Statistical Models for Extreme Weather Events

Paul Sharkey, B.Sc.(Hons.), M.Res.

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

January 2018

Abstract

Western Europe is typically prone to extreme weather events during the winter months, which typically take the form of windstorms or flooding. The storm Desmond brought strong winds and heavy rain to Ireland, northern England and Scotland in December 2015, resulting in an estimated £500 million worth of damage and extensive flooding, particularly in the region of Cumbria. Accurate modelling of such extreme weather events is necessary to ensure that the societal and infrastructural risk associated with these phenomena is minimised.

In statistical modelling, extreme value analysis is typically used to model the rate and size of extreme weather events. Typically, practitioners can use the outputs of such an analysis to design flood defences to a standard such that there is only a small probability that defences are breached in a given year. These models can be applied at individual sites or adapted to address questions related to the spatial extent of an event, which is important for policy makers eager to reduce the economic and societal impacts associated with extreme weather.

One aim of this thesis is to improve inference with regarding to existing extreme value methodology. First, we propose a reparameterisation of the likelihood corresponding to the Poisson process model for excesses above a high threshold, which improves mixing in a Bayesian framework and ensures more rapid convergence of the parameter chains in a Markov Chain Monte Carlo routine. The Poisson process model is often preferred for modelling extremes of non-stationary processes as the parameters are invariant to the choice of threshold; our approach may increase the possibility of this model being used more widely. Second, we propose an adjustment to the likelihood when implementing a spatial hierarchical model for extremes, which accounts for the dependence in the data when estimating model uncertainty. In both cases, the improvement in inference should increase confidence among practitioners of the outputs obtained from extreme value models.

The main influence of extreme weather events in winter is from the passage of low-pressure extratropical cyclones from the North Atlantic. The second aim of this thesis is to quantify the risk associated with extreme wind speed events, which we call windstorms, arising from an extratropical cyclone system. First, we develop a model capturing the spatial variation of the track associated with the cyclone, from which we can simulate synthetic tracks with the same statistical characteristics of the observed record. Second, we describe an approach for modelling the spatial extent and severity of windstorms relative to the storm track, from which we can provide improved estimates of risk associated with windstorms at individual sites and jointly over a spatial domain. The methods described in this thesis can be used to address multiple questions related to windstorm risk, that is not available using current methodology.

Acknowledgements

First of all, I would like to thank my supervisor, Jonathan Tawn, for his expert guidance over the last three years. Jon's unique blend of knowledge, wisdom, kindness and humour meant he was an absolute pleasure to work with, and I sincerely hope we cross paths again in the near future.

I would also like to acknowledge the support of the Met Office and EDF Energy, and in particular my industrial supervisors Simon Brown and Hugo Winter. I am thankful to Simon for his vast meteorological expertise and his patience as I slowly adjusted to new concepts, as well as for providing me with the benefits of his time on my visits to the Met Office. I am grateful to Hugo for his many contributions to the project, but particularly the sage advice that came from his recent experience as a PhD student. I am also thankful to Pietro Bernardara for welcoming me to EDF Energy initially and I wish him all the best in his new role.

I am very grateful for the support of the students and staff of the STOR-i Centre of Doctoral Training for providing such a stimulating and enjoyable research environment. I am particularly grateful to the Extremes Coffee group for the laughs, the joint paper, the heated punctuation arguments and the baked goods. I must also gladly acknowledge the financial support of the EPSRC.

I would like to thank my parents and brother for being a constant source of encouragement and support during my time in Lancaster. I would also like to thank Zak for celebrating with me when things were going well and suffering me when things were going badly, which I look forward to reciprocating.

Sláinte.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree. The word count for this thesis is 52,581 words.

This thesis is constructed as a series of papers and thus Chapters 3, 4, 5 and 6 should be read as separate entities. As such, background information on existing methodology may be repeated in multiple chapters.

Chapter 3 has been published as Sharkey, P. and Tawn, J. A. (2017). A Poisson process reparameterisation for Bayesian inference for extremes. *Extremes*, 20(2):239-263.

Chapter 4 has been submitted for publication as Sharkey, P. and Winter, H. C. (2017). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain.

Chapter 5 has been submitted for publication as Sharkey, P., Tawn, J. A. and Brown, S. J. (2017). A stochastic model for the lifecycle and track of extreme extratropical cyclones in the North Atlantic.

Chapter 6 has been submitted for publication as Sharkey, P., Tawn, J. A. and Brown, S. J. (2018). Modelling the spatial extent and severity of extreme European windstorms.

Paul Sharkey

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In light of a number of devastating weather events in recent years, the importance of accurate modelling, forecasting and prediction of such events is greater than ever. Hurricane Katrina decimated the city of New Orleans in August 2005, originating as a tropical depression in the Caribbean and intensifying over the Gulf of Mexico. Storm Desmond caused an estimated £500 million worth of damage to northern England and Scotland in December 2015, as an atmospheric river of moist air resulted in unprecedented levels of rainfall and consequently led to widespread flooding. More recently, Hurricane Irma caused mass damage to the Caribbean islands in September 2017, with the strongest sustained winds observed in the United States in over a decade.

These large-scale impacts are directly linked to well-defined meteorological events, in these cases, hurricanes and extratropical cyclones, that are associated with strong winds and large quantities of precipitation. These natural hazards are strongly linked to increased water depths in rivers and coastal areas, which can result in a heightened risk of flooding. Flood defences are often in place to protect large settlements; however in many cases the prevalence and durability of these are subject to cost constraints. It is therefore imperative that defensive infrastructure is designed optimally to give maximum protection against natural hazards. Models are used to inform policy by providing guidelines related to the vulnerability of a location with regard to extreme weather events, as well as information from which design codes can be derived such that the risk of failure of the flood defence is minimised. In particular, hydrological models can be used to quantify the risk of flooding given the local weather conditions and properties of a catchment. Statistical models are also used to answer questions related to the occurrence of strong winds and heavy rain; for example, the frequency at which can we expect to observe a certain amount of rain, the location at which we can expect the most amount of rain, and the locations that are likely to experience large levels of rainfall given that a particular site has experienced a certain level.

These questions are inextricably linked to the characteristics of the meteorological event itself. In western Europe, extratropical cyclones are widely regarded as the meteorological drivers behind the strong winds and heavy rain that are characteristic of winter weather. These systems are associated with low pressure and develop as a result of horizontal temperature gradients. Convergence of hot and cold airmasses create weather fronts, which can intensify winds and result in large amounts of precipitation. Knowledge regarding the structure of a cyclone should be used where possible to inform statistical models for extreme weather events, which can consequently be used to produce more realistic representations of these events, for example, relating to their spatial extent and heterogeneity. The location corresponding to the meteorological ‘centre’ of a storm has substantially different weather characteristics compared to a location on the edge of the region of influence. This motivates the need to incorporate such physical information in our modelling procedure.

Modelling extreme events represents an additional challenge as by definition, such events do not often occur within the observational record. Models are therefore required to allow extrapolation to very extreme levels in order to estimate the probability of observing events beyond the range of the data. Much of the statistical literature linked with this approach to extrapolation is based on methods from extreme value analysis, where mathematically rigorous models have been developed aimed at estimating the tail behaviour of a process. Consequently, using data from a time period significantly shorter than 200 years, one can estimate the rainfall level one would expect to see once every 200 years, for example, which can be incorporated into design of infrastructure for defence against natural hazards. These models are well-established and are widely used by practitioners. However, some issues remain with regard to inference; this thesis will address current problems relating to parameter estimation and quantification of uncertainty.

Physical information can be incorporated into extreme value models through regression of the model parameters on covariates, and thus one can estimate a high quantile conditional on some physical characteristic of the process being modelled. For example, extreme wave heights are known to vary according to what direction the waves are coming from; this variation can especially arise if there are land shadow effects on the site being studied. However, in the context of an extratropical cyclone, it is not trivial to construct a covariate that summarises the effects of, say, bearing with respect to the storm centre, lifecycle of the storm or the spatial extent of an event. As such, existing statistical models insufficiently capture the spatial and temporal characteristics of the cyclone that generate the weather extremes of interest. This represents an interesting modelling challenge that requires an alternative approach motivated by the atmospheric properties of the meteorological event.

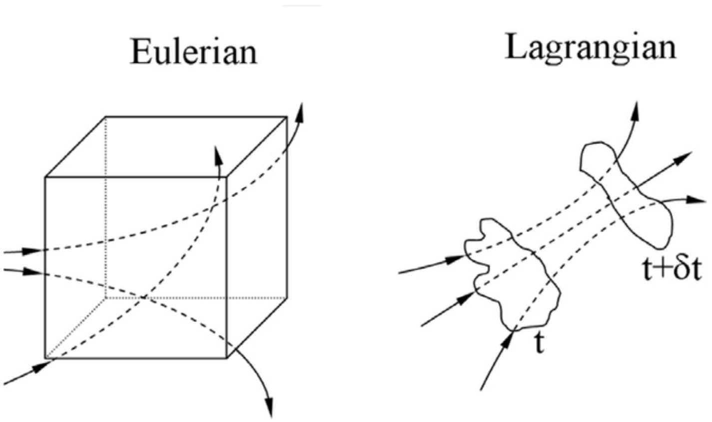


Figure 1.1.1: From Shadloo et al. (2016): a visual representation of Eulerian and Lagrangian frames of reference. In the Eulerian specification (left), measurements are recorded in a fixed location as the flow passed through. In a Lagrangian specification (right), measurements are collected along the path determined by the flow.

Existing statistical models are typically based on analysis of fixed locations and observing the process at each location as time passes. In fluid mechanics, this is commonly referred to as an *Eulerian* specification of a process. An example of this would be a person sitting on a bank of a river and watching the water flow past the fixed location. Alternatively, one can specify a *Lagrangian* specification of a process, whereby the observer follows the process as it moves through space and time (see Figure 1.1.1). This would correspond to a person sitting on a boat and drifting down the river. In oceanography, Eulerian data is typically gathered using fixed sensors at locations in the ocean, where observations are collected at regular intervals as the flow passes by. This is advantageous in the sense that long time series can be collected, but it may not perform well at addressing questions related to the pathway of the flow. Oceanographers alternatively collect Lagrangian measurements through drifters, which are used to track the transport pathways.

A Lagrangian approach to modelling extratropical cyclones seems more intuitive than an Eulerian approach, as the extremes of interest are generated relative to the path taken by the storm in space and time. An Eulerian approach to modelling was motivated by the sparse availability of observational data, e.g., from weather gauges. However, recent advances in climate modelling have resulted in the availability of high-resolution datasets that are spatially and temporally complete, from which large scale systems like extratropical cyclones can be identified and tracked. By fixing the centre of the storm as our frame of reference, we can address questions about the influence of the path and the intensity of a storm on the occurrence of extreme weather events. In particular, this approach provides a natural mechanism for exploring the likely position, spatial extent and magnitude of extreme weather events relative to the path of the storm. Because the locations affected by a storm are largely determined by its path, an Eulerian approach can only model the observed events relative to these fixed locations. However, we can improve inference by modelling the path of the storm and pooling over all events defined relative to the path. This allows us to generate events relative to synthetic paths, allowing us to estimate probabilities of events not yet observed at certain locations.

As well as being concerned with questions regarding marginal risk at a given location, practitioners are often interested in quantifying the joint risk of an extreme event occurring at multiple locations; for example, what is the probability of observing an extreme windstorm event at site B given that one occurs at site A? This is of particular interest due to the potential devastating impacts that can arise from an event whose spatial extent is especially large. In the extreme value literature (Coles et al., 1999), the quantity χ is defined such that

$$\chi = \lim_{x \to x^F} \Pr\{X_B > x \mid X_A > x\},$$

where X_A and X_B denote the wind speed at sites A and B respectively (transformed onto common margins) and x^F is the upper endpoint of the distribution. This expression is the limiting probability that X_B is extreme given that X_A is extreme. A value of $\chi > 0$ corresponds to the case of asymptotic dependence, where extremes at sites A and B can occur simultaneously, while a value of $\chi = 0$ refers to the case of asymptotic independence, where extreme events at sites A and B tend to occur at different times. Distinguishing the two cases is highly important when modelling extremal dependence, as the application of asymptotically dependent models to data exhibiting properties of asymptotic independence can lead to conservative estimates of extreme joint events. However, diagnostics used to estimate χ are often highly uncertain due to the sparsity of data near the endpoint of the distribution. By simulating a large number of synthetic windstorms, however, we can estimate χ with a greater degree of confidence. This also allows us to estimate the probability of observing extreme joint events beyond the range of the data. By modelling the processes driving the spatial extent and magnitude of windstorm events, our estimation of χ is derived from physically realistic scenarios matching the statistical characteristics of the observed record.

1.2 Thesis outline

The aims of this thesis are twofold. First, we aim to improve inference for existing extreme value methodology in the form of parameter estimation and correct quantification of uncertainty. Second, we aim to develop a statistical model for extreme windstorms arising from extratropical cyclones that is a physically consistent representation of the generating process. The thesis is split into chapters, each outlining the different approaches taken to achieve these aims.

In Chapter 2, we introduce models from univariate extreme value theory which are used as the basis of statistical methodology in this thesis. In particular we describe theory and inference for the block maxima approach as well as two approaches for modelling excesses above a high threshold. We also introduce the concept of an extratropical cyclone and describe its meteorological and climatological properties, as well as previous attempts to model these weather systems in a statistical context.

In Chapter 3, we propose a reparameterisation of the Poisson process model for excesses above a high threshold to improve estimation and inference in the Bayesian framework. This reparameterisation is based on the optimal selection of a tuning parameter in the likelihood which near-orthogonalises the model parameters, enabling better mixing in the Markov Chain Monte Carlo routine. We consider the cases of independently and identically distributed random variables along with a linear trend in the location parameter. The approach is applied to simulated data and a case study is presented on rainfall data in Cumbria, UK.

Chapter 4 outlines a spatial model in the Bayesian hierarchical framework, which borrows strength across locations in the estimation of marginal return levels. This class of models assumes conditional independence across space and time, which often results in underestimation of model uncertainty. We propose an adjustment to the likelihood, which has been used in previous studies on max-stable processes, to account for the spatial and temporal dependence in the data when making the misspecification of conditional independence. This approach is applied to precipitation data in Great Britain.

In Chapter 5, we present a model for the lifecycle and track of extreme extratropical cyclones in the North Atlantic. Specifically, we model the path of local vorticity maxima characterising the storm, referred to as a track. This model comprises submodels for the beginning, propagation and termination of a track over space and time, that preserve the dependence structure between movement and intensity and the spatial variation inherent in these characteristics. Through Monte Carlo simulation of a large number of synthetic tracks, we estimate extremal quantities such as the return period corresponding to storm Herta, a particularly damaging storm that occurred in 1990, as well as the $100-$ and $1,000$-year return levels for vorticity over space.

Chapter 6 builds on the work in Chapter 5 by constructing a model for extreme European windstorms. We approach this from a Lagrangian viewpoint, extracting wind events relative to the storm centre using a combination of spatial filtering and clustering methods. We use information from the storm track to determine the evolution of event characteristics such as its position, size and magnitude, using empirical observations to inform the statistical modelling. We simulate a large number of synthetic windstorm events, from which we can estimate the risk associated an extreme event occurring at a number of locations simultaneously.

In Chapter 7, we conclude with a summary of the contributions made by this thesis as well as some opportunities for further work.

Chapter 2

Literature Review

2.1 Extreme value methods

2.1.1 Overview

Extreme value theory is a rapidly emerging field in modern statistics, with practical significance in disciplines including finance, hydrology and medicine. The importance of analysing and predicting extreme events creates a necessity for a statistically rigorous model for the tail of a distribution of a random variable. Standard statistical modelling approaches are often biased when estimating tail behaviour, as central values tend to drive parameter estimates and model fit. This creates the need for a tail model that is not compromised by these issues. Further, one might be interested in the level corresponding to an event expected once every 200 years; in the context of flood frequency analysis, this may influence the design of defensive infrastructure. However, by definition, observations in the tails are rare, and so it is often required that inference for unobserved scenarios is made using observed data which are not as extreme as the events that must be predicted. Extreme value analysis allows this by using asymptotic models fitted only to tail data to extrapolate beyond the range of the data to unobserved levels.

This section provides an overview of theory and inference in extreme value analysis. In particular, we describe two approaches for analysing univariate extreme values that are unified by asymptotic theory. In Section 2.1.2, we describe the block maxima approach. We introduce some threshold-based methods in Section 2.1.3.

2.1.2 Block maxima approach

Consider a sequence of independently and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$ with an unknown distribution function F. Define $M_{X,n} = \max(X_1, \ldots, X_n)$ to be the maximum of this sequence of random variables. The distribution function of the sample maxima can then be expressed as

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

An analogous result for minima can be obtained by defining

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

This subsection focuses on application of extreme value analysis to sample maxima, noting that results for minima can be obtained using the above identity. Henceforth, M_n and m_n will be used in place of $M_{X,n}$ and $m_{X,n}$ respectively.

The above formula for the distribution of maxima is unhelpful in practice as the distributional form of F is typically unknown. One approach is to search for families of models for which the expression F^n converges for the tails of the distribution of F. However,

$$M_n \to x^F \text{ as } n \to \infty,$$

where

$$x^F = \sup\{x : F(x) < 1\}.$$

In other words, the distribution of M_n degenerates to a point mass on the upper end point of F. A method of overcoming this difficulty is to obtain a linear renormalisation of M_n to give a non-degenerate limit distribution. Let M_n^* be defined as

$$M_n^* = \frac{M_n - b_n}{a_n},$$

for sequences of constants $a_n > 0$ and b_n, which stabilise the location and scale of M_n^* as n increases, avoiding the issues that arise with the distribution of M_n. The Extremal Types Theorem (Fisher and Tippett, 1928) states that given appropriate choices of these normalising constants, as $n \to \infty$

$$\Pr\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x), \tag{2.1.1}$$

where G is non-degenerate and is of the same type as one of the following distributions:

- **Gumbel:** $G(x) = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty;$
- **Fréchet:** $G(x) = \begin{cases} 0 & x \le 0 \\ \exp\{-x^{-\alpha}\} & x > 0, \alpha > 0; \end{cases}$
- **Negative Weibull:** $G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \ge 0. \end{cases}$

The Unified Extremal Types Theorem (UETT) unites these distributions under one parameterisation, the generalised extreme value (GEV) distribution, with distribution function

$$G(x) = \exp\left\{ -\left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, \tag{2.1.2}$$

where $x_+ = \max(x, 0)$ and $\sigma > 0$. The parameters μ, σ and ξ are interpreted as the location, scale and shape parameters respectively. The distribution of M_n^* is of the same type as a GEV distribution as $n \to \infty$, for some value of ξ. A Gumbel distribution corresponds to $\xi = 0$, with the feature of an exponential upper tail. A Fréchet distribution corresponds to $\xi > 0$, with a heavy upper tail. A Negative Weibull distribution, for which $\xi < 0$, has the property of a finite upper end point.

In practical applications, interest lies in the estimation of a probability that extreme events are sufficiently large. The *return period* of level x_p is defined as the inverse of the probability that x_p is exceeded in any one year. The T*-year return level* is defined as the level for which the probability of exceeding this level in any one year is $1/T$. The $1/p$ return level x_p is the $1-p$ quantile of the GEV distribution for $0 < p < 1$. From this definition, distribution (2.1.2) can be rearranged to give as expression for x_p such that

$$x_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-p)\}^{-\xi}] & \text{for } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for } \xi = 0, \end{cases}$$

where (μ, σ, ξ) denote the GEV parameters.

The GEV distibution is typically used to model a series of maxima of a pre-selected number of blocks (e.g. years). Inference for the GEV model is typically carried out using maximum likelihood estimation with respect to parameters (μ, σ, ξ). There exists no analytical solution for the parameter estimates, so numerical maximisation schemes are required. The maximum likelihood estimators are asymptotically Gaussian subject to the conditions that $\xi > -1/2$ (Smith, 1985). A distribution where $\xi < -1/2$

corresponds to very short bounded upper tails and is rarely seen in practice, so these restrictions do not cause practical difficulties. The asymptotic normality of the estimators means that the uncertainty of return level estimates can be evaluated using the delta method. However, this approximation performs poorly when considering return levels corresponding to long return periods that fall beyond the scope of the data. Profile likelihood-based confidence intervals provide a more accurate representation of uncertainty when a strong degree of extrapolation is required (Coles, 2001).

An alternative approach to estimation is through Bayesian inference, which requires the specification of a prior distribution on the parameters. As no conjugate prior exists for the GEV, inference is carried out using Markov chain Monte Carlo (MCMC) methods. Bayesian inference provides flexibility for incorporating expert knowledge of the process being modelled into the analysis through the specification of the prior. Because of the data limitations in extreme value problems, this can substantially improve inference. Bayesian methods also provide a natural way of estimating parameter uncertainty without the theoretical complication of maximum likelihood. For more details, see Coles and Powell (1996) and Stephenson (2016).

2.1.3 Threshold methods

While the block maxima approach is easily used and interpretable, one of its drawbacks is its failure to capture the full behaviour of the tail of a distribution. The model is limited to analysing data selected as the maximum of a pre-selected block, despite the strong possibility of there being other observations in the same block that may be characterised as extreme (see Figure 2.1.1). Threshold methods account for the extra tail information in these observations by analysing data above a pre-determined level u. This leads to a more efficient modelling procedure as a result of the added data used in the analysis.

