<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Bayesian spatial modelling of climate extremes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors(s)</strong></td>
<td>O'Sullivan, J. (John)</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2020</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>University College Dublin. School of Biology and Environmental Science</td>
</tr>
<tr>
<td><strong>Item record/more information</strong></td>
<td><a href="http://hdl.handle.net/10197/11650">http://hdl.handle.net/10197/11650</a></td>
</tr>
</tbody>
</table>

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information, please see the item record link above.
Bayesian Spatial Modelling of Climate Extremes

John O’Sullivan
12252927

This thesis is submitted to University College Dublin in fulfillment of the requirements for the degree of

Doctor of Philosophy

School of Mathematics and Statistics
Head of School: Prof. Brendan Murphy
Supervisors: Prof. Andrew Parnell and Dr. Conor Sweeney

Doctoral Studies Panel Members:
Dr. Adrian O’Hagan, Dr. Emily Gleeson
and Assoc. Prof. Xuefeng Cui

November 2019
Acknowledgements

This thesis would not have been possible without the guidance of my supervisors, Andrew and Conor. I would like to express my sincere gratitude to them. Throughout this learning experience, they have shown immeasurable patience, expertise, motivation, and enthusiasm.

My sincere thanks also goes to my husband Séamus, and my parents John and Betty, for their encouragement and support throughout.

I would also like to thank all my family, friends, and everyone at UCD who have had to put up with me too, and shown me nothing but support and patience.
Abstract

This thesis is primarily concerned with determining effective and efficient methods to model spatial datasets of climate extremes. Climate extremes impact our technology, our infrastructure, the environment, and humans ourselves. Faced with the uncertainties and challenges of human-induced climate change, it is now more important than ever to understand the behaviour of climate extremes; whether that is for data we have already observed, or for data we have generated from physical or statistical models. Existing observations of extreme events are, by definition, rare, which makes any analysis and interpretation of this data more challenging than other analyses. Extrapolation beyond the observed extremes is necessary in order to plan for events worse than those already observed. This extrapolation needs a sound theoretical framework as mitigation and adaptation planning often involves practical and financial risks and implications. The framework applied should be flexible enough to model different variables, incorporate prior information, result in useful predictions of practical benefit, and finally be adaptable to model datasets across a dense spatial domain. It is these aims and objectives which guided and informed the research undertaken during the course of this thesis.

The first study I undertook involved using dynamically downscaled climate model output from CMIP3 and CMIP5. From this, I analysed the temperature changes projected over Ireland for the mid-21st century under different future climate change scenarios. This analysis was conducted by examining changes between a reference control period (1981-2000) and the period of interest (2041-2060), for each particular future scenario. I found that annual mean temperatures were projected to rise by
between 0.4°C and 1.8°C by the mid-century. On a seasonal basis, projected temperature changes differed by forcing scenario. Some scenarios saw future summers with the largest projected warming; others projected future winters to warm the most. Warming patterns were spatially diverse too, with the south and southeast of Ireland showing the greatest warming. An investigation into the projected change in temperature extremes found that in general there was greater projected changes in the number of cold nights and the number of hot days than for mean temperatures.

In the next study undertaken, I applied extreme value theory (EVT) in a Bayesian spatial hierarchical framework in order to model extremes of significant wave heights off the west coast of Ireland for the period 1979-2012. This model was fitted using Markov chain Monte Carlo (MCMC) algorithms. Return level surfaces across the domain and return level curves at specific sites were used in model evaluation. I found that the highest extremes of significant wave height were to be expected off the west coast of Ireland roughly between 53°N and 54.5°N, with 100-year levels close to 17 m. A comparison of the Bayesian spatial model with a simpler maximum likelihood site-specific approach found that the former resulted in smoother surfaces of posterior parameters and return level maps. In addition, there was less uncertainty associated with the Bayesian spatial model, while empirical estimates overlaid on the return level curves at selected sites showed a satisfactory fit to the data.

In the final study undertaken, I extended the Bayesian spatial hierarchical framework using a dimension reduction technique called predictive processes. This extension was necessary in order to model a spatially dense dataset of extremes of daily maximum temperature anomalies over Dublin, Ireland, for the period 1981-2010. Return level surfaces across the domain and return level curves at specific sites were used in model evaluation. The results included a posterior median 20-year return level surface for anomalies of maximum temperature ranging from 8°C to 10°C across
the domain. Highest values were observed to the west of the domain, at locations furthest inland. The corresponding 100-year return level surface ranged from 8°C to 10.7°C, with an upperbound of 12.7°C when credible intervals were examined. Additional analysis involved placing more recent extremes (2011-2018) from several synoptic stations across the domain in the context of the model results. Including this observational data in the analysis showed an increase in the frequency of extreme anomalies for this period, but not in their severity.

The methods presented in this thesis can be readily adapted to any spatially-continuous dataset. One advantage to using a Bayesian framework is the incorporation of prior information, which I showed leads to a reduction in the uncertainty of quantities such as return levels, and thus provides information on extremes which is more useful for practical applications. The main advantage of the predictive processes approach is the ability to fit a spatial model which would otherwise be too computationally expensive to fit. This allowed me to achieve results which would not have been possible by fitting the full model to the dense dataset.
Contents

1 Introduction 1
   1.1 Motivation for this thesis .............................. 1
   1.2 Overview of contributions ................................ 3
   1.3 Climate models ........................................... 4
   1.4 Introduction to the datasets used ......................... 10
   1.5 Remainder of thesis ...................................... 19
   References .................................................. 21

2 Analysis of climate model output of Irish temperatures for the mid-21st century 26
   2.1 Past research on temperature change in Ireland ............. 27
   2.2 Climate Model Validation .................................. 30
   2.3 Preliminary analysis ....................................... 35
   2.4 Results .................................................. 39
   2.5 Discussion and Conclusions ................................ 48
   References .................................................. 52

3 Using Bayesian spatial hierarchical modelling and Extreme Value Theory to predict extreme sea states 55
   3.1 Statistical background ..................................... 56
   3.2 Model details for the waves dataset ......................... 67
   3.3 Results .................................................. 76
   3.4 Discussion and Conclusions ................................ 86
   References .................................................. 90
   Appendix A .................................................. 95
## Contents

4 Bayesian spatial extreme value analysis of maximum temperatures in County Dublin, Ireland

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Overview of study</td>
<td>116</td>
</tr>
<tr>
<td>4.2 Previous relevant research</td>
<td>117</td>
</tr>
<tr>
<td>4.3 Dataset of daily maximum temperatures</td>
<td>120</td>
</tr>
<tr>
<td>4.4 Statistical background</td>
<td>125</td>
</tr>
<tr>
<td>4.5 Model details for the temperature dataset</td>
<td>129</td>
</tr>
<tr>
<td>4.6 Results</td>
<td>139</td>
</tr>
<tr>
<td>4.7 Discussion and Conclusions</td>
<td>145</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>148</td>
</tr>
<tr>
<td>References</td>
<td>148</td>
</tr>
<tr>
<td>Appendix A</td>
<td>153</td>
</tr>
<tr>
<td>Appendix B</td>
<td>160</td>
</tr>
</tbody>
</table>

5 Conclusions and further work

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Summary</td>
<td>180</td>
</tr>
<tr>
<td>5.2 Overview of contributions</td>
<td>186</td>
</tr>
<tr>
<td>5.3 Future research and challenges</td>
<td>186</td>
</tr>
<tr>
<td>References</td>
<td>193</td>
</tr>
</tbody>
</table>
Statement of Original Work

I hereby certify that the submitted work is my own work, was completed while registered as a candidate for the degree stated on the Title Page, and I have not obtained a degree elsewhere on the basis of the research presented in this submitted work.
Publications

The material in Chapters 2 and 3 has been published in peer-reviewed journals. The material in Chapter 4 has been re-submitted to a peer-reviewed journal following revisions to the original submission.

Journal Publications


Currently under review following re-submission

- O’Sullivan, J., Sweeney, C., & Parnell, A. C. Bayesian spatial extreme value analysis of maximum temperatures in County Dublin, Ireland. Submitted to: *Environmetrics*
My research resulted in a co-authored peer-reviewed publication that is relevant to my PhD but does not appear as a chapter in this thesis.

**Additional journal publication**


I also contributed to a chapter in a climate change research report undertaken by Met Éireann, Ireland’s meteorological service.

**Contribution to climate change research report**

Chapter 1

Introduction

This thesis is concerned with the following three broad topics:

- the analysis and evaluation of dynamically downscaled climate model output;
- the use of Extreme Value Theory (EVT) in a Bayesian spatial hierarchical framework to model extremes over an irregular grid;
- and the extension of a Bayesian spatial hierarchical EVT framework using predictive processes for dimension reduction of a spatially dense dataset.

This introductory chapter includes: the motivation for this research; an account of the main contributions of the work done; an overview of climate models; an introduction to the three main datasets used; and an outline of the content of the remaining chapters.

1.1 Motivation for this thesis

The month of July in 2019 was on par with the hottest month on record globally, according to the Copernicus Climate Change Service (C3S), a service implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF). The report (Copernicus Climate Change Service, 2019a) shows that, compared to the climatological reference period of 1981-2010, the global average temperature for July was about 0.56°C above average. This is approximately 1.2°C above the pre-industrial
1.1 Motivation for this thesis

level as defined by the Intergovernmental Panel on Climate Change (IPCC) and is just marginally \(0.04^\circ C\) warmer than July 2016, which was the previous warmest month on record. The same report reminds readers that the past four years (2015-2018) have been the four warmest years on record. Subsequent reports for August and September 2019 show that this trend has continued: every month so far in 2019 (January to September inclusive) has ranked among the four warmest for the month in question (Copernicus Climate Change Service, 2019b). In their climate summary for Ireland for July 2019, the Irish meteorological service (Met Éireann) report that all mean monthly air temperatures across the country were between \(0.3^\circ C\) to \(1.2^\circ C\) above their long-term average values from the base period of 1981-2000 (Met Éireann, 2019).

In the Special Report of the Intergovernmental Panel on Climate Change on managing the risks of extreme events, the authors state that there is evidence that some extremes have changed as a result of anthropogenic influences, which include increases in atmospheric concentrations of greenhouse gases. They claim that it is very likely that there has been an overall decrease in the number of cold days and nights, and an overall increase in the number of warm days and nights at the global scale, and that it is likely that these changes have also occurred at the continental scale in Europe (IPCC, 2012).

In the same IPCC report, the authors discuss how the behaviour of the atmosphere is highly interlinked with that of the hydrosphere (clouds, oceans, seas, rivers, lakes etc.). They cite a study which highlights that extreme atmospheric events (such as large cyclonic storms) may cause other rare events such as coastal flooding and damage due to severe wave action (Xie et al., 2004). Ireland has experienced quite a dramatic history of extreme wave events, from reports of the famous Night of the Big Wind in 1839 (when a storm surge is believed to have occurred, based on contemporaneous statements of waves breaking over the tops of the Cliffs of Moher - which vary in height from 120 m to 214 m) to the largest wave ever recorded in Irish waters (a 26.1 m wave recorded by the Kinsale Energy gas platform during Storm Ophelia on the 16 October 2017) (O’Brien et al., 2018).
1.2 Overview of contributions

In order to make informed decisions to help mitigate or avoid the worst effects of climate change it is vitally important for us to better understand the past and present data available to us, and that we understand the principles and the limitations of the models we use to project future data. It is this need which motivates the studies undertaken in this thesis. Our understanding starts from exploring the data we have (descriptive statistics) to testing and using statistical models in order to make inference (based on the data we have) to find out about the latent processes underlying this observed data. This approach is even more important when trying to further our understanding of extreme events, and possible changes in their severity and frequency, as these involve, by definition, very small amounts of data.

1.2 Overview of contributions

This thesis makes a number of original contributions to the existing literature. Specifically:

- In Chapter 2, I detail the first study (at such a high resolution) of dynamically downscaled climate models over Ireland. I present analysis of temperature projections for the mid-21st century. I analyse changes in the means and the extreme temperatures, using simple statistical techniques to investigate the projected changes of the daily mean, minimum, and maximum temperature across Ireland.

- In Chapter 3, I incorporate EVT into a Bayesian spatial hierarchical framework, in order to model extreme waves off the west coast of Ireland. A Bayesian approach allows prior belief and physical constraints to be incorporated into our model in a natural way, which leads to reduced uncertainties in parameter estimates. Our work improves on the previous non-spatial research in this area, by borrowing information from nearby locations to model the latent spatial process underlying the extreme behaviour of waves. This borrowing of information from across the spatial domain is particularly important when modelling extremes, due to the scarcity of available data.

- In Chapter 4, I overcome the computational difficulties in applying a Bayesian
1.3 Climate models

In this section, climate models are introduced. I provide an overview on what a climate model does; then the differences between GCMs and RCMs are presented; finally I discuss the various sources of uncertainty arising from climate model input, running, and output.

1.3.1 Preliminary note: Weather vs. Climate

As the axiom goes, climate is what we expect, but weather is what we get (first evidence of this in print is by Herbertson (1901)). Descriptions of the weather involve details such as the rainfall on a particular day, or the maximum temperature reached. Climate, on the other hand, involves describing more long-term statistics, such as the expected annual precipitation or the mean seasonal value of the daily maximum temperature. These long-term statistics are ideally calculated over a thirty-year period, resulting in climate normals. Thirty years is sufficiently long enough to filter out any interannual variation, and is sufficiently short enough to uncover longer climatic trends (World Meteorological Organization, 2017). That is, a 30-year period is long enough to ensure an abnormally cold winter, for example, will not adversely skew the climate normal obtained, while it is short enough to allow longer-term trends (such as a gradual warming) to emerge.

1.3.2 Principles of climate models

Climate models are built in a similar manner to numerical weather prediction (NWP) models, which use mathematical models of the atmosphere and oceans for short-range
1.3 Climate models

weather forecasting based on current weather conditions. However, some processes
that can be ignored over the timescales with which weather is concerned, need to
be accounted for in climate models (McGuffie and Henderson-Sellers (2014), p 70).
Examples of these include the transfer of heat and energy between the oceans and
the atmosphere, and changes in deep water formation.

Climate models are models of the climate system which are based on mathematical
and physical principles. In these models, the globe is divided into three-dimensional
boxes, chosen in order to cover the world both horizontally and vertically (McGuffie
and Henderson-Sellers (2014), p 282). Climate models are the only tools currently
available to describe the ‘complex set of processes’ that will determine ‘future climate
change’, both globally and regionally (Murphy et al., 2004).

A climate model works by simulating the evolution of the ‘three-dimensional state
of the atmosphere, ocean, cryosphere, and land surface’ (Lambert and Boer, 2001).
The values of the seven quantities temperature, water, pressure, humidity and the
three components of wind velocity at each time-step and in each box are then calcu-
lated - statistics obtained from these values define the model climate.

1.3.3 GCMs and RCMs

Climate models can be either global (global climate models - GCMs) or regional
(regional climate models - RCMs) in extent. GCMs resolve the governing equations
in boxes which cover the globe horizontally and extend vertically upwards into the
atmosphere and downwards into the ocean. GCMs are very computationally expensive
due to their complexity and size. They are typically run at a lower resolution than
RCMs, i.e. with larger and therefore fewer boxes. RCMs focus their attention on
a particular region, e.g., over Europe only (see Figure 1.1). They can therefore be
run with a significantly higher resolution (smaller and therefore more boxes in a given
area) than GCMs.
1.3 Climate models

Figure 1.1: Above is a schematic of a GCM on the left, where the globe is divided into boxes, horizontally and vertically. The RCM on the right uses data from the GCM at its boundaries, and has a finer resolution (more and smaller boxes). (Credit: WMO)

Since the weather at any point depends on the points around it, which in turn have weather depending on the points around them and so on, it follows that RCMs need information about what is happening at their spatial boundaries (e.g., at the edge of Europe and along North Africa, as illustrated in Figure 1.1). RCMs need this information at each time-step, as otherwise they will fail to run. Providing information to an RCM at its spatial boundaries allows, for example, storms and other weather systems to enter and exit the area covered by the RCM. These boundary conditions are provided by a GCM, meaning that a GCM needs to be run before an RCM.

1.3.4 Uncertainty in climate models

There are many factors which contribute to the uncertainty for which all climate models attempt to account (Fronzek et al., 2012). Three of these are discussed below.

Firstly, there is uncertainty due to the natural variability of the climate system (aleatoric uncertainty). A common method that attempts to deal (in part) with this uncertainty typically involves the creation of an ensemble: i.e. combining the results of a model run several times with different initial conditions (multiple realisations for

---

1https://public.wmo.int/en/our-mandate/climate
1.3 Climate models

one GCM), or from several different models (GCM/RCM pairings) or a combination of both (Déqué et al., 2007). However, this introduces additional uncertainties. Different GCM/RCM pairings may result in different variability, while the timing of natural variability can be different in different GCM realisations. These parallel runs of a climate model, which only differ in formulation due to the state of the climate at the first step, are called ensembles. They may eventually produce different climatologies, which can be considered separately as equally valid projections, or combined to form one climatology.

Secondly, uncertainties arise due to the formulation of the models themselves (epistemic uncertainty). There are a lot of processes which drive our climate. Some of these are quite well understood (e.g., warming due to carbon dioxide absorbing and emitting radiation), whereas there is significant uncertainty surrounding others (e.g., cloud reflectivity (or albedo) and its effects) (Forster et al., 2007). Different climate models therefore have different ways to mathematically express these processes, their interactions, and their complex feedbacks. Epistemic uncertainty can be accounted for by using several different models with different parameterisation schemes and dynamical cores, which also results in an ensemble. (Note that accounting for these two uncertainties often leads to further issues because, for example, ensemble members from related climate models will not be statistically independent.)

The relative coarseness of GCMs, with typical grid sizes of the order of 100 km, leads to related uncertainties in future regional climate. Downscaling the data over the region of interest can reduce this problem. The resulting higher resolution output is more useful for focused climate impact studies. There are two general approaches to downscaling data: statistical and dynamical. Statistical downscaling involves fitting a statistical model to the GCM output to find changes in climate at a local scale. It is quick and inexpensive to perform (Yang et al., 2012). Dynamical downscaling uses RCMs to focus on a particular region (see Figure 1.2) in order to generate projections of future climate at a higher resolution. The GCM outputs are used as boundary conditions to drive the RCM (in contrast to the freely evolving original GCM). The higher resolution allows for a better representation of coastlines, general topography (Chan et al., 2013) and land use. The physically based RCMs explicitly resolve more smaller-scale transient-dynamical features of atmospheric flow (e.g., squalls) than the
1.3 Climate models

coarser GCMs (Wilby et al., 2002).

Figure 1.2: Different model resolutions over Britain and Ireland are shown above. Successively higher resolutions [from 125 km (a) to 50 km (b) to 18 km (c) to 4 km (d)] allow coastlines and topography to be modelled in greater detail, and smaller-scale atmospheric features to be resolved.

In addition to the uncertainty concerning various atmospheric processes, uncertainty about the composition of the atmosphere leads to further difficulties. The composition of the atmosphere changes over time due to both natural causes (e.g., volcanic explosions) and anthropogenic factors (e.g., fossil fuel combustion). Different greenhouse gases and aerosols (suspended solid or liquid particles) have different average lifetimes, spatial distributions, direct and indirect warming effects on the atmosphere, and effects on cloud formation. Knowledge of the exact composition of the atmosphere is quite complicated, both for past periods and the present. This
1.3 Climate models

creates a problem for climate models, as the composition of the atmosphere needs to be defined in order for changes in temperature, for example, to be calculated.

There is also uncertainty concerning the future atmospheric composition, which affects the radiative balance of the earth. If the earth is in radiative equilibrium, then incoming solar energy is balanced by an equal flow of heat outwards to space. Under this condition, then global temperatures will be relatively stable. However, changes in the atmospheric composition affects and upsets this balance, sometime in poorly understand ways. Rates of change in quantities such as future fossil fuel consumption, future volcanic explosions, wildfires, adoption of green technology, world population, arable land etc. cannot be known in advance. Yet they and many other complex things affect climate evolution, and so must somehow be accounted for in climate models. In order to account for this uncertainty, the IPCC previously recommended using six pre-defined scenarios (Nakićenović et al., 2000). These scenarios (or storylines) are derived from four families of possible futures and are referred to as Special Report on Emissions Scenarios (SRES). For its most recent Assessment Report 5 (AR5), the IPCC recommended using four Representative Concentration Pathways (RCPs) in the simulation of future climate projections (Moss et al., 2010). Four RCPs (2.6, 4.5, 6 and 8.5, seen in Figure 1.3) are defined, each of which results from ‘different combinations of economic, technological, demographic, policy, and institutional futures’ (IPCC, 2013). The number of an RCP represents its total radiative forcing by the year 2100 (measured in units of Watts per square metre). In other words, the different RCPs vary in the measure of the effect that humans will have on the Earth’s systems by the year 2100. The higher an RCP is, the greater its radiative forcing on the climate. Higher RCPs result in more energy (heat) trapped in the Earth’s systems, and thus lead to greater warming. No SRES or RCP is treated as more likely or unlikely than another, meaning that all projections are considered as equally plausible futures regardless of the driving scenario or RCP.
1.4 Introduction to the datasets used

The following three subsections provide an overview of the datasets used in this thesis, along with the background motivation for undertaking these studies.

1.4.1 Climate model data: An introduction

In collating current research on the future European climate for their regional chapter in the most recent IPCC report (AR5), Kovats et al. (2014) state that there will be warming all across the continent. At a seasonal level, the strongest warming is projected across Southern Europe in summer, and in Northern Europe in winter. Even under scenarios with global temperature increases limited to 2°C, the climate of Europe is projected to depart significantly in the next decades from today’s climate.

Figure 1.3: The image above shows the four Representative Concentration Pathways (RCPs) recommended by the IPCC for use in climate models. Higher RCPs result in more energy (heat) trapped in the Earth’s systems, and thus lead to greater warming. (Credit: IPCC Data Distribution Centre)
1.4 Introduction to the datasets used

(Van der Linden and Mitchell, 2009). They claim that there will be a marked increase in high temperature extremes, with an increase in the number of warm days, warm nights, and heat waves projected. They recognise the limitations of global climate projections, and point out that using inter-model comparisons has provided a more robust range of future climates. In addition, they state that the climate models used show significant agreement for all emission scenarios in the magnitude and rate of warming across Europe. In looking at the impacts of these projections, the authors warn (with a medium level of confidence) that ‘climate change will increase the likelihood of systemic failures across European countries caused by extreme climate events affecting multiple sectors’ (Kovats et al., 2014).

With such significant changes and consequences, it is vitally important that Irish policymakers and the general public have access to robust and validated climate model projections, to help them to mitigate or avoid the worst of these effects. In order to do so effectively, access to quality data and projections at a fine (detailed) resolution for projected changes of mean and extreme temperatures across Ireland is necessary.

It is this need which motivates our dynamical downscaling and analysis of the temperature output from three high-resolution (4-7 km grid size) RCMs over Ireland. The three RCMs, driven by four GCMs from the Coupled Model Intercomparison Projects 3 and 5 (CMIP3 and CMIP5), were run under different SRES and RCP future scenarios.

The study domain here is shown in Figure 1.4. The RCMs were initially driven by GCM boundary conditions (achieving a ≈50 km grid size on the domain shown), and were then nested twice in succession (one-way nesting in the case of downscaling using the RMCs CCLM3 and CCLM4, and two-way nesting using the RCM WRF) to achieve the finest resolution (ranging from 4 km to 7 km grid size). An exception was the downscaling of the ECHAM5 (GCM) data, which had already been downscaled by the CLM community, and was available at a grid size of 18 km (validated by Hollweg et al. (2008)). The 50 and 18 km domains are large enough to allow changes to synoptic scales. Although information may not change further at synoptic scales for the finest grid size, the high-resolution representation of topography and dynamic processes may result in changes for variables near the land surface. Ideally, the domain
1.4 Introduction to the datasets used

for the finest grid size would be larger in order to allow the RCM to fully develop
small-scale dynamical structures in the interior of the domain, superposed on the
coarse-scale information that enters through the lateral boundaries. However, this
was not possible due to computational constraints.

Figure 1.4: The study domain is shown, with the nest and corresponding grid for each
run of the RCMs.

An overview of the simulations are included in Table 1.1: each row includes
information on a driving GCM; the corresponding RCM used to downscale it; the
SRES/RCP used for future simulations (including the number of realisations); the time
slice simulated; and the finest grid size achieved. Data from two time slices 1981-2000
(the control) and 2041-2060 were used for analysis of projected temperature changes
in the mid-21st century over Ireland. These periods were chosen as these are the
longest decadal time periods that were common to all simulations (see Table 1.1).
1.4 Introduction to the datasets used

<table>
<thead>
<tr>
<th>Driving GCM</th>
<th>RCM</th>
<th>Scenario (number of realisations)</th>
<th>Time slice</th>
<th>Finest grid size (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECHAM5</td>
<td>CCLM3</td>
<td>Historical (2)</td>
<td>1961-2000</td>
<td>7</td>
</tr>
<tr>
<td>ECHAM5</td>
<td>CCLM3</td>
<td>SRES A1B (2), SRES B1 (1)</td>
<td>2021-2060</td>
<td>7</td>
</tr>
<tr>
<td>ECHAM5</td>
<td>CCLM4</td>
<td>Historical (2)</td>
<td>1961-2000</td>
<td>7</td>
</tr>
<tr>
<td>ECHAM5</td>
<td>CCLM4</td>
<td>SRES A1B (2)</td>
<td>2021-2060</td>
<td>7</td>
</tr>
<tr>
<td>CGCM3.1</td>
<td>CCLM4</td>
<td>Historical (1)</td>
<td>1961-2000</td>
<td>4</td>
</tr>
<tr>
<td>CGCM3.1</td>
<td>CCLM4</td>
<td>SRES A1B (1), SRES A2 (1)</td>
<td>2021-2060</td>
<td>4</td>
</tr>
<tr>
<td>HadGEM2_ES</td>
<td>CCLM4</td>
<td>Historical (1)</td>
<td>1961-2000</td>
<td>4</td>
</tr>
<tr>
<td>EC_Earth</td>
<td>CCLM4</td>
<td>RCP4.5 (1), RCP8.5 (1)</td>
<td>2021-2060</td>
<td>4</td>
</tr>
<tr>
<td>EC_Earth</td>
<td>CCLM4</td>
<td>Historical (3)</td>
<td>1981-2009</td>
<td>4</td>
</tr>
<tr>
<td>EC_Earth</td>
<td>WRF</td>
<td>RCP4.5 (3), RCP8.5 (3)</td>
<td>2021-2060</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1.1: Details of all simulations used are shown, including the driving GCM, the RCM used to downscale it, the SRES/RCP used for future simulations (including the number of realisations), the time slice simulated, and the finest grid size achieved.

GCMs were chosen from two generations of the Coupled Model Intercomparison Project, CMIP3 and CMIP5. Detailed surveys of all CMIP5 models have been conducted over particular areas in other studies (such as by van den Hurk et al. (2014), who develop a regression technique to use all available CMIP5 projections). However, as the main concern of this study was to investigate the effect of downscaling models at a high resolution, a subset of these had to be selected due to computational constraints. Within CMIP3, CGCM3.1 (Scinocca et al., 2008) was chosen due to its superior ability to capture the spatial and temporal behaviour of the primary modes that drive weather in the Euro-Atlantic region (Casado and Pastor, 2012). ECHAM5 (Roeckner et al., 2003) was chosen due to its previously verified good performance when downscaled at 18 km grid size over Europe by the RCM CLM3 (Hollweg et al., 2008). Within CMIP5, EC-Earth (Hazeleger et al., 2012) and HadGEM2 (Collins et al., 2011) were chosen due to their complementary performance and biases. By comparing 33 models from CMIP5 with the E-OBS dataset over continental Europe for winters and summers separately, Cattiaux et al. (2013) found that both EC-Earth
and HadGEM2-ES had biases within one standard deviation of the ensemble mean. Moreover, EC-Earth showed a slight warm bias in winter and a slight cold bias in summer, whereas HadGEM2-ES showed the opposite. Consequently, EC-Earth was found to be slightly under-dispersive and HadGEM2-ES over-dispersive. Within the computational constraints of this study and the limitation to two models from CMIP5, these represent a good choice for performance over Europe, with complementary (opposing) biases, while ensuring adequate variability.

After choosing the four models, the individual realisations were selected. These realisations result from running the same GCM with different initial conditions. As they are all viewed as equally probable, the choice of realisation can often be pragmatic rather than theoretical. For CGCM3.1, the fourth realisation of CGCM3.1 T47 was used. For ECHAM5, realisations 1 and 2 were chosen. For EC-Earth, the three realisations performed by the Irish meteorological service (Met Éireann) were chosen (realisations 1, 13 and 14: r1i1p1, r13i1p1 and r14i1p1). For HadGEM2-ES, r1i1p1 was the chosen realisation.

The output from the four selected GCMs was downscaled using three RCMs: CCLM3, CCLM4 (both Rockel et al. (2008)) and WRF (Skamarock et al., 2008). The selected GCM-RCM pairings were conducted separately for each realisation - e.g., for the pairing EC-Earth-WRF, three separate downscaling computations were run: WRF downscaled EC-Earth realisation r1i1p1, WRF downscaled r13i1p1 and WRF downscaled r14i1p1. There are more GCMs than RCMs used (recommended by Déqué et al. (2007)). Due to computational constraints, some (but not all) GCM-RCM combinations or future scenarios/ RCPs were simulated.

In order to analyse this set of datasets, we need to know what methods we can use to validate multi-model climate model output, and how we should interpret their projections for future climates. Methods to do this are presented and discussed in Chapter 2, followed by the results of applying these methods to the set of datasets above.
1.4 Introduction to the datasets used

1.4.2 Waves dataset: An introduction

Knowledge of the extreme sea states affecting a region is essential for any marine activity. For shipping, and for offshore and coastal installations, it is crucial to have accurate information on the extremes likely to be encountered during operational lifetimes (Clancy et al., 2016). There is particular interest in modelling the significant wave height, \( H_s \), defined here as \( H_s = 4\sqrt{m_0} \) where \( m_0 \) is the zeroth moment of the directional wave spectrum.

Using the third-generation spectral WAVEWATCH III version 4.11 model (Tolman, 2014) with an unstructured grid (Roland, 2008), the wave climate around Ireland for the period from 1979 to 2012 was simulated. The unstructured triangular grid consisted of approximately 15,000 nodes (Figure 1.5) with horizontal resolution varying from 250 m in the nearshore to 10 km further offshore. The hourly fields produced were validated with observations from wave buoys and satellite altimeter data.

![Figure 1.5: The computational grid used for the wave climate hindcast, with the region of study outlined.](image)

Gallagher et al. (2014) reported a strong spatial variability in \( H_s \) across the do-
1.4 Introduction to the datasets used

main. Gallagher (2014) carried out a preliminary extreme value analysis of this dataset using the annual maxima at each point across the domain. The highest extremes of $H_s$ were found to occur off the west coast of Ireland, but the limitations of the approach meant that there was a high level of uncertainty in the parameter estimates. Clancy et al. (2015) found similar results, using a dataset from a coarser-resolution hindcast.

To build on this analysis, we choose to focus on the region off the west coast of Ireland indicated in Figure 1.5. This domain contains 334 gridpoints and has a depth ranging from 39 m to 1902 m. Figure 1.6 shows the mean significant wave height $H_s$ from the hindcast dataset. A clear spatial pattern is evident: the lowest values are those nearest the shore in the east of the domain, with a general increase in $H_s$ the farther west the gridpoint is.

![Figure 1.6: The mean significant wave heights in metres, for the 1979-2012 hindcast. The lowest values are those nearest the shore in the east of the domain, with a general increase in $H_s$ the farther west the gridpoint is.](image)

A model which uses the spatial information provided by nearby points and useful covariates should result in posterior estimates of desirable statistics (e.g., the significant wave height associated with a 100-year return period) which have greatly reduced uncertainty, compared to a more straightforward point-wise analysis of this dataset.
1.4 Introduction to the datasets used

It is this observation which leads us to explore spatial EVT and its application to this dataset, which forms the focus of Chapter 3.

1.4.3 Gridded temperature dataset: An introduction

With an increase in maximum temperature extremes through the current century projected by the IPCC (IPCC, 2013), it is important to have a better understanding of recent observations of climate extremes. Ireland is ‘completely off course in terms of achieving its 2020 and 2030 emission reduction targets’, according to the Climate Change Advisory Council’s Annual Review 2018. Without urgent action leading to substantial reductions in greenhouse gas emissions, they warn that Ireland is unlikely to deliver on its national, European Union (EU) and international obligations (Climate Change Advisory Council, 2018). A better understanding of our current climate extremes is essential in order to mitigate or avoid the many impacts of climate change; whether social, infrastructural, environmental, or economic (IPCC, 2014).

In order to understand better our extremes of temperature, we make use of a gridded dataset of daily maximum temperatures over Ireland. The data is on a $1 \times 1 \text{ km}^2$ grid (see Figure 1.7), consisting of daily values of maximum temperature for the 30-year period 1981-2010. Due to the large number of gridpoints in the dataset, we decide to focus on a smaller section to ensure that running the model is computationally feasible. We focus on an area covering County Dublin, which includes the capital city and its surroundings, located in the east of the island (the image on the right of Figure 1.7). This dataset contains approximately 1,700 gridpoints.
1.4 Introduction to the datasets used

Figure 1.7: The data domain showing the outline of the state of Ireland (in black, with county outlines included) is on the left of the image. The area within the black box is our study region, and is blown-up and shown on the right. The grey circles mark the gridpoints of the spatial domain.

