 871–932.

Engelke, S., Kabluchko, Z. & Schlather, M. (2011), 'An equivalent representation of the Brown-Resnick process', *Statistics and Probability Letters* **81**(8), 1150–1154.

Engelke, S., Malinowski, A., Kabluchko, Z. & Schlather, M. (2015), 'Estimation of Hüsler–Reiss distributions and Brown-Resnick processes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(1), 239–265.

England, M. R., Eisenman, I., Lutsko, N. J. & Wagner, T. J. (2021), 'The recent emergence of arctic amplification', *Geophysical Research Letters* **48**(15), e2021GL094086.

Falk, M., Hüsler, J. & Reiss, R.-D. (2010), *Laws of Small Numbers: Extremes and Rare Events* , Vol. 23, 3 edn, Birkhäuser, Basel.

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

Finkelshteyn, B. V. (1953), 'Limiting distribution of extremes of a variational series of a two-dimensional random variable', *Dokl. Ak. Nauk. SSSR* **91**(2), 209–211.

Fischer, E. M. & Knutti, R. (2015), 'Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes', *Nature Climate Change* **5**(6), 560–564.

Fisher, R. & Tippett, L. (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Cambridge Philosophical Society* **24**(1), 180–190.

Francis, J. A. & Vavrus, S. J. (2012), 'Evidence linking Arctic amplification to extreme weather in mid-latitudes', *Geophysical Research Letters* **39**(6).

Francis, J. A. & Vavrus, S. J. (2015), 'Evidence for a wavier jet stream in response to rapid Arctic warming', *Environmental Research Letters* **10**(1).

Fréchet, M. (1927), 'Sur la loi de probabilité de l'écart maximum', *Annales de la Société Polonaise de Mathématique* **6**(1), 93–116.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Gürses, , Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., Mcgrath, M. J., Metzl, N., Monacci, N. M., Munro, D. R., Nakaoka, S. I., Niwa, Y., O'brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., Van Der Werf, G. R., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, A., Wright, R., Yuan, W., Yue, C., Yue, X., Zaehle, S., Zeng, J. & Zheng, B. (2022), 'Global carbon budget 2022', *Earth System Science Data* **14**(11), 4811–4900.

Fu, C. & Sayed, T. (2022), 'A multivariate method for evaluating safety from conflict extremes in real time', *Analytic Methods in Accident Research* **36**, 100244.

Fuentes, M., Henry, J. & Reich, B. (2013), 'Nonparametric spatial models for extremes: Application to extreme temperature data', *Extremes* **16**(1), 75–101.

Gan, R., Liu, Q., Huang, G., Hu, K. & Li, X. (2023), 'Greenhouse warming and internal variability increase extreme and central Pacific El Niño frequency since 1980', *Nature Communications 2023 14:1* **14**(1), 1–11.

Gao, Y., Merz, C., Lischeid, G. & Schneider, M. (2018), 'A review on missing hydrological data processing', *Environmental Earth Sciences* **77**(2), 1–12.

García, C., C.A, W., Dwyer, N. & Gault, J. (2022), Climate Status Report for Ireland 2020, Technical report, Environmental Protection Agency, Dublin.

Garrabou, J., Coma, R., Bensoussan, N., Bally, M., Chevaldonné, P., Cigliano, M., Diaz, D., Harmelin, J. G., Gambi, M. C., Kersting, D. K., Ledoux, J. B., Lejeusne, C., Linares, C., Marschal, C., Pérez, T., Ribes, M., Romano, J. C., Serrano, E., Teixido, N., Torrents, O., Zabala, M., Zuberer, F. & Cerrano, C. (2009), 'Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave', *Global Change Biology* **15**(5), 1090–1103.

Geffroy, J. (1958), 'Contribution à la théorie des valeurs extrêmes', *Publications de l'Institut de Statistique de l'Université de Paris* **7**(1), 37–185.

Gelman, A. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Analysis* **1**(3), 515–534.

Genest, C., Ghoudi, K. & Rivest, L. P. (1995), 'A semiparametric estimation procedure of dependence parameters in multivariate families of distributions', *Biometrika* **82**(3), 543–552.

Genton, M. G., Ma, Y. & Sang, H. (2011), 'On the likelihood function of Gaussian max-stable processes', *Biometrika* **98**(2), 481–488.