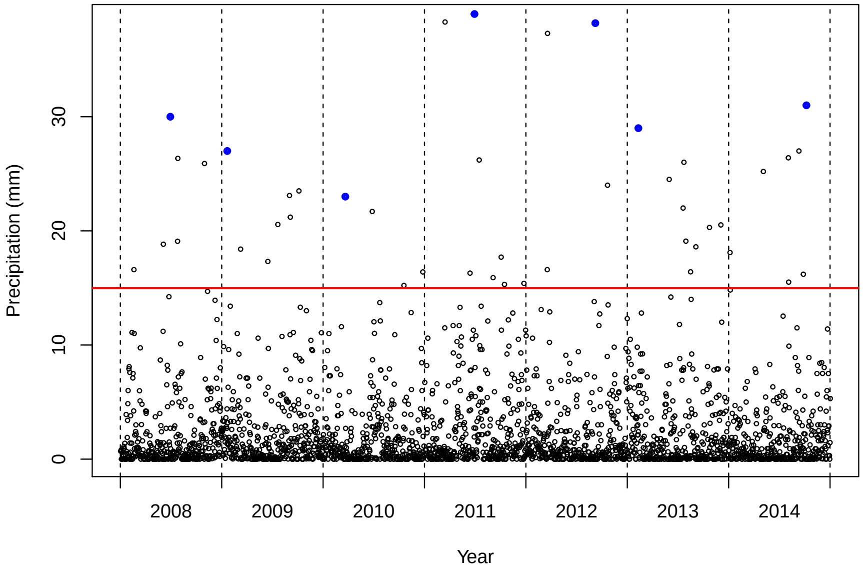


Figure 2.1.1: Rainfall accumulations in a grid cell in southeast England from 2008-2014, taken from reanalysis data used in Chapter 4. The blue observations represent the values used in a block maxima approach, whereas all observations above the red line would be used in a threshold-based model.

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with distribution function F. Assuming the asymptotic theory of equations (2.1.1) and (2.1.2) hold, then a sequence of point processes P_n can be constructed on $[0,1] \times \mathbb{R}$ by

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

and the behaviour as $n \to \infty$ can be examined. The limit process is non-degenerate as the distribution of the normalised maxima is non-degenerate. Large points of the process are retained in the limit process while small points are normalised to the same value b_l, with

$$b_l = \lim_{n \to \infty} \frac{x_F - b_n}{a_n}.$$

Under these conditions on P_n, on the set $[0,1] \times (b_l, \infty)$

$$P_n \to P \text{ as } n \to \infty, \tag{2.1.3}$$

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

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

This result motivates the idea that the behaviour of large values from F are determined asymptotically by the characteristics of a_n, b_n and ξ, as with the block maxima approach. Under the conditions for the limit in (2.1.3) to hold, Pickands (1975) shows that for $x > 0$ and $X \sim F$

$$\Pr(X > u_n + a_n x \mid X > u_n) \to \left(1 + \frac{\xi x}{\psi}\right)_+^{-1/\xi}, \tag{2.1.4}$$

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

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

corresponds to the generalised Pareto distribution (GPD), with scale parameter $\psi > 0$ and shape parameter $\xi \in \mathbb{R}$, parameterised as GPD(ψ, ξ). Expression (2.1.4) shows that the scaled excesses of a threshold tends to a GPD as the threshold tends to the upper endpoint of the distribution.

The limiting result in (2.1.4) motivates the use of a GPD for an approximate distribution of excesses above a suitably high threshold, such that $(X - u) \mid X > u \sim \text{GPD}(\sigma_u, \xi)$, with

$$\Pr(X > x | X > u) = \left(1 + \xi \frac{x-u}{\sigma_u}\right)_+^{-1/\xi}, \; x > u.$$

The threshold stability property of the GPD states that if $X - u \mid X > u \sim \mathrm{GPD}(\sigma_u, \xi)$ for some high threshold u, then for a higher threshold $v \geq u$, $X - v \mid X > v \sim \mathrm{GPD}(\sigma_u + \xi(v-u), \xi)$. Thus, ξ is invariant to threshold choice, but σ_u is not (Davison and Smith, 1990).

We can obtain return levels in a similar way to the block maxima approach, but first we must undo the condition of exceeding the threshold to calculate the marginal survival function $\Pr(X > x)$. We define the parameter $\lambda_u = \Pr(X > u)$ to be the rate of exceedance above u. It follows that

$$\Pr(X > x) = \lambda_u \left[1 + \xi \left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}, \tag{2.1.5}$$

for $x > u$. The m-observation return level x_m, that is, the value one would expect to exceed once every m observations, is then calculated by inverting (2.1.5) such that

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

The T-year return level can be obtained by setting $m = Tn_y$, where n_y is the average number of observations in a year.

A concern in practical implementation of the GPD for modelling threshold exceedances is the choice of threshold u. The asymptotic approximation of the GPD model may not be valid if the threshold is too low, while a threshold that is too high will reduce the size of the dataset, which leads to greater parameter uncertainty. An ideal threshold choice is based on this trade-off between bias and variance. While there are no exact methods for threshold selection, diagnostics are available to guide selection based on properties of the GPD distribution. Classical methods include the analysis of mean residual life plots and parameter stability plots (Coles, 2001). The former is based on the idea that if a GPD model is a good fit, then the sample mean excess over a threshold should be a linear with respect to the threshold. The latter is used based on the idea that ξ and a reparameterised scale parameter $\sigma^* = \sigma_u - \xi u$ are constant with respect to the threshold. Recent advances in threshold selection include Wadsworth and Tawn (2012b), Wadsworth (2016) and Northrop et al. (2017). Other methods are outlined comprehensively in Scarrott and MacDonald (2012).

The point process framework provides an alternative approach for characterising threshold excesses. We assume the limit Poisson process P defined in (2.1.3) is a reasonable approximation to the behaviour of P_n on $A_u = [0,1] \times [u, \infty)$, where u is a suitably high threshold. It is common practice to multiply the intensity function by a factor m (Smith, 1989), where m is free, such that the likelihood can be expressed as

$$L(\mu_m, \sigma_m, \xi) = \exp\left\{-m\left[1+\xi\left(\frac{u-\mu_m}{\sigma_m}\right)\right]_+^{-1/\xi}\right\}\prod_{j=1}^{r}\frac{1}{\sigma_m}\left[1+\xi\left(\frac{x_j-\mu_m}{\sigma_m}\right)\right]_+^{-1/\xi-1}, \tag{2.1.6}$$

where (μ_m, σ_m, ξ) denote the rescaled parameters (that now depend on m), r denotes the number of excesses above the threshold u and $x_j > u, j = 1, \dots, r$ denote the threshold excesses. This point process characterisation is a generalisation of both the GEV and GPD models. The GEV parameters are recovered when m is set to be the number of years of observation, whereas a choice of $m = r$ recovers the GPD parameters. For general m, the parameters (μ_m, σ_m, ξ) are invariant to the choice of threshold. This can be beneficial when adapting the model to account for non-stationarity by modelling the parameters as functions of covariates.

Inference for both the GPD and Poisson process models is carried out typically in the same way as the block maxima approach, where maximum likelihood estimation can be used subject to the restriction that $\xi > -1/2$. Bayesian methods are also often applied as in the block maxima approach. Typically, more precise inferences are obtained from threshold models due to there being more data available than in the block maxima approach. Parameter estimation is often difficult for the Poisson process model, as the parameters are highly non-orthogonal. Numerical maximisation schemes are highly dependent on starting values, while MCMC schemes are often associated with poor mixing and failed convergence. In contrast, the GPD model is easier to fit as the rate parameter λ_u is orthogonal to both σ_u and ξ. In Chapter 3, we explore options for reparameterising the Poisson process model to improve parameter estimation in the Bayesian framework.

2.2 Extratropical cyclones

2.2.1 Overview

In this section, we review the physical processes known as extratropical cyclones, which form the basis for the material in Chapters 5 and 6. Section 2.2.2 features a broad summary of the meteorological and climatological properties of extratropical cyclones, examining their structure, the factors influencing their evolution, and the weather effects they can cause. In Section 2.2.3, we discuss previous studies aimed at modelling these systems in a statistical context.

2.2.2 Meteorology and climatology

Extratropical cyclones are low pressure systems in the mid-latitudes that are largely responsible for stormy weather conditions in the United Kingdom and Northern Europe. For example, Storm Desmond devastated northern England and Scotland in December

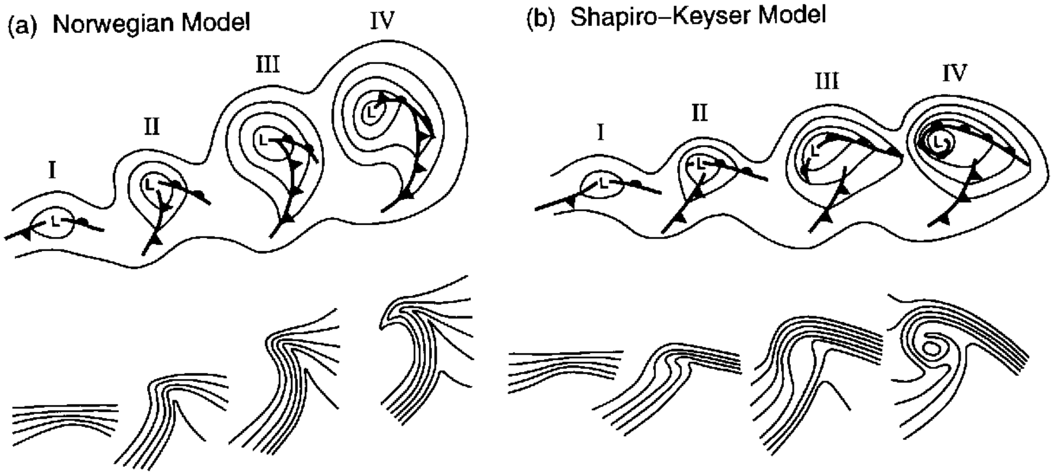


Figure 2.2.1: From Schultz et al. (1998): (a) Norwegian cyclone model: (I) incipient frontal cyclone, (II) and (III) narrowing warm sector, (IV) occlusion; (b) Shapiro-Keyser cyclone model: (I) incipient frontal cyclone, (II) frontal fracture, (III) frontal T-bone and bent-back front, (IV) frontal T-bone and warm seclusion.

2015, bringing severe winds and torrential rain, resulting in record-breaking water levels that caused considerable damage. This motivates the need for understanding the factors relating the presence of a cyclone to the impact on the ground, so that infrastructure can be suitably designed for defence against these natural hazards.

In meteorological terms, an extratropical cyclone forms when the interface between warm and cold air masses develops into a wave form with its apex located at the centre of the low-pressure area. This behaviour was formalised by meteorologists behind the Norwegian Cyclone Model (Bjerknes, 1919) (see Figure 2.2.1), who developed methods for classifying the boundaries between air masses, otherwise known as fronts, into four types: cold, warm, stationary and occluded. High levels of precipitation and wind speeds are characteristic of the locations of warm and cold fronts. This model is generally characteristic of oceanic extratropical cyclone formation, but analyses following cyclone formation over land found substantial departures from the Norwegian model.

The ideology behind the Norwegian model was born from analysis of surface weather maps over Europe in a time before routine air observations began. In recent years, due to inconsistencies between data and the Norwegian model, revisions have been made to the original configuration, such as the Shapiro-Keyser model (Shapiro and Keyser, 1990) (see Figure 2.2.1). As with the Norwegian cyclone model, an incipient cyclone develops cold and warm fronts, but in this case, the cold front moves roughly perpendicular to the warm front such that the fronts never meet, the so-called 'T-bone' (see Figure 2.2.1). This is followed by seclusion, the mature phase of the cyclone life-cycle, which may result in hurricane winds and torrential rain. Not all extratropical cyclones originate as frontal waves. Some begin as tropical cyclones before moving into the mid-latitudes, where different types of behaviour have been observed; see Bengtsson et al. (2006) for a detailed overview.

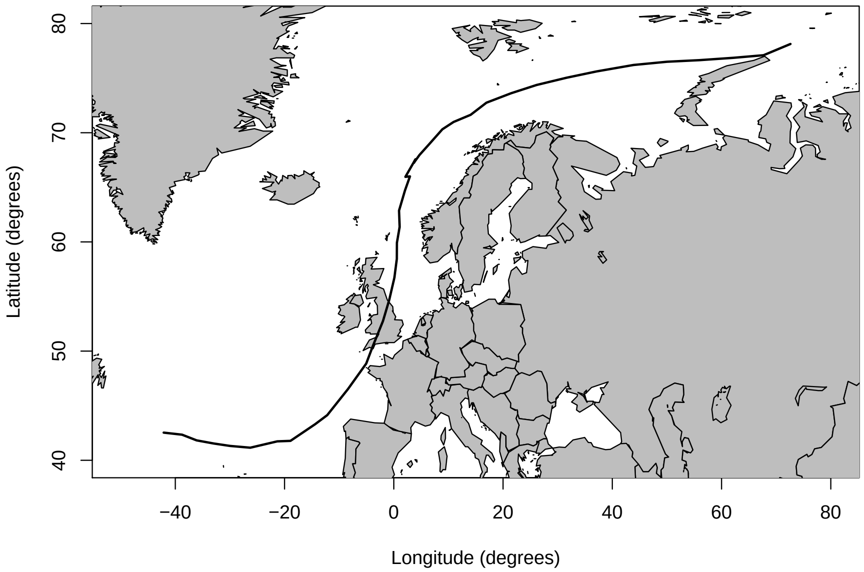


Figure 2.2.2: The storm track of the Great Storm of 1987 extracted using the feature tracking algorithm of Hoskins and Hodges (2002). The Great Storm of 1987 caused devastating wind gusts leading to 22 fatalities and £2bn worth of damage in the UK.

An extratropical cyclone is typically characterised in a spatial sense by a path known as a track (see Figure 2.2.2), which represents the movement of the centre of the cyclone. Many factors affect the movement of the track, and consequently, the complexity of quantifying the intensity of extratropical cyclones makes their occurrence difficult to predict. Cyclones are usually identified as minima in mean surface level pressure (MSLP), either by searching for closed isolines on weather maps (Schinke, 1993) or with an automated approach. However, in some regions, low pressure systems have characteristics different from mid-latitude cyclones, such as monsoon depressions.

Automated approaches are usually based on extracting features of a cyclone from reanalysis data. A reanalysis dataset is the output from a numerical weather prediction model where the past atmospheric observations have been re-analysed using fixed, up-to-date data assimilation systems. This provides a dynamically consistent estimate of the climate state at each timestep. Reanalysis projects have produced global datasets of weather variables with continuously improving temporal and spatial resolution. However, biases in models and observations can induce spurious trends and variability into outputs. Examples of reanalysis projects are featured in Uppala et al. (2005) and Dee et al. (2011).

With regard to extratropical cyclones, a number of identification schemes exist that extract information about cyclones from reanalysis data using vorticity as an indicator. Vorticity measures the rotation of the atmosphere about some vertical axis. It has units of "rotations" per second, where "rotations" are dimensionless. Absolute vorticity refers to the spin caused by local conditions and the rotation of the earth. Relative vorticity refers to the spin caused by local atmospheric flow only. For the remainder of the thesis, relative vorticity is referred to as vorticity, for simplicity. Vorticity is high in atmospheric troughs and low on atmospheric ridges. Hoskins and Hodges (2002) suggest that vorticity is a better indicator than surface-level pressure for identifying synoptic-scale systems such as extratropical cyclones. Once cyclones are detected, the track is identified using techniques based on a nearest-neighbour search, accounting for the expected movement of a cyclone based on atmospheric conditions (Murray and Simmonds, 1991). Chapters 5 and 6 feature statistical models constructed based on the cyclone tracking algorithm of Hoskins and Hodges (2002) for detecting local vorticity maxima in space. Further details on this feature extraction algorithm can be found in Chapter 5.

There have been numerous studies on relating the weather conditions in the vicinity of the extratropical cyclone system to the physical structure of the cyclone itself. Catto et al. (2010) analysed the distribution of surface weather features in a spherical cap of a 15° radius centred at the storm track, identified as the local vorticity maximum using the approach of Hoskins and Hodges (2002). In particular, they assessed the behaviour of surface-level pressure, temperature and wind speeds and matched this behaviour to cyclonic features such as cold fronts, and the cold and warm conveyor belts. Rudeva and Gulev (2011) conduct a similar analysis, in which the relationship between the surface-level pressure associated with an extratropical cyclone and its weather features are explored. They found that the highest precipitation rates tend to occur east of the storm centre, where the warm conveyor belt is typically located, while the rate tends to increase as the pressure gradient increases.

Bengtsson et al. (2006) provide a compehensive analysis of the climatology of extratropical cyclones affecting Northern Europe. Regions where storms begin, known as cyclogenesis regions, tend to be concentrated in the North Atlantic, with some activity on the east coast of the United States. It is suggested that high orography may influence the triggering of a cyclone on land, whereas at sea, cyclones may be activated by local ocean heat fluxes. Tracks generally move in a north/northeasterly direction towards Northern Europe, with evidence to suggest a 'rebirth' of sorts in the eastern North Atlantic. Largest mean intensities tend to be observed south of Greenland in the North Atlantic.

The climatology of extratropical cyclones has been the subject of numerous studies related to how it might vary under anthropogenic climate change. Ulbrich et al. (2009) provide a comprehensive review. Many studies have found that the number of extreme cyclones will increase in winter, with an overall reduction in the total number of cyclones (Lambert and Fyfe, 2006). Several studies emphasise that increased mean intensities only manifest in specific regions, such as the British Isles (Bengtsson et al., 2006), without an increase in the track density at these sites. Jiang and Perrie (2008) see an increase of cyclone radius and a northward shift of the track density, but only small increases in intensity. Many studies (Bengtsson et al., 2006; Pinto et al., 2007) give evidence of a reduced impact from cyclones on the Mediterranean region in winter. Despite no overwhelmingly broad consensus on the issue, there is little doubt that the characteristics of extratropical cyclones are changing in response to climate change induced by human activity.

2.2.3 Statistical modelling

Literature on statistical modelling of cyclonic behaviour in the extratropics is sparse. Rather than explicitly modelling the events themselves, authors have used many of the extreme value methods introduced in Section 2.1 to gain insight into how the extremal behaviour of various measures of storm intensity is changing with respect to climatic variables. Lionello et al. (2008) fit a GEV model to monthly pressure minima derived from regional climate models in order to analyse the impact of a changing climate on cyclone intensities. This paper found that, in two climate change scenarios, it is projected that North Atlantic regions will have stormier winters and milder summers, consistent with the predicted effects of a poleward shift of storm tracks. Similarly, Della-Marta and Pinto (2009) used a GPD model to assess future changes in extreme wind intensity, with the result that the frequency of intense wind events in western Europe is predicted to increase. Sienz et al. (2010) extended this model to incorporate a temporal trend and the North Atlantic Oscillation (NAO) index. However, these analyses fail to account for the spatial variability of both the storm track and the extremes over Europe, a key issue as local estimation of extreme weather conditions is of specific interest to practitioners.

Bonazzi et al. (2012) uses a bivariate extreme value copula to assess the tail dependence of wind intensities between pairs of locations over Europe. Tail dependence exhibits stronger coupling in the west-east direction, which is consistent with the dominant track of extratropical cyclones. The analysis did not incorporate any direct modelling of the dependence between extreme wind and the storm track, however. Economou et al. (2014) specify a spatial extreme value model for pressure minima in the Bayesian hierarchical framework, in which the parameters of the Poisson process model (2.1.6) are dependent on a spatio-temporal covariate structure and spatial random effects. The model enables a regional analysis of pressure minima and the upper endpoint of the distribution of pressure minima is estimated. However, this approach largely ignores the spatial and temporal evolution of the storm track itself, which could be critical in terms of the extreme weather that it generates. A Bayesian hierarchical model of this type is constructed and applied to rainfall data in Chapter 4.

Most developments in storm track modelling have come from the tropical cyclone literature. Casson and Coles (2000) generate synthetic storm tracks by sampling from observed data with random perturbations. The authors incorporate various aspects of cyclone structure, including the distribution of wind speeds in the vicinity of the track as a function of the pressure minimum. Rumpf et al. (2007) sampled from kernel density estimates of speed and direction increments to propagate the track in space, while Hall and Jewson (2007) use a first-order autoregressive process to accomplish this. Neither approach incorporates a model for cyclone intensity.

Youngman and Stephenson (2016) develop a geostatistical approach to simulating the spatio-temporal development of wind gusts generated from extratropical cyclones in Europe. A GPD is used to model the marginal tail behaviour, while the residual dependence across locations is modelled using a Gaussian process. This approach allows fast simulation of synthetic windstorm events over Europe. While the model incorporates the velocity of the storm track in propagating the wind gust events, it lacks a direct link between the storm and the process being modelled. In particular, one might be concerned with how the spatial distribution of wind gusts evolves with respect to the storm centre, and how the strength of the storm affects the magnitude and spatial extent of wind gust events.

