# Extreme Value Modelling of Heatwaves

Hugo Charles Winter

Submitted for the degree of Doctor of
Philosophy at Lancaster University.

September 2015

# Abstract

Since the turn of the century record temperatures have been observed in at least 20 different countries across Europe. Isolated hot days are not often an issue; most devastation occurs when hot temperatures persist over many days. For example, the 2003 heatwave over Europe caused 40,000 deaths over a four week period at a cost of € 13.1 million to the agriculture sector. It is clear that accurate models for the risks associated with heatwaves are important to decision makers and planners who wish to reduce the number of people affected by these extreme events.

Extreme value theory provides a statistical framework for modelling extreme events. Extreme value models for temperature data tend to focus solely on the intensity, overlooking how long periods of hot weather will last and what the spatial extent of the event will be. For heatwaves, it is vital to explicitly model extremal dependence in time and space.

An aim of this thesis is to develop extreme value methods that can accurately capture the temporal evolution of heatwaves. Specifically, this is the first to use a broad class of asymptotically motivated dependence structures that can provide accurate inferences for different types of extremal dependence and over different orders of lagged dependence. This flexibility ensures that these models are less likely to dramatically under- or over-estimate the risks of heatwave events. Climate change is now widely regarded as a driving force behind increased global temperatures. Extending the extreme value

heatwave models to include covariate structure permits answers to critical questions such as: *How will a 1ºC warming in the global temperature increase the chance of a 2003 style event?*

The 2009 heatwave over Australia highlighted issues posed when multiple cities are affected simultaneously. Both Adelaide and Melbourne observed record temperatures during the same event which led to 374 deaths and 2000 people being treated for heat related illness. It is not enough for heatwave models to account for temporal dependence, they also need to explicitly model spatial dependence. Large-scale climatic phenomena such as the El Niño-Southern Oscillation are known to affect the temperatures across Australia. This thesis develops new spatial extreme value methods that account for covariates, which are shown to model the 2009 event well. A novel suite of spatial and temporal risk measures is designed to better understand whether these covariates have an effect on the spatial extent and duration of heatwaves. This provides important information for decision makers that is not available using current methodology.

# Acknowledgements

First and foremost I would like to thank Jonathan Tawn, without who this thesis might not exist and certainly would not have been so fun (for want of a better word...) to write. His steady guidance and willing to contribute so much time when he is so busy has been much appreciated and provided great motivation for me to push myself to work harder. I would also like to thank him for the opportunities that he has given me to meet influential people in the extreme value theory community and present my work at the largest conferences in the area.

My supervisor at the Met Office, Simon Brown, has also made vital contributions to the thesis by providing the motivation and application behind the statistical theory. Without this I would have been far less motivated to fight through the morass of theory involved in a statistical thesis. I am honoured to have been his first PhD student and thank him for his infinite patience as I struggled to get acquainted with the meteorological concepts in the thesis and then bombarded him with statistical theory in return!

I would also like to acknowledge the funding from EPSRC through the STOR-i Centre for Doctoral Training. One of the main aims of the STOR-i programme is to have a large and vibrant community of students engaged with cutting edge research. During my stay in Lancaster I have found everyone to be lovely and it has lived up to and beyond my expectations when I started. I would especially like to thank my year

group: Emma, Rhian, Tom, Ben, George, Dave, James and Rob. Without all of you my four years at Lancaster would have been much less fun and the support that each of you have provided through the course of my thesis has been vital.

From a work perspective, I must also give thanks to the research group at EPFL for fantastic hospitality when I visited Lausanne. These trips were very interesting and the discussions I had whilst there were most informative. I would especially like to thank Thomas Lugrin and wish him the best in future work with Jon at Lancaster. I would also like to thank my intern Kathryn Turnbull for the interesting work that she did. I am glad to see that she has accepted an offer to start on the STOR-i programme and it makes me certain that the future of STOR-i is bright.

Away from my work, there are a couple of people I would like to thank. My parents David and Linda have been a great support throughout my time in Lancaster. I doubt I shall ever be able to adequately put into words how much I respect them and appreciate all they have done for me. I am proud to be the first member of the Winter family to go through university, even if my high qualifications in Mathematics have not imbued with better mental arithmetic skills ('real maths') than my father!

Lastly (and by no means least!) I would like to thank my girlfriend Charlotte. Her love and support has been an ever present throughout my time in Lancaster and my undergraduate years in Exeter. I have mentioned patience a lot in these acknowledgements but few have probably had to put up with more than Charlotte. I will be forever appreciative of her decision to move up away to Lancaster and accompany me for the last few years. The thesis has been a roller-coaster of emotions from the crushing lows (luckily few and far between) to the soaring highs and through them all Charlotte has been there to keep me sane. It may sound hyperbolic, but without her I do not think I would be where I am now and the person I am today.

# Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Chapter 3 has been accepted for publication as Winter, H.C. and Tawn, J.A. (2016). Modelling heatwaves in central France: a case study in extremal dependence. *Journal of the Royal Statistical Society: Series C.*

All analysis in this thesis has been undertaken in the statistical computing language R.

Hugo Charles Winter

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Climate modelling

The modelling of weather and climate phenomena has been a very important area of research over the last 60 years and continues to generate many interesting modelling challenges. Progress in the field of climate modelling has occurred in many areas, broadly summarised in Figure 1.1.1. To accurately model the climate system advances in each of these areas have been required. To start with accurate observations of the climate system are needed to inform the behaviour of climate models. The rise of technology has made it easier to collect accurate and more comprehensive observational records. These observations inform the design of climate models based upon in depth knowledge about the physics of the climate. Increased computational power has also permitted the creation of more complicated climate models that attempt to accurately model the behaviour of observed climatic phenomena. These climate observations and models can be combined with a risk assessment model to estimate the risk posed by certain climatic conditions. This information is of vital importance for changing policy and mitigation regarding climatic phenomena.

One recently developing area concerns the use of statistical models to provide ad-
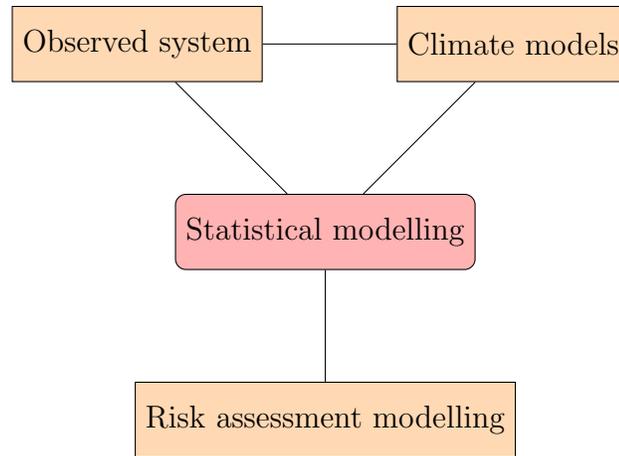
Figure 1.1.1: The main steps of climate modelling. Accurate modelling of the climate system requires development of all areas here. During this thesis we focus on developing new approaches within the class of statistical modelling.

ditional insight when modelling the climate system. These models provide mathematically motivated approaches to estimate the inherent uncertainty in the climate system. With the rapid increase in the amount of available data, statistical methods for big data provide an opportunity to obtain more information. In particular here, statistical models for data obtained from observations and climate models are used to estimate the probability of the occurrence of different climatic phenomena of interest. These estimates can then be used to assess the risk posed by certain important climatic events, using specific models of how climatic phenomena impact upon people discussed previously. Statistical modelling can also be used to provide insight into the reliability of climate models; a process called verification. This area of research is especially important within the climate community since climate models are required to assess potential changes that occur due to climate change.

In this thesis we concentrate on the statistical modelling of particular weather phenomena. In particular we are interested in the behaviour of a group of rare and potentially destructive phenomena called natural hazards. Understanding the risk

posed by natural hazards is especially important for risk assessors since these events lead to higher death tolls and larger economic losses. The statistical modelling of natural hazards is important for a number of different specialists within the field of climate science. Most obviously, decision makers would like to know how likely certain types of natural hazards are to permit better preparation and mitigation. Statistics can also provide additional insight for climate modellers. Currently, extremes of climatic phenomena are often poorly represented by climate models. This revelation is especially troubling when policy decisions regarding future climate change need to be made based upon results coming from climate models. The information obtained from statistical models could be used to improve the representation of extreme values in climate models and inform governmental policy.

The statistical modelling of any type of rare event represents a challenge for statisticians since by definition rare events do not occur often within the observational record. As such approaches are required that can be applied in situations where data are sparse. We would also like to be able to estimate events that are larger than have been observed previously. Models that permit extrapolation to very extreme levels are thus required. One well researched statistical approach for modelling rare events is extreme value theory. At the most basic level values that are extreme, usually defined as either the maximum of some block of time or as exceedances of some critical level, are modelled and used to estimate the behaviour of extreme values. Many different types of natural hazard can be modelled by extreme value theory. These can range from hazards like tornadoes that occur on a short timescale to droughts and heatwaves that result from the persistence of certain conditions over long time scales. Throughout the thesis we focus on developing extreme value models that can be used on a wide class of natural hazards. It is noted that in many situations weather generators could be used to model the risks associated with natural hazards. Such approaches will be useful when interest is in moderately extreme events, but are not

so useful when extrapolating to very extreme events. For the interested reader, a modern review of weather generators is given in Maraun et al. (2010).

## 1.2 Heatwave modelling

Throughout the thesis we restrict our attention to heatwaves as the natural hazard of interest. IPCC (2012) provide a general definition of a heatwave event:

> *A set of hot days and/or nights that are associated with a marked short-term increase in mortality.*

This definition is very general but highlights the most important point that any definition of a heatwave is constructed in terms of a critical level and a set of temperature values (often consecutive) that exceed this level. A temperature observed above the critical level, often chosen as a sufficiently high quantile, is said to be extreme; the choice of this level is discussed further in later chapters. However, singular hot days do not often cause many deaths or economic damage. During a heatwave most casualties are caused by heat exhaustion where core body temperature exceeds healthy levels (37-39ºC). This is often caused by multiple consecutive hot days without respite, which does not allow the body to recover. As such, the duration of a heatwave is of great importance and motivates the need for novel statistical approaches that account for this aspect. Figure 1.2.1 shows important heatwave quantities that are referred to throughout the thesis.

It is also important to consider the location and spatial extent of heatwave events. A heatwave event that occurs over an urbanised region is likely to pose different risks than one that occurs over a sparsely populated agricultural region. Estimating the risk of a heatwave event occurring at multiple locations at the same time is also necessary. Heatwaves can often stretch the capabilities of medical services. Estimates of the probability of having a heatwave striking multiple locations could provide vital
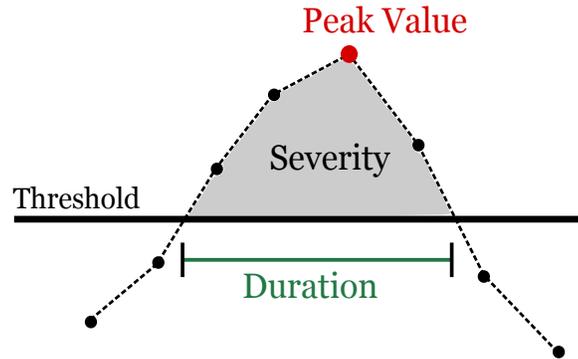
Figure 1.2.1: Definition of heatwave terminology for example event at a single location.

information to mitigate for the potential increase in hospital admissions.

One especially important consideration concerns how heatwave events might change under climate change and other important climatic phenomena. Climate change is now widely regarded as contributing to recent increases in global temperatures. Many papers within the field of climate science have investigated changes in global mean temperatures and shown a warming climate in the future. However, the behaviour of extreme events is less well studied and many results in this area are subject to great uncertainty. As mentioned previously, there is also debate about how well climate models reproduce climatic extremes. For heatwaves, we are most interested in whether heatwaves will occur more frequently or will last longer under climate change. Figure 1.2.2 shows potential changes that could occur: events that become more frequent (right top), events that last longer (right center) and the combination of more frequent and longer heatwave events (right bottom). Each of these situations would require different mitigation and therefore it is important to be able to accurately capture all such changes.
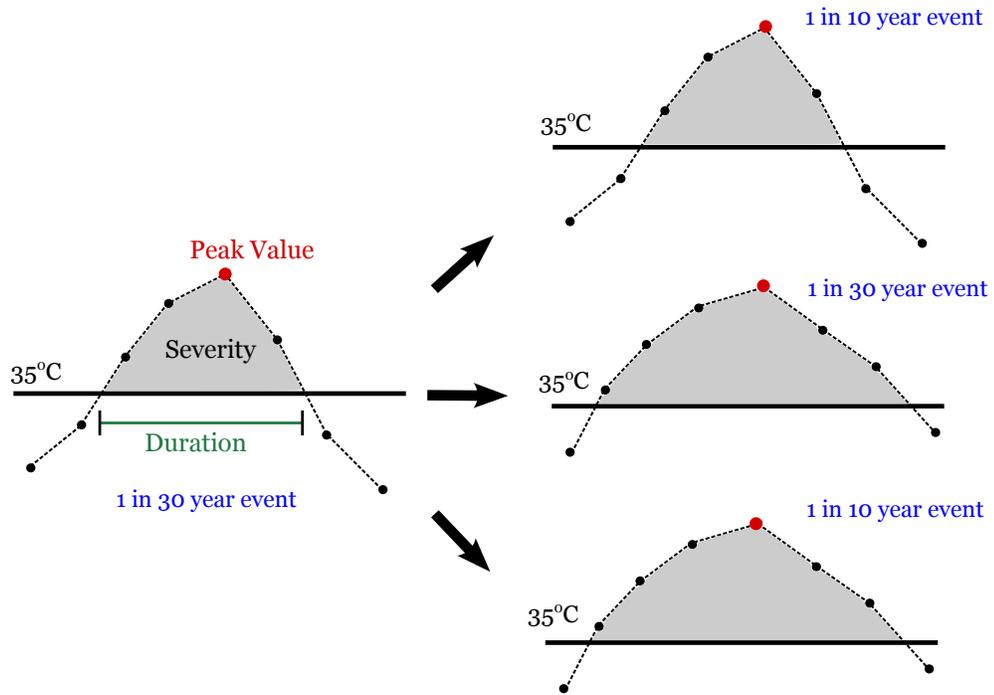
Figure 1.2.2: How the behaviour of heatwave events at a single location could change with future climate change for a example event occurring above a critical level of 35°C. A change in the frequency of heatwaves (right top), a change in the duration of heatwaves (right centre) and a change in the frequency and duration of heatwaves (right bottom).

## 1.3 Thesis outline

The aim of the thesis is to provide statistical methodology, based upon extreme value theory, to model the behaviour of heatwave events. To do this we shall extend basic extreme value techniques to account for the spatial and temporal nature of heatwaves. The thesis is split into different chapters that investigate different aspects of modelling heatwaves using extreme value theory.

Chapter 2 gives an introduction to univariate extreme value theory and sets out important notation in the area. We also review current approaches within the field of multivariate extreme value theory, focusing on the important concept of extremal dependence. This chapter ends with three extensions to current approaches for modelling extremal dependence. One extension aims to take advantage of information not currently used in the estimation of extremal dependence for more efficient estimation of parameters and extremal measures. The second extension provides an approach for estimating important extremal quantities directly without the need of repeated simulation. Finally, if simulation is required, we provide an approach for more accurate estimation of extremal measures in high dimensions.

In Chapter 3 an extreme value model for heatwaves at a single location is built based upon a first-order Markov assumption. This assumption permits the use of bivariate extreme value results and generally simplifies the modelling procedure. We derive an approach for estimating the probability of specific heatwave events which permits a broad set of dependence characteristics. This flexible approach is based upon repeated simulation of heatwave events which can then be used to estimate any measure of interest. This approach is compared against more restrictive pre-existing techniques and then used to estimate the probability of an event being more extreme than the 2003 heatwave event at a site in France. For this study we investigate daily maximum temperature observations taken at Orleans in central France for the years

1946-2012.

The first-order Markov assumption is restrictive and Chapter 4 introduces models for deriving the probability of specific heatwave events at a single site that take advantage of higher-order structure. This approach is based upon a multivariate kernel density estimation approach incorporated into the framework from Chapter 3. Specific diagnostic tests are developed to estimate how much higher-order structure should be included to capture specific heatwave behaviour while avoiding inefficient overparameterisation. Finally the higher-order approach is also used to obtain estimates of the probability of an event more extreme than the 2003 heatwave and these are compared with the results in Chapter 3.

Up until this point of the thesis, the effect of climate change has been ignored. Chapter 5 provides an approach, under the first-order Markov assumption for a single site, for incorporating covariate structure into the approach outlined in Chapter 3. We analyse the behaviour of daily maximum temperature values taken from an ensemble of different general circulation models for the years 2006-2090 forced with the RCP8.5 climate scenario. We investigate the change in the behaviour of heatwaves with a 1°C and 5°C increase in the mean global temperature, estimating whether heatwaves will become more frequent or last longer.

Chapter 6 introduces spatial structure to investigate the spatial extent of hot events over Australia. Here, we ignore the issues raised concerning temporal dependence to focus on a spatial model. The nature of the problem motivates a new set of extremal measures for summarising the spatial extent of extreme temperature events. Using this model we estimate the effect of the El Niño Southern Oscillation (ENSO), a large-scale climatic phenomenon, on the spatial extent of heatwave events. In this chapter, data are gridded daily maximum temperature observations across Australia

for the years 1957-2011, along with a monthly measure of the strength of ENSO.

In Chapter 7 we bring together the models from the previous chapters to create a full space-time model for assessing the risk of heatwave events. We also consider how the methods outlined during the thesis could be applied to drought as opposed to heatwaves. A discussion of the outcomes of the thesis is also included.

The thesis is structured as a set of papers and as such Chapters 3, 4, 5 and 6 can be read as separate entities. As such some important sections outlining the extreme value models used will be repeated in multiple chapters.

# Chapter 2

# Extreme value theory

## 2.1 Introduction

Extreme value theory is a field developed throughout the twentieth century starting from asymptotic arguments derived by Fisher and Tippett (1928) and formalised into statistical methods by Gumbel (1958). The area of extreme value theory is driven by the desire to accurately model the probabilistic behaviour of events that are by definition rare. Framing this problem in terms of the distribution function, the tails of the distribution are of greatest concern. Most data are concentrated in the centre of the distribution, which means that estimates such as the mean and standard deviation are typically driven by these central values. A fit to the body of the data permits many different extrapolations to tail regions, a situation that reduces confidence in estimates of high quantiles and other associated measures. Observations in the tail of the distribution are scarce which makes inference hard via standard methods that rely on having a large sample. There is often the desire to be able to estimate levels that are beyond the range of the current data. Extreme value models provide an asymptotically justified approach for exactly this type of extrapolation.

Section 2.2 reviews existing theory and inference procedures for univariate extreme

value approaches. In Section 2.3 we introduce multivariate extreme value theory and in particular the concept of extremal dependence. Two different approaches for modelling extremal dependence, the joint tail model (Ledford and Tawn, 1997) and the conditional extremes model (Heffernan and Tawn, 2004), are discussed in Sections 2.5 and 5.3.2 respectively. Useful extensions to these approaches are derived and Section 2.7 gives a brief simulation study to compare the two methods and assess the utility of the different extensions.

## 2.2 Univariate Extreme Value Theory

### 2.2.1 Overview

In many situations we are interested in modelling the extremes of a particular variable at a single location. For this purpose, univariate extreme value distributions exist based upon asymptotically derived theory. Two types of modelling approach are discussed below, modelling block maxima and exceedances of a high threshold.

### 2.2.2 Block maxima approach

**Theory**

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (IID) random variables with distribution function $F$. The maximum value is defined as

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

For example, the annual maxima $M_{X,365}$ could be obtained from the daily values $X_1, \ldots, X_{365}$. Any distributional results that concern the minima can be derived from

the results obtained for maxima since

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

As such, only methods for the upper tail are presented in the thesis. The subscript $X$ is also dropped from $M_{X,n}$ for notational simplicity. The distribution of $M_n$ can be derived exactly as

$$\mathrm{P}(M_n \leq z) = \mathrm{P}(X_1 \leq z, \ldots, X_n \leq z)$$
$$= \mathrm{P}(X_1 \leq z) \ldots \mathrm{P}(X_n \leq z)$$
$$= [F(z)]^n.$$

This result is not immediately useful since $F$ is unknown. Instead the behaviour of $F^n$ can be studied as $n \to \infty$. However as $n \to \infty$, $M_n$ tends to the upper end point of $F$; the asymptotic distribution of $M_n$ is termed degenerate. If it is possible to define sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that as $n \to \infty$

$$\mathrm{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z), \tag{2.2.1}$$

where $G$ is a non-degenerate distribution function, then $G$ belongs to the family of extreme value distributions. The Extremal Types Theorem (Leadbetter et al., 1983) gives three different families of limit distribution that maximum values could take (Gumbel, Fréchet and Negative Weibull). Up to type $G$ is of the form of one of the following distributions:

$$G(z) = \exp\{-\exp(-z)\} \qquad \text{for} \quad \infty < z < \infty \tag{2.2.2}$$

$$G(z) = \begin{cases} 0 & z \leq 0 \\ \exp(-z^{-\alpha}) & z > 0, \alpha > 0 \end{cases} \tag{2.2.3}$$

$$G(z) = \begin{cases} \exp[-(-z)^{-\alpha}] & z < 0, \alpha > 0 \\ 1 & z \geq 0. \end{cases} \qquad (2.2.4)$$

Equations (2.2.2), (2.2.3) and (2.2.4) are distributions from the Gumbel, Fréchet and Negative Weibull families respectively. A detailed proof of the theorem is given in Leadbetter et al. (1983). A distribution $G$ is said to be max-stable if for every $n > 0$ there exist constants $A_n > 0$ and $B_n$ such that

$$G(A_n z + B_n) = \{G(z)\}^n.$$

Max-stability is a property that is only satisfied by the Gumbel, Fréchet and Negative Weibull families.

Since it is inconvenient to work with three distinct classes, a parameterisation which unifies the distributions is commonly used. This Generalised Extreme Value (GEV) distribution, often written as $\text{GEV}(\mu, \sigma, \xi)$, is defined as

$$G(z; \mu, \sigma, \xi) = \exp\left\{ -\left[ 1 + \xi\left( \frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, \qquad (2.2.5)$$

where $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$ and $x_+ = \max(x, 0)$. The density function associated with equation (2.2.5) exists on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$. The parameters $(\mu, \sigma, \xi)$ are termed the location, scale and shape parameters respectively. The value of the shape parameter $\xi$ determines the tail behaviour and family of the limit distribution, where

- $\xi > 0$ corresponds to the Fréchet distribution which has a heavy upper tail.

- $\xi = 0$ corresponds to the Gumbel distribution which has an exponential tail.

- $\xi < 0$ corresponds to the Negative Weibull distribution which has a tail with a finite upper limit.

The GEV distribution is used to model the distribution of block maxima. It is assumed that the limit in equation (2.2.1) holds exactly for some finite value of $n$. The assumption relies on the choice of values for $n$ and $k$ which split the data of length $nk$ into $k$ blocks that each contain $n$ data points. A trade-off is required since to assume that limit form in equation (2.2.1) holds exactly for some finite value of $n$, it is necessary to take the maximum of sufficiently many observations, i.e. set $n$ as high as possible. However to ensure as many independent maxima as possible we also wish to set $k$ as high as possible. In many practical applications the length of block $n$ is given by the context, i.e. in many environmental applications taking annual maxima ensures stationarity of the resulting maxima. Methods, either likelihood based or moment based, are then required to obtain estimates for the parameters $(\mu, \sigma, \xi)$.

**Return levels**

When considering extreme values, it is important to make inferences based upon the time to or the severity of the next extreme event. For a stationary series this can be expressed by return levels and return periods. The return period of level $z_p$ is the expected waiting time until $z_p$ is next exceeded. This is related to the $1/p$ year return level which is the level for which the expected waiting time between annual maxima exceedances is $1/p$ years. Therefore, the $1/p$ year return level $z_p$ is the $1 - p$ quantile of the GEV distribution where $0 < p < 1$. An estimate for the $1/p$ return level is thus given as

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

where $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ are the maximum likelihood estimates of $(\mu, \sigma, \xi)$.

### 2.2.3 Threshold exceedance approach

**Motivation**

Up until this point, we have modelled extreme values using the block maxima approach. However this can be a wasteful approach to modelling extreme values that can lead to inefficient statistical procedures. If more than one large value occurs in a block, only the biggest will be used even if these other events are large enough to be called extreme. To see this, note that between the smallest and largest block maxima in the data there are likely to be other observations that are not block maxima. In block maxima approaches these tail values are being ignored despite the fact they are more extreme than some of the block maxima. Figure 2.2.1 shows daily rainfall accumulations from a location in South West England for 1952-1962. The solid dots are the values that come above a threshold at 35mm. It is clear that some blocks contain more than one value above the threshold. Using block maxima these data would be discarded, whereas a threshold method uses this additional information. Methods that do not organise the data into blocks can therefore provide a better alternative.
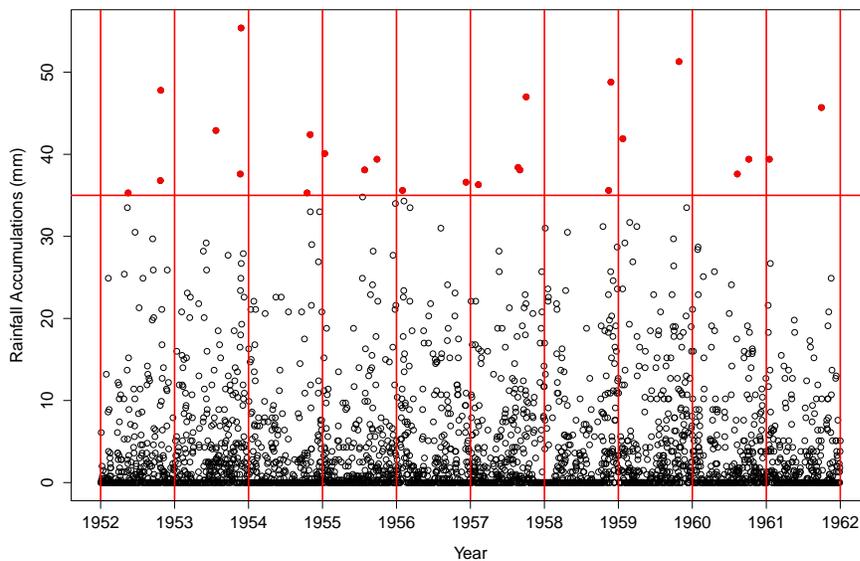


Figure 2.2.1: Rainfall accumulations (mm) at a location in South West England 1952-1962.

**Theory**

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed variables with distribution function $F$. Under the assumptions of the asymptotic theory of equations (2.2.1) and (2.2.5), construct a sequence of point processes $P_1, P_2, \ldots$ on $[0, 1] \times \mathbb{R}$ using $P_n$

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

and examine the limit process as $n \to \infty$. This process is non-degenerate and normalises small points to the same value ($b_l$) whilst retaining all the large points of the process in the limit process. A description of the limit process asymptotically motivates a model for all large values. Under the definition of $P_n$ above, the limiting point process is defined on the set $[0, 1] \times (b_l, \infty)$

$$P_n \to P \qquad \text{as} \quad n \to \infty, \tag{2.2.6}$$

where $P$ is a non-homogeneous Poisson process with intensity

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

for $(t, x) \in [0, 1] \times (b_l, \infty)$. The Poisson process limit result shows that the behaviour of all large values can be determined asymptotically by the characteristics of $a_n$, $b_n$ and $\xi$. Under the conditions for the limit in equation (2.2.6) to hold, Pickands (1975) and Smith (1989) show that for $x > 0$ and $X \sim F$

$$P(X > u_n + a_n x \mid X > u_n) \to \left[ 1 + \xi \frac{x}{\psi} \right]_+^{-1/\xi}, \tag{2.2.7}$$

as $n \to \infty$, where $u_n \to x^F$ as $n \to \infty$ with $x^F$ being the upper endpoint of $F$ and $\psi > 0$ and $\xi \in \mathbb{R}$. The distribution function

$$G(x) = 1 - \left[ 1 + \xi \frac{x}{\psi} \right]_+^{-1/\xi} \qquad x > 0,$$

is known as the Generalised Pareto Distribution, denoted GPD($\psi$, $\xi$), with scale parameter $\psi > 0$ and shape parameter $\xi \in \mathbb{R}$. The limit in equation (2.2.7) shows that

under weak conditions, in the limit as the threshold tends to the upper endpoint of the distribution, the scaled exceedances of the threshold tend to a $\text{GPD}(\psi, \xi)$.

If we assume that the limit in equation (2.2.7) holds exactly for a sufficiently large threshold $u_n$, this gives

$$\text{P}(X > x \mid X > u) = \left(1 + \xi \frac{x - u}{\sigma_u}\right)_{+}^{-1/\xi}, \tag{2.2.8}$$

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

$$X - u \mid X > u \sim \text{GPD}(\sigma_u, \xi). \tag{2.2.9}$$

Note that the absorption of the scaling $a_n$ into the scale parameter of the limiting GPD gives a scale parameter

$$\psi a_n = \psi a_{g(u_n)} =: \sigma_{u_n} = \sigma_u,$$

where $g(u_n) = n$. The GPD has the threshold stability property which states that if $X - u$ is distributed as in equation (2.2.9), for any higher threshold $v > u$

$$X - v \mid X > v \sim \text{GPD}(\sigma_u + \xi(v - u), \xi). \tag{2.2.10}$$

As such $\xi$ is constant with threshold, once a GPD is appropriate, but the scale parameter $\sigma_v = \sigma_u + \xi(v - u)$ is not (Davison and Smith, 1990). The shape parameter of the GPD is equal to the shape parameter of the corresponding GEV distribution defined in equation (2.2.5). This property means that the shape parameter determines the behaviour in the same way as for the GEV distribution.

**Return levels**

Return levels are calculated using a similar process to the block maxima approach, however since the data are conditional upon having exceeded a sufficiently high threshold $u$ we must undo this conditioning by multiplying by the rate of exceedance

$\lambda_u = \mathrm{P}(X > u)$ such that

$$\mathrm{P}(X > x) = \mathrm{P}(X > u)\mathrm{P}(X > x \mid X > u)$$

$$= \lambda_u \left[1 + \xi \left(\frac{x - u}{\sigma_u}\right)\right]_+^{-1/\xi}, \qquad (2.2.11)$$

for $x > u$. The unconditional probability distribution function given in equation (2.2.11) can be inverted to give the return value that is exceeded once every $m$ observations

$$x_m = \begin{cases} u + \sigma_u/\xi \left[(m\lambda_u)^\xi - 1\right] & \text{for } \xi \neq 0 \\ u + \sigma_u \log(m\lambda_u) & \text{for } \xi = 0, \end{cases}$$

where $m$ must be sufficiently large, i.e. $m > \lambda_u^{-1}$, to ensure that $x_m > u$.

**Threshold choice**

When modelling data that come in the form of threshold exceedances, an important issue concerns the choice of threshold $u$. The choice of the threshold $u$ directly affects the number of threshold exceedances $n_u$ and creates a bias-variance trade-off. Setting a low threshold increases the amount of data that can be used, i.e. increases $n_u$, which makes the statistical inference more efficient by reducing uncertainty in the estimation of the model parameters. However, the asymptotic arguments that motivate the use of threshold approaches will break down if the threshold is set too low, thus introducing bias. The choice of threshold needs to balance these two opposing demands. Choice of threshold using diagnostics is an area of research that is of much interest (Tancredi et al. (2006), Wadsworth and Tawn (2012a)). The application of interest may motivate a sensible choice of threshold. However there are no specific rules about choosing the best threshold, with threshold selection diagnostics available to infer threshold choice. Two commonly used diagnostic plots are mean residual life (MRL) plots and parameter stability plots.

In order to construct the MRL plot we assume that for a given set of threshold

exceedances $X_i \mid X_i > u$, for $i = 1, \ldots, n_u$, that the $X_i$ follow a generalised Pareto distribution with scale parameter $\sigma_u$ and shape parameter $\xi$. Then the expected value of the threshold excesses $X - u$ is given as

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

if $\xi < 1$. By considering a higher threshold $v > u$, we also derive the expectation

$$\mathbb{E}\left[X - v \mid X > v\right] = \frac{\sigma_u + \xi(v - u)}{1 - \xi}, \tag{2.2.12}$$

if $\xi < 1$. In order for the threshold to be suitable for modelling data points above, the mean excesses given in equation (2.2.12) should be linear in $v$ for all $v > u$ if $u$ is large enough.

Another method that is used to select the threshold is the parameter stability plot. If $X$ follows a generalised Pareto distribution above the threshold $u$ as in equation (2.2.9) then for any higher threshold $v \geq u$, the $X$ above the higher threshold $v$ has a GPD distribution as in equation (2.2.10). From equation (2.2.10) the shape parameter is found to be constant for the higher threshold but the scale parameter is threshold variant. So that we are able to assess parameter stability we work with the modified scale $\sigma^*$ for the higher threshold $v$, where the modified scale $\sigma^* = \sigma_v - \xi v$, this reparameterisation results in the modified scale being threshold invariant. The choice of the threshold from the parameter stability is determined by the selecting the lowest possible value of $u$ for which both the estimates of the modified scale and shape parameter remains constant (excluding sampling variability) above this level.

## 2.3 Multivariate extreme value theory

### 2.3.1 Motivation

In previous sections it has been assumed that a set of observations have been obtained from an IID set of random variables. However for many types of data this assumption

is not realistic. For example, if certain weather conditions have occurred on a given day it is likely that the conditions on the following day will be closely related to the conditions on the current day. To model these situations more realistically there exists methodology to deal with multivariate extremes that is outlined in the rest of the section. Two types of modelling approach are discussed below, modelling using componentwise maxima of multiple variables, a natural extension of the block maxima framework, and multivariate threshold approaches, an extension of univariate threshold approaches. In Section 2.3.2 both approaches are outlined; throughout the rest of the thesis we focus on multivariate threshold approaches since these are more efficient in their use of data, permit a broader class of dependence structures and enable estimation of any joint tail feature unlike with componentwise maxima.

## 2.3.2 Theory

**Componentwise maxima**

Let $(X_{j,1}, \ldots, X_{j,d})$, where $j = 1, \ldots, n$, be a collection of $d$-dimensional vectors which for each $j$ has the joint distribution function $G$ and is independent over $j$. We define componentwise maxima $M_{nk} = \max \{X_{1,k}, \ldots, X_{n,k}\}$ for $k = 1, \ldots, d$. If there exist normalising constants $a_{n,k} > 0$ and $b_{n,k}$ for $k = 1, \ldots, d$ such that

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

as $n \to \infty$ where $F$ is a distribution with all non-degenerate marginals then $F$ is a multivariate extreme value distribution of dimension $d$. Each marginal

$$Z_k = \lim_{n \to \infty} \frac{M_{n,k} - b_{n,k}}{a_{n,k}} \qquad k = 1, \ldots, d,$$

follows a GEV distribution with parameters $(\mu_k, \sigma_k, \xi_k)$. To focus on the dependence structure it is assumed that each margin follows a standard Fréchet distribution, i.e. GEV(1,1,1), such that

$$\mathrm{P}(Z_1 \le \infty, \ldots, Z_k \le z_k, \ldots, Z_d \le \infty) = \exp(-1/z_k),$$

for $z_k > 0$. To simplify notation we now consider a pair of random variables $Z_1$, $Z_2$ with common Fréchet marginal distributions. The multivariate extreme value distribution function $F$ is written as

$$F(z_1, z_2) = \exp\left\{-V(z_1, z_2)\right\} \qquad \text{for} \quad z_1 > 0, z_2 > 0, \tag{2.3.1}$$

where the exponent measure $V$ is defined as

$$V(z_1, z_2) = \int_0^1 \max\left(\frac{w}{z_1}, \frac{1-w}{z_2}\right) 2dH(w), \tag{2.3.2}$$

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

$$\int_0^1 wdH(w) = 1/2.$$

An important property of this distribution is that the quantity

$$V(1, 1) = \int_0^1 \max\left(w, 1 - w\right) 2dH(w),$$

is bounded on the range $[1, 2]$, with the lower bound occurring when $H(\{1/2\}) = 1$ and the upper bound occurring when $H(\{0\}) = H(\{1\}) = 1/2$. It is noted here that in standard extreme value notation $H$ is often taken as a measure as well as a distribution function. One common bivariate extreme value distribution has a logistic dependence structure (here often shortened to BEVL), parameterised by $\gamma$ (Tawn, 1990). This can be written in terms of the exponent measure as

$$V(z_1, z_2) = \left(z_1^{-1/\gamma} + z_2^{-1/\gamma}\right)^\gamma, \tag{2.3.3}$$

where $\gamma \in (0, 1]$, $z_1 > 0, z_2 > 0$ and where the distribution function $H$, from equation (2.3.2), is given as

$$H(w) = \frac{1}{2}\left[\left\{w^{(1-\gamma)/\gamma} - (1-w)^{(1-\gamma)/\gamma}\right\}\left\{w^{1/\gamma} + (1-w)^{1/\gamma}\right\}^{\gamma-1} + 1\right].$$

Thus by equation (2.3.1) the joint distribution function $F$ is given as

$$F(z_1, z_2) = \exp\left\{-\left(z_1^{-1/\gamma} + z_2^{-1/\gamma}\right)^\gamma\right\}. \tag{2.3.4}$$

Independent variables correspond to $\gamma = 1$ which reduces equation (2.3.4) to

$$F(z_1, z_2) = \exp\left\{-\left(z_1^{-1} + z_2^{-1}\right)\right\},$$

where $H$ now consists of half-unit mass atoms at $\{0\}$ and $\{1\}$, i.e. $H(\{0\}) = H(\{1\}) = 1/2$. Perfectly dependent variables are given as $\gamma \to 0$ and equation (2.3.4) becomes

$$F(z_1, z_2) = \exp\left\{-\max\left(z_1^{-1}, z_2^{-1}\right)\right\},$$

where $H$ is a unit mass atom at $\{1/2\}$, i.e. $H(\{1/2\}) = 1$.

**Multivariate threshold approaches**

When analysing data from a multivariate extreme value distribution the extremal dependence structure is important. Taking a pair of variables $(X_1, X_2)$ with common marginal distribution, we are interested in the extremal dependence structure of the pair which are not necessarily componentwise maxima. The joint tail model of Ledford and Tawn (1997) models the asymptotic form of the joint survivor function $\bar{F}$ directly. As a result we explicitly model only the points for both variables that fall above a high level $v = v_p$, often defined as the $(1 - p)$th quantile i.e.

$$P(X_1 > v_p) = p. \tag{2.3.5}$$

The form of the joint tail is given in Ledford and Tawn (1996) on Fréchet margins, but more generally is given as

$$P(X_1 > v_p, X_2 > v_p) \sim \mathcal{L}(1/p)\, p^{1/\eta}, \tag{2.3.6}$$

as $p \to 0$, where $\mathcal{L}$ is a slowly varying function at infinity and $\eta$ is named the coefficient of tail dependence; defined over the range $\eta \in (0, 1]$. In equation (2.3.6), the slowly varying function $\mathcal{L}$ satisfies

$$\frac{\mathcal{L}(t/p)}{\mathcal{L}(1/p)} \to 1 \qquad \text{as} \qquad p \to 0, \tag{2.3.7}$$

for all fixed $t > 0$. The value of the coefficient of tail dependence $\eta$ gives the level of extremal dependence between the marginal variables where $0 < \eta < 1/2$ implies negative association, $\eta = 1/2$ implies independence and $1/2 < \eta \leq 1$ implies positive association. If $\eta = 1$ and $\mathcal{L}(1/p) \nrightarrow 0$ as $p \to 0$ the pair of variables $(X_1, X_2)$ are asymptotically dependent. Any values of $\eta < 1$ imply asymptotic independence. Definitions of asymptotic dependence and asymptotic independence follow in Section 2.3.3.

The difference between asymptotically dependent and independent distributions will be of great importance later; see Section 2.3.3 for more details. As such, we define another important bivariate distribution, which is asymptotically independent. It is found by looking at the lower tail of the distribution given in equation (2.3.4). The distribution function for this inverted bivariate extreme value distribution (IBEV) is

$$F(z_1, z_2) = F_1(z_1) + F_2(z_2) - 1 + \exp\left\{-V\left(\left[-\log\bar{F}_1(z_1)\right]^{-1}, \left[-\log\bar{F}_2(z_2)\right]^{-1}\right)\right\},$$

(2.3.8)

where $F_i$, $i = 1, 2$, are univariate marginal distribution functions and $\bar{F}_i(z_i) = 1 - F_i(z_i)$ for $i = 1, 2$. Under the logistic dependence structure (IBEVL) given in equation (2.3.3), equation (2.3.8) becomes

$$\begin{aligned}
F(z_1, z_2) =\ & \exp\{-1/z_1\} + \exp\{-1/z_2\} - 1 \\
& + \exp\left\{-\left(\left[-\log(1 - \exp\{-1/z_1\})\right]^{1/\gamma} + \left[-\log(1 - \exp\{-1/z_2\})\right]^{1/\gamma}\right)^\gamma\right\}.
\end{aligned}$$

(2.3.9)

### 2.3.3 Extremal dependence measures

If two variables $(X_1, X_2)$ are asymptotically dependent it means that if $X_1$ is large it is possible for $X_2$ to be simultaneously extreme. Asymptotic independence is broadly the opposite case where the extreme values of variables $X_1$ and $X_2$ are unlikely to occur simultaneously. One common measure of the level of extremal dependence is given by the threshold dependent extremal dependence measure $\chi(x)$ (Coles et al.,

1999). The measure is defined as

$$\chi(x) = \mathrm{P}(X_2 > x \mid X_1 > x), \tag{2.3.10}$$

if $X_1$ and $X_2$ are on common margins, its limiting form $\chi$ is

$$\chi = \lim_{x \to x^*} \chi(x), \tag{2.3.11}$$

with $x^*$ being the upper limit of the support of the common marginal distribution. Dependence structures can be broadly split into those with asymptotic dependence and those with asymptotic independence (Sibuya (1960), Ledford and Tawn (1996)) determined by the value of $\chi$ in equation (2.3.11). In the case when $\chi = 0$ the variables $(X_1, X_2)$ are said to be asymptotically independent and $\chi > 0$ corresponds to asymptotic dependence. From equation (2.3.6) it is possible to construct the extremal dependence measure equation (2.3.10) as

$$\chi(v_p) = \mathrm{P}\left(X_2 > v_p \mid X_1 > v_p\right) \sim \mathcal{L}(1/p)p^{1/\eta - 1}, \quad \text{as } p \to 0,$$

with $v_p$ defined by condition (2.3.5). As outlined in Section 2.3.2, if $\eta = 1$ and $\mathcal{L}(1/p) \to c$ as $p \to 0$ then $X_1$ and $X_2$ are asymptotically dependent with $\chi = c$. If $\eta < 1$ or $\eta = 1$ and $\mathcal{L}(1/p) \to 0$ then $X_1$ and $X_2$ are asymptotically independent.

The extremal dependence measure in equation (2.3.11) gives the level of dependence within the asymptotic dependence class, but is not informative under asymptotic independence. A different measure of extremal dependence within the asymptotic independence class is available (Coles et al., 1999) in terms of $\eta$, i.e.

$$\bar{\chi} = 2\eta - 1. \tag{2.3.12}$$

Since $\eta \in (0, 1]$ it follows that $-1 < \bar{\chi} \leq 1$. Different values of $\bar{\chi}$ determine the level

of asymptotic independence. We have

$$\bar{\chi} = \begin{cases} 1 & \text{if asymptotically dependent} \\ (0,1) & \text{if asymptotically independent with positive association} \\ 0 & \text{if independent} \\ (-1,0) & \text{if asymptotically independent with negative association.} \end{cases}$$

Furthermore, if $\bar{\chi} = 1$ and $\mathcal{L}(1/p) \not\to 0$ the variables are asymptotically dependent and if $\bar{\chi} < 1$ the variables are asymptotically independent. All bivariate extreme value distributions either exhibit asymptotic dependence or independence. For the bivariate extreme value distribution, the exponent measure in equation (2.3.2) links to $\chi$ from equation (2.3.11) by

$$\chi = 2 - V(1,1), \tag{2.3.13}$$

and thus $\chi > 0$ unless $H$ puts all mass at $\{0\}$ or $\{1\}$, i.e. the case of independence from Section 2.3.2. The BEVL given in equation (2.3.4), can only account for asymptotic dependence, except in the case of perfect independence when $\gamma = 1$. The logistic dependence parameter links to the extremal dependence measure via equation (2.3.13) and since for this distribution $V(1,1) = 2^\gamma$ it is found that $\chi = 2 - 2^\gamma$. Under asymptotic dependence the coefficient of tail dependence is given as $\eta = 1$ and as such by equation (2.3.12) $\bar{\chi} = 1$.

For the asymptotically independent inverted bivariate extreme value distribution the same extremal measures can be calculated. Unless the variables are perfectly dependent, i.e. if $V(1,1) \neq 1$, we have that $\chi = 0$ and $\bar{\chi} = 2/V(1,1) - 1$, since $\eta = 1/V(1,1)$. Therefore, under the logistic dependence structure the strength of the subasymptotic dependence is given as $\bar{\chi} = 2^{1-\gamma} - 1$.

## 2.3.4 Copulas

In previous sections the behaviour of the dependence structure has been investigated for Fréchet margins. One more general way to express the dependence structure between random variables is via copulas. The copula is a joint distribution function with the property that every marginal distribution is uniform on the interval $[0, 1]$. Take a joint distribution $F$ for a set of random variables $X_1, \ldots, X_d$ each with univariate marginal distributions $F_1(x_1), \ldots, F_d(x_d)$ and with corresponding quantile functions $F_1^{-1}, \ldots, F_d^{-1}$. The copula $C$ can be expressed as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{2.3.14}$$

The copula in equation (2.3.14) can also be expressed as

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)),$$

where $0 \leq u_i \leq 1$ for $i = 1, \ldots, d$. Many different copulas are available, comprehensive reviews are given in Joe (1997) and Nelson (2007). One simple example is the independence copula which occurs when all marginals are independent

$$C^{ind}(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i.$$

The multivariate extreme value distribution with logistic dependence structure, given in equation (2.3.4) for the bivariate case, can be written in terms of the copula as

$$C^{log}(u_1, \ldots, u_d) = \exp\left\{-\left[\sum_{i=1}^{d} (-\log u_i)^{1/\gamma}\right]^{\gamma}\right\}, \tag{2.3.15}$$

where $0 < \gamma \leq 1$ is the logistic dependence parameter. The lower tail of the multivariate extreme value distribution with logistic dependence structure is asymptotically independent and this motivates an inverted multivariate extreme value distribution with logistic dependence structure, given in equation (2.3.9) for the bivariate case. The copula is given most easily in terms of the survival copula

$$\bar{C}^{ilog}(u_1, \ldots, u_d) = \exp\left\{-\left[\sum_{i=1}^{d} (-\log (1 - u_i))^{1/\gamma}\right]^{\gamma}\right\}, \tag{2.3.16}$$

where $\bar{C}(u_1, \ldots, u_d) = \mathrm{P}\left(F_1(x_1) > u_1, \ldots, F_d(x_d) > u_d\right)$.

The threshold dependent extremal measure in equation (2.3.10) can now be defined via the copula as

$$\chi(q) = \frac{\bar{C}(q, q)}{1 - q},$$

where $q \in [0, 1]$. In the limit

$$\chi = \lim_{q \to 1} \chi(q).$$

Expressions for $\chi(q)$ for the two extreme value distributions of interest follow from the copulas given in equations (2.3.15) and (2.3.16). The extremal dependence measure for the BEVL is

$$\chi(q) = \frac{1 - 2q + q^{2^\gamma}}{1 - q}. \tag{2.3.17}$$

The measure is obtained for the IBEVL as

$$\chi(q) = (1 - q)^{2^\gamma - 1}. \tag{2.3.18}$$

In Section 2.7, equations (2.3.17) and (2.3.18) will provide 'true' values for $\chi(q)$ for the bivariate extreme value distributions outlined in Section 2.3.4 which can be used to measure the performance of approaches for modelling the joint extreme tails; these approaches are given in Sections 2.5 and 2.6, one is the joint tail approach of Ledford and Tawn (1997) and the other is the conditional extremes approach of Heffernan and Tawn (2004).

## 2.4   Marginal transformations

Estimation of extremal tail properties using any multivariate dependence approach requires a model for the marginal characteristics of the data prior to modelling the dependence structure on common margins. Different marginal choices are necessary

for different methods outlined later so a general framework for marginal transformation is required.

Consider a set of $n$ pairs of random variables $(X_{11},\ X_{21}),\ldots,(X_{1n},\ X_{2n})$ that occur simultaneously such that the first value of each pair has corresponding marginal distribution $F_1$ and the second value has distribution $F_2$. A high threshold $u_i$, $i = 1, 2$, is chosen and points falling below this threshold are modelled using the empirical cumulative distribution function. Any points that lie above the threshold are modelled using the GPD, as outlined in Section 2.2.3, thus $F_i$ is given by

$$F_i(x) = \begin{cases} 1 - \lambda_{u_i}\left(1 + \xi_i \frac{x - u_i}{\sigma_{u_i}}\right)_+^{-1/\xi_i}, & x \geq u_i \\ \tilde{F}_i(x), & x < u_i, \end{cases} \tag{2.4.1}$$

for the marginal threshold $u_i$, where $\lambda_{u_i} = 1 - F_i(u_i)$ and $\tilde{F}_i(x)$ is the empirical cumulative distribution function of $\{X_i\}$. A transformation onto an appropriate margin $Y_i$ is made such that

$$Y_i = T^{-1}\{F_i(X_i)\},$$

for $i = 1, 2$, where the inverse distribution function $T^{-1}$ transforms to the appropriate marginal distribution. At different points in the thesis, we require Fréchet, Pareto, Gumbel and Laplace margins to be defined. When necessary the appropriate margins are defined in later sections.