Gerber, E. P. & Vallis, G. K. (2009), 'On the zonal structure of the North Atlantic oscillation and annular modes', *Journal of the Atmospheric Sciences* **66**(2), 332–352.

Ghosh, J., Li, Y. & Mitra, R. (2018), 'On the use of Cauchy prior distributions', *Bayesian Analysis* **13**(2), 359–383.

Giorgi, F. (2019), 'Thirty years of regional climate modeling: where are we and where are we going next?', *Journal of Geophysical Research: Atmospheres* **124**(11), 5696–5723.

Glasbey, C. A. (1995), 'Imputation of missing values in spatio-temporal solar radiation data', *Environmetrics* **6**(4), 363–371.

Gleeson, E., McGrath, R. & Treanor, M., eds (2013), *Ireland's Climate: the road ahead*, Met Éireann, Dublin.

Gnedenko, B. (1943), 'Sur la distribution limite du terme maximum d'une serie aleatoire', *Annals of Mathematics* **44**(3), 423–453.

Gneiting, T. & Katzfuss, M. (2014), 'Probabilistic forecasting', *Annual Review of Statistics and Its Application* **1**, 125–151.

Gorman, C. E., Torsney, A., Gaughran, A., McKeon, C. M., Farrell, C. A., White, C., Donohue, I., Stout, J. C. & Buckley, Y. M. (2023), 'Reconciling climate action with the need for biodiversity protection, restoration and rehabilitation', *Science of the Total Environment* **857**(1), 159316.

Government of Ireland (2019), Climate Action Plan 2019, Technical report, Department of the Environment, Climate and Communications.

Graham, J. W. (2009), 'Missing data analysis: Making it work in the real world', *Annual Review of Psychology* **60**, 549–576.

Griffith, D. A. & Liau, Y. T. (2021), 'Imputed spatial data: Cautions arising from response and covariate imputation measurement error', *Spatial Statistics* **42**(1), 100419.

Gudendorf, G. & Segers, J. (2010), 'Extreme-Value Copulas', *Springer* **198**(Copula Theory and Its Applications, Lecture Notes in Statistics 198), 93–109.

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

Gumbel, E. J. (1960), 'Distributions de valeurs extrêmes en plusieurs dimensions', *Publications de l'Institut de Statistique de l'Université de Paris* **9**(1), 171–173.

H. L. Ip, R. & W. K., L. (2017), 'On some Matérn covariance functions for spatio-temporal random fields', *Statistica Sinica* **27**(2), 805–822.

Hahsler, M., Hornik, K. & Buchta, C. (2008), 'Getting things in order: an introduction to the R package seriation', *J3ournal of Statistical Software* **25**.

Hall, R., Erdélyi, R., Hanna, E., Jones, J. M. & Scaife, A. A. (2015), 'Drivers of North Atlantic polar front jet stream variability', *International Journal of Climatology* **35**(8), 1697–1720.

Hallam, S., Josey, S. A., McCarthy, G. D. & Hirschi, J. J. (2022), 'A regional (land–ocean) comparison of the seasonal to decadal variability of the Northern Hemisphere jet stream 1871–2011', *Climate Dynamics* **59**(1), 1897–1918.

Harrison, S. (2014), 'Never mind the gap: climate, rather than insularity, may limit Ireland's species richness', *The Irish Naturalists' Journal* **33**(1), 107–123.

Hastie, T., Tibshirani, R. & Friedman, J. (2001), *The Elements of Statistical Learning*, Springer, New York.

Hawkins, S. J., Sugden, H. E., Mieszkowska, N., Moore, P. J., Poloczanska, E., Leaper, R., Herbert, R. J. H., Genner, M. J., Moschella, P. S., Thompson, R. C., Jenkins, S. R., Southward, A. J. & Burrows, M. T. (2009), 'Consequences of climate-driven biodiversitychanges for ecosystem functioning of North European rocky shores', *Marine Ecology Progress Series* **396**(1), 245–259.

Hazra, A., Huser, R. & Bolin, D. (2023), 'Realistic and fast modeling of spatial extremes over large geographical domains', *arXiv preprint arXiv:2112.10248* pp. 1–36.

He, Y., Fang, J., Xu, W. & Shi, P. (2022), 'Substantial increase of compound droughts and heatwaves in wheat growing seasons worldwide', *International Journal of Climatology* **42**(10), 5038–5054.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J. N. (2020), 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society* **146**(730), 1999–2049.