Chapter 3

A Poisson process reparameterisation for Bayesian inference for extremes

3.1 A Poisson process model for extremes

The aim of extreme value analysis is to model rare occurrences of an observed process to extrapolate to give estimates of the probabilities of unobserved levels. In this way, one can make predictions of future extreme behaviour by estimating the behaviour of the process using an asymptotically justified limit model. Let $X_1, X_2, \ldots, X_n$ be a series of independent and identically distributed (i.i.d.) random variables with common distribution function F. Defining $M_n = \max\{X_1, X_2, \ldots, X_n\}$, if there exists sequences of normalising constants $a_n > 0$ and b_n such that:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \to G(x) \qquad \text{as } n \to \infty, \tag{3.1.1}$$

where G is non-degenerate, then G follows a generalised extreme value (GEV) distribution, with distribution function

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}, \tag{3.1.2}$$

where $x_+ = \max(x, 0)$, $\sigma > 0$ and $\mu, \xi \in \mathbb{R}$. Here, μ, σ and ξ are location, scale and shape parameters respectively.

Using a series of block maxima from a dataset $\{X_t\}$, typically with blocks corresponding to years, the standard inference approach to give estimates of (μ, σ, ξ) is the maximum likelihood technique, which requires numerical optimisation methods. In these problems, particularly when covariates are involved, such methods may converge to local optima, with the consequence that parameter estimates are largely influenced by the choice of starting values. The standard asymptotic properties of the maximum likelihood estimators are subject to certain regularity conditions outlined in Smith (1985), but can give a poor representation of true uncertainty. In addition, flat likelihood surfaces can cause identifiability issues (Smith, 1987a). For these reasons, we choose to work in a Bayesian setting. Bayesian approaches have been used to make inferences about $\boldsymbol{\theta} = (\mu, \sigma, \xi)$ using standard Markov Chain Monte Carlo (MCMC) techniques as no closed-form posterior distribution exists. They have the advantage of being able to incorporate prior information when little is known about the extremes of interest, while also better accounting for parameter uncertainty when estimating functions of $\boldsymbol{\theta}$, such as return levels (Coles and Tawn, 1996). For a recent review, see Stephenson (2016).

An approach to inference that is considered to be more efficient than using block maxima is to consider a model for threshold excesses, which is superior in the sense that it reduces uncertainty due to utilising more extreme data (Smith, 1987b). Given a high threshold u, the conditional distribution of excesses above u can be approximated by a generalised Pareto distribution (GPD) (Pickands, 1975) such that

$$\Pr(X - u > x | X > u) = \left(1 + \frac{\xi x}{\psi_u}\right)_+^{-1/\xi}, \; x > 0,$$

where $\psi_u > 0$ and $\xi \in \mathbb{R}$ denote the scale and shape parameters respectively, with ψ_u dependent on the threshold u, while ξ is identical to the shape parameter of the GEV distribution. This model conditions on an exceedance, and a third parameter λ_u, denoting the rate of exceedance of X above the threshold u, must also be estimated.

Both of these extreme value approaches are special cases of a unifying limiting Poisson process characterisation of extremes (Smith, 1989; Coles, 2001). Let P_n be a sequence of point processes such that

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

where $a_n > 0$ and b_n are the normalising constants in limit (3.1.1). The limit process is non-degenerate since the limit distribution of $(M_n - b_n)/a_n$ is non-degenerate. Small points are normalised to the same value $b_L = \lim_{n \to \infty} (x_F - b_n)/a_n$, where x_F is the lower endpoint of the distribution F. Large points are retained in the limit process. It follows that P_n converges to a non-homogeneous Poisson process P on regions of the form $A_y = (0, 1) \times [y, \infty)$, for $y > b_L$. The limit process P has an intensity measure on A_y given by

$$\Lambda(A_y) = \left[1 + \xi \left(\frac{y - \mu}{\sigma}\right)\right]_+^{-1/\xi}. \qquad (3.1.3)$$

It is typical to assume that the limit process is a reasonable approximation to the behaviour of P_n, without normalisation of the $\{X_i\}$, on $A_u = (0, 1) \times [u, \infty)$, where u is a sufficiently high threshold and a_n, b_n are absorbed into the location and scale parameters of the intensity (3.1.3). It is often convenient to rescale the intensity by a factor m, where $m > 0$ is free, so that the n observations consist of m blocks of size n/m with the maximum M_m of each block following a GEV(μ_m, σ_m, ξ) distribution, with ξ invariant to the choice of m. The Poisson process likelihood can be expressed as

$$L(\boldsymbol{\theta}_m) = \exp\left\{-m\left[1+\xi\left(\frac{u-\mu_m}{\sigma_m}\right)\right]_+^{-1/\xi}\right\}\prod_{j=1}^{r}\frac{1}{\sigma_m}\left[1+\xi\left(\frac{x_j-\mu_m}{\sigma_m}\right)\right]_+^{-1/\xi-1}, \quad (3.1.4)$$

where $\boldsymbol{\theta}_m = (\mu_m, \sigma_m, \xi)$ denotes the rescaled parameters, r denotes the number of excesses above the threshold u and $x_j > u, j = 1, \ldots, r$, denote the exceedances. It is possible to move between parameterisations associated with different numbers of blocks. If for k blocks the block maximum is denoted by M_k and follows a GEV distribution with the parameters $\boldsymbol{\theta}_k = (\mu_k, \sigma_k, \xi)$, then for all x

$$\Pr(M_k < x) = \Pr\left(M_m < x\right)^{m/k}.$$

As M_k is GEV(μ_k, σ_k, ξ) and M_m is GEV(μ_m, σ_m, ξ) it follows that

$$\begin{aligned} \mu_k &= \mu_m - \frac{\sigma_m}{\xi}\left(1-\left(\frac{k}{m}\right)^{-\xi}\right) \\ \sigma_k &= \sigma_m\left(\frac{k}{m}\right)^{-\xi}. \end{aligned} \quad (3.1.5)$$

In this chapter, we present a method to improve inference for $\boldsymbol{\theta}_k$, the parameterisation of interest. For an ‘optimal’ choice of m we first undertake inference for $\boldsymbol{\theta}_m$ before transforming our results to give inference for $\boldsymbol{\theta}_k$ using the mapping in expression (3.1.5).

In many practical problems, k is taken to be n_y, the number of years of observation, so that the annual maximum has a GEV distribution with parameters $\boldsymbol{\theta}_{n_y} = (\mu_{n_y}, \sigma_{n_y}, \xi)$. Although inference is for the annual maximum distribution parameters $\boldsymbol{\theta}_{n_y}$, the Poisson process model makes use of all data that are extreme, so inferences are more precise than estimates based on a direct fit of the GEV distribution to the annual maximum data as noted above.

To help see how the choice of m affects inference, consider the case when $m = r$, the number of excesses above the threshold u. If a likelihood inference was being used with this choice of m, the maximum likelihood estimators $(\hat{\mu}_r, \hat{\sigma}_r, \hat{\xi}) = (u, \hat{\psi}_u, \hat{\xi})$, see Appendix A.1 for more details. Therefore, Bayesian inference for the parameterisation of the Poisson process model when $m = r$ is equivalent to Bayesian inference for the GPD model.

Although inference for the Poisson process and GPD models is essentially the same approach when $m = r$, they differ in parameterisation, and hence inference, when $m \neq r$. The GPD model is advantageous in that λ_u is globally orthogonal to ψ_u and ξ. Chavez-Demoulin and Davison (2005) achieved local orthogonalisation of the GPD model at the maximum likelihood estimates by reparameterising the scale parameter as $\nu_u = \psi_u(1 + \xi)$. This ensures all the GPD tail model parameters are orthogonal locally at the likelihood mode. However, the scale parameter is still dependent on the choice of threshold. Unlike the GPD, the parameters of the Poisson process model are invariant to choice of threshold, which makes it more suitable for covariate modelling and hence suggests that it may be the better parameterisation to use. In contrast, it has been found that the parameters are highly dependent, making estimation more difficult.

As we are working in the Bayesian framework, strongly dependent parameters lead to poor mixing in our MCMC procedure (Hills and Smith, 1992). A common way of overcoming this is to explore the parameter space using a dependent proposal random walk Metropolis-Hastings algorithm, though this requires a knowledge of the parameter dependence structure *a priori*. Even in this case, the dependence structure potentially varies in different regions of the parameter space, which may require different parameterisations of the proposal to be applied. The alternative approach is to consider a reparameterisation to give orthogonal parameters. However, Cox and Reid (1987) show that global orthogonalisation cannot be achieved in general.

This chapter illustrates an approach to improving Bayesian inference and efficiency for the Poisson process model. Our method exploits the scaling factor m as a means of creating a near-orthogonal representation of the parameter space. While it is not possible in our case to find a value of m that diagonalises the Fisher information matrix, we focus on minimising the off-diagonal components of the covariance matrix. We present a method for choosing the 'best' value of m such that near-orthogonality of the model parameters is achieved, and thus improves the convergence of MCMC and sampling from the joint posterior distribution. Our focus is on Bayesian inference but the reparameterisations we find can be used to improve likelihood inference as well, simply by ignoring the prior term.

The structure of the chapter is as follows. Section 3.2 examines the idea of reparameterising in terms of the scaling factor m and how this can be implemented in a Bayesian framework. Section 3.3 discusses the choice of m to optimise the sampling from the joint posterior distribution in the case where $X_1, \ldots, X_n$ are i.i.d.. Section 3.4 explores this choice when allowing for non-identically distributed variables through covariates in the model parameters. Section 3.5 describes an application of our methodology to extreme rainfall in Cumbria, UK, which experienced major flooding events in November 2009 and December 2015.

3.2 Bayesian inference

Bayesian estimation of the Poisson process model parameters involves the specification of a prior distribution $\pi(\boldsymbol{\theta}_m)$. Then using Bayes Theorem, the posterior distribution of $\boldsymbol{\theta}_m$ can be expressed as

$$\pi(\boldsymbol{\theta}_m|\mathbf{x}) \propto \pi(\boldsymbol{\theta}_m)L(\boldsymbol{\theta}_m),$$

where $L(\boldsymbol{\theta}_m)$ is the likelihood as defined in (3.1.4) and $\mathbf{x}$ denotes the excesses of the threshold u. We sample from the posterior distribution using a random walk Metropolis-Hastings scheme. Proposal values of each parameter are drawn sequentially from a univariate Normal distribution and accepted with a probability defined as the posterior ratio of the proposed state relative to the current state of the Markov chain. In all cases throughout the chapter, each individual parameter chain is tuned to give the acceptance rate in the range of $20\% - 25\%$ to satisfy the optimality criterion of Roberts and Rosenthal (2001). For illustration purposes, results in Sections 3.2 and 3.3 are from the analysis of simulated i.i.d. data. A total of 300 exceedances above a threshold $u = 30$ are simulated from a Poisson process model with $\boldsymbol{\theta}_1 = (80, 15, 0.05)$. Figure 3.2.1 shows individual parameter chains for $\boldsymbol{\theta}_k$ from a random walk Metropolis scheme

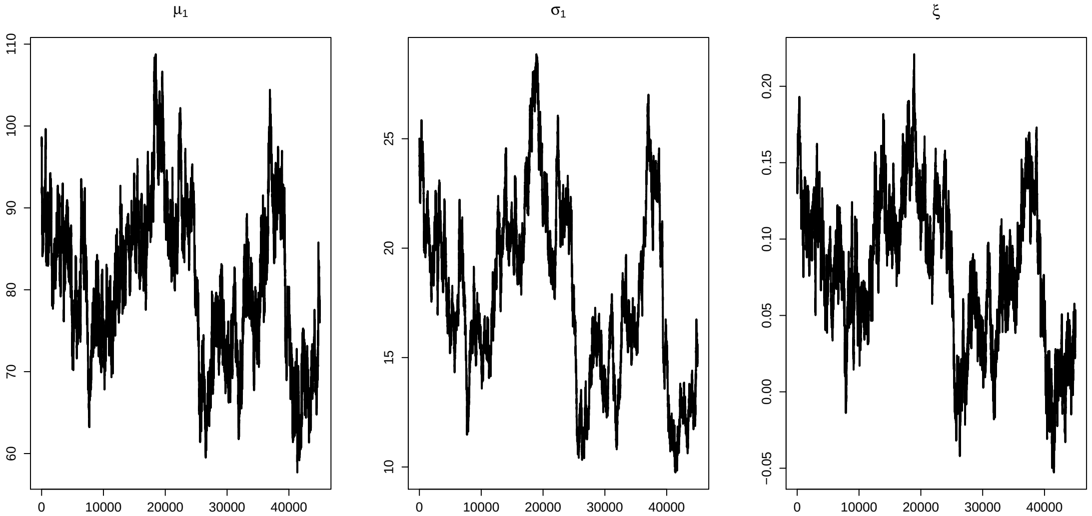


Figure 3.2.1: Random-walk Metropolis chains run for each component of $\boldsymbol{\theta}_1$.

run for $50,000$ iterations with a burn-in of $5,000$ removed, where $k = 1$ and a chosen $m = 1$. This figure shows the clear poor mixing of each component of $\boldsymbol{\theta}_1$, indicating non-convergence and strong dependence in the posterior sampling.

We explore how reparameterising the model in terms of m can improve sampling performance. For a general prior on the parameterisation of interest $\boldsymbol{\theta}_k$, denoted by $\pi(\boldsymbol{\theta}_k)$, Appendix A.2 derives that the prior on the transformed parameter space $\boldsymbol{\theta}_m$ is

$$\pi(\boldsymbol{\theta}_m) = \left(\frac{m}{k}\right)^{-\xi} \pi(\boldsymbol{\theta}_k). \tag{3.2.1}$$

In this example, independent Uniform priors are placed on μ_1, $\log \sigma_1$ and ξ, which gives

$$\pi(\boldsymbol{\theta}_1) \propto \frac{1}{\sigma_1}; \quad \mu_1 \in \mathbb{R}, \sigma_1 > 0, \xi \in \mathbb{R}. \tag{3.2.2}$$

This choice of prior results in a proper posterior distribution, provided there are at least 4 threshold excesses (Northrop and Attalides, 2016). By finding a value of m that near-orthogonalises the parameters of the posterior distribution $\pi(\boldsymbol{\theta}_m|\boldsymbol{x})$, we can run an efficient MCMC scheme on $\boldsymbol{\theta}_m$ before transforming the samples to $\boldsymbol{\theta}_k$. It is noted in Wadsworth et al. (2010) that setting m to be the number of exceedances above the threshold, i.e. $m = r$, improves the mixing properties of the chain, as is illustrated in Figure 3.2.2. This is approximately equivalent to inference using a GPD model, as discussed in Section 3.1.

Given this choice of m, the MCMC scheme is run for $\boldsymbol{\theta}_m$ before transforming to estimate the posterior of $\boldsymbol{\theta}_1$ using the mapping in (3.1.5), where $k = 1$ in this case. Figure 3.2.3 shows contour plots of estimated joint posterior densities of $\boldsymbol{\theta}_1$ based on 5,000 and 50,000 run lengths (with 1,000 and 5,000 samples removed as burn-in respectively). It compares the samples from directly estimating the posterior of $\boldsymbol{\theta}_1$ with that from

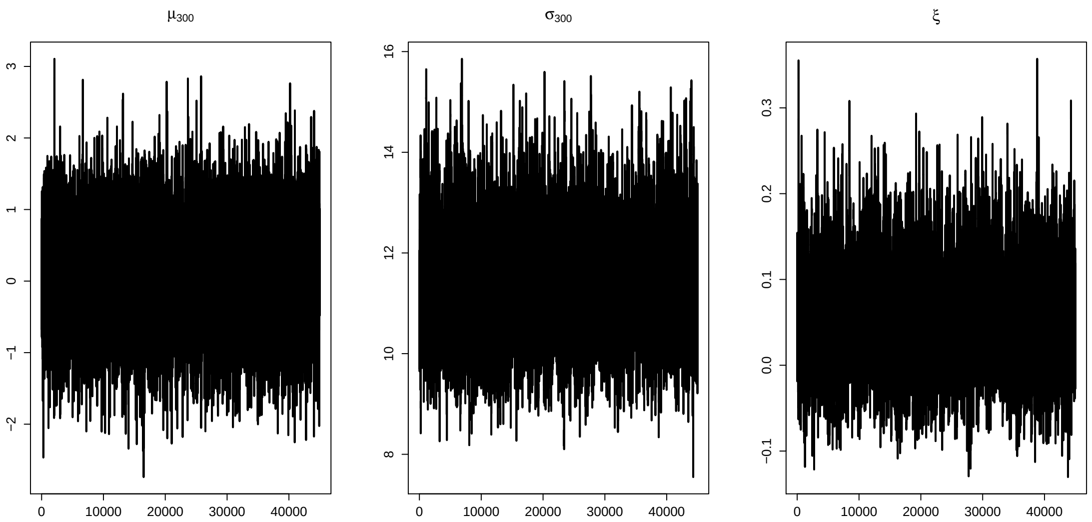


Figure 3.2.2: Random-walk Metropolis chains run for parameters $\boldsymbol{\theta}_r$, where $r = 300$ is the number of exceedances in the simulated data.

transforming from the MCMC samples of the posterior of $\boldsymbol{\theta}_m$ to give a posterior sample for $\boldsymbol{\theta}_1$. Figure 3.2.3 indicates that $\boldsymbol{\theta}_1$ are highly correlated, with the result that we only sample from a small proportion of the parameter space when exploring using independent random walks for each parameter. This explains the poor mixing if we were to run the MCMC without a transformation. In particular, very different estimates of the joint posterior are achieved for the 5,000 and 50,000 run lengths. Even with 50,000 iterations the estimated density contours are very rough, indicating considerable Monte Carlo noise as a result of poor mixing. In contrast, it is clear that, after back-transforming to $\boldsymbol{\theta}_1$, the reparameterisation enables a more thorough exploration of the parameter space, with almost identical estimated joint density contours based on both 5,000 and 50,000 iterations. This shows a very rapid mixing of the associated MCMC. In fact, we found that the reparameterisation yielded smoother density contours for $5,000$ iterations than for 5 million iterations without the transformation. However, while this transformation is a useful tool in enabling an efficient Bayesian inference procedure, further investigation is necessary in the choice of m to achieve near-orthogonality of the parameter space and thus maximising the efficiency of the MCMC procedure.

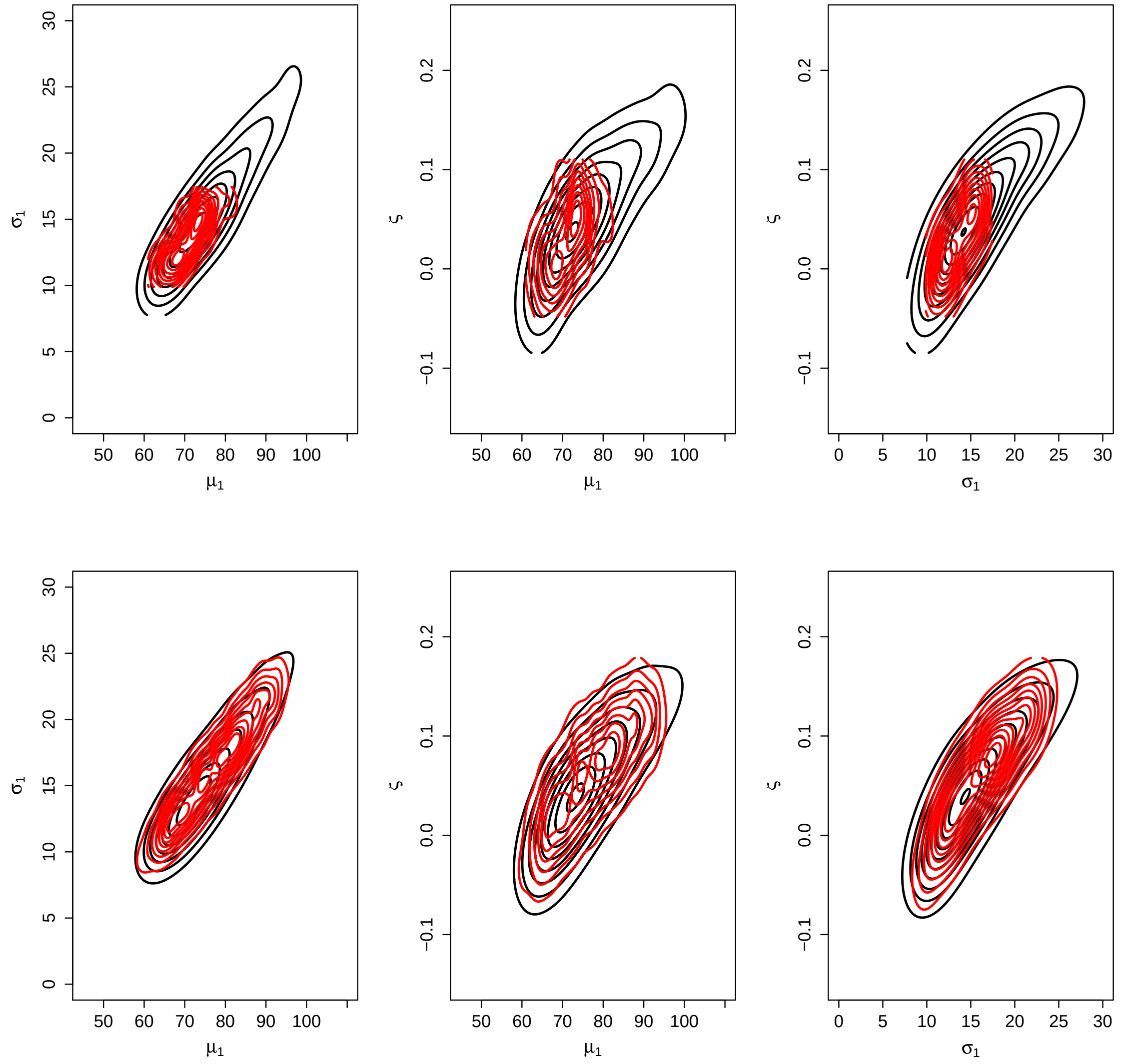


Figure 3.2.3: Contour plots of the estimated joint posterior of $\boldsymbol{\theta}_1$ for 4,000 iterations (top) and 45,000 iterations (bottom) created from the transformed samples drawn from the MCMC procedure for $\boldsymbol{\theta}_m$ (in black) and samples of $\boldsymbol{\theta}_1$ drawn directly (in red).