## 2.5 Parametric joint tail approach

The joint tail model of Ledford and Tawn (1997) has already been introduced in Section 2.3.2 as a model for the distribution for the extremes of a pair of random variables that can account for asymptotic dependence and asymptotic independence. In this section inference for this approach is discussed with different techniques for

deriving uncertainty bounds. An extension to the original model that incorporates more information about values that are extreme in only one variable is proposed.

### 2.5.1 Inference

The modelling assumption $\mathcal{L}(1/p) \to c$ as $p \to 0$ is made, where $c \in (0, 1]$ and $\mathcal{L}(1/p)$ is a slowly varying function that satisfies equation (2.3.7). In the limit this assumption could introduce a small amount of bias, but at the subasymptotic levels we are interested in we shall take the slowly varying function as a constant. As such, our model may be mis-specified but this can be investigated by goodness-of-fit tests. Under the above assumption, the problem of calculating extremal dependence comes down to the estimation of parameters $(c, \eta)$ which fully explain the dependence in the joint tail. Define $Y = \min(Y_1, Y_2)$ and let $(Y_1, Y_2)$ have Pareto margins, i.e.

$$Y_i = \left[1 - F_i(X_i)\right]^{-1},$$

for $i = 1, 2$ and thus $Y_i > 1$. For large $v$, Fréchet and Pareto margins are equivalent up to first order in the limit since

$$\exp\{-1/v\} = 1 - \frac{1}{v} + \mathcal{O}(v^{-2})$$

Under Pareto margins we have that $p = v^{-1}$, see equation (2.3.5), and under the assumption that the limit form in equation (2.3.6) holds above some sufficiently high threshold $u$ we write

$$\mathrm{P}\left(Y > v\right) = cv^{-1/\eta} \qquad \text{for} \quad v > u. \tag{2.5.1}$$

Equation (2.5.1) is used to construct the likelihood in terms of the parameters $\boldsymbol{\theta} = (c, \eta)$

$$L(\boldsymbol{\theta}) = \mathrm{P}(Y < u)^{n-n_u} \prod_{i=1}^{n_u} \left( \frac{c}{\eta y_i^{1/\eta+1}} \right), \tag{2.5.2}$$

where $n_u$ is the number of $Y$ exceeding the threshold $u$, $y_1, \ldots, y_{n_u}$ are data points of $Y$ that exceed the threshold $u$ and $n$ is the length of the data. The likelihood

contribution for each data point that falls below one of the marginal thresholds is given as $P(Y < u)$. The maximum likelihood estimates $\hat{\boldsymbol{\theta}} = (\hat{c}, \hat{\eta})$ are

$$\hat{\eta} = \min\left(\frac{1}{n_u}\sum_{i=1}^{n_u}\log\left(\frac{y_i}{u}\right), 1\right) \tag{2.5.3}$$

$$\hat{c} = \frac{n_u}{n}u^{\hat{\eta}}. \tag{2.5.4}$$

The derivations of equations (2.5.3) and (2.5.4) follow in a similar manner to those derived in Appendix A. If the variables are asymptotically dependent, i.e. $\eta = 1$, and since $\eta = 1$ is a boundary value, with probability $1/2$ we have $\hat{\eta} = 1$ (Self and Liang, 1987) and the level of asymptotic dependence is given by $\hat{c} = un_u/n$; related to $\chi(u)$ via $\hat{c} = u\tilde{\chi}(u)$, where $\tilde{\chi}(u)$ is an empirical version of $\chi(u)$. An estimate of the threshold dependent extremal dependence measure at a threshold $v > u$ can be obtained as

$$\hat{\chi}(v) = \hat{c}v^{1-1/\hat{\eta}} \qquad\qquad v > u. \tag{2.5.5}$$

Estimates of the uncertainty in $\hat{\chi}(v)$ can be obtained by deriving confidence intervals, either based upon estimates of the standard error or via the profile likelihood. For notational simplicity the dependence measure $\chi(v)$ is rewritten as $\phi$, which is still dependent on the threshold $v$. A $(1 - \alpha)100\%$ confidence interval for $\hat{\phi}$ is given by

$$\left(\hat{\phi}_l, \hat{\phi}_u\right) = \left(\hat{\phi} - z_{1-\alpha/2}\text{var}(\hat{\phi})^{1/2}, \hat{\phi} + z_{1-\alpha/2}\text{var}(\hat{\phi})^{1/2}\right),$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the normal distribution and the variance of $\hat{\phi}$ is given by

$$\text{var}(\hat{\phi}) = \nabla g(\boldsymbol{\theta})^T I(\boldsymbol{\theta})^{-1} \nabla g(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $\phi = g(\boldsymbol{\theta}) = cv^{1-1/\eta}$, the inverse information matrix $I(\boldsymbol{\theta})^{-1}$ is obtained in practice as the inverse of the Hessian matrix and the vector derivatives $\nabla g(\boldsymbol{\theta})$ can be calculated using finite differencing. Although easy to calculate, one drawback of this type of confidence interval is that some values within the range are not attainable

by the parameter. In this situation we have $\phi \in [0, 1]$, so negative values of the parameter do not make sense. Approaches based upon profile likelihood avoid this problem and could provide more accurate estimates of the uncertainty as a result. Let $\ell(\boldsymbol{\theta}) = \ell(c, \eta)$ be the log-likelihood derived from equation (2.5.2) and invert equation (2.5.5) such that

$$c = \phi v^{1/\eta - 1}.$$

Define the profile log-likelihood for $\phi$ as

$$P\ell(\phi) = \max_{\eta} \ell_{\phi}(\phi, \eta),$$

where $\ell_{\phi}$ is the log-likelihood re-parameterised in terms of the pair $(\phi, \eta)$ as opposed to $(c, \eta)$. The profile deviance function is thus defined as

$$D^*(\phi) = 2(\ell_{\phi}(\hat{\phi}, \hat{\eta}) - P\ell(\phi)).$$

The profile deviance function has an approximate $\chi_1^2$ distribution under the null hypothesis that $\phi = \phi_0$, if $\phi_0$ is the true value of $\phi$, which can be used to form the basis of confidence interval construction. Uncertainty estimates obtained using both approaches are compared via a simulation study in Section 2.7.2.

## 2.5.2   Extension

We propose an extension to the joint tail model that incorporates more information about values that are extreme in one variable but not the other. When constructing the likelihood in equation (2.5.2) the $n_u$ points that fall in the extreme quadrant in the top right are modelled explicitly and all other values are simply modelled as not being in the extreme quadrant. In this way values that exceed only one of the marginal thresholds are treated in the same way as values that are small in both margins. To overcome this, define the four different quadrants as

$$R_{00} = \{Y_1 \leq u, Y_2 \leq u\} \qquad R_{01} = \{Y_1 \leq u, Y_2 > u\}$$
$$R_{10} = \{Y_1 > u, Y_2 \leq u\} \qquad R_{11} = \{Y_1 > u, Y_2 > u\},$$

such that the whole bivariate space is partitioned into $R_{00}$, $R_{01}$, $R_{10}$ and $R_{11}$. The probability of falling in $R_{11}$ under the joint tail model is given directly by equation (2.5.1). The probability of falling in $R_{00}$ is derived through simple inclusion/exclusion arguments, i.e.

$$P(Y_1 \leq u, Y_2 \leq u) = 1 - P(Y_1 \leq u) - P(Y_2 \leq u) + P(Y_1 > u, Y_2 > u)$$
$$= 1 - 2/u + cu^{-1/\eta},$$

and the regions $R_{01}$ and $R_{10}$ follow by similar arguments. With probabilities of falling in all regions we can now construct the likelihood function. Let $n_{00} = \#\{R_{00}\}$, $n_{01} = \#\{R_{01}\}$, $n_{10} = \#\{R_{10}\}$ and $n_u = \#\{R_{11}\}$ and therefore $n = n_{00} + n_{01} + n_{10} + n_u$. The likelihood given in equation (2.5.2) can be re-written as

$$L(c, \eta) = \left(1 - \frac{2}{u} + cu^{-1/\eta}\right)^{n_{00}} \left(\frac{1}{u} - cu^{-1/\eta}\right)^{n_{01}} \left(\frac{1}{u} - cu^{-1/\eta}\right)^{n_{10}} \prod_{i=1}^{n_u} \left(\frac{c}{\eta y_i^{1/\eta+1}}\right).$$

Obtaining the log-likelihood and calculating derivatives leads to an analytical expression for $\hat{\eta}$ given by

$$\hat{\eta} = \min\left(\frac{1}{n_u} \sum_{i=1}^{n_u} \log\left(\frac{y_i}{u}\right), 1\right), \tag{2.5.6}$$

which is the same as $\hat{\eta}$ for the original joint tail model in equation (2.5.3); we have $\min(\cdot, 1)$ to ensure that $\eta$ cannot be greater than 1. We also obtain an analytical expression for $\hat{c}$

$$\hat{c} = \frac{-\omega \pm \sqrt{\omega^2 - 4\rho\kappa}}{2\rho}, \tag{2.5.7}$$

where

$$\rho = nu^{-2/\hat{\eta}}$$
$$\omega = (n_{01} + n_{10})\left(u^{-1/\hat{\eta}} - 2u^{-1/\hat{\eta}-1}\right) - n_{00}u^{-1/\hat{\eta}-1} - n_u\left(3u^{-1/\hat{\eta}-1} - u^{-1/\hat{\eta}}\right)$$
$$\kappa = -n_u\left(\frac{1}{u} - \frac{2}{u^2}\right).$$

From the terms above we note that $\kappa\rho < 0$ and we have the constraint that $c \in (0, 1]$ and as such we need to take the positive root of equation (2.5.7). From equations (2.5.7) and (2.5.6) we observe that any differences in the values of $\hat{\chi}(v)$ between the two estimation methods will be driven by changes in the value of $\hat{c}$ rather than $\hat{\eta}$. Derivations of the estimates in equations (2.5.7) and (2.5.6) are given in Appendix A. A simulation study testing whether the new approach provides efficiency gains is undertaken in Section 2.7.1.

## 2.6 Semi-parametric conditional extremes approach

The conditional extremes approach was proposed by Heffernan and Tawn (2004) as a separate method for modelling extremal dependence which avoided the limiting arguments of the joint tail methods in which all variables must become large at the same rate. Standard copula based methods, as discussed in Section 2.3.4, can typically only handle one form of extremal dependence, either asymptotic dependence or asymptotic independence. As such the form of the dependence structure has to be chosen in advance before the model is fitted. The conditional extremes approach estimates the form of the extremal dependence structure as part of the fitting procedure so removes the need to choose the form of the dependence structure in advance. The conditional extremes approach can also be used to model high-dimensional data with greater ease than for copula based methods, although all theory in this section is given for the bivariate case.

This section is split into four parts. Firstly the conditional extremes model is presented in the bivariate case. Then inferential considerations are outlined as well as approaches for generating simulated data sets from the fitted model. Finally two different extensions to the approach are proposed. One extension aims to generate accurate estimates of extremal dependence measures without simulation. The second

extension aims to generate more realistic simulated data sets in high dimensions.

## 2.6.1 Theory

To estimate the dependence structure of $(X_1, X_2)$ using the conditional extremes approach both variables must be transformed onto common marginal distributions; see Section 2.4. The classical representation from Heffernan and Tawn (2004) is given on Gumbel margins, i.e.

$$Y_i = -\log[-\log\{F_i(X_i)\}] \qquad \text{for} \quad i = 1, 2$$

where $F_i$ is the marginal distribution function for $X_i$, $i = 1, 2$. The transformation to Gumbel margins means that $P(Y_i \leq y) = \exp\{-\exp(-y)\}$, and that as $y \to \infty$, $P(Y_i > y) \sim \exp(-y)$ for $i = 1, 2$. Therefore, both random variables $(Y_1, Y_2)$ now have an approximately exponential upper tail, which is of importance when we consider the convergence of the conditional distribution in equation (2.6.1). A different formulation is given by Keef et al. (2013) on Laplace margins, i.e.

$$Y_i = \begin{cases} \log\{2F(X_i)\} & \text{if} \quad X_i < F_i^{-1}(0.5) \\ -\log\{2[1 - F(X_i)]\} & \text{if} \quad X_i \geq F_i^{-1}(0.5), \end{cases}$$

for $i = 1, 2$. Again both variables have an exponential upper tail, but under this marginal choice both variables also have an exponential lower tail. As such negative extremal dependence can be characterised as well as positive extremal dependence. In later sections the transformation to Laplace margins is preferred for simplicity, but for completeness here the dependence model is provided for both marginal choices.

Having made the marginal transformation, the desire is to model $(Y_1, Y_2)$ using the distribution of $Y_2$ given that $Y_1$ is large (defined as exceeding a high threshold). A requirement for modelling the conditional distribution $P\{Y_2 \leq y_2 \mid Y_1 = y_1\}$ is that this distribution should be non-degenerate as $y_1 \to y^*$, where $y^*$ is the upper endpoint of

the common marginal distribution. As such the conditional extremes approach aims to identify normalizing functions $a : \mathbb{R}_+ \to \mathbb{R}$ and $b : \mathbb{R}_+ \to \mathbb{R}_+$ that are defined such that for $y > 0$

$$P \left( \frac{Y_2 - a\,[Y_1]}{b\,[Y_1]} \le z, Y_1 - u > y \mid Y_1 > u \right) \to \exp(-y)G(z), \qquad (2.6.1)$$

as $u \to \infty$, where $G(z)$ is a non-degenerate distribution function. The first term of the limit in equation (2.6.1) arises from the fact that $Y_i$ for $i = 1, 2$ now both have an exponential upper tail. The second term in the limit characterises the behaviour of $Y_2 \mid Y_1 > u$ in terms of the limiting distribution $G(z)$ along with location and scale norming functions $a(Y_1)$ and $b(Y_1)$ respectively. From equation (2.6.1), $G(z)$ is defined to be the limiting conditional distribution of

$$Z = \frac{Y_2 - a(Y_1)}{b(Y_1)}, \qquad (2.6.2)$$

given $Y_1 > u$ as $u \to \infty$. One result that follows from equations (2.6.1) and (2.6.2) is that $Z$ and $Y_1$ are independent in the limit as $u \to \infty$ given that $Y_1 > u$.

Equation (2.6.2) defines the limiting distribution $G$ but to fully characterise the second term in the limit of equation (2.6.1) normalising functions $a(Y_1)$ and $b(Y_1) > 0$ must be defined. Heffernan and Tawn (2004) work on Gumbel margins and, for a broad class of copula families, in this situation the normalising functions are found to be special cases of the parametric family, i.e.

$$a(y) = \alpha y + \mathbb{I}_{\alpha=0, \beta<0}\{c - d\log(y)\}$$
$$b(y) = y^\beta,$$

where $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$, $c \in (-\infty, \infty)$ and $d \in [0, 1]$. The specification of Laplace margins ensures that the upper and lower tails are both modelled by a symmetric distribution with exponential tails and permits the definition of a single unified class of normalising functions

$$a(y) = \alpha y \qquad \text{and} \qquad b(y) = y^\beta, \qquad (2.6.3)$$

where $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1)$. This form of the normalising functions does not affect the limiting dependence model in Heffernan and Tawn (2004) and simplifies the inference for variables which are either negatively or weakly associated. For the rest of the thesis data are transformed onto Laplace margins before modelling extremal dependence using the conditional extremes approach. This permits easier interpretation of extremal dependence using the parameters in equation (2.6.3).

Different values of $\alpha$ and $\beta$ characterise different forms of tail dependence. In the case where $\alpha = 1$ and $\beta = 0$, variables $(Y_1, Y_2)$ exhibit asymptotic positive dependence. Due to the exponential lower tail specified by the Laplace margins, the case of asymptotic negative dependence is given when $\alpha = -1$ and $\beta = 0$. If $\alpha = \beta = 0$ and $G(z)$ is the Laplace distribution function the variables are independent. Dependence parameters can also be estimated for the distributions outlined in Section 2.3.2. Since the BEVL is asymptotically dependent, the conditional extremes dependence parameters are $\alpha = 1$ and $\beta = 0$. For the IBEV, $\alpha = 0$ and $0 < \beta < 1$ with the value of $\beta$ determined by the tail features of of the spectral measure of the multivariate extreme value distribution; for the IBEVL $\beta = 1 - \gamma$. Keef et al. (2013) give the form of the dependence parameters for other distributions.

## 2.6.2   Inference

Modelling using the conditional extremes approach requires the assumption that the limiting form of equation (2.6.1) holds exactly for all values of $Y_1 > u$ given that $u$ is a sufficiently high threshold. Given this assumption it is possible to write the form of $Y_2$ given that $Y_1 > u$ as

$$Y_2 = \alpha Y_1 + Y_1^{\beta} Z, \tag{2.6.4}$$

where $Z$ is a random variable with distribution function $G$, as defined in equation (2.6.2), and is independent of $Y_1$. As $G$ does not take any simple parametric

form, a false working assumption is made in Keef et al. (2013) that $Z \sim N(\mu, \sigma^2)$ and as such

$$Y_2 \mid \{Y_1 = y\} \sim N\left(\alpha y + \mu y^\beta, \sigma^2 y^{2\beta}\right) \qquad \text{for} \quad y > u.$$

The working assumption permits the estimation of the set of parameters $(\alpha, \beta, \mu, \sigma)$ via standard likelihood approaches. At this stage the estimates for $(\mu, \sigma)$ are discarded and a non-parametric distribution for $Z$ is formed by inverting equation (2.6.4) to give estimated values of $Z$. Specifically, let $y_{i,k}$ be the $k$th data value for variable $i$ and $k_1, \ldots, k_{n_u}$ be the indices of $k = 1, \ldots, n$ where $y_{1,k} > u$ then

$$\hat{z}_j = \frac{y_{2,k_j} - \hat{\alpha} y_{1,k_j}}{(y_{1,k_j})^{\hat{\beta}}}, \tag{2.6.5}$$

for $j = 1, \ldots, n_u$, where $n_u$ is the number data points exceeding the threshold $u$. In this way a non-parametric estimate $\hat{G}$ to the distribution function $G$ is formed using $\hat{z}_j$, $j = 1, \ldots, n_u$.

In many situations throughout the thesis the likelihood ratio test is used to assess different aspects about the fitted conditional model; for example, whether we can assume asymptotic dependence ($\alpha = 1$, $\beta = 0$) holds or not. This approach is a standard way in statistics of testing whether a certain model provides a significantly better fit than another and as such more detail is not provided. However, it is noted that the test of the model will be made under the false working assumption given above and as such the model will be misspecified. A consequence of the misspecification of the model is that standard asymptotic arguments do not hold and instead of a $\chi^2$ distribution on $q$ degrees of freedom say, the limiting distribution is that of a weighted sum of $q$ independent $\chi_1^2$ variables. Our incorrect use of the standard limit distribution may induce inefficiency, i.e. this may cause us to make the wrong decision when performing a likelihood ratio test, but evidence suggests the loss of efficiency can be slight; see Chandler and Bate (2007).

### 2.6.3 Simulation

The motivation behind fitting a model for the joint extremes is to understand the dependence between variables at extreme levels, especially levels beyond previously observed levels. Section 2.3.3 presented measures such as $\chi(v)$ for summarising extremal dependence and being able to estimate such measures from our conditional extremes model provides important information. For this reason, the conditional extremes approach can be used to generate simulated data sets from the fitted extremal dependence model. Here we provide an equivalent simulation scheme to Heffernan and Tawn (2004) and Keef et al. (2013):

1. Pick critical level $v$ and simulate an exceedance $Y_1^*$ from an Exponential distribution with rate 1.

2. Sample $Z^*$ from $\hat{z}_j$, $j = 1, \ldots, n_u$ independently of $Y_1^*$.

3. Obtain $Y_2^* = \hat{\alpha}Y_1^* + (Y_1^*)^{\hat{\beta}} Z^*$.

By repeating the steps above $m$ times we obtain a sample of pairs of size $m$, denoted $(Y_1^{**}, Y_2^{**})$ which has the desired conditional distribution $Y_2 \mid Y_1 > u$. The simulated sample can be used to estimate the threshold dependent extremal dependence measure $\chi(v)$ given in equation (2.3.10) as

$$\hat{\chi}(v) = \frac{\#\{Y_2^{**} > v\}}{m}.$$

### 2.6.4 Bootstrapping

Throughout the thesis, estimates of the uncertainty in the dependence parameters and other extremal quantities will be derived by bootstrapping. In the most simple case, a bootstrap sample can be constructed by resampling pairs of data from $(Y_1, Y_2)$ with replacement until a sample of the same length as the original data set has been constructed. Then the dependence parameters and residuals can be estimated using the approaches in Section 2.6.2 and we can simulate from this fitted model using the

method outlined in Section 2.6.3.

In many situations later in the thesis we shall be explicitly modelling the temporal dependence of a single variable $Y_t$ and as such it will be necessary to have a different bootstrap approach. In this situation we shall construct bootstrap samples by splitting the original sample into periods of exceedances and periods of non-exceedances. We then consequently pick randomly from the sets of non-exceedances and exceedances until we have constructed a new dataset of the same length as the original sample. This new sample retains the temporal dependence features of the original data and as such is valid for our purposes. The marginal and dependence characteristics of the bootstrap sample can then be assessed using the approaches outlined above and multiple replications can be used to build up uncertainty estimates.

### 2.6.5 Extensions

In most situations the simulation scheme proposed above can generate a simulated sample from the desired conditional distribution with little computational expense. However situations may arise where computational power is at a premium and two tricks can be used to obtain more accurate estimates of extremal quantities. The first approach completely removes the need to simulate in the bivariate case for a broad class of extremal dependence measures, in particular we focus on the threshold dependent extremal dependence measure $\chi(v)$. Then we use kernel density estimation to provide a more flexible simulation approach in two or more dimensions.

**Direct estimation of extremal quantities**

Firstly we propose an approach that can be used to obtain an estimate of the extremal dependence measure $\chi(v)$ without the need to simulate. Equation (2.6.4) gives the form of $Y_2$ given that $Y_1 > u$, which can be directly used to obtain the threshold dependent extremal measure $\chi(v)$ for any $v > u$. Due to the decomposition of terms

in equation (2.6.1), $\chi(v)$ can be estimated as

$$\chi(v) = P(Y_2 > v \mid Y_1 > v)$$

$$= P(\alpha Y_1 + Y_1^\beta Z > v \mid Y_1 > v). \tag{2.6.6}$$

As outlined in Section 2.6.2, during inference the distribution of the random variable $Z$ is estimated by the empirical distribution of the sample $\hat{z}_j$ for $j = 1, \ldots, n_u$ as in equation (2.6.5). From step 2 of the simulation algorithm in Section 2.6.3 we know that random draws are taken from $\hat{z}_j$, $j = 1, \ldots, n_u$ to construct the simulated sample, i.e. each $\hat{z}_j$ has a probability of $n_u^{-1}$ of being picked. Let $\hat{z}_{(j)}$ be the ordered values of the sample $(\hat{z}_1, \ldots, \hat{z}_{n_u})$, given by equation (2.6.5), such that $\hat{z}_{(1)} \geq \hat{z}_{(2)} \geq \cdots \geq \hat{z}_{(n_u)}$. For sufficiently large $v$ we may rewrite equation (2.6.6) as

$$
\begin{aligned}
\chi(v) &= \sum_{j=1}^{n_u} P(\alpha Y_1 + Y_1^\beta Z > v \mid Y_1 > v, Z = z_{(j)}) P(Z = z_{(j)}) \\
&= \frac{1}{n_u} \sum_{j=1}^{n_u} P(\alpha Y_1 + Y_1^\beta z_{(j)} > v \mid Y_1 > v) \\
&= \begin{cases} \frac{1}{n_u} \sum_{j=1}^{n_u} P(Y_1 > \tilde{y}_j \mid Y_1 > v) & \text{for } \alpha > 0 \\[2mm] \frac{1}{n_u} \sum_{j=1}^{n_u} P(Y_1 < \tilde{y}_j \mid Y_1 > v) & \text{for } \alpha < 0, \end{cases}
\end{aligned} \tag{2.6.7}
$$

where $\tilde{y}_j \geq v > 0$ is the root of the equation

$$\alpha \tilde{y}_j + (\tilde{y}_1)^\beta z_{(j)} - v = 0,$$

for $j = 1, \ldots, n_u$. Note that if $\alpha\beta < 0$ and $z_{(j)} > 0$ then the terms $\alpha Y_1$ and $Y_1^\beta z_{(j)}$ change in opposite directions as $Y_1$ increases. Unless $v$ is large enough, the final step of equation (2.6.7) does not hold. Figure 2.6.1 provides motivation for the distinction depending on the sign of $\alpha$ in equation (2.6.7). When $\alpha > 0$ each $z_{(j)}$ for $j = 1, \ldots, n_u$ defines a ray that increases with $Y_1$ and as such any $Y_1 > \tilde{y}_j$ contributes to the value of $\chi(v)$ for that particular $z_{(j)}$. The opposite is true when $\alpha < 0$. Since $Y_1$ is on

Laplace margins equation (2.6.7) simplifies, for large $v$, to

$$\chi(v) = \begin{cases} \frac{1}{n_u} \sum_{j=1}^{n_u} \exp\left\{-(\tilde{y}_j - v)\right\} & \text{for } \alpha > 0 \\ \frac{1}{n_u} \sum_{j=1}^{n_u} 1 - \exp\left\{-(\tilde{y}_j - v)\right\} & \text{for } \alpha < 0. \end{cases}$$
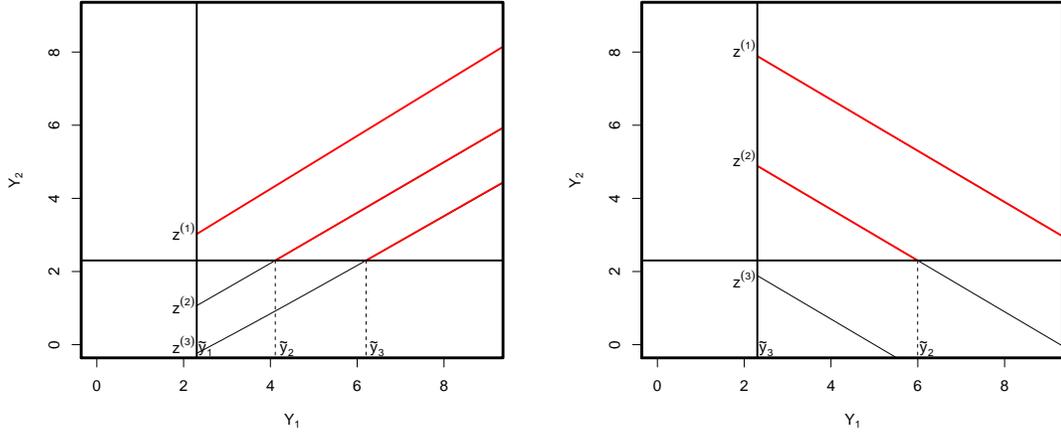


Figure 2.6.1: Diagram of how to directly simulate $\chi(v)$ for random variables $(Y_1, Y_2)$ without the need for repeated simulation for $\alpha > 0$ (left) and $\alpha < 0$ (right) with $\beta = 0$. Red segments highlight areas for each value $z^{(j)}$ for $j = 1, \ldots, n_u$ which fall within the extremal region of interest.

A simulation study designed to investigate the efficiency of the direct estimation approach for estimating the threshold dependent extremal measure is given in Section 2.7.3.

## Using kernel smoothing in extremal simulation

As a different extension we propose an improvement to step 2 of the conditional simulation scheme given in Section 2.6.3. We note that sampling from a kernel smoothed version of $\hat{G}$ instead of directly from the original sample could lead to more accurate simulations and therefore more accurate estimates of extremal measures. We envision two situations in which this approach will have benefits over the standard simulation approach outlined in Section 2.6.3. Firstly, in the situation where $n_u$ is small, simple sampling can lead to misleading extremal simulations and kernel smoothing could lead

to a more realistic simulated sample; see Figure 2.6.2. Secondly, in high-dimensional problems sampling directly from the equivalent multivariate version of $\hat{G}$ could lead to simulations that provide poor coverage of the whole multivariate space. This coverage could be improved by a kernel density approach. For the rest of this section we concentrate on the first use as this can be outlined using bivariate results derived in the chapter so far. Kernel smoothing for multivariate extreme value problems are covered in more detail in Chapter 4.

It was highlighted in Section 2.2.3 there is a bias-variance trade-off when selecting the threshold at which to model marginal extremes. A similar trade-off exists when modelling the dependence structure using the conditional extremes approach. Setting a higher modelling threshold will provide a more accurate representation of the tail but at the cost of having little data to model. When simulating from the conditional extremes model this manifests itself as a small $n_u$ which will restrict the size of the set $\hat{z}_j$ upon which to generate replications. The left plot of Figure 2.6.2 illustrates the potential problem. A simulated sample, on Gumbel margins, generated from a conditional extremes model fitted at a high modelling threshold (black solid line) is plotted. The simulated data (black circles) tend to line up along rays which could lead to inaccurate estimates of $\chi(v)$ as the simulated values may not adequately cover the bivariate space.

We propose to use a kernel smoothed version of the distribution $\hat{G}$ during the simulation procedure outlined in Section 2.6.3. Specifically instead of sampling directly from $\hat{z}_j$ we sample from a kernel density estimate

$$\hat{g}_h(z) = \frac{1}{n_u h} \sum_{i=1}^{n_u} K\left(\frac{z - \hat{z}_i}{h}\right), \tag{2.6.8}$$

where $h$ is the bandwidth and $K$ is the kernel function. There exist many choices for
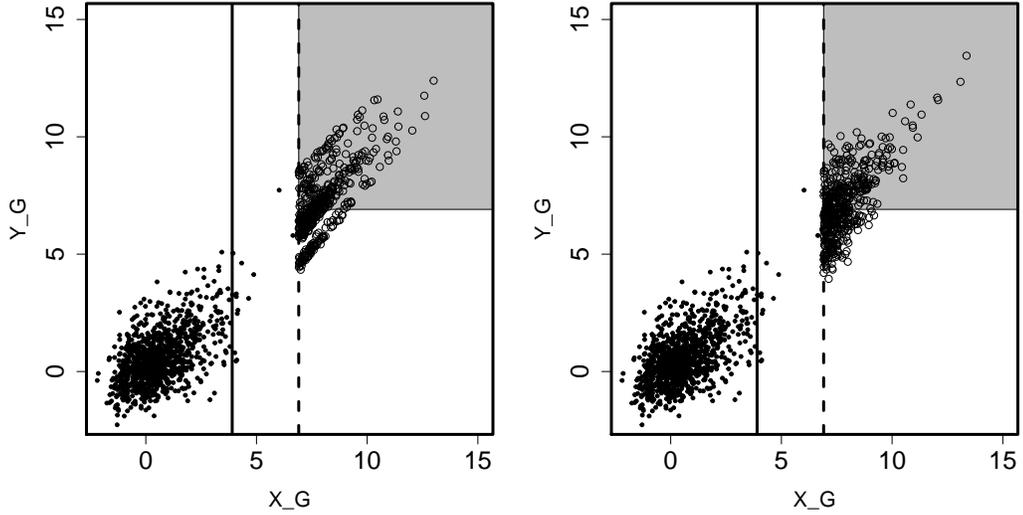
Figure 2.6.2: A simulated sample generated on Gumbel margins using current conditional extremes method (left) and with additional Gaussian noise and shrinkage steps (right), fitted at the modelling threshold (solid line). Notice the exceedances on the left tend to group along rays emanating from the critical level (dotted line) which is remedied with the additional steps.

the kernel $K$, here we take $K$ to be a standard Normal kernel, i.e.

$$\hat{g}_h(z) = \frac{1}{n_u h} \sum_{i=1}^{n_u} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(z - \hat{z}_i)^2}{2h^2} \right\}.$$

The choice of bandwidth to use is not trivial and an important area of research within the field of kernel density estimation (Silverman (1986), Jones et al. (1996)). Setting $h$ high increases the amount of smoothing and leads to a smoother $\hat{f}_h(z)$; setting $h$ smaller leads to a noisier density function. A standard choice of bandwidth is the default in `R` (Scott, 1992), i.e.

$$h = \left( \frac{4\sigma_z^5}{3n} \right)^{1/5},$$

where $\sigma_z$ is the standard deviation of the sample $\hat{z}_j$. When increasing the value of $h$ we must be careful to not artificially inflate the variance of the original sample. For extremal quantities such as $\chi(v)$, inflating the variance of our sample could lead to misleading estimates. Liu and West (2001) introduced an approach for shrinking the

variance of a kernel smoothed version of a sample that we use here. Let $\mu_z$ and $\sigma_z$ be the mean and variance of $\hat{z}_j$ respectively. We can obtain a new kernel density function similar to equation (2.6.8) with kernel shrinkage (Liu and West, 2001)

$$\tilde{g}_h(z) = \frac{1}{n_u h} \sum_{i=1}^{n_u} K\left(\frac{z - a\hat{z}_j - (1-a)\mu_z}{h}\right),$$

where $a = \sqrt{1 - h^2}$.

## 2.7 Simulation studies

In this section simulation studies are provided to investigate the extensions to the joint tail approach and conditional extreme approach proposed in Sections 2.5 and 2.6. We also compare the performance of the two approaches against one another on simulated data. In all studies we use bivariate extreme value distributions with logistic dependence structure (BEVL) and the inverted version of this distribution (IBEVL) to investigate performance under distributions with asymptotic dependence and asymptotic independence respectively. By default we shall follow the setting of the simulation study in Heffernan and Tawn (2004). Unless stated we shall simulate 200 replicate data sets each containing 5000 data points from the relevant distribution with $\gamma = 0.5$ and a critical level set at the 90th quantile, such that 10% of data points exceed the threshold.

### 2.7.1 Testing the effect of marginal information on the joint tail approach

In Section 2.5 an extension to the joint tail approach of Ledford and Tawn (1997) was suggested in which we could incorporate additional information about points that are extreme in at least one variable. We saw that this led to a new estimate for the parameter $c$ which changed from the form in equation (2.5.4) to the form given in equation (2.5.7). The estimate for the parameter $\eta$ stayed consistent across both

approaches. Via a simulation study we aim to observe whether this change in the estimate of $c$ leads to more accurate and efficient estimates of the extremal dependence measure $\chi(v)$.

The simulation study has the standard form outlined above. True values for $\chi(v)$ (denoted $\chi_{true}(v)$) for the extreme value distributions of interest are given by equations (2.3.17) and (2.3.18). These are used to obtain estimates of the root mean squared error (RMSE), i.e.

$$\text{RMSE}\left[\chi(v)\right] = \sqrt{\mathbb{E}\left\{\left[\hat{\chi}(v) - \chi_{true}(v)\right]^2\right\}}.$$

Figure 2.7.1 shows estimates of the RMSE for the two different approaches with data simulated from the BEVL and the IBEVL for a selection of $\gamma$ values. The joint tail model is fitted at the modelling threshold $u$, here set at the 90th quantile. Results are given at two different critical levels, the 90th quantile and the 99th quantile. We observe that the approach that includes the additional marginal information reduces the RMSE for both distributions, at all values of $\gamma$ and at both critical levels. The most noticeable difference occurs for the BEVL with $\gamma$ close to zero. In this situation we have a very strong positive association and as such $\hat{\eta} \to 1$. Equation (2.5.5) implies we have $\hat{\chi}(v) \to \hat{c}$ and therefore any improvements in the estimation of $c$ are directly reflected in more accurate estimates of the extremal measure $\chi(v)$.

Theoretically, since the BEVL is asymptotically dependent we should have $\eta = 1$ at all values of $\gamma < 1$. On the left side of Figure 2.7.1 estimates of $\chi(v)$ are given with $\eta$ fixed at 1. This highlights discrepancies between the estimate $\hat{\eta}$ and the limit value. Let $\hat{c}_\eta$ be the estimate of $c$ obtained with $\eta$ fixed at 1. The estimate $\hat{c}_\eta$ is obtained at the modelling threshold $u$ and by definition $\hat{\chi}(v) = \hat{c}_\eta$ for any higher $v$. Therefore, when the critical level $v$ is set at the 90th quantile, such that $v = u$, the estimates of $\hat{\chi}(v)$ overlap. However at the 99th quantile the different estimates show very different behaviour. Here, when $\gamma$ is closer to zero the simulated data are highly dependent

and therefore assuming the asymptotic limit form can lead to reduced RMSE of $\hat{\chi}(v)$. However, as $\gamma \to 1$ the data become less dependent and as such imposing the asymptotically dependent limit form leads to a large value of the RMSE for $\hat{\chi}(v)$.
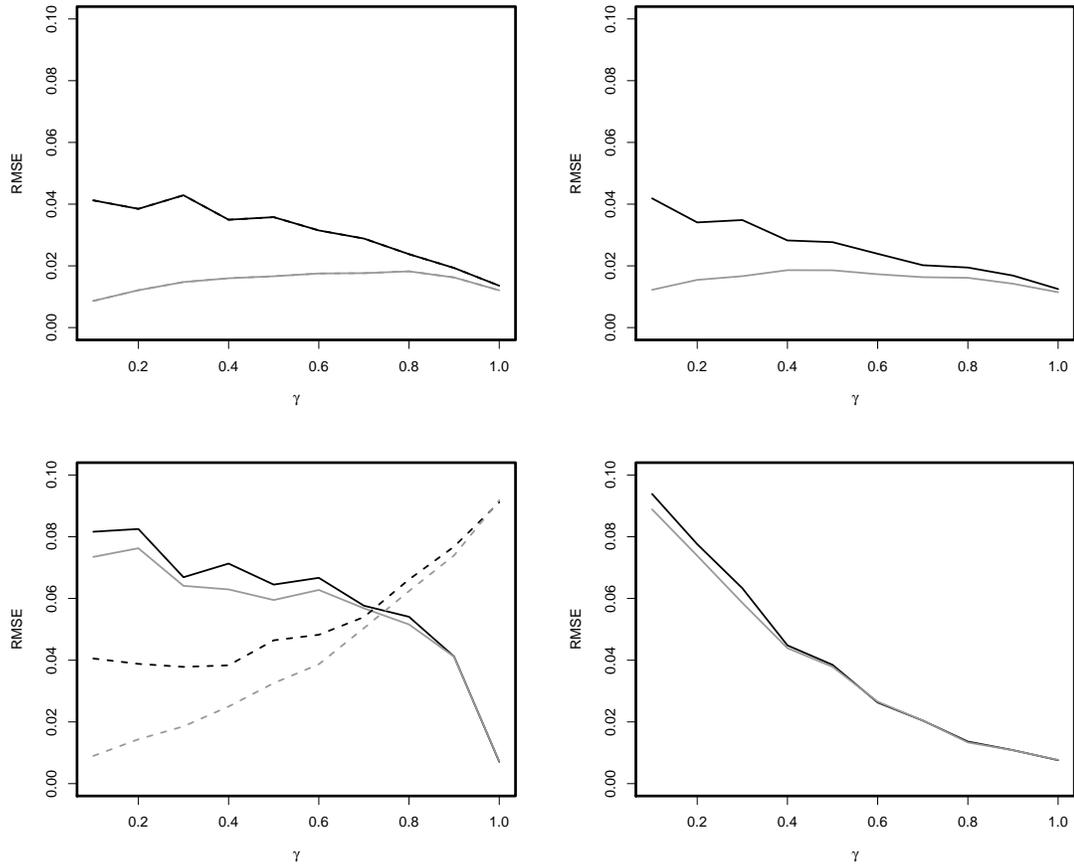


Figure 2.7.1: Estimates of the RMSE of $\hat{\chi}(v)$ for the joint tail approach (black) and the joint tail approach with additional marginal information (grey). Estimates given for data simulated from extreme value distributions with logistic dependence structure (left) and inverted logistic dependence structure (right) at critical level set at 90th quantile (top) and 99th quantile (bottom) for a selection of different values for the logistic dependence parameter $\gamma$; see Section 2.3.2. Dotted lines (shown on the left hand plot for the logistic dependence structure) show RMSE with coefficient of tail dependence $\eta$ fixed at 1, i.e. assuming the limit behaviour (asymptotic dependence) for all values of $\gamma$.

A similar pattern is obtained for the data simulated from the IBEVL at the given

|  | BEVL | | IBEVL | |
|---|---|---|---|---|
| $\gamma$ | 0.5 | 0.75 | 0.5 | 0.75 |
| $\chi(v)$ | 0.616 | 0.376 | 0.385 | 0.208 |
| $\hat{\chi}(v)$ JT | 0.616 | 0.374 | 0.388 | 0.208 |
| $\hat{\chi}(v)$ CE | 0.624 | 0.384 | 0.389 | 0.210 |
| 95% SE CI | (0.548, 0.683) | (0.321, 0.427) | (0.333, 0.442) | (0.167, 0.250) |
| 95% PL CI | (0.552, 0.685) | (0.324, 0.429) | (0.336, 0.443) | (0.171, 0.250) |
| 95% BT CI | (0.587, 0.671) | (0.337, 0.436) | (0.341, 0.442) | (0.166, 0.251) |

Table 2.7.1:  Simulation results for the estimation of $\chi(v)$ under the BEVL and IBEVL with logistic dependence parameters $\gamma = 0.5$ and $\gamma = 0.75$.  The critical level $v$ is fixed at the 90th quantile.  Estimates of $\chi(v)$ for joint tail (JT) and conditional extremes (CE) models evaluated as the mean over 200 replicate data sets of 5000 points.  Standard error (SE) and profile likelihood (PL) confidence intervals are given for the joint tail approach and bootstrapped (BT) confidence intervals are given for the conditional extremes approach.  All confidence intervals are obtained as the averaged endpoints from the different samples.

threshold level.  We observe that the RMSE is improved at all levels by the approach with additional marginal information, although any benefit is diluted at higher critical levels when the effect of $\hat{c}$ on $\hat{\chi}(v)$ is balanced by changes in the other terms of equation (2.5.5).

## 2.7.2    Comparing different approaches to modelling extremal dependence

We now compare directly the joint tail approach and conditional extremes approach through a simulation study with the standard settings.  The aim is to assess whether one approach provides significantly more accurate or reliable results when estimating the extremal dependence measure $\chi(v)$.

| | BEVL | | IBEVL | |
|---|---|---|---|---|
| $\gamma$ | 0.5 | 0.75 | 0.5 | 0.75 |
| $\chi(v)$ | 0.589 | 0.324 | 0.148 | 0.043 |
| $\hat{\chi}(v)$ JT | 0.562 | 0.303 | 0.150 | 0.046 |
| $\hat{\chi}(v)$ CE | 0.561 | 0.276 | 0.144 | 0.042 |
| 95% SE CI | (0.495, 0.629) | (0.251, 0.357) | (0.095, 0.205) | (0.005, 0.087) |
| 95% PL CI | (0.428, 0.671) | (0.206, 0.403) | (0.091, 0.237) | (0.021, 0.093) |
| 95% BT CI | (0.485, 0.634) | (0.187, 0.341) | (0.084, 0.219) | (0.014, 0.083) |

Table 2.7.2: Same as Table 2.7.1 but with critical level $v$ set at the 99th quantile.

Results for data simulated from an extreme value distribution with logistic dependence structure are summarised in Tables 2.7.1 and 2.7.2. The average estimate of $\chi(v)$ from the 200 replicate data sets is close to the true value for both the joint tail and conditional extremes models; although in most situations the joint tail approach provides a slightly more accurate estimate. The true value is contained within both types of 95% confidence interval for the former method and the bootstrapped confidence intervals for the latter method. Confidence intervals based upon the standard error seem to coincide the profile likelihood intervals for the joint tail method at the lower critical level. The width of the boostrapped confidence intervals are also very similar for the conditional extremes approach. At the higher critical level the profile likelihood confidence intervals appear to be wider than the standard error confidence intervals for the joint tail, with boostrapped intervals for the conditional extreme approach in general being wider than the standard error intervals but narrower than the profile likelihood intervals.

The conclusions drawn from the simulation study are that both the joint tail and conditional extremes methods outlined in this chapter model bivariate dependence

well in the joint tail region. Investigations into behaviour for a much richer class of distributions can be found in papers by Ledford and Tawn (1996) and Wadsworth and Tawn (2013). As outlined in Section 2.6, the conditional extremes method has greater flexibility. In the bivariate setting the conditional extremes approach gives accurate estimates of a wider class of extremal dependence measures, especially in cases where we might be interested in different joint exceedance regions than $\{X_1 > u, X_2 > u\}$; see Heffernan and Tawn (2004) for an example. The conditional extremes approach can also be extended into a multivariate setting more easily. The similarity of the estimates obtained from the different methods coupled with the benefits of the conditional extremes approach will motivate the use of the conditional extremes method in later chapters.

### 2.7.3 Direct estimation of extremal dependence using conditional extremes approach without simulation

We now investigate an extension of the conditional extremes approach outlined in Section 2.6 that allows for accurate estimation of extremal dependence measures without the need to simulate. This extension is useful in situations where computational power is at a premium. For this study we take the standard settings and vary the number of simulated exceedances generated from the conditional extremes approach to estimate $\chi(v)$. The reliability of this estimate is then compared against direct calculation of $\chi(v)$ from the conditional extremes approach without simulating.

Figure 2.7.2 shows estimates for the RMSE of $\chi(v)$ for the two different approaches with data simulated from the BEVL. We observe that for small numbers of simulated exceedances the RMSE is much larger than for direct estimation. As the number of simulated exceedances is increased the method based upon simulations becomes slightly more reliable. The result for data simulated from the IBEVL is similar and is omitted.
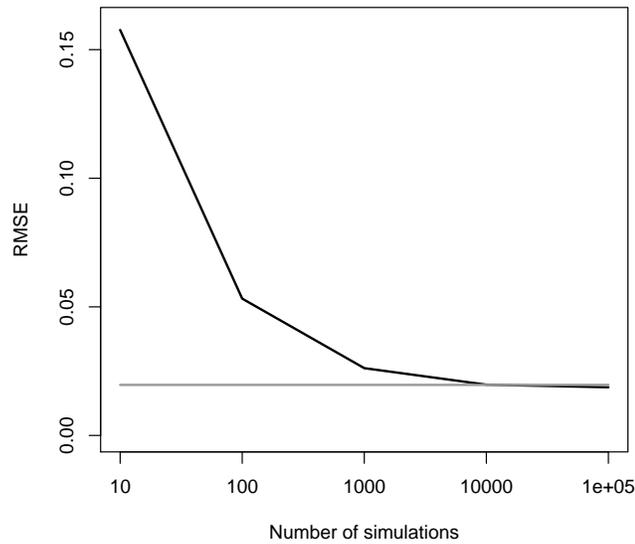
Figure 2.7.2: Estimates of the RMSE for $\chi(v)$ for data simulated from extreme value distribution with logistic dependence structure using the conditional extremes approach with simulated exceedance data sets of different sizes (black, on log scale) and via direct estimation (grey).

We have shown that estimating extremal measures using simulation introduces unnecessary variability which can be controlled by simulating a large number of values. In most practical situations it is not computationally expensive to simulate many exceedances and as such in applications this approach may not have much use. But it is clear that in situations where computational simulations come at a cost, then direct estimation can provide a more reliable alternative.

# Chapter 3

# Modelling heatwaves in central France: a case study in extremal dependence

## 3.1 Introduction

When modelling heatwaves decision makers are most interested in mitigating for disruption and fatalities. The heatwave across Europe in 2003 that caused around 40,000 heat related deaths (Fischer and Schär, 2010) and cost the farming industry around € 13.1 billion highlights the potential large scale effects of such an event. High temperatures reduce the capacity of the human body for heat loss and are likely to cause core body temperature to exceed healthy limits (37-39°C). Most casualties in a heatwave are caused by heat exhaustion which leads to heat stroke. Heat exhaustion increases the blood pressure and leads to cardiovascular stress, which if not relieved results in cellular damage and an increased risk of mortality (Donaldson et al., 2003). Young and old people are particularly vulnerable during heatwave events. A day of strong heat could disrupt certain services for a couple of days but is unlikely to cause many fatalities. Conversely, a long sustained period of moderate to high heat may

not disrupt services but can lead to many fatalities.

A heatwave is defined as a set of hot days and/or nights that are associated with a marked short-term increase in mortality. To make this definition precise we need to clarify what is meant by a hot day and a set of days. A hot day is when the temperature, or a related variable, exceeds a critical threshold level for health. Koppe et al. (2004) proposed threshold definitions based upon air temperature or indices based upon air temperature and relative humidity. Clearly health impacts increase with both the extent of the temperature excess over the critical threshold and the number of days that such an event lasts for. One way to measure the severity of the heatwave is to count the total number of days that the temperature series exceeds the critical level during a meteorological event, which we refer to as the duration of the heatwave event. During an extreme event, a set of days with temperatures below the critical level could allow respite from heat exhaustion and dramatically change the impact of the event so the duration of the heatwave is an insufficient measure for assessing some health implications. In these cases metrics such as the maximum consecutive sequence of exceedances or aggregated temperatures over the event are more appropriate. Abaurrea et al. (2007), Stefanon et al. (2012) and Fischer and Schär (2010) all define a heatwave using a critical temperature threshold corresponding a fixed percentile of daily maximum summer temperatures (in the range 90%-95%) and a specified minimum duration (in the range 1-6 days). Relative critical levels are typically preferred to absolute levels when defining a heatwave since temperature can vary by geographical location and humans are able to adapt to local climate (Nitschke et al., 2011). Although heatwave definitions vary, all correspond to different but well-defined functionals of a meteorological event having temperatures which exceed a critical level.