Hickey, K. (2011), 'The historic record of cold spells in Ireland', *Irish Geography* **44**(2-3), 303–321.

Hiscock, K., Southward, A., Tittley, I. & Hawkins, S. (2004), 'Effects of changing temperature on benthic marine life in Britain and Ireland', *Aquatic Conservation: Marine and Freshwater Ecosystems* **14**(4), 333–362.

Hobday, A. J., Oliver, E. C., Gupta, A. S., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Holbrook, N. J., Moore, P. J., Thomsen, M. S., Wernberg, T. &

Smale, D. A. (2018), 'Categorizing and naming marine heatwaves', *Oceanography* **31**(2 Special Issue), 162–173.

Hooker, K. V., Coxon, C. E., Hackett, R., Kirwan, L. E., O'Keeffe, E. & Richards, K. (2008), 'Evaluation of cover crop and reduced cultivation for reducing nitrate leaching in Ireland', *Journal of Environmental Quality* **37**(1), 138–145.

Hoskins, B. J. & James, I. N. (2014), *Fluid Dynamics of the Midlatitude Atmosphere*, John Wiley & Sons, Ltd.

Huang, J., Mendoza, B., Daniel, J. S., Nielsen, C. J., Rotstayn, L. & Wild, O. (2013), 'Anthropogenic and natural radiative forcing', *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* **9781107057**, 659–740.

Huser, R. & Davison, A. C. (2013), 'Composite likelihood estimation for the Brown-Resnick process', *Biometrika* **100**(1), 1–9.

Huser, R. & Davison, A. C. (2014), 'Space-time modelling of extreme events', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 439–461.

Huser, R. & Genton, M. G. (2016), 'Non-stationary dependence structures for spatial extremes', *Journal of Agricultural, Biological, and Environmental Statistics* **21**(3), 470–491.

Huser, R., Opitz, T. & Thibaud, E. (2017), 'Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures', *Spatial Statistics* **21**(Part A), 166–186.

Huser, R. & Wadsworth, J. L. (2022), 'Advances in statistical modeling of spatial extremes', *WIREs Computational Statistics* **14**(1), e1537.

Hüsler, J. & Reiss, R.-D. (1989), 'Maxima of normal random vectors: between independence and complete dependence', *Statistics and Probability Letters* **7**(1), 283–286.

IFRC (2022), Extreme heat: preparing for the heatwaves of the future, Technical report, United Nations Office for the Coordination of Humanitarian Affairs, Geneva, Switzerland.

IPCC (2021), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, *in* V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu & B. Zhou, eds, 'Cambridge University Press', Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 2061–2086.

IPCC (2023), AR6 Synthesis Report: Climate Change 2023, Technical report, Intergovernmental Panel on Climate Change, Interlaken, Switzerland.

Irish Medical Organisation (2020), 'Chronic shortage of doctors will have devastating implications if not addressed'.
URL: *https://imo.ie/news-media/news-press-releases/2020/chronic-shortage-of-docto/index.xml*

Jackson, L. C., Kahana, R., Graham, T., Ringer, M. A., Woollings, T., Mecking, J. V. & Wood, R. A. (2015), 'Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM', *Climate Dynamics* **45**(11-12), 3299–3316.

Jean-Baptiste, A. & Jean-Luc, B.-K. (2014), Analysis of continuous time series in urban hydrology: filling gaps and data reconstitution, *in* 'Proceedings of the METMA VII and GRASPA14 conference', Torino, Italy.

Jiménez-Hernández, J. D. C., López-Cerino, M. & Aguirre-Salado, A. I. (2020), 'A Bayesian hierarchical model for the spatial analysis of Carbon Monoxide pollution extremes in Mexico City', *Mathematical Problems in Engineering* **2020**, 1–11.

Joe, H. (1994), 'Multivariate extreme-value distributions with applications to environmental data', *Canadian Journal of Statistics* **22**(1), 47–64.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, CRC press, New York.

Kelleher, O. (2023), 'Supply of vegetables to Ireland disrupted by poor weather and energy costs'.
URL: *https://irishtimes.com/food/2023/02/21/supply-of-vegetables-to-ireland-disrupted-by-poor-weather-and-energy-costs/*

Kennedy, J. J., Rayner, N. A., Atkinson, C. P. & Killick, R. E. (2019), 'An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set', *Journal of Geophysical Research: Atmospheres* **124**(14), 7719–7763.