In order to model extremes of daily maximum temperature, the chosen statistical models must use EVT because extremes of temperature are rare (by definition), and occur in the tails of the distribution. Since extremes of temperature vary by location on any given day, the statistical models used should account for this spatial dependence. And since these distributions are governed by parameters that depend not only on the data, but on other (e.g., physical or mathematical) principles and constraints, it is natural to use a Bayesian framework. Where datasets are very dense, as here, spatial models become computationally infeasible. Dimension-reduction techniques such as predictive processes can help to overcome this issue. With this in mind, in Chapter 4 we apply a Bayesian spatial hierarchical EVT model incorporating predictive processes to this dataset of daily maximum temperatures in Ireland.
1.5 Remainder of thesis

The rest of this thesis is organised as follows:

**Chapter 2**

In this chapter, I explain why we need to validate climate model output, and I discuss various ways in which we can do that. Following this, I then present and discuss the results of the high-resolution, dynamically downscaled multi-model dataset of Irish temperatures for the mid-21st century.

**Chapter 3**

In this chapter, I discuss some of the theory underlying spatial statistics (the methods used when analysing data at different points in space), and EVT (the methods used when dealing with extreme data). These two fields of statistics are combined in forming a Bayesian spatial hierarchical model, and then applied to the dataset of extreme waves off the west coast of Ireland. Results from this application are then presented and discussed.

**Chapter 4**

In this chapter, I discuss some theory of dimension-reduction techniques used when dealing with dense datasets. I then present a model which incorporates one of these techniques (predictive processes) in the Bayesian spatial hierarchical framework presented in the previous chapter, in order to perform inference on daily maximum temperature data over Ireland for 1981-2010. Following the results and discussion, more recent data (2011-2018) is incorporated into the final analysis to assess if there has been an increase in the severity and/or frequency of the observed extremes.

**Chapter 5**

The final chapter provides an overview of the results of this thesis. I discuss the contributions made here, and include several comments and suggestions for possible
future directions of research.
References


REFERENCES


Van der Linden, P., & Mitchell, J. F. B. (2009). *ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project*. 160 Met Office Hadley Centre FitzRoy Road, Exeter EX1 3PB, UK.


Chapter 2

Analysis of climate model output of Irish temperatures for the mid-21st century

There is a paucity of data from dynamically downscaled climate models at a high resolution over Ireland, and subsequent analysis of temperature projections for the mid-21st century. To address this gap in the research, this chapter analyses the high-resolution, multi-model climate model output introduced in Section 1.4.1. Prior to presenting this analysis, relevant previous studies on temperature in Ireland are discussed, leading to the motivation for the research undertaken in this chapter. In order to analyse this multi-model dataset, it is first necessary to understand how climate model output is validated. With this in mind, I explain the rationale for climate model validation, and describe some validation methods. In addition to the more usual focus of analysing the projected mean change in a variable, I also discuss a simple technique to analyse the projected extreme changes. Following this contextual background, the projected temperature changes from the output of three high-resolution regional climate models (RCMs) over Ireland, driven by four global climate models (GCMs), under five different possible futures are analysed. This chapter contains material from a paper we wrote and submitted to the International Journal of Climatology. It was accepted and published as an article entitled "A high-resolution, multi-model analysis of Irish temperatures for the mid-21st century" (O’Sullivan et al., 2016).
2.1 Past research on temperature change in Ireland

The following two subsections include an overview of previous research on projected changes in mean and extreme temperatures over Ireland and, more generally, over Europe. The final subsection explains the rationale for the current study.

2.1.1 Changes in mean temperature

The IPCC reports an estimated rise in global mean surface temperatures by the late 21st century of between 0.3°C and 4.8°C (IPCC, 2013b). The rise in European mean temperatures is projected to exceed the rise in the global mean. In fact, the median temperature over Ireland for the period 2046-2065 is projected to increase by 1-1.5°C in future summers and by 0.5-1.5°C in future winters under RCP 4.5 (IPCC, 2013a).

Heinrich and Gobiet (2012) used eight RCMs (all approximately of 25 km grid size) from the ENSEMBLES project to analyse projected changes in mean temperature over Europe between 1961-1990 and 2021-2050 under SRES A1B. Examining the multi-model mean change seasonally, they found that warming is projected across all seasons and all areas; these are most pronounced in the northeast of Europe in winter and southern Europe in summer. Temperatures over Britain and Ireland are projected to increase uniformly across all areas and all seasons in that period by approximately 1-1.5°C. This uniform warming is possibly an artefact of the coarse model resolutions (approximately 25 km), as differences across small scales are less likely to emerge at this level.

Using data resulting from the PRUDENCE project, Déqué et al. (2007) compared the results of 25 simulations driven by three GCMs under two different driving scenarios (SRES A2 and B2), in order to estimate the uncertainty in using RCMs for future climate projections over Europe. They investigated the changes projected between 1961-1990 and 2071-2100, while attempting to separate the various sources of uncertainty from each other. They note that the choice of the driving GCM generally introduces more uncertainty than the other sources. They suggest using at least as many GCMs as RCMs in any further studies. However, despite the uncertainties, they
2.1 Past research on temperature change in Ireland

concluded that the projected warming evident across Europe is statistically significant (e.g., the lower value of all 99% confidence intervals for projected warming exceeds 1°C).

The ENSEMBLES project (results summarised by Van der Linden and Mitchell (2009)) involved the downscaling of seven GCMs over Europe by eight RCMs at a 25 km grid size. Under SRES A1B for 2021-2050, projected annual temperature changes over Ireland were found to be in the range of 1-1.2°C above the control period of 1961-1990. For the same time period, winter and summer changes were both projected to lie in the range 1-1.5°C.

The Community Climate Change Consortium for Ireland (C4I) downscaled data from five GCMs over Ireland and Britain, using all SRES scenarios, achieving a finest grid size of 14 km. Looking at seasonal projections for 2021-2060, they found the greatest change in mean temperatures projected for summers and autumns (1.2-1.4°C) (McGrath et al., 2008).

Statistical downscaling was used by Fealy and Sweeney (2008) to assess projected temperature changes at several sites in Ireland. They used output from three GCMs under two SRES scenarios (A2 and B2) and found that by the 2050s Irish temperatures are projected to increase by 1.4-1.8°C above the control period of 1961-1990. In addition, they found that the greatest warming is projected for future autumns. Mullan et al. (2012) also used statistical downscaling, but over Northern Ireland, and achieved similar results.

2.1.2 Changes in extreme temperature

The IPCC report that it is now ‘very likely’ that human-induced climate change has contributed ‘to the observed changes in the frequency and intensity of daily temperature extremes on the global scale’ (IPCC, 2013b). This confirms what was already suggested both in IPCC AR4 (2007) and an IPCC Special Report: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (2012). They warn that near-term projections suggest that increases in temperature
2.1 Past research on temperature change in Ireland

Extremes are likely. In Europe, high-percentile summer temperatures are projected to rise faster than mean temperatures (IPCC, 2013c).

Beniston et al. (2007) used the RCM output from the PRUDENCE project to examine how extreme temperature events in Europe are projected to change by the end of the 21st century. They found that an increase in temperature variability in the future over the interior of the continent implies that the intensity of extreme temperature events (as characterised by having relatively large deviations from the mean) will increase more rapidly than the intensity of more moderate temperatures.

In their study over Northern Ireland, Mullan et al. (2012) examined the projected changes in extremes of temperature. They used a threshold approach to define percentile values for the control period that depend on each particular location: the 30-year 90th percentile of maximum temperature to examine changes in hot days and the 30-year 10th percentile of minimum temperature to examine changes in cold nights. The projected changes in these thresholds at each site were then calculated. Both the hot-day threshold and the cold-night threshold are projected to increase by similar amounts, and show an increasing trend through the 2020s, 2050s and 2080s. There is a large uncertainty range for both changes, though all ranges are non-negative.

2.1.3 Rationale for current study

Following a review of prior research, it is clear that there is a paucity of data from dynamically downscaled climate models at a high resolution over Ireland, and subsequent analysis of temperature projections for the mid-21st century. Existing studies have focused on larger domains with coarser resolutions (≈14 km grid size, McGrath et al. (2008)), or have instead used statistical downscaling. In Sections 2.3 and 2.4, the outputs of three RCMs driven by four GCMs under different SRESs and RCPs are presented and analysed in relation to projected temperature changes. But before this can be done, I explain why climate model output needs to be validated, and how the projected changes should then be analysed.
2.2 Climate Model Validation

Model validation involves comparing a climate model's output to an independent set of observations to come to some quantitative or qualitative assessment of its skill in reproducing the observed climate. It is a test which every climate model should undergo, and in which it should succeed. It is, to date, the only method with which model confidence can be assigned. Climate models are not typically run from the present day forward, but start from 1850 (the first year of the instrumental record of global surface temperature). This provides model data for a historical period which can then be compared against observations during the validation process.

2.2.1 Validation Methods

It is a far from straightforward task to measure and quantify how similar or dissimilar historical output from a particular climate model is to observed reality. The first issue encountered is that reality, as defined by observations, is only known from weather station readings in particular places and at particular times. In addition, the values of these quantities are not known exactly, but instead have probability ranges depending on instrument accuracy. Furthermore, they are distributed very non-uniformly: some areas may have a dense concentration of weather stations (e.g., across Europe), whereas others (e.g., over oceans) may be very sparsely populated with data from weather stations (see Figure 2.1, which shows the spatial variation of the U.S. National Oceanic and Atmosphere Administration’s (NOAA’s) Integrated Surface Database, which consists of global hourly and synoptic observations). Many weather stations also suffer from a lack of continuity in their data, leading to broken time series (Alexander et al., 2006).
2.2 Climate Model Validation

Figure 2.1: This figure shows the spatial variation of the U.S. National Oceanic and Atmosphere Administration's (NOAA’s) Integrated Surface Database, which consists of global hourly and synoptic observations (Credit: NOAA\(^1\)).

The second issue arises due to the many different ways to statistically interpolate observational data from a non-uniform and sometimes time-discontinuous distribution onto a regularly spaced grid, with values of quantities given within each box, at each time-step.

Once the observed data have been interpolated onto a grid, methods of comparison with model output need to be decided upon. If the grids do not match each other exactly (that is, if the boxes are in different positions and/or are different sizes), then interpolation once again needs to be carried out. This involves either moving the model output onto the gridded observations, or the gridded observations onto the model output grid. Each interpolation may involve smoothing of the data or model output, which means that extreme values or ‘spikes’ in the data may be reduced. This needs to be considered when analysis is undertaken, particularly in the area of extremes.

\(^1\)https://www.ncdc.noaa.gov/sites/default/files/Integrated-Surface-Database-Station-Distribution-Map.gif
There is no single best method to assess how similar two sets of data are, when they consist of many quantities distributed in space and time. The method chosen in each study should follow careful research and consideration of methods from other studies. These can then be tailored to the individual requirements or focus of the particular study in question.

The most common initial approach to the problem is to assess mean values (averages) over both space and time. For example, the mean observed surface temperature for winter could be compared to the mean modelled winter temperature. This can be done at each geographical location (comparing means on time-averaged area plots). Alternatively single values can be obtained for particular winters for both modelled data and observations, and then compared in a time series (comparing means on area-averaged time plots).

After a comparison of means, attention typically turns to the range of values a quantity takes, and how often it takes those values. This is related to the spread of the distribution of a quantity, often measured as standard deviations in the same units as the quantity itself.

2.2.2 Extremes
The validation of mean values of variables alone is insufficient. A climate model which produces a mean value identical to the observed mean does not necessarily accurately reproduce the rest of the values that the variable may take (that is, the distribution of the variable). A change in the distribution of a variable (e.g., a change in the spread of values) may not change the mean, but will lead to a change in the number of extreme events (Frich et al., 2002). If there were no future change expected in the annual mean temperature over Ireland, for example, that does not exclude the possibility of an increase in the number of heatwaves and big freezes. If both occur to the same extent, averaging over the course of a year could cause them to ‘cancel each other out’ leaving the annual mean temperature unaffected. Changes in extreme events are arguably of more immediate importance to people, since they have an abrupt and
2.2 Climate Model Validation

much larger impact on lives and livelihoods than a gradual change in mean values (Easterling et al., 2000).

2.2.2.1 Percentiles

One approach to the analysis of extremes involves a percentile-based method. A 90th percentile value of a set of numbers, for example, is the number above which only 1 in 10 numbers lie. Using data from the Hadley Centre RCM (following downscaling of the GCM HadCM2), Jones and Reid (2001) follow a percentile approach in order to identify threshold amounts, specific to a location, which must be exceeded in order to be in the top wettest 10% of days. Their assessment of future changes in extreme precipitation over Britain concludes that ‘the number of events above [these] fixed thresholds increases’.

2.2.3 Distributions

A more comprehensive picture of how well a climate model captures the mean, the extremes, and all values in between can be built up through the probability density method employed by Perkins et al. (2007). In their paper, the authors evaluate the climate models used in the IPCC AR4 regarding their ability to simulate daily maximum and minimum temperatures, and daily precipitation over Australia. For each variable, this approach involves assigning each day to a particular ‘bin’ depending on the value of that variable. For example, a typical bin may count all days with a maximum temperature of between 25°C and 25.5°C. These bins and their totals are then plotted as histograms to reveal the shape of the distribution of the variable. In this way, all values of a variable are included, allowing for a more complete assessment of the fit between modelled and observed precipitation.

The authors then proceed to scale the histograms to have an area of 1. They then introduce a scoring method for the goodness of fit between the observations and a particular model which involves summing the minimum value of these two at each bin. Since each sums individually to 1, the resultant score will be between 0 and 1. 1 represents complete agreement and 0 means that there is no overlap between the two histograms. This scoring system has the advantage of allowing a quantitative
2.2 Climate Model Validation

measure of the agreement between two histograms which is both ‘easily interpreted... and comparable to what would be assessed by eye’. It also allows for easy comparison of performance across different ensembles, different models, and even different quantities.

This is illustrated in Figure 2.2, taken from Perkins et al. (2007) (Figure 3 in that article). Figure 2.2(a) shows modelled data and observations in very close agreement: summing over the minima here gives a very high skill-score of 0.9. In contrast, Figure 2.2(b) shows very little agreement between the modelled and observed data: summing over the minima here gives a very poor skill-score of 0.02.

![Figure 2.2](image)

Figure 2.2: This diagram illustrates a modelled vs. an observed distribution of values (for temperature, in this schematic). (a) has a skill-score of 0.9, while (b) has a skill-score of 0.02 (Perkins et al., 2007).

A disadvantage to this methodology employed by Perkins et al. (2007) is that as an event becomes more extreme (or less frequent) in the observed and the modelled world, ‘failure of the model to simulate [this event] becomes less important to the skill score’. This problem can be countered. Rather than summing up all minima as
explained above (essentially with even weights for each bin), larger weights in the tails of the distribution of observations would penalise models more severely for failing to model extremes accurately, while still taking the other values into account.

However, as Osborn and Hulme (1997) stress (and as mentioned in Section 2.2.1), interpolation of observations onto a uniform grid may cause some detail (e.g., extremes) to be lost or smoothed in the averaging process. A disadvantage to the weighted bins approach suggested above is that it assumes that the observed gridded distribution used is the ‘true’ distribution, even though the observed distribution itself may be skewed from reality and under-represent the extremes in the variable.

2.3 Preliminary analysis

The data analysed for the remainder of this chapter is the high-resolution multi-model dataset which is described in Section 1.4.1. The reader should refer back to this section for a detailed description, but included here is a brief reminder: Three RCMs were initially driven by one of four different GCM boundary conditions (achieving a ≈50 km grid size), and were then nested twice in succession to achieve the finest resolution (ranging from 4 km to 7 km grid size). Data from two time slices 1981-2000 (the control) and 2041-2060 are used for analysis of projected temperature changes in the mid-21st century over Ireland. The downscaling involved future scenarios defined from three SRESs (SRES A1B, SRES B1 and SRES A2) and future forcing defined from two RCPs (RCP 4.5 and RCP 8.5). Due to computational constraints, some (but not all) GCM-RCM combinations or future scenarios/RCPs were simulated.

2.3.1 Comparison of GCM and RCM data

In order to justify the computational expense of downscaling an ensemble of GCMs at a high resolution, a preliminary analysis was conducted into the effects of downscaling over Ireland. Two representative cases were considered: a low scenario (SRES B1) and a high scenario (RCP 8.5). For SRES B1, the GCM is ECHAM5 and the RCM is CCLM3. For RCP 8.5, the GCM is EC-Earth (realisation r1i1p1) and the RCM is CCLM4. The RCM data over Ireland (‘RCM land’) was compared with the
2.3 Preliminary analysis

corresponding GCM data both over Ireland (‘GCM land’) and over the Atlantic to the west of Ireland with the same domain size (‘GCM sea’). The temperature response was calculated by subtracting the mean state of the historical period from each corresponding future dataset (i.e. the response gives the deviation of each future climate from its historical mean state).

Figure 2.3 illustrates the empirical density (achieved by binning the temperature responses) for RCP 8.5. It clearly shows that there is greater variability present in ‘GCM land’ compared with ‘GCM sea’, which is to be expected due to the higher specific heat capacity of the ocean. It also shows that there is greater variability present in ‘RCM land’ compared with ‘GCM land’. This is also to be expected as the higher resolution of the RCM allows for more land-only grid points, thus reducing the moderating effect of the ocean for inland points. This is a clear illustration of one motivation for using high-resolution RCMs: they allow extremes to be captured that are not represented in the lower resolution GCMs. Results for SRES B1 are comparable, showing a similar increase in variability from ‘GCM sea’ to ‘GCM land’ to ‘RCM land’.
2.3 Preliminary analysis

Figure 2.3: The empirical densities of temperature responses for EC-Earth over the Atlantic (GCM sea - solid bold line), EC-Earth over Ireland (GCM land - dotted line) and EC-EARTH downscaled over Ireland by CCLM4 (RCM land - solid thin line) are shown, under RCP 8.5.

Further evidence that the high-resolution RCMs reduce the moderating effect of the ocean is seen in Table 2.1, which lists the mean temperature responses of ‘GCM sea’, ‘GCM land’ and ‘RCM land’. For EC-Earth realisation 1-CCLM4 in scenario RCP 8.5, there is a progressive increase in the response value from ‘GCM sea’ (1.03 °C) to ‘GCM land’ (1.20°C) to ‘RCM land’ (1.38°C). More information about how to interpret these values can be gained by including data from other realisations. Realisation 13 shows a larger response than realisation 1, while realisation 14 shows a lower response. The spread for the EC-Earth-CCLM4 response over land is 0.68°C (aleatoric uncertainty), while the mean response is 1.51°C. The equivalent EC-Earth-WRF values are 0.45°C and 1.51°C, respectively. These responses of all such pairings are included in Table 2.1, where it can be seen that the response is always greater over
land than sea, and that the response is always greater than the aleatoric uncertainty. Responses for SRES B1 also progressively increase (but are of a lower magnitude as it is a low emissions scenario). Both the increased variability and the greater warming evident in the high-resolution RCMs in comparison to the driving GCMs prompted the use of high-resolution dynamical downscaling to investigate both mean and extreme projected temperature changes over Ireland.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>GCM</th>
<th>GCM sea response</th>
<th>GCM land response</th>
<th>RCM land response</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRES B1</td>
<td>ECHAM5</td>
<td>0.26°C</td>
<td>0.61°C</td>
<td>CCLM3: 0.67°C</td>
</tr>
<tr>
<td>RCP 4.5</td>
<td>EC-Earth: (1,13,14)</td>
<td>0.79°C, 1.10°C, 0.90°C</td>
<td>0.93°C, 1.29°C, 1.12°C</td>
<td>CCLM4: 1.06°C, 1.32°C, 1.13°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.67°C</td>
<td></td>
<td>WRF: 1.13°C, 1.29°C, 1.11°C</td>
</tr>
<tr>
<td>SRES A1B</td>
<td>ECHAM5: (1,2)</td>
<td>0.72°C, 0.49°C</td>
<td>1.23°C, 1.06°C</td>
<td>CCLM3: 1.35°C, 1.20°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.03°C, 1.54°C, 0.82°C</td>
<td>1.20°C, 1.77°C, 1.18°C</td>
<td>CCLM4: 1.37°C, 1.20°C</td>
</tr>
<tr>
<td>RCP 8.5</td>
<td>EC-Earth: (1,13,14)</td>
<td>1.03°C, 1.54°C, 0.82°C</td>
<td>1.03°C, 1.54°C, 0.82°C</td>
<td>CCLM4: 1.38°C, 1.92°C, 1.24°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.38°C</td>
<td></td>
<td>WRF: 1.38°C, 1.80°C, 1.35°C</td>
</tr>
</tbody>
</table>

Table 2.1: The results of the analysis of the effect of downscaling on mean two-metre temperature are shown. The results include the particular future scenario in the first column, the GCM and its realisations (if applicable) in the second column, the mean temperature response for the ‘GCM sea’ data in the third column, the mean temperature response for the ‘GCM land’ data in the fourth column, and each RCM and its corresponding mean temperature response (‘RCM land’) in the fifth column. Responses in bold are those analysed earlier in Section 2.3.1.

### 2.3.2 Analysis methods

In order to analyse temperature changes, responses to the model forcings were calculated as described above in the test case comparing GCM driving data and the resulting downscaled RCM data (Section 2.3.1). That is, the mean of each historical period was subtracted from the corresponding future period of each model within the same realisation. This resulted in temperature responses for each realisation of a GCM-RCM pair; that is, the difference between future and past. In this way, cold or warm biases of particular models will not skew results, and each response can be meaningfully compared with the other groups.

In order to reduce the GCM-RCM pairings to a more manageable number to ease comparison, the future model data were grouped together by SRES scenario or RCP. This resulted in five groups, hereafter denoted A1B, B1, A2, RCP45 and RCP85.
Analysis has been conducted on the mean two-metre temperature (T2m), the daily minimum temperature (Tmn), and the daily maximum temperature (Tmx).

2.4 Results

In this section, the outputs of three RCMs driven by four GCMs, grouped into five different future scenarios, are analysed in relation to projected temperature changes. The final model outputs are at a horizontal resolution of 4-7 km. Annual and seasonal mean changes are examined, both averaged across the island and analysed spatially. In addition to examining projected mean changes, changes in the behaviour of temperature extremes are analysed. This research also investigates projected changes in the spatially averaged distribution of temperature over Ireland.

Figure 2.4 shows the time series of annual mean temperature responses for the five groups for 2041-2060. These are calculated using values averaged over Ireland. Dashed lines are lines of regression, calculated by the least-squares algorithm to show the linear trend over the two decades.

Figure 2.4: Time series of the annual mean temperature responses for each of the five groups are shown, averaged over Ireland. Dashed lines are lines of regression fitted using the least-squares method.

There is a broad range of values for temperature changes projected across the
2.4 Results

five groups. Group B1 shows the least warming (its mean response is 0.66°C) and the greatest interannual variability. Its large spread, seen through its annual mean responses ranging from -0.4°C to greater than 1.4°C, is most likely due to it being composed of a single-member ensemble. The remaining four groups show greater warming than B1, with A2, A1B and RCP45 having large areas of overlap, and similar 20-year means (1.1°C, 1.2°C and 1.24°C, respectively). RCP85 shows the greatest warming of the groups considered, with a trend line always exceeding the other groups (its mean response is 1.56°C).

As annual trends can hide or smooth larger seasonal trends, seasonal mean temperature responses have also been examined. Seasonal means from the control period 1981-2000 were calculated for each group, then subtracted from the 20 corresponding seasons in 2041-2060, resulting in seasonal responses for each group (showing their deviation from the mean state of the control period, as explained in Section 2.3.1). The spread of these responses within each group, for each of the four seasons, could then be examined. The box and whisker plots for each group are shown (Figure 2.5) for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). The whisker ends mark the maximum and minimum seasonal values within the 20-year sample, while the box marks the median value (central line) and the first and third quartile values (lower and upper sides, respectively).
2.4 Results

Figure 2.5: The seasonal T2m responses [winter (DJF), spring (MAM), summer (JJA) and autumn (SON)] are shown for the five groups. The bottom and top whiskers represent the minimum and maximum group values, respectively, while the bottom and top of the box represents each group’s first and third quartiles respectively, while the middle line represents its median.

The seasonal responses in Figure 2.5 again illustrate that there is a lot of uncertainty in future projections, both within each season and within each group. Within summer, for example, temperature responses range from approximately -1°C in one particular year (group B1) to almost 6°C in another (RCP45). The responses below 0°C are indicative of the natural variability inherent in the climate system. A particularly cold spring projected in group A1B, for example, has a seasonal temperature response below -2°C. However, the overall trend of an increase in temperature is evident across all seasons. With the exception of group B1 in spring and summer, all first quartiles across all seasons exceed 0°C, suggesting a definitive movement upwards in T2m across all groups and all seasons, relative to the control period 1981-2000.
2.4 Results

There are also clear differences between results across the different groups. Future summers have the largest projected warming in RCP85, whereas winters are projected to warm the most in groups A1B and A2. However, all three agree on autumn as the season with the second-greatest projected warming.

Changes can be spatially dependent, so seasonal responses for each gridpoint over Ireland were calculated for each group. Figure 2.6 shows the mean seasonal temperature responses for the group RCP85 (the white areas are an artefact resulting from the regridding process applied to land-sea masks under different projections; only those boxes with sufficient land were included in the final analysis). The single number included on each plot is the value of the response at that particular point, and is included as a visual reference to aid interpretation of the figure.
2.4 Results

The greatest change is for future summers, where a uniform warming of between 1.5°C and 2°C is projected across the island, with the exception of the southeast. Here, projected warming is greater, in the 2-2.5°C range. There is a uniform projected warming in the 1.5-2°C range for autumns, while spring is projected to warm the least, with projected temperature increases of between 1°C and 1.5°C. In future winters, the greatest warming is projected for the northeast of Ireland, in the range 1.5-2°C. There are many possible factors that could have influenced this uneven warming for future summers and winters, such as a change in storm tracks or the North Atlantic Oscillation (NAO) relative to the control period 1981-2000. However, further investigation of these factors is necessary to attribute causation, and is beyond...
2.4 Results

the scope of this study. Moreover, both patterns fail to appear in the other groups. Groups A1B and A2 both project relatively uniform warming across the island within each season. Unlike RCP85 however, the greatest warming is projected for future winters (1.5-2°C for both groups, with the exception of the extreme southwest of the country, which has projected warming of 1-1.5°C), followed by future autumns (1-1.5°C for both groups). This supports the seasonal trends already evidenced from the boxplot analysis of Figure 2.5. Similarly to these two groups, relatively uniform warming is projected in group RCP45. However, the greatest warming for this group is projected for future autumns.

Changes in the extreme values of temperature have also been examined. The IPCC consider extremely unlikely events as those that have a less than 5% chance of occurring (IPCC, 2012). Therefore, changes in the 5 percentile value of the daily minimum temperature (hereafter Tmn05) and the 95 percentile value of the daily maximum temperature (hereafter Tmx95) were examined. Any changes in these values relate to changes in the number of cold nights and the number of hot days, respectively. Events that are considered extremely unlikely to occur in the historical period may occur more or less frequently in the projected futures. Figure 2.7 shows the RCP85 winter and summer response of Tmn05 and Tmx95, respectively.
2.4 Results

Figure 2.7: The response for Tmn05 for winter (DJF) and the response for Tmx95 for summer (JJA) are shown above for group RCP85. The number included on each plot as an aid for visual analysis represents the value of the response at that point.

There is evidence here (Figure 2.7) of greater warming of the two percentile values considered than for the mean seasonal temperature changes projected in Figure 2.6. Projected changes in the mean temperature do not exceed 2°C (apart from a small area in the southeast of the country in summer), but projected changes in Tmn05 and Tmx95 exceed 2°C across almost the entire island, with many places exceeding 3°C.

For future winters, the projected change in Tmn05 exceeds 1.5°C everywhere, but is considerably greater in the north of the island (2-3.5°C). This means that, during the period 2041-2060 under RCP85, the threshold of daily minimum temperature below which only 5% of nights fall is projected to increase by at least 1.5°C and by a much greater amount in the north by mid-century.

Across almost the entire island, Tmx95 is projected to increase by more than 2°C - for most of the south of the country, this increase exceeds 2.5°C. This means that the threshold of daily maximum temperature that is exceeded only by 5% of days is
projected to increase by approximately 2.5°C in future summers.

Under all groups, projections for Tmn05 during winter increased by a greater magnitude than those for T2m. In some cases, the difference was quite considerable. For example, in group A2 mean winter temperatures are projected to increase uniformly across the island by less than 2°C. However, projections for Tmn05 in the same group exceed 3°C everywhere in future winters, with a maximum increase of 7°C. Similarly, the magnitudes of the projections for changes in Tmx95 for future summers exceed the mean changes across all groups. For example, T2m in future summers is projected to increase by 1-1.5°C in group RCP45, whereas Tmx95 in future summers is projected to increase by 1.5-2°C for the same group across most of Ireland, with projections greater than 2°C in parts of the southwest.

A more comprehensive step in moving from analysing mean values to the examination of extreme events is to consider the distribution of a variable. The distribution of a variable involving discrete values (as here) can be represented by its histogram. Seasonal histograms were calculated by binning (in bins of width 1°C) responses over all grid points and all group members, and scaling so the area summed to 1, resulting in control and future histograms. Group RCP85 has been included, as this was the RCP with the greater projected change. Figure 2.8 shows the daily temperature responses by season (drawn as densities to aid visual interpretation), where the group’s control and its respective future are compared. The density of the historical control period is shown as a continuous line with open circles, and the density of projected temperature for the group RCP85 is shown as a dashed line with crosses. Overlap scores were calculated, which assess the similarity between each group’s past and future. This score was obtained by summing the minimum value at each bin, across the entire combined range of the two densities (Equation (12) in Cha (2007)). By then multiplying by 100, this results in a score between 0 and 100%, with 100% indicating perfect agreement (climate completely unchanged) and 0% indicating no agreement at all (past and future climates have no values in common).
2.4 Results

Figure 2.8: Seasonal densities for group RCP85 illustrating the distribution of historical model temperature responses (solid line, with open circles) and the future model temperature responses (dotted line, with crosses) are shown. A measure of overlap indicates how much the two distributions have changed [(a) DJF=81%; (b) MAM=86%; (c) JJA=71%; (d) SON=83%]. Means are shown as vertical lines for historical (solid line) and future (grey line) densities.

There is evidence in Figure 2.8 of temperature increases across all seasons, as each future density mean is shifted to the right of its historical density mean. In addition to this mean increase, the future density values across the entire range are shifted upwards, strengthening the evidence from all of the previous analysis of increases in both the mean values and the tails of the temperature distribution.

The overlap scores range from 71% in summer (showing the largest increase
across the distribution) to 86% in spring (showing the least change between historical and future densities). These reinforce the mean changes suggested by RCP85 in Figure 2.6, which had the greatest increases in summer, and the smallest increases projected for future springs.

2.5 Discussion and Conclusions

Examination of Figure 2.4 reveals a mean temperature increase across all groups, evidenced by positive responses (0.66-1.56°C). Also, the regression lines suggest that temperatures are projected to continue increasing between 2041 and 2060. That is, the regression coefficients (slopes of the lines) are also positive for all groups [0.03°C decade\(^{-1}\) (RCP45) to 0.4°C decade\(^{-1}\) (A1B)]. However, 95% confidence intervals for the regression coefficients are positive only for groups A1B and RCP85 [(0.16°C decade\(^{-1}\), 0.63°C decade\(^{-1}\)] and [(0.02°C decade\(^{-1}\), 0.39°C decade\(^{-1}\)], respectively. The regression line for RCP45 is quite flat (confidence interval: -0.17°C decade\(^{-1}\) to 0.23°C decade\(^{-1}\)), suggesting neither a cooling nor a warming trend over the period considered.

The annual results for group A1B (1.2°C response) broadly agree with those of the ENSEMBLES project (Van der Linden and Mitchell, 2009) for SRES A1B, which projected annual temperature changes to be in the range of 1-1.2°C (though the time slices are not identical). The seasonal changes are similar in summer (both studies project changes to be in the range 1-1.5°C), but the ENSEMBLES winter results (1-1.5°C) differ somewhat from this particular study (1.5-2°C warming projected across the island).

The results for C4I (McGrath et al., 2008) are averaged across different SRES, but in general their magnitude (1-1.4°C across all seasons) is similar to the results observed in Figure 2.5.

Though the time slices differ slightly, it is worth comparing the results with those obtained at sites over Ireland by Fealy and Sweeney (2008) (see Section 2.1.1). Their projections are for temperature increases of between 1.4°C and 1.8°C by the 2050s.
under SRES A2 and B2. These projections exceed those in Figure 2.4, where the group A2 has a regression line projecting temperature increases of approximately 0.9-1.4°C over the 20-year period.

The magnitude of the results for group RCP45 (shown in Figure 2.5) broadly agrees with a study on a larger ensemble analysed in IPCC AR5, which projected the median temperature over Ireland to increase by 1-1.5°C in future summers, and by 0.5-1.5°C in future winters under RCP 4.5, for the period 2046-2065.

The results illustrated in Figure 2.6 disagree in part with Fealy and Sweeney (2008). They found that the greatest warming was projected for future autumns, in contrast with the analysis above projecting summers (in RCP85) or winters (in A1B and A2) to warm the most. They also note that the inter-GCM range for seasonal change is relatively large, again supporting the idea of using multiple GCMs (as this study does).