3.3 Choosing m optimally

As illustrated in Section 3.2, the choice of m in the Poisson process likelihood can improve the performance of the MCMC required to estimate the posterior density of model parameters $\boldsymbol{\theta}_k$. We desire a value of m such that near-orthogonality of $\boldsymbol{\theta}_m$ is achieved, before using the expressions in (3.1.5) to transform to the parameterisation of interest, e.g. $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_{n_y}$. As a measure of dependence, we use the asymptotic expected correlation matrix of the posterior distribution of $\boldsymbol{\theta}_m|\boldsymbol{x}$. In particular, we explore how the off-diagonal components of the matrix, that is, the correlation between parameters, changes with m. The covariance matrix associated with $\boldsymbol{\theta}_m|\boldsymbol{x}$ can be derived analytically by inverting the Fisher information matrix of the Poisson process log-likelihood (see Appendix A.3). The correlation matrix is then obtained by normalising so that the matrix has a unit diagonal.

Other choices for the measure of the dependence of the posterior could have been used, such as the inverse of the Hessian matrix (or the expected Hessian matrix) of the log-posterior, evaluated at the posterior mode. For inference problems with strong information from the data relative to the prior there will be limited differences in the approach and similar values for the optimal m will be found. In contrast, if the prior is strongly informative and the number of threshold exceedances is small then the choice of m from using our approach could be far from optimal. Also the use of the observed, rather than expected, Hessian may better represent the actual posterior distribution of $\boldsymbol{\theta}_m$ and deliver a choice of m that better achieves orthogonalisation, see Efron and Hinkley (1978) and Tawn (1987) respectively.

We prefer our choice of measure of dependence as for i.i.d. problems it gives closed form results for m which can be used without the computational work required for other approaches, and this gives valuable insight into the choice of m to guide future implementation without the need for detailed computation of an optimal m. Furthermore, informative priors rarely arise in extreme value problems, and so information in the data typically dominates information in the prior, particularly around the posterior mode. It should be pointed out however, that the prior is used in the MCMC so there is no loss of prior information in our approach. Also standard MCMC diagnostics should be used even after the selection of an optimal m, so if the asymptotic posterior correlations differ much from the posterior correlations, making our choice of m poor, this will be obvious and a more complete but computationally burdensome analysis can be conducted using the methods described above.

In this section, we use the data introduced in Section 3.2. For all integers $m \in [1, 500]$, maximum posterior mode estimates $\hat{\boldsymbol{\theta}}_m$ are computed and pairwise asymptotic posterior correlations calculated by substituting $\hat{\boldsymbol{\theta}}_m$ into the expressions for the Fisher information matrix, in Appendix A.3, and taking the inverse. Figure 3.3.1 shows how parameter correlations change with the choice of m, illustrating that the asymptotic posterior distributions of μ_m and ξ are orthogonal when $m = r$, the number of excesses above a threshold, which explains the findings of Wadsworth et al. (2010).

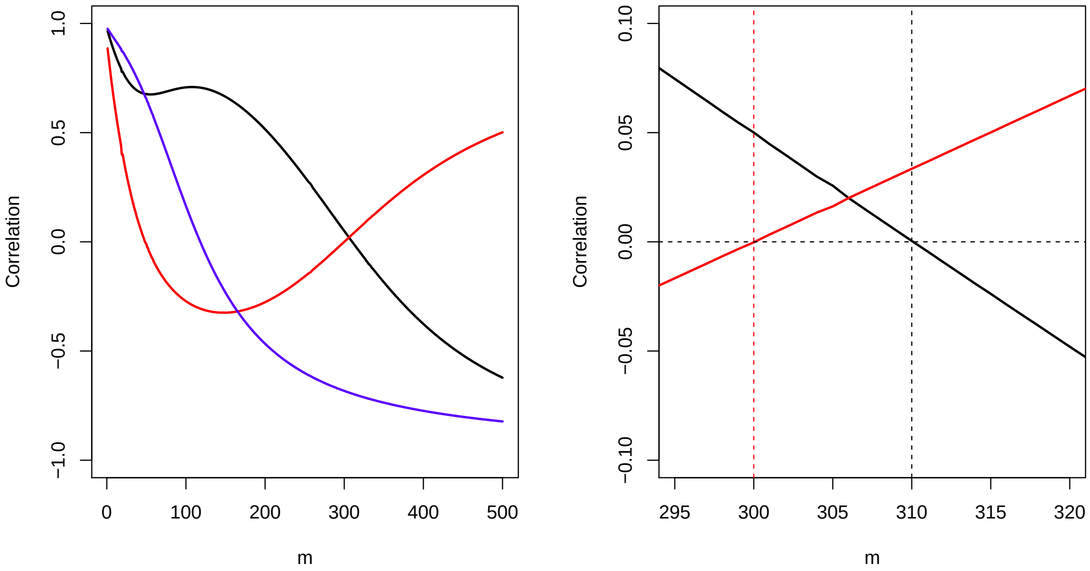


Figure 3.3.1: Left: Estimated parameter correlations changing with m: ρ_{μ_m,σ_m} (black), $\rho_{\mu_m,\xi}$ (red), $\rho_{\sigma_m,\xi}$ (blue). Right: Expanded region of the graph showing $\rho_{\mu_m,\xi} = 0$ for m close to r where $r = 300$ is the number of excesses above the threshold, while $\rho_{\mu_m,\sigma_m} = 0$ when $m \approx 310$.

It is proposed that MCMC mixing can be further improved by minimising the overall correlation in the asymptotic posterior distribution of $\boldsymbol{\theta}_m$. Therefore, we would like to find the value of m such that $\rho(\boldsymbol{\theta}_m)$ is minimised, where $\rho(\boldsymbol{\theta}_m)$ is defined as

$$\rho(\boldsymbol{\theta}_m) = |\rho_{\mu_m,\sigma_m}| + |\rho_{\mu_m,\xi}| + |\rho_{\sigma_m,\xi}|, \tag{3.3.1}$$

where ρ_{μ_m,σ_m} denotes the asymptotic posterior correlation between μ_m and σ_m for example. We also look at the sum of the asymptotic posterior correlation terms involving

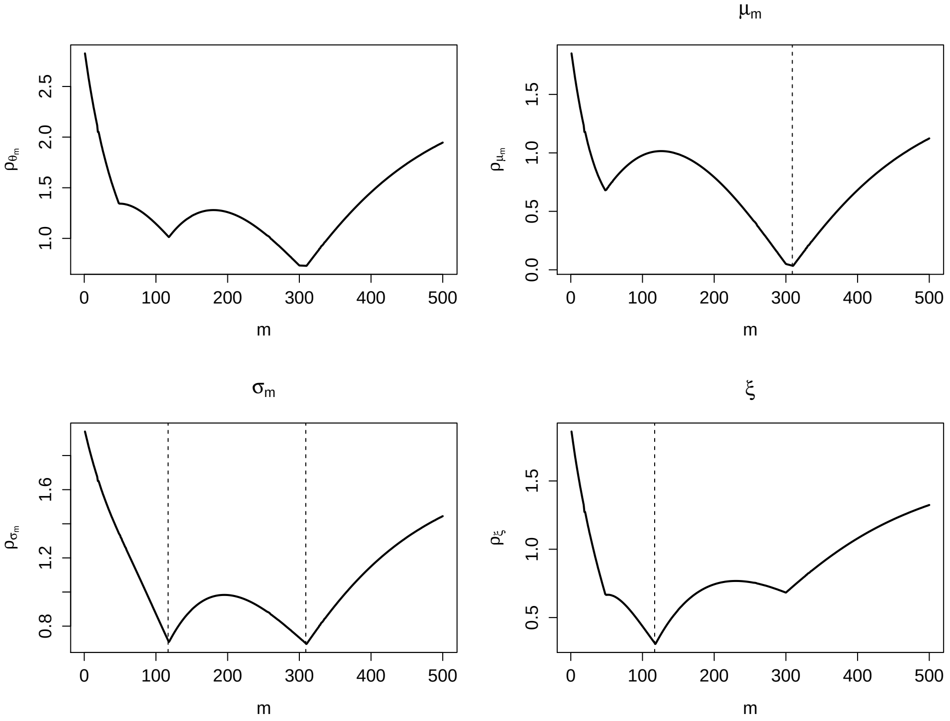


Figure 3.3.2: How $\rho(\boldsymbol{\theta}_m)$ changes with m (top left) and how correlations in each individual estimated parameter, as measured by $\rho_{\mu_m}, \rho_{\sigma_m}$ and ρ_ξ, change with m.

each individual parameter estimate. For example, we define ρ_{μ_m}, the asymptotic posterior correlation associated with the estimate of μ_m, to be:

$$\rho_{\mu_m} = |\rho_{\mu_m,\sigma_m}| + |\rho_{\mu_m,\xi}|. \tag{3.3.2}$$

Figure 3.3.2 shows how the asymptotic posterior correlation associated with each parameter varies with m. From Figure 3.3.2 we see that while ρ_{μ_m} is minimised at the value of m for which $\rho_{\mu_m,\sigma_m} = 0$ (see Figure 3.3.1), ρ_{σ_m} and ρ_ξ have minima at the value of m for which $\rho_{\sigma_m,\xi} = 0$. We denote the latter minimum by m_1 and the former by m_2. In terms of the covariance function, this can be written as:

$$\text{ACov}(\sigma_{m_1}, \xi|\boldsymbol{x}) = \text{ACov}(\mu_{m_2}, \sigma_{m_2}|\boldsymbol{x}) = 0, \tag{3.3.3}$$

where ACov denotes the asymptotic covariance. Figure 3.3.2 shows that m_2 also minimises the total asymptotic posterior correlation in the model.

One would expect that the values of m for which $\rho(\boldsymbol{\theta}_m)$ is minimised would correspond to the MCMC chain of $\boldsymbol{\theta}_m$ with good mixing properties. We examine the effective sample size (ESS) as a way of evaluating this objectively. ESS is a measure of the equivalent number of independent iterations that the chain represents (Robert and Casella, 2009). MCMC samples are often positively autocorrelated, and thus are less precise in representing the posterior than if the chain was independent. The ESS of a parameter chain ϕ is defined as

$$\text{ESS}_\phi = \frac{n}{1 + 2\sum_{i=1}^{\infty} \nu_i}, \tag{3.3.4}$$

where n is the length of the chain and ν_i denotes the autocorrelation in the sampled chain of ϕ at lag i. In practice, the sum of the autocorrelations is truncated when ν_i drops beneath a certain level. Figure 3.3.3 shows how ESS varies with m for each parameter in $\boldsymbol{\theta}_m$. For these data the ESS follow a pattern we found to typically occur. We see that ESS_{μ_m} is maximised at $m = m_2$ due to the near-orthogonality of μ_{m_2} with σ_{m_2} and ξ. We find that ESS_{σ_m} is maximised for $m_1 < m < m_2$, as σ_{m_1} remains substantially positively correlated with μ_{m_1} and σ_{m_2} is negatively correlated with ξ. Similarly, ESS_ξ is maximised at a value of m close to m_1, but ξ is negatively correlated with μ_{m_1}, which explains the slight distortion. From these results, we postulate that a selection

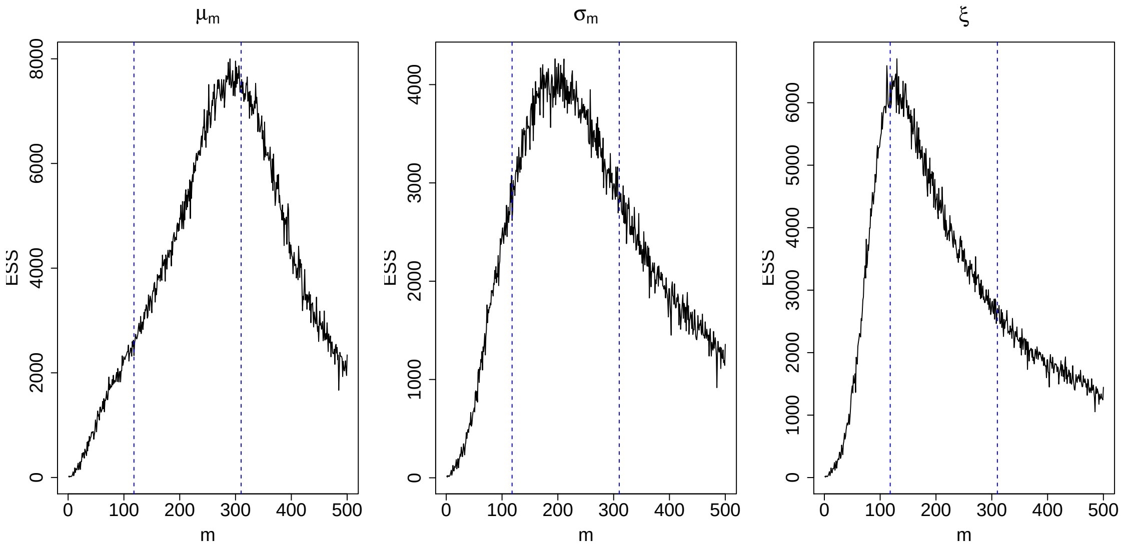


Figure 3.3.3: How ESS varies with m for each parameter in $\boldsymbol{\theta}_m$. The blue dashed lines represent $m = m_1$ (left) and $m = m_2$ (right) in the simulated data example for 45,000 iterations of the MCMC, where m_1 and m_2 are defined by property (3.3.3). In the calculations, the sum of the autocorrelations were truncated when the autocorrelations in the chain drop below 0.05.

of m in the interval $(m_1, m_2) = (118, 310)$ would ensure the most rapid convergence of the MCMC chain of $\boldsymbol{\theta}_m$, thus enabling an effective sampling procedure from the joint posterior. Figure 3.3.3 shows clearly the benefits of the proposed approach. For example, $\text{ESS}_{\mu_{310}} = 7459$ and $\text{ESS}_{\mu_1} = 24$, illustrating that the former parameterisation is over 300 times more efficient than the latter. In addition, by introducing the interval (m_1, m_2), this approach gives a degree of flexibility to the choice of m and giving a balance of mixing quality across the model parameters.

Quantities m_1 and m_2 can be found by numerical solution of equations $\text{ACov}(\sigma_m, \xi|\boldsymbol{x}) = 0$ and $\text{ACov}(\mu_m, \sigma_m|\boldsymbol{x}) = 0$ respectively, using the asymptotic covariance matrix of the posterior of $\boldsymbol{\theta}_m$, which is given by the inverse of the Fisher information (see Appendix A.3). Approximate analytical expressions for m_1 and m_2 can be derived using Halley's method for root-finding (Gander, 1985) applied to equations (3.3.3). This

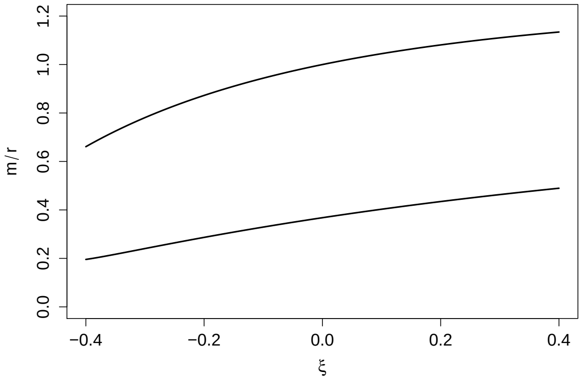


Figure 3.3.4: How $\hat{m}_1$ and $\hat{m}_2$ change as a multiple of r with respect to $\hat{\xi}$: $\hat{m}_1/r$ (bottom), $\hat{m}_2/r$ (top).

method yields the following approximations of m_1 and m_2:

$$\hat{m}_1 = r\frac{(2\xi+1)\left(1+2\xi+(\xi+1)\log\left[\frac{2\xi+3}{2\xi+1}\right]\right)}{(2\xi+3)\left(3+2\xi-(\xi+1)\log\left[\frac{2\xi+3}{2\xi+1}\right]\right)} \tag{3.3.5}$$

$$\hat{m}_2 = r\frac{2\xi^2+13\xi+8}{2\xi^2+9\xi+8}. \tag{3.3.6}$$

In practice, the values of $\hat{m}_1$ and $\hat{m}_2$ are estimated by using an estimate of ξ, such as the maximum likelihood or probability weighted moments estimates. Figure 3.3.4 shows how $\hat{m}_1$ and $\hat{m}_2$ change relative to r for a range of ξ. This illustrates that for negative estimates of the shape parameter, r is not a suitable candidate to be the 'optimal' value of m as it is not in the range (m_1, m_2). In the simulated data used in this section, although a selection of $m = r$ is reasonable, Figure 3.3.3 shows that this may not be wise if one was primarily concerned about sampling well from ξ, for example. In this case, $\hat{m}_2$ is relatively close to r, but Figure 3.3.4 shows that this is not the case for models with a larger positive estimate of ξ.

A simulation study was carried out to assess the suitability of expressions $\hat{m}_1$ and $\hat{m}_2$ as approximations to m_1 and m_2 respectively. A total of $1,000$ Poisson processes were simulated with different values of $\boldsymbol{\theta}_m$. The approximations were calculated and compared with the true values of m_1 and m_2, which were obtained exactly by numerical methods. It was found that $|\hat{m}_i - m_i| < 0.1$ for $i = 1, 2$ always, while $|\hat{m}_i - m_i| < 0.01$ for 78% and 88.2% of the time for $i = 1, 2$ respectively. Both quantities were compared to the performance of other approximations derived using Newton's method, which unlike Halley's method does not account for the curvature in a function. Simulations show that the root mean square errors are significantly smaller for estimates of m_i using Halley's method (0.2% and 5% smaller than Newton's method for $i = 1, 2$ respectively). A summary of the reparameterisation method is given in Algorithm 1.

Algorithm 1: Sampling from the posterior distribution of the Poisson process model parameters $\boldsymbol{\theta}_k = (\mu_k, \sigma_k, \xi)$ or $\boldsymbol{\theta}_k = (\mu_k^{(0)}, \mu_k^{(1)}, \sigma_k, \xi)$ after reparameterising

```
Data: Threshold excesses $\boldsymbol{x}$
Result: Samples from the posterior distribution $\pi(\boldsymbol{\theta}_k|\boldsymbol{x})$
Choose parameterisation of interest $\boldsymbol{\theta}_k$;
if $\boldsymbol{\theta}_k = (\mu_k, \sigma_k, \xi)$ then
    Obtain an estimate of shape parameter $\xi$ using maximum likelihood, for example;
    Compute $\hat{m}_1$ and $\hat{m}_2$ as defined in (3.3.5) and (3.3.6);
    Choose $m$ in range $(\hat{m}_1, \hat{m}_2)$;
else
    Choose $m$ to be the value of $m$ that numerically solves $\rho_{\mu_m^{(0)}, \sigma_m} = 0$;
Obtain MCMC samples for posterior distribution $\pi(\boldsymbol{\theta}_m|\boldsymbol{x})$;
Transform to obtain samples from $\pi(\boldsymbol{\theta}_k|\boldsymbol{x})$ using expression (3.1.5).
```

3.4 Choosing m in the presence of non-stationarity

In many practical applications, processes exhibit trends or seasonal effects caused by underlying mechanisms. The standard methods for modelling extremes of non-identically distributed random variables were introduced by Smith (1989) and Davison and Smith (1990), using a Poisson process and generalised Pareto distribution respectively. Both approaches involve setting a constant threshold and modelling the parameters as functions of covariates. In this way, we model the non-stationarity through the conditional distribution of the process on the covariates. We follow the Poisson process model of Smith (1989) as the parameters are invariant to the choice of threshold if the model is appropriate. We define the covariate-dependent parameters $\boldsymbol{\theta}_m(z) = (\mu_m(z), \sigma_m(z), \xi(z))$, for covariates z. Often in practice, the shape parameter ξ is assumed to be constant. A log-link is typically used to ensure positivity of $\sigma_m(z)$.