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P. & Gray, L. J. (2015), 'Stratospheric influence on tropospheric jet streams, storm tracks and surface weather', *Nature Geoscience* **8**(6), 433–433.

Kim, S. E., Hashizume, M., Armstrong, B., Gasparrini, A., Oka, K., Hijioka, Y., Vicedo-Cabrera, A. M. & Honda, Y. (2023), 'Mortality risk of hot nights: a nationwide population-based retrospective study in Japan', *Environmental Health Perspectives* **131**(5), 057005.

Kotz, S. & Nadarajah, S. (2000), *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

Leadbetter, M. R. (1991), 'On a basis for 'Peaks over Threshold' modeling', *Statistics & Probability Letters* **12**(4), 357–362.

Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes.*, Springer, New York.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., Mcinnes, K., Risbey, J., Schuster, S., Jakob, D. & Stafford-Smith, M. (2014), 'A compound event framework for understanding extreme impacts', *Wiley Interdisciplinary Reviews: Climate Change* **5**(1), 113–128.

Li, J., Wasko, C., Johnson, F., Evans, J. P. & Sharma, A. (2018), 'Can regional climate modeling capture the observed changes in spatial organization of extreme storms at higher temperatures?', *Geophysical Research Letters* **45**(9), 4475–4484.

Lindgren, F. & Håvard, R. (2015), 'Bayesian spatial modelling with R-INLA', *Journal of Statistical Software* **63**(19), 1–25.

Liou, K. N. (2002), Solar radiation at the top of the atmosphere, *in* 'An Introduction to Atmospheric Radiation', number C, Academic Press, pp. 37–64.

Lugrin, T., Tawn, J. A. & Davison, A. C. (2021), 'Sub-asymptotic motivation for new conditional multivariate extreme models', *Stat* **10**(1), e401.

Lyon, C., Saupe, E. E., Smith, C. J., Hill, D. J., Beckerman, A. P., Stringer, L. C., Marchant, R., McKay, J., Burke, A., O'Higgins, P., Dunhill, A. M., Allen, B. J., Riel-Salvatore, J. & Aze, T. (2022), 'Climate change research and action must look beyond 2100', *Global Change Biology* **28**(2), 349–361.

Madley-Dowda, P., Hughes, R., Tilling, K. & Heron, J. (2019), 'The proportion of missing data should not be used to guide decisions on multiple imputation', *Journal of Epidemiology* **110**(1), 63–73.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A. & Mearns, L. O. (2017), 'Towards process-informed bias correction of climate change simulations', *Nature Climate Change* **7**(11), 764–773.

Massicotte, P., South, A. & Hufkens, K. (2023), 'naturalearth: World map data from natural earth'.

Mateus, C. & Coonan, B. (2022), Isotherman maps of maximum and minimum shade air temperatures in Ireland, Technical report, Met Éireann, Dublin.

Matthes, H., Rinke, A. & Dethloff, K. (2015), 'Recent changes in Arctic temperature extremes: warm and coldspells during winter and summer', *Environmental Research Letters* **10**, 114020.

McElwain, L. & Sweeney, J. (2003), 'Climate change in Ireland- recent trends in temperature and precipitation', *Irish Geography* **36**(2), 97–111.

McElwain, L. & Sweeney, J. (2007), Key Meteorological Indicators of Climate Change in Ireland, Technical report, Environmental Protection Agency, Wexford, Ireland.

Met Éireann (2019), Winter report 2018/19, Technical report, Met Éireann, Dublin.

Met Office (2012), 'Met Office Integrated Data Archive System (MIDAS) land and marine surface stations data (1853-current)', *NCAS British Atmospheric Data Centre* **2**, 1–30.

Mishra, S. & Khare, D. (2014), 'On comparative performance of multiple imputation methods for moderate to large proportions of missing data in clinical trials: a simulation study', *Article in Journal of Medical Statistics and Informatics* **2**(9).

Monteiro, A., Ankrah, J., Madureira, H., Pacheco, M. O., Perry, J., Monteiro, A., Ankrah, J., Madureira, H. & Pacheco, M. O. (2022), 'Climate risk mitigation and adaptation concerns in urban areas: a systematic review of the impact of IPCC assessment reports', *Climate* **10**(8), 115.