As the mean values of T2m are projected to increase by less than either Tmn05 or Tmx95 for the seasons considered (comparing the relevant seasons in Figures 2.6 and 2.7), this suggests a change in the variability and/or skewness of the distributions. Seasonal density overlap scores were calculated for each season to investigate the source of the changes observed in group RCP85 (Figure 2.8). Summer temperatures show the greatest projected change between future and past, with the lowest agreement score of 71% for this season. Winter has the second-lowest agreement between future and past (81%), indicative of the second-greatest projected change. This agrees with the analysis of Figures 2.6 and 2.7, which shows that Tmn05 in winter and Tmx95 in summer are projected to increase by more than the corresponding changes in T2m. This is in line with the findings of Beniston et al. (2007), who concluded that the intensity of extreme temperature events will increase more rapidly than that of moderate temperatures (though their conclusion was for continental Europe; Section 2.1.2). Their conclusion was due to an increase in temperature variability, which is observed in this study in group RCP85 for summer (a 20% increase in standard deviation) and spring (7.6% increase), whereas autumn only shows a 0.1% increase and winter shows a 2.7% decrease. Skewness measures the asymmetry of a
distribution, and is also an important statistic to consider. An increase in skewness indicates that either the right tail has become more elongated or the left tail more compressed, or both. A decrease in skewness implies the opposite. Skewness is a dimensionless measure, but the standard deviation of the skewness measure can be used to assess how large the change is. The greatest change in skewness is seen for future springs, where the skewness changed from 0.11 to 0.56, an increase of 0.45. This is much greater than 0.023, the standard deviation of the skewness (this value is the standard deviation of the skewness for all four seasons). The change in mean temperature for future springs is the least of all seasons however, meaning the change in skewness is hidden by the highest similarity score of 86%. Changes in skewness were also positive for winters and autumns, though of a smaller magnitude (from -0.05 to 0.07 for winters, a change of +0.12, and from -0.26 to -0.21 for autumns, a change of +0.05), while the skewness for future summers decreased slightly (from 0.98 to 0.9, a change of -0.08).

In a statistical downscaling of GCM data at specific sites over Ireland, Mullan et al. (2012) examine changes in Tmn10 and Tmx90 for the 2050s. It is difficult to compare those results with this study due to their averaging across scenarios and seasons, and choice of a different threshold. Analogous to this study, they found Tmx90 projected to increase more than the mean temperature. In contrast, however, their results for Tmn10 show it projected to change less than the mean temperature.

This study presented results obtained from a high-resolution dynamically down-scaled dataset over Ireland, involving three RCMs and four GCMs. The scale and domain of this study is unique, and has addressed a topic that previously lacked adequate research. Analysis conducted to justify using high-resolution RCM data illustrated the added value of such an approach. Greater variability in the climate signal was captured by the RCMs, which were better able to capture and explicitly resolve more small-scale features that affect the evolution of two-metre temperature. Outputs from our selected ensemble of models and anthropogenic forcing scenarios were then examined. Temperature changes for the mid-21st century have been projected and analysed in detail. A comparison to previous relevant studies was carried out.
2.5 Discussion and Conclusions

Time slices of 1981-2000 (control) and 2041-2060 (anthropogenically forced future scenarios) were compared in order to determine the nature and magnitude of projected temperature changes. All groups showed an increase in annual mean temperature: this increase was greatest in group RCP85 (1.56°C), with the greatest trend projected in group A1B (0.4°C decade$^{-1}$).

Seasonally, a large spread of temperature responses across groups was evident. However, the warming signal was apparent from each median seasonal temperature exceeding 0.5°C for every group and season (with the exception of B1 in spring). Groups differed in their warmest projected seasons, from winter (A1B and A2) to summer (RCP85).

On a spatial scale, there are discrepancies between the five groups as to which season has the greatest projected warming, and also whether this warming will be uniform or vary across Ireland. As expected, however, the greatest warming is evident from those groups with higher emissions or RCPs: A1B, A2 and RCP85.

Examination of the projected changes in Tmn05 in winter and those of Tmx95 in summer again show differences across the five groups considered. However, a strong pattern of greater increases in both of these values when compared with the projected mean changes in T2m is evident across all groups. This is reinforced by examining the densities of seasonal changes.
References


Chapter 3

Using Bayesian spatial hierarchical modelling and Extreme Value Theory to predict extreme sea states

As part of the analysis presented in Chapter 2, changes in the projected extremes of temperature were investigated. This involved examining how particular percentile levels were projected to change at each gridpoint of the domain. While this is informative, it failed to take account of the spatial dependence in the data - nearby points are likely to experience similar extremes, as well as points which share other physical attributes such as a similar altitude, proximity to the coast, etc. Making use of this spatial dependence across the domain could greatly benefit any analysis of spatially-distributed data. This is true in all cases, but even more important when it comes to analysis of extremes, for which the available data is rare. In order to model spatial dependence in extreme data, we need to use more sophisticated statistical techniques. This chapter introduces some such techniques, and is structured as follows:

- In Section 3.1, I present a background in the theory needed to build a spatial model to allow modelling of data extremes over a spatial field.
- In Section 3.2, I describe a Bayesian spatial hierarchical model which can be used to model these extremes and return (samples from) posterior distributions
3.1 Statistical background

This section provides a background in the theory that is needed to construct the more complex models used to analyse and conduct inference on extremes of spatially-distributed data.

3.1.1 Extreme Value Theory

A comprehensive introduction to the field of Extreme Value Theory (EVT), the branch of statistics dealing with the description and inference of extreme values from a distribution, can be found in Coles (2001). The examples and theory in this section are adapted from this book. The fundamental challenge of EVT is that it typically involves estimating the probability of events that are more extreme than any that have already been observed. Coles explains that, though the objective may be, for example, to estimate the maximum sea-level expected in 100 years in order to aid in the design of a new sea-wall, the local data record may only be available for a much shorter time period, for example, 10 years. In order to estimate what sea-levels might occur over the next 100 years, given this 10-year history, we can extrapolate using the framework of EVT (see, e.g., Tawn (1992)). Without empirical or physical rules or guidelines to form an extrapolation rule, standard statistical models are derived from asymptotic
arguments. These models can then be used to provide a solid theoretical framework from which to make projections of the maximum sea-level expected in the next 100 years, along with the uncertainty associated with this estimate, and other statistics of interest to the issue at hand. In the following sections, I discuss the modelling of maxima, but note that there are simple steps that can be taken to ‘reverse’ the data or the model in order to model minima.

3.1.2 Block maxima approach

One approach to model extremes involves what is commonly called the block maxima approach. Continuing with the example above, assume we have hourly sea-levels $Z_1, Z_2, \ldots$. From these, we define:

$$M_n = \max\{Z_1, \ldots, Z_n\}$$

to be the maximum sea-level over an 'n-observation' period. Often $n$ is chosen so that $M_n$ is the annual maximum. Coles points out that, were the exact behaviour of the $Z_i$ known, then the distribution of the set of $M_n$’s could be calculated exactly. However, in practice the behaviour of $Z_i$ is not known. But under suitable assumptions and limit arguments by letting $n \to \infty$, a family of three models emerges which can be calibrated by the observed values of $M_n$. The models differ based on whether they are supported on an infinite or semi-infinite interval. These three models can be re-formulated into a single family of three-parameter models having cumulative distribution functions of the form:

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\varsigma} \right) \right]^{-1/\xi} \right\} \quad (3.1)$$

which are defined on the set \{ $z : 1 + \xi(z - \mu)/\varsigma > 0$\}, and with parameters which satisfy $-\infty < \mu < \infty, \varsigma > 0$, and $-\infty < \xi < \infty$. This is referred to as the Generalised Extreme Value (GEV) family of distributions. The three parameters define the location ($\mu$), the scale ($\varsigma$) and the shape ($\xi$) of the distribution. A value of $\xi < 0$ implies that the distribution is bounded to the right (with a support of $(-\infty, z']$ for some $z' \in \mathbb{R}$), while a value of $\xi > 0$ implies that the distribution is bounded to the left (with a support of $[z', \infty)$ for some $z' \in \mathbb{R}$). Taking the limit
of equation (3.1) as $\xi \to 0$ leads to the third case, $\xi = 0$, where the cumulative distribution function has a support of $(-\infty, \infty)$:

$$G(z) = \exp \left[ - \exp \left\{ - \left( \frac{z - \mu}{\xi} \right) \right\} \right]$$

This re-formulation of three different distributions into one distribution simplifies statistical inference, as the user doesn’t have to introduce subjective judgement to decide, a priori, on the most appropriate distribution; through inference on $\xi$, the data themselves will decide on the most appropriate type of tail behaviour.

### 3.1.3 Threshold models

The approach to modelling block maxima can be wasteful of data - in the motivating example of the sea-wall where 10 years of data are available, it may be the case that the second-highest sea-level in a particular year is higher than the maximum observed for all other years. With the block maxima approach, this valuable piece of data will not be used.

A second general approach to modelling extremes is to use threshold models - so called because they involve retaining only those observations above a particular threshold (Pickands, 1975). Given a sequence of independent and identically distributed random variables $Z_1, Z_2, \ldots$, it is natural to regard those of the $Z_i$ that exceed some high threshold $u$ as being extreme events (Coles, 2001). Assuming that this sequence of variables satisfies the conditions necessary for the extremal types theorem in Section 3.1.2, then for a large enough threshold $u$, the cumulative distribution function of the excesses $Y = Z - u$, conditional on $Z > u$, is approximately given by the Generalised Pareto Distribution (GPD):

$$H(y) = 1 - \left( 1 + \frac{\xi y}{\sigma} \right)^{-1/\xi}$$

which is defined on $\{y : y > 0 \text{ and } (1 + \xi y/\sigma) > 0\}$. Asymptotic arguments again show that equation (3.2) is a limiting distribution as $u$ increases. Here, $\sigma$ and $\xi$ are known as the scale and the shape parameters, respectively, and have ranges $\sigma > 0$ and $-\infty < \xi < \infty$. As before, the sign of $\xi$ indicates the shape of the distribution:
3.1 Statistical background

if $\xi < 0$, the distribution of excesses has a finite upperbound; whereas if $\xi > 0$, the
distribution of excesses has a finite lowerbound. For the limiting value when $\xi = 0$,
the distribution is unbounded, and corresponds to an exponential distribution with
parameter $1/\sigma$.

3.1.4 Return levels

Once we have the parameters of a distribution (methods to find these will be discussed
later in this chapter), we can find values of interest, such as the $N$-year return level,
the level for which the probability that the annual maximum exceeds this value is $1/N$.

From equation (3.2), we have that:

$$P(Z > z|Z > u) = \left(1 + \frac{\xi(z - u)}{\sigma}\right)^{-1/\xi} \tag{3.3}$$

Using the basic laws of probability, we can write:

$$P(Z > z \cap Z > u) = P(Z > z|Z > u)P(Z > u)$$

Since $z > u$, this simplifies to:

$$P(Z > z) = P(Z > z|Z > u)P(Z > u) \tag{3.4}$$

Writing $\zeta_u = P(Z > u)$, and using (3.3) and (3.4), we can then find the return
level $z_m$, the level which is exceeded on average once every $m$ observations, by solving:

$$P(Z > z_m) = \zeta_u \left(1 + \frac{\xi(z_m - u)}{\sigma}\right)^{-1/\xi} = \frac{1}{m}.$$ 

Letting $m = Nn_y$, where $n_y$ is the number of observations per year, this leads to
the following expression for the $N$-year return level:

$$z_N = u + \frac{\sigma}{\xi} \left[(Nn_y\zeta_u)^{\xi} - 1\right]$$
3.1 Statistical background

3.1.5 Likelihood-based inference

Given a set of data (spatially distributed or otherwise), assume we want to fit a statistical model to it. The method of maximum likelihood estimation (MLE) is a common way to do this. Consider a set of \( n \) independent values, \( z_1, z_2, \ldots, z_n \), to which we wish to fit a probability density function (pdf) \( f(z; \theta) \) (which, for the GEV or GPD, would be the derivative of the corresponding cumulative densities presented in Sections 3.1.2 and 3.1.3), where \( \theta \) is the (possibly multiple-valued) parameter of the distribution. For example, \( \theta = (\sigma, \xi) \), the two parameters for the GPD in equation (3.2). Note that these two parameters may themselves be vectors, if the data is spatial in nature. Therefore \( \theta \) is used for brevity and to simplify notation, and represents all parameters of the pdf. The likelihood function involves viewing the pdf as a function of the parameter \( \theta \) rather than as a function of the data. The data, having been observed, are considered now to be fixed. Assuming the \( n \) observations are independent (and therefore the joint pdf is the product of the individual pdfs) leads to the following expression for the likelihood of \( \theta \):

\[
L(\theta) = f(z|\theta) = \prod_{i=1}^{n} f(z_i; \theta)
\]  

The maximum likelihood estimator \( \hat{\theta} \) is found by maximising the likelihood function in equation (3.5). Practically, this often involves maximising the logarithm of \( L(\theta) \) instead, since this is a more computationally stable task (and as the logarithm is a monotonic function, both expressions will be maximised by the same value). The outcome of this is that we find the value(s) of the parameter(s) so that, under the assumed statistical model, the observed data is most probable. Once the maximum likelihood estimator \( \hat{\theta} \) is found, its asymptotic properties (which assume Gaussian behaviour) can then be used to compute confidence intervals. Confidence intervals for quantities derived from the parameter estimates (such as return levels calculated using equation (3.3)) can also be found using the so-called delta method. Details of this approach is given in Coles (2001) (pp. 33-36), along with a discussion of other methods for fitting and analysing uncertainty, such as the profile likelihood method.
3.1 Statistical background

3.1.6 Bayesian inference

An alternative to the MLE approach to parameter estimation is instead to use Bayesian inference (Gelman et al., 2013). Continuing the notation above, we can use Bayes’ Theorem to write:

\[ f(\theta|z) \propto f(z|\theta)f(\theta) \]  (3.6)

The left-hand side of equation (3.6) is a posterior distribution, \( f(\theta|z) \), which is proportional to the product of the likelihood of the data \( f(z|\theta) \) and a given prior distribution \( f(\theta) \). This Bayesian approach yields a complete distribution for the parameter(s) of interest, rather than the point estimate resulting from the MLE approach. From these posterior distributions, percentile values can be used to quantify parameter uncertainty in a natural way, rather than relying on the asymptotic arguments needed under the MLE approach. The Bayesian approach essentially allows the conversion of an initial set of beliefs about \( \theta \) into a complete posterior distribution, that includes the additional information provided by the data \( z \) (Coles (2001), p. 170).

A detailed description of Bayesian theory and methods may be found in Gelman et al. (2013). Coles (2001) provides a brief introduction and motivating example to its application in EVT (pp. 169-177). In a study on precipitation in Venezuela, Coles et al. (2003) demonstrate the superior performance of a Bayesian approach over a classical MLE approach to modelling precipitation extremes there. The authors argue that the Bayesian approach is the most natural way to take all uncertainties into account.

3.1.7 Metropolis-Hastings Markov chain Monte Carlo

Finding the posterior distribution \( f(\theta|z) \) in equation (3.6) involves solving (often multidimensional) integrals. Typically, these integrals are intractable, and an alternative approach to finding the posterior distribution is needed. Markov chain Monte Carlo (MCMC) methods are a common solution to this problem of intractability. Rather than trying to find the analytical form of the posterior distribution, MCMC methods involve a collection of algorithms which draw samples from the target posterior distri-
3.1 Statistical background

If enough independent samples are collected, then these can be used to approximate the posterior distribution, and subsequent analysis can be conducted based on these samples, using histograms, posterior means, medians, variances, percentile values etc. Conducting Bayesian inference on spatial data using MCMC typically involves the need to invert or factor matrices (these matrices arise from the spatial model used to describe the dependence between the \( n \) gridpoints). Without introducing any simplifying assumptions about this relationship across the spatial domain, the required computations are \( O(n^3) \) (Gelfand et al., 2010). This can become computationally challenging or even impractical as the dimension of the spatial field grows.

One approach to drawing samples of the parameter \( \theta \) (from the target posterior distribution) is to use the Metropolis-Hastings MCMC algorithm to sample from an approximation to the densities (described in Geyer (2011)). Where a new sample \( \theta^* \) is proposed to replace \( \theta \), it is accepted or rejected at random according to the rule outlined below. Of course, \( \theta \) may in fact be a vector. A common approach to MCMC is that each element of \( \theta \) is updated separately according to its conditional distribution, given the data and all other parameters in the model. Another approach involves block updates, where blocks of (usually highly correlated) parameters are updated in the same step (Sargent et al., 2000). This approach can be inefficient however (with slow convergence), and the authors point out that success stories in blocking are often application-specific. When each parameter is updated separately, \( \theta \) and \( \theta^* \) only differ in one entry, which may then allow simplification of the calculations.

In order to draw samples, we need to calculate the Hastings ratio:

\[
\alpha(\theta, \theta^*) = \min \left( 1, \frac{L(\text{data}|\theta^*, \text{covariates}, \text{other parameters})p(\theta^*)q_i(\theta^* \rightarrow \theta)}{L(\text{data}|\theta, \text{covariates}, \text{other parameters})p(\theta)q_i(\theta \rightarrow \theta^*)} \right)
\]

Following this calculation, we always accept proposed sample \( \theta^* \) when \( \alpha(\theta, \theta^*) \) equals 1, and we reject accordingly when the ratio is smaller than 1 by simulating a uniform random variable \( u \sim \mathbb{U}[0,1] \), and accepting proposed value \( \theta^* \) when \( u \leq \alpha(\theta, \theta^*) \). Here, \( L(\text{data}|\theta^*, \text{covariates}, \text{other parameters}) \) is viewed as a function of the parameter \( \theta^* \) and gives the likelihood of the data, under the assumed model,
given all covariates and the values of all other parameters; \( p(\theta^*) \) is the prior probability that the parameter \( \theta = \theta^* \); and \( q_t(\theta^* \rightarrow \theta) \) is the transition probability of proposing old value \( \theta \) given newly proposed value \( \theta^* \). This is often taken to be a normal distribution, where a random sample \( \theta^* \) is drawn from the distribution \( \mathcal{N}(\theta, \nu) \) where \( \nu \) is a manually set tuning parameter used to control the size of the proposed steps. If this is too big, values too far from the current value will be proposed, and most likely rejected, and so the algorithm takes a long time to target the posterior density. If this is too small, most of the proposed values will be accepted, but the algorithm will take too long to fully explore the parameter space.

### 3.1.8 Spatial methods

In order to take advantage of the spatial dependence in datasets of extremes observed across a spatial field, it is necessary to establish some of the background theory underlying traditional spatial methods. Typically when dealing with random variables recorded as point-referenced (or geostatistical) data, the location index \( s \) is assumed to vary continuously over \( D \), a fixed subset of \( \mathbb{R}^d \) (Banerjee et al., 2014). Let \( T(s) \) be a vector of random variables (a random vector) at locations \( s \). This could be, for example, measurements of daily maximum wave height at locations \( s \). While it is sensible to conceptually assume such values exist at all possible sites in the spatial domain, in practice the data is a partial realisation of this continuous spatial process. Given this partial realisation, the problem then becomes inference about this spatial process \( T(s) \) as well as prediction at new locations. To this end, it is assumed that the covariance between the random variables at two locations depends on the distance between these locations. That is, \( \text{Cov}(T(s_i), T(s_j)) = C(s_i, s_j) = C(d_{ij}) \) is a function of \( d_{ij} \) where \( d_{ij} \) is the distance between locations \( s_i \) and \( s_j \) (for brevity, the dependence on sites \( i \) and \( j \) is dropped below). The method of calculating this distance must be specified (with Euclidean distance the most common approach). There are many choices of covariance functions (see e.g., Diggle et al. (1998), Gelfand et al. (2010) (pp. 22-26) or Banerjee et al. (2014) (pp. 25-30) for a description of several parametric models, and their relative merits). A popular and flexible choice is the Matérn class of covariance functions (Matérn, 1960), with univariate form:
3.1 Statistical background

\[ C(d) = \begin{cases} \frac{\varsigma^2}{2^{\nu-1}(\nu \pi)^{\nu/2}} K_\nu(\phi d) & \text{if } d \neq 0 \\ \tau^2 + \varsigma^2 & \text{if } d = 0. \end{cases} \]

Here, \( \varsigma^2 \) is the partial sill (variance of the spatial effect), \( \tau^2 \) is the nugget (variance of the non-spatial effect), \( \nu \) is a parameter controlling the smoothness of the spatial field, \( \phi \) is a spatial decay parameter controlling how quickly the covariance decreases with distance, \( \Gamma() \) is the gamma function, and \( K_\nu \) is the modified Bessel function of the second kind of order \( \nu \).

There are alternatives to using the Matérn class of covariance functions, such as kernel convolution (or moving average) models, and convolutions of covariance models (Gelfand et al., 2004). However, the Matérn class of functions is a flexible class, with parameters that have attractive interpretations, and includes as special cases the exponential (by setting \( \nu = 1/2 \)) and the squared exponential (which arises in the limit as \( \nu \to \infty \) with the scale parameter set at \( \phi = 2\sqrt{\nu} \)) covariance functions (Gelfand et al., 2010). In this study, it is the squared exponential covariance function with which we choose to work.

3.1.9 Gaussian Processes

The process \( T(s) \) is said to be Gaussian if, for any \( n \geq 1 \) and any set of sites \( \{s_1, s_2, \ldots s_n\} \), \( T = (T(s_1), T(s_2), \ldots T(s_n))^T \) has a multivariate normal distribution (Banerjee et al., 2014). Gaussian Processes (GPs) can be thought of as extending the finite multivariate normal distribution to infinitely many random variables; in other words, a GP is an infinite collection of variables such that every finite subset follows a multivariate normal distribution. This is a very flexible framework for modelling spatial data, as the covariance matrix can be specified using any valid covariance function.

Given realisations of the process \( T(s) \), and \( p \) spatially-referenced covariates at the same locations \( s \), let \( X(s) \) be the \( n \times (p + 1) \) matrix associated with the spatial regression model:

\[ T(s) = X^T(s) \alpha + w(s) + \epsilon(s) \]
3.1 Statistical background

where $X^T(s)\alpha$ is the mean response, $w(s)$ is a zero-centred GP with covariance function $C(s, s')$ and $\epsilon(s) \sim N(0, \tau^2)$ is an independent measurement error (see, e.g., Finley et al. (2009) and Cressie and Wikle (2011)).

There are alternative computational approaches to fitting GPs using MCMC such as the integrated nested Laplace approximation (INLA) approach proposed by Rue et al. (2009), and INLA combined with a stochastic partial differential equation approach (INLA-SPDE) proposed by Lindgren et al. (2011). However, we have chosen to use GPs due to the ease with which they fit into a Bayesian hierarchical framework. They are flexible empirical models, which are appropriate for an irregularly fluctuating and real-valued spatial surface (Diggle et al., 2007), as we have here. The following section describes how Bayesian analysis can be conducted using hierarchical modelling.

3.1.10 Hierarchical modelling

Bayesian analysis requires us to specify a prior distribution on the parameter(s) $\theta$. This can be conceptually difficult; for example, if one component of $\theta$ is the GPD’s shape parameter $\xi$, a prior distribution for this parameter needs to encapsulate our prior knowledge on its behaviour across the spatial field. This is not at all trivial. If little knowledge is available, diffuse or even improper priors can be used, but care needs to be taken to ensure that the resulting posterior distribution is proper and we need to ensure that our concluding inference is not overly sensitive to these initial assumptions (Gelman et al., 2013).

To overcome this difficulty, hierarchical modelling offers an intuitively simpler approach. Instead of trying to fix a distribution for $\theta$, the hierarchical approach allows us to place flexible distributions on our parameters of interest (which describe the process we’re studying), where these distributions in turn depend on parameters of their own. The result is a hierarchy of distributions at different layers. These parameters could themselves depend on more parameters in an additional level, and so on ad infinitum, but typically in spatial hierarchical models, this process is stopped after three levels. Berliner (1996) describes this approach of modelling complex
3.1 Statistical background

processes in the presence of data (whether spatial, as here, or time series, as in his case), by breaking it into three primary stages:

1. Data Model: \([\text{data}|\text{process, parameters}]\)
2. Process Model: \([\text{process}|\text{parameters}]\)
3. Parameter Model: \([\text{parameters}]\)

The first stage is concerned with the model for the data, which specifies the distribution of the data, given the process of interest as well as the parameters that describe the data model. The second stage then involves specifying a distribution for the process of interest, conditional on other parameters. The final stage accounts for the uncertainty in these other parameters by endowing these with distributions. If the process is multivariate and spatial (as we have here), Berliner (1996) suggests that it might be modelled as a product of physically motivated distributions, as suggested by some scientific relationship or physical constraint. The benefit of this hierarchical approach is that it can be easier to justify distributions for the parameters in the third level (these are often called hyperparameters to avoid confusion with the parameters governing the data model) than for the initial parameters (or process).

Applying the conceptual model above using the previous notation then, and using Bayes’ Theorem, a three-layer spatial hierarchical model for the inference of the vector of parameters \(\theta\) (e.g., where \(\theta = (\sigma, \xi)\) for the GPD) is given by:

\[
p(\theta|Z(x)) \propto p_1(Z(x)|\theta)p_2(\theta|\gamma)p_3(\gamma)
\]

where the \(p_j\) are the probability densities associated with the \(j^{th}\) level of the hierarchy, \(Z(x)\) represents the data at location \(x\), and \(\gamma\) represents the (possibly vector of) hyperparameters.

With this concept of hierarchical modelling in mind, combined with the statistical theory established in the earlier sections, we now have a solid basis with which to suggest a model for the dataset of significant wave heights which was presented in Chapter 1. This leads us to the next section, where I remind the reader about this waves dataset, and then present and explain in detail the Bayesian spatial hierarchical model which we propose to use to model this data.
3.2 Model details for the waves dataset

The data analysed for the remainder of this chapter is output from the third-generation spectral WAVEWATCH III version 4.11 model, which was introduced and described in Section 1.4.2. This involves hourly measurements on wave height data simulated on an unstructured triangular grid of 15,000 nodes. The region off the west coast of Ireland which we focus on for this study contains 334 gridpoints and has a depth ranging from 39 m to 1902 m (Chapter 1; Figure 1.5). We want to model significant wave height, $H_s$, to help better inform planning for future off-shore infrastructure and development. There is clearly a spatial relationship visible across the domain (Chapter 1; Figure 1.6), where the lowest values of $H_s$ are those nearest the shore in the east of the domain, with a general increase in $H_s$ the farther west the gridpoint is. A spatial model which allows this dependence to inform the posterior distributions will reduce uncertainty over a point-based MLE approach. Though using spatial modelling to improve estimation of extremes in oceanography is not a new idea (see e.g., Dixon et al. (1998)), advances in computational resources mean that fitting Bayesian spatial hierarchical models to large datasets such as these is now feasible. To our knowledge, such a Bayesian spatial analysis has not been conducted to date on this domain off the Irish coast.

3.2.1 Model overview

The aim of this study is to produce $N$-year return levels of significant wave height. I described earlier how these could be computed using equation (3.3). Given a dataset, we require a suitable threshold $u$ (above which observations will be considered extreme), the parameters from the GPD, and the probability of exceeding the threshold $\zeta_u$. The non-trivial task of choosing a suitable threshold $u$ will be discussed in detail in Section 3.2.4.

For the exceedances of the threshold $u$, and the probability $\zeta_u$ of exceeding this threshold, we follow the approach of Cooley et al. (2007), and employ two Bayesian spatial hierarchical models with three layers. As described in Section 3.1.10, the first layer consists of linking the data to some parameters through a probability distribution.
3.2 Model details for the waves dataset

The second describes the latent spatial process underlying the extremes in the region, while the third layer consists of the prior distributions on the hyperparameters (that is, the parameters controlling the second layer). Next, I present a detailed description of the two hierarchical models, which is then followed by a graphical representation of these models; namely a directed acyclic graph (DAG), which is shown in Figure 3.1. (Further details involving the pseudocode and equations for this model are to be found in Appendix A at the end of this chapter on page 95.)

3.2.2 Modelling the threshold exceedances

3.2.2.1 Data layer

We use a GPD given by equation (3.2) to model the data at the first layer of the hierarchy. In order to ensure a positive scale parameter throughout the computations, we reparameterise with \( \phi = \log \sigma \). At this level, we thus have two spatially-varying parameters for the distribution, which we collectively write as \( \theta = [\phi(x), \xi(x)]^T \). The first term in the hierarchy given by equation (3.7) is the product of the pdfs for the GPD (found by differentiating the cumulative density in equation (3.2)) over all locations and all observations to form the likelihood function:

\[
p_1(Z(x) | \theta) = \prod_{i=1}^{n} \prod_{k=1}^{n_i} \frac{1}{\exp \phi(x_i)} \left( 1 + \frac{\xi(x_i) z_k(x_i)}{\exp \phi(x_i)} \right)^{-1-1/\xi(x_i)}
\]

(3.8)

where the indices \( i \) and \( k \) are such that \( z_k(x_i) \) refers to the \( k \)-th excess at gridpoint \( x_i \). We have denoted the number of gridpoints by \( n \) and the number of excesses at each point \( x_i \) is then \( n_i \).

3.2.2.2 Process layer

Both \( \phi(x) \) and \( \xi(x) \) are modelled as GPs, and so the second term in equation (3.7) will take the form:

\[
p_2(\theta | \gamma) = p_\phi(\phi(x) | \mu_\phi, \Sigma_\phi)p_\xi(\xi(x) | \mu_\xi, \Sigma_\xi)
\]
3.2 Model details for the waves dataset

where

\[ p_\phi(\phi(x)|\mu_\phi, \Sigma_\phi) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_\phi|}} \exp \left[ -\frac{1}{2} (\phi - \mu_\phi)^T \Sigma_\phi^{-1} (\phi - \mu_\phi) \right]. \] (3.9)

A similar expression is used for \( p_\xi(\xi(x)|\mu_\xi, \Sigma_\xi) \). Here \(| . |\) denotes the determinant and \( \gamma \) above represents all of the hyperparameters for \( \mu_\phi, \mu_\xi, \Sigma_\phi \) and \( \Sigma_\xi \), to be discussed below.

A Gaussian process characterises an infinite-dimensional smooth surface such that any finite collection of \( n \) points on the surface follows a multivariate normal distribution (above) of dimension \( n \). Such a smooth surface is an appropriate choice for the model parameters as we expect similar wave climates at nearby locations.

In addition to distance, the effect of any other covariates may be readily incorporated into the model. For \( m \) covariates \( c^{(1)}, \ldots, c^{(m)} \), we write the mean vector in the general form

\[ \mu_\phi = \mathcal{C} \alpha_\phi \] (3.10)

where \( \mathcal{C} \) is the \( n \times (m + 1) \) matrix

\[
\mathcal{C} = \begin{pmatrix}
1 & c^{(1)}_1 & c^{(2)}_1 & \cdots & c^{(m)}_1 \\
1 & c^{(1)}_2 & c^{(2)}_2 & \cdots & c^{(m)}_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & c^{(1)}_n & c^{(2)}_n & \cdots & c^{(m)}_n 
\end{pmatrix}
\]

and the vector of coefficients is \( \alpha_\phi = (\alpha_{\phi,0}, \alpha_{\phi,1}, \ldots, \alpha_{\phi,m})^T \). In this work we have used \( m = 3 \): the longitude, latitude and depth of a gridpoint, and therefore we will have four coefficients \( \alpha_{\phi,j}, j = 1 \ldots 4 \).

The use of GPs also offers great flexibility through the choice of the covariance matrix \( \Sigma_\phi \). Here, we use the matrix given by

\[ \Sigma_\phi = \kappa^2_\phi \mathcal{E} + \tau^2_\phi I \] (3.11)
3.2 Model details for the waves dataset

where \( I \) is the identity matrix. The matrix \( E \) is given by a squared exponential correlation function (which is a limiting case from the Matérn class of covariance functions, as detailed in Section 3.1.8) and has components

\[
E_{i,j} = \exp \left( -d(i,j)^T \beta^{-1} d(i,j) \right).
\]

(3.12)

For two grid-points \( x_i \) and \( x_j \), the vector \( d(i,j) \) has two components given by the differences in longitude and latitude between \( x_i \) and \( x_j \).

The \( 2 \times 2 \) matrix \( \beta \) is symmetric positive definite. Its entries measure how quickly spatial dependence drops off in the two different directions. The other parameters of the covariance matrix appearing in equation (3.11) are the partial sill \( \varsigma^2 \) and the nugget parameter \( \tau^2 \). Further details on variogram analysis may be found in Cressie and Wikle (2011). As mentioned, we assume the same Gaussian process model for the shape parameter \( \xi \). Similar expressions as those above are used for \( \mu_\xi \) and \( \Sigma_\xi \).

3.2.2.3 Priors layer

For the third and final layer in the hierarchical model, prior distributions must be assigned to the hyperparameters in the second layer, which are assumed to be independent. For those in equation (3.10), a normal distribution with large variance was selected. The covariates were re-scaled to be centred on zero and priors \( \alpha_{\phi,j}, \alpha_{\xi,j} \sim N(0,50) \) were used. A lognormal prior was employed for the positive \( \varsigma^2 \) and \( \tau^2 \) parameters in equation (3.11); that is, their logarithm was assumed to have the normal distribution \( N(0,10) \). For the entries of the matrix \( \beta \) in (3.12), a discrete uniform prior was assumed. We began with a set of proposal values \( v_\beta = \{0.001, 0.005, 0.01, 0.05, \ldots, 100, 500, 1000\} \), with all values being considered equally likely a priori. Within the MCMC algorithm, the entries of \( \beta \) are randomly proposed from \( v_\beta \). \( \beta \) is then checked to confirm it is still symmetric positive definite, and if so, this updated value is then accepted or rejected accordingly by the algorithm.