To estimate the probability of a heatwave we propose a framework based upon extreme value theory. The framework relies on asymptotically justified models for describing

the properties of the time series during an extreme temperature event. A broader and more flexible model with stronger asymptotic justification is proposed here than in previous studies. The model is used to simulate replicate extreme events that exceed a critical level permitting the evaluation of the distribution of any functional of the extreme event and hence the probability of a heatwave with specific characteristics. This approach ensures that this methodology applies to any form of heatwave definition of interest to experts from wide-ranging fields, such as heat-health researchers or those studying economic damage linked to heatwaves. Critically it enables the estimation of the probability of heatwaves occurring in a future period that are more extreme in any functional of interest than any of the observed events.

We apply these generic methods to the modelling of observed daily maximum temperatures to estimate the distribution of heatwaves at Orleans in central France, an area that was affected by the 2003 heatwave event. The hottest observed daily maximum temperature in 2003 for Orleans was 39.9ºC. The summer daily maximum temperature one year return level, defined as the level exceeded on average once every summer, for Orleans is estimated as 35ºC using standard extreme value methods (Coles, 2001). What made the 2003 event so severe for Orleans was that two heatwaves with 2 and 11 consecutive exceedances of the one year level occurred within a four week period. Pascal et al. (2013) quantified the relationship between temperatures and excess mortality over France, finding that if the average of three consecutive daily maximum temperatures exceeds 34ºC (34ºC, 35ºC) excess mortality is 47% (17%, 33%) in Paris (Limoges, Lyon) respectively. Orleans is situated between these three cities and we focus on 35ºC as the critical level for defining heatwaves. We note that the excess mortality from observing such a level is high but can vary between locations. Under the assumption that the summer daily maximum temperature at Orleans, denoted $\{X_t\}$ on day $t$, is a stationary process we will estimate multiple quantities. These quantities include the joint probability of having an event that lasts at least as long and has

peak value at least as severe as the 2003 event and that the average of three consecutive daily maximum temperatures exceeds 35°C. Pascal et al. (2013) also assessed the impact of high daily minimum temperatures coupled with high daily maximum temperatures on excess mortality. Under our framework we are also able to model the joint characteristics of temperature maxima and minima, over any time-scale, during the extreme event. We do not give specific estimates for that case, but in Section 3.6 we outline the modifications to our approach for modelling daily maxima that are required to address this broader concern.

Using empirical methods to estimate probabilities for the extreme heatwave events of interest to us is not possible so models are required. Here we need models for both the intensity and extremal dependence structure that determine properties of events. The intensity of heatwave events can be modelled by fitting an extreme value model to exceedances of a high modelling threshold $u$. The most common approach, which applies under weak conditions, is to fit a generalized Pareto distribution (GPD) to threshold excesses, i.e.

$$\mathrm{P}(X_t - u > x \mid X_t > u) = \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi} \qquad \text{for} \quad x \geq 0, \qquad (3.1.1)$$

where $c_+ = \max(c, 0)$, $\sigma_u > 0$ and $\xi$ are the scale and shape parameters of the GPD respectively (Coles, 2001).

A time-series of temperature data can be split into independent clusters where within each cluster groups of dependent exceedances occur. In the literature of extreme value theory these clusters are defined using different methods; the most popular technique is the runs method (Smith and Weissman, 1994). Under this method a cluster is ended by a sequence of $m$ consecutive non-exceedances of $u$ and a new cluster is commenced with the next exceedance of $u$. The run length $m$ can be chosen subjectively; although Ferro and Segers (2003) outline an automated method. Therefore from a time-series it is possible to obtain the number of independent clusters and the values

in each cluster. The number of clusters is Poisson distributed (Davison and Smith, 1990) so it remains to model the values within a cluster.

Standard asymptotic measures of cluster features are independent of the critical level. Examples include the distribution of the number of exceedances in a cluster, $\{\pi(i), i \geq 1\}$, associated mean cluster size $\theta^{-1}$, where $\theta \in [0, 1]$ is the extremal index (Leadbetter et al., 1983), and other cluster functionals outlined in Smith et al. (1997) and Segers (2003). The focus on heatwaves highlights the need to not only account for the number of exceedances in a cluster, but also the full profile of the event to enable estimation of features such as the distribution of the number of consecutive exceedances or the average of three consecutive values. The application motivates the study of a new distribution $\pi_C(i)$ of the longest set of consecutive exceedances within a cluster along with the associated consecutive extremal index $\theta_C$.

Under a stationary Markov process assumption, the extremal behaviour of $\{X_t\}$ can be modelled by focusing on the joint distribution of $(X_t, X_{t+1})$; more discussion of these assumptions will be given in Section 3.5. Multivariate extreme value theory leads to models for the joint tail through using separate marginal and dependence structures and can be used to assess dependence between $(X_t, X_{t+1})$. Dependence structures can be broadly split into those with asymptotic dependence and those with asymptotic independence (Sibuya (1960), Ledford and Tawn (1996)) determined by the value of $\chi$ where

$$\chi = \lim_{x \to x^*} P(X_{t+1} > x \mid X_t > x), \qquad (3.1.2)$$

with $x^*$ being the upper limit of the support of the common marginal distribution. In the case when $\chi = 0$ the variables $(X_t, X_{t+1})$ are said to be asymptotically independent and $\chi > 0$ corresponds to asymptotic dependence. The assumption of a dependence structure that is asymptotically dependent leads to the duration distribution being approximately independent of the critical level. Smith (1992), Coles et al.

(1994), Smith et al. (1997), Perfekt (1997) and Yun (1998) use a parametric Markov model for estimating extremal quantities when $\chi > 0$. In contrast if the process is an asymptotically independent Markov chain then clusters in the limit reduce to single exceedances and $\theta = 1$ (Bortot and Tawn, 1998). However, if sub-asymptotic thresholds are considered $\mathrm{P}(X_{t+1} > u \mid X_t > u) > 0$, for $u$ as in equation (3.1.1), even when $\chi = 0$ and models are required that can capture this dependence as well. In these cases the duration and level of events are not independent.

The semi-parametric conditional extremes approach of Heffernan and Tawn (2004) offers a more flexible way of estimating extremal quantities of Markov chains than existing methods. This is due to a richer class of extremal dependence properties are permitted than those of Smith et al. (1997). These properties also hold over a much broader tail region than the parametric approach of Bortot and Tawn (1998). The approach of Bortot and Tawn (1998) provided models with asymptotic justification for $(X_t, \ldots, X_{t+m})$ only in the region with $X_i > u$ for all $i = t, \ldots, t+m$ and $u$ a high threshold whereas we need models that hold for this vector subject only to $X_t > u$. The inclusion of dependence structures that also exhibit asymptotic independence permits the distribution of duration of events to change with critical level. Asymptotic dependence is a special case in the conditional extremes approach that does not require the evaluation of a parametric model; see Section 3.3.4 for more details. The non-parametric method of estimating extremal quantities of Markov chains, outlined in Section 3.3.4, can be compared to previous studies that assume a parametric dependence structure with asymptotic dependence.

In Section 3.2 the definition of a cluster and distributions of exceedances are formalised. Different approaches to model extremal dependence are outlined in Section 3.3. Section 3.4 discusses techniques for summarising the behaviour of clusters and compares the values of $\theta$ and $\theta_C$. Section 3.5 presents the temperature data for

Orleans, model fit diagnostics and results concerning the probability of observing the events of interest identified above. We focus on applying the conditional extremes approach and demonstrate how results differ from other approaches and show diagnostics that support our approach. Discussion and conclusions are presented in Section 3.6.

## 3.2   Cluster features

To understand clustering of time-series extremes it is necessary to formalise the asymptotic definition of a cluster and to provide a range of summaries. For a series $\{X_t, t = 1, \ldots, n\}$ specify a threshold level $u_n$ and a block of length $m_n$ such that as $n \to \infty$, $u_n \to x^*$, with $x^*$ as defined for equation (3.1.2), such that $nP(X_t > u_n) \to \tau > 0$ as $n \to \infty$ and $m_n = o(n)$. Under suitable long-range mixing conditions the normalised process of times of exceedances of $u_n$, i.e.

$$\left\{ \frac{t}{n+1}; t = 1, \ldots, n, X_t > u_n \right\},$$

converges to a compound Poisson process (Hsing, 1988). Since the assumption of stationarity has been made the cluster of interest can always be moved to the start of the time-series and such we can look at values such that $X_1 > u_n$. A cluster in block $\{1, \ldots, m_n\}$ of this process is a set of exceedances of $u_n$ by $X_t$ for $t = 1, \ldots, m_n$. The number of such exceedances is

$$N(u_n, m_n) = \#\{X_i > u_n \quad \text{for} \quad i = 1, \ldots, m_n\},$$

and hence a cluster occurs when $N(u_n, m_n) \geq 1$. By this definition, clusters do not need to constitute consecutive exceedances, the exceedances only need to be close in time. The cluster size distribution $\pi(i, u_n)$ is defined as

$$\pi(i, u_n) = P(N(u_n, m_n) = i \mid N(u_n, m_n) \geq 1) \qquad \text{for} \quad i = 1, \ldots, m_n.$$

Using this definition, it can be seen that $\pi(i, u_n)$ is the probability of obtaining $i$ exceedances of threshold $u_n$ in a block of $m_n$ values given that there is at least one

exceedance (i.e. there is a cluster). From this

$$\pi(i) = \lim_{n\to\infty} \pi(i, u_n) \qquad \text{for} \quad i = 1, 2, \ldots, \tag{3.2.1}$$

is the limiting probability of a cluster of size $i$. A widely discussed dependence measure is the extremal index. This measure is the reciprocal of the mean of the cluster size distribution of the extremes in a time-series (Leadbetter, 1983). In terms of equation (3.2.1) the extremal index $\theta$ can be written as

$$\theta^{-1} = \sum_{i=1}^{\infty} i\pi(i).$$

An alternate form for the extremal index is characterised in O'Brien (1987) in the form of $\theta = \lim_{n\to\infty} \theta(u_n, m_n)$ where

$$\theta(u_n, m_n) = \mathrm{P}\left(X_2 \leq u_n, \ldots, X_{m_n} \leq u_n \mid X_1 > u_n\right), \tag{3.2.2}$$

which links to the runs estimator, discussed in Section 3.1, with run length $m_n$. The distribution $\pi(i)$ can be defined (Rootzén, 1988) as

$$\pi(i) = \frac{\theta^{(i)} - \theta^{(i+1)}}{\theta^{(1)}} \qquad \text{for} \quad i = 1, 2, \ldots, \tag{3.2.3}$$

where

$$\theta^{(i)}(u_n, m_n) = \mathrm{P}(N(u_n, m_n) = i \mid X_1 > u_n) \qquad \text{for} \quad i = 1, \ldots, m_n, \tag{3.2.4}$$

defines the probability of viewing $i$ exceedances of a threshold in a block of values given that the first value ($X_1$) exceeded the threshold and

$$\theta^{(i)} = \lim_{n\to\infty} \theta^{(i)}(u_n, m_n) \qquad \text{for} \quad i = 1, 2, \ldots. \tag{3.2.5}$$

This alternative approach to evaluating $\pi(i)$ is beneficial as it requires the evaluation of the process conditional on $X_1 > u_n$, in contrast to the evaluation of $N(u_n, m_n)$ which starts from an arbitrary $X_1$, and hence it is more efficient for computational purposes.

For heatwaves it is also important to model the number of consecutive exceedances.

This can be accomplished using the distribution $\pi_C(i)$ stated in Section 3.1. Specifically, let $C^{(i)}(u_n, m_n)$ be the event

$$C^{(i)}(u_n, m_n) = \{X_1 > u_n, \ldots, X_i > u_n, X_{i+1} < u_n \cap \nexists t = 2, \ldots, m_n : X_t > u_n, \ldots, X_{t+i-1} > u_n\},$$

i.e. that at time 1 there is a run of $i$ consecutive exceedances but at no later time in the cluster does a run of this length or more occur. This leads to a measure that is analogous to equations (3.2.4) and (3.2.5), namely

$$\theta_C^{(i)}(u_n, m_n) = \mathrm{P}(C^{(i)}(u_n, m_n) \mid X_1 > u_n) \qquad \text{for} \quad i = 1, \ldots, m_n,$$

with

$$\theta_C^{(i)} = \lim_{n \to \infty} \theta_C^{(i)}(u_n, m_n) \qquad \text{for} \quad i = 1, 2, \ldots.$$

Note that both $\theta_C^{(1)}$ and $\theta^{(1)}$ are equal to $\theta$ by equation (3.2.2) since, in both situations, the event of interest is $\{X_2 < u_n, \ldots, X_{m_n} < u_n \mid X_1 > u_n\}$. The distribution of the maximum number of consecutive exceedances within a cluster $\pi_C(i)$ is defined as

$$\pi_C(i) = \frac{\theta_C^{(i)} - \theta_C^{(i+1)}}{\theta_C^{(1)}} \qquad \text{for} \quad i = 1, 2, \ldots.$$

The average length of the longest set of consecutive exceedances in a cluster is given by the reciprocal of the consecutive extremal index $\theta_C$, defined as

$$\theta_C^{-1} = \sum_{i=1}^{\infty} i \pi_C(i).$$

An event that has one exceedance in a cluster directly implies a maximum of one consecutive exceedance whereas the counter implication is not true, and hence $\pi_C(1) \geq \pi(1)$. As a consequence $\pi_C(i)$ experiences a sharper decline than $\pi(i)$ as $i$ is increased. It should be noted that $\pi_C$ is not geometric as the Markov process applies to the level as opposed to only whether the exceedance is above or below the threshold.

Smith et al. (1997) investigate the behaviour of the extremal index $\theta$ against the parameters in an underlying Markov chain model. Here interest is in $\theta_C$ and so using

the methodology outlined in Section 3.4 for evaluating cluster functions, Figure 3.2.1 compares the inverted extremal index $\theta^{-1}$ and the inverted consecutive extremal index $\theta_C^{-1}$ for a Markov chain with bivariate extreme value distribution with logistic dependence (Tawn, 1988) between consecutive values. Here a range of parameter values for the logistic dependence parameter are used, $\gamma \in (0, 1]$. A near perfect linear relationship is observed which shows that $\theta_C^{-1} \approx 0.7\theta^{-1}$. This shows that groups of consecutive exceedances are on average 30% shorter than the average cluster size for this dependence model.
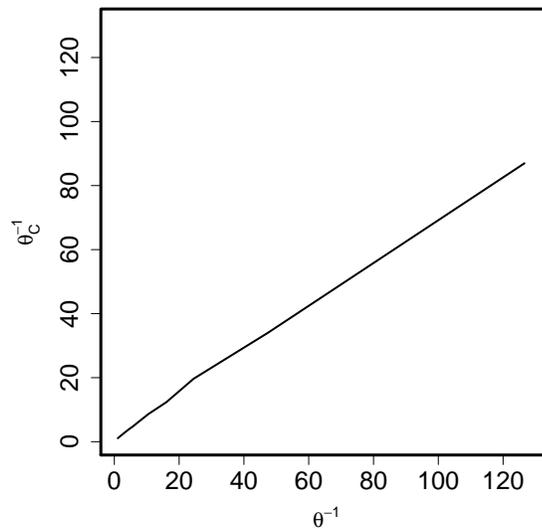


Figure 3.2.1: Relationship between inverted extremal index $\theta^{-1}$ and inverted consecutive extremal index $\theta_C^{-1}$; both are functions of logistic dependence parameter $\gamma$, for a Markov chain with bivariate extreme value distribution with $\gamma$ from 0.05 to 1.

## 3.3 Modelling temporal dependence

### 3.3.1 Markov modelling

To obtain estimates for $\pi(i)$ and $\pi_C(i)$ and their sub-asymptotic equivalents it is necessary to develop a model for the evolution of the temperature data through time.

Many types of parametric and non-parametric models could be constructed. Here, supported by exploratory data analysis, an assumption that the time series follows a first order Markov process is made. By the Markov property the distribution at each time step is only affected by the state of the system at the previous time-step with the resulting joint density of $(X_1, \ldots, X_n)$ given by

$$f(x_1) \prod_{t=1}^{n-1} f(x_{t+1}|x_t). \tag{3.3.1}$$

The assumption greatly simplifies the modelling process since to model the extremes of a time series $X_1, \ldots, X_n$ it is only necessary to model the extremes of pairs $(X_t, X_{t+1})$ for $t = 1, \ldots, n-1$, which have joint distribution function $F(x_t, x_{t+1})$.

The likelihood from equation (3.3.1) can be further simplified for our modelling approach. Specifically, we shall assume a parametric model for the marginals when the variables are above a threshold $u$ and a parametric model for the conditional density $f(y \mid x)$ for $x > u$. We denote the marginal parameters by $\boldsymbol{\phi}$ and the extra conditional (dependence) parameters by $\boldsymbol{\theta}$. Our approach is to first estimate the marginal parameters and then estimate the dependence parameters. For the latter the likelihood simplifies to

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n_u - 1} f(x_{t_i+1} \mid x_{t_i}) \tag{3.3.2}$$

where $n_u$ us the number of $\{x_t\}$ exceeding $u$ and $t_1, \ldots, t_{n_u}$ are the time indices of these exceedances. The otehr likelihood terms from equation (3.3.1) disappear into the constant of proportionality as they provide no information on $\boldsymbol{\theta}$ under our model.

The model for the marginal exceedances of the threshold $u$ is given in Section 3.3.2. Our main method for modelling dependence is presented in Section 3.3.3 and connections with other models are discussed in Section 3.3.4. Our method is based upon the conditional approach outlined in Heffernan and Tawn (2004). That model allows for a

rich class of dependence structures and most importantly allows for asymptotic independence which models the interaction between the duration distribution of an event and a critical level. Asymptotic dependence is a special case within the conditional approach, which motivates a new non-parametric approach and enables comparisons with the methods of Smith et al. (1997).

### 3.3.2 Marginal modelling

Following the assumption of stationarity of $\{X_t\}$ the marginal distributions of $F$ are identically distributed. The assumption (3.1.1) for a GPD for the marginal excesses of $u$ leads to the model for the common marginal distribution

$$F(x) = \begin{cases} 1 - \lambda_u \left(1 + \xi \frac{x-u}{\sigma_u}\right)_+^{-1/\xi}, & x \geq u \\ \tilde{F}(x), & x < u, \end{cases} \tag{3.3.3}$$

where $\lambda_u = 1 - F(u)$ and $\tilde{F}(x)$ is the empirical cumulative distribution function of $\{X_t\}_{t=1}^n$. Marginal parameters are estimated using a censored likelihood approach; see Davison and Smith (1990). For modelling extremal dependence we need to select an appropriate margin to transform onto. In copula methods (Nelson, 2007) it is common to model dependence with uniform margins, but for extremes simplifications in model form arise when focusing on a different marginal choice. Heffernan and Tawn (2004) model dependence for Gumbel margins. Keef et al. (2013) showed that a more comprehensive approach arises for Laplace margins. Following Keef et al. (2013) we transform $X_t$, $t = 1, \ldots, n$ onto Laplace margins as follows

$$T(X_t) = \begin{cases} \log\{2F(X_t)\} & \text{if} \quad X_t < F^{-1}(0.5) \\ -\log\{2\left[1 - F(X_t)\right]\} & \text{if} \quad X_t \geq F^{-1}(0.5). \end{cases}$$

### 3.3.3 Semi-parametric conditional extremes approach

The conditional extremes method of Heffernan and Tawn (2004) and Heffernan and Resnick (2007) can be used to motivate a modelling framework for which $\chi$, defined

by equation (3.1.2), can be either positive or zero. The desire is to model the joint distribution $[T(X_t), T(X_{t+1})]$ using the distribution of $T(X_{t+1})$ given that $T(X_t)$ is large (defined as exceeding a high threshold). A requirement for modelling the conditional distribution $\mathrm{P}\{T(X_{t+1}) \leq T(x_{t+1}) \mid T(X_t) = T(x_t)\}$ is that this distribution should be non-degenerate as $x_t \to x^*$. As such the Heffernan and Tawn (2004) approach aims to identify normalizing functions $a : \mathbb{R}_+ \to \mathbb{R}$ and $b : \mathbb{R}_+ \to \mathbb{R}_+$ that are defined such that for $x > 0$

$$\mathrm{P}\left(\frac{T(X_{t+1}) - a\left[T(X_t)\right]}{b\left[T(X_t)\right]} \leq z, \frac{X_t - u}{\sigma_u} > x \mid X_t > u\right) \to G(z)\left(1 + \xi x\right)_+^{-1/\xi}, \quad (3.3.4)$$

as $u \to x^*$, where $G$ is a non-degenerate distribution function and $\sigma_u$ is as in equation (3.1.1). The specification of Laplace margins ensures that the upper and lower tails are symmetric and exponential which permits the definition of a single parsimonious class of choices for the normalising functions of

$$a(y) = \alpha y \qquad \text{and} \qquad b(y) = y^\beta,$$

where $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1)$. This form of the normalising functions does not affect the limiting dependence model in Heffernan and Tawn (2004) and simplifies the inference for variables which are either negatively or weakly associated. If the variables are independent, $\alpha = \beta = 0$ and $G(z)$ is the Laplace distribution function whereas $\alpha = 1$ and $\beta = 0$ corresponds to the situation of asymptotic dependence (given by $\chi > 0$ in equation (3.1.2)) and $-1 \leq \alpha \leq 0$ to negative dependence.

Modelling using the conditional extremes approach requires the assumption that the limiting form of equation (3.3.4) holds exactly for all values of $X_t > u$ given that $u$ is a sufficiently high threshold, i.e. $\lambda_u$ is small. Given this assumption it is possible to write the form of $X_{t+1}$ given that $X_t > u$ as

$$T(X_{t+1}) = \alpha T(X_t) + T(X_t)^\beta Z_{t+1}, \quad (3.3.5)$$

where $Z_{t+1}$ is a random variable with distribution function $G$. We also have that $Z_{t+1}$ is independent of $X_t$ and, following the stationary Markov process assumption, the

sequence of $\{Z_t\}$ are independent and identically distributed. As $G$ does not take any simple parametric form, to estimate $\alpha$ and $\beta$ a false working assumption is made, as in Keef et al. (2013), that $Z_{t+1} \sim N(\mu, \sigma^2)$ and as such

$$T(X_{t+1}) \mid \{T(X_t) = y\} \sim N\left(\alpha y + \mu y^\beta, \sigma^2 y^{2\beta}\right) \qquad \text{for} \quad y > T(u). \qquad (3.3.6)$$

The working assumption permits the estimation of the set of parameters $(\alpha, \beta, \mu, \sigma)$ by standard likelihood approaches. Specifically, the likelihood in equation (3.3.2) is maximised with the conditional density given by equation (3.3.6) and parameters $(\alpha, \beta, \mu, \sigma^2)$. At this stage the estimates for $(\mu, \sigma)$ are discarded and a non-parametric estimate of the distribution for $Z$ is formed by inverting equation (3.3.5) to give estimated values of $Z_{t+1}$. Specifically, let $t_1, \ldots, t_{n_u}$ be the indices of $t = 1, \ldots, n$ where $x_t > u$ and where $n_u$ is the number of data points exceeding the threshold $u$. Then let

$$\hat{z}_j = \frac{T(x_{t_j+1}) - \hat{\alpha} T(x_{t_j})}{T(x_{t_j})^{\hat{\beta}}}, \qquad (3.3.7)$$

for $j = 1, \ldots, n_u$. In this way a non-parametric estimate $\hat{G}$ to the distribution function $G$ is formed using $\hat{z}_j$, $j = 1, \ldots, n_u$.

Note that under asymptotic dependence, i.e. $a(y) = y$ and $b(y) = 1$, the transition probability (3.3.4), when expressed in terms of the original variable $X_t$, is given by

$$P\left(X_{t+1} \leq X_t + z \left[\sigma_u + \xi(X_t - u)\right]_+ \mid X_t > u\right) \to G(z). \qquad (3.3.8)$$

Furthermore, we can rewrite $\sigma_u = \sigma_0 + \xi u$ and substitute into equation (3.3.8) which gives

$$P\left(X_{t+1} \leq X_t + z \left[\sigma_0 + \xi X_t\right]_+ \mid X_t > u\right) \to G(z),$$

which shows that the probability is well defined as $u \to \infty$. Under an asymptotic dependence assumption in the semi-parametric conditional approach it is known that

$\alpha = 1$ and $\beta = 0$. In this situation $G$ is estimated by the empirical distribution of the differences in the original data on the Laplace scale (later referred to as the non-parametric approach), i.e. using the sample $j = 1, \ldots, n_u$ of

$$\hat{z}_j = T(x_{t_j+1}) - T(x_{t_j}), \text{ for } x_{t_j} > u. \tag{3.3.9}$$

### 3.3.4 Connections with alternative approaches

Smith et al. (1997) propose a parametric Markov model for the joint distribution of consecutive values of the time series which limits the dependence structure to asymptotic dependence or exact independence. Here we show the connections between that modelling approach and the conditional approach outlined in Section 3.3.3. This enables us to show the benefits for modelling the data of relaxing the strong assumptions of asymptotic dependence and a parametric model.

Based on an asymptotic approximation for a high threshold $u$, Smith et al. (1997) propose a bivariate extreme value distribution copula with GPD marginal tails for the joint distribution function $F(x_1, x_2)$ of $(X_t, X_{t+1})$. This joint distribution is given as

$$F(x_1, x_2) = \exp\left\{-\int_0^1 \max\left(\frac{w}{z_1}, \frac{1-w}{z_2}\right) 2dH(w)\right\} \qquad \text{for} \quad x_1 > u, x_2 > u,$$

where

$$z_j = -1/\log\left[1 - \lambda_u \left(1 + \xi \frac{x-u}{\sigma_u}\right)_+^{-1/\xi}\right] \text{ for } j = 1, 2,$$

and $H$ is an arbitrary distribution function on $[0, 1]$ satisfying the moment constraint

$$\int_0^1 wdH(w) = 1/2.$$

The corresponding transition probability for extreme $X_t$ is given by

$$P\left(X_{t+1} \leq x_t + z\left[\sigma_u + \xi(x_t - u)\right]_+ \mid X_t = x_t\right) \to 2\int_{\left[1+(1+\xi z)_+^{1/\xi}\right]^{-1}}^1 wdH(w), \tag{3.3.10}$$

as $x_t \to x^*$. Following results in Heffernan and Resnick (2007) and Wadsworth et al. (2016), equations (3.3.8) and (3.3.10) are equivalent expressions, under certain smoothness assumptions, despite having slightly different conditioning and limit setups. This gives a formulation for $G$ in this case but also shows that the semi-parametric conditional approach directly extends the approach of Smith et al. (1997).

Smith et al. (1997) make the additional assumption of a parametric model for $H$, exploring a range of models; see Kotz and Nadarajah (2000) for more models. We follow Smith (1992) and assume the logistic dependence structure with parameter $\gamma$. This gives the joint distribution to be

$$F(x_1, x_2) = \exp \left\{ - \left( z_1^{-1/\gamma} + z_2^{-1/\gamma} \right)^{\gamma} \right\},$$

where $\gamma \in (0, 1]$. Independent variables correspond to $\gamma = 1$ and perfectly dependent variables are given as $\gamma \to 0$. For intermediate values of $\gamma$ there is asymptotic dependence with $\chi = 2 - 2^{\gamma}$. Inference for this parametric family is through the censored likelihood approach of Smith et al. (1997). For this parametric model it follows that $G(z) = [1 + \exp(-z/\gamma)]^{\gamma-1}$. In contrast to this parametric form for $G$ the empirical distribution of the sample given by expression (3.3.9) offers greater flexibility for the semi-parametric conditional approach even under an assumption of asymptotic dependence. For more information about non-parametric approaches for multivariate extremes under asymptotic dependence see de Haan and Ferreira (2006).

## 3.4 Cluster behaviour estimation

When analysing the behaviour of heatwaves we can look at within cluster and over cluster results. The work in the previous sections has concentrated on within-cluster behaviour since the definition of the distributions in Section 3.2 are conditional upon a cluster occurring. It is more relevant for applications to have the probability of observing a cluster with specific characteristics in a certain time period. Here we first

discuss within cluster behaviour and then discuss how this is extended to over cluster results.

Our approach to deriving the properties of clusters of a Markov chain is the repeated simulation of a segment of the chain in periods with exceedances of a critical level $v$, i.e. when the process exceeds $v$, with $v \geq u$, where $u$ is our modelling threshold. There are two different strategies for the generation of the chain in its tail state, known as the tail chain. Smith et al. (1997) suggest simulating the cluster maximum $M > v$ and then simulating forwards and backwards from this. A different method (Rootzén, 1988) involves the simulation of an exceedance of $v$, i.e. $X_1 > v$, and only requires forward simulation. Cluster properties such as $\theta(v, m)$ and $\pi(v, m)$ can be estimated empirically from repeated simulations of clusters. For example, $\theta(v, m)$ is estimated as either the reciprocal of the average cluster length using the Smith et al. (1997) approach or as the probability $\theta^{(1)}(v, m)$ in equation (3.2.4) using the Rootzén (1988) approach. For general functionals the Smith et al. (1997) approach can always be used. The Rootzén (1988) approach is easiest to implement but in practice requires additional steps to ensure non-negativity of the distributions $\pi$ and $\pi_C$; since a Monte-Carlo approach can lead to $\theta^{(i)} < \theta^{(i+1)}$ (occasionally for large $i$) which by equation (3.2.3) could lead to $\pi(i, v) < 0$. We use the pool adjacent violators (PAV) algorithm to account for this; for more information see Appendix C. For notational simplicity we define $N$ as the number of exceedances above the critical level $v$ in a cluster and $N_C$ as the maximum number of consecutive exceedances in a cluster. For the rest of this section construction of one simulated chain will be discussed; $\pi$, $\pi_C$ and other useful cluster features are evaluated using repeated simulation (2 million tail chains in Section 3.5). Importance sampling of the initial simulated exceedance is useful to obtain accurate estimates of $\pi(i, v)$ and $\pi_C(i, v)$ for large $i$ in practice but omitted from the discussion below.

The semi-parametric conditional extremes approach is used to generate realisations of a tail chain. Details of the algorithm are given in Appendix B. We require a starting exceedance of $v \geq u$ to be generated from a GPD$(\sigma_v, \xi)$, where $\sigma_v = \sigma_u + \xi(v - u)$, and step forward until a chain of length $k$ is simulated. For sufficiently high thresholds the GPD is an appropriate distribution for simulating cluster maxima; at lower levels it may be necessary to simulate cluster maxima using the distribution in Eastoe and Tawn (2012), which for high thresholds converges to the GPD. This length $k$ is chosen large enough to ensure a negligible probability of simulating any more exceedances of $v$ ($k = 40$ is found sufficient in Section 3.5). As we are specifying a model that is valid in the tail it is most appropriate to use the approach when $X_t > u$. For $X_t < u$ we continue to use the algorithm as it should still provide a reasonable approximation unless $X_t \ll u$. In this case the probability of the chain coming above $u$ again is negligible within a reasonable time horizon, as the tail chain we are using has a negative drift.

It may be of interest to work out how long a heatwave event might last given that the peak value of the cluster, $M$, is known to be greater than or equal to $\eta$ with $\eta \geq v$. Such a question cannot be evaluated efficiently using the forward tail chain methods described above. We use forward and backward tail chains starting from the peak value $M$. A simulation scheme for the conditional extremes approach to evaluate $\mathrm{P}\,(N = i \mid M = \eta)$ is outlined in Appendix B. The probability $\mathrm{P}\,(N_C = i \mid M = \eta)$ for the number of consecutive exceedances $N_C$ above $v$ given the maximum is $\eta$ can similarly be evaluated. The distribution of the number of exceedances given a maximum greater than $\eta$, where $\eta \geq v$, is given by the integral

$$\mathrm{P}\,(N = i \mid M \geq \eta) = \int_\eta^\infty \mathrm{P}\,(N = i \mid M = s) \frac{1}{\sigma_v} \left[ 1 + \xi \frac{s - v}{\sigma_v} \right]_+^{-1/\xi - 1} ds,$$

which is evaluated in practice using a Monte Carlo approximation. Similar simulation schemes can be produced for the parametric (Smith et al., 1997) and non-parametric approaches for asymptotically dependent tail chains; more details are given in Appendix B.

Extending within cluster results to over cluster results requires the assumption that clusters of the modelling threshold $u$ occur as a Poisson process (Hsing, 1988). The mean number of clusters in period $T$ is given by

$$\tau_u = \theta(u, m)\lambda_u n_T,$$

where $\theta(u, m)$ is the sub-asymptotic extremal index at $u$ for run length $m$ given by expression (3.2.2), $\lambda_u$ is the threshold exceedance probability from equation (3.3.3) and $n_T$ is the number of observations within period $T$. For example, with daily data if the rate of clusters within a summer (92 day period from June to August) is desired then $n_T = 92$. At a higher level $v > u$ the mean number of clusters in period $T$ is

$$\tau_v = \theta(v, m)\lambda_u n_T \left[1 + \xi\left(\frac{v - u}{\sigma_u}\right)\right]_+^{-1/\xi}$$

$$= \tau_u \frac{\theta(v, m)}{\theta(u, m)} \left[1 + \xi\left(\frac{v - u}{\sigma_u}\right)\right]_+^{-1/\xi},$$

where the change from $\theta(u, m)$ to $\theta(v, m)$ takes the change in mean cluster size at each level into account and the final term adjusts for the marginal rarity.

It is interesting to know the probability $\psi_v(\kappa, \eta)$ of observing at least one cluster in a period $T$ with a desired extremal property. One example is an event that lasts at least $\kappa$ days above level $v$ and attains a peak value of at least $\eta$. For this example

$$\psi_v(\kappa, \eta) = \sum_{j=0}^{\infty} \left\{1 - \left[1 - \bar{\Pi}_v(\kappa, \eta)\right]^j\right\} \frac{\tau_v^j \exp\left(-\tau_v\right)}{j!}$$

$$= 1 - \exp\left[-\tau_v \bar{\Pi}_v(\kappa, \eta)\right], \tag{3.4.1}$$

where the summation is taken over the number of clusters of the level $v$ and

$$\bar{\Pi}_v(\kappa, \eta) = P\left(N \geq \kappa, M \geq \eta \mid M > v\right).$$

Similar results can be derived with the number of consecutive exceedances $N_C$ as the quantity of interest. Such measures are required when evaluating the probability of observing events at least as severe as the 2003 heatwave.

## 3.5 Heatwave application

### 3.5.1 Data

Daily temperature observations were taken at Orleans, in central France, for the period 1946-2012. Four missing values exist in the time-series and are omitted; none occur during the 2003 event. Heatwaves are most likely to occur in summer months, here defined as the 92 day period of June-August, so summer season and yearly return levels are equivalent. These three month periods are extracted from each year to form an approximately stationary time-series for the temperature. Sample auto-correlation and partial auto-correlation functions support the assumption of a first-order Markov chain; see Figure 3.5.1 for the latter. Since the partial auto-correlation function is affected by heavy tails, we evaluated this function with the data transformed onto Gaussian margins but found no significant change. As such a first-order Markov model is adopted within each summer period and each summer period is treated as independent of others. Figure 3.5.1 shows consecutive pairs of the temperature data illustrating strong inter-day dependence.

In Figure 3.5.2 we provide a boxplot of the daily maximum temperatures to assess the validity of the assumption that the data are stationary within each year. It can be seen that the data exhibit some non-stationarity with higher temperatures at the end of July and start of August, however here for modelling simplicity we still make the assumption that the data are stationary within season. However it is noted that this unmodelled seasonal variation may be contributing to the choice of first-order Markov structure given by the PACF in Figure 3.5.1; see Chapter 4 for more information.

### 3.5.2 Problem and strategy

We want to estimate the probability of observing events such as a heatwave that is more extreme than the 2003 event or that exceeded a specified level of increased mor-
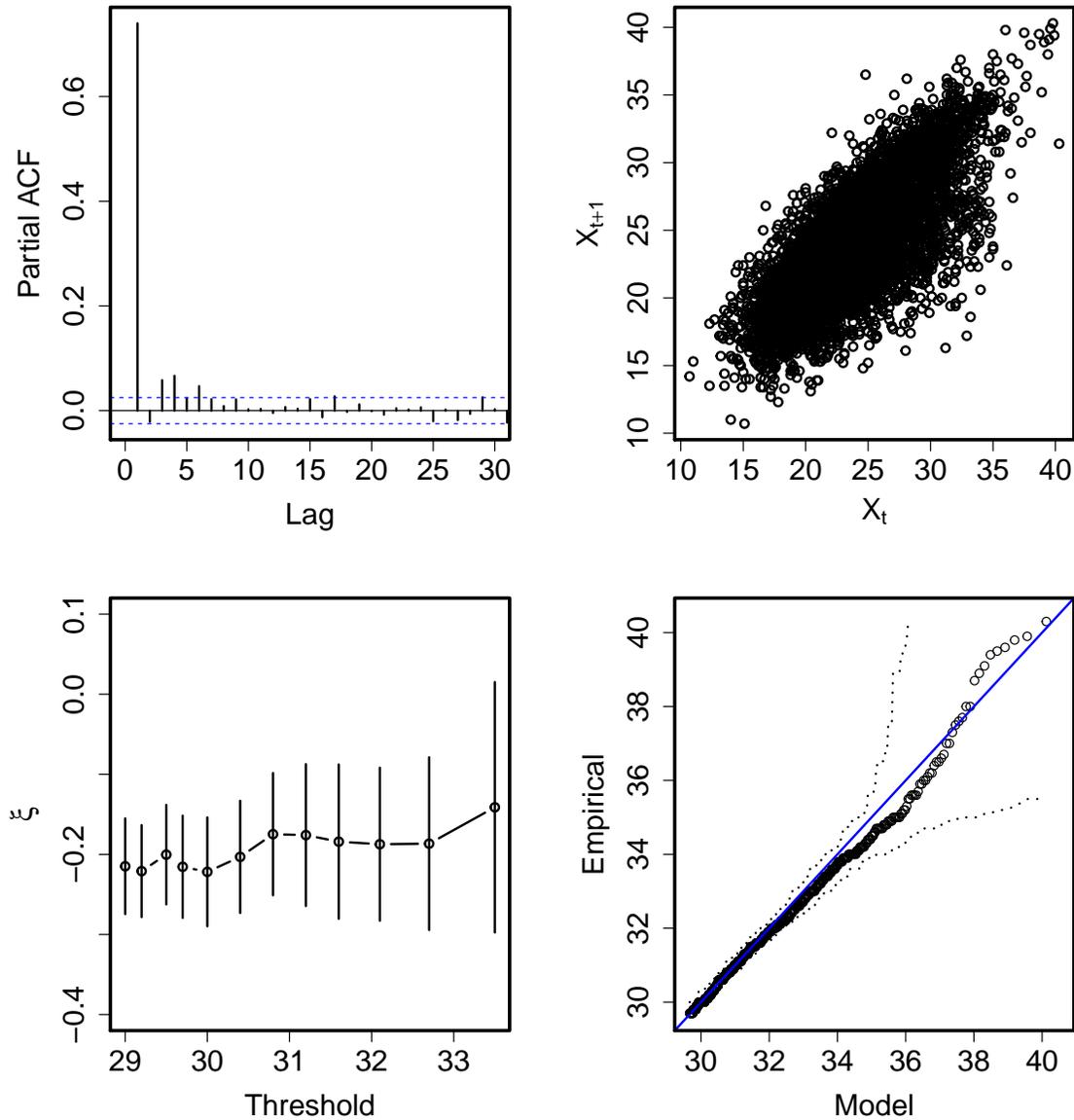
Figure 3.5.1: Diagnostic plots for June, July and August temperature data ($^o$C) for Orleans: partial autocorrelation (top left), scatter of consecutive pairs (top right), parameter stability plot for GPD shape parameter $\xi$ (bottom left) and QQ-plot of GPD fit with 95% tolerance bounds indicated by the dotted lines (bottom right).

tality. We use the empirical runs estimator with a run length of 3 days to correspond to the typical propagation of weather systems. Under this method a cluster is ended by a sequence of 3 consecutive non-exceedances of the chosen critical level and a new
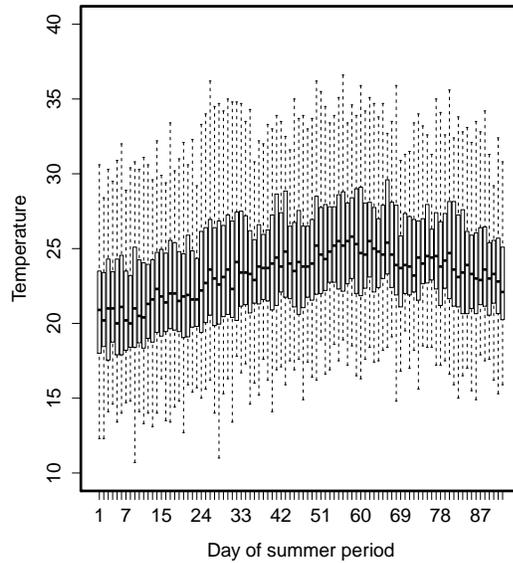
Figure 3.5.2: Boxplot of temperatures for summer days (June-August) for daily maximum temperature data (°C) at Orleans from 1946-2012; see Section 3.5.1 for more information about the data used.

cluster is commenced with the next exceedance of the critical level. A larger choice for the run length in practice will make little difference. Using the runs method with a run length of 3 and a critical level equal to the one year return level (denoted $v_1$, taking the value 35°C), two independent clusters with 2 and 11 consecutive exceedances respectively are identified within a four week period in 2003. It is expected that the daily maximum temperature series exceeds the 1 year return level on average once a summer. It is highly unlikely to observe 13 exceedances in a year, in particular in Section 3.5.3 we show that, using the runs method with the run length and critical level given above, on average we would expect to see only two exceedances for each cluster that exceed such a level.

Empirical estimates of cluster features based upon the runs method are affected by the choice of run length and cannot be used to estimate the required probabilities since they are higher levels than have been observed. We use models from Section 3.3

to provide estimates of the required extremal quantities. In this study the methods outlined in Section 3.3.3 are named the conditional and non-parametric approaches respectively and the method outlined in Section 3.3.4 is named the parametric approach. A comparison between the three approaches and the empirical estimate obtained via the runs method is used for model diagnostic purposes and is given in Section 3.5.3. Results regarding the probability of observing heatwaves with the characteristics of interest are given in Section 3.5.4.

Confidence intervals for all four approaches are generated by bootstrap methods. Runs method declustering defines $n_c$ clusters of varying length and by alternately sampling clusters of exceedances and clusters of non-exceedances from this observed set we generate a bootstrapped sample. This procedure is repeated to generate 1000 replicate data sets to which the models of Section 3.3 are fitted. Repeated simulation is used to obtain estimates of cluster functionals such as $\pi(i)$ and $\pi_C(i)$ as discussed in Section 3.4. Bootstrapped 95% confidence intervals for $\pi$ and $\pi_C$ are derived by taking the 2.5 and 97.5 percentiles of the estimates obtained from the replicate data sets.

### 3.5.3   Diagnostics

First, a GPD is fitted to exceedances of the modelling threshold $u$, with $u$ chosen using standard diagnostics (Coles, 2001). In particular, we use a parameter stability plot for $\xi$ (Figure 3.5.1) and check that estimates of the shape parameter stay consistent above the chosen threshold. Each approach is evaluated using the modelling threshold $u$, set at the 90th percentile such that 10% of days fall above the threshold (taking the value 29.7°C). Higher levels $v$ for which results are reported will be defined for each different analysis. The rate parameter $\lambda_u$ is estimated as 0.099 (0.007), where the standard error is given in the parentheses. The GPD scale parameter is estimated as $\hat{\sigma}_u = 3.002$ (0.225) and the shape parameter $\hat{\xi} = -0.215$

(0.033). A QQ-plot evaluated with the modelling threshold at the 90th percentile is provided in Figure 3.5.1 and indicates that the GPD is a reasonable fit at this threshold. Deviations from the diagonal are observed at higher thresholds but are contained within 95% tolerance intervals. Parameter stability plots at higher thresholds (not shown) do not indicate any statistically significant change in the parameter estimates.

Fitting the conditional extremes approach leads to an estimate for the dependence parameters of $\hat{\alpha} = 0.713$ (0.072) and $\hat{\beta} = 0.524$ (0.094). Parameter stability plots for the conditional extremes dependence parameters are given in Figure 3.5.3 and support that the choice of $u$ is valid. A likelihood ratio test confirms that these parameter values are significantly different from $\alpha = 1$ and $\beta = 0$ and that the data do not exhibit asymptotic dependence. Under the parametric model the logistic dependence parameter is estimated as $\hat{\gamma} = 0.578$ (0.026) with $\hat{\chi} = 0.508$ (0.027). As asymptotic dependence is the only form of dependence allowed in this model, $\hat{\chi} > 0$ despite the evidence from the conditional approach that suggests $\chi = 0$. When using peak value tail chain estimation the dependence parameters for the backward chain are also required and here $\hat{\alpha}_b = 0.816$ (0.061) and $\hat{\beta}_b = 0.512$ (0.096).

Figure 3.5.4 shows estimates of $\theta(v, m)$ and $\theta_C(v, m)$ under all approaches for return periods between 0.1 and 1 years, with $m$ set as 3 days. At these levels estimates given by the runs method are reasonably accurate and are used to assess which approach provides the best fit. The empirical estimate of $\theta(v, m)$ shows a broadly increasing pattern at lower return periods before levelling out and tailing off at higher levels, where less data are available for estimates. The estimate of $\theta(v, m)$ for the conditional extremes approach matches the behaviour of the empirical estimate the best. It is contained within the 95% confidence intervals of the empirical estimate at all levels. In contrast, both the asymptotically dependent parametric and non-parametric approaches give estimates of $\theta(v, m)$ and $\theta_C(v, m)$ that vary little over $v$. As these estimators only
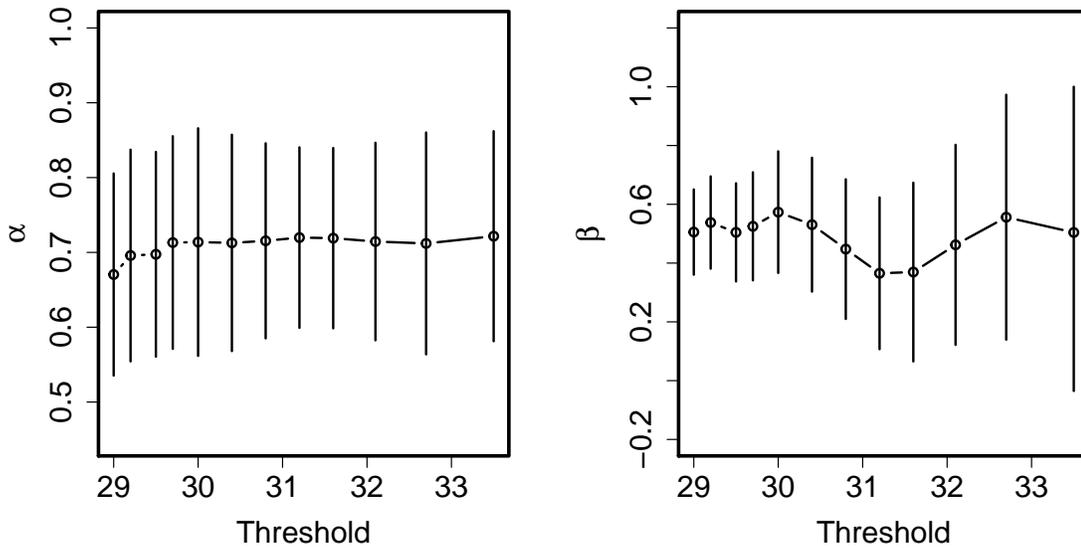
Figure 3.5.3: Parameter stability plots for conditional extremes dependence parameters $\alpha$ (left) and $\beta$ (right). Vertical bars show 95% confidence intervals obtained via bootstrapping.

give good estimates of the runs estimator at the threshold, it shows that they are highly sensitive to the threshold choice. Figure 3.5.4 illustrates that the conditional approach has a reduced sensitivity to the choice of modelling threshold $u$ compared to the asymptotically dependent parametric and non-parametric approaches as estimates of $\theta(v, m)$ can vary from $\theta(u, m)$ for $v > u$. The parametric approach is often contained within the confidence intervals, but at lower return periods it overestimates the size of the extremal index. It also fails to pick up the increase of the empirical estimate of the extremal index at lower return levels. The non-parametric approach cannot pick up this behaviour either and usually underestimates the extremal index. Similar patterns are observed for the consecutive extremal index $\theta_C(v, m)$ except that estimates are slightly higher. This pattern is similar for $N$ and $N_C$ and so whilst assessing fit we shall concentrate on estimating functions of $N$ with passing comments only made on $N_C$.