The process of choosing m is complicated when modelling in the presence of covariates. This is partially caused by a modification of the integrated intensity measure, which becomes

$$\Lambda(A) = m \int_{\mathbf{z}} \left[1 + \xi(z)\left(\frac{u - \mu_m(z)}{\sigma_m(z)}\right)\right]^{-1/\xi(z)} g(z)\mathrm{d}z, \tag{3.4.1}$$

where g denotes the probability density function of the covariates, which is unknown and with covariate space $\mathbf{z}$. The density term g is required as the covariates associated with exceedances of the threshold u are random. In addition, the extra parameters introduced by modelling covariates increases the overall correlation in the model parameters.

For simplicity, we restrict our attention to the case of modelling when the location parameter is a linear function of a covariate, that is,

$$\mu_m(z) = \mu_m^{(0)} + \mu_m^{(1)} z, \ \sigma_m(z) = \sigma_m, \ \xi(z) = \xi,$$

where we centre the covariate z, as this leads to parameters $\mu_m^{(0)}$ and $\mu_m^{(1)}$ being orthogonal. Note that the regression parameter $\mu_m^{(1)}$ is invariant to the choice of m. A total of 233 excesses above a threshold of $u = 15$ are simulated from a Poisson process model with $\mu_1^{(0)} = 75$, $\mu_1^{(1)} = 30$, $\sigma_1 = 15$, $\xi = -0.05$. We choose g to follow an Exp(2) distribution, noting that one could also choose g to be the density of a covariate that is used in practice. We impose an improper Uniform prior on the regression parameter $\mu_1^{(1)}$ and set up the MCMC scheme in the same manner as in Section 3.2.

The objective remains to identify the value of m that achieves near-orthogonality of the parameters of the posterior distribution. Like before, we run an MCMC sampler on $\boldsymbol{\theta}_m(z)$ and transform the samples back to the parameterisation of interest $\boldsymbol{\theta}_k(z)$, which can be obtained as in (3.1.5) using the relations

$$\begin{aligned}
\mu_k^{(0)} &= \mu_m^{(0)} - \frac{\sigma_m}{\xi}\left(1 - \left(\frac{k}{m}\right)^{-\xi}\right) \\
\mu_k^{(1)} &= \mu_m^{(1)} \\
\sigma_k &= \sigma_m \left(\frac{k}{m}\right)^{-\xi}.
\end{aligned} \tag{3.4.2}$$

The complication of the integral term in the likelihood for non-identically distributed variables means that it is no longer feasible to gain an analytical approximation for the optimal value of m. A referee has suggested a possible route to obtaining such expressions for m in the non-stationary case by building on results in Attalides (2015) and using a non-constant threshold as in Northrop and Jonathan (2011), but as this moves away from our constant threshold case we do not pursue this. We therefore choose a value of m that minimises the asymptotic posterior correlation in the model. Analogous to Section 3.3, the optimal m coincides with the value of m such that $\rho_{\mu_m^{(0)},\sigma_m} = 0$. Using numerical methods, we identify that this corresponds to a value of $m = 85$ for the simulated data example. Figure 3.4.1 shows contour plots of estimated posterior

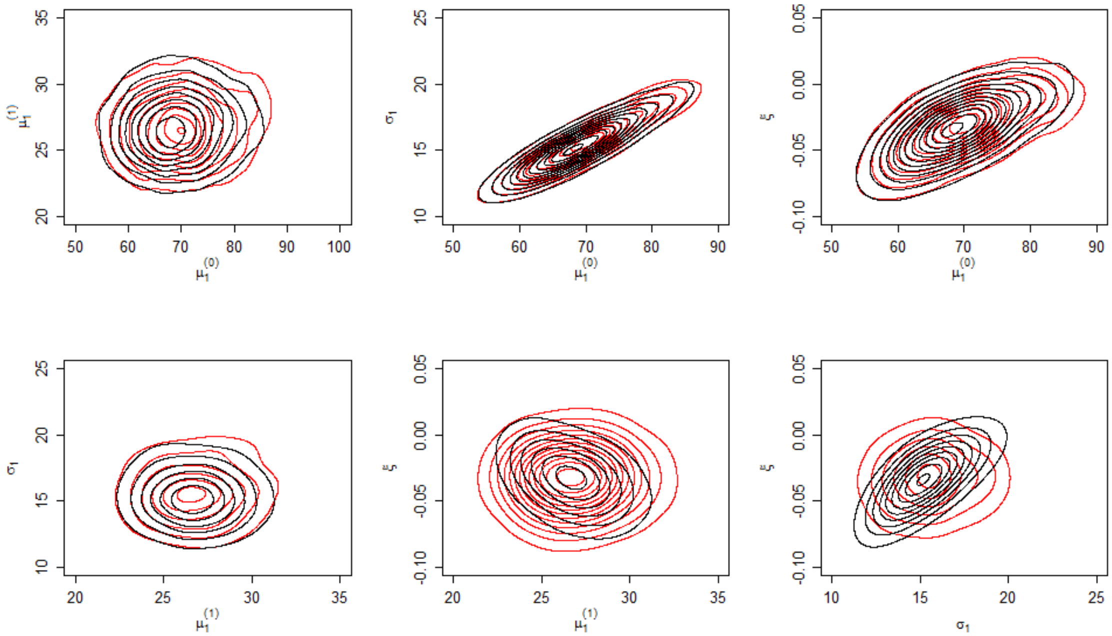


Figure 3.4.1: Contour plots of estimated posterior densities of $\boldsymbol{\theta}_1(z)$ having sampled from the joint posterior directly (red) and having transformed using (3.4.2) after reparameterising from $\boldsymbol{\theta}_{85}(z)$ (black). Both contours are constructed from 50,000 MCMC iterations with a burn-in of 5,000.

densities of $\boldsymbol{\theta}_1(z)$, comparing the sampling from directly estimating the posterior $\boldsymbol{\theta}_1(z)$ with that from transforming the samples from the estimated posterior of $\boldsymbol{\theta}_m(z)$ to give a sample from the posterior of $\boldsymbol{\theta}_1(z)$. From this figure, we see that the reparameterisation improves the sampling from the posterior $\boldsymbol{\theta}_1(z)$.

We again inspect the effective sample size for each parameter as a way of comparing the efficiency of the MCMC under different parameterisations. Figure 3.4.2 shows how the effective sample size varies with m for each parameter. This figure shows how the quality of mixing is approximately maximised in $\mu_m^{(0)}$ for the value of m that minimises the asymptotic posterior correlation. Mixing for $\mu_m^{(1)}$ is consistent across all values of m, which is unsurprising by the invariance of $\mu_m^{(1)}$ by (3.4.2). Interestingly, mixing in ξ increases as the value of m increases. Without a formal measure for the quality of mixing across the parameters, it is found that, when averaging the effective sample size

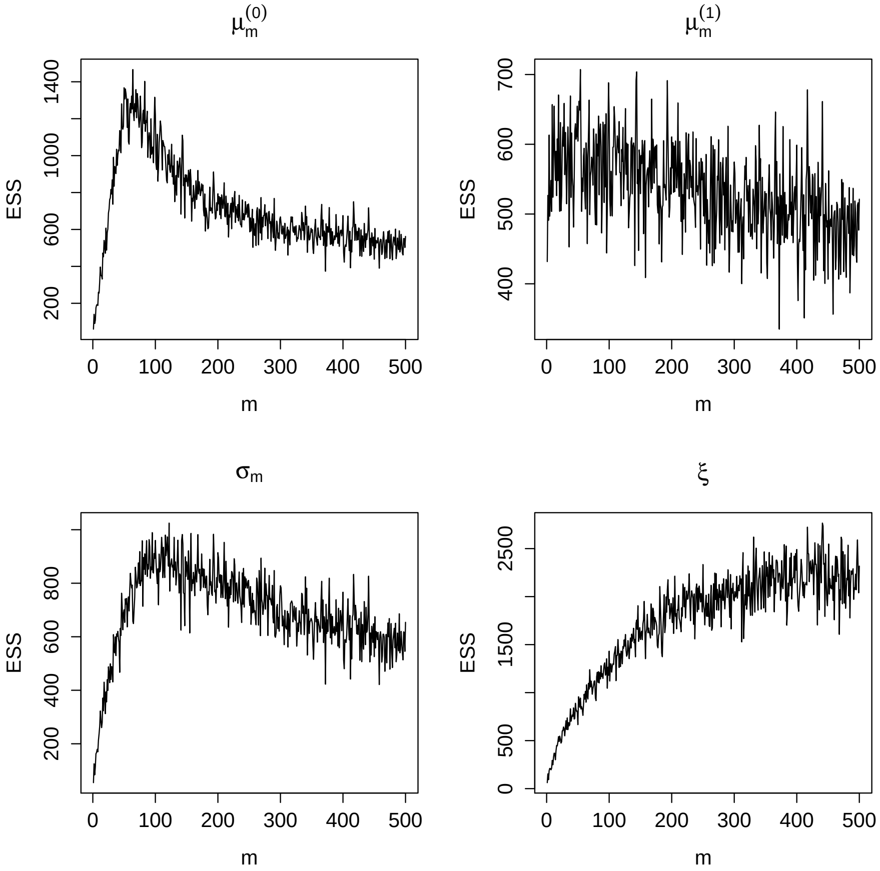


Figure 3.4.2: Effective sample size of each parameter chain of the MCMC procedure.

over the number of parameters, the ESS is stable with respect to m in the interval spanning from the value of m such that $\rho_{\mu_m^{(0)},\sigma_m} = 0$ and the value of m such that $\rho_{\sigma_m,\xi} = 0$, like in Section 3.3. For a summary of how the reparameterisation method can be used in the presence of non-stationarity, see Algorithm 1.

3.5 Case study: Cumbria rainfall

In this section, we present a study as an example of how this reparameterisation method can be used in practice. In particular, we analyse data taken from the Met Office UKCP09 project, which contains daily baseline averages of surface rainfall observations, measured in millimetres, in 25km × 25km grid cells across the United Kingdom in the period 1958-2012. In this analysis, we focus on a grid cell in Cumbria, which has been affected by numerous flood events in recent years, most notably in 2007, 2009 and 2015. In particular, the December 2015 event resulted in an estimated £500 million worth of damage, with rain gauges reaching unprecedented levels. Many explanations have been postulated for the seemingly increased rate of flooding in the North West of England, including climate change, natural climate variability or a combination of both. The baseline average data for the flood events in December 2015 are not yet available, but this event is widely regarded as being more extreme than the event in November 2009, the levels of which were reported at the time to correspond to return periods of greater than 100 years. We focus our analysis on the 2009 event, looking in particular at how a phase of climate variability, in the form of the North Atlantic Oscillation (NAO) index, can have a significant impact on the probability of an extreme event occurring in any given year.

Rainfall datasets on a daily scale are commonly known to exhibit a degree of serial correlation. Analysis of autocorrelation and partial autocorrelation plots indicates that rainfall on a day is dependent on the rainfall of the previous five days. In addition, the data may exhibit seasonal effects. However, while serial dependence affects the effective sample size of a dataset, it does not affect correlations between parameters, and is thus unlikely to influence the choice of m. For the purposes of illustrating our method, we initially make the assumption that the rainfall observations are i.i.d. and proceed with the method outlined in Section 3.3. We wish to obtain information about the parameters corresponding to the distribution of annual maxima, i.e. $\boldsymbol{\theta}_{55}$. Standard threshold diagnostics (Coles, 2001) indicate a threshold of $u = 15$ is appropriate, which corresponds to the 95.6% quantile of the data. There are $r = 880$ excesses above u (see Figure 3.5.1). We obtain bounds m_1 and m_2, then choose a value of m, with

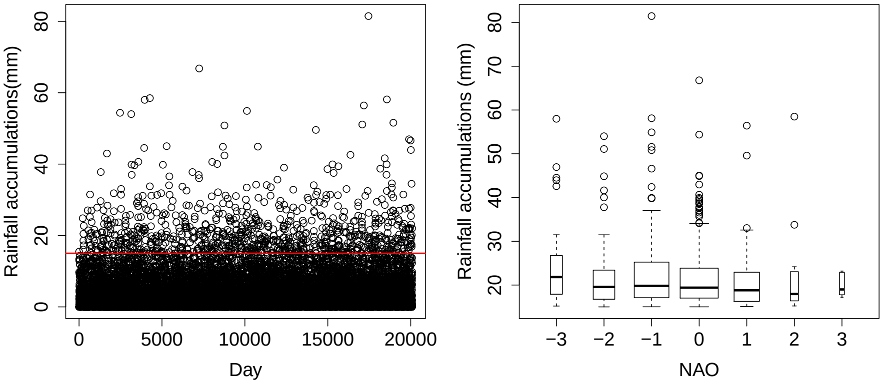


Figure 3.5.1: (Left) Daily rainfall observations in the Cumbria grid cell in the period 1958-2012. The red line represents the extreme value threshold of $u = 15$. (Right) Boxplots of rainfall above u against the corresponding monthly NAO index.

$m_1 < m < m_2$, that will achieve near-orthogonality of the Poisson process model parameters to improve MCMC sampling from the joint posterior distribution. We obtain $\hat{\xi} = 0.087$ using maximum likelihood when $m = r$, which we use to obtain approximations for m_1 and m_2 as in (3.3.5) and (3.3.6). From this, we obtain $\hat{m}_1 \approx 351$ and $\hat{m}_2 \approx 915$. We checked that $\hat{m}_1$ and $\hat{m}_2$ represent good approximations by solving equations (3.3.3) to obtain $m_1 = 350.82$ and $m_2 = 914.96$. Since $r = 880$ is contained in the interval (m_1, m_2), we choose $m = r$. We run an MCMC chain for $\boldsymbol{\theta}_{880}$ for 50,000 iterations, discarding the first 1,000 samples as burn-in. We transform the remaining samples using the mapping in (3.1.5), where $k = 55$, to obtain samples from the joint posterior of $\boldsymbol{\theta}_{55}$. The estimated posterior density for each parameter is shown in Figure 3.5.2.

To estimate probabilities of events beyond the range of the data, we can use the estimated parameters to estimate extreme quantiles of the annual maximum distribution. The quantity y_N, satisfying:

$$1/N = 1 - G(y_N), \tag{3.5.1}$$

is termed the N-year return level, where G is defined as in expression (3.1.2). The level

y_N is expected to be exceeded on average once every N years. By inverting (3.5.1) we get:

$$y_N = \begin{cases} \mu_{55} - \frac{\sigma_{55}}{\xi}[1 - \{-\log(1-1/N)\}^{-\xi}] & \text{for } \xi \neq 0 \\ \mu_{55} - \sigma_{55}\log\{-\log(1-1/N)\} & \text{for } \xi = 0. \end{cases} \tag{3.5.2}$$

The posterior density of the 100-year return level in Figure 3.5.2 is estimated by inputting the MCMC samples of the model parameters into expression (3.5.2).

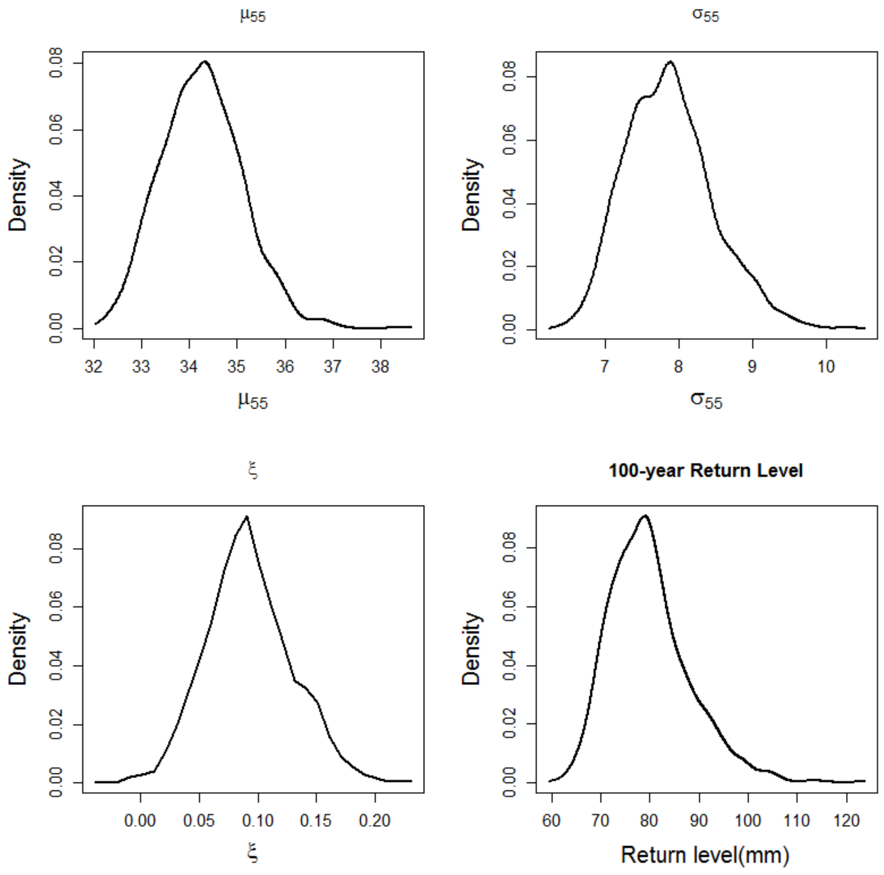


Figure 3.5.2: Estimated posterior densities of μ_{55}, σ_{55}, ξ and the 100-year return level.

We use the same methodology to explore the effect of the monthly NAO index on the probability of extreme rainfall levels in Cumbria. The NAO index describes the surface sea-level pressure difference between the Azores High and the Icelandic Low. The low frequency variability of the monthly scale is chosen to represent the large scale atmospheric processes affecting the distribution of wind and rain. In the UK, a positive NAO index is associated with cool summers and wet winters, while a negative NAO index typically corresponds to cold winters, pushing the North Atlantic storm track further south to the Mediterranean region (Hurrell et al., 2003). In this analysis, we incorporated the effect of NAO by introducing it as a covariate in the location parameter. The threshold of $u = 15$ was retained for this analysis.

To obtain the value of m that minimises the overall correlation in the model, we solve numerically the equation $\rho_{\mu_m^{(0)},\sigma_m} = 0$. We obtain a kernel density estimate of the NAO covariate, which represents g as defined in expression (3.4.1). We use this to obtain maximum posterior mode estimates $\hat{\boldsymbol{\theta}}_r$. These quantities are substituted into the Fisher information matrix. The matrix is then inverted numerically to estimate $m = 920$. This represents a slight deviation from $\hat{m}_2$ estimated during the i.i.d. analysis. We would expect this as the covariate effect is small, as shown in Figure 3.5.3. This example illustrates the benefit of numerically solving for m when modelling non-stationarity, as the range (m_1, m_2) estimated analytically during the i.i.d. analysis no longer contains the optimal value of m.

We run an MCMC chain for $\boldsymbol{\theta}_{920}$ for 50,000 iterations before discarding the first 5,000 samples as burn-in. We transform the remaining MCMC samples to the annual maximum scale using the mapping in (3.4.2) where $k = 55$. Figure 3.5.3 indicates that NAO has a significantly positive effect on the location parameter, as almost all posterior mass is distributed with $\mu_{55}^{(1)} > 0$.

We wish to estimate return levels relating to the November 2009 flood event, which is represented by a value of 51.6mm in the dataset. Return levels corresponding to the distribution of November maxima are shown in Figure 3.5.4. We can also use

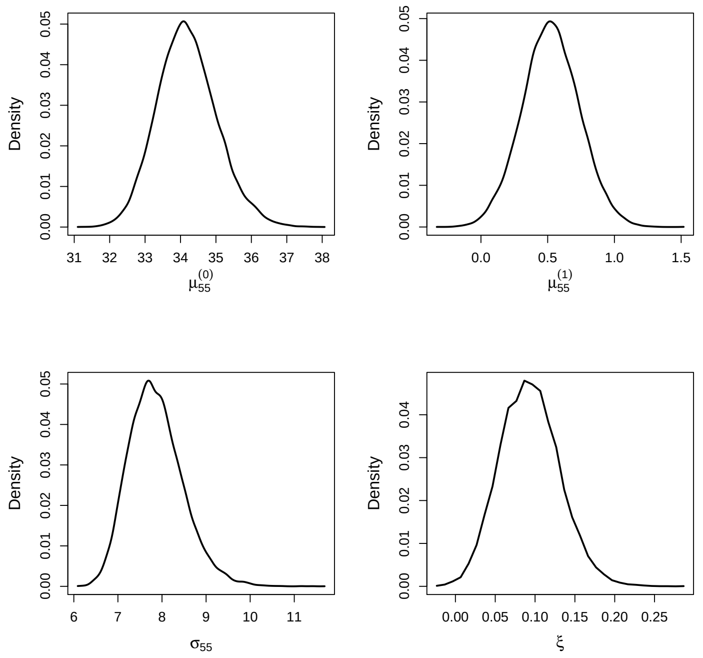


Figure 3.5.3: Estimated posterior densities of $\mu_{55}^{(0)}$, $\mu_{55}^{(1)}$, σ_{55} and ξ.

the predictive distribution in order to account for both parameter uncertainty and randomness in future observations (Coles and Tawn, 1996). On the basis of threshold excesses $\mathbf{x} = (x_1, \ldots, x_n)$, the predictive distribution of a future November maximum M is

$$\Pr\{M \leq y|\mathbf{x}\} = \int_{\boldsymbol{\theta}_{55}} \Pr\{M \leq y|\boldsymbol{\theta}_{55}\}\pi(\boldsymbol{\theta}_{55}|\mathbf{x})\mathrm{d}\boldsymbol{\theta}_{55}, \tag{3.5.3}$$

where assuming stationarity of rainfall conditional on NAO:

$$\Pr\{M \leq y|\boldsymbol{\theta}_{55}\} = \begin{cases} \exp\left\{-\frac{1}{12}\left[1+\xi\left(\frac{y-(\mu_{55}^{(0)}+\mu_{55}^{(1)}z)}{\sigma_{55}}\right)\right]_+^{-1/\xi}\right\} \\ \text{where } z \text{ is known} \\ \\ \exp\left\{-\frac{1}{12}\int_z\left[1+\xi\left(\frac{y-(\mu_{55}^{(0)}+\mu_{55}^{(1)}z)}{\sigma_{55}}\right)\right]_+^{-1/\xi} g_N(z)\mathrm{d}z\right\} \\ \text{where } z \text{ is unknown,} \end{cases} \tag{3.5.4}$$

where g_N is the density of NAO in November and the integral is evaluated numerically using adaptive quadrature methods. The integral in (3.5.3) can be approximated using a Monte Carlo summation over the samples from the joint posterior of $\boldsymbol{\theta}_{55}$. From this, we estimate the predictive probability of an event exceeding 51.6 in a typical November is 0.0112, which corresponds to an 89-year event. For November 2009, when an NAO index of -0.02 was measured, the probability of such an event was 0.0111, corresponding to a 90-year event. For the maximum observed value of NAO in November, with NAO $= 3.04$, the predictive probability of such an event is 0.0132, which corresponds to a 75-year flood event. This illustrates that the impact that different phases of climate variability can have on the probabilities of extreme events is slight but potentially important.