Moon, W., Kim, B. M., Yang, G. H. & Wettlaufer, J. S. (2022), 'Wavier jet streams driven by zonally asymmetric surface thermal forcing', *Proceedings of the National Academy of Sciences of the United States of America* **119**(38), e2200890119.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J., Osborn, T. J., Jones, P. D. & Simpson, I. R. (2021), 'An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set', *Journal of Geophysical Research: Atmospheres* **126**(3), 1–28.

Murage, P., Hajat, S. & Sari Kovats, R. (2017), 'Effect of night-time temperatures on cause and age-specific mortality in London', *Environmental Epidemiology* **1**(2), e005.

Murray, J. S. (2017), 'Log-Linear Bayesian additive regression trees for multinomial logistic and count regression models', *Journal of the American Statistical Association* **116**(534), 756–769.

Natural Environment Research Council (1975), Flood Studies Report, Technical report, Natural Environment Research Council, London.

Naveau, P., Huser, R., Ribereau, P. & Hannart, A. (2016), 'Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection', *Water Resources Research* **52**(4), 2753–2769.

Nelder, J. A. & Mead, R. (1965), 'A simplex method for function minimization', *The Computer Journal* **7**(4), 308–313.

NOAA (2023), Another Year of Record Heat for the Oceans, Technical report, National Centers for Environmental Information.

Nolan, P. & Flanagan, J. (2020), High-resolution Climate Projections for Ireland-A Multi-model Ensemble Approach, Technical report, EPA, Dublin.

Northrop, P. J., Attalides, N. & Jonathan, P. (2017), 'Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **66**(1), 93–120.

Northrop, P. J. & Jonathan, P. (2011), 'Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights', *Environmetrics* **22**(7), 799–809.

NPWS (2019), The Status of EU Protected Habitats and Species in Ireland, Technical report, Department of Culture, Heritage and the Gaeltacht, Dublin.

Ó Cléirigh, B. (2020), Energy Security in Ireland: 2020 Report, Technical report, Sustainable Energy Authority of Ireland, Dublin.

Opitz, T., Allard, D. & Mariethoz, G. (2021), 'Semi-parametric resampling with extremes', *Spatial Statistics* **42**, 100445.

Opitz, T., Huser, R., Bakka, H. & Rue, H. (2018), 'INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles', *Extremes* **21**(3), 441–462.

Orlowsky, B. & Seneviratne, S. I. (2012), 'Global changes in extreme events: regional and seasonal dimension', *Climatic Change* **110**(1), 669–696.

Osborn, T. J., Jones, P. D., Lister, D. H., Morice, C. P., Simpson, I. R., Winn, J. P., Hogan, E. & Harris, I. C. (2021), 'Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 data set', *Journal of Geophysical Research: Atmospheres* **126**(2), e2019JD032352.

Osland, M. J., Stevens, P. W., Lamont, M. M., Brusca, R. C., Hart, K. M., Waddle, J. H., Langtimm, C. A., Williams, C. M., Keim, B. D., Terando, A. J., Reyier, E. A., Marshall, K. E., Loik, M. E., Boucek, R. E., Lewis, A. B. & Seminoff,

J. A. (2021), 'Tropicalization of temperate ecosystems in North America: The northward range expansion of tropical organisms in response to warming winter temperatures', *Global Change Biology* **27**(13), 3009–3034.

O'Sullivan, J., Sweeney, C. & Parnell, A. C. (2020), 'Bayesian spatial extreme value analysis of maximum temperatures in County Dublin, Ireland', *Environmetrics* **31**(5), e2621.

Otto, F. E., Skeie, R. B., Fuglestvedt, J. S., Berntsen, T. & Allen, M. R. (2017), 'Assigning historic responsibility for extreme weather events', *Nature Climate Change* **7**(11), 757–759.

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

Palacios-Rodríguez, F., Toulemonde, G., Carreau, J. & Opitz, T. (2020), 'Generalized Pareto processes for simulating space-time extreme events: an application to precipitation reanalyses', *Stochastic Environmental Research and Risk Assessment* **34**(12), 2033–2052.

Papastathopoulos, I. & Tawn, J. A. (2013), 'Extended generalised Pareto models for tail estimation', *Journal of Statistical Planning and Inference* **143**(1), 131–143.

Pascal, M., Sweeney, J., Cullen, E., Schwartz, J. & Goodman, P. (2013), 'Heatwaves and mortality in Ireland, planning for the future', *Irish Geography* **46**(3), 203–211.

Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., Pedersen, T. L., Baston, D. & Dunnington, D. (2023), 'sf: Simple features for R'.

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* **10**(3), 119–131.

Pickands, J. (1981), Multivariate extreme value distributions, *in* 'Proc. 43rd Session I.S.I.', Buenos Aires, pp. 859–878.

Politis, D. N. & Romano, J. P. (1994), 'The stationary bootstrap', *Journal of the American Statistical Association* **89**(428), 1303–1313.

Qasmi, S. (2023), 'Past and future response of the North Atlantic warming hole to anthropogenic forcing', *Earth System Dynamics* **14**(3), 685–695.

R Core Team (2023), 'R: a language and environment for statistical computing'. **URL:** *https://www.R-project.org/*

Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S. & Schaffernicht, E. J. (2015), 'Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation', *Nature Climate Change* **5**(5), 475–480.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T. & Laaksonen, A. (2022), 'The Arctic has warmed nearly four times faster than the globe since 1979', *Communications Earth & Environment 2022 3:1* **3**(1), 1–10.

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

Richards, J., Tawn, J. A. & Brown, S. (2022), 'Modelling extremes of spatial aggregates of precipitation using conditional methods', *Annals of Applied Statistics* **16**(4), 2693–2713.

Richards, J., Tawn, J. A. & Brown, S. J. (2023), 'Joint estimation of extreme precipitation aggregates at different spatial scales through mixture modelling', *Spatial Statistics* **53**, 100725.

Richards, J. & Wadsworth, J. L. (2021), 'Spatial deformation for nonstationary extremal dependence', *Environmetrics* **32**(5), 1–22.

Ritchie, H., Roser, M. & Rosado, P. (2020), 'CO2 and Greenhouse Gas Emissions'. **URL:** *https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions*

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. & Dormann, C. F. (2017), 'Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure', *Ecography* **40**(8), 913–929.

Rohan, P. K. (1986), *The climate of Ireland*, 2 edn, Stationery Office, Dublin.

Rohmer, J., Thieblemont, R. & Le Cozannet, G. (2021), 'Revisiting the link between extreme sea levels and climate variability using a spline-based non-stationary extreme value analysis', *Weather and Climate Extremes* **33**(1), 100352.

Rootzén, H. & Tajvidi, N. (2006), 'Multivariate generalized Pareto distributions', *Bernoulli* **12**(5), 917–930.

Sabater, S., Freixa, A., Jiménez, L., López-Doval, J., Pace, G., Pascoal, C., Perujo, N., Craven, D. & González-Trujillo, J. D. (2023), 'Extreme weather events threaten biodiversity and functions of river ecosystems: evidence from a meta-analysis', *Biological Reviews* **98**(2), 450–461.

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

Schlather, M. (2002), 'Models for stationary max-stable random fields', *Extremes* **5**(1), 33–44.

Schneider, A. & Breitner, S. (2016), 'Temperature effects on health - current findings and future implications.', *EBioMedicine* **6**(1), 29–30.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J. & Brenning, A. (2019), 'Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data', *Ecological Modelling* **406**, 109–120.

Screen, J. A. & Simmonds, I. (2014), 'Amplified mid-latitude planetary waves favour particular regional weather extremes', *Nature Climate Change 2014 4:8* **4**(8), 704–709.

Sen Gupta, A. & McNeil, B. (2012), Variability and change in the ocean, *in* 'The Future of the World's Climate', Elsevier, pp. 141–165.

Sharkey, P. & Winter, H. C. (2019), 'A Bayesian spatial hierarchical model for extreme precipitation in Great Britain', *Environmetrics* **30**(1), 1–19.

Sheehan, A. & O'Sullivan, R. (2020), Ageing and Public Health – an overview of key statistics in Ireland and Northern Ireland, Technical report, Institute of Public Health, Dublin.

Shin, J. Y., Lee, T., Park, T. & Kim, S. (2019), 'Bias correction of RCM outputs using mixture distributions under multiple extreme weather influences', *Theoretical and Applied Climatology* **137**(1-2), 201–216.

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

Sibuya, M. (1960), 'Bivariate extreme statistics', *Annals of the Institute of Statistical Mathematics* **11**(1), 195–210.

Simpson, E. S., Opitz, T. & Wadsworth, J. L. (2023), 'High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields', *Extremes* .

Simpson, E. S. & Wadsworth, J. L. (2021), 'Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures', *Spatial Statistics* **41**, 100482.