Using Bayesian inference allows valuable additional information about a process to be incorporated in the form of prior information. This can be of great benefit in our case, due to the scarcity of extreme data. However, Coles and Powell (1996) argue
that with such scarce data, an expert may not be able to independently formulate prior beliefs about this process. With this in mind, our chosen prior distributions are based on physically plausible values, but with enough flexibility so that the data is not restricted from informing the posterior distributions. (In the analysis of daily maximum temperatures to follow in Chapter 4, a more thorough justification into the choice of prior distributions is given.)

3.2.3 Modelling the frequency of exceedances

We now turn to $ζ_u$, which is defined as the probability that the threshold $u$ is exceeded. This value is needed in equation (3.3) to compute return levels. For a given choice of threshold (discussed in Section 3.2.4), we let $ζ(x_i)$ be the exceedance probability at the location $x_i$. It is again assumed that there is a latent spatial process driving this and a hierarchical model is used, with data, process and prior layers.

At the data layer it is assumed that, at each gridpoint $i$, the number of threshold exceedances $N_i$ is a binomial random variable with $m_i$ trials (the total number of observations), each with a probability $ζ(x_i)$ of exceeding the threshold. That is: $N_i \sim \text{Bin}(m_i, ζ(x_i))$, where

$$P(N = N_i) = \binom{m_i}{N_i} ζ(x_i)^{N_i} (1 - ζ(x_i))^{m_i - N_i}.$$  

The process layer for $ζ(x)$ is similar to that of the GPD parameter $φ(x)$. Following the methodology in Diggle et al. (1998), we first transform $ζ(x_i)$ using a logit transformation, where

$$ζ'(x_i) \equiv \logit(ζ(x_i)) = \log\left(\frac{ζ(x_i)}{1 - ζ(x_i)}\right). \tag{3.13}$$

$ζ'$ is then modelled as a GP as in equation (3.9), with mean vector $μ_{ζ'}$ and covariance matrix $Σ_{ζ'}$ taking the same form as in equations (3.10) and (3.11), respectively. The hyperparameters are then given the same prior distributions as described above in Section 3.2.2.3.
3.2 Model details for the waves dataset

Figure 3.1: A directed acyclic graph (DAG) of the Bayesian spatial hierarchical models fitted to the wave data. On the left is the model for the excesses (using the GPD to model the data) and on the right is the model for the probability of an observation exceeding the threshold (using the Binomial distribution). The parameters of the distributions, modelled using equation (3.9), are represented as circles in the middle layer, with the hyperparameters controlling these represented in the top layer. The data, modelled using equation (3.8) for \( z_k(x_i) \) and using equation (3.13) for \( N_i \), is represented in the bottom layer (in rectangles). Arrows run into nodes from their direct predecessors (often called parents). Given its parents, each node is independent of all other nodes in the graph except its descendants (often called children). Posterior estimates of the parameters’ distributions can be used to form quantities of interest – typically return levels, as illustrated. Further details of each layer and the parameters involved may be found in the text.

3.2.4 Threshold choice and declustering

We wish to apply the GPD model of threshold exceedances to this dataset of wave heights. The choice of an appropriate threshold is a non-trivial issue in extreme value modelling and the subject of much research (see Scarrott and MacDonald (2012) and Caeiro and Gomes (2016) for recent reviews, for example). With a low threshold, the asymptotic validity of the GPD may be violated, leading to bias in the posterior estimates. On the other hand, if the threshold is set too high we will be left with too few data points for fitting the model, resulting in large variances. Coles (2001) describes commonly-used graphical methods for choosing the threshold for a single time series, based on the asymptotic properties of the GPD given in equation (3.2). However these are rather subjective and furthermore are not suited to modelling over a region with multiple locations. Numerous other methods have been described in
3.2 Model details for the waves dataset

the literature. Thompson et al. (2009) automate the choice using a goodness-of-fit test for the distribution of successive parameter estimate differences as the threshold is increased. A quantile regression model was employed by Northrop and Jonathan (2011), while Dupuis (1999) used optimal bias robust estimators to fit the model and test for the validity of the threshold. In Tancredi et al. (2006), the authors incorporate the uncertainty from the threshold choice by including this within the model estimation procedure. In Kiriliouk et al. (2018), the authors used a measure of asymptotic dependence to guide threshold selection when dealing with extreme data in a multivariate context to fit a multivariate GPD.

In this study, we adopt a more straightforward approach. At each gridpoint, the 99.8th percentile of the $H_s$ data series is selected as a threshold for modelling at that point. Taking a percentile-based threshold is convenient when dealing with a spatial array of data. In Clancy et al. (2015), the 97th percentile was used. Caires and Sterl (2005) examined both the 93rd and 97th percentile and found the higher to be more appropriate in general. Vanem (2015) tested thresholds based on even higher percentiles and found, in some cases, a value around the 99.5th to be too low. The validity of the threshold may be assessed a posteriori by examining the fit of the model, and we discuss this choice further below.

Once we have chosen our threshold, we need to decluster the exceedances of this threshold. This is necessary because the theoretical basis for the use of the GPD assumes that the exceedances are independent. The process is described in general terms in Coles (2001) (p. 99):

1. Use an empirical rule to define clusters of exceedances;
2. Identify the maximum excess within each cluster;
3. Assume this cluster maxima to be independent, with conditional excess distribution given by the GPD;
4. Fit the GPD to the cluster maximum.

Caires (2011) retains only the maxima of clusters of successive exceedances and additionally removes any peaks which occur less than 48 hours from another, these
3.2 Model details for the waves dataset

regarded as having been caused by the same storm system. Nicolae Lerma et al. (2015) varied this time between 48 and 72 hours and found no significant differences in their final results.

In this study we apply a similar, though slightly stricter, method of declustering to that of Caires (2011). Two successive sequences of exceedances are considered to be part of the same cluster and system if the time series drops below the threshold for 48 hours or less. We then use the peaks of each cluster for modelling with the GPD. With this approach, \( \zeta(x_i) \) from Section 3.2.3 then represents the probability of being a cluster maximum rather than the probability of exceeding the threshold, again following the methodology employed by Cooley et al. (2007).

3.2.5 Model fitting

The GPD was fitted to the dataset discussed above using the Bayesian spatial hierarchical model (BHM). Approximate draws from the posterior distribution of each parameter in the hierarchical model were obtained via the MCMC algorithm. Metropolis-Hastings steps were employed to update each parameter in turn, for each iteration of the MCMC algorithm. This involves drawing a potential value from an appropriate distribution and accepting or rejecting it according to the Hastings ratio of the Metropolis-Hastings algorithm (Geyer, 2011).

Three parallel chains were run for each model. Each simulation consisted of 20,000 iterations, of which the first 2,000 were considered as burn-in and consequently discarded. In order to reduce dependence amongst the remaining values, only every 10th was kept. Convergence of the resulting chains was established using the \( \hat{R} \) criterion recommended by Gelman (1996), with values below the suggested criterion of 1.2 taken to imply convergence.

The model was implemented in R using a package called Rcpp (Eddelbuettel, 2013). This interface allows integration of R with C++ code, leading to appreciable reduction in the computational burden of the Metropolis-Hastings MCMC algorithm used. All code needed to reproduce our analysis is available in a public repository.
3.2 Model details for the waves dataset

on GitHub at https://github.com/jackos13/waves. Note that this repository does not include the data; they are available upon suitable request from the corresponding author of the article “A long-term nearshore wave hindcast for Ireland: Atlantic and Irish Sea coasts (1979-2012)” (Gallagher et al., 2014).

In presenting the results in the following section, I compare the results with those obtained by fitting the distribution independently at each gridpoint with inference from an MLE approach. In addition to spatial maps of the output, we chose four locations to focus on in more detail. These are marked in Figure 3.2. Their locations and some hindcast details are listed in Table 3.1.

<table>
<thead>
<tr>
<th>Point</th>
<th>Lon.</th>
<th>Lat.</th>
<th>Depth</th>
<th>Mean</th>
<th>99.8th</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-9.66</td>
<td>52.88</td>
<td>68.84</td>
<td>2.42</td>
<td>8.70</td>
<td>13.51</td>
</tr>
<tr>
<td>2</td>
<td>-11.22</td>
<td>53.10</td>
<td>134.74</td>
<td>3.12</td>
<td>10.03</td>
<td>15.21</td>
</tr>
<tr>
<td>3</td>
<td>-11.60</td>
<td>53.97</td>
<td>329.90</td>
<td>3.21</td>
<td>10.24</td>
<td>15.46</td>
</tr>
<tr>
<td>4</td>
<td>-10.70</td>
<td>53.91</td>
<td>130.17</td>
<td>3.12</td>
<td>10.08</td>
<td>15.24</td>
</tr>
</tbody>
</table>

Table 3.1: The longitude, latitude (both in degrees) and depth (metres) of each of the four points indicated in Figure 3.2, along with the mean, 99.8th percentile and maximum significant wave height (metres) from the 34-year hindcast.
Figure 3.2: The depth of the study region (in metres) plotted on a logarithmic scale, with the four locations from Table 3.1 marked.

As discussed in Section 3.1.5 above, the MLE approach produces a single estimate of a given parameter and confidence intervals may be derived from its asymptotic properties. We considered 95% confidence intervals for our estimates and present the lower and upper bounds of these intervals. The Bayesian model, on the other hand, yields (samples from) the posterior distribution for the parameters. From the values simulated by the MCMC algorithm, we present the median value of this distribution and, again, confidence intervals bounded by the 2.5th and 97.5th percentiles (usually termed a credible interval in the Bayesian context).

3.3 Results

3.3.1 Parameter estimates

In Figures 3.3 and 3.4, we can compare the estimates of the GPD shape (\(\xi\)) and scale (\(\sigma\)) parameters, respectively. In both cases, the BHM shows less variation over the domain when compared with the MLE results. Additionally, in particular for \(\sigma\),
the spatial variation is noticeably smoother in the BHM. This is to be expected, given the latent spatial processes modelled by the BHM, where the covariance of parameter values at different gridpoints is based on the distances between them using equation (3.12). The MLE model involves simply fitting the univariate GPD independently at each gridpoint, with no account taken of the relationship between neighbouring points.

![Figure 3.3: Estimates of the shape parameter $\xi$, using the MLE method (a) and the Bayesian (BHM) method (b).](image)

![Figure 3.4: Estimates of the scale parameter $\sigma$, using the MLE method (a) and the Bayesian (BHM) method (b).](image)

The MLE estimates of the shape and scale parameters (in Figures 3.3 and 3.4)
3.3 Results

display a strong negative correlation between the two surfaces. This suggests that the
MLE method struggles to decide whether the observed extremes are better modelled
either by a large value of the shape or a large value of the scale. This uncertainty is
absent from the BHM method - both median surfaces are smoother than their MLE
counterparts, and lack the negative correlation of the latter.

The lower and upper bounds of the 95% confidence intervals for the estimates of
\( \xi \) and \( \sigma \) are plotted in Figures 3.5 and 3.6, respectively. The benefits of the Bayesian
approach are immediately clear: there is much less uncertainty in the parameter
value estimates, which is evident from the much narrower confidence intervals. This
can be seen in more detail in Figure 3.7, in which we show parameter estimates with
confidence intervals for both models for the four points detailed in Table 3.1. Looking
at Figures 3.7(a) and 3.7(b) for the shape and scale, respectively, we see that at each
of these four points the confidence intervals for the BHM are contained within the
larger intervals for the MLE approach; in fact, this is true for \( \sigma \) at every gridpoint
and \( \xi \) for all but five of the 334 gridpoints.
3.3 Results

Figure 3.5: Confidence interval bounds for the shape parameter $\xi$ using the MLE method (left) and the BHM method (right), given by the 2.5th (above) and 97.5th (below) percentiles of the estimates.
3.3 Results

Figure 3.6: Confidence interval bounds for the scale parameter $\sigma$ using the MLE method (left) and the BHM method (right), given by the 2.5th (above) and 97.5th (below) percentiles of the estimates.
3.3 Results

Figure 3.7: Comparison of estimates from both models for the points marked in Figure 3.2 and detailed in Table 3.1: (a) shape and (b) scale parameters, along with the (c) 20-year and (d) 100-year return levels. The dot marks the point estimate, with the vertical lines indicating the confidence interval bounded by the 2.5th and 97.5th percentiles.

Note that the estimates in Figure 3.3 show a positive shape parameter in general. Looking at the surfaces for the confidence intervals in Figure 3.5 and also for the particular four points in Figure 3.7(a), we see that the MLE method results in a wide range of values, both positive and negative. The narrower intervals of the Bayesian model favour a positive shape.
3.3 Results

Some previous studies fixed the range of the shape parameter in advance: to be, for example, strictly negative (Ortego et al., 2012) or equal to zero (Caires and Sterl, 2005). We have set no such restrictions a priori. This allows the model the flexibility to determine the best values of the parameters, given the data. The appropriateness of this can be addressed when validating the model. This will be discussed when the data is overlaid on model return level plots, which forms the focus of the next section.

3.3.2 Return level surfaces

We now turn to the $N$-year return levels of significant wave height, which we denote by $H_{sN}$. In Figures 3.8 and 3.9, we present spatial plots of $H_{s20}$ and $H_{s100}$, respectively. The estimates are given along with the lower and upper bounds of the 95% confidence interval. The overall patterns of the return levels are broadly similar for both models. The main differences can be seen in the size of the confidence intervals, which again are much narrower in the BHM. This is even more evident for $H_{s100}$ in Fig. 3.9.
3.3 Results

Figure 3.8: MLE (left) and BHM (right) estimates of the 20-year return level of $H_s$. The estimates are shown (top), along with confidence interval bounds given by the 2.5th (middle) and 97.5th (bottom) percentiles.
3.3 Results

Figure 3.9: MLE (left) and BHM (right) estimates of the 100-year return level of $H_s$. The estimates are shown (top), along with confidence interval bounds given by the 2.5th (middle) and 97.5th (bottom) percentiles.

As we extrapolate in time to longer return periods, we expect the uncertainty to grow. However, with the BHM we still have a much tighter confidence interval even at 100 years. Examining the values in Figures 3.7(c) and 3.7(d) makes this more explicit. At Point 4, for example, the MLE method produces a 100-year return
3.3 Results

level in the range of 12.15-21.19 m. The BHM gives a sharper result of 14.90-19.42 m.

We find that the highest extremes of significant wave height are to be expected roughly between 53°N and 54.5°N, with 100-year levels of close to 17 m. This spatial pattern is consistent with the annual maxima GEV analysis of the dataset in Gallagher (2014). Although slightly higher values were reported (up to 18 m), the annual maxima approach results in considerably higher levels of uncertainty in the estimates, due to the much smaller number of datapoints used. In Clancy et al. (2015), a larger area was studied using a coarser-resolution hindcast but, nevertheless, this region (and further west) showed the most extreme sea states.

3.3.3 Site-specific return level plots

Site-specific return level plots, as described in Coles (2001), are a useful diagnostic tool for assessing the fit of a model, in addition to illustrating the model estimates. The return level estimates, along with the bounds of the confidence intervals, are plotted against the return period. Empirical return level estimates from the data are added as follows:

1. Order the observed excesses in a vector z such that \( z_1 \leq z_2 \leq \cdots \leq z_{n_i} \) for the \( n_i \) excesses at location \( i \);
2. then calculate the corresponding vector of return periods \( x \) with entries \( x_k \), (for \( k = 1 \ldots n_i \)) given by:
   \[
   x_k = \frac{1}{1 - k/(n + 1)} \times \frac{1}{npy};
   \]
3. then add all points \( (x_k, z_k) \) to the plot (or \( \log(x_k), z_k \) if plotting on a log scale).

Here \( npy \) refers to the number of excesses per year (cf. sections 2.6.7 and 3.3.5 of Coles (2001) for further details on this process).

These are shown in Figure 3.10 for the four locations in Table 3.1 which we have been examining. The black curves are for the MLE model, while the BHM is shown.
3.4 Discussion and Conclusions

in red. The return level estimates from both models are quite similar for all return periods shown. However, once again we see clearly how the Bayesian model yields estimates with much less uncertainty. This is more evident as we extrapolate to longer return periods, such as 1000 years. Comparing with the empirical estimates shows a satisfactory fit, particularly for shorter return periods.

Figure 3.10: Model return level plots with empirical estimates (open circles) for the four points from Table 3.1. Solid curves are the return level estimates for the MLE (black) and BHM (red) methods using the 99.8th percentile threshold. The dashed curves are the corresponding lower and upper bounds of the 95% confidence intervals. Note the logarithmic scale used for the return period.

3.4 Discussion and Conclusions

In this study, we have applied a Bayesian spatial hierarchical model (BHM) to a hindcast dataset in order to study the extremes of significant wave height off the west
3.4 Discussion and Conclusions

cost of Ireland. Exceedances of $H_s$ over a high threshold were modelled with the generalised Pareto distribution (GPD). The hierarchical model includes a latent spatial process which allows us to more effectively capture the spatial variation of the extremes. This approach was compared with a model which uses maximum likelihood estimation (MLE) inference and simply carries out an independent univariate extreme value analysis on the time-series at each gridpoint in the given domain.

The parameters of the fitted GPD were used to produce spatial maps of extrapolated return levels of $H_s$. Consistent with previous studies, we found that the highest extremes are to be expected in the latitude band roughly between 53°N and 54.5°N to the west of Ireland, with phenomenal sea states of around 17 m estimated for the 100-year return level. A comparison of the two methods showed that the BHM produces smoother estimates. This is a result of modelling the latent spatial process of the parameters, whereby the model parameters at a given gridpoint are influenced by the neighbouring gridpoints.

As we have used data from a wave model hindcast, we should consider how well the model is capturing extremes when interpreting the results. A slight underestimation of the extremes of significant wave height has been reported (Gallagher et al., 2014), consistent with an underestimation in ERA-interim winds driving the model (Stopa and Cheung, 2014). As a check on the validity of our model-derived results, we may compare with those derived from satellite data in Izaguirre et al. (2011). Although at a much lower resolution, for the regions immediately to the west of Ireland they reported wintertime 20-year return levels of around 14-16 m (cf. their Figure 2). These are consistent with our 20-year return levels in Figure 3.8. In future work, nevertheless, it may be worthwhile to consider any deficiencies in numerical model output; see, for example, Caires and Sterl (2005) and Hanafin et al. (2012).

A major advantage of the Bayesian approach is the fact that it formally handles parameter uncertainty, rather than relying on the approximate normality of the MLE estimate. Parameter and return level estimates were analysed with confidence intervals bounded by the 2.5th and 97.5th percentiles. The BHM showed much narrower intervals throughout, yielding much higher levels of certainty with the results. This
3.4 Discussion and Conclusions

is of crucial importance, as a single point estimate alone is of little practical value without a meaningful measure of uncertainty. Indeed, looking at the 1000-year return levels in Figure 3.10, we see upper bounds of nearly 35 m for the MLE, which seem physically unrealistic.

The threshold eventually chosen for this work was the 99.8th percentile of the $H_s$ data at each gridpoint, though other (lower) thresholds were examined. As noted, this approach has been used by a number of authors, with various ranges of percentiles tested. Initially, we fitted the model using a threshold of the 97th percentile, following previous work in Clancy et al. (2015); we later used the 99th percentile. We subsequently analysed the validity of the model fits. We conducted this analysis both visually (by including the data on site-specific return level plots to assess the models’ predictive power) and numerically (by calculating how many of the observed data points were contained within the 95% uncertainty intervals across the domain). This analysis revealed that both models (MLE and BHM) greatly underestimated the higher return levels, particularly when using the 97th percentile. Modelling using the 99th percentile still showed a bias, with the return level intervals for the Bayesian model still failing to capture the higher extremes. Moving to the higher 99.8th percentile as a threshold improved this greatly, as seen from Figure 3.10.

As discussed earlier, a very high threshold is necessary for the asymptotic validity of the GPD model, but can result in too few points with which to fit the model. Mazas and Hamm (2011), for example, recommended a threshold yielding an average of 2-5 excesses per year. Here we have an average of 51 excesses at each location, after declustering, for a 34-year dataset. This still leaves us with many more data points than if modelling with the annual maxima approach, and the validating results outlined above attest to the appropriateness of the high threshold.

Increasing the threshold was also found to have a noticeable effect on the shape parameter. With the lower thresholds, the shape was negative, similar to other reported results (for example, Caires (2011)). But with the 99.8th percentile threshold eventually chosen for this work, we found a generally positive shape parameter. As noted above, the lower thresholds were failing to capture the highest extremes, which
3.4 Discussion and Conclusions

could be seen a result of the negative shape which describes a distribution with a finite upperbound. With the higher threshold and resulting positive shape parameter, the higher values are represented more accurately. Note that positive values of this parameter have been reported elsewhere: in other geographical regions (Izaguirre et al., 2011), for wave datasets from future climate projections (Vanem, 2015) and with the annual maxima GEV model in a similar domain off the west coast of Ireland (Clancy et al., 2015).

In addition to a correlation between points based on the distance between them, we included depth as a covariate in the mean of all second-layer parameters, as outlined in equation (3.10). Other covariates may be added to the model. For example, time-dependent covariates can be used to study the trends in extremes with a changing climate (Caires et al. (2006); Méndez et al. (2006)). Different spatio-temporal models may also be explored, in which extremes at a point are affected not just by neighbouring locations, but by recent extreme events. These more complex models may potentially further reduce the uncertainty levels in the Bayesian model, to yield estimates of increasing practical value. Such an investigation is a possible extension to this study.
References


REFERENCES

*International Conference on Ocean, Offshore and Arctic Engineering*. St John’s, NL, Canada.


REFERENCES


REFERENCES


Appendix A

This appendix includes further details on the model described in Section 3.2; namely, it describes in detail how we modelled a dataset of extreme observations using a generalised Pareto distribution and Metropolis-Hastings Markov chain Monte Carlo updates. Details for the Binomial component of the model are similar, and are omitted here for brevity.

The following pages include:

- the pseudocode in order to programme the algorithm;
- the model directed acyclic graph (DAG);
- details of the model notation;
- and the details of all equations needed for the MCMC updates of the parameters and the hyperparameters.
Algorithm details

Data: $z_k(x_i)$, the declustered threshold excesses at locations $x_i$, $i = 1 \ldots n$, with $k = 1 \ldots n_i$ excesses at location $i$;
$X_\phi$, $X_\xi$, $n \times (p + 1)$ and $n \times (q + 1)$ matrices of $p$ and $q$ covariates at locations $i = 1 \ldots n$.

Result: Samples from the posterior distributions of $\phi = \log(\sigma)$ and $\xi$ (the unknown parameters of interest), which can then be used to calculate return level estimates and other desired quantities

Initialisation;
Random starting values of $\phi$ and $\xi$;
Hyperparameter values of $\alpha_\phi$, $\beta_\phi$, $\varsigma_\phi^2$, $\tau_\phi^2$, $\alpha_\xi$, $\beta_\xi$, $\varsigma_\xi^2$, and $\tau_\xi^2$;
Number of iterations $N$;

for iterations $j$ from 1 to $N$ do
  Generate $u \sim U(0, 1)$;
  for grid locations $i$ from 1 to $n$ do
    Simulate $\phi_{\text{new},i}$;
    Set $l_{\text{new}} = \log$ full conditional of new vector $\phi_{\text{new}}$;
    Set $l_{\text{old}} = \log$ full conditional of old vector $\phi$;
    Set $a = \exp(l_{\text{new}} - l_{\text{old}})$, that is, evaluate equation (3.14);
    if $a > u$ then
      Set $\phi = \phi_{\text{new}}$;
    end
    Simulate $\xi_{\text{new},i}$;
    Set $l_{\text{new}} = \log$ full conditional of new vector $\xi_{\text{new}}$;
    Set $l_{\text{old}} = \log$ full conditional of old vector $\xi$;
    Set $a = \exp(l_{\text{new}} - l_{\text{old}})$, that is, evaluate equation (3.15);
    if $a > u$ then
      Set $\xi = \xi_{\text{new}}$;
    end
  end
end
for iterations \( j \) from 1 to \( N \) (continued) do

   for each element \( c \) of the vector \( \alpha_\phi \) do
      Simulate \( \alpha_{\phi_{\text{new},c}} \).
      Set \( a = \) result of equation (3.16);
      If \( a > u \), set \( \alpha_\phi = \alpha_{\phi_{\text{new}}} \);
   end

   for each element \( d \) of the vector \( \alpha_\xi \) do
      As above, but set \( a = \) result of equation (3.17);
   end

   for elements \( e \) in the lower-triangle of the matrix \( \beta_\phi \) do
      Simulate \( \beta_{\phi_{\text{new},e}} \), repeat until symmetric positive definite \( \beta \) results;
      Set \( a = \) result of equation (3.18);
      If \( a > u \), set \( \beta_\phi = \beta_{\phi_{\text{new}}} \);
   end

   for elements \( f \) in the lower-triangle of the matrix \( \beta_\xi \) do
      As above, but set \( a = \) result of equation (3.19);
   end

   Simulate \( \varsigma_\phi^2_{\text{new}} \);
   Set \( a = \) result of equation (3.20);
   If \( a > u \), set \( \varsigma_\phi^2 = \varsigma_\phi^2_{\text{new}} \);
   Repeat for \( \varsigma_\xi^2 \) with equation (3.21) ;

   Simulate \( \tau_\phi^2_{\text{new}} \);
   Set \( a = \) result of equation (3.22);
   If \( a > u \), set \( \tau_\phi^2 = \tau_\phi^2_{\text{new}} \);
   Repeat for \( \tau_\xi^2 \) with equation (3.23) ;
DAG

Figure 3.11: A directed acyclic graph (DAG) of the Bayesian spatial hierarchical models fitted to the wave data. On the left is the model for the excesses (using the GPD to model the data) and on the right is the model for the probability of an observation exceeding the threshold (using the Binomial distribution). The parameters of the distributions, modelled using equation (3.9), are represented as circles in the middle layer, with the hyperparameters controlling these represented in the top layer. The data, modelled using equation (3.8) for \( z_k(x_i) \) and using equation (3.13) for \( N_i \), is represented in the bottom layer (in rectangles). Arrows run into nodes from their direct predecessors (often called parents). Given its parents, each node is independent of all other nodes in the graph except its descendants (often called children). Posterior estimates of the parameters’ distributions can be used to form quantities of interest - typically return levels, as illustrated. Further details of each layer and the parameters involved may be found in the text.
Appendix A

Notation

- \( x_i \) Multivariate (bivariate or tri-variate) location values for location \( i, i = 1, \ldots, n \). Write the matrix of all locations as just \( x \)
- \( z_k(x_i) \) \( k \)-th observation at location \( i, k = 1, \ldots, n_i \) where \( n_i \) is the number of excesses at location \( i \)
- \( \sigma(x_i) \) Scale parameter for location \( x_i \)
- \( \phi(x_i) = \log(\sigma(x_i)) \) The re-parameterised scale parameter
- \( \xi(x_i) \) Shape parameter for location \( x_i \)
- \( \mu_{\phi}(x), \mu_{\xi}(x) \) Means for the Gaussian processes
- \( \Sigma, \Psi \) Covariance matrices for the Gaussian processes
- \( \tau^2_{\phi}, \tau^2_{\xi} \) Nugget parameters for the Gaussian processes
- \( \alpha_{\phi}, \alpha_{\xi} \) vectors of coefficients (including intercept) for Gaussian process means
- \( X_{\phi}, X_{\xi} \) Matrices of covariates (including column for intercept term) on the \( x \) grid
- \( \beta_{\phi}, \beta_{\xi} \) Length scale scalars or matrices
- \( \varsigma^2_{\phi}, \varsigma^2_{\xi} \) Variance (partial sill) parameters for the Gaussian processes
Appendix A

Model outline

In hierarchical notation:

\[ z_k(x_i) \sim GPD(\sigma(x_i), \xi(x_i)) \]

\[ \log(\sigma(x)) = \phi(x) \sim MVN(\mu_\phi(x), \Sigma(x, x)) \]

\[ \xi(x) \sim MVN(\mu_\xi(x), \Psi(x, x)) \]

\[ \mu_\phi(x) = X_\phi \alpha_\phi \]

\[ \mu_\xi(x) = X_\xi \alpha_\xi \]

\[ \Sigma(x_k, x_l) = \varsigma_\phi^2 \exp \left( - \left( \frac{d}{\beta_\phi} \right)^2 \right) + \tau_\phi^2 \delta_{kl} \]

\[ \Psi(x_k, x_l) = \varsigma_\xi^2 \exp \left( - \left( \frac{d}{\beta_\xi} \right)^2 \right) + \tau_\xi^2 \delta_{kl} \]

\( \delta_{kl} \) is the Kronecker delta function. In the above formulae for \( \Sigma \) and \( \Psi \), \( \beta_\phi \) and \( \beta_\xi \) are treated as scalars. When considered as matrices, where \( d \) is the distance between two gridpoints, \( (d/\beta)^2 \) is replaced by \( d^T \beta^{-1} d \).

Hyperparameter prior distributions (justifications for these values can be found in Section 3.2.2.3):

\[ \log(\varsigma_\phi^2), \log(\varsigma_\xi^2) \sim N(0, 10) \]

\[ \log(\tau_\phi^2), \log(\tau_\xi^2) \sim N(0, 10) \]

\[ \beta_\phi, \beta_\xi \sim DU(0.001, 0.005, 0.01, 0.05, \ldots, 100, 500, 1000) \]

\[ \alpha_{\phi,j} \sim N(0, 50) \]

\[ \alpha_{\xi,j} \sim N(0, 50) \]
Appendix A

Posterior distribution

The full posterior distribution is:

\[
p(\alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \varsigma_\phi^2, \varsigma_\xi^2, \tau_\phi^2, \tau_\xi^2, \phi, \xi | x, z, X_\phi, X_\xi) \propto \prod_{i=1}^n \prod_{k=1}^{n_i} p(z_k(x_i) | \sigma(x_i)) = \exp(\phi(x_i)), \xi(x_i)) \times p(\phi(x) | \mu_\phi, \Sigma) p(\xi(x) | \mu_\xi, \Psi) \times p(\alpha_\phi) p(\alpha_\xi) p(\beta_\phi) p(\beta_\xi) \times p(\tau_\phi^2) p(\tau_\xi^2) p(\varsigma_\phi^2) p(\varsigma_\xi^2)
\]

Conditional posterior distributions: Layer 2

Updating the second layer of the DAG (parameters \( \phi = \log(\sigma) \) and \( \xi \)): The conditional posterior distribution of \( \phi \) is given by:

\[
\pi(\phi | z, x, X_\phi, X_\xi, \alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \varsigma_\phi^2, \varsigma_\xi^2, \tau_\phi^2, \tau_\xi^2, \xi) \propto p(z | \phi, x, X_\phi, X_\xi, \alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \varsigma_\phi^2, \varsigma_\xi^2, \tau_\phi^2, \tau_\xi^2, \xi) \times p(\phi | x, X_\phi, X_\xi, \alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \varsigma_\phi^2, \varsigma_\xi^2, \tau_\phi^2, \tau_\xi^2, \xi) \propto p(z | \sigma, \xi) p(\phi | \mu_\phi, \Sigma)
\]

That is, the conditional posterior of \( \phi \) is proportional to the product of the likelihood of the data \( z \) given \( \phi \) and all other parameters, and the prior probability density of \( \phi \) given all of the other parameters. In the final line, most parameters have dropped out from the right-hand side, as the densities are independent of these, given the remaining terms (see the DAG in Figure 3.11 for this). Remember that \( \phi = \log(\sigma) \) throughout this, where \( \sigma \) is the scale parameter of the GPD, and note that \( \mu_\phi \) and \( \Sigma \) will have been calculated using the formulae in the model outline (above).

In briefer notation (to be used from now on):
\[ \pi(\phi \ldots) \propto p(z|\sigma, \xi) p(\phi|\mu_\phi, \Sigma) \]
\[
= \left[ \prod_{i=1}^{n} \prod_{k=1}^{n_i} p(z_k(x_i)|\sigma(x_i), \xi(x_i)) \right] p(\phi|\mu_\phi, \Sigma) \]
\[
= \left[ \prod_{i=1}^{n} \prod_{k=1}^{n_i} \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right)^{-\left(1/\xi(x_i) + 1\right)} \right] \times \]
\[
\frac{1}{\sqrt{\text{det}(2\pi\Sigma)}} e^{-\frac{1}{2} (\phi - \mu_\phi)' \Sigma^{-1} (\phi - \mu_\phi)} \]

Similarly:

\[ \pi(\xi \ldots) \propto p(z|\phi, \xi) p(\xi|\mu_\xi, \Psi) \]
\[
= \left[ \prod_{i=1}^{n} \prod_{k=1}^{n_i} p(z_k(x_i)|\sigma(x_i), \xi(x_i)) \right] p(\xi|\mu_\xi, \Psi) \]
\[
= \left[ \prod_{i=1}^{n} \prod_{k=1}^{n_i} \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right)^{-\left(1/\xi(x_i) + 1\right)} \right] \times \]
\[
\frac{1}{\sqrt{\text{det}(2\pi\Psi)}} e^{-\frac{1}{2} (\xi - \mu_\xi)' \Psi^{-1} (\xi - \mu_\xi)} \]

**Metropolis-Hastings Markov chain Monte Carlo sampling**

In order to sample from these conditional posterior distributions, we use the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. New samples are accepted or rejected at random according to the algorithm outlined below.

A new value of \( \phi_i \) is suggested: \( \phi'_i \), where \( i \) is a location on the grid, resulting in a new vector \( \phi' \).