Estimates of the probability mass function $\pi(i, v)$ for all four approaches are given in
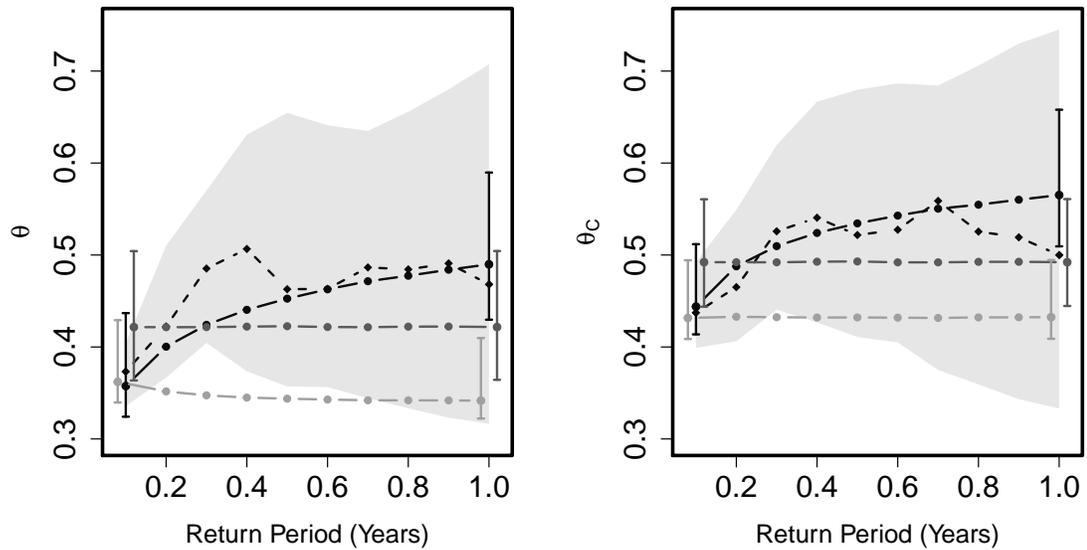
Figure 3.5.4: Estimated extremal index $\theta(v, m)$ (left) and consecutive extremal index $\theta_C(v, m)$ (right) obtained with the runs method (dashed, diamonds), parametric approach (grey dots), conditional extremes approach (black dots) and non-parametric approach (light grey dots). Confidence intervals are at the 95% level and are obtained by bootstrapping procedures for all approaches. The interval is given for runs method (light grey shading) at all return periods; for other approaches at 0.1 and 1 year return period (staggered) for visual clarity, where the return level $v$ corresponds to the return period given on the horizontal axis and $m$ is fixed at 3 days.

Figure 3.5.5 (left) for a range of $i$ for which the estimated distributions differ non-negligibly from zero. The critical level is set as $v = v_1$. All distributions are decreasing with $i$; there are only slight increases due to sampling noise for longer cluster lengths where results become sparse. The empirical estimate based upon the runs method shows the greatest amount of variability and all other approaches have narrower confidence intervals. The results obtained from the parametric and conditional extremes approaches tend to coincide but show some varying behaviour. At most values of $i$ the confidence intervals for the parametric and conditional extremes approaches are contained within those of the runs estimator. The result suggests that both methods are adequately modelling the data at this critical level. The non-parametric approach

gives estimates that seem to generally coincide with the empirical runs estimate and there is general agreement with the other approaches, though the estimate of $\pi(1, v_1)$ is lower.
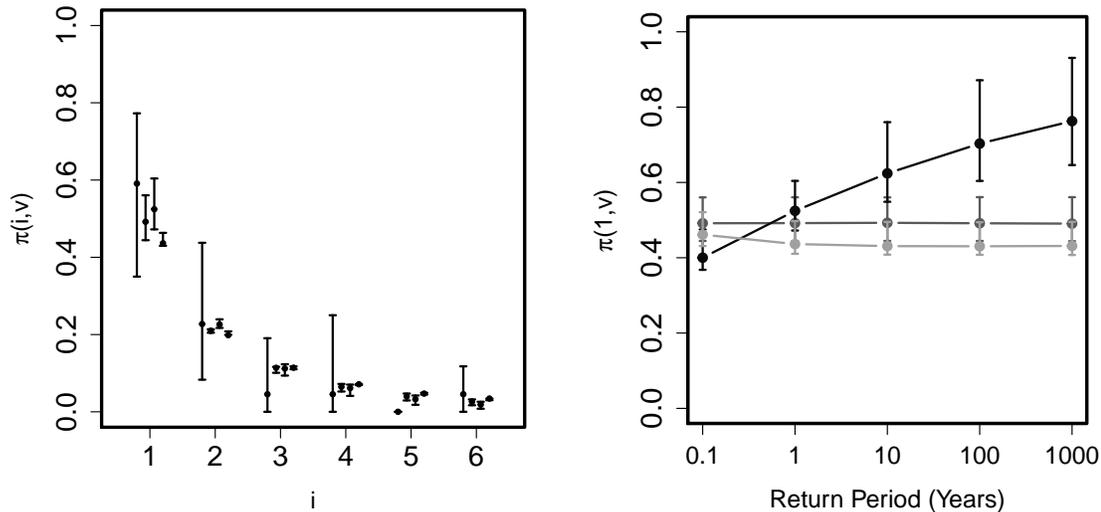


Figure 3.5.5: Left: Estimated distribution of $\pi(i, v_1)$ obtained with runs method (far left values), parametric approach (centre left), conditional extremes approach (centre right) and non-parametric approach (far right). Right: The effect of varying the return period (threshold) on $\pi(1, v)$ for parametric approach (grey, middle), conditional extremes approach (black, top) and non-parametric approach (light grey, bottom) plotted against $v$ on a log return period scale. In both plots confidence intervals are at the 95% level and are obtained by bootstrapping procedures for all approaches.

The effect of changing critical level $v$ on $\pi(1, v)$ is presented in Figure 3.5.5 (right). The runs estimator has been omitted from the plot since at these high levels the estimates obtained in this way become unreliable with wide confidence bands. Figure 3.5.5 (right) confirms that for the parametric and non-parametric approaches the value of $\pi(1, v)$ remains constant at all levels whereas for the conditional extremes approach $\pi(1, v)$ increases as the critical level is increased. The same pattern can be observed for $\pi_C(1, v)$ (not shown). This occurs since the parametric and non-parametric approaches are restricted to asymptotic dependence which does not allow for interaction
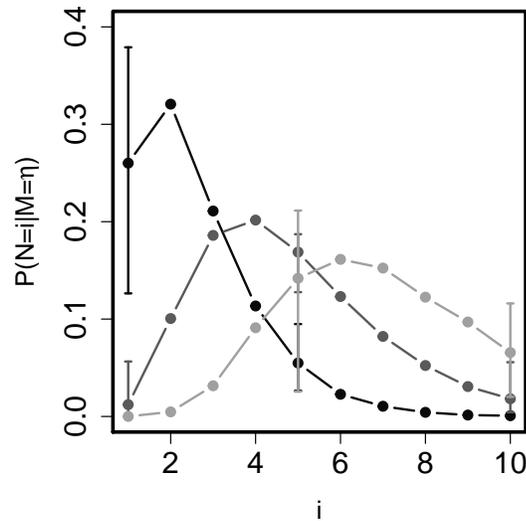
Figure 3.5.6: The effect of different cluster maximum sizes on distributions $P(N = i \mid M = \eta)$ at critical level $v = v_1$ when estimated using the conditional extremes approach. Cluster maxima $\eta$ at level that would occur on average once in 5 years (black), once in 50 years (grey) and once in 1000 years (light grey). Confidence intervals are given at the 95% level, constructed from 100 replicate data sets with 10000 forward/backward simulated chains and presented at only three durations for visual clarity.

between duration distribution and critical level. The conditional extremes method can allow for the asymptotically independent behaviour of the series and therefore can have interaction between duration distribution and critical level. The parametric and non-parametric approaches average the dependence over observed levels which leads to the constant behaviour. The confidence intervals of these solely asymptotically dependent methods tend to overlap.

It may also be of interest to know the duration distribution of a cluster given that the peak value was recorded at a specific level. The conditional extremes approach has highlighted that the behaviour of clusters changes with the critical level used to define them. Peak value chain simulation for the conditional extremes approach can be used as outlined in Section 3.4. Setting $v = v_1$ it is possible to analyse cluster

characteristics of events that have a larger peak value. The plot of $P(N = i \mid M = \eta)$ for different peak values $\eta$ is given in Figure 3.5.6. The shape of the distribution is positively skewed for lower peak values. This result is anticipated since a peak value nearer to the critical level will typically yield a cluster with fewer exceedances of the critical level than for clusters with a larger peak value.

### 3.5.4   Results

Empirical analysis in Section 3.5.2 identified events of length 2 and 11 above the critical level, corresponding to the one year return level. Tools from Section 3.4 allow us to estimate how likely each event was. In what follows we present estimated probabilities and in parentheses the associated bootstrapped 95% confidence intervals. The probability of observing at least one event in a year that lasts at least 2 days is 0.208 (0.200, 0.216) for the conditional extremes approach, 0.193 (0.183, 0.199) for the parametric approach and 0.175 (0.172, 0.188) for the non-parametric approach. Similarly, the probability of observing at least one event that lasts at least 11 days in a year is 0.001 ($1 \times 10^{-4}$, 0.004) for the conditional extremes approach, 0.005 (0.002, 0.009) for the parametric approach and 0.012 (0.007, 0.013) for the non-parametric approach. The asymptotically dependent parametric and non-parametric approaches give a much higher probability of observing a long event than the asymptotically independent conditional extremes approach. The same analysis can be completed for the probability of observing at least one event in a year that lasts longer than 11 consecutive days. For the conditional extremes approach the probability is $6 \times 10^{-4}$ ($4 \times 10^{-5}$, 0.002), for the parametric approach 0.004 (0.002, 0.006) and for the non-parametric approach this increases to 0.007 (0.004, 0.009).

In Section 3.1 we noted that periods of 3 days with an average daily maximum temperature above 35°C could lead to an excess mortality which varies over local cities between 17-47%. Using all approaches we can estimate the probability of observing

at least one such event in a year. For the conditional extremes approach the proba-
bility is given as 0.199 (0.181, 0.226), equivalent to an event that happens on average
once every five years. The same probability is given as 0.174 (0.161, 0.180) for the
parametric approach and 0.169 (0.157, 0.183) for the non-parametric approach. For
the remainder of the analysis we focus on the conditional extremes approach.

The maximum temperature in Orleans in 2003 was recorded at 39.9°C which corre-
sponds to a 1 in 50 year event. The peak value chain estimation method in Section 3.4
is used to assess the joint probability of an event with a hotter maximum temperature
and longer duration than the 2003 heatwave event. The probability of observing a
cluster with at least 11 exceedances conditional on a peak value greater than the 2003
temperature is 0.06 (0.008, 0.23). The joint probability for the cluster functionals can
be obtained by multiplying the conditional probability by the probability of observing
a peak value greater than the 2003 temperature and is estimated as 0.001 ($7 \times 10^{-5}$,
0.013). Application of equation (3.4.1) allows the derivation of over cluster results
from the within cluster results given above. As such the probability of observing at
least one event in a year that both lasts longer than 11 days and has a peak value
greater than 39.9°C is $6 \times 10^{-4}$ ($4 \times 10^{-5}$, $6 \times 10^{-3}$), approximately equivalent to the
1650 year return period. The equivalent probability for 11 consecutive exceedances is
$4 \times 10^{-4}$ ($3 \times 10^{-5}$, $5 \times 10^{-3}$) for the conditional extremes approach.

## 3.6 Discussion and conclusion

The results given in Section 3.5 show that the interaction between the duration distri-
bution of heatwave events and a critical level is only modelled realistically by methods
that account for asymptotic independence. At high critical levels this leads to a reduc-
tion in the probability of observing longer events when using the conditional approach
over other approaches that can account only for asymptotic dependence. Model se-

lection diagnostics indicate that these lower estimates of the probability of observing longer events at high critical levels reflect the characteristics of the data better. If a user is especially averse to the risk of longer heatwave events then they could be willing to mitigate for such an event using one of the asymptotically dependent approaches. However, our analysis shows that this can considerably inflate the estimated risk. If such a conservative approach is to be taken we have found that two different asymptotically dependent modelling approaches, parametric and non-parametric, give very similar risk estimates. It is noted that there are some asymptotically dependent models that can account for further weakening of dependence above the selected critical level of interest; see Wadsworth and Tawn (2012b) and Wadsworth et al. (2016) for more details.

The assumption that the temperature time-series follows a first order Markov process has been made to permit the modelling process outlined in the paper. Such an assumption was supported by an exploratory data analysis but might be an unrealistic assumption in other applications or for such extreme events as in 2003. Specifically, our approach gives the return period of an event rarer than in 2003 as 1650 years; either this really was an exceptional event or there are subtleties in higher order dependence for the extreme temperature process that are not captured by our Markov model. Making an assumption of higher order Markov processes has not been considered in this chapter, but is investigated further in Chapter 4. Alternatively no Markov structure assumptions could be made, e.g. as in Eastoe and Tawn (2012), but this comes at the cost of large numbers of parameters and a high dimensional non-parametric distribution $G$ to estimate which is likely to lead to very poor estimates of events more extreme than the event observed in 2003.

Our approach has focused on daily maximum temperatures. As outlined in Section 3.1, Pascal et al. (2013) point out that extremely hot night time temperatures

during a heatwave can also be important in raising mortality. Thus we may be interested in extremes of the series $X_t, Y_t, X_{t+1}, Y_{t+1}, \ldots$ where $X_t$ is the daily maximum temperature and $Y_t$ is the daily minimum temperature on day $t$. A Markov model is likely to be appropriate for the series. Although the series is non-stationary its components $X_t$ and $Y_t$ may be individually stationary with marginal distributions $F_X$ and $F_Y$. Applying the marginal methods of Section 3.3 we can transform the $X_t$ and $Y_t$ series to have an identical marginal Laplace distributions. We can then model the dependence structure for the transitions between the series using the conditional extremes approach. These transitions may have parameters that vary between the pairs $(X_t, Y_t)$ and $(Y_t, X_{t+1})$ but otherwise the methodology developed in the paper can be extended easily to this more general situation. Our approach has also focused on heatwaves at a single site whereas the spatial nature of an event is also critical for the economy and health. Davison and Gholamrezaee (2012) look at the heatwave problem from a spatial perspective, focusing on asymptotically dependent models only and ignoring temporal aspects. Therefore a future open line of research is to draw together our approach with theirs, requiring a fully space-time model for extremes. The first approaches to space-time extremes models are Huser and Davison (2014) and Davis et al. (2013), but these are restricted to asymptotic dependence in both space and time.

The underlying effect of climate change has been ignored during this paper and is an important future extension for each approach. Stott et al. (2004) have investigated the human contribution to the European heatwave of 2003. They suggest that it is very likely that anthropogenic climate change has at least doubled the risk of a heatwave as intense as the event in 2003 in comparison to pre-industrial times. In Chapter 5 we apply the approaches in this chapter to assess the affect of human induced climate change which can affect both marginal and dependence characteristics of the process.

# Chapter 4

# $k$th-order Markov extremal models for assessing heatwave risks

## 4.1 Introduction

Many devastating natural hazards are caused by events that are extreme and rare. Extreme value theory provides a general framework for modelling extreme values. In many situations a singular extreme observation does not have a great effect, whereas combinations and runs of extreme values can cause widespread devastation. For example when estimating risks attributed to heatwaves, one hot day may not cause a large increase in excess mortality whereas a run of consecutive hot days is far more damaging. Therefore any extreme value model utilised must be able to reliably capture such behaviour. In the terminology of extreme value theory this requires a model that can explicitly capture the extremal temporal dependence structure alongside marginal tail characteristics.

Methods for modelling multivariate extreme events are often separated between the two aforementioned components, the margins and dependence structure. Let $\{Y_t\}$ be a stationary time-series, during this paper taken to be a series of maximum daily

temperature values at a single site. The most common approach to model the extreme values of the margins is to fit a generalized Pareto distribution (GPD) to exceedances of a high modelling threshold $u_Y$, i.e.

$$P(Y_t - u_Y > y \mid Y_t > u_Y) = \left(1 + \frac{\xi y}{\sigma_{u_Y}}\right)_+^{-1/\xi} \qquad \text{for} \quad y \geq 0,$$

where $c_+ = \max(c, 0)$, $\sigma_{u_Y} > 0$ and $\xi$ are the scale and shape parameters of the GPD respectively (Coles, 2001), with the scale parameter being threshold dependent.

A heatwave is defined as a set of consecutive days and/or nights that lead to an increase in mortality. As such, an important quantity to model is the number of exceedances of a critical level during a block of time. It is also necessary to be able to estimate other important extremal quantities, here named cluster functionals. Empirical methods exist to split a time-series of temperature data into independent clusters of exceedances of the threshold $u_Y$ where within each cluster groups of dependent exceedances occur; the most popular technique is the runs method (Smith and Weissman, 1994). As such from a time-series we can obtain the number of independent clusters and the values in each cluster. The number of clusters are Poisson distributed (Davison and Smith, 1990); here we wish to accurately model the values within a cluster, i.e. the local time-series during an extreme event.

Different approaches exist for modelling the dependence structure, which broadly split into methods that attempt to model the joint tail of multivariate extremes and methods that condition upon one extreme variable. The former group include Ledford and Tawn (1997) and Smith et al. (1997) and are the classical models for modelling joint extremes. However these models can be hard to implement in a high-dimensional setting and are often restricted to the case of asymptotic dependence, i.e. for two random variables $(Y_t, Y_{t+\tau})$ the extremal dependence measure

$$\chi_\tau = \lim_{y \to y^*} P(Y_{t+\tau} > y \mid Y_t > y),$$

is strictly greater than zero, where $y^*$ is the upper endpoint of the common distribution function and $\tau$ is a time lag. The conditional extremes approach of Heffernan and Tawn (2004) provides a more flexible approach that permits both asymptotic dependence ($\chi_\tau > 0$) and asymptotic independence ($\chi_\tau = 0$). The approach can also more easily be generalised to higher-dimensional problems.

A range of temporal dependence structure models have been proposed, some specific to heatwave applications. Smith et al. (1997) provide a framework for modelling threshold exceedances using first-order Markov chain approaches, but are restricted to the situation of asymptotic dependence. Yun (2000) outline an approach to analyse the distribution of cluster functionals of extreme events in an asymptotically dependent $k$th-order Markov chain. More recently, Reich et al. (2014) formulate an asymptotically dependent max-stable process using random effects within a Bayesian framework, incorporating dependence within 10 day windows. Bortot and Tawn (1998) use theory from Ledford and Tawn (1997) to derive a class of models for first-order Markov chains that permits asymptotic independence. An asymptotically independent Gaussian copula model is proposed in Dupuis (2012), who considers a pre-processing approach for the margins and an AR model with lags up to lag-8 for the temporal dependence structure.

In Chapter 3 we built a first-order Markov approach based upon the conditional extremes approach of Heffernan and Tawn (2004) that can account for both asymptotic dependence and asymptotic independence. For the daily maximum temperature data analysed in Chapter 3, it was found that standard diagnostics, e.g. PACF and comparison of observed and modelled cluster functionals, suggest that the first-order Markov assumption was reasonable. However, the physical mechanisms of heatwaves suggest that this is perhaps an oversimplification that could lead to underestimation of the risk of a heatwave event. This paper seeks to take advantage of the higher-order

structure to give a $k$th-order Markov model and to provide more accurate estimates of the risk of a heatwave event.

We also seek to develop diagnostic tests to choose an appropriate order Markov process to fit to extreme events. Standard time-series diagnostics for choosing an appropriate Markov process are often misleading when considering the behaviour of extremes. In some situations the extremal structure is less complicated than in the body and standard diagnostics suggest using a more complicated model than is required. When there is more complicated structure in the extremes than in the body, models that are too simple are selected by standard techniques. Ledford and Tawn (2003) developed diagnostic tools to test long and short range dependence assumptions within extreme events of both asymptotically dependent and asymptotically independent processes. However, these methods were unable to detect the order of the process. Here, we seek to extend the tools of Ledford and Tawn (2003) under the conditional extremes framework to provide greater insight into our modelling of heatwaves, without being restricted to the assumption of a first-order Markov process. Our work naturally extends Papastathopoulos and Tawn (2013) to give the first formal methods for testing for conditional independence in extreme values when the variables can be either asymptotically dependent or asymptotically independent.

Section 4.2 sets out theory for modelling the extremes of Markov chains by modelling the margins and dependence structure separately using a threshold exceedance approach for the former and the conditional extremes approach for the latter. Inference for the dependence structure of the conditional extremes model and extensions to account for higher-order structure are presented in Section 4.3 along with a discussion of diagnostic methods for order choice. Simulation of cluster features is discussed in Section 4.4 along with an extension to the algorithm for generating replicate tail chains set out in Appendix B. Section 4.5 gives results for a temperature data set over

central France and compares to the results in Chapter 3. Discussion and conclusions are presented in Section 4.6.

## 4.2   Modelling temporal dependence

### 4.2.1   Markov modelling

Throughout this paper we develop an approach that uses $k$th-order Markov chains to model the behaviour of the temperature series during heatwave events. Under the assumption that a time-series $\{Y_t\}$ follows a $k$th-order Markov process, the joint density function $f_n$ of $(Y_1, \ldots, Y_n)$ can be written as

$$f_n(y_1, \ldots, y_n) = f_k(y_1, \ldots, y_k) \prod_{t=1}^{n-k} f(y_{t+k} \mid y_{t+k-1}, \ldots, y_t),$$

where $f(\cdot \mid \cdot)$ is the conditional density function associated with $f_k$. This assumption permits us to model the extremes of the whole joint distribution by analysing the extremes of $(Y_t, \ldots, Y_{t+k})$ for $t = 1, \ldots, n-k$ and studying the conditional distribution of $Y_{t+k} \mid (Y_t, \ldots, Y_{t+k-1})$. The model for the marginal exceedances of the threshold $u_Y$ is given in Section 4.2.2. Our main method for modelling extremal dependence is presented in Section 4.2.3 and is based upon the conditional approach outlined in Heffernan and Tawn (2004). In Section 4.2.4 we propose an extension to the conditional extremes approach for time-series.

### 4.2.2   Marginal modelling

Here, we take $\{Y_t\}$ to be a stationary series and as such the marginal distributions $F$ are identical. The marginal excesses of $u_Y$ are assumed to follow a GPD which leads to the model for the common marginal distribution

$$F(y) = \begin{cases} 1 - \lambda_{u_Y} \left(1 + \xi \frac{y - u_Y}{\sigma_{u_Y}}\right)_+^{-1/\xi}, & y \geq u_Y \\ \tilde{F}(y), & y < u_Y, \end{cases}$$

where $\lambda_{u_Y} = 1 - \tilde{F}(u_Y)$ and $\tilde{F}(y)$ is the empirical cumulative distribution function of $\{Y_t\}_{t=1}^n$. Standard censored likelihood approaches are used to estimate the marginal parameters. Having estimated the marginal structure, a choice of transformation onto common margins is required prior to modelling extremal dependence. Different marginal choices exist for the conditional extremes approach. Heffernan and Tawn (2004) model dependence for Gumbel margins, but Keef et al. (2013) showed that a more comprehensive approach arises for Laplace margins. Following Keef et al. (2013) we transform $Y_t$, $t = 1, \ldots, n$ into Laplace margins as follows

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

and subsequently model dependence on $\{X_t\}$ which can be easily translated back to $\{Y_t\}$.

### 4.2.3 Conditional extremes approach for multivariate context

Heffernan and Tawn (2004) motivate an approach to modelling the extremes of a vector $\mathbf{X}_{t:t+\tau} = (X_t, \ldots, X_{t+\tau}) = (X_t, \mathbf{X}_{-t})$ for $t = 1, \ldots, n - \tau$ and for fixed $\tau$, with $X_t$ large and where $\mathbf{X}_{-t}$ are all components of the vector $\mathbf{X}_{t:t+\tau}$ without $X_t$, i.e. $(X_{t+1}, \ldots, X_{t+\tau})$. This approach is called the conditional extremes method. The desire is to model the joint behaviour of $\mathbf{X}_{t:t+\tau}$ using the distribution of $\mathbf{X}_{-t}$ given that $X_t$ exceeds some high threshold $u$. A requirement for modelling the conditional distribution $\mathrm{P}\{\mathbf{X}_{-t} \leq \mathbf{x}_{-t} \mid X_t = x\}$ is that this distribution should be non-degenerate as $x \to \infty$, hence $\mathbf{x}_{-t}$ needs to be a function of $x$. Below, all vector calculations are to be interpreted componentwise. Following a characterisation in Heffernan and Resnick (2007) and results in Heffernan and Tawn (2004), under weak assumptions on $\mathbf{X}_{t:t+\tau}$ and the specification of Laplace margins, there exist dependence parameters

$\boldsymbol{\alpha}_{1:\tau} = (\alpha_1, \ldots, \alpha_\tau) \in [-1, 1]^\tau$ and $\boldsymbol{\beta}_{1:\tau} = (\beta_1, \ldots, \beta_\tau) \in (-\infty, 1)^\tau$ such that for $x > 0$

$$\mathrm{P}\left(\frac{\mathbf{X}_{-t} - \boldsymbol{\alpha}_{1:\tau}X_t}{X_t^{\boldsymbol{\beta}_{1:\tau}}} \leq \mathbf{z}, X_t - u > x \;\middle|\; X_t > u\right) \to G_{1:\tau}(\mathbf{z})\exp(-x), \qquad (4.2.1)$$

as $u \to \infty$, where $G_{1:\tau}$ is non-degenerate in each margin, i.e. for $j = 1, \ldots, \tau$ the $j$th margin $G_j$ of $G_{1:\tau}$ is non-degenerate, with $\mathbf{z} = (z_1, \ldots, z_\tau) \in \mathbb{R}^\tau$. Different types of dependence lead to different values of the aforementioned dependence parameters. If the variables $(X_t, X_{t+j})$ are independent then $\alpha_j = \beta_j = 0$ and $G_j$ is the Laplace distribution function, for $j \leq \tau$ whereas $\alpha_j = 1$ and $\beta_j = 0$ corresponds to the situation of asymptotic dependence, $-1 \leq \alpha_j \leq 0$ to negative dependence and $0 < \alpha_j < 1$ or $\alpha_j = 0$ and $\beta_j > 0$ corresponds to asymptotic independence with positive dependence. For more information see Keef et al. (2013).

## 4.2.4 Conditional extremes approach for time-series

We propose an extension to the statistical implementation of this approach that takes into account the value of intermediary values of the time-series, where interest is in the conditional distribution of the process given the past, taken here as $X_{t+\tau} \mid \mathbf{X}_{t:t+\tau-1}$ for large $X_t$. As such we are interested in an analogous expression to equation (4.2.1) given as

$$\lim_{\delta z_1 \to 0, \ldots, \delta z_\tau \to 0} \lim_{u \to \infty} \mathrm{P}\left(\frac{X_{t+\tau} - \alpha_\tau X_t}{X_t^{\beta_\tau}} \leq z, X_t - u > x \;\middle|\; \frac{\mathbf{X}_{t+1:t+\tau-1} - \boldsymbol{\alpha}_{1:\tau-1}X_t}{X_t^{\boldsymbol{\beta}_{1:\tau-1}}} \in \delta\mathbf{z}_{1:\tau-1}, X_t > u\right)$$

$$= G_{\tau \mid 1:\tau-1}(z \mid \mathbf{z}_{1:\tau-1})\exp(-x),$$

$$(4.2.2)$$

for $x > 0$, where $\mathbf{X}_{t+1:t+\tau-1} = (X_{t+1}, \ldots, X_{t+\tau-1})$ and $\delta\mathbf{z}_{1:\tau-1} = (z_1, z_1 + \delta z_1) \times \cdots \times (z_{\tau-1}, z_{\tau-1} + \delta z_{\tau-1})$. By undoing the conditioning, the limit on the left hand side of equation (4.2.2) can be rewritten as

$$\lim_{\delta z_1 \to 0, \ldots, \delta z_\tau \to 0} \lim_{u \to \infty} \frac{\mathrm{P}\left(\frac{X_{t+\tau} - \alpha_\tau X_t}{X_t^{\beta_\tau}} \leq z, \frac{\mathbf{X}_{t+1:t+\tau-1} - \boldsymbol{\alpha}_{1:\tau-1}X_t}{X_t^{\boldsymbol{\beta}_{1:\tau-1}}} \in \delta\mathbf{z}_{1:\tau-1}, X_t - u > x \;\middle|\; X_t > u\right)}{\mathrm{P}\left(\frac{\mathbf{X}_{t+1:t+\tau-1} - \boldsymbol{\alpha}_{1:\tau-1}X_t}{X_t^{\boldsymbol{\beta}_{1:\tau-1}}} \in \delta\mathbf{z}_{1:\tau-1} \;\middle|\; X_t > u\right)}.$$

$$(4.2.3)$$

Under the assumption that the non-limit form on the left hand side of equation (4.2.1) is $\tau$-times differentiable in $\mathbf{z}$, we have that equation (4.2.3) can now be represented as

$$\lim_{\delta z_1 \to 0, \ldots, \delta z_\tau \to 0} \lim_{u \to \infty} \frac{\int_{-\infty}^{z} \int_{z_1}^{z_1 + \delta z_1} \cdots \int_{z_{\tau-1}}^{z_{\tau-1} + \delta z_{\tau-1}} \bar{g}_{1:\tau}(\mathbf{s}_{1:\tau-1}, s; u) \exp(-x) ds_1 \ldots ds_{\tau-1} ds}{\int_{z_1}^{z_1 + \delta z_1} \cdots \int_{z_{\tau-1}}^{z_{\tau-1} + \delta z_{\tau-1}} \bar{g}_{1:\tau-1}(\mathbf{s}_{1:\tau-1}; u) ds_1 \ldots ds_{\tau-1}},$$

(4.2.4)

where $\bar{g}_{1:\tau}(\cdot; u)$ is the joint density function of $\left(\mathbf{X}_{-t} - \boldsymbol{\alpha}_{1:\tau} X_t\right) / X_t^{\boldsymbol{\beta}_{1:\tau}} \mid X_t > u$. Under the assumption of exchangeability of limits and integrals, equation (4.2.4) is given as

$$\lim_{\delta z_1 \to 0, \ldots, \delta z_\tau \to 0} \frac{\int_{-\infty}^{z} \int_{z_1}^{z_1 + \delta z_1} \cdots \int_{z_{\tau-1}}^{z_{\tau-1} + \delta z_{\tau-1}} g_{1:\tau}(\mathbf{s}_{1:\tau-1}, s) \exp(-x) ds_1 \ldots ds_{\tau-1} ds}{\int_{z_1}^{z_1 + \delta z_1} \cdots \int_{z_{\tau-1}}^{z_{\tau-1} + \delta z_{\tau-1}} g_{1:\tau-1}(\mathbf{s}_{1:\tau-1}) ds_1 \ldots ds_{\tau-1}}, \quad (4.2.5)$$

where $g_{1:\tau}$ is the joint density function of $G_{1:\tau}$ in equation (4.2.1). By repeated application of L'Hôpital's rule equation (4.2.5) can be rewritten as

$$\exp(-x) \frac{\int_{-\infty}^{z} g_{1:\tau}(\mathbf{z}_{1:\tau-1}, s) ds}{g_{1:\tau-1}(\mathbf{z}_{1:\tau-1})} = \exp(-x) \int_{-\infty}^{z} g_{\tau|1:\tau-1}(s \mid \mathbf{z}_{1:\tau-1}) ds, \quad (4.2.6)$$

and therefore from equation (4.2.2) we have that

$$G_{\tau|1:\tau-1}(z \mid \mathbf{z}_{1:\tau-1}) = \int_{-\infty}^{z} g_{\tau|1:\tau-1}(s \mid \mathbf{z}_{1:\tau-1}) ds,$$

where $g_{\tau|1:\tau-1}(s \mid \mathbf{z}_{1:\tau-1})$ is the conditional density function from $G_{1:\tau}$. We also note that if $\tau > k$, where $k$ is the order of the Markov process used, any terms of $\mathbf{z}_{1:\tau-1}$ separated by greater than lag $k$ will be redundant and thus can be ignored. Our approach to estimating the distribution function $G_{\tau|1:\tau-1}$ is to obtain an estimate of the joint density function $g_{1:\tau}$ and use this to derive the conditional distribution. For this purpose we propose an approach based upon kernel density estimation which is given in Section 4.3.1. It is noted at this stage that such an approach, if used in very high dimensions, can be affected by the curse of dimensionality. We do not directly deal with this problem here, since our estimates of important extremal quantities in Section 4.5 seem sensible compared to the first-order estimates given in Chapter 3.

## 4.3 Inference for dependence structure

### 4.3.1 Estimation of extremal quantities

Modelling using the conditional extremes approach requires the assumption that the limiting form of equation (4.2.1) holds exactly for all values of $X_t > u$ given that $u$ is a sufficiently high threshold. We write the form of $\mathbf{X}_{t+1:t+\tau}$ given that $X_t > u$ as

$$\mathbf{X}_{t+1:t+\tau} = \boldsymbol{\alpha}_{1:\tau} X_t + X_t^{\boldsymbol{\beta}_{1:\tau}} \mathbf{Z}_{1:\tau}, \qquad (4.3.1)$$

where $\mathbf{Z}_{1:\tau} = (Z_1, \ldots, Z_\tau)$ is a dependent random variable, independent of $X_t$, with distribution function $G_{1:\tau}$ as in equation (4.2.1). Dependence parameters $\alpha_j$ and $\beta_j$, for $j = 1, \ldots, \tau$, are estimated using pairwise data on $(X_t, X_{t+j})$ via standard likelihood approaches. Constraints on the parameters caused by stationarity assumption are given in Heffernan and Tawn (2004). Since $G_j$ does not take any simple parametric form, in order to fit $\alpha_j$ and $\beta_j$, for $j = 1, \ldots, \tau$, we make a temporary working assumption that $Z_j \sim \mathrm{N}(\mu_j, \sigma_j)$ (Keef et al., 2013) and as such

$$X_{t+j} \mid \{X_t = x\} \sim N\left(\alpha_j x + \mu_j x^{\beta_j}, \sigma_j^2 x^{2\beta_j}\right) \qquad \text{for} \quad x > u,$$

where $j = 1, \ldots, \tau$. In this way we estimate the set of parameters $(\alpha_j, \beta_j, \mu_j, \gamma_j)$ by standard likelihood approaches. At this stage the Gaussian assumption is discarded and a non-parametric estimate of the distribution $G_{1:\tau}$ is formed by inverting equation (4.3.1). Specifically, let $t_1, \ldots, t_{n_u}$ be the indices of $t = 1, \ldots, n$ where $x_t > u$ then let

$$\hat{z}_j^{(i)} = \frac{(x_{t_i+j} - \hat{\alpha}_j x_{t_i} - \hat{\mu}_j x_{t_i}^{\hat{\beta}_j})}{\hat{\sigma}_j x_{t_i}^{\hat{\beta}_j}}, \qquad (4.3.2)$$

for $i = 1, \ldots, n_u$ and $j = 1, \ldots, \tau$, where $n_u$ is the number data points exceeding the threshold $u$. The sample in equation (4.3.2) have been normalised using $(\mu_j, \sigma_j)$ to ensure all values have mean 0 and variance 1 through time. Here, similar to Papastathopoulos and Tawn (2013), we estimate the density $g_{1:\tau}$ of $G_{1:\tau}$ using a

multivariate kernel density based on data $\hat{\mathbf{z}}^{(i)} = \left( \hat{z}_1^{(i)}, \ldots, \hat{z}_\tau^{(i)} \right)$, $i = 1, \ldots, n_u$, to obtain the final estimate $\tilde{g}_{1:\tau}$, i.e.

$$\tilde{g}_{1:\tau}(\mathbf{z}) = \frac{1}{n_u} \sum_{i=1}^{n_u} K_{\mathbf{H}} \left( \mathbf{z} - \hat{\mathbf{z}}^{(i)} \right),$$

where $K_{\mathbf{H}}$ is the multivariate kernel function, $\mathbf{H}$ is a symmetric and positive definite bandwidth matrix such that $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{1/2}\mathbf{x})$. A common choice of the kernel function is the standard independent Normal multivariate kernel such that

$$K(\mathbf{x}) = (2\pi)^{-\tau/2} \exp \left\{ -\frac{\mathbf{x}^\mathsf{T} \mathbf{x}}{2} \right\},$$

and $\mathbf{H}$ is taken to be diagonal. As such the joint density $\tilde{g}_{1:\tau}$ has the form

$$\tilde{g}_{1:\tau}(\mathbf{z}) = \tilde{g}(z_1, \ldots, z_\tau) = \frac{1}{n_u} \sum_{i=1}^{n_u} \prod_{j=1}^{\tau} \frac{1}{h_j} \phi \left( \frac{z_j - \hat{z}_j^{(i)}}{h_j} \right),$$

where $\hat{z}_j^{(i)}$, $j = 1, \ldots, \tau$ and $i = 1, \ldots, n_u$, are the values given by equation (4.3.2), $\phi(\cdot)$ is the standard Normal density function and $h_j$ for $j = 1, \ldots, \tau$ are associated bandwidths. Under the assumption that the limit form in equation (4.2.2) holds for a sufficiently high value of threshold $u$, using equation (4.2.6), the distribution function $G_{\tau|1:\tau-1}$ is

$$\hat{G}_{\tau|1:\tau-1}(z \mid \mathbf{z}_{1:\tau-1}) = \sum_{i=1}^{n_u} w_i \Phi \left( \frac{z - \hat{z}_\tau^{(i)}}{h_\tau} \right),$$

where weights $w_i$, $0 \leq w_i \leq 1$ and $\sum_{i=1}^{n_u} w_i = 1$, are given as

$$w_i = \prod_{j=1}^{\tau-1} \phi \left( \frac{z_j - \hat{z}_j^{(i)}}{h_j} \right) \Big/ \sum_{r=1}^{n_u} \prod_{j=1}^{\tau-1} \phi \left( \frac{z_j - \hat{z}_j^{(r)}}{h_j} \right) \qquad i = 1, \ldots, n_u \qquad (4.3.3)$$

In this way the weights are given as the density of standardised differences of $\mathbf{z}_{\tau-1}$ from the observed values $\mathbf{z}_{\tau-1}^{(i)}$, for each observation $i$ with $X_{t_i} > u$.

## 4.3.2 Selection of the order of Markov process

We have motivated an approach that takes into account higher-order Markov structure when simulating clusters of extreme events. Ideally we wish to test whether the

incorporation of higher-order structure provides any significant improvement over a lower-order Markov model. As we incorporate higher-order information beyond the true order of the Markov process in the extremes additional parameters and distributional assumptions are introduced and as such we would like not to select too large an order to avoid inefficiency and potentially decreasing the precision of estimates. If the order selected is too small, we may not adequately capture the extremal dependence structure which leads to inaccurate inferences.

Based upon simulations from our fitted model, we outline two diagnostics that can be used to pick a suitable order. Both diagnostics are motivated by standard univariate threshold selection diagnostics (Coles, 2001); essentially these diagnostics are equivalent to threshold stability plots. Neither diagnostic provides a specific value for the order $k$ but will suggest a set of sensible values. Such an approach also allows the order to be tailored to the type of extremal quantity that we want to estimate. The first diagnostic tests whether there is any significant difference between the threshold dependent extremal dependence measure $\chi_j(v)$, i.e.

$$\chi_j(v) = \mathrm{P}(X_{t+j} > v \mid X_t > v) \qquad \text{for} \quad j \geq k, \qquad (4.3.4)$$

estimated empirically and with the $k$th-order conditional extremes model, denoted by $\tilde{\chi}_j(v)$ and $\hat{\chi}_j^{(k)}(v)$ respectively. It explores how the differences in these estimates changes with $k$. The $k$th-order Markov model that provides the best fit for a range of $j \geq k$ is the one that lies closest to the empirical estimate for a range of critical levels.

A different diagnostic for the order of Markov chain aims to identify an order above which estimates of important extremal quantities are stable other than sampling variability. Here, important extremal quantities are defined as the extremal index, see Section 4.4, and short, medium and long runs of exceedances of a high level, see Section 4.5 for a more precise definition. For threshold selection we wish to pick the

lowest threshold for which parameter estimates are stable at higher levels. Here, if the estimates of extremal quantities do not change above a certain order, it is not necessary to incorporate higher-order structure above the defined order.

We also propose an approach that compares whether a $k$th-order model provides a significant advantage over a first-order model using a hypothesis test. Reich et al. (2014) perform such a test to check differences between their higher-order model and a first-order Markov model; to take a similar approach in our setting we exploit limit results for first-order Markov chains in Papastathopoulos et al. (2015). The test of a first-order model against a second-order model is outlined below. Under the assumption of a first-order Markov process we have that

$$X_{t+1} = \alpha_1 X_t + X_t^{\beta_1} Z_{1|0},$$

for $X_t > u$, where $\{Z_{i+1|i}\}$ are independent and identically distributed over $i$ with the same distribution as $Z_1$ in equation (4.3.1), and as such when $X_{t+1} > u$ and $\alpha_1 > 0$ we have that

$$
\begin{aligned}
X_{t+2} &= \alpha_1 X_{t+1} + X_{t+1}^{\beta_1} Z_{2|1} \\
&= \alpha_1 \left( \alpha_1 X_t + X_t^{\beta_1} Z_{1|0} \right) + \left( \alpha_1 X_t + X_t^{\beta_1} Z_{1|0} \right)^{\beta_1} Z_{2|1} \\
&= \alpha_1^2 X_t + \alpha_1 X_t^{\beta_1} Z_{1|0} + \left( \alpha_1 X_t + X_t^{\beta_1} Z_{1|0} \right)^{\beta_1} Z_{2|1} \\
&= \alpha_1^2 X_t + \alpha_1 X_t^{\beta_1} Z_{1|0} + \alpha_1^{\beta_1} X_t^{\beta_1} Z_{2|1} + \mathcal{O}_p \left( X_t^{2\beta_1 - 1} \right).
\end{aligned}
$$

As such, under the first-order Markov assumption, for $x > 0$

$$\mathrm{P}\left( \frac{X_{t+2} - \alpha_1^2 X_t}{X_t^{\beta_1}} \le z, X_t > u + x \mid X_t > u \right) \to G_2(z) \exp(-x), \qquad (4.3.5)$$

as $u \to \infty$ where $G_2$ is the distribution of $Z_{2|0} = \alpha_1 Z_{1|0} + \alpha_1^{\beta_1} Z_{2|1}$. The result is extended for lag $\tau$ in Papastathopoulos and Tawn (2013) and Papastathopoulos et al. (2015) such that a similar test can be constructed by fixing $\alpha_\tau = \alpha_1^\tau$ and $\beta_\tau = \beta_1$. In the situation where $\alpha_\tau = 1$ and $\beta_\tau = 0$, i.e. asymptotic dependence, this expression is

equivalent to the random walk given in Smith (1992). When $\alpha_1 = 0$, we have that

$$\text{P}\left(\frac{X_{t+\tau}}{X_t^{\beta_1^\tau}} \leq z, X_t > u + x \mid X_t > u\right) \to G_\tau(z)\exp(-x), \qquad (4.3.6)$$

as $u \to \infty$, where $G_\tau$ is the distribution of $Z_{\tau|0} = \prod_{i=1}^\tau Z_{i|i-1}^{\beta_1^{\tau-i}}$. Under the assumption that the limit form in equations (4.3.5) and (4.3.6) hold exactly at some high threshold, a comparison of the suitability of the first- and $k$th-order approaches is constructed via a likelihood ratio test by testing a model with fixed parameters $\alpha_k = \alpha_1^k$ and $\beta_k = \beta_1$ against a model where $\alpha_k$ and $\beta_k$ are allowed to vary. For completeness the model where $\alpha_k = 0$ and $\beta_k = \beta_1^k$ must also be tested. If the $k$th-order approach is found to obtain a significantly better fit than the first-order approach, a natural next step is to ask whether the $k$th-order result is a better fit than the $j$th-order result for all $j = 2, \ldots, k - 1$. Such a set of nested tests exists when modelling time-series using AR models (Brockwell and Davis, 2006), but has not been previously investigated within the framework here. It is difficult to derive theoretical results that are as interpretable as testing against the first-order approach and are only considered briefly in Section 4.6.

## 4.4  Cluster simulation

We simulate chains with the desired extremal properties to derive the required properties of the chain using Monte Carlo methods. Common approaches for defining clusters have already been mentioned in Section 4.1; here we focus on within cluster behaviour. Two main strategies exist for the generation of the Markov chain in its tail, known as the tail chain. We can either simulate forwards and backwards from a cluster maximum $M > v$ (Smith et al., 1997) or simulate an initial exceedance of $v$ and simulate forwards only (Rootzén, 1988). The choice of approach is informed by the properties of the tail chain that we wish to estimate. In Section 4.3.2 the sub-asymptotic extremal dependence measure $\chi_j(v)$ was introduced in equation (4.3.4)

as a useful summary of the level of dependence in the tail. However, we are also interested in estimating other quantities, often an important quantity is the number of exceedances of a critical level, i.e.

$$D_v = \sum_{t \in C} \mathbb{I}\,(X_t - v)_+ \,,$$

where $\mathbb{I}(.)$ is the indicator function and $C$ is a set of values constituting a cluster. In Chapter 3 we introduced an important quantity denoted the duration distribution

$$\pi(i, v) = \mathrm{P}(D_v = i \mid M > v), \qquad (4.4.1)$$

where $M$ is a random variable denoting the cluster maximum. A common measure that follows from equation (4.4.1) is the subasymptotic extremal index $\theta(v)$ (Leadbetter et al., 1983), with $\theta(v) \in [0, 1]$. The reciprocal of the subasymptotic extremal index gives the mean of the cluster size distribution of the extremes in a time-series at a level $v$, i.e.

$$\theta(v)^{-1} = \sum_{i=1}^{\infty} i\pi(i, v),$$

To estimate the above cluster functionals it is sufficient to use the forwards simulation scheme of Rootzén (1988). This approach is easier to implement since it only requires forward simulation and does not require the initial simulation to be the cluster maximum. For the $k$th-order approach outlined here this is the more efficient computationally.

The simulation approach to apply the model in Sections 4.2 and 4.3 to generate realisations of a tail chain with $k$th-order structure is as follows. To commence simulating the tail chain, we simulate a starting exceedance, $X_0^*$, of $v > u$ to be generated as $X_0^* = v + E_0$ where $E_0$ is from an Exponential distribution with rate parameter 1. For all time-steps $0 < j \le k$ the tail-chain is stepped forward using the appropriate $j$th order scheme by setting

$$X_j^* = \alpha_j X_0^* + (X_0^*)^{\beta_j}\, Z_{j|0:j-1}^*, \qquad (4.4.2)$$

where $Z^*_{j|0:j-1}$ is sampled from $\hat{G}_{j|1:j-1}$ with weights $w_i$, for $i = 1, \ldots, n_u$, and $\hat{G}_{1|0} = \hat{G}_1$. At all subsequent time-steps $k < j < m$ the $k$th-order scheme is used to complete the simulation of the tail chain, i.e.

$$X^*_j = \alpha_k X^*_{j-k} + \left(X^*_{j-k}\right)^{\beta_k} Z^*_{j|j-k+1:j-1}, \tag{4.4.3}$$

where $Z^*_{j|j-k+1:j-1}$ is sampled from $\hat{G}_{k|1:k-1}$ with weights $w_i$, for $i = 1, \ldots, n_u$. The tail chain length $m$ needs to be chosen large enough to ensure a negligible probability of obtaining a chain with $X^*_m > v \mid X^*_1 > v$ (in Section 4.5, $m = 40$ is found to be sufficient); see Section 3.4 for a similar discussion of the tail chain. A complete algorithm for the forward simulation approach with higher-order structure is given in Appendix D. The asymptotic justification for the algorithm is satisfied only when $X^*_j > u$, for $0 \le j \le m - k$. For $X^*_j < u$ the algorithm is used as it should still provide a reasonable approximation unless $X^*_j \ll u$. In this case the probability of the tail chain coming above $u$ again is negligible and thus the chain can be terminated.

When reporting results it is more instructive to give estimates of the probability of a particular event occurring within a given time period (often taken to be a year). In Chapter 3 we outlined an approach for deriving over cluster results from within cluster quantities based upon the assumption that clusters of the modelling threshold $u$ occur as a Poisson process (Hsing, 1988). The probability of observing at least one cluster in time period $T$ with at least $i$ days above the critical level $v$ is given by

$$\psi_v(i) = 1 - \exp\left\{-\tau_v \bar{\pi}(i, v)\right\},$$

where $\bar{\pi}(i, v) = \mathrm{P}(D_v \ge i \mid M > v)$ and

$$\tau_v = \theta(v)\lambda_u n_T \left[1 + \xi \left(\frac{v - u}{\sigma_u}\right)\right]^{-1/\xi}_+,$$

with $n_T$ the number of observations in a time period of interest.

## 4.5   Data analysis

Daily temperature observations are taken at Orleans, in central France, for the period 1946-2012. Four missing values exist in the time-series and are omitted, none occur during the 2003 heatwave event that we focus aspects of our analysis on. Heatwaves are most likely to occur in summer months, here defined as the 92 day period of June-August, so summer season and yearly return levels are equivalent. These three month periods are extracted from each year to form an approximately stationary time-series for the temperature.