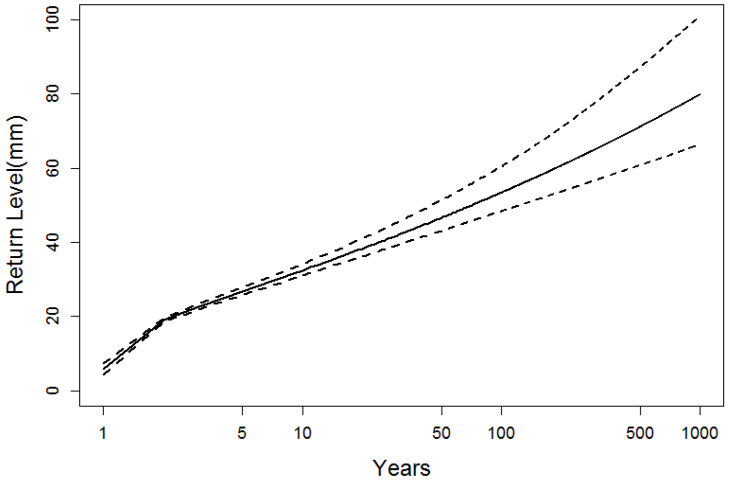


Figure 3.5.4: Return levels corresponding to November maxima. The full line represents the posterior mean and the two dashed lines representing 95% credible intervals.

Chapter 4

A Bayesian spatial hierarchical model for extreme precipitation in Great Britain

4.1 Introduction

In a changing climate with an increased frequency of intense precipitation events (Trenberth, 2011), modelling the rate and size of such events has become increasingly important. In Great Britain, extreme events can arise from the presence of extratropical cyclones evolving from the North Atlantic Ocean. They can also originate from short-term localised convective behaviour. Extreme precipitation levels are commonly associated with an increased risk of flooding (Kunkel et al., 1999), which can contribute to substantial infrastructural damage. Consequently, estimation of extreme precipitation is a vital component of hydrological models for assessing flood risk and therefore needs to be modelled carefully.

Extreme value analysis is used in practice to model rare events by extrapolating beyond observed data to give probability estimates of events occurring at unobserved levels. In this way, one can make predictions of future extreme behaviour by estimating the behaviour of the process using an asymptotically justified limit model. Let $X_1, \ldots, X_n$ be a sequence of independently and identically distributed (i.i.d.) random variables with distribution function F. Defining $M_n = \max(X_1, \ldots, X_n)$, if there exists sequences of constants $a_n > 0$ and b_n, such that, as $n \to \infty$

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

for some non-degenerate distribution G, then G is a generalised extreme value (GEV) distribution with distribution function

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

where $x_+ = \max(x, 0)$, $\sigma > 0$ and $\mu, \xi \in \mathbb{R}$. In this formulation, μ, σ and ξ denote location, scale and shape parameters respectively.

An alternative to modelling the maxima of random variables is to model excesses above a high threshold. Conditional on a high threshold u, the distribution of excesses above u can be approximated by a generalised Pareto distribution (GPD) (Pickands, 1975) such that

$$\Pr(X - u > x | X > u) = \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi}, \ x > 0, \tag{4.1.1}$$

where $\sigma_u > 0$ denotes the threshold-dependent scale parameter and ξ denotes the shape parameter, identical to that of the GEV distribution. A third parameter λ_u, denoting the rate of exceedance, must also be estimated. In practice, this approach to inference is often preferred to analysing block maxima as parameter uncertainty is reduced by utilising more extreme data. The threshold u is typically chosen using standard diagnostics outlined in Coles (2001). An alternative, but equivalent approach for modelling threshold excesses is the Poisson process model (Coles, 2001), which is theoretically more suited to modelling extremes in the presence of covariates, but has more issues in implementation (see Chapter 3). This work uses the GPD as our asymptotically justified threshold excess model.

When modelling spatial extremes, a naive approach would be to analyse the tails independently at each site, with no inherent spatial structure built into models and instead relying on the data to reveal any spatial similarity in the marginal distributions. Precipitation tends to affect clusters of sites simultaneously as a result of a generating storm system, for example. Intuitively, after accounting for physical effects relating to geography, one would expect the probability of extreme precipitation events to be more similar for neighbouring sites than for sites separated by large distances. Analysis of individual sites independently can produce very different probability estimates for neighbouring sites, particularly when the data record is short, which justifies the need for an extreme value model that incorporates spatial information.

A natural class of models for spatial extremes are max-stable processes, which are the extension of univariate and multivariate extreme value theory to the infinite-dimensional setting. In particular, the limiting process of the componentwise maxima of a sequence of normalised stochastic processes is a max-stable process. Max-stable processes are commonly used to model spatial extremes, but are often difficult to fit due to the number of terms required in the likelihood computation. Max-stable processes are unsuitable for modelling variables that are independent in their extremes as they assume extremal dependence (Kereszturi et al., 2016). In addition, they are limited in the sense that they are only suitable for observations that are componentwise (e.g. annual) maxima. Recently, however, generalised Pareto processes have been used to extend the concept of threshold excess models to the space of continuous functions (Ferreira and De Haan, 2014). For more details on max-stable processes, see Smith (1990a), Schlather (2002), Padoan et al. (2010) and Davison et al. (2012).

Another method widely used in the hydrology community is regional frequency analysis (RFA). The aim of RFA is to borrow strength across neighbouring locations with homogeneous statistical behaviour to reduce uncertainty in parameter estimates. It has been extensively studied in Hosking and Wallis (2005). It is a useful approach in extreme value analysis as the pooled information can increase confidence in return level estimates, which can often be highly uncertain due to the scarce nature of the data used in the analysis. The specification of homogeneous regions can be somewhat restrictive as it imposes artificial spatial boundaries on the quantity to be estimated, which can lead to very different extremal characteristics at neighbouring sites that are separated by a boundary. A further disadvantage is that covariates cannot be implemented as part of the L-moments scheme used for estimation, meaning that physical information cannot be incorporated into the model in this manner.

Spatial Bayesian hierarchical models have been used in extreme value analysis with a similar aim to RFA - to use information from neighbouring locations to produce similar return level estimates at neighbouring sites and reduce uncertainty in these estimates. These methods are often used to model spatial count and binary data (Diggle et al., 1998). For a comprehensive overview of such methods, see Banerjee et al. (2004). In recent years, these methods have been extended for use in extreme value analysis. Cooley et al. (2007) modelled threshold excesses at 56 sites using the GPD with an underlying latent spatial model for the GPD parameters. This model was used to interpolate the GPD parameters over the entire domain. Sang and Gelfand (2009) built a hierarchical model for extreme precipitation on a lattice, using an intrinsic autoregressive Model

(IAR) as the latent process, but assumed that the shape parameter was constant over the region they studied. In this work, we use the model of Cooley and Sain (2010), which builds on the model of Sang and Gelfand (2009) by additionally modelling the shape parameter using a latent spatial process. Other recent studies include Schliep et al. (2010), Apputhurai and Stephenson (2013) and Wang and So (2016).

For extreme value analysis, Bayesian hierarchical models are advantageous in their flexibility and incorporation of physical and spatial information through covariates and random effects respectively. Because small proportions of data records are used in extreme value problems, the reduction in uncertainty gained from pooling over space is particularly useful. The Bayesian hierarchical framework relies on the assumption that the extremes are independent conditional on the covariate structure and latent process. While this makes the model unsuitable for modelling joint extremes, that is, extremes that occur simultaneously across locations, our interest lies in how the extremal characteristics of the marginal distribution of precipitation varies across locations. However, when some degree of spatial and/or temporal dependence is apparent in the data, any marginal inference approach must be adjusted to account for this dependence under the misspecified model, otherwise the standard errors derived from the model will be falsely small. In previous studies using a hierarchical model, however, inference is not adjusted to account for dependence in the data, with the consequence that model uncertainty is underestimated.

Previous studies (Smith, 1990b; Fawcett and Walshaw, 2007) have used a modified covariance matrix for computing standard errors to correct for this in the presence of dependence in space and time, due to the fact that the standard asympotic properties of maximum likelihood estimators do not hold for a misspecified model. Ribatet et al. (2012) propose two adjustments to the likelihood function in the Bayesian framework based on asymptotic arguments that account for spatial dependence in a similar manner. In their work, these adjustments were applied to pairwise likelihoods used in the estimation of max-stable processes to correct for overly precise inferences which arise from the pairwise approximation of the spatial dependence structure. However, we apply this reasoning to a model assuming conditional independence not only over space, but over time as well. Our approach utilises the likelihood adjustment of Ribatet et al. (2012) in a hierarchical model, in which we assume conditional independence over space and time but correctly account for both spatial and temporal dependence in the data when quantifying model uncertainty. Our study finds a quantifiable difference in the estimated uncertainty when dependence is accounted for, which serves as a warning to practitioners to ensure inference is performed correctly when false assumptions are made. In some applications, the effect of this can be severe, which highlights the importance of performing such an analysis to determine how standard errors are inflated as a result of dependence in the data.

In this chapter, we describe a spatial extreme value model using the Bayesian hierachical modelling framework, using an adjusted likelihood to account for the spatial and temporal dependence in the data when performing inference on the model parameters. By imposing a condition of spatial similarity on the model parameters, we can produce a map of probabilities of extreme events that exhibits similar behaviour at neighbouring sites. Section 4.2 details the precipitation data along with a previous study using the RFA approach. Section 4.3 gives a comprehensive overview of our modelling strategy. In Section 4.4, we apply the hierarchical model to the data and compare the results with other approaches. We conclude in Section 4.5 with some discussion.

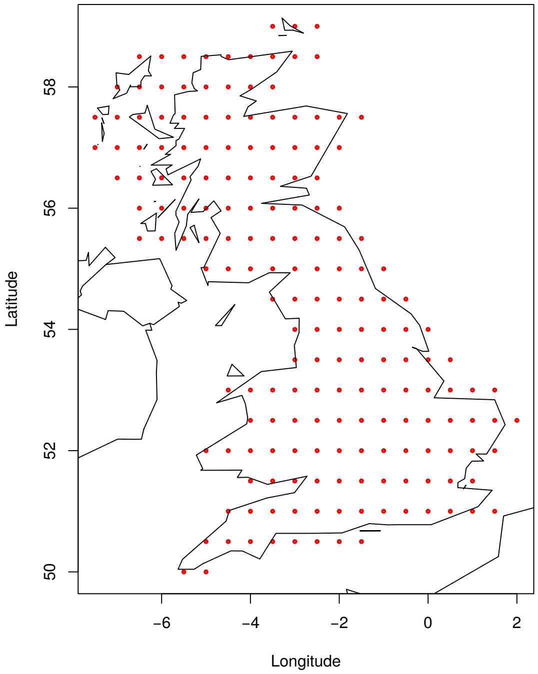


Figure 4.2.1: Locations of gridded reanalysis precipitation data used in the analysis.

4.2 Data

4.2.1 Data description

Reanalysis precipitation data were taken from the Climate Forecast System Reanalysis dataset (CFSR), supplied by the National Centres for Environmental Prediction (NCEP). Input models for the reanalysis include various atmospheric, ocean and land models. Data are available on a daily scale from January 1979 - August 2016 on a 0.5° resolution grid (see Figure 4.2.1). The data are spatially and temporally complete, thus do not contain missing values; this avoids any issues resulting from the treatment of missing data, however, there are limitations. Due to complex processing and the combination of data sources it is often difficult to ascertain the source of error, making uncertainty in the data difficult to quantify.

Figure 4.2.2 shows that the largest precipitation events tend to occur in northwest Scotland, with a general tendency for larger events on the west coast of Great Britain,

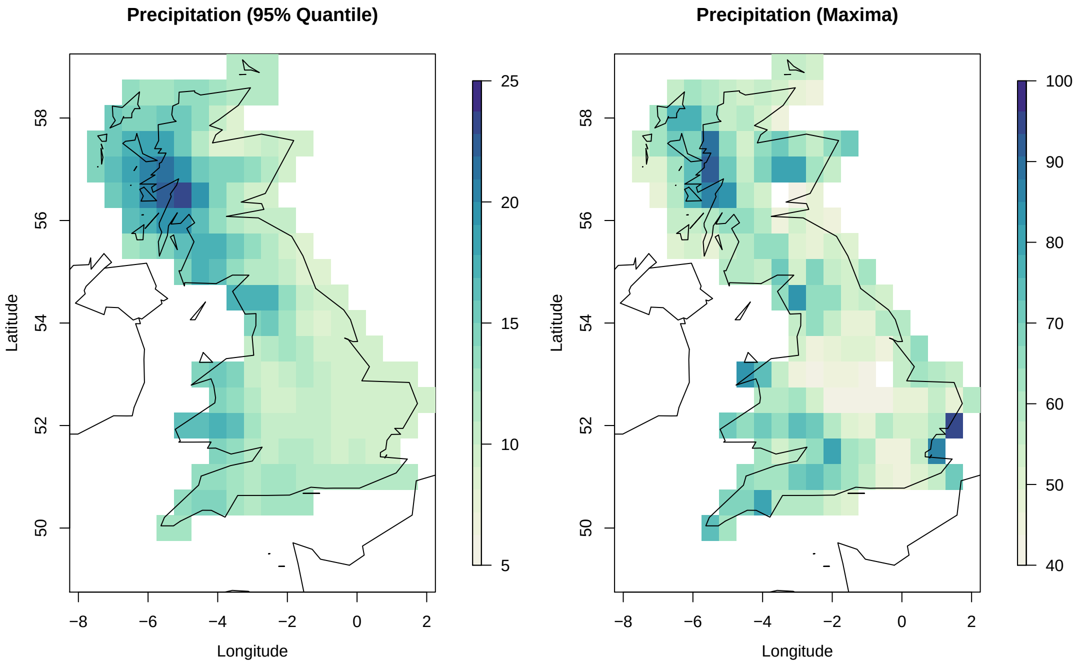


Figure 4.2.2: The 95% quantile (left) and maximum (right) of the daily precipitation data in each cell in the period $1979 - 2014$.

which one may expect due to the west-east passage of weather systems over the Atlantic Ocean and Irish Sea. The spatial distribution of maxima is more varied. The global maximum value occurs in southeast England, in a region that is typically dry, at least compared to Scotland. This event, which occurred during summer, was most likely a result of a short-term, localised convective event, whereas the precipitation in Scotland is typically the result of large-scale synoptic storms arising from the Atlantic Ocean.

4.2.2 Regional frequency analysis

A study of this dataset is conducted in Winter et al. (2017) using a regional frequency analysis (RFA) approach. This aims to reduce uncertainty in return level esimation by pooling information across regions with statistically homogeneous behaviour. The first step in this approach is to define such regions using techniques in Weiss et al. (2014). Observations are standardised by a cell-specific threshold and a regional GPD

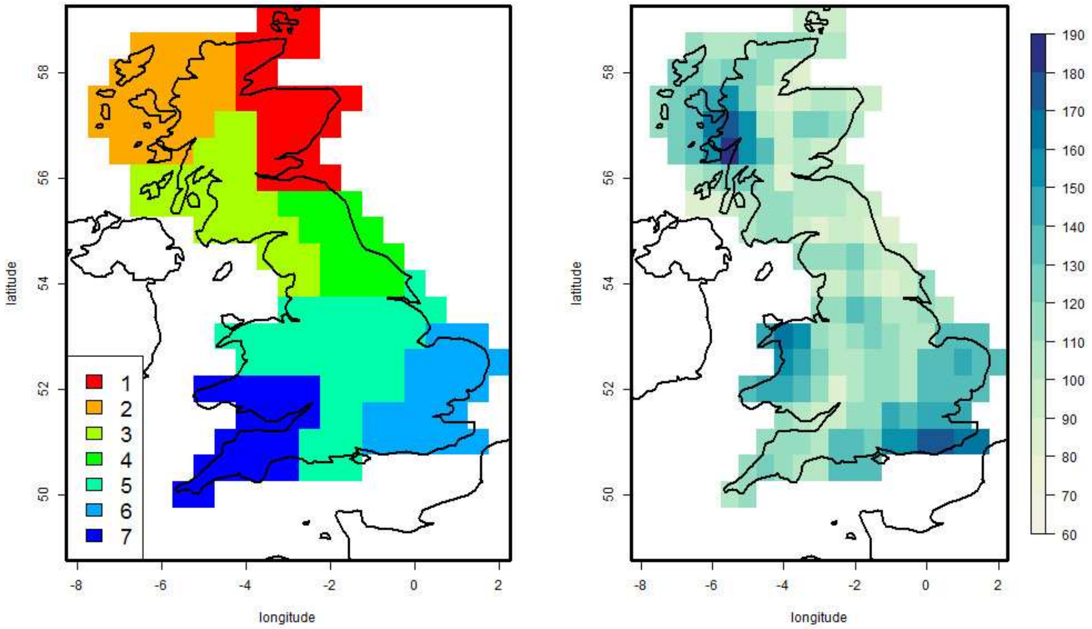


Figure 4.2.3: The homogeneous regions specified for RFA (left) and the 10,000 year return levels (right) estimated using this approach (Winter et al., 2017).

model (4.1.1) is fitted. The cell-specific distribution is obtained by rescaling the regional model using the cell-specific threshold. Inference is carried out using L-moments estimation.

The main drawback of RFA is the fixed specification of regions deemed to be statistically homogeneous. Cells within each homogeneous region are pooled together for parameter estimation. However, this process creates artificial boundaries, meaning that the cells along these boundaries can potentially have very different characteristics from neighbouring cells that have been assigned to other regions. Unless there is a physical analogue of these statistical boundaries, for example, a mountain range, then the intuition behind this specification of spatial similarity begins to break down. Figure 4.2.3 (left panel) shows how the classification algorithm of Weiss et al. (2014) defines the homogeneous regions over our spatial domain.

RFA assumes a constant shape parameter over all cells in each homogeneous region, which is an arguably simplistic assumption to make. When cell-level parameter estimates are used to determine return levels, very different estimates can be produced at neighbouring sites separated by boundaries of homogeneous regions (see Figure 4.2.3, right panel). The large return levels obtained for the southeast region are unrealistic as they are less likely to experience precipitation at the same level as Scotland on a regular basis. This anomaly is most likely caused by the influence of the maxima caused by convective events, as observed in Figure 4.2.2. When the data over the southeast region are pooled together, these events are determining the heaviness of the tail over the entire region under the assumption of a common shape parameter.

This approach brings about substantial reductions in the statistical estimate of return level uncertainty. However, this uncertainty has not been adjusted to account for dependence in the data, as the misspecified model assumes independence over both space and time. Failure to account for this dependence in a model built on independence assumptions can result in misleadingly narrow estimated return level confidence intervals, and perhaps inaccurate design specifications for practitioners as a consequence.

4.3 Model

Our approach uses a Bayesian spatial hierarchical model, with an assumption of conditional independence, to induce similarity of parameter estimates of an extreme value model between neighbouring sites. We describe a method for accounting for spatial and temporal dependence in data when computing standard errors of estimates from this model. The model consists of data, process and prior levels.