Suggested updates are drawn from a Normal distribution centred on the old value and with a variance of a manually set tuning parameter used to control the size of the proposed steps.
Appendix A

We calculate:

$$\rho(\phi_i, \phi'_i) = \min \left( 1, \frac{\pi(\phi'|\ldots)q_t(\phi'_i \rightarrow \phi_i)}{\pi(\phi|\ldots)q_t(\phi_i \rightarrow \phi'_i)} \right)$$

where $\pi(\phi|\ldots)$ is as defined above, and $q_t(a \rightarrow b)$ is the transition probability of proposing value $b$ given value $a$. Since updates are proposed using a Normal distribution, these transition probabilities above and below the line will always cancel, so the above simplifies to:

$$\rho(\phi_i, \phi'_i) = \min \left( 1, \frac{\pi(\phi'|\ldots)}{\pi(\phi|\ldots)} \right)$$

Following this calculation, we always accept proposed value $\phi'_i$ when $\rho(\phi_i, \phi'_i)$ equals 1 and we reject accordingly when the ratio is smaller than 1 by simulating a random variable $u \sim U[0, 1]$ and accepting proposed value $\phi'_i$ when $u \leq \rho(\phi_i, \phi'_i)$.

Evaluating $\rho(\phi_i, \phi'_i)$ typically involves products and quotients of many terms which may be close to 0. In order to work with something far more computationally stable, we use the property that $x = \exp(\log(x))$. Note also that since the vectors $\phi$ and $\phi'$ differ in only one entry, the GPD calculations above and below will cancel for all other gridpoints apart from the one being updated, $i$.

Following these observations we need to evaluate:
\[
\exp \left( \log \left( \frac{\pi(\phi|\ldots)}{\pi(\phi|\ldots)} \right) \right) \\
= \exp \left[ \log(\pi(\phi|\ldots)) - \log(\pi(\phi|\ldots)) \right] \\
= \exp \left[ \log \left( \prod_{i=1}^{n} \prod_{k=1}^{n_i} p(z_k(x_i)|\sigma'(x_i), \xi(x_i)) p(\phi'|\mu_\phi, \Sigma) \right) - \\
\log \left( \prod_{i=1}^{n} \prod_{k=1}^{n_i} p(z_k(x_i)|\sigma(x_i), \xi(x_i)) p(\phi|\mu_\phi, \Sigma) \right) \right] \\
= \exp \left[ \log \left( \prod_{k=1}^{n_i} p(z_k(x_i)|\sigma'(x_i), \xi(x_i)) p(\phi'|\mu_\phi, \Sigma) \right) - \\
\log \left( \prod_{k=1}^{n_i} p(z_k(x_i)|\sigma(x_i), \xi(x_i)) p(\phi|\mu_\phi, \Sigma) \right) \right] \\
\]

Filling in the distribution details, this becomes:
\begin{align*}
\exp \left[ \log \left( \prod_{k=1}^{n_i} \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) \times 
\frac{1}{\sqrt{\det(2\pi \Sigma)}} e^{-\frac{1}{2} (\phi' - \mu_\phi)' \Sigma^{-1} (\phi' - \mu_\phi)} \right] 
- \log \left( \prod_{k=1}^{n_i} \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) \times 
\frac{1}{\sqrt{\det(2\pi \Sigma)}} e^{-\frac{1}{2} (\phi - \mu_\phi)' \Sigma^{-1} (\phi - \mu_\phi)} \right] 
= \exp \left[ \sum_{k=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) \right] + 
\log \left( \frac{1}{\sqrt{\det(2\pi \Sigma)}} \right) - \frac{1}{2} (\phi' - \mu_\phi)' \Sigma^{-1} (\phi' - \mu_\phi) 
- \sum_{k=1}^{n_i} \log \left( \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) - 
\log \left( \frac{1}{\sqrt{\det(2\pi \Sigma)}} \right) + \frac{1}{2} (\phi - \mu_\phi)' \Sigma^{-1} (\phi - \mu_\phi) 
= \exp \left[ \sum_{k=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) \right] - 
\frac{1}{2} (\phi' - \mu_\phi)' \Sigma^{-1} (\phi' - \mu_\phi) - 
\sum_{k=1}^{n_i} \log \left( \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) + 
\frac{1}{2} (\phi - \mu_\phi)' \Sigma^{-1} (\phi - \mu_\phi) 
\end{align*}

The common term \( \log(1/\sqrt{\det(2\pi \Sigma)}) \) in both numerator and denominator above dropped out as it will be equal in both (it has no dependence on \( \phi \)).

The GPD can clearly be simplified further using the properties of logs. Just taking one of the functions on its own for clarity:
\[
\sum_{k=1}^{n_i} \log \left( \left[ \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i) + 1\right)} \right] \right) = \\
\sum_{k=1}^{n_i} \left[ \log \left( \frac{1}{\sigma'(x_i)} \right) + \log \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i) + 1\right)} \right] = \\
\sum_{k=1}^{n_i} \left[ - \log(\sigma'(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right) \right]
\]

Using this form, the full term we need to evaluate in order to update \( \phi \) is:

\[
= \exp \left[ \sum_{k=1}^{n_i} \left[ - \log(\sigma'(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma'(x_i)} \right) \right] - \right. \\
\left. \frac{1}{2}(\phi - \mu_\phi)'\Sigma^{-1}(\phi - \mu_\phi) - \sum_{k=1}^{n_i} \left[ - \log(\sigma(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right) \right] + \right. \\
\left. \frac{1}{2}(\phi - \mu_\phi)'\Sigma^{-1}(\phi - \mu_\phi) \right] (3.14)
\]

Similar reasoning leads to the update for \( \xi \). We need to evaluate:

\[
\rho(\xi_i, \xi'_i) = \min \left( 1, \frac{\pi(\xi' | \ldots) q_t(\xi'_i \rightarrow \xi_i)}{\pi(\xi | \ldots) q_t(\xi_i \rightarrow \xi'_i)} \right)
\]

The full term we need to evaluate is:
\[
= \exp \left[ \sum_{k=1}^{n} \left[ -\log(\sigma(x_i)) - \left( \frac{1}{\xi'(x_i)} + 1 \right) \log \left( 1 + \xi'(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right) \right] - \right.
\]
\[
\frac{1}{2} (\xi' - \mu_{\xi})' \Psi^{-1} (\xi' - \mu_{\xi}) -
\]
\[
\sum_{k=1}^{n} \left[ -\log(\sigma(x_i)) - \left( \frac{1}{\xi'(x_i)} + 1 \right) \log \left( 1 + \xi(x_i) \frac{z_k(x_i)}{\sigma(x_i)} \right) \right] +
\]
\[
\frac{1}{2} (\xi - \mu_{\xi})' \Psi^{-1} (\xi - \mu_{\xi}) \right] \right]
\]

(3.15)

**Conditional posterior distributions: Layer 3**

\( \alpha \)

Updating the third layer of the DAG (that is, all hyperparameters) starting with \( \alpha_\phi \):

\[
\pi(\alpha_\phi | z, \ldots) \propto p(z | \alpha_\phi, \ldots) p(\alpha_\phi | \ldots) \\
\propto p(z | \sigma, \xi) p(\phi | \mu_\phi, \Sigma) p(\alpha_\phi) \\
\propto p(\phi | \mu_\phi, \Sigma) p(\alpha_\phi)
\]

Essentially the GPD component is independent of \( \alpha_\phi \) once the other parameters are known, and so can be absorbed into the constant of proportionality. What we’re left with is the MVN piece (since \( \alpha_\phi \) features in the calculation of \( \mu_\phi \)) and the prior on \( \alpha_\phi \).

Then we have:
\[ \pi(\alpha_{\phi} | \ldots) \propto p(\phi | \mu_{\phi}, \Sigma)p(\alpha_{\phi}) \]
\[ \propto \frac{1}{\sqrt{\det(2\pi \Sigma)}} e^{-\frac{1}{2}(\phi - \mu_{\phi})'\Sigma^{-1}(\phi - \mu_{\phi})} \times \]
\[ \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(\alpha_{\phi} - m)^2}{2s^2}} \]

where \( m \) is the prior mean and \( s^2 \) is the prior variance. In a slight abuse of notation in the final line above, \( \alpha_{\phi} \) is a scalar, the single element which is being updated from the vector \( \alpha_{\phi} \).

Similarly:

\[ \pi(\alpha_{\xi} | \ldots) \propto p(\xi | \mu_{\xi}, \Psi)p(\alpha_{\xi}) \]
\[ \propto \frac{1}{\sqrt{\det(2\pi \Psi)}} e^{-\frac{1}{2}(\xi - \mu_{\xi})'\Psi^{-1}(\xi - \mu_{\xi})} \times \]
\[ \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(\alpha_{\xi} - m)^2}{2s^2}} \]

where \( m \) is the prior mean and \( s^2 \) is the prior variance, and a similar slight abuse of notation applies in the final line for \( \alpha_{\xi} \).

As with the parameters \( \phi \) and \( \xi \), updates for \( \alpha_{\phi} \) are suggested element-wise:
\( \alpha_{\phi,\kappa} \rightarrow \alpha'_{\phi,\kappa} \), where the index \( \kappa \) runs over the vector of coefficients. The ratio \( \rho(\alpha_{\phi,\kappa}, \alpha'_{\phi,\kappa}) \) is then calculated. Similar manipulations to those used in the previous section lead to the following calculation needed to update \( \alpha_{\phi} \):
Appendix A

\[ \exp \left[ -\frac{1}{2} \log(\det(2\pi \Sigma)) - \frac{1}{2} (\phi - \mu_\phi)'\Sigma^{-1}(\phi - \mu_\phi) ight. \\
- \frac{1}{2} \log(2\pi s^2) - \frac{(\alpha_{\phi,\kappa} - m)^2}{2s^2} \\
+ \frac{1}{2} \log(\det(2\pi \Sigma)) + \frac{1}{2} (\phi - \mu_\phi)'\Sigma^{-1}(\phi - \mu_\phi) \\
+ \frac{1}{2} \log(2\pi s^2) + \frac{(\alpha_{\phi,\kappa} - m)^2}{2s^2} \right] = \\
\] (3.16)

And for \( \alpha_\xi \):

\[ \exp \left[ -\frac{1}{2} (\xi - \mu'_\xi)'\Psi^{-1}(\xi - \mu'_\xi) - \frac{(\alpha'_{\xi,\kappa} - m)^2}{2s^2} \\
+ \frac{1}{2} (\xi - \mu_\xi)'\Psi^{-1}(\xi - \mu_\xi) + \frac{(\alpha_{\xi,\kappa} - m)^2}{2s^2} \right] \] (3.17)

Updates for the other hyperparameters are similar. The matrices \( \Sigma \) and \( \Psi \) will change with any change in \( \beta, \varsigma^2 \) or \( \tau^2 \). The remaining hyperparameters either have a discrete update (in which case the prior probability will be \( 1/\#\text{discrete.values} \) and will cancel above and below the line, or have a univariate Normal prior on them (or their log).

\( \beta \)

For \( \beta_\phi \) we have:

\[ \pi(\beta_\phi | \ldots) \propto p(\phi | \mu_\phi, \Sigma)p(\beta_\phi) \]

Since \( \beta_\phi \) has a discrete prior, to update \( \beta_\phi \) we need to evaluate:
\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi \Sigma')) - \frac{1}{2}(\phi - \mu_\phi)' \Sigma'^{-1} (\phi - \mu_\phi) \\
+ \frac{1}{2} \log(\det(2\pi \Sigma)) + \frac{1}{2}(\phi - \mu_\phi)' \Sigma^{-1} (\phi - \mu_\phi) \right] \tag{3.18}
\]

where \( \Sigma' \) has been formed using \( \beta'_\phi \), the new proposal value or matrix.

Then for \( \beta_\xi \) we have:

\[
\pi(\beta_\xi | \ldots) \propto p(\xi | \mu_\xi, \Psi)p(\beta_\xi)
\]

To update \( \beta_\xi \) we need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi \Psi')) - \frac{1}{2}(\xi - \mu_\xi)' \Psi'^{-1} (\xi - \mu_\xi) \\
+ \frac{1}{2} \log(\det(2\pi \Psi)) + \frac{1}{2}(\xi - \mu_\xi)' \Psi^{-1} (\xi - \mu_\xi) \right] \tag{3.19}
\]

where \( \Psi' \) has been formed using \( \beta'_\xi \), the new proposal value or matrix.

\( \varsigma^2 \)

\( \varsigma^2_\phi \) has a conditional posterior distribution of:

\[
\pi(\varsigma^2_\phi | \ldots) \propto p(\phi | \mu_\phi, \Sigma)p(\varsigma^2_\phi)
\]

The prior distribution of the log of \( \varsigma^2_\phi \) is a univariate Normal. So to update \( \varsigma^2_\phi \) we need to evaluate:
\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi\Sigma')) - \frac{1}{2}(\phi - \mu_{\phi})'\Sigma'^{-1}(\phi - \mu_{\phi}) - \frac{(\log(\varsigma_{\phi}'^2) - m)^2}{2s^2} + \frac{1}{2} \log(\det(2\pi\Sigma)) + \frac{1}{2}(\phi - \mu_{\phi})'\Sigma^{-1}(\phi - \mu_{\phi}) + \frac{1}{2} \log(\det(2\pi\Sigma)) + \frac{1}{2}(\phi - \mu_{\phi})'\Sigma'^{-1}(\phi - \mu_{\phi}) - \frac{(\log(\varsigma_{\phi}'^2) - m)^2}{2s^2} \right]
\]

(3.20)

where \(\Sigma'\) has been formed using \(\varsigma_{\phi}'\), the new proposal value, and \(m\) and \(s\) are the prior mean and standard deviation respectively.

\(\varsigma_{\xi}^2\) has a conditional posterior distribution of:

\[
\pi(\varsigma_{\xi}^2|\ldots) \propto p(\xi|\mu_{\xi}, \Psi)p(\varsigma_{\xi}^2)
\]

To update \(\varsigma_{\xi}^2\) we need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi\Psi')) - \frac{1}{2}(\xi - \mu_{\xi})'\Psi'^{-1}(\xi - \mu_{\xi}) - \frac{(\log(\varsigma_{\xi}'^2) - m)^2}{2s^2} + \frac{1}{2} \log(\det(2\pi\Psi)) + \frac{1}{2}(\xi - \mu_{\xi})'\Psi^{-1}(\xi - \mu_{\xi}) + \frac{1}{2} \log(\det(2\pi\Psi)) + \frac{1}{2}(\xi - \mu_{\xi})'\Psi'^{-1}(\xi - \mu_{\xi}) - \frac{(\log(\varsigma_{\xi}'^2) - m)^2}{2s^2} \right]
\]

(3.21)

where \(\Psi'\) has been formed using \(\varsigma_{\phi}'\), the new proposal value, and \(m\) and \(s\) are the prior mean and standard deviation respectively.
Appendix A

\( \tau^2 \)

\( \tau^2_\phi \) has a conditional posterior distribution of:

\[
\pi(\tau^2_\phi | \ldots) \propto p(\phi | \mu_\phi, \Sigma)p(\tau^2_\phi)
\]

The prior distribution of the log of \( \tau^2_\phi \) is a univariate Normal. So to update \( \tau^2_\phi \) we need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\text{det}(2\pi \Sigma')) - \frac{1}{2}(\phi - \mu_\phi)'\Sigma'^{-1}(\phi - \mu_\phi) - \frac{(\log(\tau^2_\phi) - m)^2}{2s^2} + \frac{1}{2} \log(\text{det}(2\pi \Sigma)) + \frac{1}{2}(\phi - \mu_\phi)'\Sigma^{-1}(\phi - \mu_\phi) + \frac{(\log(\tau^2_\phi) - m)^2}{2s^2} \right]
\]

where \( \Sigma' \) has been formed using \( \tau^2_\phi \), the new proposal value, and \( m \) and \( s \) are the prior mean and standard deviation respectively.

\( \tau^2_\xi \) has a conditional posterior distribution of:

\[
\pi(\tau^2_\xi | \ldots) \propto p(\xi | \mu_\xi, \Psi)p(\tau^2_\xi)
\]

To update \( \tau^2_\xi \) we need to evaluate:
\begin{align}
\exp \left[ - \frac{1}{2} \log(\det(2\pi \Psi')) - \right. \\
&\left. \frac{1}{2}(\xi - \mu_\xi)'\Psi'^{-1}(\xi - \mu_\xi) - \frac{(\log(\tau_{\xi}'') - m)^2}{2s^2} + \right. \\
&\left. \frac{1}{2} \log(\det(2\pi \Psi)) + \right. \\
&\left. \frac{1}{2}(\xi - \mu_\xi)'\Psi^{-1}(\xi - \mu_\xi) + \frac{(\log(\tau_{\xi}'') - m)^2}{2s^2} \right]
\end{align}

(3.23)

where $\Psi'$ has been formed using $\tau_{\xi}''$, the new proposal value, and $m$ and $s$ are the prior mean and standard deviation respectively.
Chapter 4

Bayesian spatial extreme value analysis of maximum temperatures in County Dublin, Ireland

In the previous chapter, Bayesian spatial hierarchical models were applied to a dataset of extreme wave heights in order to learn more about the parameters of the latent spatial processes underlying this data. This was done with a focus on reducing the uncertainty surrounding the posterior estimates of those parameters (or more specifically, reducing the uncertainty of quantities calculated using those parameters). Two Bayesian hierarchical models were fitted and their output used to generate estimates of return level surfaces over the spatial domain. Throughout this analysis, the Bayesian spatial approach was observed to show a superior fit to the data and reduced uncertainty of posterior estimates, in comparison with a maximum-likelihood estimation (MLE) point-based analysis. This borrowing of information from across the spatial field of data by the Bayesian hierarchical model is particularly important when it comes to analysing extremes, for which the available data is rare. While the Bayesian hierarchical model in Chapter 3 demonstrated a superior performance to the MLE point-based approach, this comes at an expensive computational cost, particularly when the number of gridpoints in the spatial domain increases. In order to continue to use such models, we need to build on the methods from Chapter 3 and introduce more sophisticated statistical tools to overcome this computational prob-
In this chapter, I define and explain a particular dimension-reduction technique called predictive processes, and I then proceed to apply this technique in an adapted Bayesian spatial hierarchical model to a dense dataset of extremes of daily maximum temperature. The rest of this chapter is structured as follows:

- In Section 4.1, I include a brief overview of this research, the motivation for undertaking it, and our aims for doing so.
- In Section 4.2, I review the body of literature of previous work in areas relevant to this study.
- In Section 4.3, I remind the reader of the nature and the domain of the daily maximum temperatures dataset introduced in Section 1.4.3, and then proceed to describe the calculation of temperature anomalies which we model in this study.
- In Section 4.4, I provide an overview of the statistical techniques which are needed in order to build on and extend the statistical theory from Chapter 3. Following a brief reminder of the areas presented in the previous chapter, the concepts and the theory underlying predictive processes (a dimension-reduction technique) are described in detail.
- In Section 4.5, I describe in detail the Bayesian spatial hierarchical model using predictive processes which we use in order to make it computationally feasible to model our dataset over its dense spatial domain. This includes details of all three layers in the two hierarchies; details on the prior distributions on the hyperparameters; a discussion on threshold selection; and the specifics of the model implementation process.
- Following the implementation of the model outlined in the previous section, in Section 4.6 I outline the main results from this model fitting. Posterior parameter surfaces and return level surfaces will be presented. In addition to this, some site-specific return level curves will be used for model evaluation, along with a comparison with more recent observations at several sites close to synoptic weather stations.
- Finally, in Section 4.7, I discuss the results in further detail, and draw conclusions based on these.
4.1 Overview of study

In this study, we produce return levels of anomalies of daily maximum temperature extremes for an area over Ireland, for the 30-year period of 1981-2010. We apply Bayesian hierarchical modelling (Gelman et al., 2013) combined with a reduced-rank method called predictive processes to solve the computationally difficult problem of modelling data over a very dense spatial field (Banerjee et al., 2008). The role of extreme value theory (EVT) (Coles, 2001) is increasingly important in furthering our understanding of climate change and climate extremes. With an increase in maximum temperature extremes through the current century projected by the Intergovernmental Panel on Climate Change (IPCC), it is important to have a better understanding of recent observations of climate extremes (IPCC, 2013).

In order to model extremes of daily maximum temperature, the chosen statistical models must use EVT because extremes of temperature are rare (by definition), and occur in the tails of the distribution. Since extremes of temperature vary by location on any given day, the statistical models used should account for this spatial dependence. And since these distributions are governed by parameters that depend not only on the data, but on other (e.g., physical or mathematical) principles and constraints, it is natural to use a Bayesian framework. With this in mind, we apply a Bayesian hierarchical spatial extreme value model to a dataset of daily maximum temperatures in Ireland.

Fitting Bayesian hierarchical models using Markov chain Monte Carlo (MCMC) methods involves matrix factorisations of the order of $n^3$, where $n$ is the number of locations (Guhaniyogi et al., 2011). In our case, with a dataset of daily maximum temperature from 1981 to 2010 at more than $\sim 72,000$ locations, this becomes computationally infeasible. In order to overcome this problem, we focus our attention on the capital city of Ireland and its surroundings, and employ reduced-rank spatial models. These are popular models for analysing large spatial datasets (Banerjee et al., 2008). In particular, we work with a flexible class of low-rank models called predictive process models.
4.2 Previous relevant research

With this study, we aim to begin a comprehensive characterisation of temperature extremes in Ireland for the period 1981-2010. Our contribution here expands upon existing research on historical temperature extremes in Ireland, which to our knowledge, consists only of site-specific analysis with no spatial component to the models used (e.g., Met Éireann (2012) and Osman et al. (2015); described in the next section). All code needed to reproduce our analysis is available in a public repository on GitHub at https://github.com/jackos13/extremes. Note that this repository does not include the data; they are available upon suitable request from Met Éireann.

4.2 Previous relevant research

4.2.1 Temperature extremes in Ireland

In a study of Ireland’s climate from 1981 to 2010, Met Éireann (2012) produces a table of monthly values of mean daily maximum temperatures for two synoptic stations, one at Casement Aerodrome and one at Valentia Observatory. At Casement Aerodrome (in Dublin), the annual mean daily maximum temperature is found to be 13.4°C. The monthly mean daily maximum temperature at this station ranges from 8°C in January to 19.8°C in July. A further breakdown by individual days is not provided. Highest single observations of daily maximum temperature are also given, broken down by month. December has the lowest of these individual maximum values (14.8°C) while the overall highest daily maximum temperature for this period was recorded in the month of July (31°C on the 19th of July, 2006). Osman et al. (2015) investigate temperature extremes at six different locations in Ireland. They fit a generalised Pareto distribution (GPD) to the observed extremes at each location (for the period 1961-2000), allowing the scale parameter to vary as a function of large-scale climate variables derived from reanalysis data. They then use future values of these climate variables from global climate model (GCM) output in order to produce return-level plots for all six locations, leading them to conclude that significant changes in extreme temperature events are projected to occur in Ireland over the course of the 21st century. These projected changes include hotter summers and milder winters. As detailed in Chapter 2, O’Sullivan et al. (2016) investigate projected changes in the extremes of Irish temperatures for the mid-21st century using a high-resolution
multi-model analysis. Analysing changes in percentiles they find that, at a seasonal level, there is greater projected warming of temperature extremes than for mean temperatures. Across almost the entire island, the daily maximum temperature for future summers under a high emissions scenario is projected to increase by more than 2°C (in the southeast of the island, this increase exceeds 2.5°C).

4.2.2 Spatial modelling for large data sets

In the context of using Bayesian hierarchical spatial models to analyse the large spatial datasets that occur frequently in environmental sciences, Guhaniyogi et al. (2011) discuss predictive process models. They point out the flexibility of these models, explaining how they can overcome the ‘big n’ problem of large datasets, where matrix factorisations of the order of $n^3$ (where $n$ is the number of locations) can make the direct estimation of hierarchical spatial models infeasible. Finley et al. (2012), using a Bayesian hierarchical framework, apply a predictive process model to mean temperature data from 2000-2005 from weather stations situated across the northeastern U.S.A. They found that the low-dimensional predictive process model was highly effective at borrowing information over space to make accurate and precise predictions for new locations. In addition to this, the choice of a Bayesian framework meant that the authors could determine the full posterior predictive distribution at any new location.

4.2.3 Spatial extreme value analysis

In order to produce maps of precipitation return levels and uncertainty measures over a region in Colorado in the U.S.A., Cooley et al. (2007) (whose methodology informed the model for the waves dataset in Chapter 3) construct two separate Bayesian hierarchical models for extreme precipitation events: one for the intensity (using the GPD), and another for the frequency (using the Binomial distribution) of such events. The assumption underlying both models is that regional extreme precipitation is driven by a latent spatial process, characterised by geographical and climatological covariates. Their approach involves pooling all of the information from different stations in order to produce a 25-year daily precipitation return level map, which directly takes into account the parameter and interpolation uncertainty in the method itself. Shaby and
4.2 Previous relevant research

Reich (2012) apply a Bayesian hierarchical spatial extreme value model to temperature extremes across Europe in order to assess the changes in the risk of widespread extremely high temperatures across agricultural land. For their data, they use annual maximum temperatures on a selected subgrid of 985 locations. They find that the risk of large percentages of cropland exceeding a high temperature threshold has probably increased in the last century, but only slightly so. Lehmann et al. (2016) use a Bayesian hierarchical framework and a block-maxima approach to model extremes of precipitation of different durations from over 1,300 weather stations in two different regions in Australia. The parameters of the distribution are modelled as spatial Gaussian processes. They found that pooling the data across space, and thus borrowing strength from neighbouring stations, leads to more precise parameter estimates and therefore superior posterior inference. This borrowing of strength is particularly important when dealing with extremes. In a similar study, Dyrrdal et al. (2015) also use spatial Gaussian processes in a Bayesian hierarchical model in order to produce spatial maps of extreme hourly precipitation over Norway. They specify a generalised extreme value distribution at each point in space, and allow the parameters to depend on location-specific geographic and meteorological variables, a structure similar to generalised linear modelling. Variable uncertainty is accounted for using Bayesian model averaging. They find that their approach performs well at estimating extreme hourly precipitation return levels, both in terms of magnitude and spatial distribution. Ghosh and Mallick (2011), using a Bayesian hierarchical approach, compare two separate data models (a GEV distribution and a variation of the GPD) to model precipitation extremes over continental U.S.A. from 1900 to 1998. They find that the peaks-over-threshold approach (i.e. using the GPD) shows better fit and improved forecast ability than the block-maxima approach (i.e. the GEV).

To our knowledge, no study to date has combined these two approaches from Sections 4.2.2 and 4.2.3; that is, no study has used EVT and predictive processes to model large spatial datasets of climate extremes.
4.3 Dataset of daily maximum temperatures

The dataset analysed for the remainder of this chapter is the gridded dataset of daily maximum temperatures over Ireland, first introduced and described in Section 1.4.3. The data is on a $1 \times 1$ km$^2$ grid ($\sim$72,000 gridpoints), consisting of daily values of maximum temperature for the 30-year period 1981-2010. This was produced by Met Éireann’s Climatology and Observations division, using observational data from 138 stations and independent variables available at each grid point, such as elevation, latitude, and longitude. This observational data was interpolated onto the regular grid, using an inverse distance weighted algorithm (Met Éireann, 2017). (This dataset is available following a suitable request to the Climate Enquiries Office at Met Éireann: https://www.met.ie/.)

4.3.1 Study domain

Figure 4.1 shows the domain of the dataset. The image on the left illustrates the extent of the full dataset, which covers the state of Ireland. Geographic covariates of latitude, longitude and altitude are available on the same $1 \times 1$ km$^2$ grid as the temperature data. Due to the large number of gridpoints in the dataset, we decided to focus on a smaller area to ensure that running the model was computationally feasible. We selected an area covering County Dublin, which includes the capital city and its surroundings, located in the east of the island. This study domain is shown on the right panel of Figure 4.1. It contains approximately 1,700 gridpoints.
4.3 Dataset of daily maximum temperatures

4.3.2 Data verification

We performed various checks of the gridded data in order to compare it to the underlying synoptic station observations which were used in generating it. Two of these site-specific checks are illustrated in Figures 4.2 and 4.3. The data here are of the daily maximum temperature. Observations are taken from the synoptic station at Casement Aerodrome (location shown in Figure 4.1), while the gridded data is taken from the nearest gridpoint (approximately 540 m away).

Firstly, we created a scatterplot of observed vs. gridded data. This shows (Figure 4.2) a very strong positive correlation between the observed and the gridded data, as would be expected. At all 9 sites checked across the full domain (namely, Bal-

Figure 4.1: The data domain showing the outline of the state of Ireland (in black, with county outlines included) is on the left of the image. The area within the black box is our study region, and is blown-up and shown on the right. The grey circles mark the gridpoints of the spatial domain. A red cross marks the location of Casement Aerodrome, for which we have synoptic station data that is used in the next section for data verification, and later in the chapter for model evaluation purposes.
4.3 Dataset of daily maximum temperatures

Ilyhaise, Belmullet, Carlow Oakpark, Casement Aerodrome, Dublin Airport, Fermoy Moorepark, Malin Head, Roches Point, and Valentia Observatory), there was a correlation of $>0.985$ between the observations at the synoptic station there and the gridded data at the nearest gridpoint. For the data illustrated here, the correlation is 0.997.

![Scatterplot of observed and gridded data at Casement Aerodrome](image)

Figure 4.2: Shown is a scatterplot of observed daily maximum temperatures at Casement Aerodrome vs. the gridded data at the closest gridpoint (approximately 540 m away). A very strong positive correlation between the observed and the gridded data is clearly visible (correlation $= 0.997$).

Secondly, in order to focus on the tail of the data (which is where our interest lies), we then considered only the data above the 99th percentile in both cases: observations and gridded data. Figure 4.3 shows these two smoothed densities overlaid (where the respective thresholds were subtracted from the two datasets in order to have them both begin at 0 to aid comparison). A very close relationship can be seen between the two densities: the observations and the gridded data have a very similar shape (the area of overlap calculated from the two kernel density estimations from the empirical data $= 0.95$). For all 9 sites nationwide checked in this manner, there was an overlap of $>0.89$ between the overlaid densities of exceedances of the 99th percentile for the observations and the gridded data from the nearest gridpoint.
4.3 Dataset of daily maximum temperatures

Casement Aerodrome:
exceedances of the 99th percentile

Degrees Celsius above 99th percentile

Figure 4.3: Shown are the two overlaid densities (of the daily maximum temperature gridded data and observations) of the exceedances of the 99th percentile at Casement Aerodrome (area of overlap = 0.95).

4.3.3 Temperature Anomalies

In order to focus our research on maximum temperatures which would be considered extreme relative to the time of year in which they occurred, we chose to create and then analyse temperature anomalies. Adapting a procedure used by Brown et al. (2008), we created the anomalies by removing the mean annual cycle at each location. To do this, we first calculated a 31-day moving average for each location over the 30 years of data, using only values at that location. We then averaged by calendar date in order to get a climate value for the first of January, a climate value for the second of January etc., at that location. Using these 366 values provided a distinct climate curve for each location - an annual curve at that gridpoint. (As mentioned in Section 1.3.1, the World Meteorological Organization recommend using 30-year periods of reference in order to define climate normals (World Meteorological Organization, 2017).) Each of these values from the climate curve were then subtracted from the 30 years of raw data for the corresponding day, resulting in a dataset of anomalies at each location. It is the extreme values of these anomalies that we want to analyse. (Further details and an illustrative example of the steps involved in this calculation are to be found in Appendix A on page 153, along with an exploration of whether the different seasons have an effect on this approach.)
4.3 Dataset of daily maximum temperatures

We first selected those anomalies which exceeded the 99th percentile at their gridpoint, in order to further explore the values which would be considered extreme under this approach. Figure 4.4(a) illustrates this threshold surface: it shows the value at each point above which only 1% of anomalies at that point lie, and ranges here from 5.2 - 6.2°C. The threshold is generally lower near the coast - this is unsurprising, as the temperatures here are moderated by their proximity to the sea. Examining the standard deviation of the excesses at each point (Figure 4.4(b)) shows a similar trend. Not only are values of the threshold higher as you move away from the coast, but the excesses above this threshold are also more variable, indicating more extreme temperature anomalies inland. Some of the highest values in both figures are observed close to Casement Aerodrome (note the yellow spot at ≈ 53.3°N and -6.4°E).

Figure 4.4: The figure on the left (a) is the the 99th percentile value - that is, the value for each location above which only 1% of observed anomalies lie; the figure on the right (b) illustrates the standard deviation of the excesses of this threshold. Both scales are in degrees Celsius. Some of the highest values in both figures are observed close to Casement Aerodrome (note the yellow spot at ≈ 53.3°N and -6.4°E).
4.4 Statistical background

At this stage, the reader is directed back to Section 3.1 to the comprehensive overview presented there in the areas of: Extreme Value Theory (EVT); Bayesian inference; the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm; and hierarchical modelling. Knowledge of this theory is assumed, and for reasons of brevity, won’t be repeated here. A reminder of the fundamentals underlying spatial statistics and Gaussian processes (GPs) is included below. After this reminder, I can then describe the new technique which we use to model the dense dataset described above in Section 4.3, using notation consistent with that from Chapter 3, where possible.