Skendžić, S., Zovko, M., Živković, I. P., Lešić, V. & Lemić, D. (2021), 'The impact of climate change on agricultural insect pests'.

Sklar, M. (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229–237.

Smith, R. L. (1984), Threshold methods for sample extremes, *in* J. T. de Oliveira, ed., 'Statistical Extremes and Applications', Vol. 131, Springer, Dordrecht, pp. 621–638.

Smith, R. L. (1990), Max-stable Processes and Spatial Extremes, Technical report, University of Surrey, Guildford.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M. & Santos, J. A. (2018), 'European temperature responses to blocking and ridge regional patterns', *Climate Dynamics* **50**, 457–477.

Stendel, M., Francis, J., White, R., Williams, P. D. & Woollings, T. (2021), The jet stream and climate change, *in* 'Climate Change: Observed Impacts on Planet Earth, Third Edition', Elsevier, pp. 327–357.

Sun, X., Ren, G., You, Q., Ren, Y., Xu, W., Xue, X., Zhan, Y., Zhang, S. & Zhang, P. (2019), 'Global diurnal temperature range (DTR) changes since 1901', *Climate Dynamics* **52**(5-6), 3343–3356.

Sweeney, J. (2020), Climate change in Ireland: science, impacts and adaptation, *in* D. Robbins, D. Torney & P. Brereton, eds, 'Ireland and the Climate Crisis', 1 edn, Palgrave Macmillan, pp. 15–36.

Sweeney, J., Donnelly, A., Mcelwain, L. & Jones, M. (2002), Climate change indicators for Ireland, Technical report, Environmental Protection Agency, Dublin.

Tajvidi, N. (1996), Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions, PhD thesis, University of Gothenburg, Gothenburg.

Tawn, J. A. (1988), 'Bivariate extreme value theory: Models and estimation', *Biometrika* **75**(3), 397–415.

The Royal Society (2014), Resilience to Extreme Weather, Technical report, The Royal Society, London.

Thibaud, E. & Opitz, T. (2015), 'Efficient inference and simulation for elliptical Pareto processes', *Biometrika* **102**(4), 855–870.

Thompson, D. W. & Wallace, J. M. (1998), 'The Arctic oscillation signature in the wintertime geopotential height and temperature fields', *Geophysical Research Letters* **25**(9), 1297–1300.

Thompson, D. W. & Wallace, J. M. (2001), 'Regional climate impacts of the Northern Hemisphere annular mode', *Science* **293**(5527), 85–89.

Thorne, P. W., Menne, M. J., Williams, C. N., Rennie, J. J., Lawrimore, J. H., Vose, R. S., Peterson, T. C., Durre, I., Davy, R., Esau, I., Klein-Tank, A. M. & Merlone, A. (2016), 'Reassessing changes in diurnal temperature range: A new data set and characterization of data biases', *Journal of Geophysical Research* **121**(10), 5115–5137.

Tian, M., El Khoury, R. & Alshater, M. M. (2023), 'The nonlinear and negative tail dependence and risk spillovers between foreign exchange and stock markets

in emerging economies', *Journal of International Financial Markets, Institutions and Money* **82**, 101712.

Todorovic, P. & Rousselle, J. (1971), 'Some problems of flood analysis', *Water Resources Research* **7**(5), 1144–1150.

Todorovic, P. & Zelenhasic, E. (1970), 'A stochastic model for flood analysis', *Water Resources Research* **6**(6), 1641–1648.

van der Wiel, K. & Bintanja, R. (2021), 'Contribution of climatic changes in mean and variability to monthly temperature and precipitation extremes', *Communications Earth & Environment* **2**(1), 1.

Van Oldenborgh, G. J., Mitchell-Larson, E., Vecchi, G. A., De Vries, H., Vautard, R. & Otto, F. (2019), 'Cold waves are getting milder in the northern midlatitudes', *Environmental Research Letters* **14**(11).

Vignotto, E., Engelke, S. & Zscheischler, J. (2021), 'Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland', *Weather and Climate Extremes* **32**(1), 100318.

Vihma, T., Graversen, R., Chen, L., Handorf, D., Skific, N., Francis, J. A., Tyrrell, N., Hall, R., Hanna, E., Uotila, P., Dethloff, K., Karpechko, A. Y., Björnsson, H. & Overland, J. E. (2020), 'Effects of the tropospheric large-scale circulation on European winter temperatures during the period of amplified Arctic warming', *International Journal of Climatology* **40**(1), 509–529.