4.3.1 Data level

Model specification

On the top layer of the hierarchy, we fit a threshold excess model to the tails of the precipitation distribution in each cell. Let $Y_{j,t}$ be the daily precipitation level in cell $j \in \{1, \ldots, d\}$ on day t. We assume that the excesses above a threshold u_j follow a GPD, defined in (4.1.1). The threshold u_j is typically chosen using standard selection diagnostics (Coles, 2001) or by choosing an appropriately large quantile. We can thus write

$$(Y_{j,t} - u_j) \mid (Y_{j,t} > u_j, \tilde{\sigma}_j, \xi_j) \sim \text{GPD}(\tilde{\sigma}_j, \xi_j), \tag{4.3.1}$$

where $\tilde{\sigma}_j = \sigma_{uj} - \xi_j u_j$ denotes the threshold-independent scale parameter and ξ_j denotes the shape parameter in cell j. We assume that $Y_{j,t} \mid Y_{j,t} > u_j$ given $(\tilde{\sigma}_j, \xi_j)$ is conditionally independent of $Y_{i,t} \mid Y_{i,t} > u_i$ for all $i \neq j$ and all t. While this conditional independence assumption is common in hierarchical modelling, it is often not well-supported in precipitation-based applications as storms can affect multiple locations simultaneously, for example. We also make a working assumption that the precipitation data are i.i.d. over observations in each cell. Partial autocorrelation plots (not shown) show that these data exhibit signs of a third-order temporal dependence structure, meaning that the level of precipitation on a given day will be influenced by levels on the previous three days. Declustering methods are typically used to identify independent clusters of precipitation events, see Ferro and Segers (2003) for details of such a method. We proceed with the assumption of conditional independence in both space and time.

Adjusting for dependence

Given that the data are not independent; rather, we observe clear evidence of spatial and temporal dependence, we require a modification in our approach to ensure correct inference under misspecification of the model. We define the misspecified likelihood to be

$$L(\boldsymbol{\theta}, \boldsymbol{y}) = \prod_{j=1}^{d} L(\boldsymbol{\theta}_j, \boldsymbol{y}_j), \tag{4.3.2}$$

the product of GPD likelihoods over all sites. We denote the log-likelihood as $l(\boldsymbol{\theta}, \boldsymbol{y})$. In a frequentist setting, a modified covariance matrix is used to compute standard errors, which arises from the asymptotic properties of maximum likelihood estimators under misspecified models (Kent, 1982; Chandler and Bate, 2007). An application to extreme value modelling is featured in Smith (1990b). In a misspecified likelihood, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is asymptotically distributed such that

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}^0, n^{-1}H^{-1}VH^{-1}), \tag{4.3.3}$$

where $\boldsymbol{\theta}^0$ is the true value of $\boldsymbol{\theta}$. The quantity $H^{-1}VH^{-1}$ is the modified covariance matrix under misspecification of the model, where $H = -\mathbb{E}\,\nabla^2\, l(\boldsymbol{\theta}^0, \boldsymbol{y})$ and $V = \operatorname{Cov} \nabla\, l(\boldsymbol{\theta}^0, \boldsymbol{y})$, where ∇^i denotes the ith derivative. Both of these quantities can be estimated empirically by setting $\boldsymbol{\theta}^0 = \hat{\boldsymbol{\theta}}$. Smith (1990b) proposed decomposing the log-likelihood into approximately independent yearly contributions in the estimation of V in order to remove possible confounding of dependence in time when accounting for spatial dependence in the data. Fawcett and Walshaw (2007) use a similar approach to account for temporal dependence in the data at one location. In the Bayesian framework, Ribatet et al. (2012) propose two adjustments to the likelihood as a means of allowing correct inference under misspecification of the model. The first, a curvature adjustment, ensures that the asymptotic posterior distribution of $\boldsymbol{\theta}$ is the same as in (4.3.3). The second, a magnitude adjustment, doesn't ensure this but inflates the asymptotic variance compared to an approach where an independent model is assumed correct. Ribatet et al. (2012) show that the choice between the two adjustments makes little difference in application, however. With this in mind, we proceed with a description of the magnitude adjustment due to its simplicity and ease of use.

To be precise, we define the adjusted likelihood $L^*(\boldsymbol{\theta}, \boldsymbol{y})$ such that

$$L^*(\boldsymbol{\theta}, \boldsymbol{y}) = L(\boldsymbol{\theta}, \boldsymbol{y})^k, \tag{4.3.4}$$

where $k > 0$ and $L(\boldsymbol{\theta}, \boldsymbol{y})$ is defined in (4.3.2). This represents a magnitude adjustment of the likelihood that leaves parameter estimates unchanged, but scales the uncertainty of these estimates to account for model misspecification. The adjustment is based on the property that $D_{full} \to \chi^2_p$ as $n \to \infty$, where D_{full} denotes the deviance function corresponding to the full likelihood, which is unknown, and p is the number of parameters and n is the number of data points. For the deviance D_{mis} corresponding to the misspecified likelihood (4.3.4), $D_{mis} \to k \sum_{i=1}^{p} \lambda_i X_i$ as $n \to \infty$, where $X_1, \ldots, X_p$ are independent χ^2_1 random variables and $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of the modified covariance matrix $H^{-1} V H^{-1}$. Setting $k = p / \sum_{i=1}^{p} \lambda_i$ ensures that $\mathrm{E}[D_{mis}]$ converges to $\mathrm{E}[\chi^2_p] = p$. However, in general, higher moments do not match those of χ^2_p. Ribatet et al. (2012) show that this adjustment inflates the asymptotic variance of the posterior, and in practice performs well compared to the curvature adjustment, which ensures the asymptotic posterior matches with (4.3.3).

Estimation of the adjustment factor k requires estimation of the modified covariance matrix. The matrix H is approximated by the observed information matrix $-\nabla^2\, l(\hat{\boldsymbol{\theta}}, \boldsymbol{y})$. We follow a similar approach to estimate V as in Smith (1990b) and Fawcett and Walshaw (2007), which decompose the log-likelihood into its contributions by year. Our approach differs by decomposition of the log-likelihood into contributions over 7-day periods, which analysis of partial autocorrelations suggests are approximately independent, thus it captures the short-range temporal dependence in the data. We can then obtain an empirical estimate of V that accounts for both the spatial and temporal dependence structure in the data. The eigenvalues $\hat{\lambda}_i$ for $i = \{1, \dots, p\}$ of the estimated modified covariance matrix are obtained and k is estimated such that $\hat{k} = p / \sum_{i=1}^{p} \hat{\lambda}_i$. The quantity k can be interpreted as the effective proportion of the dataset containing independent information.

4.3.2 Process level

This layer of the hierarchical model borrows strength across locations. We assume an underlying spatial process in the mean of the distribution of both GPD parameters, such that the parameters are more likely to be similar in neighbouring cells. As well as these spatial effects, we can also incorporate fixed climate or physical effects in a cell through the inclusion of covariates in the model. Formally, we assume

$$\boldsymbol{\theta}_j \sim N(\boldsymbol{X}_j \boldsymbol{\beta} + \boldsymbol{\phi}_j, T_{\boldsymbol{\theta}}^{-1}), \tag{4.3.5}$$

where $\boldsymbol{\theta}_j = (\log \tilde{\sigma}_j, \xi_j)$ is the vector of transformed GPD parameters in cell j, $\boldsymbol{X}_j$ is the design matrix of fixed covariates in cell j, $\boldsymbol{\beta}$ is the vector of regression coefficients, $\boldsymbol{\phi}_j$ is a spatial random effect in cell j and $T_{\boldsymbol{\theta}}$ is the common precision matrix for the transformed GPD parameters over all cells. We include a log-link in the estimation of the scale parameter to ensure positivity in the presence of covariate information.

The latent spatial process is induced through fixed effect terms $\boldsymbol{X}_j$ and the random effect term $\boldsymbol{\phi}_j$. Since we are working with lattice data, the natural choice of prior for $\boldsymbol{\phi}_j$ is a multivariate conditional autoregressive (CAR) model. We make the assumption that the lattice structure of the region represents a Markov random field, and that specification of local conditional relationships can be used to specify a global joint distribution (Banerjee et al., 2004). The intrinsic autoregressive (IAR) model is a special case of the CAR model, with the conditional prior for $\boldsymbol{\phi}_j$ written as

$$\boldsymbol{\phi}_j \mid \boldsymbol{\phi}_{(-j)} \sim N\left(\frac{1}{m_j}\sum_{i\in\delta_j}\boldsymbol{\phi}_i, \frac{1}{m_j}T_{\boldsymbol{\phi}}{}^{-1}\right),$$

where δ_j denotes the set of neighbours of cell j, m_j is the number of neighbours and $T_{\boldsymbol{\phi}}$ denotes the common precision matrix of random effects over all cells. The IAR model has an improper density and does not represent a legitimate probability distribution. For this reason, it cannot be applied directly to data and is often used as a prior, since inference can still be made as long as the posterior distribution is proper. With IAR priors, the fixed effects are identifiable as long as the spatial random effects are centred so that $\sum_{j=1}^{d}\boldsymbol{\phi}_j = 0$ (Banerjee et al., 2004). This can be implemented in practice by centering each joint iteration of the MCMC sampling of the random effect terms. An alternative to the IAR approach is to specify a propriety parameter ρ, which controls the level of spatial association between cells. Cooley and Sain (2010) use a separable formulation to specify the joint prior density of $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_d)$ using matrix decomposition techniques from Rue and Held (2005).

4.3.3 Prior level

With the exception of the random effect parameters inducing spatial dependence, we have no prior knowledge regarding any of the parameters in the model. We choose to assign uninformative priors where possible. We also choose to assign conjugate priors where possible to allow us to sample from the posterior distribution using Gibbs sampling. For the regression coefficients, we use an empirical Bayes approach, such that

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, T_{\boldsymbol{\beta}}^{-1}),$$

where $\boldsymbol{\beta}_0$ and $T_{\boldsymbol{\beta}}^{-1}$ denote the prior mean and precision matrix of the regression coefficients respectively, which are fixed. A fully Bayesian approach could be implemented but we have no extra information regarding these upper-level parameters. Thus, in the interest of parsimony and computational benefit, we proceed with the empirical Bayes approach. We take the intercept terms to be the mean of the cell-wise maximum likelihood estimates for $\log \tilde{\boldsymbol{\sigma}}$ and $\boldsymbol{\xi}$ while the covariate coefficients are given a mean of 0. The precision matrix is chosen so that the intercept terms have a precision of 0.01 and the covariate effects have a precision of 0.1. These values are typically chosen to represent the levels of variability one would expect in the parameter estimates. Because we cannot achieve conjugacy for the GPD parameters, we impose a flat joint prior for $(\log \sigma_j, \xi_j)$ such that $\pi(\boldsymbol{\theta}_j) \propto 1/\sigma_j$.

We assign conjugate Inverse-Wishart priors to both ${T_{\boldsymbol{\theta}}}^{-1}$ and ${T_{\boldsymbol{\phi}}}^{-1}$, such that:

$$\begin{aligned} {T_{\boldsymbol{\theta}}}^{-1} &\sim \text{Inv-Wishart}\left(\nu_\theta, \Omega_\theta\right) \\ {T_{\boldsymbol{\phi}}}^{-1} &\sim \text{Inv-Wishart}\left(\nu_\phi, \Omega_\phi\right). \end{aligned}$$

Matrices Ω_θ and Ω_ϕ were chosen to reflect the levels of variability found within each cell's parameter estimates and also between each cell's parameter estimates.

4.3.4 Implementation

As we have used conjugate priors for some parameters, we can construct closed-form full conditionals from these and thus construct these components of the joint posterior distribution using a Gibbs sampler. The full conditionals are

$$\begin{aligned} &\boldsymbol{\beta} \mid \ldots \sim N\left(\boldsymbol{\mu}_\beta, V_\beta\right) \\ &\quad V_\beta = \left[T_{\boldsymbol{\beta}} + \boldsymbol{X}^T T_{\boldsymbol{\theta}} \boldsymbol{X}\right]^{-1} \end{aligned}$$

$$\boldsymbol{\mu}_\beta = V_\beta \left[T_{\boldsymbol{\beta}}\boldsymbol{\beta}_0 + \boldsymbol{X}^T T_{\boldsymbol{\theta}}(\boldsymbol{\theta} - \boldsymbol{\phi})\right]$$

$$\boldsymbol{\phi}_j \mid \ldots \sim N\left(\boldsymbol{\mu}_\phi, V_\phi\right)$$

$$V_\phi = \left[m_j T_{\boldsymbol{\phi}} + T_{\boldsymbol{\theta}}\right]^{-1}$$

$$\boldsymbol{\mu}_\phi = V_\phi \left[T_{\boldsymbol{\theta}}(\boldsymbol{\theta}_j - \boldsymbol{X}_j\boldsymbol{\beta}) + T_{\boldsymbol{\phi}} \sum_{i \in \delta_j} \boldsymbol{\phi}_i\right]$$

$${T_{\boldsymbol{\theta}}}^{-1} \mid \ldots \sim \text{Inv-Wishart}\left(\nu_\theta + d, \Omega_\theta + (\boldsymbol{\theta} - \boldsymbol{\phi} - \boldsymbol{X\beta})^T(\boldsymbol{\theta} - \boldsymbol{\phi} - \boldsymbol{X\beta})\right)$$

$${T_{\boldsymbol{\phi}}}^{-1} \mid \ldots \sim \text{Inv-Wishart}\left(\nu_\phi + d, \Omega_\phi + \boldsymbol{\phi}^T W \boldsymbol{\phi}\right),$$

where W is a matrix defining spatial proximity between cells. The matrix W has off-diagonal elements $w_{ij} = -1$ if cells i and j are adjacent and $w_{ij} = 0$ otherwise and diagonal elements $w_{ii} = -\sum_{i \neq j} w_{ij}$, being the number of neighbours of cell i.

Proposals for $\boldsymbol{\beta}$, $\boldsymbol{\phi}_j$, ${T_{\boldsymbol{\theta}}}^{-1}$ and ${T_{\boldsymbol{\phi}}}^{-1}$ can be generated using a Gibbs sampler step. The GPD parameters are updated using a Metropolis-Hastings step. We use a random walk Metropolis scheme with a multivariate Gaussian proposal distribution, tuning the proposal covariance matrix appropriately to give the optimal acceptance rate of Roberts and Rosenthal (2001).

4.4 Results

In this section, we compare output from the Bayesian hierarchical modelling approach described in Section 4.3 with two alternative approaches. First, we compare this analysis with results from a model whereby a GPD model is fitted to each cell independently with an uninformative prior on the model parameters. We henceforth refer to this approach as the ‘uninformative’ model. The comparison with the spatial hierarchical model allows us to assess whether there is any value in including a latent spatial process in the modelling procedure. Second, we briefly compare against the regional frequency analysis approach detailed in Section 4.2.2 and describe the strengths of the Bayesian model relative to this method.

In both the uninformative and spatial hierarchical model, we select the 95% quantile as our extreme value threshold for each cell. Threshold stability plots were checked at a handful of cells and the aforementioned threshold represented a sensible choice in all cases. In the absence of spatial information, the uninformative model is specified as in (4.3.1), but where uninformative Uniform priors are imposed on $\log \tilde{\sigma}_j$ and ξ_j for $j = \{1, \ldots, d\}$. Previous studies on extreme rainfall have used the prior density defined by Martins and Stedinger (2000), which is useful in small-sample cases as it constrains the shape parameter to be in a sensible interval. In this analysis, however, we obtained realistic shape parameters for estimating rainfall, so we chose not to use this prior.

To account for spatial and temporal dependence in the data under the misspecified conditional independence model, we use the adjusted likelihood (4.3.4) in the hierarchical model. The constant $\hat{k} = 0.784$ is estimated using methods described in Section 4.3.1 with the adjusted likelihood being used to form the posterior. When accounting for spatial dependence only, that is, when data at each location are assumed to be independent in time, the estimated $\hat{k} = 0.8$. This difference highlights the merits of adjusting for both spatial and temporal dependence in the data in an analysis. We proceed with the value $\hat{k} = 0.784$, which corresponds to a moderate level of spatial and temporal dependence within the data. In both models, we run the MCMC for 20,000 iterations and discard the first 5,000 as burn-in. We assess the convergence of the MCMC chains using the Gelman-Rubin diagnostic (Gelman and Rubin, 1992), which indicates convergence in $> 99\%$ of parameter chains. Inspection of trace plots (not shown) indicates that the parameter chains are mixing well. Parameter estimates in this analysis are taken to be posterior means.

As we are working on a relatively coarse lattice, there is a limited amount of physical information we can incorporate into the problem without being overly simplistic regarding the assumptions we make. For example, including elevation as a covariate would be unwise because various different landscapes can be observed in a single cell, and assigning an average to that would mask out any meaningful effect. We assessed whether adding latitude and longitude to the hierarchical model as covariates as in (4.3.5) would improve the overall fit using the deviance information criterion (DIC) (Spiegelhalter et al., 2002). We included both linear and second-order terms in model estimation. The best-fitting model included a significant negative linear relationship between the scale parameter and longitude. This result is consistent with the tendency for high precipitation events to occur on the west coast (see Figure 4.2.2). Investigation of the random effect terms (see Appendix B.1) shows a north-south trend in the scale, which may correspond to a non-significant latitude fixed effect. There is some evidence of west-east dependence in the random effect term corresponding to the shape parameter, though this is less clear and may instead correspond to localised behaviour. For example, southeast England sees the highest values of this component, which could reflect its tendency to experience short-term convective events, as discussed in Section 4.2.

Figure 4.4.1 (top and centre) compares the GPD parameter estimates from the spatial and uninformative models. In both models, the scale parameter reflects the claim from the exploratory analysis that larger precipitation events tend to occur on the west coast of Britain, with particular impact in Scotland. The heavy tails in the southeast of England are likely determined by short-term, localised convective events. This parameter has a clear geographical structure, with higher values in the southeast of England and smaller values on the south coast and parts of the Midlands, implying that an

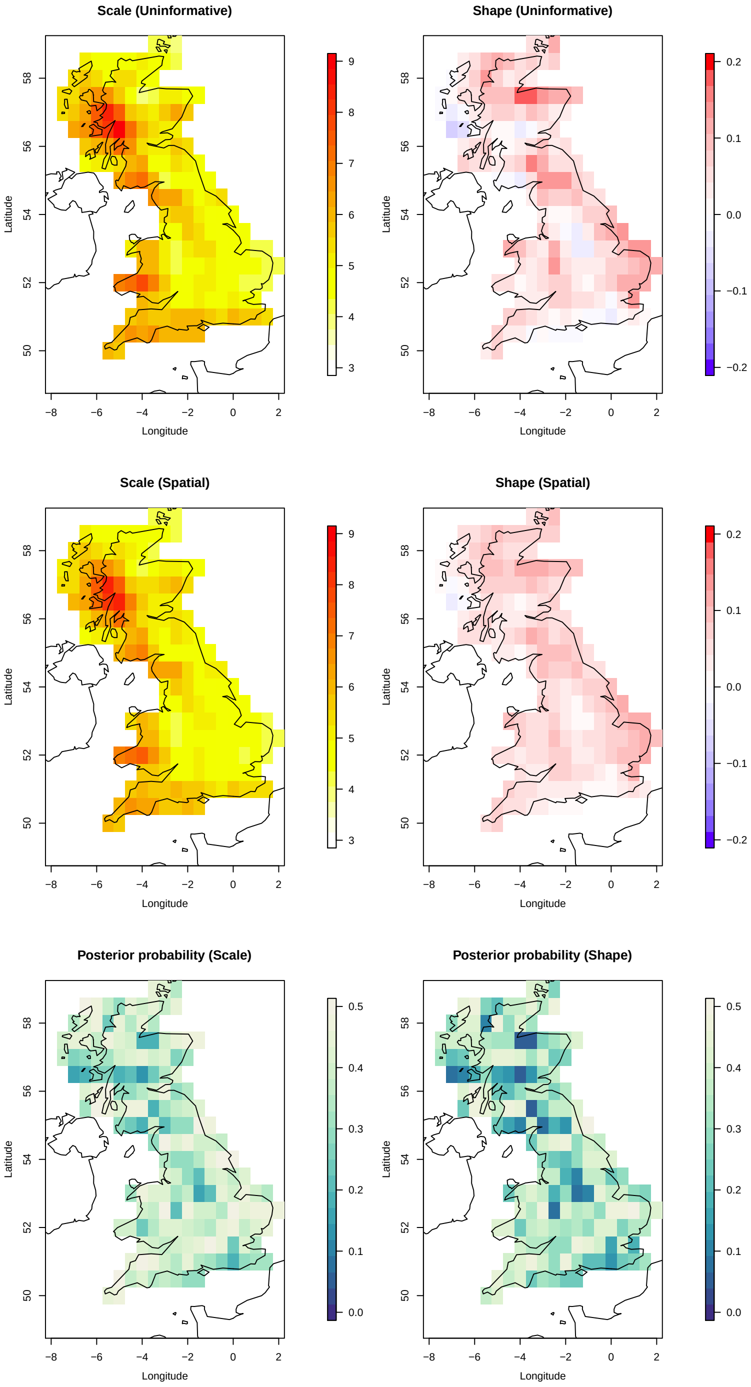


Figure 4.4.1: Parameter estimates of the threshold-dependent scale and shape parameters for all cells from the uninformative model (top) and the spatial model (centre). Also shown (bottom) are the posterior probabilities of the uninformative model estimate occurring under the spatial model.

assumption of constancy is overly simplistic. The shape parameter is typically difficult to estimate given short data records and benefits from the extra information supplied through the spatial prior. While there is a little visible difference in estimation of the scale parameter, the spatial model produces a much smoother surface for the shape parameter. We also investigated the posterior probability of the parameter estimates of the uninformative model (defined by the posterior mean) occurring under the spatial model (see Figure 4.4.1). This shows that while there is little significant difference in estimation of the scale parameter between the two models, the shape parameter is significantly different in multiple cells. In Appendix B.1, the 95% posterior credible intervals for the shape and scale parameters estimated under the spatial model are shown for reference.