4.4.1 Spatial methods

Typically when dealing with random variables recorded as point-referenced (or geo-statistical) data, the location index $s$ is assumed to vary continuously over $D$, a fixed subset of $\mathbb{R}^d$ (Banerjee et al., 2014). Let $T(s)$ be a vector of random variables (a random vector) at locations $s$. This could be, for example, measurements of daily maximum temperatures at locations $s$. While it is sensible to conceptually assume such values exist at all possible sites in the spatial domain, in practice the data is a partial realisation of this continuous spatial process. Given this partial realisation, the problem then becomes inference about this spatial process $T(s)$ as well as prediction at new locations. To this end, it is assumed that the covariance between the random variables at two locations depends on the distance between these locations. That is, $\text{Cov}(T(s_i), T(s_j)) = C(s_i, s_j) = C(d_{ij})$ is a function of $d_{ij}$ where $d_{ij}$ is the distance between locations $s_i$ and $s_j$ (for brevity, the dependence on sites $i$ and $j$ is dropped below). The method of calculating this distance must be specified (with Euclidean distance the most common approach). There are many choices of covariance functions (see e.g., Diggle et al. (1998), Gelfand et al. (2010) (pp. 22-26) or Banerjee et al. (2014) (pp. 25-30) for a description of several parametric models, and their relative merits). In this work, we use the Matérn class of covariance functions (Matérn, 1960) with univariate form:
4.4 Statistical background

\[ C(d) = \begin{cases} \frac{\varsigma^2}{\Gamma(\nu)} \phi d \nu K_{\nu}(\phi d) & \text{if } d \neq 0 \\ \tau^2 + \varsigma^2 & \text{if } d = 0. \end{cases} \]

Here, \( \varsigma^2 \) is the partial sill (variance of the spatial effect), \( \tau^2 \) is the nugget (variance of the non-spatial effect), \( \nu \) is a parameter controlling the smoothness of the spatial field, \( \phi \) is a spatial decay parameter controlling how quickly the covariance decreases with distance, \( \Gamma() \) is the gamma function, and \( K_{\nu} \) is the modified Bessel function of the second kind of order \( \nu \).

There are alternatives to using the Matérn class of covariance functions, such as kernel convolution (or moving average) models, and convolutions of covariance models (Gelfand et al., 2004). However, we decided to work with the Matérn class of functions as it is a flexible class, with parameters that have attractive interpretations, and includes as special cases the exponential (by setting \( \nu = 1/2 \)) and the squared exponential (which arises in the limit as \( \nu \to \infty \) with the scale parameter set at \( \phi = 2\sqrt{\nu} \)) covariance functions (Gelfand et al., 2010).

4.4.2 Gaussian Processes

The process \( T(s) \) is said to be Gaussian if, for any \( n \geq 1 \) and any set of sites \( \{s_1, s_2, \ldots, s_n\}, T = (T(s_1), T(s_2), \ldots, T(s_n))^T \) has a multivariate normal distribution (Banerjee et al., 2014). Gaussian Processes (GPs) can be thought of as extending the finite multivariate normal distribution to infinitely many random variables; in other words, a GP is an infinite collection of variables such that every finite subset follows a multivariate normal distribution. This is a very flexible framework for modelling spatial data, as the covariance matrix can be specified using any valid covariance function.

Given realisations of the the process \( T(s) \), and \( p \) spatially-referenced covariates at the same locations \( s \), let \( X(s) \) be the \( n \times (p + 1) \) matrix associated with the spatial regression model:

\[ T(s) = X^T(s)\alpha + w(s) + \epsilon(s) \quad (4.1) \]
where $X^T(s)\alpha$ is the mean response, $w(s)$ is a zero-centred GP with covariance function $C(s,s')$ and $\epsilon(s) \overset{iid}{\sim} N(0, \tau^2)$ is an independent measurement error (see, e.g., Finley et al. (2009) and Cressie and Wikle (2011)).

There are alternative computational approaches to fitting GPs using MCMC such as the integrated nested Laplace approximation (INLA) approach proposed by Rue et al. (2009), and INLA combined with a stochastic partial differential equation approach (INLA-SPDE) proposed by Lindgren et al. (2011). However, we have chosen to use GPs due to the ease with which they fit into a Bayesian hierarchical framework. They are flexible empirical models, which are appropriate for an irregularly fluctuating and real-valued spatial surface (Diggle et al., 2007), as we have here.

Bayesian inference using GPs typically involves the need to invert or factor the covariance matrix arising from the covariance function $C(s,s')$ in equation (4.1) which forms part of the zero-centred $w(s)$. This becomes computationally impractical as the dimension $n$ becomes large (that is, a large number of gridpoints), particularly when using an algorithm such as Markov chain Monte Carlo (MCMC) which involves inverting or factoring this matrix hundreds of thousands of times in one run of a model fitting. For this reason, we decided to examine reduced-rank representations, and focus our attention on Gaussian predictive process models.

### 4.4.3 Predictive Process Models

A comprehensive overview of hierarchical Gaussian predictive process models is given in Banerjee et al. (2014). I present a brief description below. Following the notation from Section 4.4.2, we can avoid dealing with the dense covariance matrix induced by the zero-centred GP $w(s)$ by projecting it onto a subspace spanned by its realisation over the $n^*$-dimensional $S^*$ where $n^* \ll n$. An optimal projection of the process $w$ at location $s$, based upon its realisation over $S^*$ is given by the kriging equation:

$$\tilde{w}(s) = E(w(s)|w^*)$$

(4.2)

where $w^* = (w(s_1^*), w(s_2^*), \ldots, w(s_n^*))^T$. $\tilde{w}(s)$ is referred to as the predictive process derived from the parent process $w(s)$. 

127
Further expanding the above:

\[
\hat{w}(s) = C(s, S^*)^T \cdot C(S^*, S^*)^{-1} \cdot w(S^*)
\]

where \( C(s, S^*) \) is the \( n^* \times 1 \) vector with \( C(s, s^*_j) \) as the \( j \)th element; \( C(S^*, S^*)^{-1} \) is the \( n^* \times n^* \) covariance matrix on the subspace \( S^* \); and \( w(S^*) \) is the \( n^* \times 1 \) values of the process \( w \) on the subspace \( S^* \). The important point is that inference on the process \( w \) will now involve inverting or factoring an \( n^* \)-dimensional matrix rather than an \( n \)-dimensional one - in other words, moving to the predictive process model immediately reduces computational complexity from \( O(n^3) \) to \( O(n^{*3}) \) and hence reduces run-time (Finley et al., 2012).

A key consideration when using predictive processes (and other low-rank models) is the choice of the number and the location of the knots. The number of knots \( n^* \) is often a pragmatic choice, depending on the time and computational resources available for the study. For a given \( n^* \) then, the first decision concerning the location of the knots is whether to use a subset of the sampled locations, or a disjoint set of locations across the spatial domain (Finley et al., 2012). If a subset is to be used, the next decision is whether this should be drawn at random or selected based on spatial coverage considerations. If a subset is not to be used, this is reduced to a design problem, with the additional knowledge of already having observations at \( n \) locations. A thorough overview of spatial design optimisation is offered by Xia et al. (2006). One possible approach to selecting the location of these knots is a space-filling design, which typically involves a regular grid over the extent of the domain (Royle and Nychka, 1998). More complex approaches can also be employed, which place knots at locations specifically chosen to minimise a chosen predictive variance criterion (see, e.g., Finley et al. (2009)). This approach involves assuming knowledge of the covariance function across the domain. However, the covariance function is typically not known (precisely), and it is important to have designs that are not overly sensitive to the exact form of the covariance (Nychka et al., 1997). The location of the knots themselves may vary using even more complex algorithms - e.g., adaptive Gaussian predictive process models which, along with modelling the data.
in a hierarchical model, include stochastic modelling of the knot locations (see, e.g., Guhaniyogi et al. (2011)). In practice though, there can be relatively little difference in inference based on knot locations chosen using a grid, a space-filling design, or some other criterion, as long as the observed locations are evenly distributed across the domain (Finley et al., 2012). Instead, it is the number of knot locations which has a greater impact on parameter estimates and reducing the uncertainties in quantities derived from them.

4.5 Model details for the temperature dataset

The aim of this study is to produce $N$-year return levels of anomalies of extreme temperature ($T_{max}$). The dataset we want to model has been described in Section 4.3. We have chosen to analyse temperature anomalies in order to consider all daily maximum temperatures which would be considered extreme relative to the location and the time of year in which they occur. To prepare the data to be modelled, we did the following:

1. Temperature anomalies were calculated at each location (as described in Section 4.3.3, and in more detail in Appendix A on page 153);
2. then site-specific thresholds ($u_i$) were calculated (a discussion on final threshold selection is included in Section 4.5.4), and values exceeding this threshold were selected;
3. only declustered excesses above this threshold were kept for modelling (that is, if two or more consecutive days were considered extreme after step 2, then only the maximum of these values was retained).

Declustering is a process described by Coles (2001) (p. 99), and has already been discussed in some detail in Section 3.2.4. It is a necessary procedure as the theoretical basis for the use of the GPD assumes that the exceedances are independent.

We then require location-specific estimates of the parameters from the GPD ($\sigma$ and $\xi$), as well as the probability of exceeding the threshold ($\zeta_u$) in order to calculate $N$-year return levels using the equation:
4.5 Model details for the temperature dataset

\[ z_N = u + \frac{\sigma}{\xi} \left[ (N n_y \zeta_u)^\xi - 1 \right] \]  

(4.3)

where \( n_y \) is the number of observations per year.

Following the approach of Cooley et al. (2007) (described in some detail in Sections 3.2.1 and 4.2.3), we employ two Bayesian spatial hierarchical models with three layers. The first hierarchy models the parameters of the GPD; the second models the probability of exceeding the threshold. As described in Section 3.1.10, the first layer consists of linking the data to some parameters through a probability distribution. The second layer describes the latent spatial process underlying the extremes in the region, while the third layer consists of the prior distributions on the hyperparameters (that is, the parameters controlling the second layer). Next, I present a detailed description of the two hierarchical models, which is then followed by a graphical representation of these models; namely a directed acyclic graph (DAG), which is shown in Figure 4.5.

Let the declustered extreme anomalies of \( T_{max} \) be denoted by \( y_j(x_i) \) where the indices \( i \) and \( j \) are such that \( y_j(x_i) \) refers to the \( j \)th exceedance (\( j = 1 \ldots n_i \)) at gridpoint \( x_i \) (\( i = 1 \ldots n \)). Let \( M_i \) refer to the number of (declustered) exceedances. The layers in the hierarchies are described below. (Further details involving the pseudocode and equations for this model are to be found in Appendix B on page 160.)

### 4.5.1 Layer 1: Extremal Data and Probability of Exceedance

We model the extremal data \( y_j(x_i) \) using the GPD. This is the first layer in the first hierarchy. To ensure a positive scale parameter throughout the computations, we reparameterise \( \phi = \log(\sigma) \). We have two spatially-varying parameters for the GPD distribution. The first layer in the first hierarchy is then given by:

\[ y_j(x_i) \sim GPD(u(x_i), \sigma(x_i), \xi(x_i)) \]

The first layer in the second hierarchy involves the parameter \( \zeta_{u_i} \) (the probability of exceeding the threshold \( u_i \) - or more precisely, following declustering, the probability of being a cluster maximum at gridpoint \( i \)). This needs to be modelled in order
4.5 Model details for the temperature dataset

to calculate return level surfaces using equation (4.3). Following the methodology of Cooley et al. (2007), we model this as a binomial random variable. Here, the probability of being a cluster maximum is modelled using the empirical probability as our data. It is assumed that the observed number of cluster maxima $M_i$ at gridpoint $i$ is a binomial random variable with $m_i$ trials (the total number of observations in the period of study), each with a probability $\zeta(x_i)$ of being a cluster maximum:

$$M_i \sim \text{Bin}(m_i, \zeta(x_i))$$

From this point onwards, I omit all dependence on location $x_i$ for ease of notation, unless emphasis at a particular point is necessary.

4.5.2 Layer 2: Process

We assume that the GPD parameters vary smoothly over space and thus model the two variables ($\phi = \log(\sigma)$ and $\xi$) as GPs. Following our decision to model these parent processes using the reduced-rank representation of predictive processes, at this layer we thus directly model $\phi^* = \log(\sigma^*)$ and $\xi^*$ (where $\phi^*$ and $\xi^*$ are defined on a smaller grid $z$, and are thus lower-dimensional, than $\phi$ and $\xi$, which are defined on the full grid $x$). The second layer for $\phi^*$ in this hierarchy then is:

$$\phi^* \sim \text{MVN}(\mu_\phi, \Sigma_\phi)$$

Here, $\mu_\phi = Z\alpha_\phi$ where $Z$ is a matrix of covariates (at locations $z$) for the linear regression component of the model, and $\alpha_\phi$ is a vector of coefficients. $\Sigma_\phi$ is the covariance matrix of the spatial process, and is modelled using the Matérn covariance function described earlier. This has parameters $\beta_\phi$ consisting of a matrix of range parameters (controlling how quickly the correlation drops off in different directions), $\varsigma_\phi$ is the partial sill, $\tau_\phi$ is the nugget, and $\nu_\phi$ is the smoothness parameter. The $n$-dimensional $\tilde{\phi}$ is then calculated using the kriging equation (4.2) detailed in Section 4.4.2. The layer for $\xi^*$ is similar.

In a similar manner (and again following the methodology of Cooley et al. (2007)), we assume $\zeta$ to vary smoothly over space. We let $\zeta^*$ be a reduced-rank representation
of $\zeta$, and then apply the transformation $\zeta' \equiv \logit(\zeta^*)$, and use predictive processes to model this reduced-rank transformed representation. As before, the kriging equation is used to calculate $\tilde{\zeta}$ given $\zeta'$.

### 4.5.3 Layer 3: Hyperparameters

The third and final layers of the hierarchies consist of the 15 prior distributions on the parameters in the second layers - that is, the distributions of the $\alpha$, $\beta$, $\varsigma$, $\tau$ and $\nu$ hyperparameters, for each of the three parameters $\phi^*$, $\xi^*$ and $\zeta'$.

Using Bayesian inference allows additional information about a process to be incorporated in the form of prior information. This could be of great benefit in our case, due to the scarcity of extreme data. Coles and Tawn (1996) argue that careful elicitation of prior expert information can supplement the data, which then leads to improved posterior estimates of extremal behaviour. However, with such scarce data, an expert may not be able to independently formulate prior beliefs about this process (Coles and Powell, 1996). With this advice in mind, we aim to use semi-informative prior distributions which are based on physically plausible values, but with enough flexibility so that the data is not restricted from informing the posterior distributions.

We constructed the regression matrix $Z$ to model an intercept parameter in addition to the covariates of latitude, longitude, and elevation. As elevation had a strong positive skew (skewness = 2.36), we transformed it using a log-transform. The three covariates were then scaled to be centred on 0 and with a standard deviation of 1.

For the prior distributions for the corresponding vectors of regression coefficients $\alpha_{\phi}$, $\alpha_{\xi}$, and $\alpha_{\zeta}$, we chose Normal distributions centred on 0, 0 and -6 respectively for the intercept parameters. Remembering that $\phi^*$ is the log of the scale parameter, this corresponds to a prior distribution for the scale parameter $\sigma$ with high density values in the neighbourhood of 1. The prior for the intercept for $\xi^*$ is centred on 0, which assumes the data to have infinite support. The prior for the intercept for $\zeta'$ is centred on -6, which is approximately the logit of the probability of an observation selected at random exceeding the chosen threshold. Standard deviations of 2 are used for all
4.5 Model details for the temperature dataset

three intercepts. This is arguably a little too wide in some cases (e.g., Coles and Powell (1996) point out that a shape of -1 rarely occurs when modelling the maxima of environmental data), but as the shape parameter is particularly difficult to model, we preferred to err on the side of caution, and allow the data play the dominant role in informing the posterior distribution for this parameter.

The following details regarding the non-intercept $\alpha$ coefficients, and the $\beta$, $\nu$, $\varsigma^2$ and $\tau^2$ hyperparameters are identical for the $\phi^\ast$, $\xi^\ast$ and $\zeta'\ast$ parameter surfaces. For the non-intercept $\alpha$ coefficients for the scaled covariates of latitude, longitude and altitude, we chose $N(0, 1)$ distributions (as proposed by Dyrrdal et al. (2015)). Although we expect to see a relationship between the covariates and the parameter surfaces (e.g., more extreme excesses are expected further from the sea, and so the scale surface is likely to increase as the distance from the sea increases), we wanted to assume no relationship a priori and allow the data to inform the relationship between the parameter surfaces and the covariates. A standard deviation of 1 is, again, arguably a little too wide (with the scaled covariates, a value of 1 would imply a parameter surface which increases by at least 3 across the fields of latitude, longitude or altitude - a slope of this magnitude is unrealistic for all three parameters under consideration in this study). We have aimed to use physical arguments in informing the prior distributions, but again have chosen to err on the side of caution in order to allow the data play the dominant role in informing the posterior distributions.

The $\beta$ and $\nu$ priors needed to be considered together. Looking again at the univariate Matérn function for the covariance between two gridpoints (where the spatial decay parameter $\phi$ has been re-parameterised for convenience here as the range parameter $\beta = 1/\phi$), where $d$ is the distance between the gridpoints $s$ and $s'$:

\[
C(s, s') = C(d) = \begin{cases} 
\frac{\varsigma^2}{2^{\nu-1}\Gamma(\nu)} \left( \frac{d}{\beta} \right)^\nu K_\nu \left( \frac{d}{\beta} \right) & \text{if } d \neq 0 \\
\varsigma^2 + \tau^2 & \text{if } d = 0,
\end{cases}
\]

it can be seen that the shape of the parameter surface depends on $\beta$ (which controls how quickly the covariance decreases with distance) and $\nu$ (which controls the smoothness of the spatial field).
In order to allow $\beta$ to differ depending on the direction of its two-dimensional coordinates (longitude and latitude), it is necessary to extend the univariate case above to incorporate a $2 \times 2$ positive definite matrix $\beta$. With this approach, the $(1, 1)$ entry in the matrix represents the spatial range in the direction of longitude, the $(2, 2)$ entry represents the spatial range in the direction of latitude, and the $(1, 2)$ entry represents the covariance between the two. The covariance function between two points $s$ and $s'$ now becomes:

$$C(s, s') = C(d) = \begin{cases} \frac{\varsigma^2}{2^{\nu-1}} \left( \sqrt{d^T \beta^{-1} d} \right)^\nu K_\nu \left( \sqrt{d^T \beta^{-1} d} \right) & \text{if } d \neq 0 \\ \varsigma^2 + \tau^2 & \text{if } d = 0, \end{cases}$$

where $d$ is now a 2-dimensional vector of the distance between the two points.

In order to construct a set of prior matrices $\beta$, all possible combinations of matrices were formed using values from the set $\{0, 0.05, 1, 10\}$, from which only the positive definite ones were retained. This led to a set of 20 matrices. The smoothness parameter $\nu$ is assigned a prior support of $\{0.5, 2.5\}$. It is common to use such values for $\nu$, as the data can rarely inform about smoothness of higher orders (Finley et al., 2009). This means there are $20 \times 2 = 40$ prior $(\beta, \nu)$ combinations to model the spatial range across the parameter field and the smoothness of the resulting field.

Though it is possible for the spatial dependence of individual extreme observations to have a short range, the surfaces being modelled here refer to a climatological quantity rather than a weather quantity - and so it is reasonable to assume the climate will be similar at two nearby locations (Cooley et al., 2007). With this in mind, the $(\beta, \nu)$ combinations above allow for either very short or very long effective spatial ranges, which we take to be the distance at which the correlation equals 0.05 (Finley et al., 2013). The extent of these combinations means that correlation can either drop off very quickly with distance (with an effective spatial range of 10 km) or else reduce very slowly (reducing from 1 to 0.67 between the two furthest points on the grid). This latter combination means all points can be well within the effective spatial
range of all other points in the domain of the study. We feel that this represents a sufficiently broad selection to allow for great flexibility in modelling the spatial fields of the parameters. Though physically plausible (but a conservative approach), we will check that the posteriors for $\beta$ assign negligible probabilities to the extreme combinations in the specified prior range (Diggle et al., 2007).

The remaining two parameters in the Matérn covariance function, $\varsigma^2$ and $\tau^2$, represent the partial sill (variance of the spatial effect) and the nugget (variance of the non-spatial effect - essentially representing the measurement error in repeated measurements at any site) respectively. It is difficult to have information on these parameters a priori, so we chose relatively uninformative priors. As both of these quantities are positive, we chose to model their log: $\log(\varsigma^2) \sim N(0, 1)$ and $\log(\tau^2) \sim N(-2.3, 1)$ (i.e. we assume $\varsigma^2$ has high density values in the neighbourhood of 1, and $\tau^2$ has high density values in the neighbourhood of $\sim 0.1$). The standard deviations of 1 (for the priors on $\log(\varsigma^2)$ and $\log(\tau^2)$) are again sufficiently large to allow the data play the dominant role in informing the posterior distributions for these parameters, without allowing for unrealistic larger values to occur.
4.5 Model details for the temperature dataset

Figure 4.5: A directed acyclic graph (DAG) of the Bayesian hierarchical model fitted to the spatial dataset of daily maximum temperatures. On the left is the model for the excesses (using the GPD) and on the right is the model for the probability of an observation being a cluster maximum (using the Binomial distribution). The parameters of the distributions are modelled on a subgrid using Gaussian predictive processes as detailed in Section 4.5.2. These are represented as circles in the middle layer, with the hyperparameters controlling these represented in the top layer. The data is represented in the bottom layer (in rectangles). Arrows run into nodes from their direct predecessors (often called parents). Given its parents, each node is independent of all other nodes in the graph except its descendants (often called children). Posterior estimates of the parameters’ distributions can be used to form quantities of interest - typically return levels, as illustrated. Though omitted for clarity of the diagram, it should be remembered that the three parameters ($\phi^*$, $\xi^*$, and $\zeta'$) are defined on a subgrid of the domain, but give rise to the predictive processes ($\phi$, $\xi$, and $\zeta$, respectively) defined on the full grid of the domain, using equation (4.2). Further details of each layer and the parameters involved may be found in the text. Further details involving the pseudocode and equations for this model are to be found in Appendix B on page 160.

4.5.4 Threshold selection

In order to ensure independence of observations in time, we declustered the dataset of extremes anomalies by removing all observations which occurred in clusters except for the maximum of this cluster (a process described earlier in this section). That
4.5 Model details for the temperature dataset

is, if two or more consecutive days at any point exceeded the threshold, only the maximum of these values was retained. We then ran our model for various thresholds from the 98th percentile upwards. As detailed in Coles (2001), threshold selection involves a trade-off between bias and variance (higher thresholds lead to reduced bias, but increased variance). In our case, choosing too low a threshold will result in many less-extreme exceedances (i.e. extremes which are marginally above the threshold), which will threaten the asymptotic nature of the GPD model. Choosing too high a threshold will result in too few datapoints, and result in large uncertainties in the posterior estimates. Here, we selected a threshold of the 99.5th percentile - results for simulations below this tended to be overly dominated by the large number of excesses marginally above the threshold, and failed to model the extreme excesses sufficiently well. The posterior predictive intervals of return level curves produced from these model runs were typically too narrow, with the extreme excesses being poorly accounted for, frequently falling outside the range of the predictive intervals. Following declustering of excesses above the 99.5th percentile threshold, there was a median number of 40 independent excesses retained across the domain (the 2.5th percentile value of the number of independent excesses retained was 35, and the corresponding number for the 97.5th percentile value was 50 (see Table 4.1)).

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Threshold</th>
<th>98th</th>
<th>98.5th</th>
<th>99th</th>
<th>99.5th</th>
<th>99.7th</th>
<th>99.9th</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%</td>
<td>136</td>
<td>105</td>
<td>73</td>
<td>35</td>
<td>21</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>148</td>
<td>114</td>
<td>76</td>
<td>37</td>
<td>22</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>152</td>
<td>117</td>
<td>79</td>
<td>40</td>
<td>27</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>162</td>
<td>126</td>
<td>86</td>
<td>46</td>
<td>29</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>97.5%</td>
<td>176</td>
<td>137</td>
<td>93</td>
<td>50</td>
<td>31</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: This table shows, for six selected threshold surfaces of the 98th percentile and above, the remaining number of declustered excesses across the surface for the 2.5%, 25%, 50%, 75% and 97.5% percentile levels. For example, the 98th percentile threshold has a median number of excesses at a gridpoint of 152, with a 2.5th percentile value of 136 excesses and a 97.5th percentile value of 176 excesses.
4.5 Model details for the temperature dataset

4.5.5 Model implementation

We implemented our model using the programming language R (R Core Team, 2017) and the package Rcpp (Eddelbuettel, 2013). A Metropolis-Hastings MCMC algorithm was used to draw samples from the posterior distributions of all parameters in the hierarchy. For both components of the model, three chains were run for 50,000 iterations. A burn-in of 10,000 iterations was discarded from each chain. The remaining chains were then thinned by retaining only every 10th sample to reduce auto-correlation. Convergence was then assessed using the $\hat{R}$ criterion recommended by Gelman (1996), with values below the suggested criterion of 1.2 taken to imply convergence. The resulting simulations are presented in the next section. All code needed to reproduce this analysis is available in a public repository on GitHub at https://github.com/jackos13/extremes.

On Figure 4.6, the full grid (grey circles) over the spatial domain is displayed again, but now also shown is subgrid that we selected for this study (black circles). As the data was already located on a regular $1 \times 1$ km$^2$ grid, there were no 'gaps' in the domain, so we felt that the natural choice was to choose a subset of the sampled locations, rather than a disjoint set of locations across the spatial domain. Using a space-filling approach, we constructed the subgrid ($n^* = 426$) by selecting every second gridpoint from the full domain ($n = 1,699$) in both the horizontal and vertical directions. This spatial design choice ensures that the observed locations are evenly distributed across the domain, as suggested by Finley et al. (2012), who found that it is the number of knot locations which has a greater impact on parameter estimates and reducing the uncertainties in quantities derived from them. With this caveat in mind, along with our computational constraints, we felt that choosing 25% of the points from the full grid, and selecting knots which were regularly spaced across the domain, represented both the optimal design and the best use of our limited resources.
4.6 Results

Figure 4.6: The domain of the study is shown again. The full grid is displayed with grey circles and the selected subgrid is shown with black circles. The subgrid is constructed by selecting every second gridpoint in both the horizontal and vertical directions. The location of four synoptic stations are indicated in bold lettering: Casement Aerodrome (CA - already displayed in Figure 4.1, and used in Section 4.3.2 to verify the gridded dataset); Dun Laoghaire (DL); Dublin Airport (DA); and Phoenix Park (PP). These four sites are used in the site-specific analysis in Section 4.6.3.

4.6 Results

4.6.1 Posterior parameter estimates

Posterior median surfaces for the scale and shape parameters of the GPD are shown in Figure 4.7. The scale parameter (indicating the variability of the distribution) increases with increasing distance from the coast. Along the coastline, it has a median value of between 0.75 and 1. Further inland, it exceeds 1 at most locations,
4.6 Results

extending upwards of 1.5 in the west and south-west of the study domain. This is consistent with the surface of the standard deviation in Figure 4.4(b), which showed that the excesses of the threshold farther from the coast exhibited larger variance. In other words, more variable temperature extremes are observed farther from the coast, due to the increased distance from the moderating effect of the Irish Sea. This corresponds to standard meteorological theory where the diurnal range of temperature generally increases with distance from the sea (Rafferty, 2011).

The shape parameter is slightly more difficult to interpret. A negative shape indicates a finite upper-bound to the corresponding posterior distribution; a shape of 0 indicates that the data has infinite support; while a positive shape indicates a finite lower-bound to the corresponding posterior distribution. The scale and shape are known to be generally negatively correlated (Cooley and Sain, 2010). This is evident here, where the median shape is seen to have the opposite trend to the median scale - the shape surface generally decreases with increasing distance from the coast (Figure 4.7(b)).

The median surface of the zeta parameter $\zeta_u$ is shown in Figure 4.8. The relatively
simple nature of this parameter (essentially, it models the binomial probability that a randomly chosen day is a cluster maximum) led it to converge quite quickly, with very little uncertainty in its posterior distribution. The general pattern here is for higher values nearer the coast. This indicates that extreme temperatures exceeding the threshold are more likely to be isolated incidents here, whereas further inland extremes are more likely to occur in clusters (and therefore there are fewer independent excesses retained here). The higher frequency of isolated excesses near the coast is due to the moderating effect of the sea, which makes it more unlikely for prolonged periods where successive days exceed the threshold to occur.

Figure 4.8: Shown is the posterior estimate for the median surface of the zeta parameter, the probability of a randomly selected day being a cluster maximum.

### 4.6.2 Posterior return level estimates

20- and 100-year return-level median surfaces are shown in Figure 4.9. The 20-year median surface is seen to range from just below 8°C to almost 10°C. The lowest levels are seen along the north-eastern coast of the study domain, while higher return levels are observed inland. The highest return levels are observed on the western side of the domain. The 100-year median surface ranges from 8°C to just under 10.7°C. A similar pattern to the 20-year return level is observed - lowest values appear along the sea in the north-east of the domain, with highest values on the western boundary. (Upper and lower surfaces of the credible intervals for both 20- and 100-year return
4.6 Results

levels show a similar pattern to the median surfaces and are omitted here for brevity. The lower 20-year return level surface has values ranging from $6.5^\circ \text{C}$ to $9.6^\circ \text{C}$, while the upper 20-year return level surface has values ranging from $7.9^\circ \text{C}$ to $10.3^\circ \text{C}$. The lower 100-year return level surface has values ranging from $7.5^\circ \text{C}$ to $10.3^\circ \text{C}$, while the upper 100-year return level surface has values ranging from $8.3^\circ \text{C}$ to $12.7^\circ \text{C}$.

Amongst the highest values in both figures are those close to Casement Aerodrome (note the yellow spot at $\approx 53.3^\circ \text{N}$ and $-6.4^\circ \text{E}$). The highest part of the mean climate curve calculated for this location exceeds $20^\circ \text{C}$ in July. This means that if the median temperature anomaly return levels seen in Figure 4.9 occur at this time of year, daily maximum temperatures will be in excess of $30.5^\circ \text{C}$. Since observations began at Casement Aerodrome in 1944, there have been only two days recorded here with temperatures exceeding $30^\circ \text{C}$: $30.5^\circ \text{C}$ on the 3rd of August 1975 (before the time period of this dataset), and $31^\circ \text{C}$ on the 19th of July 2006. If temperature anomalies at the upper end of the credible interval for the 100-year surface were to occur here in July, daily maximum temperatures would be almost $33^\circ \text{C}$.

Figure 4.9: Shown are the 20- and 100-year return level median surfaces of the temperature anomaly. Highest values are seen for the west of the domain. Amongst the highest values in both figures are those close to Casement Aerodrome (note the yellow spot at $\approx 53.3^\circ \text{N}$ and $-6.4^\circ \text{E}$).

4.6.3 Comparison with recent observations

Figure 4.10 shows the return level curve at Casement Aerodrome (location shown in Figure 4.6), with the observations from the synoptic station overlaid on the plot. The
return period (measured in years) is displayed on a log scale for ease of interpretation. The median return level curve is shown in black; a 95% credible interval is contained within the upper and lower bounds in grey. Observations are included by plotting their empirical return period against their return level anomaly. Empirical return level estimates from the data are added as follows:

1. Order the observed excesses in a vector $z$ such that $z_1 \leq z_2 \leq \cdots \leq z_n$, for the $n_i$ excesses at location $i$;
2. then calculate the corresponding vector of return periods $x$ with entries $x_k$, (for $k = 1 \ldots n_i$) given by:
\[
x_k = \frac{1}{1 - k/(n + 1)} \times \frac{1}{npy};
\]
3. then add all points $(x_k, z_k)$ to the plot (or $(\log(x_k), z_k)$ if plotting on a log scale).

Here $npy$ refers to the number of excesses per year (cf. sections 2.6.7 and 3.3.5 of Coles (2001) for further details on this process).

Figure 4.10: Shown is the return level curve for Casement Aerodrome showing the median (black) and the 95% credible interval (grey) curves, with observations super-imposed for 1981-2010 (blue dots) and 2011-2018 (orange dots).
4.6 Results

Observed excesses from the time period of the study (1981-2010) are plotted with blue circles. Observed excesses from the most recent eight years (2011-2018) are plotted with orange circles. These new data are obtained from the synoptic station at Casement Aerodrome, and are selected in the same way as the original dataset, as detailed earlier. Namely: the climate curve at that location (the curve calculated from the 1981-2010 dataset) is subtracted to get a time series of anomalies; the 99th percentile is found; and independent excesses above this threshold are retained to form the dataset. There were more declustered excesses per year in the period 2011-2018 (12 in total - 1.5 per year) than there were in the period of the study (34 in total - 1.33 per year). However, this increase in the frequency of threshold excesses does not appear to be due to an increase in the severity of threshold excesses: as can be seen in the plot, all of the more recent excesses appear in the lower region of the return level plot. For reasons of clarity, the graph only includes those points with a return period of 2 years or greater. 11 points from 1981-2010 are included in this set, and 7 points from 2011-2018. The greatest empirical return period calculated for an anomaly (almost 39 years) is for an observation from 1981-2010 (indicated by the rightmost blue circle on the graph), whereas the greatest empirical return period calculated for an anomaly from the 2011-2018 period is just under 4 years (indicated by the rightmost orange circle). This again demonstrate the increase in the frequency of threshold excesses at this location from 1981-2010 to 2011-2018, but shows that there is no corresponding increase in their severity.

This increase in the frequency of threshold excesses from recent observational data is seen at more than one location. Figure 4.11 shows posterior 95% credible intervals (black line segments, with a filled circle indicating the median) of the probability of a particular day being a cluster maximum at four synoptic stations across the domain of the study: Casement Aerodrome, Dublin Airport, Dun Laoghaire, and the Phoenix Park (locations shown in Figure 4.6). Superimposed on these credible intervals is the data used to fit the model - the observed site-specific probability of being a cluster maximum for the period 1981-2010 (blue diamonds). Also superimposed is the more recent (2011-2018) observed site-specific probability of being a cluster maximum (red diamonds). From this plot, it can be seen that the greatest increase in the frequency of threshold excesses is at Casement Aerodrome (where the 2011-2018 observed prob-
4.7 Discussion and Conclusions

ability is 20% greater than the upperbound of the credible interval at that site). The next largest increase is seen at the Phoenix Park station, where the 2011-2018 observed probability is 7% greater than the upperbound of the corresponding credible interval. Of the remaining two stations, the 2011-2018 observed probability at Dun Laoghaire is marginally outside the credible interval bounds (it is 2% greater than the upperbound at this location), while the corresponding probability for Dublin Airport is within the credible interval bounds.