First, a GPD is fitted to exceedances of the modelling threshold $u_Y$, set at 29.7°C, chosen using standard diagnostics (Coles, 2001). Diagnostic plots for this data set and justification of the GPD model and threshold choice are given in Chapter 3. The rate parameter $\lambda_u$ is estimated as 0.099 (0.007), the GPD scale parameter is estimated as $\hat{\sigma}_u = 3.002$ (0.225) and the shape parameter $\hat{\xi} = -0.215$ (0.033); the standard errors are given in the parentheses.

Estimates for the conditional extremes dependence parameters $(\alpha_j, \beta_j)$ and dependence measure $\chi_j(v_1)$ are given in Table 4.5.1 for $j = 1, \ldots, 10$. The estimates of $\chi_j(v)$ are obtained from the pairwise conditional model for $X_{t+j} \mid X_t$ and are denoted by $\hat{\chi}_j(v)$. Throughout this section, when not stated otherwise the critical level is set at the one-year return level, denoted $v_1$ and taking the value 35°C. Here, estimates of $\chi_j(v_1)$ are obtained using the standard conditional extremes approach for a bivariate vector from Section 4.2.3. We observe that $\hat{\chi}_j(v_1)$ decreases as $j$ is increased, showing a reduction in the level of extremal dependence with lag. Such a pattern is confirmed by the value of $\hat{\alpha}_j$ which moves further from 1 as $j$ is increased. We note that estimates of $\chi_j(v_1)$ decrease monotonically with lag whereas this is not the case for the conditional extremes dependence parameters. This highlights that $\alpha_j$ and $\beta_j$ are correlated and as such there is often a trade-off between the two parameters.

| $j$ | $\hat{\alpha}_j$ | $\hat{\beta}_j$ | $\hat{\chi}_j(v_1)$ |
|-----|------------------|-----------------|---------------------|
| 1 | 0.713 (0.072) | 0.524 (0.094) | 0.508 (0.027) |
| 2 | 0.576 (0.080) | 0.538 (0.126) | 0.276 (0.042) |
| 3 | 0.440 (0.084) | 0.514 (0.163) | 0.186 (0.041) |
| 4 | 0.342 (0.083) | 0.400 (0.182) | 0.144 (0.037) |
| 5 | 0.395 (0.082) | 0.301 (0.201) | 0.117 (0.031) |
| 6 | 0.288 (0.077) | 0.286 (0.226) | 0.095 (0.026) |
| 7 | 0.313 (0.069) | 0.253 (0.210) | 0.076 (0.018) |
| 8 | 0.259 (0.053) | 0.280 (0.193) | 0.067 (0.016) |
| 9 | 0.198 (0.040) | 0.091 (0.158) | 0.036 (0.011) |
| 10 | 0.162 (0.037) | -0.061 (0.143) | 0.019 (0.008) |

Table 4.5.1: Estimates for the extremal dependence parameters $(\alpha_j, \beta_j)$ and extremal dependence measure $\chi_j(v_1)$ for a set of different lag values $j = 1, \ldots, 10$ given at the one year return level $v_1$. The estimates of $\chi_j(v)$ are obtained from the pairwise model for $X_{t+j} \mid X_t$ and are denoted by $\hat{\chi}_j(v)$. Standard errors are given in parentheses.

A standard approach to estimate the order of a Markov chain is to identify the largest lag at which the partial auto-correlation function (PACF) is deemed to be significantly different from zero (Chatfield, 2003), since this function gives the strength of the dependence between $(X_t, X_{t+j}) \mid \mathbf{X}_{t+1:t+j-1}$, for $j = 1, \ldots, \tau$, for some maximum lag of interest $\tau$. In Figure 4.5.1 the auto-correlation function and PACF are plotted for the Orleans daily maximum temperature data. The auto-correlation shows a decay of dependence, which is near exponential. There is a large spike in the PACF at lag 1 with smaller values at all larger lags. This motivated the first-order Markov model used in Chapter 3. However, there are some values that lie outside the confidence intervals up to lag 6 which suggest that a first-order Markov model might omit some important higher-order structure.
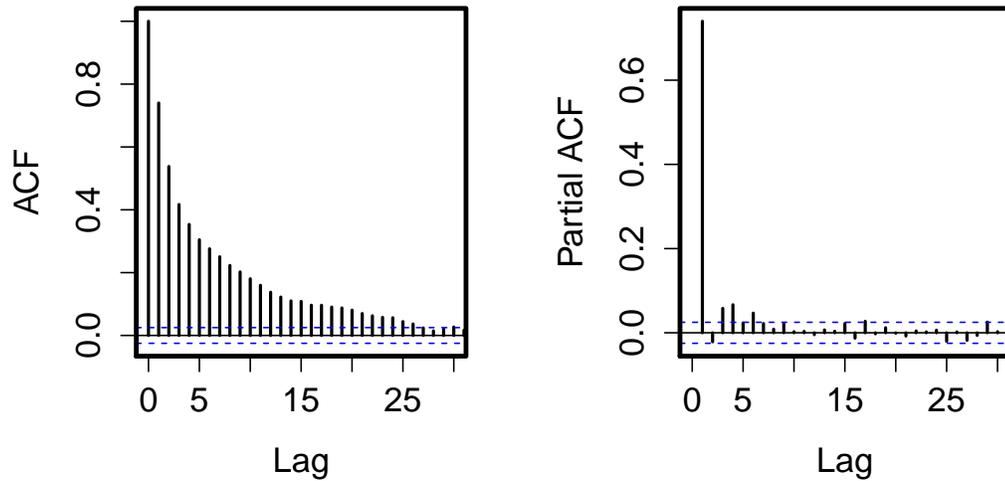
Figure 4.5.1: Auto-correlation and partial auto-correlation functions for Orleans daily maximum temperature data. Dashed intervals represent a 95% confidence interval.

In Section 4.3 we provided a set of different diagnostics to estimate the order Markov chain when focusing on the extreme values. Figure 4.5.2 gives estimates of $\chi_j(v)$ empirically (denoted $\tilde{\chi}_j(v)$) and for $j \geq k$ for different $k$th-order Markov models (denoted $\hat{\chi}_j^{(k)}(v)$) at two different values of the critical level $v$. If the process is $k$th-order then we should find that $\hat{\chi}_j^{(k)}(v)$ is close to $\tilde{\chi}_j(v)$ for all $j \geq k$. With $v$ corresponding to the 90% quantile, it appears that the third-order scheme comes closest to the pattern observed in the empirical estimates. First and second order schemes seem to underestimate the strength of the dependence whereas higher order estimates seem to lead to an overestimation. The performance of the different order Markov schemes is similar when the threshold is set at the 95% quantile, although the higher-order schemes seem to be contained within the empirical confidence bands for higher values of $j$. When the critical level is increased to the 99% quantile (not shown) there are little data available for the empirical approach and as such the 95% confidence intervals are wide and do not provide much information. In this situation, $\hat{\chi}_j^{(k)}(v)$ was compared

against the unstructured estimate $\hat{\chi}_j(v)$ to provide a more reliable diagnostic. From the diagnostic in Figure 4.5.2 we conclude that the third-order scheme seems to provide the most reliable estimates of $\chi_j(v)$ at all levels. The very high order approaches ($6 \leq k \leq 14$) seem to be greatly overestimating the dependence.

Figure 4.5.3 shows estimates of different extremal quantities related the duration of heatwave events for different order Markov models. We aim to identify the lowest order for which these extremal quantities remain consistent for all higher orders. Although the estimates suggest that the average number of exceedances within a cluster is larger when a higher-order Markov chain is used, the estimates of the subasymptotic extremal index $\theta(v_1)$ in Figure 4.5.3 seem to provide little information about order choice. The uncertainty bounds in Figure 4.5.3 are obtained via a modified bootstrap, used since the higher-order algorithm outlined in Appendix D is computationally intensive it is not feasible to run many bootstrap replications (i.e. 1000). We run a reduced number of replications (here 20) to approximate the standard error for the measure of interest and then construct symmetric confidence intervals around the point estimate using this standard error. If the computational cost of the algorithm could be reduced, more bootstrapped replications could be used and the diagnostic might have more power. The use of conditional kernels here is the cause of this inefficiency. We do not specifically seek to improve the efficiency of the code here, but it is noted that this is an important direction for future work.

Along with this measure, the probability of a cluster with 2, 6 and 11 exceedances of $v_1$ is estimated. These estimates give probabilities for short, medium and long events respectively and we aim to ensure that an order is chosen that can capture each type of event sufficiently. The short and long event types coincide with the observed duration of the 2003 European heatwave; two separate events of 2 and 11 days that were observed within a four week period at Orleans in the summer of 2003. The
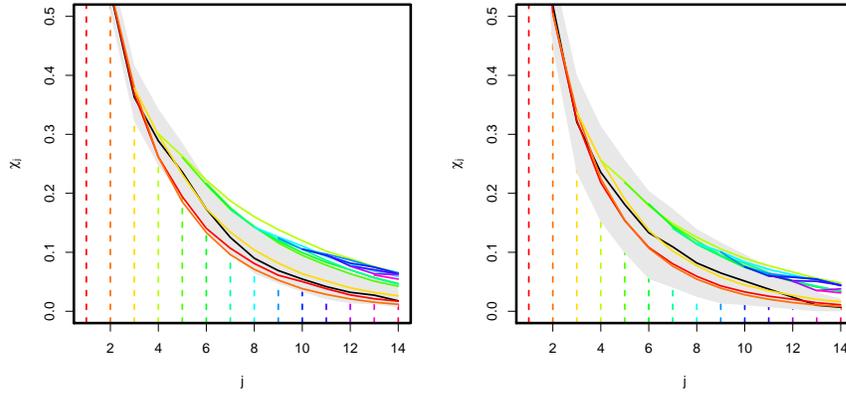
Figure 4.5.2: Estimates of the threshold dependent extremal measure $\chi_j(v)$ using empirical approach (black) and different order Markov chains (rainbow) with $v$ set at 90% (left) and 95% (right) quantiles respectively. Grey shaded region corresponds to 95% confidence interval for empirical obtained via a block bootstrap approach.
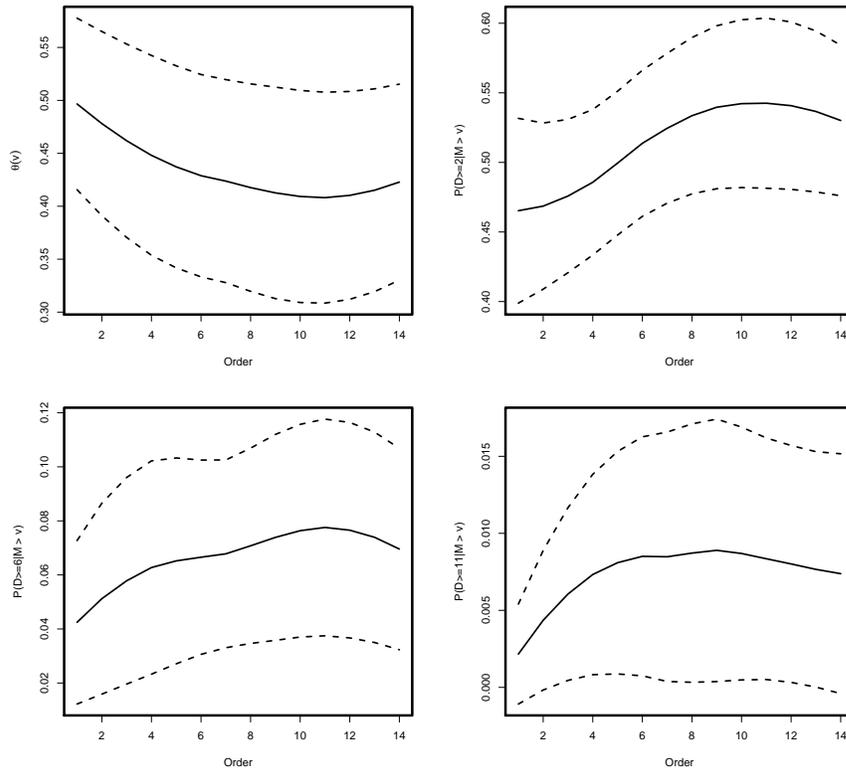


Figure 4.5.3: Estimates of within cluster extremal quantities for different higher-order schemes with $v$ set at the one-year return level $v_1$. Modified bootstrapping approach used to obtain 95% confidence intervals (dotted). Estimates have been smoothed using loess method for clarity.

selected quantities are estimated for the $k$th-order scheme where $k = 1 \ldots, 14$. For probabilities associated with shorter events, e.g. $P(D_{v_1} \geq 2 \mid M > v_1)$, our diagnostic in Figure 4.5.3 suggests that we need $k \geq 3$, since a horizontal line drawn from the central estimate at $k = 3$ would not intersect with the confidence intervals for higher $k$. For longer events the diagnostic suggests that lower order schemes ($k \geq 1$) could be used. However, very long events are rarely observed and the behaviour of these measures is highly uncertain for all orders.

To support the visual diagnostics in Figures 4.5.2 and 4.5.3, a hypothesis test is constructed to test whether a $k$th-order dependence structure provides a significantly better fit than a first-order approach. Under a first-order model ($\alpha_j$, $\beta_j$) are constrained as $\alpha_j = \alpha_1^j$ and $\beta_j = \beta_1$ whereas both parameters are allowed to vary for the $k$th-order model. To counteract any problems associated with multiple testing, the Bonferroni correction is used (Dunn, 1961). Tests are constructed for $j = 2, \ldots, 10$ and as such the significance level is set at $0.05/9$. All tests for which $j \geq 7$ are found to be significant at the 5% significance level. Such a test only suggests that the true order is $k \geq 2$.

Taking into account the diagnostic plots and hypothesis test we take the 3rd- and 7th-order schemes and estimate further extremal quantities for a Markov model of these orders. In Chapter 3 we estimated the probability of an event that lasts longer than the heatwave event in 2003 using a first-order Markov chain approach. In the 2003 event there were two events, of length 2 and 11, above the critical level $v_1$ during a four week period. In Chapter 3 we estimated the probability of observing at least one event in a year that lasts at least 2 days as 0.208 (0.200, 0.216), where the 95% confidence interval is given in parentheses; for 11 days the equivalent probability is 0.001 ($1 \times 10^{-4}$, 0.004), equivalent to the 1000 year return level. Using the 3rd-order Markov model yields the estimated probabilities of observing at least one event in

a year that lasts at least 2 and 11 days as 0.196 (0.171, 0.221) and 0.002 (0, 0.004) respectively. The equivalent probabilities for the 7th-order Markov model are 0.201 (0.179, 0.224) and 0.003 (0, 0.007) respectively. Thus it appears that the inclusion of higher order structure does not greatly affect the probability of smaller events but can lead to a 3-fold increase in the point estimates of the probability of very long duration extreme events. Uncertainty estimates are again wider for the higher-order approaches. This increase may be partially due to the use of the modified bootstrap but also reflects the increased number of parameters to be estimated.

## 4.6 Discussion and conclusion

This paper provides a new framework for incorporating higher-order Markov models for temporal dependence when modelling extreme events. Such an approach is motivated by the application to heatwave events, since models under a first-order Markov assumption do not adequately capture the prevailing conditions. For this purpose we have developed a $k$th-order Markov model framework for incorporating higher-order information using the conditional extremes approach. The new framework has necessitated new diagnostics for choosing the 'best' order scheme to use for extreme events from an efficiency perspective. Our results show that using standard time series diagnostics can lead to the identification of an order Markov process that does not adequately capture extremal features; e.g. for the extreme temperature data over Orleans, a PACF suggests a first-order process is adequate whereas our diagnostics suggest that a third-order process may be more appropriate. Specifically, standard time-series diagnostics ignore more complex structure in the extremes, which if not captured leads to an underestimation in the probability of longer and potentially devastating heatwave events. One area for further work is to formalise and unify our range of visual diagnostic methods for estimating the order of the extremal Markov process in order to provide one best approach that could be used by decision makers

and practitioners when faced with a similar problem.

As in Chapter 3, daily maximum temperatures have been analysed instead of looking at the joint distribution of daily maximum and minimum temperatures. A Markov model would still be appropriate in such a situation but a different order scheme might be required. The effect of climate change and other large scale climatic phenomena have not been incorporated into this paper. In Chapter 5 we illustrate how the tail chain simulation approach with first-order dependence structure can be altered to take into account the effect of covariates. A similar extension could be applied to the method outlined here.

# Chapter 5

# Detecting changing behaviour of heatwaves with climate change

## 5.1 Introduction

Heatwaves are events that are characterised by a set of hot days and nights which are associated with a marked increase in the mortality rate (IPCC, 2012). High temperatures reduce the capacity of the human body for heat loss and are likely to cause core body temperature to exceed healthy limits (37-39ºC). Most casualties in a heatwave are caused by heat exhaustion which leads to heat stroke. Heat exhaustion increases the blood pressure and leads to cardiovascular stress, which if not relieved results in heat stroke, cellular damage and an increased risk of mortality (Donaldson et al., 2003). Young and old people are particularly vulnerable during heatwave events. Food security is also adversely affected by climatic extremes such as heatwaves (Falloon and Betts, 2010). For instance the 2003 heatwave in Europe was estimated to reduce maize yields by up to 36% in Italy (Stott et al., 2004) while the 2012 heatwave in the USA reduced maize production by 13% in 2012 compared to the reported 2011 production (USDA, 2013). Determining future food security impacts will therefore require an understanding of how the intensity and duration characteristics of heatwaves

might change. Critical thresholds (Falloon et al., 2014) are crop dependent and may vary regionally even for the same crop (Wheeler et al. (2000), Koehler et al. (2013), Asseng et al. (2013)).

The importance of understanding how climate change will affect heatwaves is highlighted by the number of papers that investigate this issue, especially detection and attribution of factors that lead to an increased risk of extreme events. Stott et al. (2004) give an attribution study that suggests that the 2003 European heatwave was 2-4 times more likely when including anthropogenic climate forcings as opposed to just considering natural climate forcings. A similar study by Christidis (2005) detects significant human influence on extremely warm nights but not for extremely warm days. Changes in the behaviour of heatwaves can manifest themselves in different ways; this paper focuses on how changes in the behaviour of heatwaves can be modelled using extreme value theory. Within this framework climate change could cause an increase in critical levels (such as a return level), or could affect the duration and severity characteristics of events. If the duration and severity of heatwaves increase, this will have an effect on mortality rates and there would be a need to mitigate for these effects.

For this study we have daily maximum temperature values from a single grid-box over Orleans, in central France, from 13 state of the art general circulation models (GCMs) from phase 5 of the Coupled Model Intercomparison Project (CMIP5) database (Taylor et al., 2012). Each of the models is forced with the RCP8.5 'business as usual' high emissions scenario for the period 2006-2090. GCMs are complex computer simulations designed to replicate observed climate variables. Many different climate forcings can be included (e.g. greenhouse gas emissions, cloud cover) in a climate model run and each model will have different behaviour. Our ensemble of GCMs is selected using available GCMs of similar resolution that give a reliable

estimate of the variability in the climate system. We aim to detect any change in the behaviour of heatwaves over central France for one member of this ensemble, the Met Office's HadGEM2-ES GCM (Martin, 2011), and then use the whole ensemble to test whether these changes are consistent across GCMs.

Operational definitions of heatwaves are generally split into three different categories (Koppe et al., 2004) based upon: (i) an air temperature threshold; (ii) an air temperature threshold and minimum duration; (iii) indices based upon air temperature and relative humidity; in this paper we focus on cases (i) and (ii). Abaurrea et al. (2007) use values that exceed the 95th percentile of June-August daily maximum temperatures from 1971-2000. The same critical level is used in Stefanon et al. (2012), but a minimum duration of 4 days is introduced. Fischer and Schär (2010) use a local 90th percentile with a minimum duration of at least 6 consecutive days. In each approach only the exceedances of a threshold are used during the modelling process. Relative thresholds are preferred to absolute thresholds when defining a heatwave since temperature can vary by geographical location and humans are able to adapt to local climate (Nitschke et al., 2011). Although heatwave definitions vary the importance of estimating the duration and severity of events accurately is universally recognised.

Let $\{Y_t\}$ denote the time-series of daily maximum temperatures over the summer period. The intensity of a heatwave can be modelled by fitting an extreme value model to exceedances by a stationary series $\{Y_t\}$ of a high modelling threshold $u^Y$. The most common method is to fit a generalized Pareto distribution (GPD)

$$\mathrm{P}\left(Y_t - u^Y > y \mid Y_t > u^Y\right) = \left(1 + \tilde{\xi} y / \tilde{\sigma}_{u^Y}\right)_+^{-1/\tilde{\xi}} \qquad \text{for} \quad y \geq 0,$$

where $c_+ = \max(c, 0)$ and $\tilde{\sigma}_{u^Y} > 0$ and $\tilde{\xi}$ are the scale and shape parameters of the GPD respectively (Coles, 2001). However under climate change $\{Y_t\}$ is non-stationary, so approaches that model exceedances above a constant threshold $u^Y$ can be problematic since the sample of exceedances will be dominated by values from certain points

in the time-series (for example there are likely to be more exceedances in future years under climate change). Pre-processing (Eastoe and Tawn, 2009) can be used to obtain a standardised time-series $\{Y_t^s\}$ from the original series $\{Y_t\}$ which is approximately marginally stationary over all $\{Y_t^s\}$. A GPD is fitted to the exceedance by $\{Y_t^s\}$ of a high threshold $u^s$ to obtain the scale and shape parameters $(\sigma_{u^s}, \xi)$. Davison and Smith (1990) explain how to incorporate covariates into an extreme value framework by allowing the parameters of the GPD to vary as a function of covariates. As is common in extreme value studies the standardised series is then transformed onto a common scale, here Laplace, to give the series $\{X_t\}$.

It is important to stress the difference between marginal non-stationarity and non-stationarity in the extremal dependence structure. Pre-processing will remove non-stationarity in the marginal distribution, however it does not account for non-stationarity between consecutive time points. In the framework of Markov chains, incorporating covariates into the conditional distribution of $X_{t+1} \mid X_t$ will allow assessment of how the dependence between values on successive days changes with a covariate. A previous study of the effect of covariates on dependence structure appears in Jonathan et al. (2013) for estimating wave heights in the North Sea as a function of the wave direction. In the heatwave problem, a change in the marginal characteristics leads to a change in the overall strength of a heatwave whereas a change in the dependence characteristics leads to a change in the persistence of events. Both of these factors are important when mitigating for heatwave events.

The main aim of this paper is to provide a coherent extreme value framework for investigating the effect that climate change will have on the behaviour of heatwave events. Reich et al. (2014) have previously modelled heatwaves using a GPD to capture marginal characteristics and a max-stable process to model dependence in a Bayesian hierarchical framework. Many other previous studies have focused on the

occurrence of singular hot days and how this might vary with climate change while ignoring changes in the persistence of events. Our approach incorporates analysis of the former through pre-processing of the margins while providing a framework for testing for significant changes in the latter through the conditional extremal dependence approach. The pre-processing approach of Eastoe and Tawn (2009) provides a natural framework for modelling temperature series where a constant threshold approach cannot be applied; the conditional extremal dependence approach offers a more flexible way of estimating extremal dependence properties of Markov chains than previous methods since it covers all types of extremal dependence. In comparison, Smith (1992) applies only for a restricted special case and the approach Bortot and Tawn (1998) only holds over a much narrower tail region. In practice, this means that the duration and severity of a heatwave event is permitted to change with critical level (i.e. a heatwave exceeding the 1 year return level will have different dependence characteristics than a heatwave exceeding the 50 year return level).

Section 5.2 introduces the temperature data set from the HadGEM2-ES GCM and the pre-processing technique used to obtain a marginally stationary series. The conditional extremal dependence approach is outlined and extended to include covariates in Section 5.3. Methods for simulating clusters of extreme values to derive heatwave properties are briefly mentioned in Section 5.4. Results for the HadGEM2-ES GCM are given in Section 5.5 and results over the rest of the GCM ensemble are given in Section 5.6. Discussion and conclusions are provided in Section 5.7.

## 5.2 Heatwave application

### 5.2.1 HadGEM2-ES Data

General circulation models are large scale computer simulations that aim to replicate the known physical processes of the climate system. Here, we have daily maxi-

mum temperature values for a single grid-box over Orleans, in central France, for 13 GCMs over the period 2006-2090. Since heatwaves are most likely to occur in summer months, here defined as the 90 day period of June-August (the HadGEM2 climate models are built upon months consisting of 30 days), these three month periods are extracted from each year and as such summer season and yearly return levels are equivalent. The top left plot of Figure 5.2.1 shows the HadGEM2-ES GCM temperature data represented as a time series which clearly shows marginal non-stationarity, but within each year (summer period) values of the time-series are approximately marginally stationary. Since data are taken from a GCM the problem of missing values is avoided for this analysis. We are going to use global mean temperature ($g_t$), as produced by the GCM, as a covariate through this study to capture non-stationarity in the margins and dependence structure. The bottom right plot of Figure 5.2.1 shows the time-series of global mean temperature from HadGEM2-ES. A full analysis is undertaken for this series in Section 5.5.

## 5.2.2 Pre-processing

Figure 5.2.1 shows that the data are non-stationary which poses problems when trying to use threshold methods with a constant modelling threshold. Eastoe and Tawn (2009) give a framework for transforming marginally non-stationary data such that constant threshold approaches can be used. Specifically, taking the original non-stationary time-series $\{Y_t\}$ the transformation

$$[Y_t^{\kappa(g_t)} - 1]/\kappa(g_t) = \psi(g_t) + \tau(g_t)Y_t^s,$$

yields the approximately stationary standardised sequence $\{Y_t^s\}$, where $(\psi(g_t), \tau(g_t))$ are location-scale parameters, $\kappa(g_t)$ is the Box-Cox parameter and $g_t$ is the global mean temperature. In this paper all covariates are included in a linear manner, i.e.

$$\kappa(g_t) = \kappa_0 + \kappa_1 g_t \qquad \psi(g_t) = \psi_0 + \psi_1 g_t \qquad \log \tau(g_t) = \tau_0 + \tau_1 g_t.$$
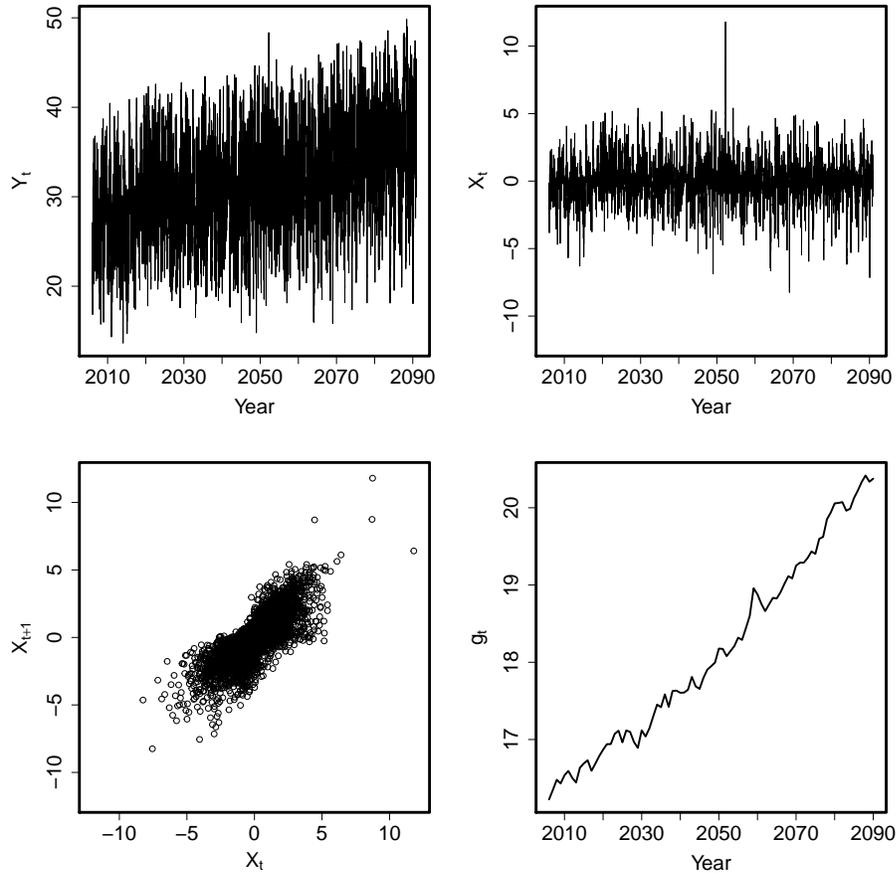
Figure 5.2.1: Original June, July and August daily temperature data ($^{o}$C) from HadGEM2-ES GCM represented as a time-series (top left), same data on Laplace margins after pre-processing to ensure marginal stationarity (top right), pre-processed data as a set of consecutive pairs (bottom left) and global mean temperature taken as covariate (bottom right). Data from separate years have been concatenated for the time-series plot to show only relevant data. As such continuity of data from year to year is induced but not considered during modelling procedure.

Higher-order covariate relationships are possible but not investigated here. In practice the Box-Cox location-scale model may not completely capture all of the non-stationarity in the extremes and a GPD model is fitted to the upper tail of the

margins of the standardised series $\{Y_t^s\}$ such that

$$
F_{Y_t^s}(y) = \begin{cases} 1 - \lambda_{u^s}(g_t)\left[1 + \xi(g_t)\left(y - u^s\right)/\sigma_{u^s}(g_t)\right]_+^{-1/\xi(g_t)} & \text{if} \quad y \geq u^s \\ \tilde{F}(y) & \text{if} \quad y < u^s, \end{cases} \tag{5.2.1}
$$

where $u^s$ is the modelling threshold on the part pre-processed margins (i.e. margins that have undergone the location-scale transform but not the GPD part of the transform), $(\sigma_{u^s}(g_t), \xi(g_t))$ are scale and shape parameters that depend on the covariate such that $\log \sigma_{u^s}(g_t) = \sigma_0 + \sigma_1 g_t$ (where $\sigma_0$ and $\sigma_1$ depend on the threshold $u^s$ but subscript is dropped for notational simplicity) and $\xi(g_t) = \xi$, $\lambda_{u^s}(g_t) = 1 - \tilde{F}(u^s)$ and $\tilde{F}(y)$ is the empirical cumulative distribution function of $\{Y_t^s\}_{t=1}^n$. It is assumed that non-stationarity in the body of the distribution is accounted for using the original location-scale transform and as such the stationary empirical distribution function is appropriate for values that fall below or equal to the modelling threshold $u^s$. To study the extremal dependence structure it is common to transform the marginal to a standard form. We transform $Y_t^s$, $t = 1, \ldots, n$ onto Laplace margins as follows

$$
X_t = \begin{cases} \log\left\{2F_{Y_t^s}(Y_t^s)\right\} & \text{if} \quad Y_t^s < F_{Y_t^s}^{-1}(0.5) \\ -\log\left\{2\left[1 - F_{Y_t^s}(Y_t^s)\right]\right\} & \text{if} \quad Y_t^s \geq F_{Y_t^s}^{-1}(0.5), \end{cases} \tag{5.2.2}
$$

where $F_{Y_t^s}$ is given in equation (5.2.1). The top right plot in Figure 5.2.1 shows a plot of the transformed data on Laplace margins showing an assumption of marginal stationarity to be appropriate for $\{X_t\}$. Estimates of all the parameters for our data set are given in Section 5.5.

## 5.3 Modelling temporal dependence

### 5.3.1 Markov modelling

To obtain estimates for the duration and severity of heatwave events it is necessary to develop a model for the evolution of the temperature time-series. Here, supported by

data analysis, an assumption that the time series follows a first order Markov process is made. By the Markov property the distribution at each time step is only affected by the state of the system at the time-step before. As such to model the extremes of the transformed stationary time series $X_1, \ldots, X_n$ it is only necessary to model the extremes of pairs $(X_t, X_{t+1})$ for $t = 1, \ldots, n - 1$.

As mentioned in Section 5.1, different methods for modelling bivariate data have been presented. Broadly these methods are split into two categories depending on whether heatwave characteristics change with the critical level. The category is determined by the value of $\chi$ where

$$\chi = \lim_{x \to \infty} \mathrm{P}(X_{t+1} > x \mid X_t > x).$$

When $\chi = 0$ the variables $(X_t, X_{t+1})$ are said to be asymptotically independent and $\chi > 0$ defines asymptotic dependence. The assumption of a dependence structure that is asymptotically dependent leads to heatwave characteristics that are independent of the critical level. The joint tail approach outlined in Smith et al. (1997) uses a bivariate extreme value distribution with a parametric dependence structure to model the extremal dependence of $(X_t, X_{t+1})$. That approach assumes that an asymptotically dependent dependence structure holds which can be restrictive in many applications. The semi-parametric conditional extremal dependence approach outlined in Heffernan and Tawn (2004) allows for a richer class of dependence structures and most importantly allows for both asymptotic dependence and asymptotic independence, see Chapter 3 for details of how these two methods differ. The additional flexibility of the conditional extremal dependence approach is useful for our application and as such is used through the rest of this paper.

## 5.3.2   Semi-parametric stationary conditional extremal dependence approach

The conditional extremal dependence approach of Heffernan and Tawn (2004) and Heffernan and Resnick (2007) can be used to model the extremes of pairs $(X_t, X_{t+1})$ for $t = 1, \ldots, n-1$. Heffernan and Tawn (2004) gave their representation for Gumbel margins, but Keef et al. (2013) showed that a more comprehensive approach arises for Laplace margins (equation (5.2.2)). The desire is to model $(X_t, X_{t+1})$ using the distribution of $X_{t+1}$ given that $X_t$ is large (defined as exceeding a high threshold). A requirement for modelling the conditional distribution $\mathrm{P}\{X_{t+1} \leq x_{t+1} \mid X_t = x_t\}$ is that this distribution should be non-degenerate as $x_t \to \infty$ and hence $x_{t+1}$ needs to change appropriately with $x_t$. The specification of Laplace margins ensures that the upper and lower tails are both modelled by a symmetric distribution with exponential tails and permits the definition of a single unified class of normalising functions such that the conditional distribution from Heffernan and Tawn (2004) is given as

$$\mathrm{P}\left([X_{t+1} - \alpha X_t]/X_t^{\beta} \leq z, X_t - u^X > x \mid X_t > u^X\right) \to G(z)\exp(-x), \qquad (5.3.1)$$

as $u^X \to \infty$, where $u^X$ is the modelling threshold on Laplace margins, $G$ is a non-degenerate distribution function, $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1)$. This form of the normalising functions does not affect the limiting dependence model in Heffernan and Tawn (2004) and simplifies the inference for variables which are either negatively or weakly associated. If the variables are independent then $\alpha = \beta = 0$ and $G(z)$ is the Laplace distribution function whereas $\alpha = 1$ and $\beta = 0$ corresponds to the situation of asymptotic dependence, $-1 \leq \alpha \leq 0$ to negative extremal dependence and $0 < \alpha < 1$ or $\alpha = 0$ and $\beta > 0$ corresponds to asymptotic independence with positive extremal dependence. Thus unless $\alpha = 1$ and $\beta = 0$ this representation extends the asymptotic dependence class of Smith et al. (1997).

Modelling using the conditional extremal dependence approach requires the assump-

tion that the limiting form of equation (5.3.1) holds exactly for all values of $X_t > u^X$ given that $u^X$ is a sufficiently high threshold. Given this assumption it is possible to write the form of $X_{t+1}$ given that $X_t > u^X$ as

$$X_{t+1} = \alpha X_t + X_t^\beta Z_{t+1},$$

where $Z_{t+1}$ is a random variable with distribution function $G$ and is independent of $X_t$. As $G$ does not take any simple parametric form, a false working assumption is made as in Keef et al. (2013) that $Z_{t+1} \sim N(\mu, \gamma^2)$ and as such

$$X_{t+1} \mid \{X_t = x\} \sim N\left(\alpha x + \mu x^\beta, \gamma^2 x^{2\beta}\right) \qquad \text{for} \quad x > u^X.$$

The working assumption permits the estimation of the set of parameters $(\alpha, \beta, \mu, \gamma)$ by standard likelihood approaches. At this stage the Gaussian assumption is discarded and a non-parametric estimate of the distribution for $Z$ is formed. Specifically, let $t_1, \ldots, t_{n_{u^X}}$ be the indices of $t = 1, \ldots, n$ where $x_t > u^X$ then let

$$\hat{z}_j = (x_{t_j+1} - \hat{\alpha} x_{t_j} - \hat{\mu} x_{t_j}^{\hat{\beta}})/\hat{\gamma} x_{t_j}^{\hat{\beta}}, \tag{5.3.2}$$

for $j = 1, \ldots, n_{u^X}$, where $n_{u^X}$ is the number data points exceeding the threshold $u^X$. In this way a non-parametric estimate $\hat{G}$ to the distribution function $G$ is formed using $\hat{z}_j$, $j = 1, \ldots, n_{u^X}$.

### 5.3.3 Incorporating covariates

The process of incorporating covariates into the marginal parameters was highlighted in Section 5.2.2. However for a more complete analysis of the extremal behaviour it is necessary to ascertain whether the global temperature covariate has an effect on the level of extremal dependence by testing if the covariate has a significant effect on the dependence parameters. As in Jonathan et al. (2013) covariates are introduced into the set of parameters $(\alpha, \beta, \mu, \gamma)$ such that

$$\tanh^{-1}[\alpha(g_t)] = \alpha_0 + \alpha_1 g_t \qquad \tanh^{-1}[\beta(g_t)] = \beta_0 + \beta_1 g_t$$

$$\mu(g_t) = \mu_0 + \mu_1 g_t \qquad \log \gamma(g_t) = \gamma_0 + \gamma_1 g_t.$$

An inverse tanh link function is used for $\beta(g_t)$ as well as $\alpha(g_t)$ in this situation since in practice it is very unlikely that $\beta(g_t) < -1$ as this corresponds to $X_{t+1} - \alpha(g_t)X_t$ tending rapidly to zero for large $X_t$, i.e. $X_{t+1}$ is essentially deterministic given $X_t$. The impact of the covariate on the dependence structure is assessed using a likelihood ratio test. A non-parametric estimate of the distribution $G$ is formed using equation (5.3.2) with the covariate dependent set of parameters $(\alpha(g_t), \beta(g_t), \mu(g_t), \gamma(g_t))$ and the resulting $\{Z_t\}$ are assumed to be independent and identically distributed.

## 5.4 Cluster behaviour estimation

With a marginally stationary time-series obtained by pre-processing techniques we wish to estimate whether heatwave events become longer and more severe with climate change. We define a critical level, $v^Y$ on original margins and $v^X$ on Laplace margins, as some level of interest above which events are extreme. Such a level will often be related to the $T$ year return level and denoted $v_T^Y$ or $v_T^X$ depending on the margin of interest, with $v_T^Y$ time dependent. A cluster is defined as a set of points which exceed the critical level $v^X$, preceded and followed by a set amount of non-exceedances (Smith and Weissman, 1994). The common measure linked to clusters is the threshold dependent extremal index $\theta(v^X)$ (Leadbetter, 1983), defined as the reciprocal of the average number of exceedances of $v^X$ in a cluster (the duration of events above $v^X$). Since we assume stationarity within years the value of $\theta(v^X)$ and $\theta(v^Y)$ will be approximately equivalent. For the rest of the paper $D$ will relate to the duration of a heatwave event in days and $S$ the severity in degrees Celsius (°C). There are varying definitions of the severity of any type of extreme event (e.g. Mishra and Singh (2010)). In this paper we shall refer to the severity of an event as the sum of all excesses of a critical level within an event on the original temperature scale, i.e.

$$S = \sum_{t \in C} \left(Y_t - v^Y\right)_+ ,$$

where $C$ is a set of values comprising a cluster. The duration of an event is defined as the number of days above the critical level within an event, i.e.

$$D = \sum_{t \in C} \mathbb{I} \left( Y_t - v^Y \right)_+ ,$$

where $\mathbb{I}(.)$ is the indicator function. In this study we look to estimate $P(D > d \mid M > \eta)$ and $P(S > s \mid M > \eta)$; given a peak value of a cluster $M$ is greater than some critical value $\eta$ with $\eta \geq v^Y$ these represent the probability of an event that has more than $d$ exceedances of $v^Y$ or has a severity greater than $s$ respectively.

Our approach to deriving the properties of clusters of a Markov chain is repeated simulation of the chain in periods with exceedances of a critical value, i.e. when the process exceeds $v^Y$, with $v^Y \geq u^Y$. We adopt the approach outlined in Chapter 3, an extension of Smith et al. (1997), called peak value chain estimation, by simulating the cluster maximum $M > v^Y$ and then simulating forwards and backwards from this peak value using the conditional model. Estimation of the forward chain is implicit in the approach in Section 5.3.2, estimation of the backward chain requires dependence parameters $(\alpha_b, \beta_b, \mu_b, \gamma_b)$ for $X_t \mid X_{t+1} > u^X$ to be estimated similarly. The approach behind peak value chain estimation allows full extreme events to be simulated permitting easy estimation of severity and duration characteristics. From the peak value tail chain estimation approach, we obtain estimates of $P(D > d \mid M > \eta)$ and $P(S > s \mid M > \eta)$ where $\eta$ is some cluster maximum of interest and $d$ and $s$ are critical values of duration and severity respectively. The joint probability of an event exceeding a given duration and severity will also be evaluated.

## 5.5   HadGEM2-ES Results

We fit the pre-processing method and test whether the global mean temperature covariate has a significant effect on the Box-Cox parameter and the location-scale parameters. The log-link function is used to ensure the non-negativity of the scale

parameter $\tau(g_t)$. At each stage of the pre-processing likelihood ratio tests at the 5%
significance level are used to assess whether the covariate effect is significant. We
find that the covariate has a significant effect on the Box-Cox parameter and the
location-scale parameters. The estimates of the Box-Cox parameters are $\hat{\kappa}_0 = 0.992$
$(0.040)$ and $\hat{\kappa}_1 = -0.018$ $(6 \times 10^{-4})$, where standard errors are given in parenthe-
ses. Estimates for the location-scale parameters are given as $\hat{\psi}_0 = 12.512$ $(0.292)$ and
$\hat{\psi}_1 = 0.037$ $(0.016)$ and $\hat{\tau}_0 = 2.596$ $(0.241)$ and $\hat{\tau}_1 = -0.082$ $(0.013)$ respectively. A
$\text{GPD}(\sigma_{u^s}(g_t), \xi)$ is fitted to the upper tail of the standardised data to assess whether
there is still any residual marginal non-stationarity in the extremes. Throughout this
study the modelling threshold $u^s$ is set at the 90th percentile. The global tempera-
ture covariate does not seem to have an effect on the estimate of the rate parameter
$\hat{\lambda}_{u^s} = 0.10$ $(0.007)$ or the shape parameter with $\hat{\xi} = -0.250$ $(0.020)$, the effect has
been removed by the first stage of the pre-processing, but does have an effect on
the scale parameter. Estimates of the scale parameters are $\hat{\sigma}_0 = 0.397$ $(0.379)$ and
$\hat{\sigma}_1 = -0.071$ $(0.021)$. The estimates above from the full pre-processing method are
used to transform the non-stationary series in the top left of Figure 5.2.1 into the
transformed marginally stationary series given in the top right of Figure 5.2.1.

The marginally non-stationary nature of the time-series means that the value of a
$T$-level return level varies with the value of the covariate. Below, the 95% confidence
intervals given in parentheses have been obtained by bootstrapping. The critical level
associated with the 1 year return period (denoted $v_1^Y$) is 39.5°C (38.9, 40.5) for the
global mean temperature in 2006, a value that increases by 1.8°C (1.5, 2.1) or 11.7°C
(9.7, 13.6) with an increase in the global mean temperature of 1°C or 5°C respectively.
The maximum value of the 2003 heatwave event over this region for observed values
was 39.9°C, equivalent to the 50 year return level for the observed series which shows
that the HadGEM2-ES GCM is significantly hotter than the observed series. Return
levels are obtained for the 50-year return period that increase from 43.5°C (41.4, 45.1)

by 1.7°C (1.1, 2.1) or 11.3°C (8.6, 13.4) under the same change in the covariate, which highlights that extreme levels change at a different rate to mean levels.

Estimates are now provided for the extremal dependence between consecutive days using the conditional extremal dependence approach. Each approach is evaluated using the modelling threshold $u^X$, set at the 90th percentile. We use a likelihood ratio test to determine whether the global mean temperature covariate has a significant effect on the key dependence parameters $\alpha$ and $\beta$. It is found that the data do not exhibit any change in the dependence structure with the covariate and as such the stationary model from Chapter 3 is used to analyse extremal dependence. Estimates of the dependence parameters $\alpha$ and $\beta$ are given as 0.168 ($-0.503$, 0.532) and 0.680 (0.496, 0.810) respectively, with bootstrapped 95% confidence intervals in parentheses. Parameter values for the backward chain are given as $\alpha_b = 0.681$ (0.529, 0.855) and $\beta_b = 0.445$ (0.023, 0.677) and since different pattern is detected in the dependence parameters than for the forward chain this suggests non time-reversibility. A likelihood ratio test confirms that the parameter values for both the forward and backward chain are significantly different from ($\alpha = 1, \beta = 0$) an ($\alpha_b = 1, \beta_b = 0$) respectively, the situation of asymptotic dependence, and as such the data exhibit asymptotic independence.

Since mean global temperature does not appear to have a significant effect on the extremal dependence of consecutive days we estimate the probability of observing an event with a specific duration using the stationary dependence parameters. As in Chapter 3 we estimate three quantities: the extremal index to give an estimate of the average behaviour of a cluster; the probability of an event whose 3 day average exceeds the one year return level; the probability of observing an event longer than the 2003 heatwave event, i.e. where there is an event of 11 days above the 1-year return level. We feel this triple provides information about the average heatwave event expected

as well as giving probabilities for very severe and potentially devastating events. To estimate all quantities described above we use the methods from Section 5.4. The extremal index $\theta(v^X)$ is given as 0.467 (0.410, 0.565) and suggests an average of just over 2 exceedances in a cluster. An event with a 3 day average that exceeds $v_1^Y$ is given as an important quantity in terms of mortality (Pascal et al., 2013), suggesting a potential excess mortality of up to 50% and triggering heat health warnings. The estimate of the probability of such event happening in a year is given as 0.178 (0.175, 0.241), equivalent to an event that happens on average once every 5.6 years. Finally, we estimate the probability of an event that lasts longer than the 2003 heatwave event in a given year as 0.001 ($2 \times 10^{-4}$, 0.002), an event that happens on average once every 1000 years.

One extension on Chapter 3 is to consider the severity of a heatwave alongside the duration since the non-linear nature of the marginal transformation means that the severity of an event could increase despite there being no difference in the duration of an event. We estimate quantiles of the distribution $S \mid M > v_1^Y$ to see if there is any change in the severity with an increase in the global mean temperature. At the 2006 global mean temperature level, the median of the distribution $S \mid M > v_1^Y$ is given as 2.2°C (1.5, 3.2). An increase of 1°C in the covariate leads to a change in the median severity of $-0.1$°C ($-0.5$, 0.2) and a 5°C increase leads to a change in the median severity of $-0.3$°C ($-1.7$, 1.1). The respective values for the 99th quantile are 6.0°C (4.0, 8.9) with a change of $-0.2$°C ($-1.3$, 0.6) and $-0.7$°C ($-4.6$, 3.1). The results show that the severity of heatwaves decreases at different high quantiles, however the confidence intervals suggest that the pattern is not certain with zero contained within all intervals for a change in severity with climate change. It must be stressed that since we are using a stationary extremal dependence model all these variations are coming from the effect of the covariate on marginal parameters. Previously we observed that $v_1^Y$ increases by 1.84°C for a 1°C increase in the global mean temperature

which swamps any estimated change of the severity distribution relative to this level; it is clear that marginal changes in return levels are more important than any change induced in the severity.

## 5.6 Results across GCM ensemble

### 5.6.1 Marginal results

Time-series for the collection of 12 additional GCMs to be analysed alongside HadGEM2-ES (Figure 5.2.1) are given in Figure 5.6.1. These particular GCMs have been chosen for their similarity to HadGEM2-ES in terms of spatial resolution. One runs per GCM has been used for simplicity to assess the methodology. Each GCM data set is taken through the pre-processing steps outlined in Section 5.2.2. Global mean temperatures are available for each of the GCMs and the respective values will be used as the covariate $g_t$. For each GCM the most general form of covariate dependence is assumed for each of the pre-processing parameters, except for the shape parameter $\xi$ which is assumed to remain constant over covariates but differs over GCMs. This ensures consistency across the models and any change in marginal parameters is directly comparable.