Von Mises, R. (1936), 'La distribution de la plus grande de n valuers', *Rev. math. Union interbalcanique* pp. 141–160.

Vose, R., Easterling, D., Kunkel, K., LeGrande, A. & Wehner, M. (2017), Temperature changes in the United States. Climate Science Special Report: Fourth National Climate Assessment, Volume I, Technical report, U.S. Global Change Research Program, Washington, DC.

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

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

Walsh, S. (2017), Long-term rainfall averages for Ireland, 1981-2010, Technical report, Met Éireann, Dublin.

Wazneh, H., Arain, M. A., Coulibaly, P. & Gachon, P. (2020), 'Evaluating the dependence between temperature and precipitation to better estimate the risks of concurrent extreme weather events', *Advances in Meteorology* **12**, 8763631–16.

White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., Harley, C. D. G., Henderson, S. B., Jakob, M., Lau, C.-A., Mareshet Admasu, L., Narinesingh, V., Rodell, C., Roocroft, E., Weinberger, K. R. & West, G. (2023), 'The unprecedented Pacific Northwest heatwave of June 2021', *Nature Communications* **14**(1), 727.

Wingler, A., Cawkwell, F., Holloway, P., Misra, G., De La, R., Cerro, T. & Sweeney, C. (2022), PhenoClimate: Impact of Climate Change on Phenology in Ireland, Technical report, Environmental Protection Agency, Dublin.

Winter, H. C. & Tawn, J. A. (2016), 'Modelling heatwaves in central France: a case-study in extremal dependence', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(3), 345–365.

Winter, H. C., Tawn, J. A. & Brown, S. J. (2016), 'Modelling the effect of the El Niño-southern oscillation on extreme spatial temperature events over Australia', *Annals of Applied Statistics* **10**(4), 2075–2101.

Wood, S. N. (2006), *Generalized Additive Models: an Introduction with R*, Chapman and Hall/CRC, New York.

Wu, R., Hamshaw, S. D., Yang, L., Kincaid, D. W., Etheridge, R. & Ghasemkhani, A. (2022), 'Data imputation for multivariate time series sensor data with large gaps of missing data', *IEEE Sensors Journal* **22**(11), 10671–10683.

Yang, Q., Zhao, Y., Wen, Q., Yao, J. & Yang, H. (2018), 'Understanding Bjerknes compensation in meridional heat transports and the role of freshwater in a warming climate', *Journal of Climate* **31**(12), 4791–4806.

Yin, J. & Zhao, M. (2021), 'Influence of the Atlantic meridional overturning circulation on the U.S. extreme cold weather', *Communications Earth and Environment* **2**(1), 218.

Youngman, B. D. (2022), 'evgam: An R package for generalized additive extreme value models', *Journal of Statistical Software* **103**(3), 1–26.

Yozgatligil, C., Aslan, S., Iyigun, C. & Batmaz, I. (2013), 'Comparison of missing value imputation methods in time series: The case of Turkish meteorological data', *Theoretical and Applied Climatology* **112**(1-2), 143–167.

Yu, K. & Moyeed, R. A. (2001), 'Bayesian quantile regression', *Statistics and Probability Letters* **54**(4), 437–447.

Zamo, M. & Naveau, P. (2018), 'Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts', *Mathematical Geosciences* **50**(2), 209–234.

Zeka, A., Browne, S., McAvoy, H. & Goodman, P. (2014), 'The association of cold weather and all-cause and cause-specific mortality in the island of Ireland between 1984 and 2007', *Environmental Health: A Global Access Science Source* **13**(1), 1–9.

Zhang, L., Shaby, B. A. & Wadsworth, J. L. (2022), 'Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations', *Journal of the American Statistical Association* **117**(539), 1357–1369.

Zhong, P., Brunner, M., Opitz, T. & Huser, R. (2022), 'Spatial modeling and future projection of extreme precipitation extents', *arXiv preprint arXiv:2212.03028* pp. 1–35.

Zhong, P., Huser, R. & Opitz, T. (2022), 'Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude', *The Annals of Applied Statistics* **16**(1), 272–299.

Zhou, P. & Liu, Z. (2018), 'Likelihood of concurrent climate extremes and variations over China', *Environmental Research Letters* **13**(9), 094023.