Figure 4.11: Shown in black is the 95% credible interval of the probability of being a cluster maximum ($\zeta$) at four synoptic stations: Casement Aerodrome, Dublin Airport, Dun Laoghaire, and the Phoenix Park (locations shown in Figure 4.6); black dots show the posterior median values; blue diamonds show the observed (1981-2010) probability of being a cluster maximum - that is, the data used to fit the model; red diamonds show the more recent (2011-2018) observed probability of being a cluster maximum.

4.7 Discussion and Conclusions

In this research, we began a comprehensive characterisation of temperature extremes in Ireland for the period 1981-2010. We produced return-level surfaces of daily maximum temperature anomalies across County Dublin, the domain of the study. We also produced site-specific return-level curves at synoptic stations, and super-imposed data from 2011-2018 onto these plots. To our knowledge, this is the first study to combine
predictive processes and EVT in the manner we have used here.

We modelled a spatial dataset of daily maximum temperatures over the domain of County Dublin, Ireland, in order to better understand the nature of temperature extremes there. We first created a dataset of anomalies to focus attention on temperatures which would be considered extreme relative to the time of year in which they occurred (adapting a procedure used by Brown et al. (2008)). In order to make the best use of this dataset, we chose a peaks-over-threshold approach, and declustered the exceedances of this threshold to remove successive days of extremes, retaining only the maximum excess from a cluster (Coles, 2001). We then used the GPD and the reduced-rank representation of predictive processes in the first component of two three-level Bayesian hierarchical models. This resulted in (samples of) posterior densities for the scale and shape surfaces, the parameters which uniquely determine the distribution and behaviour of the temperature anomaly excesses. The second Bayesian hierarchical model was applied in a similar manner to the surface of the probability that a day selected at random is a cluster maximum which exceeds the threshold. Our chosen Bayesian approach means that uncertainty is accounted for directly in the estimate of the full posterior density, in contrast to approaches which yield only summary statistics from the target density (Gelman et al., 2013). Values from these posterior densities for the three surfaces were drawn at random in order to calculate estimates of return-level surfaces using equation (4.3). This direct accounting for uncertainty in the parameter distributions means that the uncertainty of the return-level surfaces can be quantified directly, as the process yields posterior densities for these values too. We used the reduced-rank representation of predictive processes to significantly reduce the computational burden of MCMC by modelling the surfaces of interest directly on a subgrid of the domain, while still being able to incorporate and make use of the data at every gridpoint in the domain (Banerjee et al., 2008).

Following model fitting and convergence diagnostics, posterior parameter estimates and return level surfaces were produced. These were presented in Section 4.6. The median return level surfaces showed that, for example, for both 20- and 100-year time periods, exceedances of these values in July would mean maximum daily tem-
4.7 Discussion and Conclusions

Temperatures in excess of 30.5°C at Casement Aerodrome. Further site-specific analysis here showed that, for the period 2011-2018, an increase in the frequency of extreme anomalies, but not the severity, was observed. This increased frequency of extreme anomalies at Casement Aerodrome is such that the observed probability of being a cluster maximum was 20% greater here than the upperbound of the credible interval produced by the model using the 1981-2010 data. This increase was also present to a lesser extent at the Phoenix Park and Dun Laoghaire stations (where the observed probability of being a cluster maximum was 7% and 2% respectively above their corresponding upperbounds).

Extremes are, by definition, rare - our approach ensured that valuable information about the observed extremes was not neglected in pursuit of a relatively fast model-fitting algorithm. Modelling the GPD and binomial parameters as GPs which vary continuously over space also allowed us to make the best use of the limited data which we had - parameter values at a gridpoint are not only informed by the data at that gridpoint, but by the data at surrounding gridpoints too, with nearby gridpoints having more influence than those at a greater distance. Using the space-filling design of a regularly-spaced subgrid (Royle and Nychka, 1998) (by selecting every second gridpoint in both directions) we ensured that every point on the full grid of the domain was close to a point on the subgrid. We felt that this amount of subgrid gridpoints was still a sufficiently large number to retain and made the best use of our computational resources in order to have a positive impact on parameter estimates and in reducing the uncertainties in quantities derived from them (Finley et al., 2012). This reduces the (large) uncertainties which result in parameter estimates from a single-site analysis (e.g, maximum-likelihood estimation at a single point). This was particularly helpful when it came to the shape parameter, the parameter about which there is usually greatest uncertainty and on which it can be difficult to perform inference (Coles and Powell, 1996).

One of the problems with modelling large spatial datasets with a likelihood-based approach using MCMC algorithms is the need to invert large and dense matrices thousands of times (Guhaniyogi et al., 2011). We tackled this problem with the use of reduced-rank representations of the latent spatial processes, and hence the
computational burden was vastly reduced. The problem of modelling extremes (that is, the unavoidable scarcity of data) was approached by using spatial models to make best use of the information contained in the data, and a Bayesian hierarchy in order to set prior distributions for the parameters which were based on physical principles. In Chapter 5, I suggest possible future work that could be undertaken in order to extend the research presented here.

**Acknowledgements**

I am grateful to Séamus Walsh and Sandra Spillane for making the gridded dataset available and for their initial input into the research contained in Chapter 4.
References


Appendix A

This appendix includes additional details and some further data exploration on the procedure of forming anomalies from the temperature dataset.

As described in Section 4.3.3, we decided to focus our research on maximum temperatures which would be considered extreme relative to the time of year in which they occurred. In order to do this, we chose to create and then analyse temperature anomalies. Adapting a procedure used by Brown et al. (2008), we created the anomalies with the following steps:

1. We calculated a 31-day moving average for each location over the 30 years of data, using only values at that location.

2. We then averaged by calendar date in order to get a climate value for the first of January, a climate value for the second of January etc., at that location. These 366 values provided a distinct climate curve for each location - an annual curve of expected values (i.e. the climate) of daily maximum temperature.

3. We then removed this mean annual cycle at each location, by subtracting this climate curve from the 30 years of raw data for the corresponding days - that is, the climate value for the first of January was subtracted from the 30 observed values for the first of January at that location, resulting in 30 observed anomalies, this was then repeated for the second of January, etc.

The following figures illustrate how this process works for a sample year and gridpoint. The domain of the study is shown again in Figure 4.12. The grey circles mark the gridpoints of the spatial domain. Also shown is a red cross which marks the 1000th gridpoint across the domain. (Gridpoints were numbered from south to north, from the bottom-left corner up to the top-left corner, then south to north for the next column right, etc.) This gridpoint is central in the domain, and is used in the sample figures below.
Figure 4.12: The domain of the study is shown. The grey circles mark the gridpoints of the spatial domain. A red cross marks the location of the 1000th gridpoint, which will be used for demonstrative purposes in the sample figures which follow.

Figure 4.13 shows the time series of daily maximum temperature values for a randomly selected year (1997, roughly mid-way through the dataset) at the 1000th gridpoint (as marked on Figure 4.12). Overlaid on this is the climate curve for this particular gridpoint. Marked in red are the three points which (in the subsequent figure) will be found to be considered extreme, as they are above the 99th percentile as calculated from all anomalies at this gridpoint. Judging from this figure alone, it is clear that they are quite high above the climate curve - that is, they are extreme values, relative to the time of year in which they occur. Considered as an isolated piece of data, the value of 13.9°C which occurs on the 11th of January would not be considered an extreme value of maximum temperature in Ireland. But given the time of year in which this occurs, as can be seen from the climate curve, it is a lot higher
than the value which would be expected for the 11th of January at this location, and so is considered to be extreme relative to the time of year and location in which it occurs.

Figure 4.13: Shown above is the time series of daily maximum temperatures from 1997 at the 1000th gridpoint of the domain. The heavy black line shows the underlying climate curve calculated using the 30 years of data available at that gridpoint. Three particular days (seen in Figure 4.14 to be above the 99th percentile threshold of the anomalies, and are therefore considered to be extreme observations) are highlighted in red.

The next figure (Figure 4.14) then shows how these three values were selected as the extreme anomalies from this year at this gridpoint. The anomalies were calculated as described above, and then all 30 years of anomalies at the particular gridpoint were used to calculate the 99th percentile value. This value is then used as the threshold - anomalies in excess of this value are considered as extreme observations. In Figure 4.14, again the data for 1997 at the 1000th gridpoint is shown, but this time the climate curve has been subtracted from the raw data. The time series plotted
now is that of the anomalies of daily maximum temperature. The red line indicates the 99th percentile as calculated over all 30 years of data at this gridpoint. The anomalies indicated in red are the three observations in this year which exceed this threshold.

Figure 4.14: Shown above is the time series of daily maximum temperature anomalies from 1997 at the 1000th gridpoint of the domain. The horizontal red line shows the 99th percentile of the anomalies, as calculated using the 30 years of data at this gridpoint. Anomalies which exceed this threshold are considered as extreme observations at this threshold. Three particular days (the same days as already seen in Figure 4.13) are above this threshold, and are therefore considered to be extreme observations.

Removing the mean annual cycle at each location in this manner allowed us to focus on modelling temperatures which would be considered extreme relative to the time of year in which they occur. In order to ensure that the data had no seasonal biases despite this anomaly approach, we did some further data exploration. Across all locations, we found that observations in summer were more likely to exceed the
threshold and therefore be considered as extreme observations. The mean proportion of excesses across the grid which occurred during the months of summer (June, July and August) was 40%. Spring (March, April, May) had the next largest mean proportion of excesses across the grid (32%), autumn (September, October, November) had a mean of 14%, and finally winter (December, January, February) had a mean of 13%. Different numbers of observations from different seasons is not in itself a problem as long as the underlying distributions to which the excesses asymptotically converge are the same.

To investigate if this was the case, we calculated the maximum-likelihood estimates (MLEs) of the scale and shape parameters of the GPD at each location across the domain, after splitting the data into bins by season. Figure 4.15 shows the spread of these MLEs across the domain, summarised in boxplots for each season, in addition to an annual boxplot where all observations are considered together. The larger range for both parameters for autumn and winter (which includes some very unrealistic outlier values of both the scale and shape) is due to the smaller number of excesses for these seasons. This is also why the annual boxplot is the least variable in both cases - it has the most observations used to inform the parameter estimates, and therefore these estimates vary less across the spatial domain. Overall, the median or central behaviour of the MLEs across the grid is not very different for all four seasons. The generally higher estimates of the scale parameter for summer excesses is countered by generally lower estimates of the shape parameter for summer excesses when compared to the other seasons. These parameters are known to be negatively correlated, where a particular dataset may be explained similarly well by a high scale/low shape or low scale/high shape parameter pair (Cooley and Sain, 2010).
Figure 4.15: Shown are boxplots of the maximum-likelihood estimates (MLEs) of the scale (top) and shape (bottom) parameters of the GPD at each location across the domain, where the dataset is binned by season. Also shown for comparison is an annual boxplot where all observations are considered together.
Following this data exploration, we feel confident that we can proceed with our analysis of extreme anomalies of daily maximum temperature, where extreme temperatures are considered relative to the time of year and the location in which they occur, using the procedure described above.
Appendix B

This appendix includes further details on the model described in Section 4.5; namely, it describes in detail how we modelled a dataset of extreme observations using a generalised Pareto distribution with predictive processes and Markov chain Monte Carlo updates. Details for the Binomial component of the model are similar, and are omitted here for brevity.

The following pages include:

- the pseudocode in order to programme the algorithm;
- the model directed acyclic graph (DAG);
- details of the model notation;
- and the details of all equations needed for the MCMC updates of the parameters and the hyperparameters.
Appendix B

Algorithm details

Data: $y_{ij}$, the declustered threshold excesses at locations $i = 1 \ldots n$, with $j = 1 \ldots n_i$ excesses at location $i$;
$X_\phi$, $X_\xi$, $n \times (p + 1)$ and $n \times (q + 1)$ matrices of $p$ and $q$ covariates at locations $i = 1 \ldots n$.

Result: Samples from the posterior distributions of $\phi = \log(\sigma)$ and $\xi$ (the unknown parameters of interest), which can then be used to calculate return level estimates and other desired quantities.

Initialisation:
Random starting values of $\phi^*$ and $\xi^*$;
Projection of $\phi$ and $\xi$ from $\phi^*$ and $\xi^*$ respectively;
Hyper-parameter values of $\alpha_\phi$, $\beta_\phi$, $\varsigma^2_\phi$, $\tau^2_\phi$, $\nu_\phi$, $\alpha_\xi$, $\beta_\xi$, $\varsigma^2_\xi$, $\tau^2_\xi$, and $\nu_\xi$;
Number of iterations $N$;
for iterations $i$ from 1 to $N$ do
  Generate $u \sim U(0, 1)$;
  for sub-grid locations $k$ from 1 to $m$ do
    Simulate $\phi^*_{\text{new},k}$;
    Project $\phi_{\text{new}}$ from $\phi^*_{\text{new}}$;
    Set $l_{\text{new}} = \log$ full conditional of new vector $\phi^*_{\text{new}}$;
    Set $l_{\text{old}} = \log$ full conditional of old vector $\phi^*$;
    Set $a = \exp(l_{\text{new}} - l_{\text{old}})$, that is, evaluate equation (4.4);
    if $a > u$ then
      Set $\phi^* = \phi^*_{\text{new}}$;
    end
    Simulate $\xi^*_{\text{new},k}$;
    Project $\xi_{\text{new}}$ from $\xi^*_{\text{new}}$;
    Set $l_{\text{new}} = \log$ full conditional of new vector $\xi^*_{\text{new}}$;
    Set $l_{\text{old}} = \log$ full conditional of old vector $\xi^*$;
    Set $a = \exp(l_{\text{new}} - l_{\text{old}})$, that is, evaluate equation (4.5);
    if $a > u$ then
      Set $\xi^* = \xi^*_{\text{new}}$;
    end
  end
end
for iterations $i$ from 1 to $N$ (continued) do

for each element $c$ of the vector $\alpha_\phi$ do

Simulate $\alpha_{\phi_{\text{new}},c}$;

Set $a =$ result of equation (4.6);

If $a > u$, set $\alpha_\phi = \alpha_{\phi_{\text{new}}}$;

end

for each element $d$ of the vector $\alpha_\xi$ do

As above, but set $a =$ result of equation (4.7);

end

Simulate $\beta_{\phi_{\text{new}}}$;

Set $a =$ result of equation (4.8);

If $a > u$, set $\beta_\phi = \beta_{\phi_{\text{new}}}$;

Repeat for $\beta_\xi$ with equation (4.9);

Simulate $\nu_{\phi_{\text{new}}}$;

Set $a =$ result of equation (4.10);

If $a > u$, set $\nu_\phi = \nu_{\phi_{\text{new}}}$;

Repeat for $\nu_\xi$ with equation (4.11);

Simulate $\varsigma_{\phi_{\text{new}}}^2$;

Set $a =$ result of equation (4.12);

If $a > u$, set $\varsigma_\phi^2 = \varsigma_{\phi_{\text{new}}}^2$;

Repeat for $\varsigma_\xi^2$ with equation (4.13);

Simulate $\tau_{\phi_{\text{new}}}^2$;

Set $a =$ result of equation (4.14);

If $a > u$, set $\tau_\phi^2 = \tau_{\phi_{\text{new}}}^2$;

Repeat for $\tau_\xi^2$ with equation (4.15);

end
DAG

Figure 4.16: A directed acyclic graph (DAG) of the Bayesian hierarchical model fitted to the spatial dataset of daily maximum temperature anomalies. On the left is the model for the excesses (using the GPD) and on the right is the model for the probability of an observation being a cluster maximum (using the Binomial distribution). The parameters of the distributions are modelled on a subgrid using Gaussian predictive processes as detailed in Section 4.5.2. These are represented as circles in the middle layer, with the hyperparameters controlling these represented in the top layer. The data is represented in the bottom layer (in rectangles). Arrows run into nodes from their direct predecessors (often called parents). Given its parents, each node is independent of all other nodes in the graph except its descendants (often called children). Posterior estimates of the parameters’ distributions can be used to form quantities of interest - typically return levels, as illustrated. Though omitted for clarity of the diagram, it should be remembered that the three parameters (φ*, ξ*, and ζ′) are defined on a subgrid of the domain, but give rise to the predictive processes (φ, ξ, and ζ, respectively) defined on the full grid of the domain, using equation (4.2). Further details of each layer and the parameters involved may be found in the text and in this appendix.
Appendix B

Notation

- $x_i$ Multivariate (bivariate or tri-variate) location values for location $i$, $i = 1, \ldots, n$. Write the matrix of all locations as just $x$.

- $y_{ij}$ Excess $j$ for observation $i$, $j = 1, \ldots, n_i$ where $n_i$ is the number of excesses at location $i$.

- $\sigma(x_i)$ Scale parameter for location $x_i$.

- $\phi(x_i) = \log(\sigma(x_i))$ The re-parameterised scale parameter.

- $\xi(x_i)$ Shape parameter for location $x_i$.

- $z_k$ Sub-grid locations $k = 1, \ldots, m$. Together written as $z$.

- $A, B$ Projection matrices of dimension $n \times m$.

- $\phi^*, \xi^*$ Gaussian processes of $\phi$ and $\xi$ defined on sub-grid $z$.

- $\mu_{\phi}(z), \mu_{\xi}(z)$ Means for the Gaussian processes.

- $\Sigma, \Psi$ Covariance matrices for the Gaussian processes.

- $\tau^2_{\phi}, \tau^2_{\xi}$ Nugget parameters for the Gaussian processes.

- $\alpha_{\phi}, \alpha_{\xi}$ Vectors of coefficients (including intercept) for Gaussian process means.

- $X_{\phi}, X_{\xi}, Z_{\sigma}, Z_{\xi}$ Matrices of covariates (including column for intercept term) on the $x$ grid and the $z$ sub-grid.

- $\nu_{\phi}, \nu_{\xi}$ Matern smoothness parameters for the Gaussian processes.

- $\beta_{\phi}, \beta_{\xi}$ Matern length scale scalars or matrices for the Gaussian processes.

- $\varsigma^2_{\phi}, \varsigma^2_{\xi}$ Variance parameters (partial sills) for the Gaussian processes.
Appendix B

Model outline

In hierarchical notation:

\[ y_{ij} \sim GPD(\sigma(x_i), \xi(x_i)) \]

\[ \log(\sigma(x)) = \phi(x) = A(x, z)\Sigma^{-1}(z, z)\phi^*(z) \]

\[ \xi(x) = B(x, z)\Psi^{-1}(z, z)\xi^*(z) \]

\[ A(x_i, z_k) = \sigma^2 \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \left( \sqrt{\frac{\nu_\phi}{\beta_\phi}} \right)^{\nu_\phi} \left( \sqrt{\frac{\nu_\phi}{\beta_\phi}} \right) \left( \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \right) \left( \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \right) \]

\[ B(x_i, z_k) = \xi^2 \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \left( \sqrt{\frac{\nu_\xi}{\beta_\xi}} \right)^{\nu_\xi} \left( \sqrt{\frac{\nu_\xi}{\beta_\xi}} \right) \left( \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \right) \left( \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \right) \]

\[ \phi^* \sim MVN_m(\mu_\phi(z), \tau_\phi^2 I_m + \Sigma(z, z)) \]

\[ \xi^* \sim MVN_m(\mu_\xi(z), \tau_\xi^2 I_m + \Psi(z, z)) \]

\[ \mu_\phi(z) = Z_\phi \alpha_\phi \]

\[ \mu_\xi(z) = Z_\xi \alpha_\xi \]

\[ \mu_\phi(x) = X_\phi \alpha_\phi \]

\[ \mu_\xi(x) = X_\xi \alpha_\xi \]

\[ \Sigma(z_k, z_l) = \sigma^2 \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \left( \sqrt{\frac{\nu_\phi}{\beta_\phi}} \right)^{\nu_\phi} \left( \sqrt{\frac{\nu_\phi}{\beta_\phi}} \right) \left( \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \right) \left( \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \right) \]

\[ \Psi(z_k, z_l) = \xi^2 \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \left( \sqrt{\frac{\nu_\xi}{\beta_\xi}} \right)^{\nu_\xi} \left( \sqrt{\frac{\nu_\xi}{\beta_\xi}} \right) \left( \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \right) \left( \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \right) \]

In the above formulae for \( A, B, \Sigma, \) and \( \Psi, \beta_\phi \) and \( \beta_\xi \) are treated as scalars. When considered as matrices, where \( d \) is the distance between two gridpoints, \( d^T\beta^{-1}d \) is replaced by \( \sqrt{d^T\beta^{-1}d} \).
Appendix B

Hyperparameter prior distributions (justifications for these values can be found in Section 4.5.3):

\[
\begin{align*}
\log(\varsigma_\phi^2), \log(\varsigma_\xi^2) &\sim N(0, 1) \\
\log(\tau_\phi^2), \log(\tau_\xi^2) &\sim N(-2.3, 1) \\
\beta_\phi, \beta_\xi &\sim DU(0, 0.05, 1, 10) \\
\nu_\phi, \nu_\xi &\sim DU(0.5, 2.5) \\
\alpha_\phi &\sim MVN(\vec{\eta}_\phi, H_\phi) \\
\alpha_\xi &\sim MVN(\vec{\eta}_\xi, H_\xi)
\end{align*}
\]

\(\vec{\eta}_\phi\) and \(\vec{\eta}_\xi\) are vectors of the appropriate length with all entries equal to 0. \(H_\phi\) and \(H_\xi\) are the relevant covariance matrices - to begin, these are diagonal matrices with 2 on the diagonal. The code is designed to allow the relationship between the coefficients to be modelled (if desired) by adjusting this covariance matrix.

**Posterior distribution**

The full posterior distribution is:

\[
\begin{align*}
p(\alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \varsigma_\phi^2, \varsigma_\xi^2, \tau_\phi^2, \tau_\xi^2, \nu_\phi, \nu_\xi, \phi^*, \xi^* | y, x, z, X_\phi, X_\xi, Z_\phi, Z_\xi) &\propto \\
&\left[ \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(y_{ij} | \sigma(x_i) = \exp(\phi(x_i)), \xi(x_i)) \right] \times \\
p(\phi^*(x) | \mu_\phi, \tau_\phi^2, \Sigma)p(\xi^*(x) | \mu_\xi, \tau_\xi^2, \Psi) \times \\
p(\alpha_\phi)p(\alpha_\xi)p(\beta_\phi)p(\beta_\xi)p(\nu_\phi)p(\nu_\xi) \times \\
p(\tau_\phi^2)p(\tau_\xi^2)p(\varsigma_\phi^2)p(\varsigma_\xi^2)
\end{align*}
\]

**Conditional posterior distributions: Layer 2**

Updating the second layer of the DAG (parameters \(\phi^* = \log(\sigma^*)\) and \(\xi^*)\): The conditional posterior distribution of \(\phi^*\) is given by:
\[ \pi(\phi^* | y, x, z, X_\phi, X_\xi, Z_\phi, Z_\xi, \alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \sigma^2_\phi, \sigma^2_\xi, \tau^2_\phi, \tau^2_\xi, \nu_\phi, \nu_\xi, \xi^*) \propto \\
p(y | \phi^*, x, z, X_\phi, X_\xi, Z_\phi, Z_\xi, \alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \sigma^2_\phi, \sigma^2_\xi, \tau^2_\phi, \tau^2_\xi, \nu_\phi, \nu_\xi, \xi^*) \times \\
p(\phi^* | x, z, X_\phi, X_\xi, Z_\phi, Z_\xi, \alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \sigma^2_\phi, \sigma^2_\xi, \tau^2_\phi, \tau^2_\xi, \nu_\phi, \nu_\xi, \xi^*) \propto \\
p(y | \sigma, \xi)p(\phi^* | \mu_\phi, \tau^2_\phi, \Sigma) \]

That is, the conditional posterior of \( \phi^* \) is proportional to the product of the likelihood of the data \( y \) given \( \phi^* \) (and its associated projection on the full grid, \( \phi \)) and all other parameters, and the prior probability density of \( \phi^* \) given all of the other parameters. In the final line, most parameters have dropped out from the right-hand side, as the densities are independent of these, given the remaining terms (see the DAG in Figure 4.16 for this). Remember that \( \phi = \log(\sigma) \) throughout this. Of the remaining terms, \( \xi, \mu_\phi \) and \( \Sigma \) are deterministic given the other parameters. These will have been calculated using the formulae in the model outline (above).

In briefer notation (to be used from now on):

\[
\pi(\phi^* | \ldots ) \propto p(y | \sigma, \xi)p(\phi^* | \mu_\phi, \tau^2_\phi, \Sigma) \\
= \left[ \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(y_{ij} | \sigma(x_i), \xi(x_i)) \right] p(\phi^* | \mu_\phi, \tau^2_\phi, \Sigma) \\
= \left[ \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma(x_i)} \right)^{-\frac{1}{2}} \right] \times \\
\frac{1}{\sqrt{\det(2\pi(\Sigma + \tau^2_\phi I_m))}} e^{-\frac{1}{2}((\phi^* - \mu_\phi)'(\Sigma + \tau^2_\phi I_m)^{-1}(\phi^* - \mu_\phi))} \\
\]
\[ \pi(\xi^* | \ldots) \propto p(y | \sigma, \xi)p(\xi^* | \mu_\xi, \tau_\xi^2, \Psi) \]
= \left[ \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(y_{ij} | \sigma(x_i), \xi(x_i)) \right] p(\xi^* | \mu_\xi, \tau_\xi^2, \Psi)
= \left[ \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sigma(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma(x_i)} \right)^{-\left(1/(\xi(x_i)+1)\right)} \right] \times
\frac{1}{\sqrt{\det(2\pi(\Psi + \tau_\xi^2 I_m))}} e^{-\frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi)}

**Metropolis-Hastings Markov chain Monte Carlo sampling**

In order to sample from these conditional posterior distributions, we use the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. New samples are accepted or rejected at random according to the algorithm outlined below.

A new value of \( \phi^*_k \) is suggested: \( \phi'^*_k \), where \( k \) is a location on the subgrid. The new vector \( \phi' \) is then calculated on the full grid using the projection formula.

Though \( \phi'^* \) differs from \( \phi^* \) in just one location \( k \), \( \phi \) and \( \phi' \) can be different from each other in many locations \( i \) due to this projection.

Suggested updates are drawn from a Normal distribution centred on the old value and with a variance of a manually set tuning parameter used to control the size of the proposed steps.

We calculate:

\[ \rho(\phi'^*_k, \phi^*_k) = \min \left( 1, \frac{\pi(\phi'^* | \ldots)q_t(\phi'^*_k \to \phi^*_k)}{\pi(\phi^* | \ldots)q_t(\phi^*_k \to \phi'^*_k)} \right) \]

where \( \pi(\phi^* | \ldots) \) is as defined above, and \( q_t(a \to b) \) is the transition probability of proposing value \( b \) given value \( a \). Since updates are proposed using a Normal
distribution, these transition probabilities above and below the line will always cancel, so the above simplifies to:

$$\rho(\phi^*_k, \phi^\prime_k) = \min \left( 1, \frac{\pi(\phi^\prime \mid \ldots)}{\pi(\phi^* \mid \ldots)} \right)$$

Following this calculation, we always accept proposed value \( \phi^\prime_k \) when \( \rho(\phi^*_k, \phi^\prime_k) \) equals 1 and we reject accordingly when the ratio is smaller than 1 by simulating a random variable \( u \sim U[0, 1] \) and accepting proposed value \( \phi^\prime_k \) when \( u \leq \rho(\phi^*_k, \phi^\prime_k) \).

Evaluating \( \rho(\phi^*_k, \phi^\prime_k) \) typically involves products and quotients of many terms which may be close to 0. In order to work with something far more computationally stable, we use the property that \( x = \exp(\log(x)) \).

Following this observation we need to evaluate:

$$\exp \left( \log \left( \frac{\pi(\phi^\prime \mid \ldots)}{\pi(\phi^* \mid \ldots)} \right) \right)$$

$$= \exp \left[ \log(\pi(\phi^\prime \mid \ldots)) - \log(\pi(\phi^* \mid \ldots)) \right]$$

$$= \exp \left[ \log \left( \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(y_{ij} \mid \sigma'(x_i), \xi(x_i)) \right) p(\phi^\prime \mid \mu_\phi, \tau^2_\phi, \Sigma) \right] -$$

$$\log \left( \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(y_{ij} \mid \sigma(x_i), \xi(x_i)) \right) p(\phi^* \mid \mu_\phi, \tau^2_\phi, \Sigma) \right]$$

Filling in the distribution details, this becomes:
\[
\exp \left[ \log \left( \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma'(x_i)} \right)^{-(1/\xi(x_i)+1)} \right) \times \right.
\]
\[
\frac{1}{\sqrt{\det(2\pi(\Sigma + \tau^2_{\phi} I_m))}} e^{-\frac{1}{2}(\phi'-\mu_{\phi})' (\Sigma + \tau^2_{\phi} I_m)^{-1} (\phi' - \mu_{\phi})} \]
\[
- \log \left( \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma'(x_i)} \right)^{-(1/\xi(x_i)+1)} \right) \times \]
\[
\frac{1}{\sqrt{\det(2\pi(\Sigma + \tau^2_{\phi} I_m))}} e^{-\frac{1}{2}(\phi^*-\mu_{\phi})' (\Sigma + \tau^2_{\phi} I_m)^{-1} (\phi^* - \mu_{\phi})} \right]
\]
\[
= \exp \left[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma'(x_i)} \right)^{-(1/\xi(x_i)+1)} \right) \right] +
\]
\[
\log \left( \frac{1}{\sqrt{\det(2\pi(\Sigma + \tau^2_{\phi} I_m))}} \right) + \frac{1}{2}(\phi^* - \mu_{\phi})' (\Sigma + \tau^2_{\phi} I_m)^{-1} (\phi^* - \mu_{\phi}) \]
\[
- \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma'(x_i)} \right)^{-(1/\xi(x_i)+1)} \right) -
\]
\[
\log \left( \frac{1}{\sqrt{\det(2\pi(\Sigma + \tau^2_{\phi} I_m))}} \right) + \frac{1}{2}(\phi^* - \mu_{\phi})' (\Sigma + \tau^2_{\phi} I_m)^{-1} (\phi^* - \mu_{\phi}) \right]
\]
\[
= \exp \left[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma'(x_i)} \right)^{-(1/\xi(x_i)+1)} \right) \right] -
\]
\[
\frac{1}{2}(\phi^* - \mu_{\phi})' (\Sigma + \tau^2_{\phi} I_m)^{-1} (\phi^* - \mu_{\phi}) -
\]
\[
\sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma'(x_i)} \right)^{-(1/\xi(x_i)+1)} \right) +
\]
\[
\frac{1}{2}(\phi^* - \mu_{\phi})' (\Sigma + \tau^2_{\phi} I_m)^{-1} (\phi^* - \mu_{\phi}) \right]
\]

The common term \( \log(1/\sqrt{\det(2\pi(\Sigma + \tau^2_{\phi} I_m))}) \) in both numerator and denominator above dropped out as it will be equal in both (it has no dependence on \( \phi^* \)).

The GPD can clearly be simplified further using the properties of logs. Just taking one of the functions on its own for clarity:
\[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \left( \frac{1}{\sigma'(x_i)} \left( 1 + \frac{\xi(x_i) \cdot y_{ij}}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right) = \]
\[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \log \left( \frac{1}{\sigma'(x_i)} \right) + \log \left( 1 + \frac{\xi(x_i) \cdot y_{ij}}{\sigma'(x_i)} \right)^{-\left(1/\xi(x_i)+1\right)} \right] = \]
\[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ -\log(\sigma'(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \frac{\xi(x_i) \cdot y_{ij}}{\sigma'(x_i)} \right) \right] \]

Using this form, the full term we need to evaluate in order to update \( \phi^* \) is:

\[ = \exp \left[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ -\log(\sigma'(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \frac{\xi(x_i) \cdot y_{ij}}{\sigma'(x_i)} \right) \right] - \frac{1}{2} (\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) \right] - \]
\[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ -\log(\sigma(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \frac{\xi(x_i) \cdot y_{ij}}{\sigma'(x_i)} \right) \right] + \]
\[ \frac{1}{2} (\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) \] (4.4)

Similar reasoning leads to the update for \( \xi^* \). We need to evaluate:

\[ \rho(\xi_k^*, \xi_k'^*) = \min \left( \frac{\pi(\xi^*) \ldots q_k(\xi_k' \rightarrow \xi_k^*)}{\pi(\xi^*) \ldots q_k(\xi_k^* \rightarrow \xi_k'^*)} \right) \]

The full term we need to evaluate is:
Appendix B

\[
= \exp \left[ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ -\log(\sigma(x_i)) - \left( \frac{1}{\xi'(x_i)} + 1 \right) \log \left( 1 + \xi'(x_i) \frac{y_{ij}}{\sigma(x_i)} \right) \right] - \frac{1}{2} (\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) - \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ -\log(\sigma(x_i)) - \left( \frac{1}{\xi(x_i)} + 1 \right) \log \left( 1 + \xi(x_i) \frac{y_{ij}}{\sigma(x_i)} \right) \right] + \frac{1}{2} (\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) \right] \tag{4.5}
\]

**Conditional posterior distributions: Layer 3**

\[\alpha\]

Updating the third layer of the DAG (that is, all hyperparameters) starting with \(\alpha_\phi\):

\[
\pi(\alpha_\phi|y, \ldots) \propto p(y|\alpha_\phi, \ldots)p(\alpha_\phi|\ldots)
\]

\[
\propto p(y|\sigma, \xi)p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)p(\alpha_\phi)
\]

\[
\propto p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)p(\alpha_\phi)
\]

Essentially the GPD component is independent of \(\alpha_\phi\) once the other parameters are known, and so can be absorbed into the constant of proportionality. What we’re left with is the \(MVN\) piece (since \(\alpha_\phi\) features in the calculation of \(\mu_\phi\)) and the prior on \(\alpha_\phi\).