The change in the global mean temperatures for the period 2006-2090 range from 1-5°C across the GCMs and as such it is sensible to give results in terms of the change in behaviour for a 1°C and 5°C increase in the global mean temperature from the 2006 level. As the global mean temperature for each GCM varies over time, presenting results for a 1°C and 5°C increase is more consistent than giving values for particular years. It is noted that not all GCMs exhibit a 5°C increase in their global mean temperature and any results presented at this level will be extrapolations and wider confidence bands are expected.
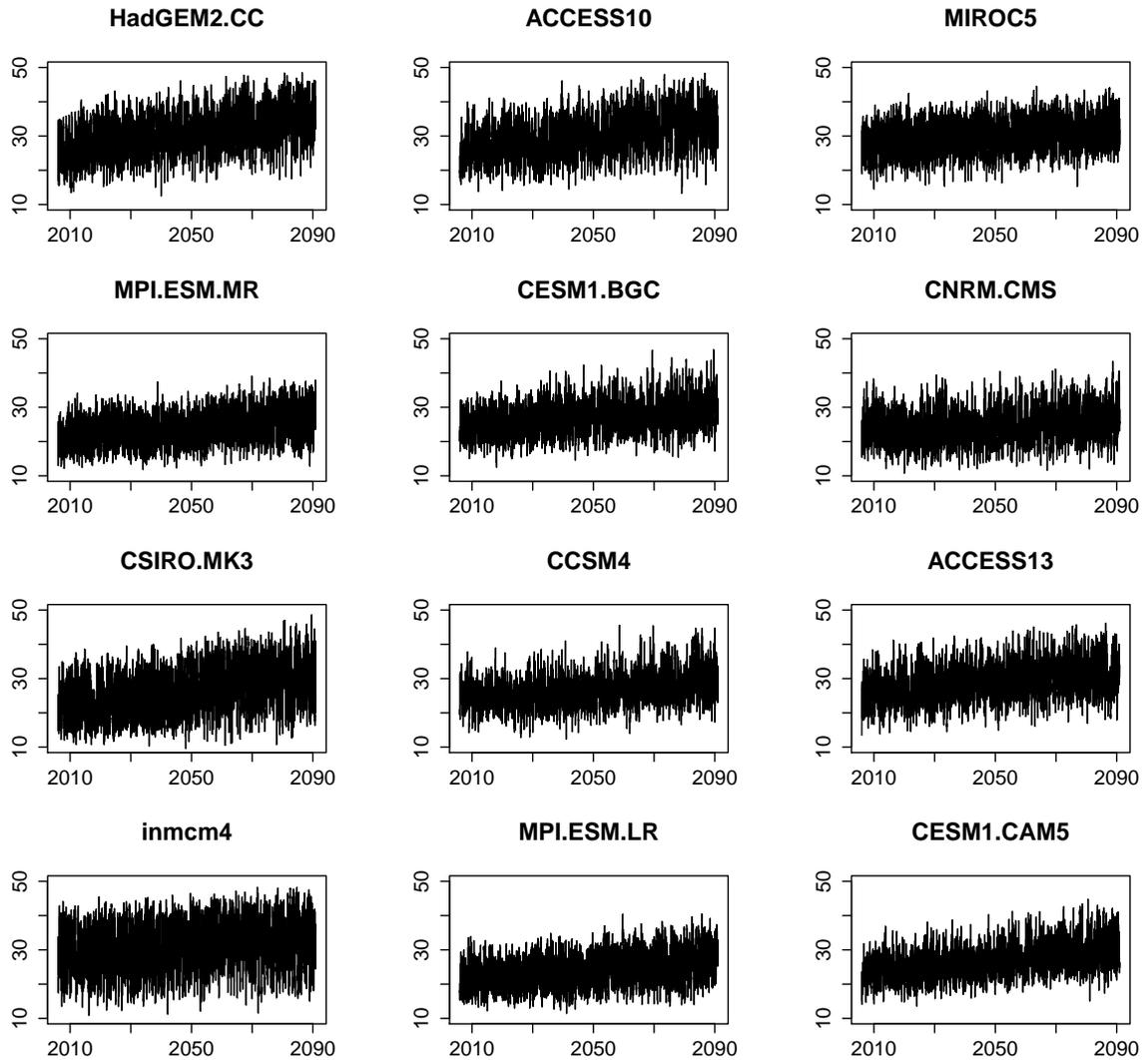
Figure 5.6.1: Temperature series (ºC ) for the period 2006-2090 for 12 different GCM models for a grid box containing Orleans in central France..

Table 5.6.1 shows the average marginal parameters for the 13 GCMs along with the range of different values estimated from the different series. For each parameter we also show the number of series that are identified to have a significant change with mean global temperature and the number of series that show a positive trend. The Box-Cox parameter $\kappa$ shows the clearest trend of all the parameters with the covariate being significant and negative for all GCMs. The location parameter shows an expected increase with the covariate, a pattern that was expected due to the upward trend in temperature values seen in Figure 5.6.1, although this pattern is not replicated for all series. The scale parameter $\tau$ and the scale parameter of the GPD, $\sigma_{u^s}$, both exhibit a slight decrease with an increase in the covariate and are broadly significant, which suggests that the spread of standardised temperature values is decreasing. The shape parameter is negative as expected when looking at the upper tail of temperature values.

|  | Estimate (GCM range) |  | Estimate (GCM range) | # Sig | # Pos |
|---|---|---|---|---|---|
| $\kappa_0$ | 0.649 (−0.110, 1.001) | $\kappa_1$ | −0.017 (−0.027, −0.009) | 13 | 0 |
| $\psi_0$ | 7.681 (2.143, 13.201) | $\psi_1$ | 0.020 (−0.017, 0.146) | 6 | 11 |
| $\tau_0$ | 0.195 (−4.667, 3.479) | $\tau_1$ | −0.069 (−0.120, 0.007) | 11 | 1 |
| $\sigma_0$ | 0.450 (−2.294, 2.935) | $\sigma_1$ | −0.069 (−0.213, 0.106) | 10 | 3 |
| $\xi$ | −0.278 (−0.381, −0.182) |  |  |  |  |

Table 5.6.1: Average marginal parameter estimates given across all 13 GCMs, range over models are given in parentheses. Penultimate column shows the number of series that show a significant change in the gradient of respective parameter with the covariate. Final column shows how many series show a positive change in respective parameter with the covariate.

The results in Table 5.6.1 are not conclusive across GCMs, with some series showing different behaviour to others. This result occurs for two main reasons. Firstly, each of the GCMs exhibit different increases in the temperature over time as well as a dif-

ferent amount of variability. As such different GCMs have a different behaviour with the covariate. Secondly, the marginal parameters are non-orthogonal and therefore reporting the separate parameter values does not give the whole picture. In Table 5.6.2 we report the behaviour of certain return levels under a change in the global mean temperature. An increase in the one and fifty year return levels can be observed for all GCMs in the ensemble with an increase in the covariate. It is interesting to note that a 1°C and 5°C increase in the global mean temperature can lead to a varied increase at different return levels, confirming that an analysis of the extremes is essential alongside any analysis of the average behaviour. It also seems that higher critical levels increase at a faster rate.

| $g$ | $g_{2006} + 1$ | $g_{2006} + 5$ |
|---|---|---|
| $v_1^Y(g) - v_1^Y(g_{2006})$ | 2.0 | 13.8 |
| | (1.3, 2.6) | (8.0, 18.8) |
| $v_{50}^Y(g) - v_{50}^Y(g_{2006})$ | 2.0 | 14.3 |
| | (1.2, 3.6) | (8.0, 26.0) |

Table 5.6.2: Change in the one year $(v_1^Y)$ and 50 year $(v_{50}^Y)$ return levels from $v_1^Y(g_{2006}) = 39.5$°C and $v_{50}^Y(g_{2006}) = 43.5$°C respectively after a 1°C increase and a 5°C increase from the 2006 global mean temperature. Results given for the average for the ensemble of GCMs with range given across all 13 GCMs in parentheses.

## 5.6.2 Dependence results

Having noted the significant increase in the magnitude of return levels, we explore whether there is any difference in the duration and severity characteristics of heatwave events. In Section 5.5 a likelihood ratio test showed that the global temperature covariate had no effect on the dependence parameters for the HadGEM2-ES series and

this pattern is observed across all GCMs within our ensemble. We repeat the analysis of the dependence between the temperature on consecutive days from Section 5.5 for the 13 GCMs after standardising each series to an approximately marginally stationary time-series. Below, average estimates for the GCM ensemble are reported with the range over all 13 GCMs given in parentheses. The extremal dependence parameters take the values $\alpha = 0.400$ (0.168, 0.775) and $\beta = 0.580$ (0.387, 0.743). The respective values for the backward chain dependence parameters are given as $\alpha_b = 0.624$ (0.472, 0.735) and $\beta_b = 0.408$ (0.126, 0.590).

### 5.6.3 Duration and severity results

We estimate the three measures of the duration outlined in Section 5.5 for each of the GCMs from our ensemble. We find the average extremal index to be 0.481 (0.406, 0.579), where the range across the 13 GCMs is given in parentheses. The probability of observing at least one event in a year where the 3 day average exceeds the one year return level is estimated as 0.177 (0.159, 0.189). The probability of at least one event in a year lasting for 11 days above the one year level is given as 0.001 ($1 \times 10^{-4}$, 0.002).

As in Section 5.5 we also estimate whether heatwave events will become more severe with an increase in global mean temperature. We generate a distribution of the severities, $S \mid M > v_1$, and quantiles of this distribution can be estimated. Below, average estimates for the GCM ensemble are reported again with the range given in parentheses. At the 2006 global mean temperature level, the median of the distribution $S \mid M > v_1$ is given as 2.2ºC (1.4, 2.8). An increase of 1ºC in the covariate leads to a change in the median severity of 0ºC ($-0.3$, 0.5) and a 5ºC increase leads to a change in the median severity of 0.3ºC ($-1.1$, 4.2). The respective values for the 99th quantile are 5.9ºC (3.7, 7.7) with a change of 0ºC ($-0.9$, 1.4) and 0.8ºC ($-3.1$, 11.4). It is noted that 7 of the 13 members of our ensemble show a decrease in severity with an increase in the mean global temperature, the average across the ensemble is skewed

by the CCSM4 GCM that shows a very large increase in severity.

## 5.7 Discussion and Conclusion

The results in Sections 5.5 and 5.6 show that an increase in the global mean temperature due to climate change is likely to change the behaviour of heatwaves. This change has manifested itself in a significant increase in marginal quantities such as the return level as opposed to increases in the severity and duration that are driven by the dependence structure. One explanation behind this pattern could lie with the data used. Each GCM used has been forced with forcings equivalent to the RCP8.5 future emissions scenario, a scenario that is based upon 'business as usual' with slow development of renewable energy and increased use of fossil fuels. Figure 5.6.1 shows that each GCM has a distinct upward trend, borne out by the increase in return levels with increased global mean temperature. With such a large increase, any changes in the dependence structure might be getting hidden. Also, the use of GCM data might be a barrier to obtaining any significant results for what might be subtle changes in the dependence structure. For further work it would be interesting to investigate data for different climate scenarios and on different spatial scales to see if changes in the dependence structure become more important.

The assumption that the temperature time-series follows a first order Markov process has been made to permit the modelling process outlined in the paper. In Chapter 3 we suggested that using such a model might ignore subtleties in higher order dependence for the extreme temperature process. Making an assumption of higher order Markov processes has not been considered in this paper, but the extension to higher order Markov processes is given in Chapter 4 and this could be extended to incorporate covariate structure. Alternatively no Markov structure assumptions could be made, e.g. as in Eastoe and Tawn (2012), but this comes at the cost of large num-

bers of parameters and a high dimensional non-parametric distribution $G$ to estimate.

In this paper global mean temperature has been used as a covariate that indicates a change in the climate; a choice that is often used in reports such as IPCC (2000). Victor and Kennel (2014) suggest that the global mean temperature alone might not be the best way of measuring the level of climate change and put forward a set of measures that include greenhouse gas concentrations and ocean heat content. We note here that the framework developed in the paper could be extended to incorporate any such covariates of interest.

# Chapter 6

# Modelling the effect of ENSO on extreme temperatures over Australia

## 6.1 Introduction

The 2009 heatwave event was one of the most extreme to hit south-eastern Australia. Melbourne recorded its highest temperature, since records began in 1859, at 46.4ºC and Adelaide its third highest temperature over the same observational period at 45.7ºC. In total there were 374 heat related deaths with over 2,000 people treated for heat related illness. A particular challenge when modelling any environmental process across Australia is the spatial distribution of the population and agricultural activity across the country. Four of the five largest cities are located on the coast in the south-eastern region and most agriculture occurs in the south-eastern region. A hot event occurring over this region will lead to increased mortality and potential economic losses. As such, for mitigation purposes, it is necessary to be able to give accurate estimates of the risk posed by high temperatures over specific regions of interest. Extreme value theory provides a statistical framework for modelling rare events.

To model this problem sufficiently using extreme value statistics we require not only a univariate extreme value model that focuses on very high temperatures, but also a flexible model that accurately captures the spatial dependence between different sites.

There is much interest in how certain large-scale climatic phenomena will affect extreme events; both currently and under future climate change. One particular phenomenon known to affect the climate of Australia is the El Niño-Southern Oscillation (ENSO). It is a large-scale naturally occurring fluctuation in sea surface temperatures (SSTs) in the equatorial Pacific. Two limiting cases, corresponding to higher and lower SSTs in the equatorial Pacific Ocean, are called El Niño and La Niña respectively. During El Niño conditions, weaker easterly trade winds blowing across the Pacific can cause warm surface water to flow eastwards. This leads to increased convection in the central Pacific and reduces the amount of precipitation over Australia and other countries in southern Asia. In contrast, during La Niña conditions stronger trade winds blow warmer surface water to the west Pacific and cooler SSTs are observed in central and eastern Pacific regions (Wang and Picaut, 2004).

There is a clear consensus that large-scale climatic modes have an effect on the temperatures observed over Australia (Kenyon and Hegerl, 2008). El Niño conditions will lead to increased temperatures over eastern and northern regions of Australia whereas during La Niña conditions the opposite will be true. Strong El Niño conditions do not guarantee higher temperatures and patterns are not uniform across space. The 2009 heatwave event over south-eastern Australia occurred during a moderate La Niña event. The event covered a large area and as such had a great impact leading to record temperatures in certain places; this was not a uniform pattern across the whole of Australia with some regions affected by only moderate heat.

Many studies have attempted to quantify the effect of ENSO and other large scale

climatic modes of variability on temperatures and precipitation. The effect of ENSO on mean global temperatures has been well studied, see Kenyon and Hegerl (2008). The impact of ENSO on temperature extremes is less well studied. From a global perspective, Kenyon and Hegerl (2008) and Alexander et al. (2009) show that ENSO has a significant effect on temperature extremes around the Pacific Rim and over the US and also note that many other large climatic nodes (e.g. Pacific Decadal Oscillation and Northern Atlantic Oscillation) have a significant effect on extreme temperatures in different parts of the globe. However, no explicit modelling using statistical extreme value methods is undertaken, with most results being empirical and only related to observed levels.

Looking at Australia specifically, Perkins and Alexander (2013) review the occurrence of heatwaves for a selection of different heatwave indices. Min et al. (2013) use extreme value theory to estimate the effects of ENSO, the Indian Ocean Dipole and the Southern Annular Mode on seasonal temperatures over Australia. They use a GEV distribution with covariates in the location and scale parameters. Alexander and Arblaster (2009) analyse the change in different climatic extremes over Australia, including extreme temperatures, alluding to the potential effect that ENSO could have on these climatic extremes. None of these papers explicitly model spatial dependence, deriving spatial patterns by mapping univariate results obtained at multiple sites. Therefore, these approaches cannot be used to estimate the probability of heatwave events occurring at multiple sites over space.

The aim of this study is to develop a better understanding of whether ENSO has an effect on extreme temperatures over Australia. As such we analyse the effect of ENSO not only on temperatures at singular sites but also on the spatial extent of extreme temperature events. To investigate this we have gridded daily maximum temperatures for the years 1957-2011. Heatwave events are most commonly charac-

terised in terms of hot temperatures that impact a certain area and last for many days. Figure 6.1.1 shows spatial cross-correlation functions for Australia when conditioning on the grid square that contains Melbourne. On the left the cross-correlation between Melbourne and all other grid boxes at time lag 0 is given and on the right we show the difference in the cross-correlation at the time lag at which the maximum cross-correlation is obtained and the cross-correlation at time lag 0. Figure 6.1.1 shows little difference between the spatial structure over south-eastern Australia when considering lags other than lag 0. For simplicity, during this analysis we ignore the effect of temporal dependence to focus on the impact of ENSO on spatial dependence. It is noted that this may lead to confidence intervals that are too tight and significant hypothesis tests may not be completely correct. To correct for this we could use sandwich estimators or use a block bootstrap procedure taking the whole spatial grid for blocks of time; see Chapter 7 for initial work on building a full space-time model.

We have found no studies that consider the effect of ENSO on the spatial extent of extreme temperatures, using a statistical framework based upon extreme value theory. Many statistical approaches explicitly analyse the spatial extent of environmental processes. The broad area of geostatistics provides the most basic approaches for spatial modelling, but these tend to focus on the main body of data and as such can lead to misleading results when analysing rare events such as extreme temperatures. From the area of extreme value theory the most popular approach to spatial modelling is to fit a max-stable process. A max-stable process arises as a limiting process derived by taking componentwise maxima pointwise over independent and identically distributed replicates. Early examples of these are given in Smith (1990), Coles (1993) and Schlather (2002); a thorough review of such techniques is given in Davison et al. (2012). Max-stable models are widely used since they are a flexible class of models which can be fitted at multiple sites and used to estimate values at other sites across a spatial field. However max-stable models are often computationally intensive
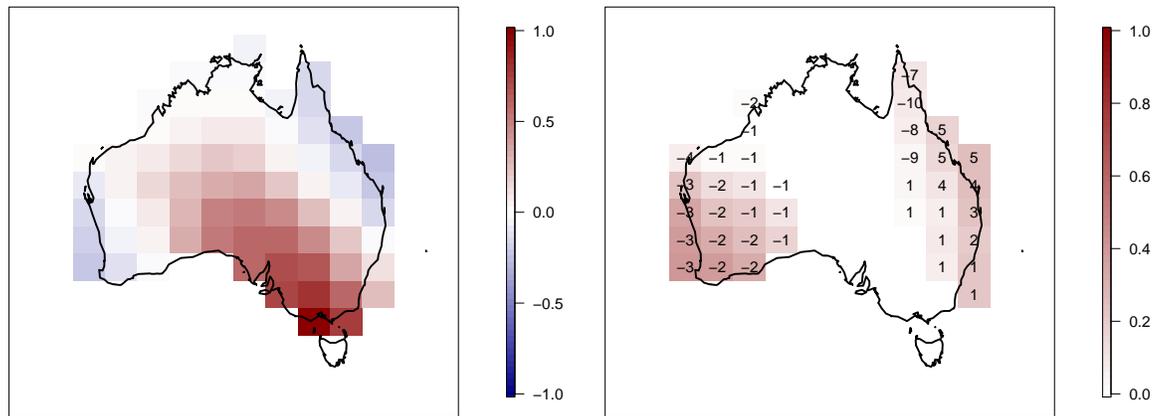
Figure 6.1.1: Conditioning on grid square that contains Melbourne: the lag 0 spatial cross-correlation function (left) and the difference between the maximum value of the cross-correlation function taking into account a 20 day window and value of the lag 0 cross-correlation function (right). Data are daily maximum temperatures across Australia for the years 1957-2011. Numbers in squares represent the lag value at which the maximum cross-correlation occurs; a blank square represents lag 0.

to fit (Davison et al., 2012) and difficult to conditionally simulate from (Wang and Stoev (2011); Dombry et al. (2013)).

One important concept in extreme value theory concerns the measure,

$$\chi = \lim_{p \to 1} P(F_2(Y_2) > p \mid F_1(Y_1) > p), \qquad (6.1.1)$$

where $Y_1$ and $Y_2$ are random variables with distribution functions $F_1$ and $F_2$ respectively. In the situation where $\chi > 0$, $Y_1$ and $Y_2$ are asymptotically dependent, i.e. the conditional probability of concurrent extremes in $Y_1$ and $Y_2$ has some non-zero probability in the limit. The variables $Y_1$ and $Y_2$ are asymptotically independent when $\chi = 0$. Max-stable processes are restricted to the case of asymptotic dependence, a restriction that can lead to incorrect inferences if the data exhibit asymptotic independence. Max-stable approaches model the behaviour of componentwise maxima across a space. Since they are often used to model annual maxima, it is possible that

when modelling using a max-stable approach the spatial pattern could be driven by separate events occurring within a given block of time (e.g. a year).

To accurately model extremal dependence we build a flexible multivariate model based upon the conditional extremes approach (Heffernan and Tawn, 2004), that fully takes into account spatial dependence on a lattice within the framework of extreme value theory. The conditional extremes model leads to a class of multivariate distributions that allow for asymptotic dependence and asymptotic independence. Since asymptotically independent forms are permitted, the conditional extremes approach covers the class of Gaussian processes (Ledford and Tawn, 1996); this is not the case for max-stable processes. The estimation of the dependence structure is driven by the observed data and does not require the more restrictive asymptotic dependence class to be chosen in advance. Most importantly the conditional extremes framework permits the estimation of not only extremes at singular sites, but also how ENSO affects the spatial extent of a hot event.

To analyse the effect of ENSO on extreme temperatures we shall estimate not only the change in return levels and other marginal quantities at singular sites, but also introduce a collection of new spatial risk measures. At singular sites many measures exist to quantify the effect of heat on mortality and other factors; see Alexander and Arblaster (2009) and Chapter 3. Let $\mathbf{Y} = (Y_1, \ldots, Y_l)$ be the daily maximum temperatures at $l$ sites, often we denote this as belonging to the set of all sites $S$ such that $|S| = l$. A commonly used measure is the extremal dependence measure (Coles et al., 1999), the sub-asymptotic form of equation (6.1.1), which is given as

$$\chi_{s'|s}(p) = \mathrm{P}\left(F_{s'}(Y_{s'}) > p \mid F_s(Y_s) > p\right), \tag{6.1.2}$$

where $s$ and $s'$ are two different sites from the set $S$ and $p$ is some high critical level, i.e. $p$ is close to one, often taken to be the non-exceedance probability associated to a critical return level. This measure only describes the dependence between pairs of

sites and does not represent an adequate measure of the spatial risk. We find that a summary measure, often used in the analysis of droughts, called the severity-area-frequency (SAF) curve (Henriques and Santos, 1999) provides a more informative measure for spatial risk. To construct a SAF curve over the region $S$, elements of $(F_1(Y_1), \ldots, F_l(Y_l))$ are ordered from largest to smallest. Defining a set of ordered random variables $\left(F_{(1)}\left(Y_{(1)}\right), \ldots, F_{(l)}\left(Y_{(l)}\right)\right)$ where $F_{(1)}\left(Y_{(1)}\right) \geq \cdots \geq F_{(l)}\left(Y_{(l)}\right)$, the SAF curve $\{j, \gamma_j; j = 1, \ldots, l\}$ is given by

$$\gamma_j = \frac{1}{j} \sum_{i=1}^{j} \left[1 - F_{(i)}\left(Y_{(i)}\right)\right]^{-1}. \tag{6.1.3}$$

The measure permits spatial information to be compressed into a single curve that is easily interpretable by climate scientists. Broadly, $\gamma_j$ gives the average marginal return period of an event at the $j$ worst affected sites. Figure 6.1.2 shows these curves for the day with the most extreme temperature during the 2009 heatwave event, with marginal and dependence structures both fixed and changing with ENSO. By fixing the maximum value at the observed value we have simulated replicate days under the conditional extremes model and a model that is restricted to asymptotic dependence, showing the mean and 95% confidence intervals for $\gamma_j$; for details of these models see Section 6.3. The model that allows for asymptotic independence provides a better fit to the observed curve, especially when using a non-stationary model to account for the effect of ENSO; for more details see Section 6.6. This highlights the need to account for asymptotic independence and ENSO in the spatial dependence structure.

Measures closely linked to equation (6.1.2) are also introduced to better understand other features of spatial dependence. One important measure is the expected number of concurrently extreme sites, i.e. the expected number of sites in a region, $R \subseteq S$, affected by an extreme event given that at least one site in the region $R$ is extreme. This type of measure does not require the definition of a particular conditioning site and as such is not as restrictive as equation (6.1.2).
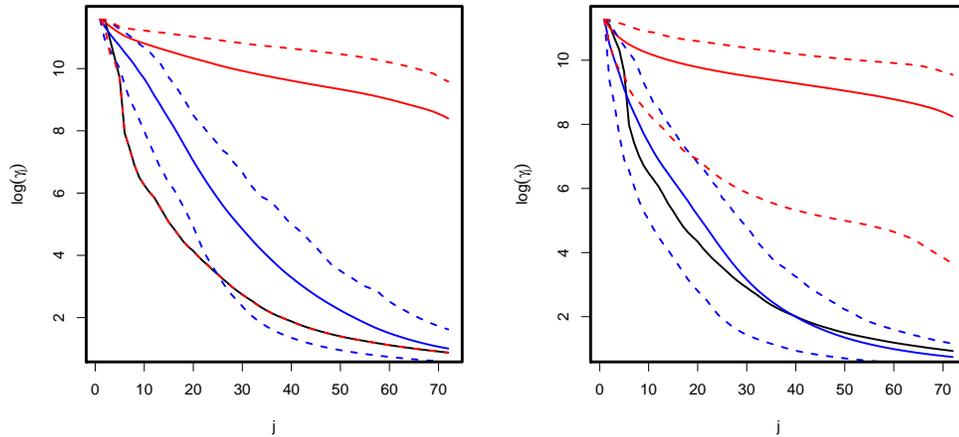
Figure 6.1.2: Severity-area-frequency (SAF) curves for hot day during the 2009 heatwave event on a log scale; observed empirical curve (black), $\mathbb{E}(\gamma_j)$ under conditional extremes model (blue solid) and $\mathbb{E}(\gamma_j)$ under asymptotically dependent model (red solid) using stationary model (left) and for the non-stationary model with covariate fixed at level observed on the given day (right). Dotted lines are 95% confidence intervals obtained from 10000 repeated simulations from each respective model.

By analysing this suite of measures we aim to explore the spatial extent of temperature extremes across Australia and see how the measures alter under a change in ENSO conditions. We also aim to test the validity of our approach by comparing observations from the record breaking heatwave event in 2009 to simulations of hot days generated by our model, thus demonstrating that our model can capture these events accurately. We then illustrate how our approach can be used to estimate extremal features for rarer events than have been previously observed.

In Section 6.2 we introduce the gridded daily maximum temperature data set we are going to use along with the ENSO covariate. Section 6.3 presents the models for the margins and dependence structure. A selection of measures for assessing spatial risk are developed in Section 6.4. In Section 6.5 an approach for simulating spatial fields

using the conditional extremes model is given. Results regarding the marginal and dependence parameters are provided in Section 6.6 along with estimates of important extremal measures. Finally, discussion and conclusions are given in Section 6.7.

## 6.2   Data

Daily maximum near-surface air temperatures for Australia are taken from HadGHCND, a global gridded dataset (`http://hadobs.metoffice.com/hadghcnd/`) of quality-controlled station observations compiled by the U.S. National Climatic Data Center (Caesar et al., 2006). An angular distance weighting technique is used to interpolate observed anomalies onto a 2.5º latitude by 3.75º longitude grid which results in 72 boxes covering Australia and spanning 1957-2011. Whilst this is a relatively coarse resolution heatwaves are large meteorological phenomena and surface air temperatures have long correlation length scales, for which Caesar et al. (2006) found values of between 700km and 1400km for the 0ºS to 30ºS latitude band. Avila et al. (2015) find for Austraian surface air temperatures their extremal characteristics and their correlations with ENSO are preserved across a range of gridding resolutions from 0.25º to 2.5º. The use of such global datasets also facilitates any future comparison with other regions. Hot days are most likely to occur in summer months, here defined as the 90 day period from December to February (91 day period for a leap year); these three month periods are extracted from each year. No missing data values exist within the summer months of the years for which the data are provided.

To measure the effect of ENSO the Niño3.4 index is used. This is a measurement of the monthly SST anomaly, with respect to the average SST for 1981-2010, in a region bounded by 5ºN to 5ºS and 170ºW to 120ºW. Other ways of measuring ENSO variability are available; for example the Southern Oscillation Index which is based on atmospheric changes as opposed to changes in SSTs (Jones and Trewin, 2000). How-

ever, Niño3.4 is commonly used to characterise ENSO (Kenyon and Hegerl, 2008). Large positive values of this index indicate El Niño events, whereas large negative values correspond to La Niña events. In this paper values of $+1^{\circ}$C and $-1^{\circ}$C are used to define El Niño and La Niña events respectively. Our framework permits estimates for any value of Niño3.4 that is of interest. It is noted that ENSO is a coupled atmosphere-ocean phenomenon which has a timescale of 10-12 months; this behaviour is well captured by a monthly index.

## 6.3 Modelling extreme values

Our strategy for modelling the probabilistic behaviour of extreme temperatures is two-fold. Firstly, we model the marginal structure using a threshold based approach at each site separately. Different approaches are available to model the effect of a covariate on tail behaviour (Davison and Smith (1990); Northrop and Jonathan (2011)). Here pre-processing (Eastoe and Tawn, 2009) is used to model the effect of the ENSO covariate $g_t$, which varies with time but not space, on temperatures at each separate site. The pre-processing step removes covariate effects from the body of the distribution and then residual influences of the covariates on the tails are accounted for using the methods of Davison and Smith (1990). This approach has close parallels with Northrop and Jonathan (2011) since the threshold for the extreme value modelling is derived to be covariate dependent. Once the marginal structure has been modelled, we transform the data from each site onto common margins and model the extremal dependence structure using the conditional extremes approach. It is noted that greater precision could be achieved by using the spatial structure at this stage as opposed to modelling each site separately, but we do not investigate this here.

## 6.3.1 Marginal modelling

To understand the effect that ENSO phase is having on margins we use pre-processing. Specifically, we choose to fit the location-scale model in the margins, i.e. for daily maximum temperature $Y_{s,t}$ at location $s$ and time $t$ we have

$$Y_{s,t} = \psi_s(g_t) + \tau_s(g_t)Y_{s,t}^{\dagger},$$

for $t = 1, \ldots, n$ and $s \in S$, where $(\psi_s(g_t), \tau_s(g_t))$ are the location-scale parameters, $g_t$ is a time-varying covariate and $Y_{s,t}^{\dagger}$ is the residual. In this paper all covariates are included linearly with an appropriate link function such that

$$\psi_s(g_t) = {}_0\psi_s + {}_1\psi_s g_t \qquad \log \tau_s(g_t) = {}_0\tau_s + {}_1\tau_s g_t,$$

with parameters ${}_0\psi_s$, ${}_1\psi_s$, ${}_0\tau_s$ and ${}_1\tau_s$ each in $\mathbb{R}$. In practice the location-scale model may not completely capture all the non-stationarity in the extremes and such a GPD tail model is fitted, above a high threshold $u_s^{\dagger}$, to the margins of the residuals $Y_{s,t}^{\dagger}$ such that

$$F_s(y) = \begin{cases} 1 - \lambda_{s,u_s^{\dagger}} \left[ 1 + \xi_s(g_t)\left(y - u_s^{\dagger}\right)/\sigma_{s,u_s^{\dagger}}(g_t) \right]_+^{-1/\xi_s(g_t)} & \text{if} \quad y \geq u_s^{\dagger} \\ \tilde{F}_s(y) & \text{if} \quad y < u_s^{\dagger}, \end{cases} \tag{6.3.1}$$

where $(\sigma_{s,u_s^{\dagger}}(g_t), \xi_s(g_t))$ are scale and shape parameters that depend on the covariate, $\lambda_{s,u_s^{\dagger}} = 1 - \tilde{F}_s(u_s^{\dagger})$ and $\tilde{F}_s(y)$ is the empirical cumulative distribution function of $\{Y_{s,t}^{\dagger}\}_{t=1}^n$ at site $s$. In this paper the covariate is included into the GPD scale parameter but the shape parameter is assumed to be constant such that

$$\log \sigma_{s,u_s^{\dagger}}(g_t) = {}_0\sigma_{s,u_s^{\dagger}} + {}_1\sigma_{s,u_s^{\dagger}} g_t \qquad \xi_s(g_t) = \xi_s,$$

with parameters ${}_0\sigma_{s,u_s^{\dagger}}$, ${}_1\sigma_{s,u_s^{\dagger}}$ and $\xi_s$ each in $\mathbb{R}$. It is assumed that temporal non-stationarity in the body of the distribution is accounted for by the original location-scale transform and as such the stationary empirical distribution function is appropriate for values that fall below the modelling threshold $u_s^{\dagger}$.

## 6.3.2 Dependence modelling

The conditional extremes method of Heffernan and Tawn (2004) is used here to model extremal dependence. Using the methods outlined in Section 6.3.1 data are transformed onto common margins. The transformation onto common margins simplifies the estimation of extremal dependence quantities. This is especially important in the spatial problems encountered here since we are interested whether different sites have rare values simultaneously irrespective of the value of these rare values on the original temperature scale. Modelling using the conditional extremes approach is simplified if the choice of common margin is assumed to be Laplace distributed (Keef et al., 2013), i.e.

$$
X_{s,t} = \begin{cases} \log\left\{2F_s(Y_{s,t}^\dagger)\right\} & \text{if} \quad F_s(Y_{s,t}^\dagger) < 1/2 \\ -\log\left\{2\left[1 - F_s(Y_{s,t}^\dagger)\right]\right\} & \text{if} \quad F_s(Y_{s,t}^\dagger) \geq 1/2, \end{cases}
$$

where $F_s$ is given in equation (6.3.1), as the margins have exponential upper and lower tails which ensures models for positive and negative dependence are symmetric.

Let $\mathbf{X}_t = (X_{1,t}, \ldots, X_{l,t})$, where $l$ is the number of sites in the region $S$, and define $\mathbf{X}_{-s,t}$ as all the components of the vector $\mathbf{X}_t$ without $X_{s,t}$, i.e. $\mathbf{X}_{s,t} = (X_{s,t}, \mathbf{X}_{-s,t})$ and in what follows all vector calculations are to be interpreted as componentwise. The aim is to model the distribution of $\mathbf{X}_{-s,t}$ given that $X_{s,t}$ exceeds some high threshold $u$. It is necessary that the conditional distribution $\mathrm{P}\left\{\mathbf{X}_{-s,t} \leq \mathbf{x}_{-s,t} \mid X_{s,t} = x_{s,t}\right\}$ is non-degenerate as $x_{s,t} \to \infty$ and hence normalising sequences are required to ensure $\mathbf{x}_{-s,t}$ changes appropriately with $x_{s,t}$. Heffernan and Tawn (2004), Heffernan and Resnick (2007) and Keef et al. (2013) show that under broad conditions there exists $\boldsymbol{\alpha}_{-s,t} = (\alpha_{1|s,t}, \ldots, \alpha_{s-1|s,t}, \alpha_{s+1|s,t}, \ldots, \alpha_{l|s,t}) \in [-1,1]^{l-1}$ and $\boldsymbol{\beta}_{-s,t} = (\beta_{1|s,t}, \ldots, \beta_{s-1|s,t}, \beta_{s+1|s,t}, \ldots, \beta_{l|s,t}) \in (-\infty, 1)^{l-1}$ such that for $\mathbf{z} \in \mathbb{R}^{l-1}$ and $x > 0$

$$
\mathrm{P}\left(\frac{\mathbf{X}_{-s,t} - \boldsymbol{\alpha}_{-s,t} X_{s,t}}{X_{s,t}^{\boldsymbol{\beta}_{-s,t}}} \leq \mathbf{z}, X_{s,t} - u > x \mid X_{s,t} > u\right) \to G_{-s,t}(\mathbf{z}) \exp(-x), \quad (6.3.2)
$$

as $u \to \infty$ where $G_{-s,t}$ is a time-varying $(l-1)$-dimensional distribution function, non-degenerate in each margin, i.e. for $j \in S\backslash\{s\}$ the $j$th margin $G_{-s,t}^{(j)}$ of $G_{-s,t}$ is non-degenerate. Different values of the dependence parameters $\boldsymbol{\alpha}_{-s,t}$ and $\boldsymbol{\beta}_{-s,t}$ arise for different types of tail dependence. If the variables $(X_{s,t}, X_{j,t})$ are independent, $\alpha_{j|s,t} = \beta_{j|s,t} = 0$ and $G_{-s,t}^{(j)}$ is the Laplace distribution function, for $j \in S\backslash\{s\}$. On the other hand for $(X_{s,t}, X_{j,t})$, $\alpha_{j|s,t} = 1$ and $\beta_{j|s,t} = 0$, for $j \in S\backslash\{s\}$, corresponds to the situation of asymptotic dependence, $-1 \leq \alpha_{j|s,t} \leq 0$ to negative extremal dependence and $0 < \alpha_{j|s,t} < 1$ or $\alpha_{j|s,t} = 0$ and $\beta_{j|s,t} > 0$ corresponds to asymptotic independence with positive extremal dependence. Here, a time-varying covariate $g_t$ is introduced into the dependence parameters such that

$$\tanh^{-1}[\boldsymbol{\alpha}_{-s,t}] = {}_0\boldsymbol{\alpha}_{-s} + {}_1\boldsymbol{\alpha}_{-s}g_t \qquad \tanh^{-1}[\boldsymbol{\beta}_{-s,t}] = {}_0\boldsymbol{\beta}_{-s} + {}_1\boldsymbol{\beta}_{-s}g_t, \qquad (6.3.3)$$

with parameters ${}_0\boldsymbol{\alpha}_{-s}$, ${}_1\boldsymbol{\alpha}_{-s}$, ${}_0\boldsymbol{\beta}_{-s}$ and ${}_1\boldsymbol{\beta}_{-s}$ each in $\mathbb{R}^{l-1}$. The inverse tanh link function is used to ensure the parameters $\boldsymbol{\alpha}_{-s,t}$ and $\boldsymbol{\beta}_{-s,t}$ are restricted to the range $[-1,1]^{l-1}$. The restriction on $\boldsymbol{\beta}_{-s,t}$ is satisfactory since in practice it is very unlikely that $\beta_{j|s,t} < -1$, for $j \in S\backslash\{s\}$, as this corresponds to $\mathbf{X}_{-s,t} - \boldsymbol{\alpha}_{-s,t}X_{s,t}$ tending rapidly to zero for large $X_{s,t}$, i.e. $\mathbf{X}_{-s,t}$ is essentially deterministic given large $X_{s,t}$.

Modelling using the conditional extremes approach requires the assumption that the limiting form of equation (6.3.2) holds exactly for all values of $X_{s,t} > u$ given that $u$ is a sufficiently high threshold, from now on called the modelling threshold. From equation (6.3.2) we have our model for $X_{s,t} > u$ that

$$\mathbf{X}_{-s,t} = \boldsymbol{\alpha}_{-s,t}X_{s,t} + X_{s,t}^{\boldsymbol{\beta}_{-s,t}}\mathbf{Z}_{-s,t},$$

where $\mathbf{Z}_{-s,t} = (Z_{1|s,t}, \ldots, Z_{s-1|s,t}, Z_{s+1|s,t}, \ldots, Z_{l|s,t})$ is a random variable with distribution function $G_{-s,t}$ that is independent of $X_{s,t}$.

The multivariate distribution $G_{-s,t}$ does not take any simple parametric form, which motivates the inclusion of a false working assumption of Gaussianity as in Keef et al.

(2013) solely for the estimation of $\alpha_{j|s,t}$ and $\beta_{j|s,t}$ with $j \neq s$.   That is $Z_{j|s,t} \sim N(\mu_{j|s,t}, \theta_{j|s,t}^2)$ and as such for each $j \in S \backslash \{s\}$

$$X_{j,t} \mid \{X_{s,t} = x\} \sim N\left(\alpha_{j|s,t}x + \mu_{j|s,t}x^{\beta_{j|s,t}}, \theta_{j|s,t}^2 x^{2\beta_{j|s,t}}\right) \qquad \text{for} \quad x > u.$$

The working assumption permits the estimation of parameters $(\alpha_{j|s,t}, \beta_{j|s,t}, \mu_{j|s,t}, \theta_{j|s,t})$ by standard likelihood approaches.   Each element of $\boldsymbol{\alpha}_{-s,t}$ and $\boldsymbol{\beta}_{-s,t}$ is estimated pairwise for a particular $s \in S$. When considering covariate effects, covariates will be included in the nuisance parameters such that

$$\boldsymbol{\mu}_{-s,t} = {}_0\boldsymbol{\mu}_{-s} + {}_1\boldsymbol{\mu}_{-s}g_t \qquad \log \boldsymbol{\theta}_{-s,t} = {}_0\boldsymbol{\theta}_{-s} + {}_1\boldsymbol{\theta}_{-s}g_t, \qquad (6.3.4)$$

where

$$\boldsymbol{\mu}_{-s,t} = (\mu_{1|s,t}, \ldots, \mu_{s-1|s,t}, \mu_{s+1|s,t}, \ldots, \mu_{l|s,t})$$
$$\boldsymbol{\theta}_{-s,t} = (\theta_{1|s,t}, \ldots, \theta_{s-1|s,t}, \theta_{s+1|s,t}, \ldots, \theta_{l|s,t}),$$

with parameters ${}_0\boldsymbol{\mu}_{-s}$, ${}_1\boldsymbol{\mu}_{-s}$, ${}_0\boldsymbol{\theta}_{-s}$ and ${}_1\boldsymbol{\theta}_{-s}$ each in $\mathbb{R}^{l-1}$. At this stage the Gaussian assumption is discarded and a non-parametric estimate of the distribution for $\mathbf{Z}_{-s,t}$ is formed.   To ensure stationarity we define a new non time-varying multivariate distribution $G_{-s}$ defined as

$$G_{-s}(\mathbf{z}) = G_{-s,t}\left(\frac{\mathbf{z} - \boldsymbol{\mu}_{-s,t}}{\boldsymbol{\theta}_{-s,t}}\right).$$

Specifically, where $n_u$ is the number data points exceeding the threshold $u$, let $t_1, \ldots, t_{n_u}$ be the indices of $t = 1, \ldots, n$ where $x_{s,t} > u$ then let

$$\hat{\mathbf{z}}_{-s,i} = \frac{\mathbf{x}_{-s,t_i} - \hat{\boldsymbol{\alpha}}_{-s,t_i}x_{s,t_i} - \hat{\boldsymbol{\mu}}_{-s,t_i}\left(x_{s,t_i}\right)^{\hat{\boldsymbol{\beta}}_{-s,t_i}}}{\hat{\boldsymbol{\theta}}_{-s,t_i}\left(x_{s,t_i}\right)^{\hat{\boldsymbol{\beta}}_{-s,t_i}}}, \qquad (6.3.5)$$

for $i = 1, \ldots, n_u$. In this way the empirical distribution of sample $\hat{\mathbf{z}}_{-s,i}$ provides a non-parametric estimate, $\tilde{G}_{-s}$, to the distribution function $G_{-s}$ for conditioning site $s$.

## 6.4 Summarising spatial dependence

To analyse the spatial behaviour of hot events, we require measures that can adequately capture spatial characteristics. In this section we set up the measures that are used for the rest of this chapter, most of which have not previously been used. By using a selection of different measures, we aim to fully characterise extremal dependence and any changes in spatial structure that may occur due to a change in ENSO. A common measure of extremal dependence is the sub-asymptotic extremal dependence measure $\chi_{s'|s}(p)$, given by equation (6.1.2), for two sites $s$ and $s'$ at a critical level $p$. If $s = s'$ then $\chi_{s'|s}(p) = 1$. In a spatial context, this measure is used by fixing $s$ at a conditioning site and estimating the quantity in equation (6.1.2) for all other sites $s' \in S$.

One issue with using $\chi_{s'|s}(p)$ is that the measure only estimates bivariate dependence and therefore does not give information about the occurrence of concurrent extremes at more than two sites at a time. To overcome this we propose a new measure of the expected number of grid boxes that exceed a critical level given that $Y_s$ exceeds the same critical level. Since we are only modelling spatial dependence the subscript $t$ is dropped from our notation from now on. Define the distribution function $F_j^P$, for the $j$th margin, that incorporates all steps of the pre-processing outlined in Section 6.3.1. Let the region of interest be denoted $R$, $R \subseteq S$ and $N_R(p) = \# \left\{ j \in R : F_j^P(Y_j) > p \right\}$ gives the number of variables that concurrently exceed the probability level $p$, where $p$ is a critical level. We are interested in the distribution $N_R(p) \mid F_s^P(Y_s) > p$ for some conditioning site $s$. A convenient summary measure of this conditional distribution is given by the expected number of sites in the region $R$ that exceed $p$ given that $F_s^P(Y_s) > p$, i.e.

$$\phi_{R|s}(p) = \mathbb{E}(N_R(p) \mid F_s^P(Y_s) > p). \tag{6.4.1}$$

The measure defined in equation (6.4.1) better takes into account the spatial structure

of extreme temperature events than $\chi_{s'|s}(p)$. However, such a measure still requires a particular conditioning site to be defined prior to estimation. In practice assuming that a hot event must strike a particular site is restrictive. We propose a new measure of the probability of an exceedance of a critical level in a region $R'$ given that there is an exceedance somewhere within a critical region $R$, i.e.

$$\omega_{R'|R}(p) = \mathrm{P}(N_{R'}(p) \geq 1 \mid N_R(p) \geq 1),$$

for some region $R' \subseteq S$. Commonly we are interested in sets of the form $R' \subset R$; but other forms such as $R' \cap R = \emptyset$ can be considered if these are of interest. A special case of this measure occurs where $R' = \{s\}$ which gives the probability of an exceedance at site $s$ given that there is an exceedance somewhere in the region $R$.

We propose a different set of new measures that are based upon severity-area-frequency (SAF) curves (Henriques and Santos, 1999). These curves originate from hydrology and are used to compress complicated spatial information into a curve that can be used to estimate the expected severity of an event covering a particular spatial extent. In the hydrology context, SAF curves are given on the probability scale, i.e. $F_s^P(Y_s)$, and a slightly modified version of this approach is used here. The SAF curve $\gamma_j$, given by equation (6.1.3), is the average return period of an event over the $j$ worst affected sites within $S$. Similar curves can be constructed on the original temperature margins, i.e. for $Y_s$, but these are not used here. SAF curves can be used for model checking and as a validation tool. Here, we look to extend their usage by introducing a new measure

$$\rho_j = \mathrm{P}\left(\gamma_j > \gamma_j^{\mathrm{obs}}\right),$$

for $j = 1, \ldots, l$, where $\gamma_j^{\mathrm{obs}}$ is the SAF curve derived for an observed event such as 2009 heatwave event. This measure gives the probability of the average return period for the $j$ largest values being higher than for an observed event. This measure

provides vital information about the probability of observing an extreme temperature event more severe than previous damaging observed events.

## 6.5  Simulating spatial fields

To estimate the measures of spatial dependence introduced in Section 6.4 we need to be able to simulate spatial gridded fields from our fitted model outlined in Section 6.3. Heffernan and Tawn (2004) and Keef et al. (2013) give simulation schemes for the conditional extremes approach conditional upon an exceedance in the conditioning variable (site). These schemes are adequate to obtain estimates of $\chi_{s'|s}(p)$ and $\phi_{R|s}(p)$ and they form the basis of the simulation scheme outlined here. Estimation of measures that condition upon an exceedance within a region require a more involved algorithm for generating simulated spatial gridded fields. The use of SAF curves for model validation also requires conditions on the value that the peak value of an event takes.

The simplest method for simulating spatial fields is used when estimating $\chi_{s'|s}(p)$ and $\phi_{R|s}(p)$. When estimating these measures we fix a site of interest $s$. To simulate a spatial field, for a particular covariate of interest $g_t$ and site $s$ being the maximum, we follow the following steps:

1. Sample $\tilde{\mathbf{z}}^*_{-s}$ from $\tilde{G}_{-s}$, i.e. the empirical distribution of the sample in equation (6.3.5).

2. Obtain $\mathbf{z}^*_{-s} = \boldsymbol{\mu}^*_{-s} + \boldsymbol{\theta}^*_{-s}\tilde{\mathbf{z}}^*_{-s}$ where $\boldsymbol{\mu}^*_{-s} = \boldsymbol{\mu}_{-s,t}$ and $\boldsymbol{\theta}^*_{-s} = \boldsymbol{\theta}_{-s,t}$ from equation (6.3.4).

3. Simulate exceedance $X^*_s > v$ as the sum of $v$ and a unit Exponential random variable.

4. Obtain spatial field $\mathbf{X}^*_{-s} = \boldsymbol{\alpha}^*_{-s} X^*_s + (X^*_s)^{\boldsymbol{\beta}^*_{-s}} \mathbf{z}^*_{-s}$, where $\boldsymbol{\alpha}^*_{-s} = \boldsymbol{\alpha}_{-s,t}$ and $\boldsymbol{\beta}^*_{-s} = \boldsymbol{\beta}_{-s,t}$ from equation (6.3.3).

5. If $\max\left(\mathbf{X}^*_{-s}\right) > X^*_s$ reject spatial field $\mathbf{X}^*_{-s}$ and repeat steps 1-4 for the selected $s$ until the simulated field is not rejected.

To estimate extremal measures of interest, steps 1-4 are repeated $m$ times to obtain $m$ spatial fields with the desired extremal dependence characteristics, $\mathbf{X}^*_1, \ldots, \mathbf{X}^*_m$, where $\mathbf{X}^*_i = (X^i_1, \ldots, X^i_l)$ for $i = 1, \ldots, m$, and all fields have $X_s > v$. For a site of interest $s' \in S$ we have

$$\hat{\chi}_{s'|s}(p) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}\left(X^i_{s'} > v_p\right),$$

where $\mathbb{I}(\cdot)$ is the indicator function and $v_p = -\log\{2(1-p)\}$ is the critical level on Laplace scale associated to the non-exceedance probability $p$. The measure $\phi_{R|s}(p)$ can also be estimated for a region $R$ by

$$\hat{\phi}_{R|s}(p) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j \in R} \mathbb{I}\left(X^i_j > v_p\right). \tag{6.5.1}$$

To estimate $\omega_{R'|R}(p)$ extensions to the simulation scheme outlined above are required since we are interested in events that are hot for at least one site over a region $R$, rather than just at a fixed site. Figure 6.5.1 illustrates an issue if we repeat the fixed site simulation scheme at each site separately; we tend to simulate too many points in the joint tail.

Our strategy for this simulation is as follows. We require an exceedance of $v_p$ by at least one site in $R$, i.e. $N_R(p) \geq 1$. We select a site that exceeds $v_p$ by picking it to be the largest value, i.e. $X_j = \max_{i \in R}(X_i)$. The probability that site $j$ is largest over $R$, given $N_R(p) > v_p$, varies with $j$ due to the changing dependence structure; we denote this probability by $q_j$, with

$$q_j = \mathrm{P}\left(X_j = \max_{i \in R}(X_i) \mid N_R(p) \geq 1\right). \tag{6.5.2}$$

Once this site has been selected, we then generate $\mathbf{X}_{-j}$ such that $X_j = \max_{i \in R}(X_i)$. Then we assess whether $\max_{i \in R'}(X_i) > v_p$. Repeated simulation enables us to derive a Monte Carlo estimate of $\omega_{R'|R}(p)$.

To estimate $q_j$ we could simply use an empirical estimate of equation (6.5.2). However, if $p$ is small this approach fails. A better approach is to use our fitted conditional model. Under the conditional extremes model we simulate $m_R$ exceedances of $v_p$, $X_j^i$ for $i = 1 \ldots, m_R$ with $X_j^i > v_p$, then use steps 1-4 to calculate the proportion of times, $\hat{q}_j$, that $X_j = \max_{i \in R}(X_i)$, i.e.

$$\hat{q}_j = \frac{1}{m_R} \sum_{i=1}^{m_R} \mathbb{I}\left(X_j^i > \max\left(\mathbf{X}_{-j}^i\right)\right) \qquad \text{for} \quad j \in R,$$

where $\mathbf{X}_{-j}^i = (X_1^i, \ldots, X_{j-1}^i, X_{j+1}^i, \ldots, X_l^i)$. To simulate the $i$th replicate event with the required structure, site $s^i \in R$ is selected with probability $\hat{q}_j$ and steps 1-5 above are applied with $s = s^i$. By repeating this simulation approach $m$ times to obtain $\mathbf{X}_1^*, \ldots, \mathbf{X}_m^*$, where $\mathbf{X}_i^* = (X_1^i, \ldots, X_l^i)$ for $i = 1, \ldots, m$, the measure $\omega_{R'|R}(p)$ is estimated in the same way as $\phi_{R|s}(p)$ in equation (6.5.1), i.e.

$$\hat{\omega}_{R'|R}(p) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j \in R} \mathbb{I}\left(X_j^i > v_p\right).$$

To use SAF curves for validation, we wish to simulate replicate events that have similar characteristics to a particular event of interest (e.g. the 2009 heatwave event). To do this we fix the maximum at the peak value observed during the event and fix the corresponding site that the maxima occurred and see whether out model can replicate similar behaviour to the observed event. The simulation scheme under a fixed maximum value and site follows steps 1-5 from the algorithm above but step 3 changes to $X_s^* = \eta$, where $\eta$ is the peak value of an event. This procedure is repeated $m$ times to generate multiple realisations of the spatial field which give $m$ different SAF curves $(\gamma_j^1, \ldots, \gamma_j^m)$ for $j = 1, \ldots, l$.
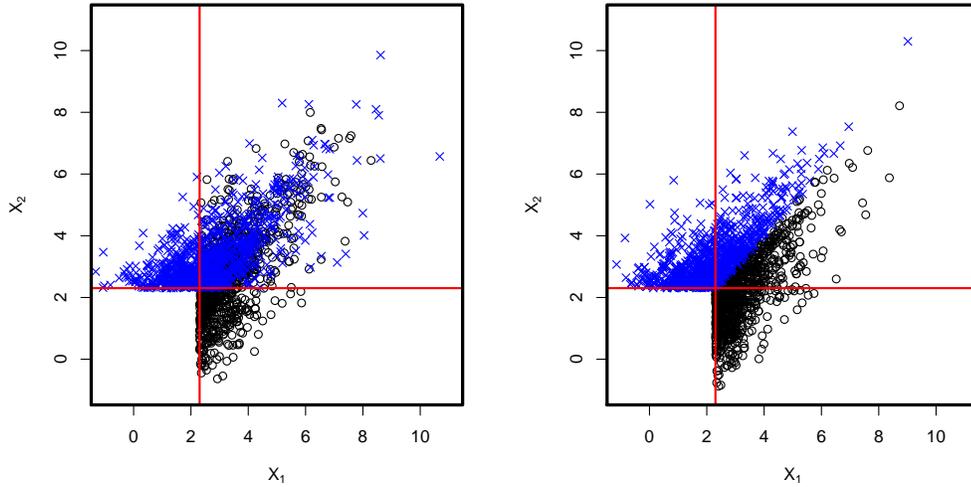
Figure 6.5.1: Bivariate data simulated using the conditional extremes approach without rejection step (left) and with rejection step (right) included to ensure that the joint extremal region is not over-sampled. Simulated data from conditional extremes method fitted to $X_2 \mid X_1 > u$ (black circles) and to $X_1 \mid X_2 > u$ (blue crosses), with critical levels at the 95th quantile (red lines).

When it comes to assessing the risk of an extreme temperature event it is necessary to generalise the restriction on the location and peak value of an event. For example, we may want to know about an event where the maximum takes a certain critical value but could occur at different sites across a region. We may also wish to analyse events that have a larger maximum value than a specified level. To address such issues we simulate the location of the maxima $M_R$ over $R$ given $M_R = \eta$ with probability

$$
\begin{aligned}
\nu_i(\eta) &= \frac{\mathrm{P}(\mathbf{X}_{-i} < \eta \mid X_i = \eta)\mathrm{P}(X_i = \eta)}{\sum_{j \in R} \mathrm{P}(\mathbf{X}_{-j} < \eta \mid X_j = \eta)\mathrm{P}(X_j = \eta)} \\
&= \frac{\mathrm{P}(\mathbf{X}_{-i} < \eta \mid X_i = \eta)}{\sum_{j \in R} \mathrm{P}(\mathbf{X}_{-j} < \eta \mid X_j = \eta)},
\end{aligned}
$$

for $i \in R$. Steps 1-5 are then followed with $s = i$. Estimates for $\nu_i(\eta)$, for $i \in R$, follow directly from the conditional extremes approach. In many situations, it might be more interesting to estimate the probability of an event with a maximum larger than a particular value $\eta$. Instead of fixing the maxima at $\eta$ we can simulate a

maximum as the sum of $\eta$ and an Exponential random variable with rate parameter 1.

## 6.6 Results

The extreme value framework built in Section 6.3 is now combined with the summary measures defined in Section 6.4 to evaluate the characteristics of hot days over Australia for the gridded observations introduced in Section 6.2. Firstly, pre-processing is applied to the original data to model the marginal structure and transform values onto consistent margins. Then the choice of the conditional extremes approach is validated by comparing against other extreme value approaches that do not account for asymptotic independence. Finally, the measures in Section 6.4 are estimated and variability between the spatial extent of hot events under El Niño and La Niña conditions is estimated. This culminates in estimating whether the framework can replicate similar events to the 2009 heatwave event over Australia and whether this event was more likely under the observed phase of ENSO.

### 6.6.1 Marginal structure

In Section 6.3.1 we outlined the pre-processing approach which is now used to estimate whether SSTs over the eastern tropical Pacific Ocean have an effect on marginal quantities such as the return level of an extreme event. The covariate used to summarise the effect of SST on temperatures is Niño3.4 as introduced in Section 6.2. At each separate site the effect of ENSO on all parameters within the pre-processing approach is assessed using likelihood ratio tests. The desire is to use the simplest model that we can whilst not ignoring any potential covariate effects. The pre-processing scheme takes the form of a location-scale transformation and then an adjustment of the extremes using a threshold exceedance model. Figure 6.6.1 gives plots of the pre-processing parameters, with shaded boxes indicating sites where the particular

parameter of interest does not exhibit any significant change with the ENSO covariate (at the 5% significance level). On the right hand plots of Figure 6.6.1 red boxes show an increase in parameter values with an increase in Niño3.4 from -1°C to +1°C. The top row gives estimates of the location parameters $(_0\psi_s,\ _1\psi_s)$; warmer temperatures are observed as expected in northern and central regions of Australia with cooler coastal areas. An increase in Niño3.4 (i.e. moving towards El Niño conditions) causes an increase in the location parameter over the most of Australia, with the largest increases in eastern and western regions. For the scale parameter $\tau_s(g_t)$, the largest changes seem to be over western regions where El Niño conditions reduce temperature variability.

For each parameter we investigate for how many grid boxes the covariate is significant using likelihood ratio tests for each site at the 5% significance level. A decision is then made as to whether the covariate effect is included in the final model. Figure 6.6.1 shows that out of a total of 72 grid boxes, 64 show a significant change in the location parameter with the ENSO covariate. This clear signal is not fully repeated by the scale parameter $\tau_s(g_t)$ which shows a significant change in 29 grid boxes out of 72. Although the result of the scale parameter is less significant we keep both covariate effects for all grid squares as we desire to have the same covariate structure incorporated in each parameter for all grid boxes.

Estimates of the GPD scale and shape parameters are given in the bottom row of Figures 6.6.1 and 6.6.2 respectively. Standard diagnostics (Coles, 2001) are run at each site separately which suggest the 90% quantile at each site is an appropriate threshold choice. As outlined in Section 6.3.1, the aim of this step is to take the approximately stationary time-series and ensure that the extremes are fully stationary. As observed in Figure 6.6.1, the previous pre-processing steps have taken account of location-scale non-stationarity in the body of the distribution at most grid boxes.
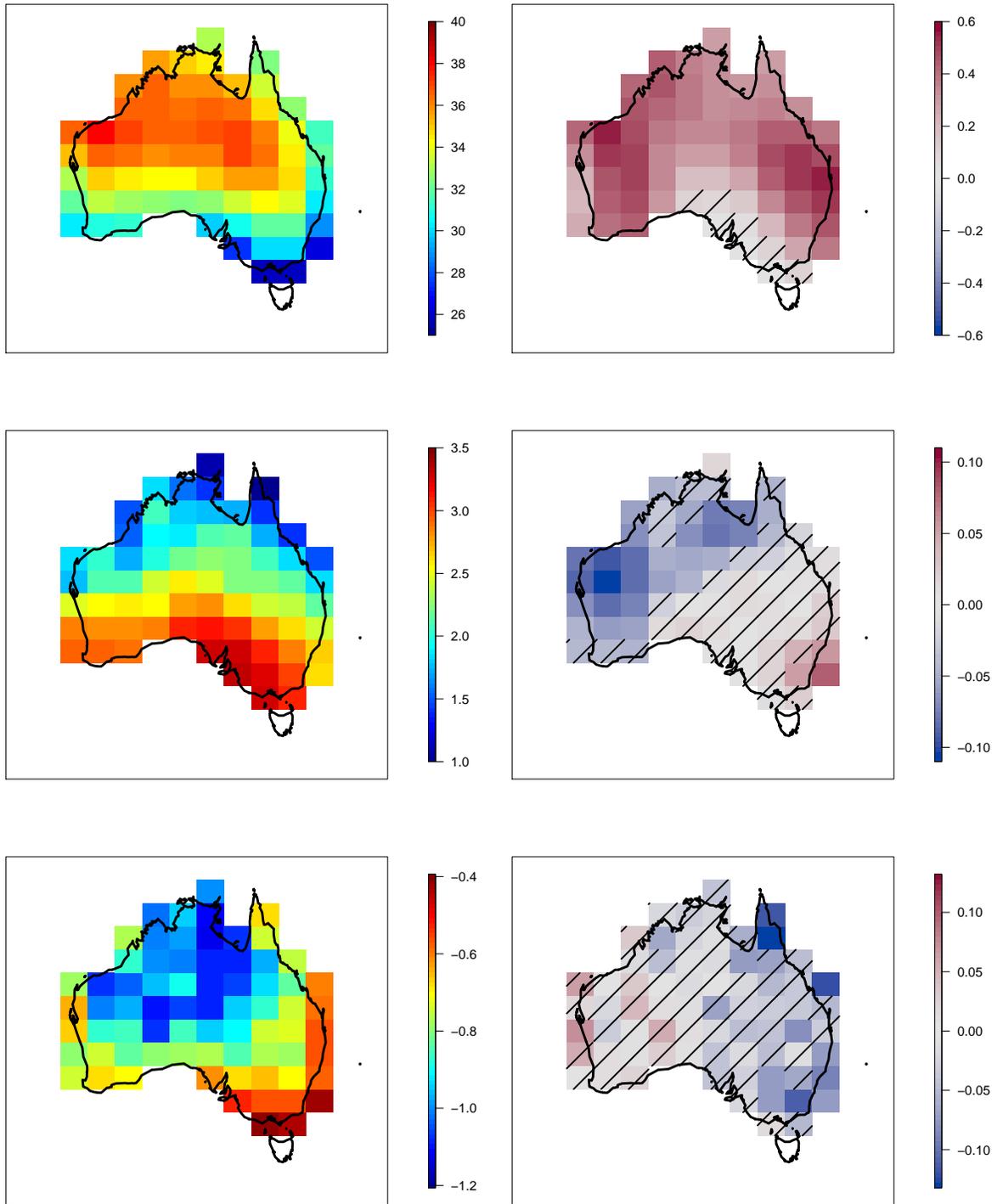
Figure 6.6.1: Estimates of pre-processing location and scale parameters $({}_0\psi_s, {}_1\psi_s)$ (top) and $({}_0\tau_s, {}_1\tau_s)$ (middle) and GPD scale parameters $({}_0\sigma_{s,u_s^\dagger}, {}_1\sigma_{s,u_s^\dagger})$ (bottom). Shaded squares correspond to boxes where the change with covariate is not significant at the 95% confidence level, tested using a likelihood ratio test.

However there are some grid-boxes on the coast that show a decrease in the dispersion of extreme values with an increase in the ENSO covariate. Since we are interested in modelling the extremes of temperature it is also important to keep this part of the pre-processing included. As such, we use the most general form of pre-processing outlined in Section 6.3.1. The estimates for $_1\tau_s$ and $_1\sigma_{s,u_s^\dagger}$ show some possibility of offsetting one another in the south-east corner. To check this we fixed the value of $\tau_s(g_t) = {_0\tau_s}$ at all sites and re-estimate $_1\sigma_{s,u_s^\dagger}$ and found that the significant changes in the south-east are still present. The changes are not brought about by an offsetting of $_1\tau_s$ and $_1\sigma_{s,u_s^\dagger}$ and as such covariate structure should be incorporated in both parameters. The shape parameter of the GPD is taken to be constant over time and is found to be negative at all sites over Australia, indicating a finite upper bound to the distribution at each site.
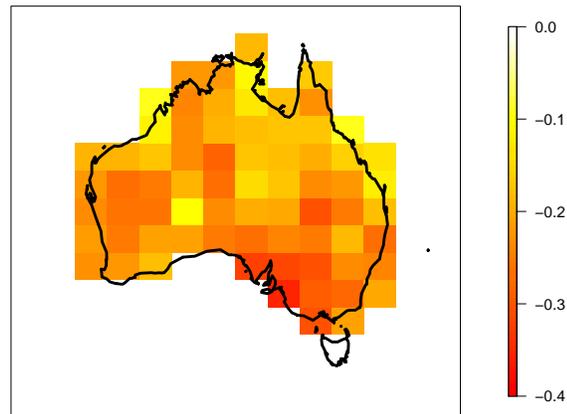


Figure 6.6.2: Estimates of the stationary GPD shape parameter $\xi_s$.

The clearest picture of the effect of the covariate can be seen when examining return levels after transforming onto the original scale. Figure 6.6.3 gives the one-year and fifty-year return levels on the original margins during an El Niño event (i.e. the value of Niño3.4 is $+1$°C) along with the change relative to a La Niña event (i.e. the value of Niño3.4 is $-1$°C). It is observed that the central regions of Australia are hotter

than coastal regions as expected. In terms of the change in the return levels with a change in the phase of ENSO, at certain sites there could be a increase of up to 1ºC in the one year return level between an El Niño event and a La Niña event. From a spatial perspective, the largest increases in the temperature occur in western and mid-eastern regions. The change in the 50-year return level is broadly similar, however southern and some northern areas show a larger decrease in temperatures with an increase in Niño3.4. This pattern is observed due to the covariate effect on the GPD scale parameter shown in Figure 6.6.1.

### 6.6.2 Spatial dependence

We now model the spatial dependence of the transformed data and see if ENSO has a significant effect. Throughout this analysis, where applicable the conditioning site (denoted $s^*$) is chosen to be the grid-box that contains Melbourne. A similar analysis could be followed through for other sites.

In Section 6.3.2 the conditional extremes model was outlined as the approach being used to estimate extremal dependence quantities. We provide justification for using this approach as opposed to other methods that can account only for asymptotic dependence. Figure 6.6.4 shows estimates of the extremal dependence measure $\chi_{s|s^*}(v)$ calculated empirically from the observed data (top), using the stationary conditional extremes approach (centre) and using a spatial version of the non-parametric asymptotically dependent estimator derived in Chapter 3 (bottom), i.e. where $\alpha_{s|s^*} = 1$ and $\beta_{s|s^*} = 0$ for all $s \in S$. With $v$ set at the 90% quantile both approaches appear to be capturing the spatial dependence well. However at higher levels, such as the one-year return level, the asymptotically dependent approach is overestimating the amount of dependence across the field, especially at sites further from the conditioning site. The conditional extremes approach permits asymptotic independence and as such can more accurately capture the dependence observed in this data set. Figure 6.6.4 also
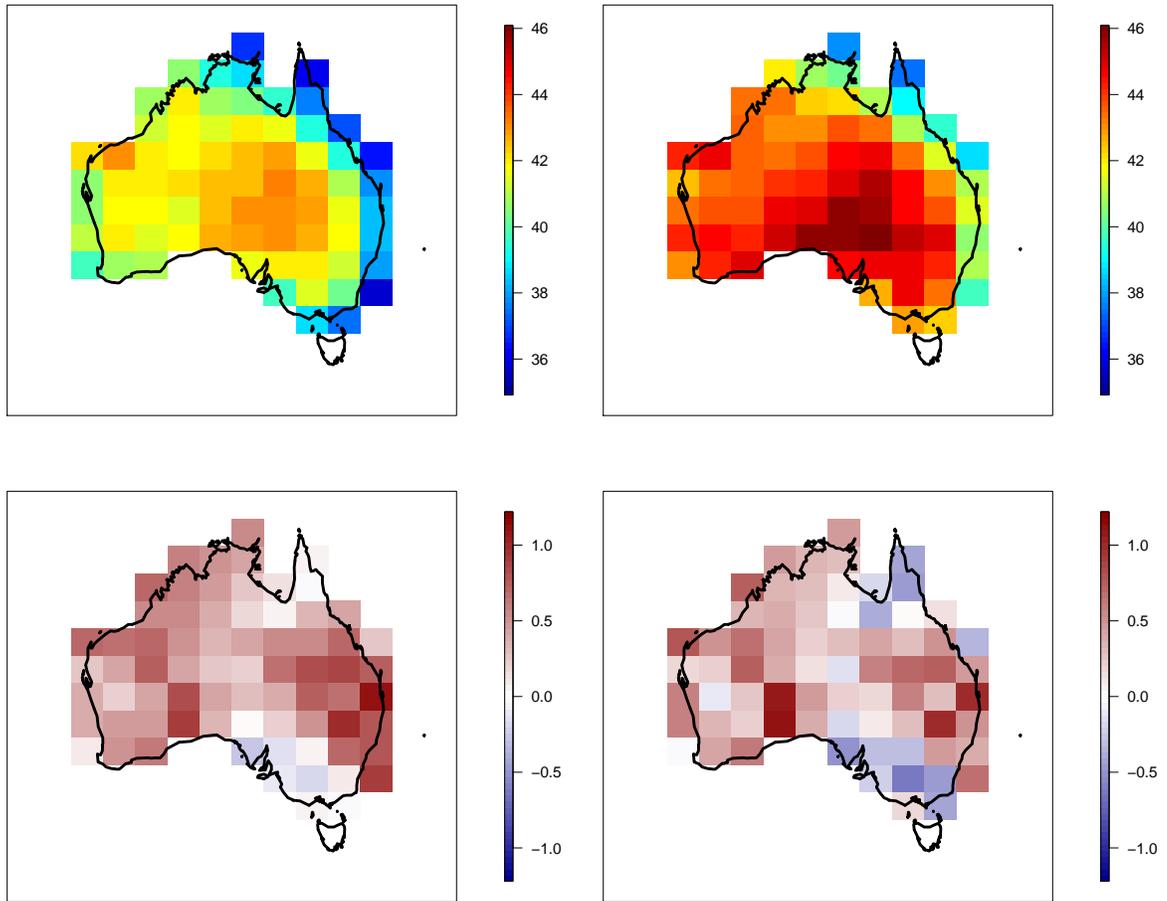
Figure 6.6.3: One-year (left) and 50-year (right) return levels plotted on original margins during El Niño conditions with SST temperature anomaly of +1°C (top) and change between return levels for El Niño and La Niña conditions under temperature anomaly of +1°C and −1°C respectively (bottom).

shows the decay of extremal dependence is not directly proportional to distance or invariant to direction. This highlights that standard geostatistical approaches and max-stable processes that are defined in terms of a consistent distance measure over space would not be reliable here; although it is noted that anisotropic dependence structures within geostatistics could be used in this situation.

Figure 6.6.5 gives estimates of the extremal dependence parameters $\boldsymbol{\alpha}_{-s^*,t}$ and $\boldsymbol{\beta}_{-s^*,t}$.
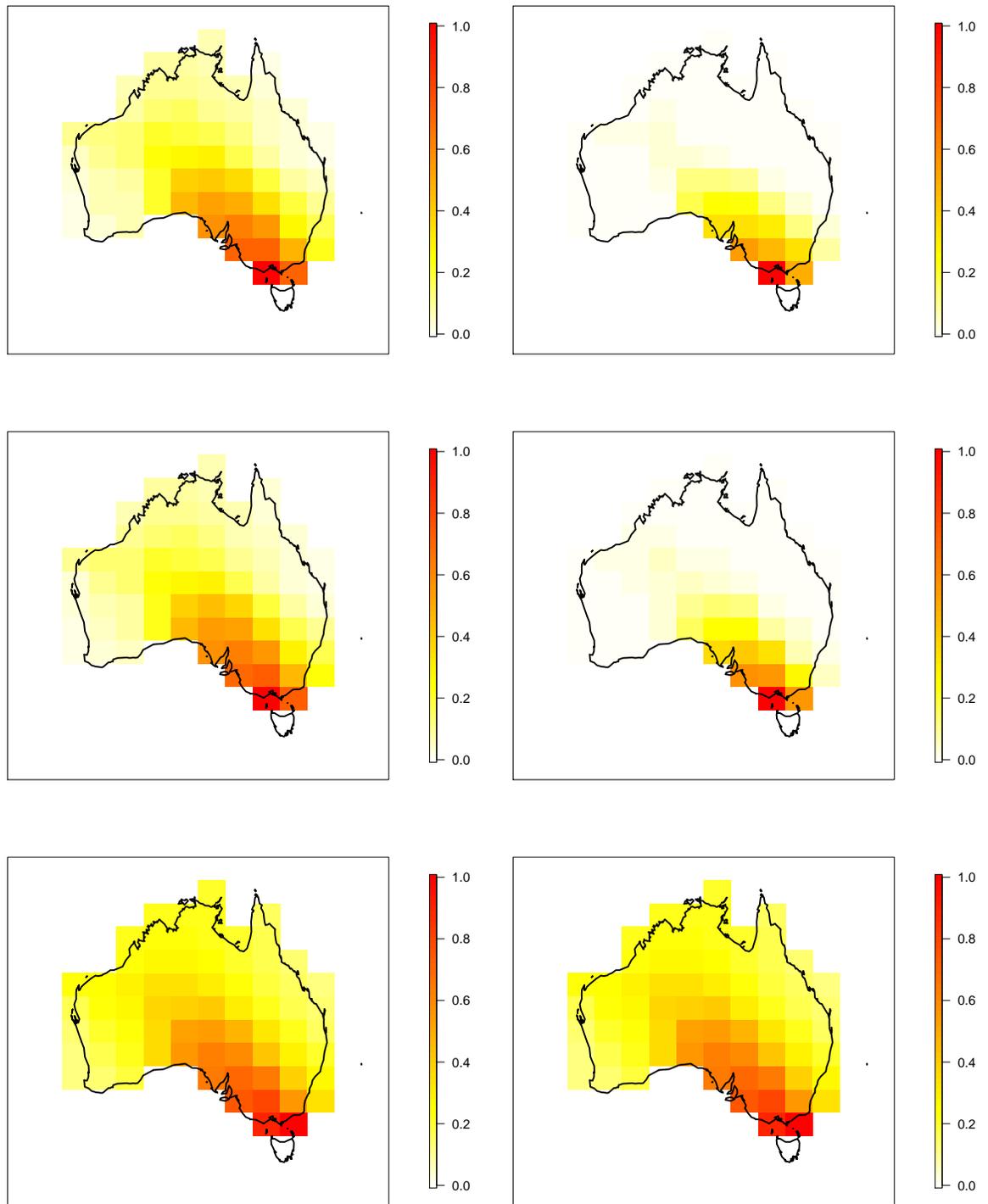
Figure 6.6.4: Values of $\chi_{s|s^*}(v)$ with $v$ set at 90th quantile (left) and one year return level (right) for empirical approach (top), conditional extremes approach (centre) and non-parametric approach that assumes asymptotic dependence (bottom). Here, the conditioning site $s^*$ is taken as the grid-box that contains Melbourne. Stationary model without covariate effect has been fitted here for dependence model assessment.

We observe that the value of $_0\hat{\alpha}_{s|s*}$ is higher for sites $s$ that are located closer to the conditioning grid square, which is expected since sites close together are likely to be more dependent than sites that are far apart. The change in $\boldsymbol{\alpha}_{-s*,t}$ with the covariate is shown by the estimate $_1\hat{\alpha}_{s|s*}$ which demonstrates an increase in extremal dependence, as $g_t$ increases, over northern regions with a slight decrease in the east. The estimates of $_1\hat{\beta}_{s|s*}$ seem to be consistently negative across northern region. These parameter estimates suggest that extreme temperature events that are occurring over Melbourne are more likely to extend over northern regions of Australia during El Niño conditions.

Drawing conclusions from the dependence parameters alone is difficult, especially since they can trade-off against one another. It is often easier to understand how the dependence parameters combine by estimating the extremal dependence measure $\chi_{s|s*}(v)$ given in equation (6.1.2). Here we set the critical level to be $v = v_1$, where $v_1$ is the one-year return level given a particular value of the covariate $g_t$. In Figure 6.6.6 a map of $\hat{\chi}_{s|s*}(v_1)$ is given for an El Niño event along with a map of the difference in $\hat{\chi}_{s|s*}(v_1)$ between El Niño and La Niña conditions. The model output suggests that during an El Niño event, conditioning on a hot event occurring over Melbourne, the spatial extent for a hot day over the south-east is likely to to increase over southern coastal regions, including Adelaide, but not cover as much of the south-eastern region.

The pattern observed in Figures 6.6.5 and 6.6.6 is dependent on the choice of conditioning site. To understand whether this pattern is consistent across all sites we estimate the new quantities defined in Section 6.4. Firstly, we estimate $\phi_{R|s}$ for all sites in Australia; results are given in Figure 6.6.7 for El Niño conditions and the change between El Niño and La Niña conditions. Here, $R$ is taken to be the set of all 72 grid-boxes over Australia. It is observed that events occurring in the middle and east of Australia seem to have a greater spatial extent than for the west side
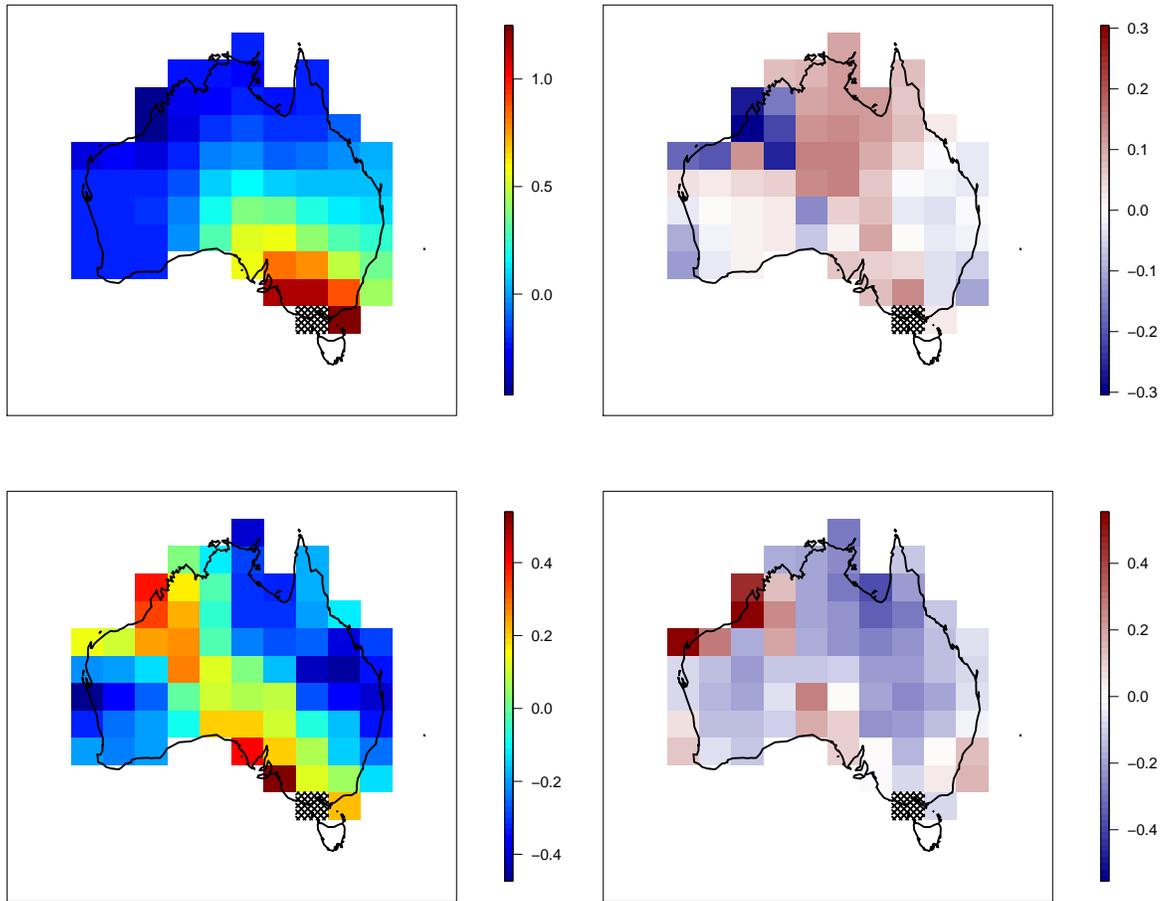
Figure 6.6.5: Conditional extremes dependence parameters $_0\hat{\alpha}_{s|s^*}$ (top left), $_1\hat{\alpha}_{s|s^*}$ (top right), $_0\hat{\beta}_{s|s^*}$ (bottom left) and $_1\hat{\beta}_{s|s^*}$ (bottom right), conditioning on site $s^*$ in south-eastern region (black hashed).

during an El Niño event. The change in $\phi_{R|s}$ between an El Niño event and a La Niña event suggests that El Niño conditions lead to a reduction in the spatial extent of hot days across most of Australia, i.e. La Niña conditions will lead to more widespread hot events. Figure 6.6.7 suggests that during La Niña conditions the difference in the spatial extent of hot days between the east and west will become more pronounced. We also observe that results obtained conditioning upon Melbourne are typical of coastal grid boxes in the south-eastern region.
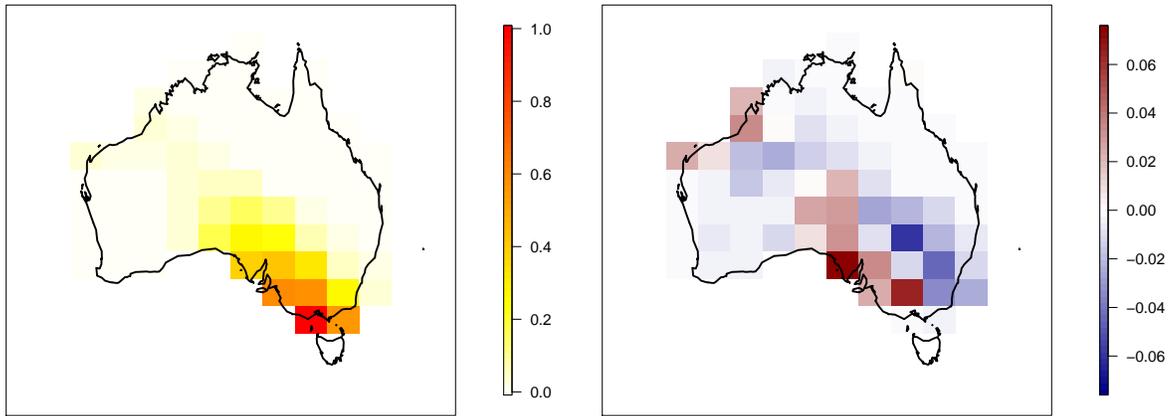
Figure 6.6.6: Extremal dependence measure $\chi_{s'|s}(v)$ for control site over Melbourne under El Niño conditions $g_t = +1$ (left) and difference between extremal dependence measures during El Niño and La Niña ($g_t = -1$) years (right).
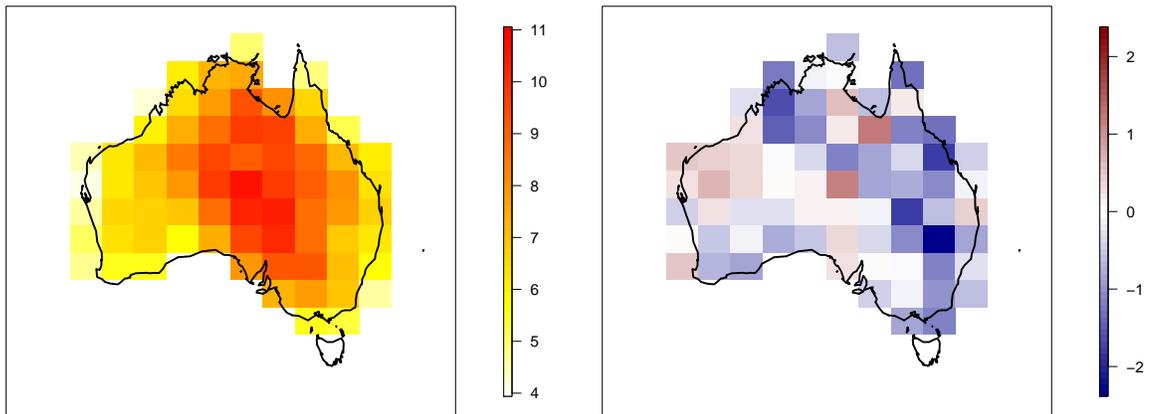


Figure 6.6.7: Estimates of $\phi_{R|s}$ across Australia under El Niño conditions (left) and the change in estimates of $\phi_{R|s}$ between an El Niño and La Niña year.

All the measures estimated in this section so far have been conditional upon an extreme temperature at a particular conditioning site. In Section 6.4 we derived the measure $\omega_{R'|R}$ which gave the probability of a hot event occurring over a region $R'$ conditional upon there being an exceedance over a specific region $R$. Here, we are interested in estimating the probability of a hot event occurring over Melbourne (de-

noted $s^*$ such that $R' = \{s^*\}$) given that an extreme temperature is observed some-where over a critical region $R$ of Australia, here defined as a set of 14 sites in the south-eastern region. Under El Niño (La Niña) conditions we have that $\hat{\omega}_{s^*|R} = 0.20$ (0.19) respectively. This result suggests that analysis based upon a single conditioning site can give significant results that average out when multiple conditioning sites are considered. As such it is important to be able to estimate both types of quantity for a complete analysis.

We also estimate $\rho_j$ for a critical region $R^*$, here taken to be 14 grid boxes in south-eastern Australia, to give an idea of how rare an event like the 2009 heatwave event would be under different ENSO conditions. Figure 6.6.8 shows estimates of $\rho_j$ under the observed La Niña ($g_t = -0.7$) and El Niño ($g_t = 1$) conditions. As mentioned previously, 2009 was a La Niña year; the estimation of the measure for the same event under El Niño conditions attempts to show hypothetically whether such an event would have been likely under El Niño conditions. In the left plot, the maximum value is taken to be greater than $v_1$ whereas in the right plot we have fixed the rarity of the peak value of all simulated events to coincide with the 2009 event; in both plots we have allowed the location at which the event occurs to vary across $R^*$. The left plot shows that at low $j$ ($1 \leq j \leq 5$) under El Niño conditions the observed event would be rarer than under La Niña conditions. As $j$ is increased there seems to be little difference between the different ENSO conditions. It is noted that irrespective of the ENSO conditions, the observed event was very rare. The right plot shows the rarity of the observed event given that the maximum is fixed at the peak of the observed event. In this situation, at all values of $j$ there is a difference between ENSO phases, with the observed event much less rare if it was to occur under El Niño conditions than for La Niña conditions.
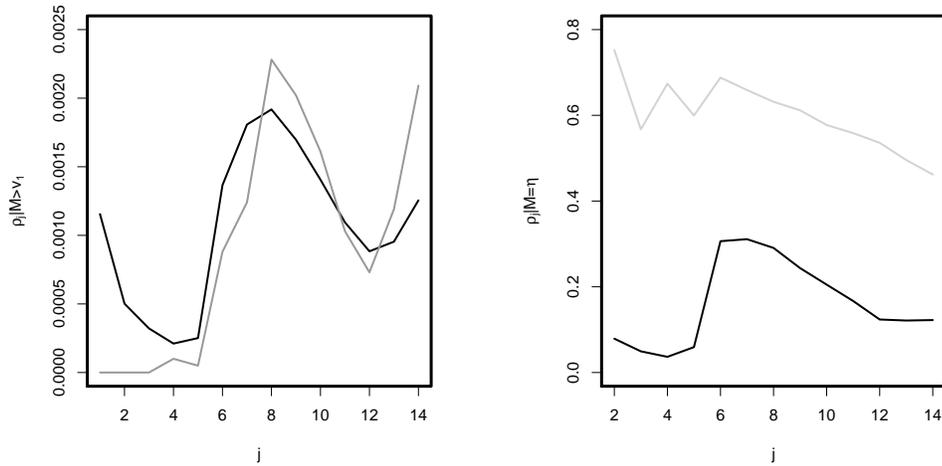
Figure 6.6.8: Estimates of $\rho_j$ conditional upon a maximum value greater than $v_1$ (left) and conditional upon a maximum value fixed as the maximum obtained during the 2009 heatwave event (right) occurring somewhere within a region of interest $R^*$ for observed La Niña conditions (black) and El Niño conditions (grey). In right plot, conditioning upon fixed maximum $j = 1$ is omitted since $\rho_j = 0$ for both methods.

## 6.7 Discussion and Conclusion

In this paper we have modelled the spatial extent of extreme temperature events over Australia and motivated an approach for modelling gridded spatial data using the conditional extremes approach. Within this framework we have included the ability to account for covariates within the margins and the dependence structure which has allowed us to understand the effect of ENSO on extreme temperatures. Our approach has confirmed that El Niño conditions lead to higher temperatures across most of Australia and that the increase in temperature might not be uniform at all return levels, i.e. the effect of ENSO does not just cause a shift in the distribution of temperatures.

Results regarding the change in the spatial extent of heatwaves with ENSO value are more subtle than the changes in marginal structure and vary for different sites.

We have shown that during La Niña conditions, a hot event over Melbourne is likely to cover more of the south-eastern region. However, this pattern is not uniform with Adelaide more likely to be concurrently hot under El Niño conditions. We have also estimated quantities that are not dependent on the choice of conditioning square, which are vital for practical use of the presented approach. These measures have highlighted drawbacks in current measures, particularly allowing us to identify asymptotic dependence and independence, and as such need to be considered in future spatial analyses. In particular, SAF curves, explored here for the first time within this context, are shown to be useful for succinctly presenting complex space-time information. They are shown to be useful for model checking and are already widely used by practitioners so should allow for easier dissemination of our findings. We have also used the observations from the 2009 heatwave event to estimate whether the event would have been more likely under El Niño or La Niña conditions. The quantities presented here are just a subset of potential measures that could be estimated, we have outlined a general approach for simulating spatial extreme temperature events that could be used to generate any quantity of interest for decision makers.

The impact of climate change on the spatial distribution of extreme temperature events has not been dealt with in this paper. This is clearly an important issue that could be included into our framework as another covariate, see Chapter 5 for a temporal framework at a single site. One problem concerns the uncertainty regarding the effect of climate change on ENSO which is currently not well known and would preclude a comprehensive study of the joint effects of ENSO and climate change on extreme temperatures.

Finally, it is also noted that from a mortality perspective runs of hot temperatures are more important than particular hot days. In Chapter 3 we show how replicate heatwave events can be simulated using the conditional extremes framework for a sin-

gle site and suggest some relevant measures of interest to estimate. The next step will be to combine these temporal approaches with the spatial approaches outlined in this paper to generate full space-time simulations on a lattice which incorporate asymptotic independence as well as asymptotic dependence, hence expanding on max-stable spatio-temporal models of Davis et al. (2013) and Huser and Davison (2014).

# Chapter 7

# Further work and outcomes

In this chapter, we present a broad summary of concepts for two extensions to the main body of work given in Chapters 3-6 and discuss outcomes from the thesis. The first extension connects the purely temporal and spatial work from the previous chapters into a coherent space-time model that can be used to model heatwave events where the data are given on a lattice. This work can be seen as the natural ending point of the thesis, with all models expounded up until this point special cases of this most general model. The second extension shows how the methods outlined in the thesis could be applied to model risks associated with droughts. Droughts occur when there is a deficit of precipitation over a period of time; in a statistical sense the problem set-up for droughts is very similar than for heatwaves. We explain where the methods work well and difficulties that arise when considering droughts as opposed to heatwaves.

## 7.1 Space-time modelling of heatwaves

In this chapter, basic extreme value theory for the margins and dependence structure is not repeated; for information on these models refer to previous chapters. Here, the aim is to outline extensions to the previous modelling framework required for a full space-time model of heatwaves on a lattice. The most important concept is

the extension of the Markov assumption over space and time. To create realisations with the desired extremal properties over space and time we need to incorporate information from all sites at previous time-steps. This requires a higher-dimensional dependence model than in previous sections which is discussed in Section 7.1.1. In Section 7.1.2 extensions to the simulation approach are outlined. Many extremal measures developed so far are unable to capture both space and time characteristics. In Section 7.1.3 we present a discussion of how to modify these extremal measures. Finally, we present some results from a preliminary analysis.

## 7.1.1  Dependence modelling and inference

Take data $\{\mathbf{X}_{S,t}\}$ at a set of locations $S$ that have been transformed onto Laplace margins from the original time-series $\{\mathbf{Y}_{s,t}\}$ using the marginal transformations outlined in previous chapters. Here, for simplicity we assume that the dependence structure is stationary through time. When fitting the conditional extremes approach we are now interested in the behaviour of $\mathbf{X}_{S,t+\tau}$ given that an extreme event has been observed at $X_{s,t}$, where $s \in S$ and $\tau$ is a time-lag. When $\tau = 0$ the problem reduces to a purely spatial problem, as in Chapter 6, and as such the limit form given in equation (6.3.2) holds. For any higher time-lag $\tau = 1, \ldots, k$, the assumption is made that there exists a single class of normalising functions such that the limiting relationship between a site $s \in S$ and all other sites in $S$ at time-lag $\tau$ is given by

$$\mathrm{P}\left(\frac{\mathbf{X}_{-s,t+\tau} - \boldsymbol{\alpha}_{-s,\tau}X_{s,t}}{X_{s,t}^{\boldsymbol{\beta}_{-s,\tau}}} \leq \mathbf{z}, X_{s,t} - u > x \;\middle|\; \frac{\mathbf{X}_{S,t+1:t+\tau-1} - \boldsymbol{\alpha}_{S|s,1:\tau-1}X_{s,t}}{X_{s,t}^{\boldsymbol{\beta}_{S|s,1:\tau-1}}} = \mathbf{z}_{S,1:\tau-1},\right.$$

$$\left.\frac{\mathbf{X}_{-s,t} - \boldsymbol{\alpha}_{-s,0}X_{s,t}}{X_{s,t}^{\boldsymbol{\beta}_{-s,0}}} = \mathbf{z}_{-s,0}, X_{s,t} > u\right)$$

$$\to G_{-s,\tau|S,0:\tau-1}(\mathbf{z} \mid \mathbf{z}_{-s,0}, \mathbf{z}_{S,1:\tau-1})\exp(-x),$$

as $u \to \infty$ and for $x > 0$, where $\mathbf{X}_{S,t} = (\mathbf{X}_{-s,t}, X_{s,t}) = (X_{1,t}, \ldots, X_{|S|,t})$ and $\mathbf{X}_{S,t+1:t+\tau-1} = (\mathbf{X}_{S,t+1}, \ldots, \mathbf{X}_{S,t+\tau-1})$, with $\mathbf{z}_{S,\tau} = (\mathbf{z}_{-s,\tau}, z_{s,\tau}) = (z_{1,\tau}, \ldots, z_{|S|,\tau})$ and

$\mathbf{z}_{S,1:\tau-1} = (\mathbf{z}_{S,1}, \ldots, \mathbf{z}_{S,\tau-1})$. Dependence parameters are given as

$$\boldsymbol{\alpha}_{S|s,1:\tau-1} = (\boldsymbol{\alpha}_{S|s,1}, \ldots, \boldsymbol{\alpha}_{S|s,\tau-1})$$

$$\boldsymbol{\beta}_{S|s,1:\tau-1} = (\boldsymbol{\beta}_{S|s,1}, \ldots, \boldsymbol{\beta}_{S|s,\tau-1}),$$

where

$$\boldsymbol{\alpha}_{S|s,\tau} = (\boldsymbol{\alpha}_{-s,\tau}, \alpha_{s,\tau}) = (\alpha_{1,\tau}, \ldots, \alpha_{|S|,\tau})$$

$$\boldsymbol{\beta}_{S|s,\tau} = (\boldsymbol{\beta}_{-s,\tau}, \beta_{s,\tau}) = (\beta_{1,\tau}, \ldots, \beta_{|S|,\tau}).$$

Ranges and limiting cases for the dependence parameters are consistent with previous chapters.

The approach taken to estimate the distribution function $G_{-s,\tau|S,0:\tau-1}$ is consistent with the approach outlined in Chapter 4 and requires an estimate of the joint density function $g_{S,1:\tau}$ obtained via kernel density estimation. All dependence parameters are estimated pairwise using standard likelihood approaches, for all pairs of sites and time-lags.

Selection of the order of Markov process to use is complicated by the space-time nature of the problem. In univariate series, standard approaches to estimate the order of the Markov process are the ACF and PACF. For multivariate problems the cross-correlation function (CCF) is an analogue for the ACF; an analogue of the PACF is less well defined. In Chapter 4 we outlined a suite of diagnostics to more accurately choose the order Markov process when we are interested in tail behaviour. In theory these approaches could be extended, however the increase in dimension of the problem, to account for space and time, would increase the computational burden of these diagnostics and exacerbate problems with wide uncertainty bounds caused by having to estimate many parameters. Measures such as the sub-asymptotic extremal dependence measure $\chi(v)$ are also less well defined over space and time; a drawback we expand on in Section 7.1.3.

### 7.1.2 Simulating heatwave events

The simulation of heatwave events using a space-time model builds upon the peak value simulation approach set out in Chapter 3 with extensions required to take into account temporal and spatial structures. The simulation approach for including temporal structure was outlined in Chapter 4 and a similar approach for including spatial structure was outlined in Chapter 6. First, we choose a location at which the peak value of an event occurs, denoted $s$, and simulate an exceedance of critical level $v > u$ from a unit Exponential distribution. The location $s$ can either be chosen in line with observed heatwave events or simulated at random from the joint distribution of a maxima occurring at a specific site over $S$. A spatial field can be simulated using the algorithm developed in Chapter 6 and then stepped forward and backward using a similar algorithm to Appendix D with additional dimensions added to account for spatial structure.

The additional dimensional complexity of the problem could lead to difficulties when simulating heatwave events. A higher-order model will often be required since events are more likely to stay within the system longer, i.e. a meteorological event could stay over Australia for many days longer than for a specific site within Australia. This effect, combined with the additional spatial dimensions, will require a more complicated conditional density to be estimated. The chain length $m$ for any algorithm will also need to be run longer to ensure a negligible probability of simulating a chain with an exceedance at greater than lag $m$.

### 7.1.3 Extremal measures

In previous chapters we have defined measures that summarise temporal and spatial dependence separately. To do a full space-time analysis it is necessary to obtain measures that combine both temporal and spatial information. However, these measures are often difficult to represent and hard to visualise. Here, we give some suggestions

on how previous measures could be extended to provide useful information.

In both temporal and spatial contexts one important and flexible measure has been the threshold dependent extremal dependence measure. Naturally, this measure can be easily extended in the space-time context for processes stationary in time, given as

$$\chi_{s'|s,\tau}(v) = \mathrm{P}(X_{s',t+\tau} > v \mid X_{s,t} > v),$$

where $v$ is the critical level on Laplace margins, $\tau$ is some time-lag and $s$ and $s'$ are two sites of interest. In Section 7.1.4 some preliminary results are given using the measure above within the space-time context.