Then we have:
\[
\pi(\alpha_\phi \mid \ldots) \propto \exp(\phi^* | \mu_\phi, \tau_\phi^2, \Sigma)p(\alpha_\phi) \\
\propto \frac{1}{\sqrt{\det(2\pi(\Sigma + \tau_\phi^2 I_m))}} e^{-\frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi)} \times \\
\frac{1}{\sqrt{\det(2\pi H_\phi)}} e^{-\frac{1}{2}(\alpha_\phi - \eta_\phi)'H_\phi^{-1}(\alpha_\phi - \eta_\phi)}
\]

where \(H_\phi\) is the covariance matrix for the prior distribution of \(\alpha_\phi\) and \(\eta_\phi\) is the prior mean.

Similarly:

\[
\pi(\alpha_\xi \mid \ldots) \propto \exp(\xi^* | \mu_\xi, \tau_\xi^2, \Psi)p(\alpha_\xi) \\
\propto \frac{1}{\sqrt{\det(2\pi(\Psi + \tau_\xi^2 I_m))}} e^{-\frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi)} \times \\
\frac{1}{\sqrt{\det(2\pi H_\xi)}} e^{-\frac{1}{2}(\alpha_\xi - \eta_\xi)'H_\xi^{-1}(\alpha_\xi - \eta_\xi)}
\]

where \(H_\xi\) is the covariance matrix for the prior distribution of \(\alpha_\xi\) and \(\eta_\xi\) is the prior mean.

As with the parameters \(\phi^*\) and \(\xi^*\), updates for \(\alpha_\phi\) are suggested element-wise: \(\alpha_{\phi,\kappa} \rightarrow \alpha'_{\phi,\kappa}\), where the index \(\kappa\) runs over the vector of coefficients. The ratio \(\rho(\alpha_{\phi,\kappa}; \alpha'_{\phi,\kappa})\) is then calculated. Remembering that \(\mu_\phi\) is calculated from \(\alpha_\phi\) (so that \(\mu'_\phi\) is the updated value, given \(\alpha'_{\phi}\)), then similar manipulations to those used in the previous section lead to the following calculation needed to update \(\alpha_\phi\):
\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi (\Sigma + \tau_\phi^2 I_m))) - \frac{1}{2} (\phi^* - \mu_{\phi}')^T(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_{\phi}') \\
- \frac{1}{2} \log(\det(2\pi H_\phi)) - \frac{1}{2} (\alpha_{\phi}' - \eta_{\phi})^T H_\phi^{-1}(\alpha_{\phi}' - \eta_{\phi}) \\
+ \frac{1}{2} \log(\det(2\pi (\Sigma + \tau_\phi^2 I_m))) + \frac{1}{2} (\phi^* - \mu_{\phi}')^T(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_{\phi}') \\
+ \frac{1}{2} \log(\det(2\pi H_\phi)) + \frac{1}{2} (\alpha_{\phi}' - \eta_{\phi})^T H_\phi^{-1}(\alpha_{\phi}' - \eta_{\phi}) \right] = \\
\exp \left[ -\frac{1}{2} (\phi^* - \mu_{\phi}')^T(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_{\phi}') - \frac{1}{2} (\alpha_{\phi}' - \eta_{\phi})^T H_\phi^{-1}(\alpha_{\phi}' - \eta_{\phi}) \\
+ \frac{1}{2} (\phi^* - \mu_{\phi}')^T(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_{\phi}') + \frac{1}{2} (\alpha_{\phi}' - \eta_{\phi})^T H_\phi^{-1}(\alpha_{\phi}' - \eta_{\phi}) \right] 
\]
(4.6)

And for \(\alpha_\xi\):

\[
\exp \left[ -\frac{1}{2} (\xi^* - \mu_{\xi}')^T(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_{\xi}') - \frac{1}{2} (\alpha_{\xi}' - \eta_{\xi})^T H_\xi^{-1}(\alpha_{\xi}' - \eta_{\xi}) \\
+ \frac{1}{2} (\xi^* - \mu_{\xi}')^T(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_{\xi}') + \frac{1}{2} (\alpha_{\xi}' - \eta_{\xi})^T H_\xi^{-1}(\alpha_{\xi}' - \eta_{\xi}) \right] 
\]
(4.7)

Updates for the other hyperparameters are very similar - although none feature the \(MVN\) prior that \(\alpha_{\phi}\) and \(\alpha_\xi\) do. In addition, the part involving the log determinant can’t be dropped from the \(m\)-dim \(MVN\), since the covariance matrix changes with any change in \(\beta, \nu, \varsigma^2\) or \(\tau^2\). The remaining hyperparameters either have a discrete update (in which case the prior probability will be \(1/\#\text{discrete.values}\) and will cancel above and below the line, or have a univariate Normal prior on them (or their log).

\(\beta\)

For \(\beta_{\phi}\) we have:
\[ \pi(\beta_\phi | \ldots) \propto p(\phi^* | \mu_\phi, \tau_\phi^2, \Sigma)p(\beta_\phi) \]

Since \( \beta_\phi \) has a discrete prior, to update \( \beta_\phi \) we need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi(\Sigma' + \tau_\phi^2 I_m))) - \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma' + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) \\
+ \frac{1}{2} \log(\det(2\pi(\Sigma + \tau_\phi^2 I_m))) + \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) \right] \tag{4.8} \]

where \( \Sigma' \) has been formed using \( \beta'_\phi \), the new proposal value.

Then for \( \beta_\xi \) we have:

\[ \pi(\beta_\xi | \ldots) \propto p(\xi^* | \mu_\xi, \tau_\xi^2, \Psi)p(\beta_\xi) \]

To update \( \beta_\xi \) we need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi(\Psi' + \tau_\xi^2 I_m))) - \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi' + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) \\
+ \frac{1}{2} \log(\det(2\pi(\Psi + \tau_\xi^2 I_m))) + \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) \right] \tag{4.9} \]

where \( \Psi' \) has been formed using \( \beta'_\xi \), the new proposal value.

\( \nu \)

\( \nu_\phi \) also has a discrete update and so will look identical to the update for \( \beta_\phi \). It has conditional posterior distribution of:

\[ \pi(\nu_\phi | \ldots) \propto p(\phi^* | \mu_\phi, \tau_\phi^2, \Sigma)p(\nu_\phi) \]
We need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi (\Sigma + \tau^2 I_m))) - \frac{1}{2} (\phi^* - \mu_\phi)'(\Sigma' + \tau^2 I_m)^{-1}(\phi^* - \mu_\phi) \\
+ \frac{1}{2} \log(\det(2\pi (\Sigma + \tau^2 I_m))) + \frac{1}{2} (\phi^* - \mu_\phi)'(\Sigma + \tau^2 I_m)^{-1}(\phi^* - \mu_\phi) \right]
\] (4.10)

where \( \Sigma' \) has been formed using \( \nu'_\phi \), the new proposal value.

In order to update \( \nu_\xi \) we have:

\[
\pi(\nu_\xi | \ldots) \propto p(\xi^* | \mu_\xi, \tau^2_\xi, \Psi)p(\nu_\xi)
\]

We therefore need to evaluate:

\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi (\Psi' + \tau^2_\xi I_m))) - \frac{1}{2} (\xi^* - \mu_\xi)'(\Psi' + \tau^2_\xi I_m)^{-1}(\xi^* - \mu_\xi) \\
+ \frac{1}{2} \log(\det(2\pi (\Psi + \tau^2_\xi I_m))) + \frac{1}{2} (\xi^* - \mu_\xi)'(\Psi + \tau^2_\xi I_m)^{-1}(\xi^* - \mu_\xi) \right]
\] (4.11)

where \( \Psi' \) has been formed using \( \nu'_\xi \), the new proposal value.

\( \varsigma^2 \)

\( \varsigma^2_\phi \) has a conditional posterior distribution of:

\[
\pi(\varsigma^2_\phi|\ldots) \propto p(\phi^* | \mu_\phi, \tau^2_\phi, \Sigma)p(\varsigma^2_\phi)
\]

The prior distribution of the log of \( \varsigma^2_\phi \) is a univariate Normal. So to update \( \varsigma^2_\phi \) we need to evaluate:
\[ \exp \left[ -\frac{1}{2} \log(\det(2\pi(\Sigma' + \tau^2_{\phi} I_m))) - \right. \\
\left. \frac{1}{2} (\phi^* - \mu_{\phi})'(\Sigma' + \tau^2_{\phi} I_m)^{-1}(\phi^* - \mu_{\phi}) - \frac{(\log(\varsigma^2_\phi) - m)^2}{2s^2} + \\
\frac{1}{2} \log(\det(2\pi(\Sigma + \tau^2_\phi I_m))) + \\
\frac{1}{2} (\phi^* - \mu_{\phi})'(\Sigma + \tau^2_\phi I_m)^{-1}(\phi^* - \mu_{\phi}) + \frac{(\log(\varsigma^2_\phi) - m)^2}{2s^2} \right] \\
(4.12) \]

where \( \Sigma' \) has been formed using \( \varsigma^2_\phi \), the new proposal value, and \( m \) and \( s \) are the prior mean and standard deviation respectively.

\( \varsigma^2_\xi \) has a conditional posterior distribution of:

\[ \pi(\varsigma^2_\xi \mid \ldots) \propto p(\xi^* \mid \mu_\xi, \tau^2_\xi, \Psi) p(\varsigma^2_\xi) \]

To update \( \varsigma^2_\xi \) we need to evaluate:

\[ \exp \left[ -\frac{1}{2} \log(\det(2\pi(\Psi' + \tau^2_\xi I_m))) - \\
\frac{1}{2} (\xi^* - \mu_\xi)'(\Psi' + \tau^2_\xi I_m)^{-1}(\xi^* - \mu_\xi) - \frac{(\log(\varsigma^2_\xi) - m)^2}{2s^2} + \\
\frac{1}{2} \log(\det(2\pi(\Psi + \tau^2_\xi I_m))) + \\
\frac{1}{2} (\xi^* - \mu_\xi)'(\Psi + \tau^2_\xi I_m)^{-1}(\xi^* - \mu_\xi) + \frac{(\log(\varsigma^2_\xi) - m)^2}{2s^2} \right] \\
(4.13) \]

where \( \Psi' \) has been formed using \( \varsigma^2_\phi \), the new proposal value, and \( m \) and \( s \) are the prior mean and standard deviation respectively.
$\tau^2$

$\tau^2_\phi$ has a conditional posterior distribution of:

$$
\pi(\tau^2_\phi|\ldots) \propto p(\phi^*|\mu_\phi, \tau^2_\phi, \Sigma)p(\tau^2_\phi)
$$

The prior distribution of the log of $\tau^2_\phi$ is a univariate Normal. So to update $\tau^2_\phi$ we need to evaluate:

$$
\exp \left[ -\frac{1}{2} \log(\det(2\pi(\Sigma + \tau^2_\phi I_m))) - \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau^2_\phi I_m)^{-1}(\phi^* - \mu_\phi) - \frac{(\log(\tau^2_\phi) - m)^2}{2s^2} + \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau^2_\phi I_m)^{-1}(\phi^* - \mu_\phi) + \frac{(\log(\tau^2_\phi) - m)^2}{2s^2} \right]
$$

(4.14)

where $m$ and $s$ are the prior mean and standard deviation repectively.

$\tau^2_\xi$ has a conditional posterior distribution of:

$$
\pi(\tau^2_\xi|\ldots) \propto p(\xi^*|\mu_\xi, \tau^2_\xi, \Psi)p(\tau^2_\xi)
$$

To update $\tau^2_\xi$ we need to evaluate:
\[
\exp \left[ -\frac{1}{2} \log(\det(2\pi(\Psi + \tau_\xi^2 I_m))) - \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) - \frac{(\log(\tau_\xi^2) - m)^2}{2s^2} + \frac{1}{2} \log(\det(2\pi(\Psi + \tau_\xi^2 I_m))) + \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) + \frac{(\log(\tau_\xi^2) - m)^2}{2s^2} \right]
\]

(4.15)

where \(m\) and \(s\) are the prior mean and standard deviation respectively.
Chapter 5

Conclusions and further work

This final chapter provides a summary of the research presented in Chapters 2-4, and a reminder of the main results and conclusions therein. I also remind the reader of the contributions made to the literature by this thesis. I then describe possible future work which could be undertaken in order to build on and extend this research.

5.1 Summary

This thesis concerned itself with the following three broad topics:

- the analysis and evaluation of dynamically downscaled climate model output;
- the use of Extreme Value Theory (EVT) in a Bayesian spatial hierarchical framework to model extremes over an irregular grid;
- and the extension of a Bayesian spatial hierarchical EVT framework using predictive processes for dimension reduction of a spatially dense dataset.

5.1.1 Climate model data

In Chapter 2, the results from a high-resolution, multi-model analysis of Irish temperatures for the mid-21st century were presented. First, a preliminary investigation of global climate model (GCM) and high-resolution regional climate model (RCM) data
showed that the latter exhibited greater variability over Ireland by reducing the dominance of the surrounding seas on the climate signal. This motivated and justified the subsequent dynamical downscaling and analysis of the temperature output from three high-resolution (4-7 km grid size) RCMs over Ireland. The three RCMs were driven by four GCMs from two generations of the Coupled Model Intercomparison Project, CMIP3 and CMIP5. These were run under different Special Report on Emissions Scenarios (SRESs) and Representative Concentration Pathway (RCP) future scenarios. In order to ease comparison, the model data were grouped together by SRES or RCP, resulting in five groups. Three groups contained output from models run using SRESs: A1B, B1, and A2; the remaining two groups contained output from models run using RCPs; RCP 4.5 and RCP 8.5. Following this grouping, projections of mean and extreme temperature changes were considered for the mid-century (2041-2060) and assessed relative to the control period of 1981-2000. Some of the main findings for projected mean temperature changes were:

- The RCM data showed that annual mean temperatures were projected to rise between 0.4°C and 1.8°C above control levels by the mid-century.

- On a seasonal basis, projected mean temperature changes differed by forcing scenario. Future summers had the largest projected warming under RCP 8.5, with the greatest warming projected for the southeast of Ireland (2-2.5°C). The remaining two high-emission scenarios (SRESs A1B and A2) projected future winters to have the greatest warming, with this warming (1.5-2°C) relatively uniformly spread across the island.

Following this investigation into projected mean temperature changes, we then decided to focus our attention on changes in the tails of the temperature distribution. Changes here would have more of an impact on society and nature than changes in the mean temperature. We examined changes in the 5th percentile value of daily minimum temperature (the value below which only 5% of days fall) and changes in the 95th percentile of daily maximum temperature (the value above which only 5% of days reach). These typically refer to the number of cold nights and the number of hot days. The main findings were that:
5.1 Summary

- In general, there was greater projected warming of the two percentile values considered than for the mean seasonal temperature changes.

- The greatest change in daily minimum temperature was projected for future winters (indicating fewer cold nights and frost days), a pattern that was consistent across all groups.

- Across almost the entire island, the daily maximum temperature for future summers under RCP 8.5 was projected to increase by more than 2°C - for most of the south of the country, this increase exceeded 2.5°C.

In addition to this analysis of the tails of the distribution, we also investigated seasonal changes in the overall distribution of the temperature variable. We calculated a measure of similarity between the past and future seasonal distributions from each group’s respective control and future empirical density. For RCP 8.5, in addition to mean increases projected for each season, the area of overlap calculated showed that future summers were projected to change the most (area of overlap: 71%) followed by future winters (area of overlap: 81%).

This investigation of projected changes in temperature extremes employed a common method of using percentiles to define extreme climate indexes: some threshold from a control period is used as a cut-off point to define where extreme observations start; and this threshold is then used as a benchmark to see if it is projected to change under future climate simulations. This is an informative and useful approach, and can be easily compared with studies in other domains which use the same index. But this approach has limitations. One can construct examples of two distributions where the behaviour of the extremes differ radically (e.g., a very long tail vs. a short fat tail), but where the 5th and 95th percentile values, for example, are identical. One step towards countering this is to consider a range of indexes: 1st percentile, 5th percentile, 10th percentile etc. But this introduces additional complexity in interpretation if different percentiles are projected to change by different amounts. The realisation that a more comprehensive statistical framework was needed to effectively model climate extremes led to our next study, which was presented in Chapter 3.
5.1 Summary

5.1.2 Waves dataset

I spent the first part of Chapter 3 outlining the theory and terminology needed in order to understand and construct a Bayesian hierarchical model, and I then explained how to fit it using a Markov chain Monte Carlo (MCMC) algorithm. I also outlined the principles of Extreme Value Theory (EVT), the branch of statistics dealing with the description and inference of extreme values from a distribution, in addition to describing Gaussian processes, a flexible framework for modelling spatial data. Once this theory was explained and the methodology was established, I then described how we used a Bayesian hierarchical framework in order to model extreme sea states. This model incorporated a latent spatial process to more effectively capture the variation in the extremes. The model was applied to a 34-year hindcast of significant wave height off the west coast of Ireland. The generalised Pareto distribution (GPD) was fitted to declustered peaks over a threshold given by the 99.8th percentile of the data. Posterior parameter estimates were used to calculate quantities of interest; namely, return levels of significant wave height. These were computed and compared against those from a point-based maximum likelihood model. The main findings were that:

- The Bayesian spatial model produced smoother surfaces of posterior parameters, which led to smoother maps of return levels. Furthermore, this approach greatly reduced the uncertainty in the estimates, thus providing information on extremes which is more useful for practical applications.

- In general, a positive shape parameter was found to best model the data. But whereas some previous studies had fixed the value of the shape \textit{a priori}, we had allowed the model the flexibility to determine the best values of the parameters, given the data.

- We found that the highest extremes of significant wave height were to be expected roughly between 53°N and 54.5°N, with 100-year levels close to 17 m. This value is similar to that suggested in some previous studies (\textit{cf.}, Gallagher (2014)), though in our case there was considerably lower uncertainty in the estimates.

- Site-specific comparisons of return level curves for both methods confirmed...
this reduction in uncertainty associated with the Bayesian spatial model, while empirical estimates overlaid on the plots showed a satisfactory fit to the data at these sites.

Results from this study were promising: using a Bayesian spatial hierarchical model, we had successfully built and improved on the previous research in the area. This approach had computational limitations though: fitting Bayesian hierarchical models using MCMC methods involves matrix factorisations of the order of $n^3$, where $n$ is the number of locations (Guhaniyogi et al., 2011). This becomes computationally infeasible as $n$ gets large. Limiting the number of gridpoints $n$ by selecting only a subset of the available data would mean throwing away valuable information - this would be a particularly poor decision when it comes to modelling extremes. We needed an approach which manages to use all of the available data, but which is computationally feasible at the same time. The realisation that more sophisticated statistical tools were needed to overcome this computational bottleneck, while still using all of the available data, led to our next study, which was presented in Chapter 4.

### 5.1.3 Gridded temperature dataset

In Chapter 4, I presented our work in beginning to characterise temperature extremes in Ireland for the period 1981-2010. I described the theory behind predictive processes, a dimension-reduction technique which can be applied to dense spatial fields. The dataset we chose to model here consisted of anomalies of daily maximum temperature on a dense regular grid ($\approx 1,700$ gridpoints) over Dublin. I explained that by modelling temperature anomalies, we were choosing to focus our research on maximum temperatures which would be considered extreme relative to the time of year in which they occurred. We then employed EVT to model the data of declustered anomalies above the 99.5th percentile threshold using the GPD as part of two three-level Bayesian hierarchical models, similar to the methodology from Chapter 3. But in addition to this, we incorporated in this framework the dimension-reduction technique of predictive processes in order to solve the computationally difficult problem of modelling data over a very dense spatial field. To our knowledge, this was the first study to combine predictive processes and EVT in this manner. The model was fit using the Metropolis-Hastings MCMC algorithm. Posterior parameter estimates and
5.1 Summary

return level surfaces were produced, in addition to specific site analysis at synoptic stations across the domain. The main results were:

- The 20-year median return level surface of temperature anomalies was seen to range from just below 8°C to almost 10°C across the domain. For the corresponding 100-year median surface, return levels ranged from 8°C to just under 10.7°C. Lowest values appeared along the sea in the north-east of the domain, with the highest values on the western boundary. Applying these 100-year median values to Casement Aerodrome (in the west of the domain) where the climatological mean just exceeds 20°C in July, if these temperature anomalies occurred at this time of year, daily maximum temperatures would be in excess of 30.5°C.

- Observational data from synoptic stations for the period 2011-2018 was included in site-specific analyses to determine if there was any evidence of a change in the observed extremes over the data from the control period. We noted here an increase in the frequency of extreme anomalies, but not in their severity.

- We found that the frequency of observed extreme anomalies from 2011-2018 at the Casement Aerodrome and Phoenix Park synoptic stations exceeded the upper bounds of their credible intervals predicted from the model by 20% and 7% respectively.

Results from this study were encouraging. We had begun a characterisation of climate extremes in Ireland. We had successfully combined predictive processes and EVT in a Bayesian spatial hierarchical model in order to produce quantities of interest; namely, return level estimates of temperature anomalies. This is the first time that this has been done, to our knowledge. In the following sections, I remind the reader of the contributions this thesis makes to the literature (Section 5.2) and I then discuss some possible future research which could extend and build on the studies presented here (Section 5.3).
5.2 Overview of contributions

This thesis makes a number of original contributions to the existing literature. Specifically:

- In Chapter 2, I detailed the first study (at such a high resolution) of dynamically downscaled climate models over Ireland. I presented analysis of temperature projections for the mid-21st century. I then analysed changes in the means and the extreme temperatures, using simple statistical techniques to investigate the projected changes of the daily mean, minimum, and maximum temperature across Ireland.

- In Chapter 3, I incorporated EVT into a Bayesian spatial hierarchical framework, in order to model extreme waves off the west coast of Ireland. A Bayesian approach allowed prior belief and physical constraints to be incorporated into our model in a natural way, which led to reduced uncertainties in parameter estimates. Our work here improved on the previous non-spatial research in this area, by borrowing information from nearby locations to model the latent spatial process underlying the extreme behaviour of waves.

- In Chapter 4, I described how we overcame the computational difficulties in applying a Bayesian spatial hierarchical EVT model to data on a very dense grid, using the intuitively appealing dimension-reduction technique of predictive processes to model daily maximum temperatures over County Dublin, Ireland. Use of this technique made results possible within the computational constraints imposed on the research, while still allowing for inference and posterior prediction at any point across the spatial domain.

5.3 Future research and challenges

There are a number of possible extensions to the methods and studies presented in this thesis. I conclude this chapter by listing some ideas that could be pursued in future work.
5.3 Future research and challenges

5.3.1 Other variables

An obvious extension to the methodology presented in Chapters 3 and 4 is to apply these methods to other climate variables. Whether predictive processes will be needed or not would depend on the number of gridpoints across the spatial domain and the computational resources available. Applying this methodology to a dataset of minimum temperature, precipitation, wind speed etc., would all prove to be interesting studies and continue the task of characterising and mapping climate extremes in the domain of the data, whether that’s across Ireland or elsewhere.

Of course, there is nothing in the methodology which is specific to climate variables. It can be applied to any continuous spatially-referenced dataset about which inference is required. This could be environmental data such as the concentration of some pollutant at sites across a domain and over time; it could be adapted to model diffusion, where matter moves from areas with high concentration to areas with less concentration, such as salt concentration moving through water.

5.3.2 Larger domain

Continuing the idea of characterising climate extremes in Ireland, another obvious extension is to extend the domains from the studies in Chapters 3 and 4. This ‘scaling-up’ process would use more of the data on the original domains of the datasets: \(\approx 15,000\) gridpoints for the waves dataset and \(\approx 72,000\) gridpoints for the gridded temperature dataset. This would extend the studies in this thesis to cover the seas all around the Irish coast (waves dataset) or cover the entire state of Ireland (gridded temperature dataset). Such studies would undoubtedly need dimension-reduction techniques, whether that is through the use of predictive processes or an alternative approach (see Section 5.3.5 for a discussion of some of these). Investigations into the optimal placement of the subgrid should be conducted too, to ensure that results are robust, and not overly sensitive to the choice of subgrid. This is more important for the waves dataset, which is on a quite irregularly spaced grid, than it is for the gridded temperature dataset, which is already on a regular grid. A trade-off between the number of gridpoints and the computational resources available would need to be considered and justified for any study too, particularly if multiple model fittings
were needed for model averaging or to investigate the effects of differing numbers of gridpoints and different subgrid structures.

5.3.3 Longer temporal component

As well as increasing the spatial domain of the study, the temporal domain could also be increased in future work. In Chapter 4, we worked with a dataset of daily maximum temperatures from 1981-2010. The 30-year length of the dataset was convenient, as it was exactly what is recommended by the World Meteorological Organization in order to define a climate normal (World Meteorological Organization, 2017). If a longer period of data were used in a similar study, the methodology could be extended to build a temporal component into the hierarchical model.

One possible approach to include the temporal component of an extreme observation could be to modify the mean response vector $\mu$ of a parameter surface $\theta$. Adapting the notation from Chapter 3: “the effect of any other covariates may be readily incorporated into the model. For $m$ covariates $c^{(1)}, \ldots, c^{(m)}$, we write the mean vector in the general form

$$\mu_{\theta} = C\alpha_{\theta}$$

where $C$ is the $n \times (m + 1)$ matrix

$$C = \begin{pmatrix}
1 & c^{(1)}_1 & c^{(2)}_1 & \ldots & c^{(m)}_1 \\
1 & c^{(1)}_2 & c^{(2)}_2 & \ldots & c^{(m)}_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & c^{(1)}_n & c^{(2)}_n & \ldots & c^{(m)}_n
\end{pmatrix}$$

and the vector of coefficients is $\alpha_{\theta} = (\alpha_{\theta,0}, \alpha_{\theta,1}, \ldots, \alpha_{\theta,m})^T$.”

Incorporating time into this mean structure could involve an idea borrowed from time series analysis: construct an autoregressive model for the mean vector of each parameter where $\mu$ is a weighted sum both of covariates and its previous value. With the notation as above, this could mean appending a final column to $C$ which includes the value of $\mu$ from the previous iteration of the algorithm. Or writing these separately
5.3 Future research and challenges

for clarity:

\[ \mu_{\theta_i} = C\alpha + \beta\mu_{\theta_{i-1}}, \]

so that the value of the mean vector \( \mu \) of parameter \( \theta \) at iteration \( i \) is given by a weighted sum of the linear regression components and the value of \( \mu \) from the previous iteration of the MCMC algorithm.

A more complex way to integrate a temporal component into the mean vector would be to let the value of \( \mu \) depend on both space and time: \( \mu_{\theta} \equiv \mu_{\theta}(s, t) \). This could be in the form of a linear component, e.g.:

\[ \mu_{\theta}(s, t) = C\alpha + \beta t \tag{5.1} \]

where \( t \) could be a year index, a decadal index, or some other time period. \( \mu \) could then be stored as a matrix, with the rows indicating the gridpoint, and the columns indicating the time index. (Equation (5.1) could be more complex, involving transformations of the temporal index, or a functional form for \( \beta \).) This approach would lead to a more complicated expression and evaluation of the GPD likelihood, for example, as it would now need to involve further product terms. In Chapter 3, the GPD likelihood had the following form:

\[
\begin{align*}
p\left(\text{data} | \text{parameters}\right) &= \prod_{s=1}^{S} \prod_{k=1}^{n_s} \frac{1}{\exp \phi(s)} \left(1 + \frac{\xi(s)z_k(s)}{\exp \phi(s)}\right)^{-1-1/\xi(s)} \\
\end{align*}
\]

where the product terms run over all gridpoints \( s = 1 \ldots S \) and also over all observations at each gridpoint \( k = 1 \ldots n_s \). Using equation (5.1) to introduce a temporal component to the parameters \( \phi \) and \( \xi \) would introduce a further product term:

\[
\begin{align*}
p\left(\text{data} | \text{parameters}\right) &= \prod_{t=1}^{T} \prod_{s=1}^{S} \prod_{k=1}^{n_s} \frac{1}{\exp \phi(s, t)} \left(1 + \frac{\xi(s, t)z_k(s, t)}{\exp \phi(s, t)}\right)^{-1-1/\xi(s, t)} \\
\end{align*}
\]

In addition to the extra computational expense of this approach, interpretation and presentation of results becomes more challenging too. If calculating return lev-
els, instead of a single median value of $\phi$ and $\xi$ at a particular gridpoint now, there will be a median value associated with each time index $t$. Return level surfaces calculated would therefore depend on the year/decade/other time period used to index $t$.

The distance between observations both in space and in time can also be accounted for using the covariance structure (as opposed to, or in addition to, using the mean response). As an initial step, this could be done by specifying a separable spatio-temporal covariance function. This is a function for which there exists purely spatial and purely temporal covariance functions such that:

$$C(s, t) = C_S(s) \times C_T(t)$$

for all $(s, t) \in \mathbb{R}^d \times \mathbb{R}$. That is, the space-time covariance function decomposes as the product of a purely spatial and purely temporal covariance function (Gneiting and Guttorp, 2010). The covariance matrix can then be represented as a Kronecker product of a purely spatial and a purely temporal covariance matrix (Genton, 2007). However, Gneiting and Guttorp (2010) warn that, while the assumption of separability reduces the number of parameters and the computational burden for large spatio-temporal datasets, such separable models do not allow for space-time interaction and frequently fail to model a physical process accurately. The alternative to this assumption is that the covariance function is non-separable. There are many such possible formulations - these include the product-sum model, where the covariance is split into a weighted sum of the product of a spatial and temporal covariance function, a spatial covariance function on its own, and a temporal covariance function on its own (De Iaco et al., 2001).

### 5.3.4 Climate model output

A previous study on which I was co-author looked at projected changes in precipitation for the mid-century using climate model output (in a similar approach to Chapter 2). As with the temperature study, analysis on the projected changes in the distribution of precipitation was conducted by assessing how similar the model past and future distributions were (Nolan et al., 2017). Both of these studies would benefit from being revisited with the theoretical framework of EVT for a more robust analysis of
model past vs. model future.

Possible approaches to doing this could involve using larger-scale climate model output variables as regression covariates for the GPD parameters (an idea used in a statistical downscaling of GCM data by Fealy and Sweeney (2008)). These larger-scale variables would allow the parameter of the distribution to vary with a dominant feature of climate, such as the globally-averaged temperature or Northern Hemisphere-averaged temperature, for example. Each GCM-RCM pairing could also feature in this analysis, with a model averaging being performed over the climate models themselves. In this situation, the mean vector $\mu$ for the parameter $\theta$ would be obtained by a weighted average over its estimate from all GCM-RCM pairings $(i,j)$ where:

$$
\mu_{\theta}(s,t) = \sum_{i,j} \beta_{ij} \mu_{\theta_{ij}}(s,t)
$$

where $\beta_{ij}$ is the weight on each GCM-RCM pairing $(i,j)$ and $\mu_{\theta_{ij}}(s,t)$ is the parameter estimate from that model pairing. Weights could be determined from the historical period by assessing the performance of each model pairing $(i,j)$ by comparison with the past observational record - in this way, pairings which perform the best at modelling extremes in the historical period have more weight on the parameter estimates for the future period.

### 5.3.5 Alternative computational approaches

As already noted in Chapters 3 and 4, there are alternative computational approaches to fitting Gaussian processes (GPs) using MCMC such as the integrated nested Laplace approximation (INLA) approach proposed by Rue et al. (2009), and INLA combined with a stochastic partial differential equation approach (INLA-SPDE) proposed by Lindgren et al. (2011). In our work, we chose to fit GPs using an MCMC algorithm due to the ease with which they fit into a Bayesian hierarchical framework. They are flexible empirical models, which were appropriate for an irregularly fluctuating and real-valued spatial surface (Diggle et al., 2007), as we had here. Following the approximation of a spatially continuous latent field by a Gaussian Markov random field on a discrete lattice, Taylor and Diggle (2014) compare INLA performance to
5.3 Future research and challenges

an MCMC algorithm. In this instance, the authors found that the MCMC approach yielded more accurate estimates of predictive probabilities, and with less bias, than the INLA approach.

However, INLA uses a combination of analytical approximations and numerical integration schemes to quickly find accurate deterministic approximations to posterior quantities of interest (Martino and Riebler, 2019), and any further research should investigate its performance across the large spatial (and possibly spatio-temporal) domains from Chapters 3 and 4. Performance should be considered both in terms of the computational expense saved (over an MCMC approach) and other chosen performance metrics for the problem in question (e.g., variability of posterior estimates, bias). Both approaches could also be combined, in a process where INLA is imbedded in a larger MCMC algorithm; in this situation, the posterior marginals of a collection of parameters (those we condition on) can be obtained via MCMC algorithms, whereas the posterior marginals of all the other parameters are estimated by averaging over several conditional marginal distributions obtained by INLA within the steps of the MCMC algorithm (Gómez-Rubio and Rue, 2018).

To conclude this section, there are many possible directions for future work based on the research I have presented here. This thesis is only the beginning. Future work may range from applying identical models to new climate variables, to new fields of study, or to larger domains; to building more complexity into the models to incorporate time-varying parameters in the mean or covariance structure; to using GCMs and RCMs as covariates in fitting the models, or averaging over output from GCM-RCM pairs; to investigating the performance of other algorithms such as INLA, INLA-SPDE, or INLA embedded in MCMC.
References