In Chapter 3 we outlined measures of the duration and severity of a heatwave event. Similar measures can be defined for each site $s \in S$, given as

$$D_v(s) = \sum_{t \in C} \mathbb{I}(X_{s,t} - v),$$

where $\mathbb{I}(\cdot)$ is the indicator function and $C$ is a set of values constituting a cluster. This measure of the duration can be calculated at each location separately across $S$ and mapped to give an idea of the duration of an event across $S$. An analogous measure of the severity of an event at a particular locations is given as

$$R_v(s) = \sum_{t \in C} (X_{s,t} - v)_+.$$

Such measures do not take into account spatial structure and as such are poor for estimating whether multiple locations are being affected by the same heatwave event simultaneously. The suite of measures for summarising spatial dependence outlined in Chapter 6 could also be extended to better understand the spatio-temporal behaviour of heatwave events.

### 7.1.4 Results

Figure 7.1.1 gives an illustrative result from a space-time analysis of temperatures across Australia. The data are the gridded daily maximum temperature values that were introduced in Chapter 6. Using the pairwise model for $X_{s,t+\tau} \mid X_{s^*,t}$ the extremal dependence measure $\chi_{s\mid s^*,\tau}$ is estimated for all $s \in S$ with $\tau = 0, 1, 2$, under El Niño conditions (left) and the difference between El Niño and La Niña conditions (right). The conditioning site $s^*$ is taken as the grid-box containing Melbourne. The plots on the top line are the same as those given in Figure 6.6.6, i.e. where $\tau = 0$. Under El Niño conditions the plots at $\tau = 1$ and $\tau = 2$ show an expected reduction in the level of $\chi_{s\mid s^*,\tau}$, since as we move further away from the original exceedance in time we would expect extremal dependence to reduce. More interesting results are provided by the difference between El Niño and La Niña conditions. At lag 1 La Niña conditions seem to lead to an increase in the level of extremal dependence and suggest that heatwave events will be more likely to persist in the south-east under La Niña conditions. This pattern is also observed for lag 2, but for later lags (not shown) there is little significant difference between the two ENSO phases. These results indicate that there is interesting behaviour that can only be captured by a full space-time model and future work in this area would be informative.

## 7.2 Modelling droughts

During the thesis we have focused on heatwaves as our natural hazard of choice. One strongly linked phenomenon is drought, which results from sustained periods of dry weather. Modelling droughts using extreme value theory can produce interesting modelling challenges since droughts are combination hazards that result from a lack of precipitation and dry ground conditions. Here, we introduce drought, discuss some of the complications that arise when trying to model droughts as opposed to heatwaves and suggest how work from the thesis could provide useful insight.
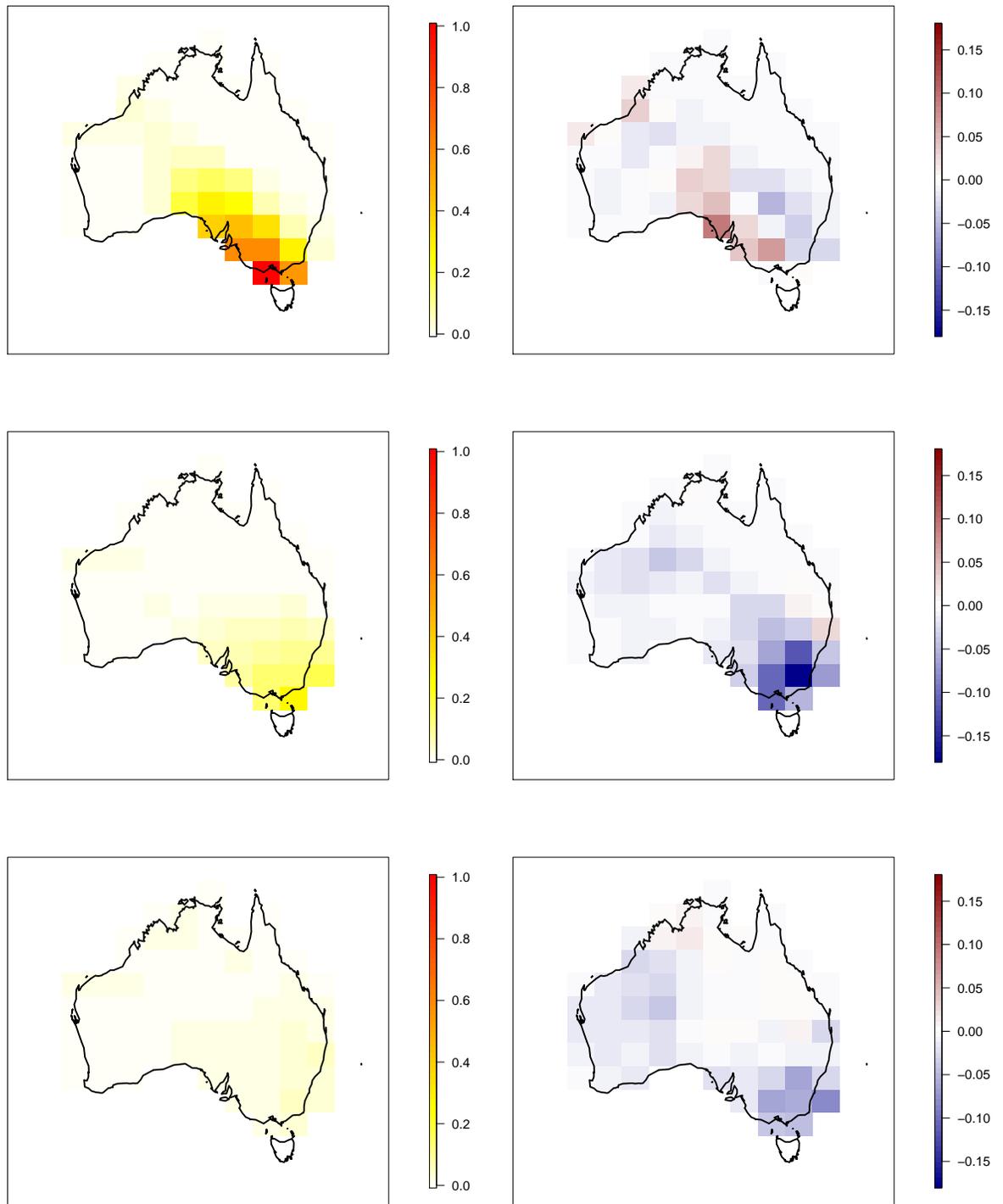
Figure 7.1.1: Estimates of $\chi_{S|s^*,\tau}(v_1)$ under El Niño conditions (left) and difference in $\chi_{S|s^*,\tau}(v_1)$ between El Niño and La Niña conditions (right) for $\tau = 0$ (top), $\tau = 1$ (middle) and $\tau = 2$ (bottom). Conditioning site $s^*$ taken to be grid-box containing Melbourne.

### 7.2.1   Introduction

The IPCC (2012) special report into extremes defines drought as:

> *A period of abnormally dry weather long enough to cause a serious hydro-logical imbalance.*

Drought is a phenomenon that has numerous and wide-ranging consequences that result from the problem of hydrological imbalance. Large scale water shortages can often cause famine which in turn has economic consequences and causes fatalities. In less well developed countries, humanitarian aid may be required to help people affected by crop failure which is a drain on the budgets of other countries. Droughts need not be so severe to be of interest to planners and management. Long-scale drought in the UK has affected the levels of reservoirs which has led to the hose-pipe bans in the South East of the country in recent years.

Droughts can be characterized in many different ways depending on the perspective. In scientific literature three different types of drought are emphasised (IPCC, 2012):

- Meteorological drought: occurs when there is a deficit of precipitation.

- Soil moisture drought: a lack of soil moisture that can affect agriculture and ecosystems.

- Hydrological drought: negative anomalies in streamflow, lakes and groundwater.

Periods of drought are driven by a combination of the three factors above. Climate variability most strongly affects the amount of precipitation and as such changing climate behaviour will manifest itself through changes in the level of meteorological drought. Natural changes in soil moisture drought are usually driven by precipitation levels, however under strong drought conditions soil moisture levels limit the amount

of evapo-transpiration which in turn further limits soil moisture drought.  Human activity such as overgrazing and deforestation can affect the capacity of soil to absorb precipitation which can lead to human-induced drought events.  The memory and water storage capacity of a system also affect measures of hydrological and soil moisture drought.  Lower precipitation levels year on year can have a severe impact on soil moisture and increase the likelihood of further drought.

The onset, duration and cessation of drought periods are complex to predict as they rely on both the amount of precipitation falling at a given time and the current state of the system.  Once dry conditions have begun a positive feedback loop can set in. The dry conditions reduce the moisture in the upper soil layers, which in turn reduces the evapo-transpiration rate.  This reduces the relative atmospheric humidity which makes rainfall less likely from clouds in the region.  The conditions are only broken when disturbances from outside the region with enough moisture produce precipitation to stop the drought.  Large-scale climatic phenomena such as ENSO and the North Atlantic Oscillation (NAO) can lead to blocking which can cause conditions to persist for longer than would be expected (Tallaksen and Lanen, 2004).

Droughts are an example of a creeping phenomenon (Mishra and Singh, 2010); i.e. it is difficult to determine when persistent dry weather becomes a drought and the impacts of drought can still be felt many years after a drought event has passed. When compared to other natural hazards such as hurricanes, tornadoes and volcanoes, drought provides some unique challenges.  More people are affected by drought which makes provision for relief harder.  A lack of agreed definition of drought can harm response times to large drought events.  Since drought can be caused by human activities, in many situations natural factors alone cannot be used to predict it.

## 7.2.2   Drought indices

Several different indices have been developed as measures of drought.  Each index attempts to model one or more of the three main types of drought whilst using available precipitation and evapo-transpiration data. Due to the relative reliability of the respective data, measures of drought tend to focus on precipitation levels while indirectly considering the effects of evapo-transpiration. As such, here we outline common drought indices for meteorological drought.

**Standardised Precipitation Index**

The Standard Precipitation Index (SPI) (Mckee et al., 1993) is the most commonly used drought index. It is obtained by fitting a probability distribution to a series of precipitation values and then transforming these values onto a Normal margin. The most common choice of probability distribution is the Gamma distribution (Mckee et al., 1993) since this often fits the body of the distribution well; see Mishra and Singh (2010) for an overview of other distributions used. However, at a practical level an empirical cumulative distribution function is often used to transform the data onto uniform margins (Wheatley, 2010). Negative values of SPI correspond to droughts, with increasing severity as the SPI becomes more negative. A surplus of rain corresponds to positive values of the SPI.

The major benefit of SPI is that it can be used to compare droughts over different accumulation time-scales. This is important since droughts over different time-scales can lead to different problems, e.g. soil moisture for agriculture is affected by droughts on a short time-scale whereas long time-scale droughts can affect the level of reservoirs and groundwater. However, there are limitations to the use of SPI as an index of drought. Fitting different distributions can lead to varying SPI estimates, especially in the tails. The use of empirical cumulative distribution functions does not permit estimation of the SPI for stronger events than have been observed previously. Finally,

in very dry climates a proliferation of zero values can cause inferential problems; although this is a common issue across all indices.

### Consecutive dry days index

The Consecutive Dry Days (CDD) index (Frich et al., 2002) considers the maximum number of consecutive dry days in a given period. The level taken as corresponding to a dry day is subjective and can be altered by the user (the common amount is 1mm in a day). The measure is based on precipitation and only indirectly takes into account evapo-transpiration. This index most closely links to the measures presented in previous chapters for heatwaves.

### Palmer drought severity index

The Palmer Drought Severity Index (PDSI) (Palmer, 1965) incorporates precipitation and temperature to estimate the departure of the moisture balance from normal conditions. The underlying model includes a two-layer soil model with a simple water balance model. The additional information incorporated is an advantage, however there are some severe limitations to using such an index. PDSI works optimally on short time-scales, making it more accurate when identifying short-term agricultural drought than long-term hydrological drought. The effects of snow fall are ignored which affects the index during winter months and at high elevation. The PDSI can also be very sensitive to changes in rainfall and temperature at different times of the year. In the winter months, rainfall dominates since evaporation is minimal whereas temperature has a greater effect in warmer months.

## 7.2.3 SPI with extreme tails

One concern with current approaches to estimating SPI is highlighted by the left plot of Figure 7.2.1. All accumulation lengths that are not a multiple of 12 months will be subjected to some form of seasonality. One approach to overcome this is to fit a
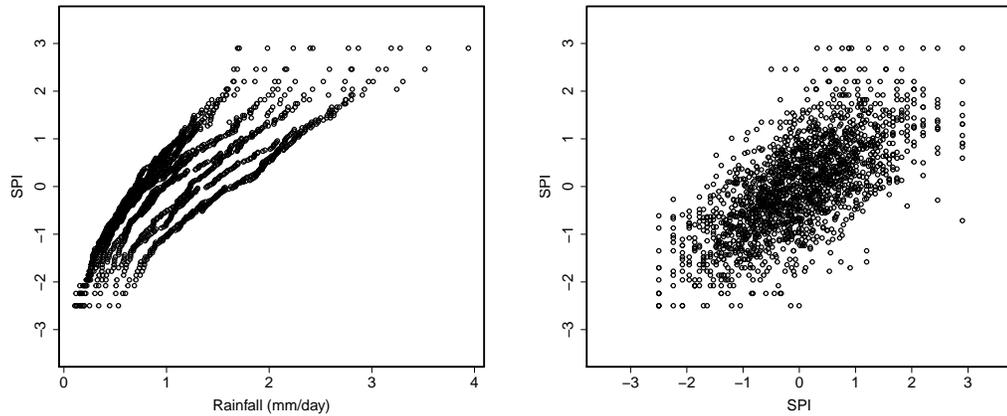
Figure 7.2.1: Comparison of model rainfall (mm/day) against SPI (left) and scatter plot of two sets of data given by SPI method (right). Data are monthly rainfall accumulations from GCM for two sites in Southern Africa. Both figures show problems with discontinuity for extreme values caused by having few values in the tail.

separate empirical cumulative distribution function to each set of months (i.e. one for January, February and so on). For a given data set each month of interest will have the same number of data points and as such we observe discontinuity in the tail when the empirical cumulative distribution function is applied. The problem is that there are few values in the tail of the distribution and so extreme rainfall values are often recorded as the same value on the SPI scale.

This issue becomes a greater problem when comparing rainfall from different sites on the SPI scale. If we are interested in estimating the level of extremal dependence, discontinuity in the upper tails of the plot on the right of Figure 7.2.1 suggests that we shall obtain poor estimates of the tail behaviour if we use estimates of SPI that is based upon the empirical cumulative distribution function.

We can produce more accurate values of the SPI by fitting a GPD to the upper tail of the rainfall data with the empirical cumulative distribution fitted to the body

of the data. For a time-series of precipitation values for a given month $\{X_i\}$ for $i = 1, \ldots, 12$ we can use equation (2.4.1), i.e.

$$F_i(x) = \begin{cases} 1 - \lambda_{u_i} \left(1 + \xi_i \frac{x - u_i}{\sigma_{u_i}}\right)_+^{-1/\xi_i}, & x \geq u_i \\ \tilde{F}_i(x), & x < u_i, \end{cases} \tag{7.2.1}$$

where $\lambda_{u_i} = 1 - \tilde{F}_i(u_i)$ and $u_i$ is some high threshold, with parameters $(\sigma_i, \xi_i)$ defined in the usual way for the GPD, to transform onto the original series onto uniform margins. The same approach can be repeated for the lower tail. In the situation where a rainfall accumulation that is a multiple of 12 months is taken equation (7.2.1) can be replaced by equation (2.4.1) since the accumulation length removes any annual cycle. By fitting the GPD above (below) a certain high (low) threshold it is possible to avoid the problem of discontinuity in the tails and also has the benefit of allowing back transformation onto the original margins for values greater than the largest observed values.

## 7.2.4  Comparison with modelling heatwaves

A natural question to pose is whether the extreme value techniques developed in previous chapters can be applied to the drought problem. As seen in Section 7.2.1, the most basic definition of drought refers to some abnormally dry weather that persists over a long enough time period to lead to hydrological imbalance. This definition closely resembles the equivalent definition of a heatwave event given in Section 1.2 and as such motivates the use of similar approaches as have been outlined throughout the thesis.

However, many complications can arise when modelling droughts as opposed to heatwaves. Primarily the occurrence of drought conditions is not only affected by the amount of precipitation, but also the amount of groundwater in the system along with other characteristics such as the run-off. Therefore, modelling the number of

consecutive dry days is important but needs to be embedded within a larger framework modelling interactions with other environmental factors. This can be accomplished by the methods proposed in the thesis since the joint extremal behaviour of precipitation and any other variable can be modelled using the extremal dependence approaches outlined previously. However, the best way to do this is not clear and would benefit from direct discussion with experts in hydrology.

When investigating the spatial structure of droughts additional complications exist. Many rainfall events occur on a small spatial scale and climate models can often miss the occurrence of these kinds of events. To capture these events we require precipitation data on a finer resolution. This would increase the computational costs associated with running the techniques previously outlined during the thesis, especially if simulating full space-time events. For a full analysis of the risk associated with droughts it is also necessary to take into account the spatial structure of river basins along with the simple Euclidean distance between locations. Asadi et al. (2015) give an approach to include this within an extreme value analysis, but incorporating this structure within our framework is not straightforward and would require further work.

Even when focusing on drought at a particular location while ignoring the effect of groundwater and run-off there are other problems. When modelling dry weather we are specifically analysing the lower tail of precipitation values. On many days within a year there is no precipitation and as such there are a proliferation of zero values within any precipitation data set. Associated problems are two-fold. Firstly, if there are many zero values, choosing a modelling threshold as a low quantile can lead to a threshold set at zero. Secondly, many climate models are not able to reproduce days without precipitation completely correctly and underestimate the amount of such days.

Two solutions exist for the problem of zero precipitation values within the data. One is to censor values that fall below a certain low precipitation value, an approach that could be incorporated into an extreme value analysis. The other approach is to analyse a precipitation accumulation instead of the precipitation values on the original time-scale. As outlined in 7.2.1 different aspects of drought behaviour can be captured by looking at different time-scales. Here the accumulation reduces the occurrence of zero values and permits the use of threshold based extreme value approaches. Precipitation accumulations are usually created using a sliding window of a certain length, often taken to be 12 months to remove the effect of seasonality. As such taking an accumulation introduces serial correlation into the precipitation values which affects estimates of the temporal dependence. It is not clear how this effects interpretation of temporal dependence results and as such these should be treated with caution.

A final consideration concerns the cessation of a drought event. When modelling heatwaves it is clear that when a temperature falls below a critical level the event has ended. However, when considering droughts, if precipitation rises above a critical level it is not clear that the drought event has finished since groundwater levels may have not been replenished to 'normal' levels. As such a model with two critical levels might be required; this will be a more complicated model but should be feasible within the framework outlined during the thesis.

## 7.3 Outcomes of thesis

The aim of the thesis, outlined in Section 1.3, was to provide extreme value methodology to model the behaviour of heatwaves. In this section, we outline the main findings contained within each chapter and how these contribute to achieving the main aims of the thesis.

When modelling heatwave events using extreme value methods the most important aspect has been developing models that can account for temporal structure. It is clear that most casualties caused during a heatwave event occur due to sustained high temperatures as opposed to singular hot days. As such univariate extreme value approaches, such as the fitting GEV and GPD models, are inadequate for our purposes. The desire to model temporal dependence has motivated the use of extremal dependence models to obtain more accurate estimates of the persistence of heatwave events. In Chapter 2, two approaches for modelling extremal dependence were compared and found to perform equally well for bivariate problems. However, the added flexibility of the conditional extremes approach suggested this as a suitable model for extremal dependence in later chapters.

Chapter 3 introduced a method for estimating the probability of heatwave events under a first-order Markov assumption. This assumption permitted the use of bivariate extreme value results and a set of different models were tested to model extreme values. One important concept was the difference between asymptotic dependence and asymptotic independence. Models that cannot account for both types of dependence structure risk producing misleading inferences in the tails. Our study of daily maximum temperatures over Orleans found asymptotic independence between temperatures on consecutive days. As such, our model based upon the conditional extremes approach produced more realistic simulations than other approaches that can only account for asymptotic dependence. Using results based upon asymptotically dependent models can still be justified if we are especially risk averse; in this situation we also found that the choice of parametric model for asymptotic dependence structure can bias results when compared to a non-parametric asymptotically dependent approach.

As part of this study we derived a set of different cluster functionals that we believe

are important when investigating heatwave events. No set definition of heatwaves exists and this can lead to many different indices presented in different papers. One important benefit of our approach is that the properties of most heatwave measures of choice can be obtained from our model, since we simulate replicate heatwave events from our model. In the work we decided to present estimates for the probability of the occurrence of an event more extreme than the 2003 European heatwave event. We found that the annual probability of observing a heatwave event lasting at least 11 days was 0.001, i.e. we would expect this to happen once in every 1000 years. Output of this type is important as it allows decision makers to understand better how often these devastating events will happen and what preparation measures need to be taken.

One drawback alluded to in Chapter 3 was that a first-order Markov assumption was made. Such an assumption does not capture the prevailing conditions of the weather system, i.e. whether previous days have seen an increase or decrease in temperatures. It was possible that, by making the first-order assumption, we were simplifying higher-order structure in the extremes. This could lead to misleading estimates of important extremal quantities, such as the 1-in-1000 year 2003 event. Chapter 4 introduced a $k$th-order Markov model for assessing heatwave risks. The most important consideration was to choose an appropriate higher-order model. Standard diagnostic approaches based upon the body of the data can often lead to misleading choices for the order. We developed a suite of different diagnostics to help infer an appropriate order Markov chain for modelling the extremal structure accurately. With these diagnostics, it was found that our estimates of the probability of exceeding the 2003 event were too low, with the observed event up to three times more likely than estimated in Chapter 3.

An important consideration when modelling any environmental process is anthropogenic climate change. This is especially true for heatwaves, where climate change

could lead to longer and more severe heatwaves in the future. Chapter 5 investigated the effect of climate change on heatwaves by proposing an extension to the model in Chapter 3 to permit the inclusion of covariate structure in the margins and dependence structure. Daily maximum temperature data were obtained from an ensemble of strongly forced general circulation models and the global mean temperature was used as a covariate for climate change. We found evidence that under an increase in the global mean temperature of 1°C, the one-year return level would increase by 2°C. This result shows that under future climate change heatwave events are likely to become more frequent. However, there was little evidence to support any change in the duration of heatwave events with future climate change.

Up until this point, we had only modelled temporal dependence at a single location. To fully understand the risks associated with heatwaves it is also important to estimate the number of sites that are experiencing hot conditions concurrently. In Chapter 6 we built a model to model the spatial structure of extreme temperature events. For this study, the effects of temporal dependence were ignored. Covariate structure was also included in the model with the aim to understand the effects of ENSO on extreme temperatures. To this end, we developed a suite of spatial risk measures to give a complete spatial characterisation of extreme temperature events. We found that El Niño conditions led to an increase in temperatures across the whole of Australia. However, if an extreme temperature event occurs during a La Niña year it had the potential to cover a greater spatial extent.

A natural extension of the methods used to model the spatial and temporal structures of extreme temperatures is to combine both into a more general space-time framework. Ideas about how to formalise this framework were outlined in Chapter 7. Preliminary results suggest that taking full spatio-temporal structure into account can provide interesting and informative insight into the heatwave problem.

The modelling of heatwaves has provided a very interesting application area that has motivated all the work throughout the thesis. However, as an applied statistician it is important that methods are not only built to answer a specific class of problems, but also have the flexibility to be used on a wider class of problems. As such, in Chapter 7 we also introduced drought and have considered how the methods developed in this thesis could be used to model this natural hazard. This points to ways in which the approaches from this thesis could be used more generally in the future across a wide class of problems associated with natural hazards.

# Appendix A

# Parametric joint tail approach with additional marginal information

The log-likelihood associated with the likelihood in equation (2.5.2) is given by

$$\ell(c, \eta) = n_{00} \log \left( 1 - \frac{2}{u} + cu^{-1/\eta} \right) + (n_{01} + n_{10}) \log \left( \frac{1}{u} - cu^{-1/\eta} \right) + n_u \log c$$
$$- n_u \log \eta - \left( \frac{1}{\eta} + 1 \right) \sum_{i=1}^{n_u} \log z_i.$$

An analytical expression for $\hat{c}$ can be obtained by differentiating the log-likelihood with respect to $c$

$$\frac{\partial \ell}{\partial c} = n_{00} \left( \frac{u^{-1/\eta}}{1 - 2/u + cu^{-1/\eta}} \right) - (n_{01} + n_{10}) \left( \frac{u^{-1/\eta}}{1/u - cu^{-1/\eta}} \right) + \frac{n_u}{c},$$

and setting the derivative equal to zero. With some rearrangement

$$0 = n_{00}\hat{c}u^{-1/\hat{\eta}} \left( \frac{1}{u} - \hat{c}u^{-1/\hat{\eta}} \right) - (n_{01} + n_{10})\hat{c}u^{-1/\hat{\eta}} \left( 1 - \frac{2}{u} + \hat{c}u^{-1/\hat{\eta}} \right)$$
$$+ n_u \left( \frac{1}{u} - \hat{c}u^{-1/\hat{\eta}} \right) \left( 1 - \frac{2}{u} + \hat{c}u^{-1/\hat{\eta}} \right), \tag{A.0.1}$$

which can be expanded to give

$$0 = n_{00} \left( \hat{c}u^{-1/\hat{\eta}-1} - \hat{c}^2 u^{-2/\hat{\eta}} \right) - (n_{01} + n_{10}) \left( \hat{c}u^{-1/\hat{\eta}} - 2\hat{c}u^{-1/\hat{\eta}-1} + \hat{c}^2 u^{-2/\hat{\eta}} \right)$$
$$+ n_u \left( \frac{1}{u} - \hat{c}u^{-1/\hat{\eta}} - \frac{2}{u^2} + 2\hat{c}u^{-1/\hat{\eta}-1} + \hat{c}u^{-1/\hat{\eta}-1} - \hat{c}^2 u^{-2/\hat{\eta}} \right),$$

and by collection of like terms

$$0 = \hat{c}^2 \left[ nu^{-2/\hat{\eta}} \right] + \hat{c} \left[ (n_{01} + n_{10}) \left( u^{-1/\hat{\eta}} - 2u^{-1/\hat{\eta}-1} \right) - n_{00} u^{-1/\hat{\eta}-1} \right.$$
$$\left. - n_u \left( 3u^{-1/\hat{\eta}-1} - u^{-1/\hat{\eta}} \right) \right] - n_u \left[ \frac{1}{u} - \frac{2}{u^2} \right].$$

Now using the quadratic formula it is possible to compute a closed form expression for $\hat{c}$ in terms of $\hat{\eta}$

$$\hat{c} = \frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha},$$

where

$$\alpha = nu^{-2/\hat{\eta}}$$

$$\beta = (n_{01} + n_{10}) \left( u^{-1/\hat{\eta}} - 2u^{-1/\hat{\eta}-1} \right) - n_{00} u^{-1/\hat{\eta}-1} - n_u \left( 3u^{-1/\hat{\eta}-1} - u^{-1/\hat{\eta}} \right)$$

$$\gamma = -n_u \left( \frac{1}{u} - \frac{2}{u^2} \right).$$

Now to obtain an analytical value for $\hat{\eta}$ we differentiate the log-likelihood with respect to $\eta$

$$\frac{\partial \ell}{\partial \eta} = n_{00} \left( \frac{c\eta^{-2} \log(u) u^{-1/\eta}}{1 - 2/u + cu^{-1/\eta}} \right) - (n_{01} + n_{10}) \left( \frac{c\eta^{-2} \log(u) u^{-1/\eta}}{1/u - cu^{-1/\eta}} \right)$$
$$- \frac{n_u}{\eta} + \eta^{-2} \sum_{i=1}^{n_u} \log(z_i),$$

from which we can obtain the MLE by setting the above equal to zero such that

$$0 = n_{00} \log(u) \hat{c} u^{-1/\hat{\eta}} \left( \frac{1}{u} - \hat{c} u^{-1/\hat{\eta}} \right) - (n_{01} + n_{10}) \log(u) \hat{c} u^{-1/\hat{\eta}} \left( 1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}} \right)$$
$$- n_u \hat{\eta} \left( \frac{1}{u} - \hat{c} u^{-1/\hat{\eta}} \right) \left( 1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}} \right)$$
$$+ \left( \frac{1}{u} - \hat{c} u^{-1/\hat{\eta}} \right) \left( 1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}} \right) \sum_{i=1}^{n_u} \log(z_i).$$

At this stage it is possible to substitute a rearranged version of equation (A.0.1) such

that

$$0 = (n_{01} + n_{10}) \log(u) \hat{c} u^{-1/\hat{\eta}} \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right)$$

$$- n_u \log(u) \left(\frac{1}{u} - \hat{c} u^{-1/\hat{\eta}}\right) \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right)$$

$$- (n_{01} + n_{10}) \log(u) \hat{c} u^{-1/\hat{\eta}} \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right)$$

$$- n_u \hat{\eta} \left(\frac{1}{u} - \hat{c} u^{-1/\hat{\eta}}\right) \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right)$$

$$+ \left(\frac{1}{u} - \hat{c} u^{-1/\hat{\eta}}\right) \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right) \sum_{i=1}^{n_u} \log(z_i)$$

which leads to

$$n_u \hat{\eta} \left(\frac{1}{u} - \hat{c} u^{-1/\hat{\eta}}\right) \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right) = \left(\frac{1}{u} - \hat{c} u^{-1/\hat{\eta}}\right) \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right) \sum_{i=1}^{n_u} \log(z_i)$$

$$- n_u \log(u) \left(\frac{1}{u} - \hat{c} u^{-1/\hat{\eta}}\right) \left(1 - \frac{2}{u} + \hat{c} u^{-1/\hat{\eta}}\right),$$

and yields the final result

$$\hat{\eta} = \frac{1}{n_u} \sum_{i=1}^{n_u} \log\left(\frac{z_i}{u}\right).$$

This value of $\hat{\eta}$ is the same as for the Ledford and Tawn (1997) joint tail model as given in equation (2.5.3).

# Appendix B

# Tail chain estimation algorithms

Algorithm 1 gives the tail chain generation method for asymptotically independent Markov chains. An exceedance of $v$ is generated from a GPD($\sigma_v$, $\xi$) and the chain is stepped forward using equation (3.3.5) by sampling from the non-parametric estimate of $\hat{G}$ given in equation (3.3.7). Particular care must be taken since negative values of the transformed tail chain, i.e. $T(X_t) < 0$, can lead to problems (since $\beta \in (-\infty, 1)$). Since the margins follow a Laplace distribution negative values correspond to values below the median and hence outside the tail region, so following a negative value all further chain values are set to zero and do not affect the cluster properties for the high levels of interest.

A non-parametric tail chain simulation scheme based upon the conditional extremes approach provides an alternate method to generate tail chains with asymptotic dependence. As outlined in Section 3.3.3 this method is a special case of the conditional extremes method with $\alpha = 1$ and $\beta = 0$. It has been shown in equation (3.3.9) that the non-parametric estimate of the distribution $G$ is given as the empirical distribution function of the set of differences between the transformed chain at times $t$ and $t + 1$ given $X_t > u$. The chain is stepped forward by sampling a value from the set of differences with replacement and adding the value to the current value of the chain.

---

**Algorithm 1:** Simulation scheme to generate tail chain using conditional extremes approach

**Input**: Dependence parameters $(\alpha, \beta)$ and non-parametric distribution $\hat{G}$

1 Set threshold $v$ and simulate exceedance using $\text{GPD}(\sigma_v, \xi)$ distribution;

2 Set exceedance as $X_1^*$;

3 **for** $i$ *in* $1 : k - 1$ **do**

4   Make draw $Z_i^*$ with replacement from $\hat{G}$;

5   Set $X_{i+1}^* = T^{-1}\left(\alpha T\left(X_i^*\right) + T\left(X_i^*\right)^\beta Z_i^*\right)$;

6 **end**

**Output**: Tail chain $X_1^*, \ldots, X_k^*$ with dependence structure given by $(\alpha, \beta)$

---

The parametric tail chain simulation scheme has a similar form to Algorithm 1. We start with dependence parameter $\gamma$ instead of $(\alpha = 1, \beta = 0)$ and instead of simulating from $\hat{G}$ in step 4 we simulate $U_i$ from Uniform(0,1) distribution, set

$$Z_i^* = -\frac{\gamma}{\sigma_u} \log\left(U_i^{1/(\gamma-1)} - 1\right),$$

and replace step 5 with $X_{i+1}^* = X_i^* + Z_i^* \left[\sigma_u + \xi\left(X_i^* - u\right)\right]_+$.

For a forward and backward simulation strategy it is necessary to know the peak value of a cluster, here denoted $M = \eta$ where $\eta > v$. By setting $X_0^* = \eta$ a tail chain of length $k$ can be simulated forwards using $(\hat{\alpha}_f, \hat{\beta}_f)$, the estimates of the conditional extremes dependence parameters for the forward chain, and the conditional extremes tail chain simulation scheme outlined above. One difference requires that a chain $X_0^*, \ldots, X_k^*$ be discarded if $X_j^* > \eta$ for any $j = 1, \ldots, k$. An additional fit of the conditional extremes model must be made for $X_t \mid X_{t+1} > u$ before the backward simulation step to obtain $(\hat{\alpha}_b, \hat{\beta}_b)$, the estimates of the conditional extremes dependence parameters for the backward chain. If the Markov chain is time-reversible $\alpha_b = \alpha_f$ and $\beta_b = \beta_f$. For the backward simulation a tail chain is constructed using $(\hat{\alpha}_b, \hat{\beta}_b)$ with

the same rejection criteria if $\eta$ is exceeded. By combining the forward and backward simulated tail chains a cluster with peak value $\eta$ is generated; see Algorithm 2.

---

**Algorithm 2:** Simulation scheme to generate realisation of cluster with peak value $\eta$ using conditional extremes approach

---

**Input**: Dependence parameters $(\alpha_f, \beta_f, \alpha_b, \beta_b)$ and non-parametric distributions $\hat{G}_f$ and $\hat{G}_b$

**1** Set the peak value $\eta$ of a cluster as $X_0^*$;

**2** **for** $i$ *in* $1 : k$ **do**

**3** $\qquad$ Make draw $Z_f^*$ with replacement from $\hat{G}_f$;

**4** $\qquad$ Set $X_i^* = T^{-1}\left(\alpha T\left(X_{i-1}^*\right) + T\left(X_{i-1}^*\right)^{\beta} Z_f^*\right)$;

**5** $\qquad$ **if** $X_i^* > X_0^*$ **then**

**6** $\qquad\qquad$ Discard current forward chain and return to step 2 ;

**7** $\qquad$ **end**

**8** **end**

**9** **for** $i$ *in* $1 : k$ **do**

**10** $\qquad$ Make draw $Z_b^*$ with replacement from $\hat{G}_b$;

**11** $\qquad$ Set $X_{-i}^* = T^{-1}\left(\alpha_b T\left(X_{-(i-1)}^*\right) + T\left(X_{-(i-1)}^*\right)^{\beta_b} Z_b^*\right)$;

**12** $\qquad$ **if** $X_{-i}^* > X_0^*$ **then**

**13** $\qquad\qquad$ Discard current backward chain and return to step 9;

**14** $\qquad$ **end**

**15** **end**

**16** Generate a cluster replicate $X^* = (X_{-k}^*, \ldots, X_0^*, \ldots, X_k^*)$;

**Output**: One realisation of a cluster $X^*$ with maximum $\eta$ and dependence structure given by $(\alpha_f, \beta_f, \alpha_b, \beta_b)$

---

# Appendix C

# Pool adjacent violators algorithm

Using differencing formulas, as in equation (3.2.3), with inputs evaluated by Monte Carlo methods from tail chains can result in negative estimates of $\pi(i)$ for certain values of $i$. Although $\theta^{(i)} \geq \theta^{(i+1)}$ for all $i$, as only a finite number of chains can be simulated by Monte Carlo, Algorithms from Appendix B provide estimates $\tilde{\theta}^{(i)}$ and $\tilde{\theta}^{(i+1)}$ which do not necessarily satisfy this constraint. Thus $\tilde{\pi}(i) < 0$ if $\tilde{\theta}^{(i)} < \tilde{\theta}^{(i+1)}$. The problem is more prevalent for large values of $i$ since the tail chain simulation scheme generates very few chains with such a large number of exceedances and hence the Monte Carlo variation is large relative to the difference between $\theta^{(i)}$ and $\theta^{(i+1)}$. A solution to this problem, to ensure that estimates $\tilde{\pi}(i)$ of $\pi(i)$ satisfy $\tilde{\pi}(i) \geq 0$ is to use the pool adjacent violators (PAV) algorithm (Robertson et al., 1988). The PAV algorithm generates a monotonically decreasing estimate of $\theta^{(i)}$ and in turn gives non-negative estimates for $\pi$ and $\pi_C$. This is achieved by checking whether $\tilde{\theta}^{(i+1)} \leq \tilde{\theta}^{(i)}$ for all $i = 1, \ldots, n$, if not these values are averaged, i.e.

$$\theta_*^{(i)} = \theta_*^{(i+1)} = (\tilde{\theta}^{(i+1)} + \tilde{\theta}^{(i)})/2,$$

and we check whether $\tilde{\theta}^{(i-1)} \geq \theta_*^{(i)}$. If not pooling is continued until decreasing monotonicity is satisfied. The algorithm can lead to ties in consecutive estimated $\theta^{(i)}$ values resulting in $\tilde{\pi}(i) = 0$ but avoids the situation where $\tilde{\pi}(i) < 0$.

# Appendix D

# $k$th-order tail chain estimation algorithm

Algorithm 3 gives the tail chain generation method for asymptotically independent Markov chains with $k$th-order dependence structure. An exceedance of $v$ is generated from a Exp(1) and the chain is stepped forward using equations (4.4.2) and (4.4.3) by sampling from the conditional distribution $G_{k|1:k-1}$ given in equation (4.2.2) where the joint distribution $G_{1:k}$ is taken to have the form in equation (4.2.1). Particular care must be taken since negative values of the transformed tail chain, i.e. $X_j^* < 0$ for $j = 1, \ldots, m$, can lead to problems (since $\beta_k \in (-\infty, 1)$). Since the margins follow a Laplace distribution negative values correspond to values below the median, hence outside the tail region, so following a negative value all further chain values are set to zero and do not effect the cluster properties for the high levels of interest.

---

**Algorithm 3:** Simulation scheme to generate tail chain of length $m$ with $k$th-order temporal dependence structure using conditional extremes approach

---

**Input**: Dependence parameters $(\alpha_j, \beta_j)$ for $j = 1, \ldots, k$ and non-parametric estimate $\hat{G}_{1:k}$ to distribution $G_{1:k}$.

**1** Set threshold $v$, simulate exceedance using Exp(1) distribution;

**2** Set exceedance as $X_0^*$;

**3** Make draw $Z_{1|0}^*$ from kernel density estimate of $\hat{z}_1^{(i)}$, i.e. $\hat{G}_1$;

**4** Set $X_1^* = \alpha_1 X_0^* + (X_0^*)^{\beta_1} Z_{1|0}^*$;

**5 for** $j$ *in* $2 : m - 1$ **do**

**6**    **if** $j \leq k$ **then**

**7**       Calculate weights $w_i$ using equation (4.3.3) and use to make draw $Z_{j|0:j-1}^*$ from $\hat{G}_{j|1:j-1}$;

**8**       Set $X_j^* = \alpha_j X_0^* + (X_0^*)^{\beta_j} Z_{j|0:j-1}^*$;

**9**    **end**

**10**    **else**

**11**       Calculate weights $w_i$ using equation (4.3.3) and use to make draw $Z_{j|j-k+1:j-1}^*$ from $\hat{G}_{k|1:k-1}$;

**12**       Set $X_j^* = \alpha_k X_{j-k}^* + \left(X_{j-k}^*\right)^{\beta_k} Z_{j|j-k+1:j-1}^*$;

**13**    **end**

**14 end**

**Output**: Tail chain $X_0^*, \ldots, X_{m-1}^*$ on Laplace margins with $k$th-order temporal dependence structure

---

# Bibliography

J. Abaurrea, J. Asin, A. Cebrian, and A. Centelles. Modeling and forecasting extreme hot events in the central Ebro valley, a continental-Mediterranean area. *Global and Planetary Change*, 57(1-2):43–58, 2007.

L. V. Alexander and J. M. Arblaster. Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *International Journal of Climatology*, 29:417–435, 2009.

L. V. Alexander, P. Uotila, and N. Nicholls. Influence of sea surface temperature variability on global temperature and precipitation extremes. *Journal of Geophysical Research*, 114(D18):D18116, 2009.

P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. pages 1–24, 2015.

S. Asseng, F. Ewert, and C. Rosenzweig. Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3:827–832, 2013.

F. B. Avila, S. Dong, K. P. Menang, J. Rajczak, M. Renom, M. G. Donat, and L. V. Alexander. Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for southeast Australia. *Weather and Climate Extremes*, 9:6–16, 2015.

P. Bortot and J. A. Tawn. Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867, 1998.

P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting.* Springer, 2006.

J. Caesar, L. Alexander, and R. Vose. Large-scale changes in observe daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research: Atmospheres*, 111(5):1–10, 2006.

R. E. Chandler and S. Bate. Inference for clustered data using the independence loglikelihood. *Biometrika*, 94:167–183, 2007.

C. Chatfield. *The Analysis Of Time Series: An Introduction.* CRC Press, 2003.

N. Christidis. Detection of changes in temperature extremes during the second half of the 20th century. *Geophysical Research Letters*, 32(20):L20716, 2005.

S. G. Coles. Regional modelling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society: Series B*, 55(4):797–816, 1993.

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

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

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

R. A. Davis, C. Kluppelberg, and C. Steinkohl. Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B*, 75 (5):791–819, 2013.

A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468:581–608, 2012.

A. C. Davison and R. L. Smith. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B*, 52(3):393–442, 1990.

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

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction.* Springer, New York, 2006.

C. Dombry, F. Éyi-Minko, and M. Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124, 2013.

G. C. Donaldson, W. R. Keatinge, and R. D. Saunders. Cardiovascular responses to heat stress and their adverse consequences in healthy and vulnerable human populations. *International Journal of Hypothermia*, 19(3):225–235, 2003.

O. J. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.

D. J. Dupuis. Modeling waves of extreme temperature: the changing tails of four cities. *Journal of the American Statistical Association*, 107(497):24–39, 2012.

E. F. Eastoe and J. A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, 58(1):25–45, feb 2009.

E. F. Eastoe and J. A. Tawn. Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55, 2012.

P. Falloon and R. Betts. Climate impacts on European agriculture and water management in the context of adaptation and mitigation - The importance of an integrated approach. *Science of the Total Environment*, 408(23):5667–87, 2010.

P. Falloon, A. Challinor, S. Dessai, L. Hoang, J. Johnson, and A. Koehler. Ensembles and uncertainty in climate change impacts. *Frontiers in Environmental Science*, 2: 1–7, 2014.

C. A. T. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B*, 65(2):545–556, 2003.

E. M. Fischer and C. Schär. Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience*, 3(6):398–403, 2010.

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

P. Frich, L. V. Alexander, and P. M. Della-Marta. Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, 19:193–212, 2002.

E. J. Gumbel. *Statistics of extremes*. New York: Columbia University Press, 1958.

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

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

A. G. Henriques and M. J. J. Santos. Regional drought distribution model. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 24(1-2): 19–22, 1999.

T. Hsing. On the extreme order statistics for a stationary sequence. *Stochastic Processes and their Application*, 29:155–169, 1988.

R. Huser and A. C. Davison. Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B*, 76(2):439–461, 2014.

IPCC. *Special Report on Emissions Scenarios*. Cambridge University Press, 2000.

IPCC. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge University Press, 2012.

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

P. Jonathan, K. Ewans, and D. Randell. Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, 79:22–31, 2013.

D. A. Jones and B. C. Trewin. On the relationships between the El Nino-Southern Oscillation and Australian land surface temperature. *International Journal of Climatology*, 20:697–719, 2000.

M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433): 401–407, 1996.

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

J. Kenyon and G. C. Hegerl. Influence of modes of climate variability on global temperature extremes. *Journal of Climate*, 21(15):3872–3889, 2008.

A. Koehler, A. Challinor, E. Hawkins, and S. Asseng. Influences of increasing temperature on Indian wheat: quantifying limits to predictability. *Environmental Research Letters*, 8(3):034016, 2013.

C. Koppe, S. Kovats, G. Jendritzky, and B. Menne. *Heat-Waves: Risks and Responses*. Number 2. World Health Organization, 2004.

S. Kotz and S. Nadarajah. *Extreme Value Distributions: Theory and Applications.* Imperial College Press, 2000.

M. R. Leadbetter. Extremes and local dependence in stationary sequences. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(2):291–306, 1983.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer Verlag, 1983.

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

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

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

J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. Technical report, 2001.

D. Maraun, F. Wetterhall, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, V. K. C. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich. Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48:1–38, 2010.

G. M. Martin. The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, 4(3):723–757, 2011.

T. B. Mckee, N. J. Doesken, and J. Kleist. The relationship of drought frequency and duration to time scales. In *Conference on Applied Climatology*, number January, pages 17–22, 1993.

S. Min, W. Cai, and P. Whetton. Influence of climate variability on seasonal extremes over Australia. *Journal of Geophysical Research: Atmospheres*, 118(2):643–654, 2013.

A. K. Mishra and V. P. Singh. A review of drought concepts. *Journal of Hydrology*, 391(1-2):202–216, 2010.

R. B. Nelson. *An Introduction to Copulas*. Springer, 2nd edition, 2007.

M. Nitschke, G. R. Tucker, A. L. Hansen, S. Williams, Y. Zhang, and P. Bi. Impact of two recent extreme heat episodes on morbidity and mortality in Adelaide, South Australia: a case-series analysis. *Environmental Health*, 10(1):42–51, 2011.

P. J. Northrop and P. Jonathan. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22:799–809, 2011.

G. L. O'Brien. Extreme values for stationary and Markov sequences. *The Annals of Probability*, 15(1):281–291, 1987.

W. Palmer. *Meteorological drought*. US Department of Commerce Weather Bureau, 1965.

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

I. Papastathopoulos, K. Strokorb, J. A. Tawn, and A. Butler. Extreme events of Markov chains. *In preparation for Journal of Applied Probability*, 2015.

M. Pascal, V. Wagner, A. Le Tertre, K. Laaidi, C. Honoré, F. Bénichou, and P. Beaudeau. Definition of temperature thresholds: the example of the French heat wave warning system. *International Journal of Biometeorology*, 57(1):21–29, 2013.

R. Perfekt. Extreme value theory for a class of Markov chains with values in Rd. *Advances in Applied Probability*, 29(1):138–164, 1997.

S. E. Perkins and L. V. Alexander. On the measurement of heat waves. *Journal of Climate*, 26(13):4500–4517, 2013.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 1975.

B. J. Reich, B. A. Shaby, and D. Cooley. A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *Journal of Agricultural, Biological, and Environmental Statistics*, 19:119–135, 2014.

T. Robertson, F. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. Wiley, New York, 1988.

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

M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.

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

J. Segers. Functionals of clusters of extremes. *Advances in Applied Probability*, 35(4): 1028–1045, 2003.

S. G. Self and K. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82:605–610, 1987.

M. Sibuya. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11:195–210, 1960.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

R. L. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–377, 1989.

R. L. Smith. Max-stable processes and spatial extremes. *Unpublished manuscript, University of Surrey*, pages 1–32, 1990.

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

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

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

M. Stefanon, F. D'Andrea, and P. Drobinski. Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, 7(1):014023, 2012.

P. A. Stott, D. A. Stone, and M. R. Allen. Human contribution to the European heatwave of 2003. *Nature*, 432:610–614, 2004.

L. M. Tallaksen and H. V. Lanen. *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*. Elsevier, 2004.

A. Tancredi, C. Anderson, and A. O'Hagan. Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9(2):87–106, 2006.

J. A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75 (3):397–415, 1988.

J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77(2): 245–253, 1990.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93:485–498, 2012.

D. G. Victor and C. F. Kennel. Ditch the 2ºC warming goal. *Nature*, 514:30–31, 2014.

J. Wadsworth, J. A. Tawn, A. Davison, and D. M. Elton. Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.

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

J. L. Wadsworth and J. A. Tawn. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):543–567, 2012b.

J. L. Wadsworth and J. A. Tawn. A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714, 2013.

C. Wang and J. Picaut. Understanding ENSO physics - a review. *Geophysical Monograph Series*, 147:21–48, 2004.

Y. Wang and S. A. Stoev. Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 43(2):461–483, 2011.

J. Wheatley. Visualizing drought, 2010. URL `http://joewheatley.net/visualizing-drought/`.

T. R. Wheeler, P. Q. Craufurd, and R. H. Ellis. Temperature variability and the yield of annual crops. *Agriculture, Ecosystems and Environment*, 82:159–167, 2000.

S. Yun. The extremal index of a higher-order stationary Markov chain. *Annals of Applied Probability*, 8(2):408–437, 1998.

S. Yun. The distributions of cluster functionals of extreme events in a dth-order Markov chain. *Journal of Applied Probability*, 37(1):29–44, 2000.