The true scientific value of this study is only clear when analysing return levels rather than the GPD parameters themselves. By rearranging the survival function and setting equal to r^{-1} such that

$$n_y \lambda_u \left(1 + \frac{\xi x}{\sigma_u}\right)^{-1/\xi} = r^{-1}, \tag{4.4.1}$$

where n_y denotes the number of observations in a given year, we solve for x to obtain the r-year return level, that is, the value that is exceeded on average once every r years. The posterior distribution of the r-year return level is obtained by applying this function to every MCMC iteration (after burn-in) of the GPD parameter chains. This enables us to extract posterior confidence intervals in a natural way by looking at the quantiles of the estimated posterior.

The r-year return level is a more intuitive quantity for practitioners to analyse, as these estimates can be considered in the design of infrastructure to defend against extreme precipitation events. Figure 4.4.2 shows that the spatial model produces similar return level estimates to that of the uninformative model when the return period is short.

As we move far beyond the range of the data, the dominance of the shape parameter means that the uninformative model can produce very different return level estimates at neighbouring sites. The spatial model performs well in the sense that we move further into the tail in a smooth and realistic way, which is also spatially coherent.

Both RFA and the hierarchical modelling approach pool information across neighbouring locations to induce spatial dependence in parameter estimates and reduce uncertainty of these estimates. The methods differ in the imposition of the spatial process on the GPD parameters. The artificial boundaries arising from the specification of homogeneous regions in RFA can result in very different return level estimates in neighbouring cells that are separated by a boundary (see Figure 4.2.3). The assumption of a constant shape parameter in each region means that the most extreme events can determine the size of the tail over the entire region, which may not be realistic. The hierarchical modelling approach, under the weaker and more intuitive assumption of similarity across neighbouring locations, produces a more realistic return level landscape, representing a smoother extrapolation into the tails of the distribution of precipitation.

As well as producing return level estimates that are spatially smooth, spatial hierarchical models borrow strength across locations, reducing uncertainty in these estimates. Figure 4.4.3 shows the standard deviation of the estimated posterior 10,000-year return level distribution in all cells. In all cases, this measure is reduced by using the spatial model, and in some cases, substantially. The largest standard errors tend to be observed at coastal locations, which is intuitively a consequence of the increased prior variance caused by a smaller number of neighbours compared to inland locations. The uncertainty in these estimates is quantified correctly under the misspecification of the model using the adjusted likelihood (4.3.4). Without this adjustment, the estimated confidence intervals of return levels are narrow and can result in misleading inference.

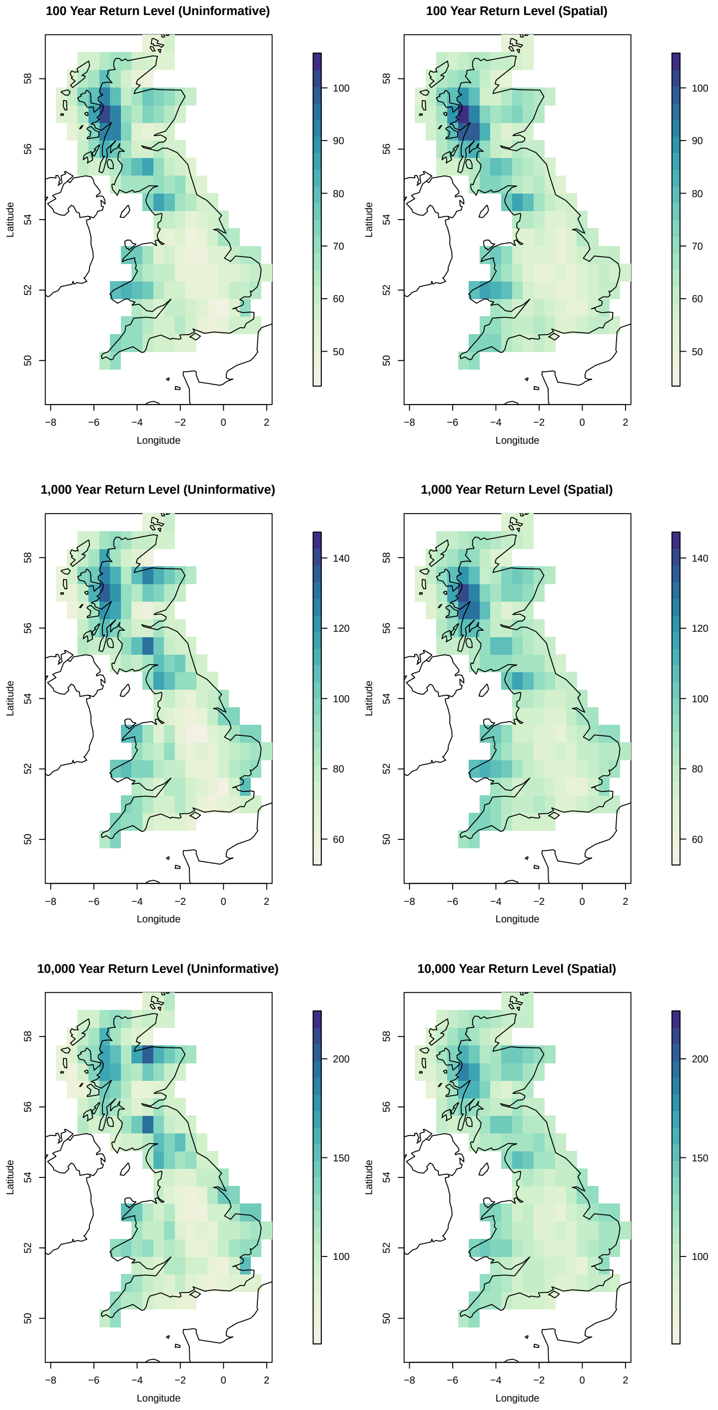


Figure 4.4.2: 100- (top), 1,000- (centre) and 10,000- (bottom) year return level estimates for each cell under the uninformative (left) and spatial hierarchical model (right).

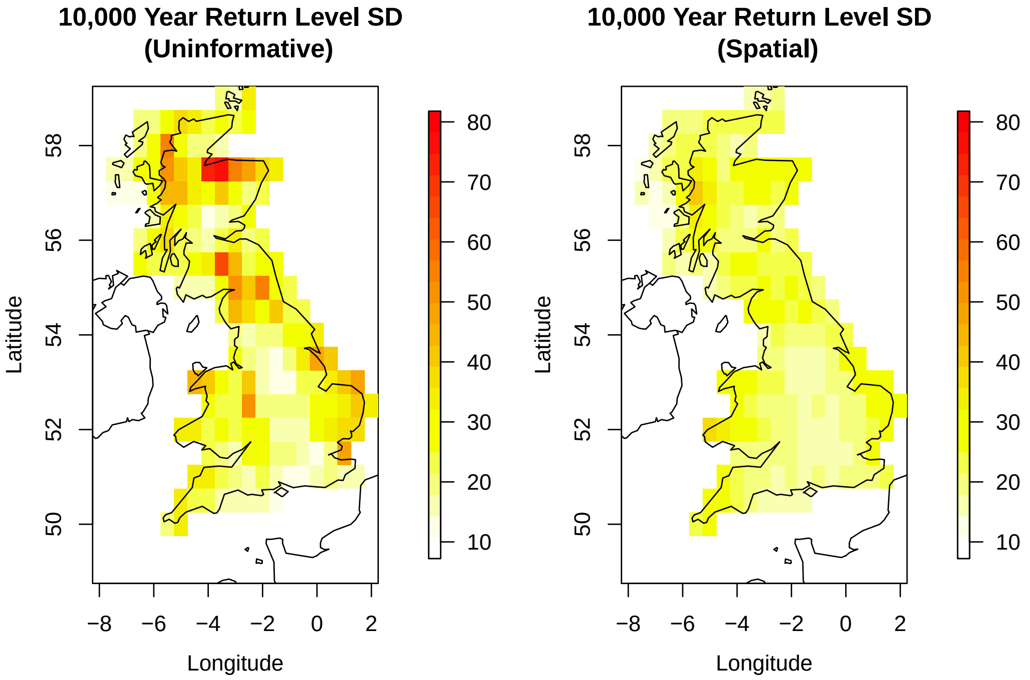


Figure 4.4.3: Standard deviations of the posterior distribution of the 10,000-year return level for the uninformative (left) and spatial model (right).

Table 4.4.1 shows the 95% credible intervals corresponding to three randomly-chosen cells, of the scale and shape parameters and the $10,000$-year return level under the uninformative model, the spatial model when $k = 0.784$, and the spatial model when $k = 1$, which corresponds to a conditional independence model with no adjustment for dependence. This shows that while the spatial model reduces model uncertainty compared to the uninformative model, the adjustment for dependence ensures that credible intervals aren't misleadingly narrow, as is the case when $k = 1$.

The Bayesian paradigm allows us to handle the issue of prediction in a natural way. We can construct the distribution of a future threshold excess - the *predictive distribution*. This incorporates both parameter uncertainty and randomness in future observations.

Table 4.4.1: The estimated 95% posterior credible intervals for three randomly-chosen cells of the scale and shape parameters and the 10, 000-year return level.

Cell	Estimate	Uninformative	Spatial ($k = 0.784$)	Spatial($k = 1$)
1	Scale	(3.660,4.596)	(3.689,4.578)	(3.731,4.559)
	Shape	(0.029,0.196)	(0.035,0.177)	(0.041,0.176)
	10,000 Year RL	(74.629,200.710)	(75.874,178.028)	(77.251,175.099)
2	Scale	(4.347,5.472)	(4.630,5.747)	(4.663,5.674)
	Shape	(0.076,0.251)	(0.045,0.166)	(0.051,0.168)
	10,000 Year RL	(119.850,365.297)	(102.635,209.140)	(106.075,210.225)
3	Scale	(4.284,5.404)	(4.382,5.456)	(4.441,5.361)
	Shape	(0.019,0.198)	(0.020,0.151)	(0.027,0.146)
	10,000 Year RL	(80.742,237.635)	(81.247,172.282)	(83.315,166.768)

For cell j, we have that

$$\Pr(Y_j \leq \tilde{y}_j \mid \boldsymbol{y}) = \int_{\boldsymbol{\theta}_j} \Pr(Y_j \leq \tilde{y}_j \mid \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j \mid \boldsymbol{y})\mathrm{d}\boldsymbol{\theta}_j.$$

We can then define the r-year predictive return level to be the value of $\tilde{y}_j$ that satisfies

$$\Pr(Y_j \leq \tilde{y}_j \mid \boldsymbol{y}) = 1 - r^{-1}.$$

While this is analytically intractible, we can approximate using a Monte Carlo summation of the samples from the estimated posterior distribution. This gives

$$\Pr(Y_j \leq \tilde{y}_j \mid \boldsymbol{y}) \approx \frac{1}{N}\sum_{i=1}^{N} \Pr(Y_j \leq \tilde{y}_j \mid \boldsymbol{\theta}_j^{(i)}),$$

where N denotes the number of MCMC samples after the burn-in has been discarded. The predictive return level estimate can then be evaluated using a standard numerical solver (Coles and Tawn, 1996). The 10, 000-year predictive return level for all cells is shown in Figure 4.4.4. The values have the same spatial pattern as that of the mean of the posterior distribution of 10, 000-year return levels, but the magnitude of the values is slightly higher. This is due to the additional parameter uncertainty being accounted

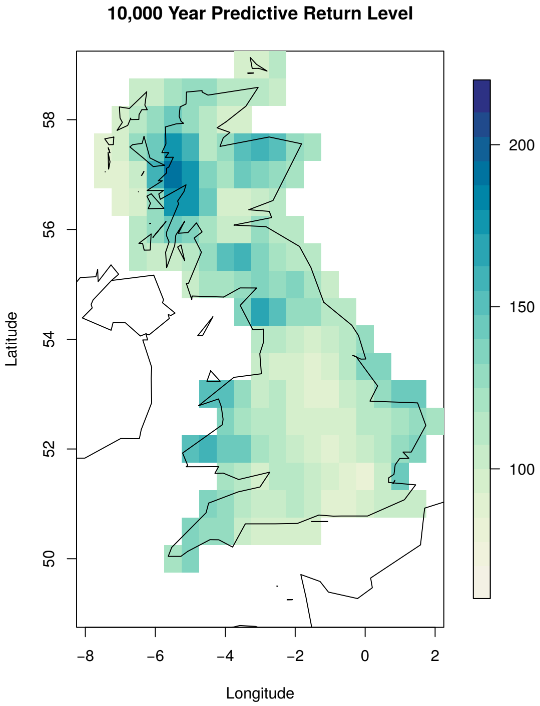


Figure 4.4.4: The 10,000-year predictive return level for all cells.

for in the predictive analysis.

4.5 Discussion

We have described a flexible spatial model for estimating marginal return levels that adjusts inference to account for dependence in the data. Inducing spatial dependence in parameter estimates through a prior model allows us to borrow strength across locations and reduce uncertainty in return level estimates. However, the conditional independence assumption required for Bayesian hierarchical modelling is often false in practical applications. We have presented an approach, through an adjustment to the likelihood, that correctly quantifies model uncertainty under misspecification of the model that accounts for the spatial and temporal dependence in the data. This approach may therefore be of benefit to practitioners who base design specifications on the posterior distribution of return level estimates, as without such an analysis of dependence, it is unclear how exactly to quantify model uncertainty.

Since the model is flexible by nature, there are many improvements that one could make to this analysis. If one was looking at cells on a finer grid, one could include more physical covariates related to the geography of the region. Previous studies have shown that elevation is a key influence on the size of a rainfall event (Cooley and Sain, 2010). Another extension of this work could be the inclusion of a seasonal component in the model. It is widely known that heavy rainfall events in the southeast are largely the result of short-term, localised convective storms that occur in summer. In contrast, heavy rainfall events in northern Scotland, for example, tend to occur as part of a more sustained, synoptic storm system from the Atlantic Ocean. It would be useful to be able to discern between these two systems in the model. This involves introducing a temporally varying covariate, adding an extra dimension to the model. Economou et al. (2014) models sea-level pressure using both spatially and temporally varying covariates in a Bayesian hierarchical model framework. Seasonality can been modelled as a periodic covariate using Fourier series (Jonathan and Ewans, 2011) in the extreme value parameters. Alternatively, Fawcett and Walshaw (2006) models extreme precipitation for multiple sites on a month-by-month basis but induces some temporal smoothness through a seasonal random effect term modelled through a conditional autoregressive prior in time.

It is also possible to consider further development of the spatial structure. Here, we assume a first order neighbourhood structure is sufficient in summarising the spatial dependence inherent in the parameters. Jalbert et al. (2017) proposes using a second-order Markov random field, which gives smoother results than the first-order structure. However, their assessment was over a wider spatial domain and thus their assumption of a higher-order structure was justified.

A drawback of previous studies using the Bayesian hierarchical framework concerned the inference arising from the misspecified conditional independence likelihood, which our approach adjusts to account for dependence in the data. An alternative approach could be to advance beyond the assumption of conditional independence. Recent studies have included a max-stable process in the data layer of a hierachical model (Ribatet et al., 2012; Reich and Shaby, 2012; Thibaud et al., 2016). This approach captures the local spatial dependence of the extremes, and provides a mechanism for simulating realistic fields of precipitation over space.

Chapter 5

A stochastic model for the lifecycle and track of extreme extratropical cyclones in the North Atlantic

5.1 Background

Although the winter climate of western Europe is typically benign, it is often subjected to extreme weather events characterised by strong winds and heavy rainfall from extratropical cyclones that pose economic, safety and environmental risks. Such events include floods and windstorms that have caused mass infrastructural damage, transport chaos and, in some instances, human fatalities. The storm Desmond, which occurred between 3rd and 8th December 2015, displaced thousands of people from their homes in northern England and Scotland, resulting in an estimated £500m worth of damage.

Storm Desmond is an example of a synoptic scale, low-pressure weather system in the North Atlantic Ocean known as an extratropical cyclone. Extratropical cyclones are usually formed as a result of horizontal temperature gradients and develop with a particular lifecycle associated with frontal systems (Shapiro and Keyser, 1990). They can be characterised by the paths of local vorticity maxima they generate, which we refer to as tracks. There has been considerable research into cyclone identification and tracking in reanalysis datasets (Murray and Simmonds, 1991; Hodges, 1995). However, this data record is relatively short and thus provides only a limited estimate of the risk from such weather systems, motivating the need for a model to provide improved information on their possible long-term and extreme characteristics. In particular, we would like to know more about the spatial distribution of these storms so that we can identify the regions with more extreme storm activity at a higher level of confidence. We might also like to assess the likelihood of observing more intense storms than those previously observed, where these might occur, and how long these might last. This chapter proposes a model from which synthetic storm tracks can be simulated and can be used to perform these assessments in a unified and coherent way.

There is limited literature relating to statistical modelling of extratropical cyclones. Sienz et al. (2010) used extreme value methods to analyse the effect of climate change on the impact of the North Atlantic Oscillation (NAO) index on cyclone intensity. Economou et al. (2014) conducted a spatial extreme value analysis of extratropical cyclone pressure minima to estimate probabilities of observing lower-pressure events and the lower endpoint of the distribution of pressure minima. While the model succeeds in capturing the spatial variation of the pressure extremes, it only uses the minimum pressure from a storm track and thus does not account for the spatial and temporal evolution of a cyclone. For example, it may be of interest to practitioners to assess where an extreme storm is likely to propagate. In addition, while the model accounts for the dependence of pressure minima on factors such as its location and the NAO index, it does not explore how it varies relative to the movement of the track. Our approach aims to incorporate both these aspects.

Most developments in track modelling have come from the tropical cyclone literature. Casson and Coles (2000) generated tropical storm tracks by sampling from historical data with random perturbations. Cyclone intensity is modelled dynamically but the history of the process is not incorporated. Rumpf et al. (2007) sampled from kernel density estimates of displacement and direction increments to propagate the track, while Hall and Jewson (2007) use a first-order autoregressive process. Neither incorporate a model for cyclone intensity. This chapter introduces a novel approach to storm track simulation incorporating various properties of extratropical cyclones. This includes the smooth propagation of the track through space, the regional differences between tracks developing at different locations, the tail behaviour of storm intensity and a stochastic termination mechanism.

Our dataset contains storm track locations at 3-hourly time steps with a vorticity measure associated with each point on the track. Storms are tracked over 36 years (1979-2014) from the ERA-Interim reanalysis dataset described in Dee et al. (2011). The identification and tracking of the cyclones is performed following the approach used in Hoskins and Hodges (2002) based on the tracking algorithm described in Hodges (1995). Before the identification and tracking progresses the data are smoothed to a resolution of approximately 2.8° and the large-scale background noise is removed. This reduces the inherent noisiness of the vorticity and makes the tracking more reliable for synoptic scale storms. The cyclones are identified by determining the vorticity maxima in the filtered data. Vorticity is preferred to mean surface-level pressure as it has been found to be more suitable for identifying synoptic systems like extratropical storms (Hoskins and Hodges, 2002). Only vorticity values above a threshold of $1.0 \times 10^{-5} s^{-1}$ are considered. Vorticity measurements are linked together through an initial nearest neighbour search that is then refined by constraints on track displacement and smoothness. Storms with a lifespan of less than one day are not considered. Our analysis is focused on storms in an extended winter period (October-March), eliminating any features that may arise due to seasonal effects and focusing on the time of year when storms are considered to be most intense. We restrict our attention to storms passing over the European domain, in particular the region defined in Figure 5.2.1. We have also removed any Mediterranean storms as these are often influenced by other factors relating to convective behaviour and are not captured well by reanalysis data (Akhtar et al., 2014).

The chapter is structured as follows; Section 5.2 details a comprehensive exploratory analysis carried out to assess the main factors influencing storm movement, intensity and termination. Section 5.3 describes the methodology to be used in model construction. In Section 5.4, submodels for cyclogenesis, propogation and cyclolysis are outlined, motivated by findings from the exploratory analysis. Section 5.5 describes the main results based on simulations from the model, followed by some conclusions and opportunities for further work.

5.2 Exploratory data analysis

An extensive exploratory data analysis was carried out in order to gain some intuition regarding the behaviour of storm tracks in the North Atlantic. As discussed in Section 1, our catalogue of observed storm tracks contains only those that have crossed the region shown in Figure 5.2.1. Our observed set contains $2,944$ storms for the 36 years of data, with approximately 31 observations per storm on average. As these observations are measured at discrete 3-hourly time points, this amounts to the average storm lasting just under four days. Previous analysis of this data (Bengtsson et al., 2006) has shown that these storms tend to begin their existence, known as cyclogenesis, in

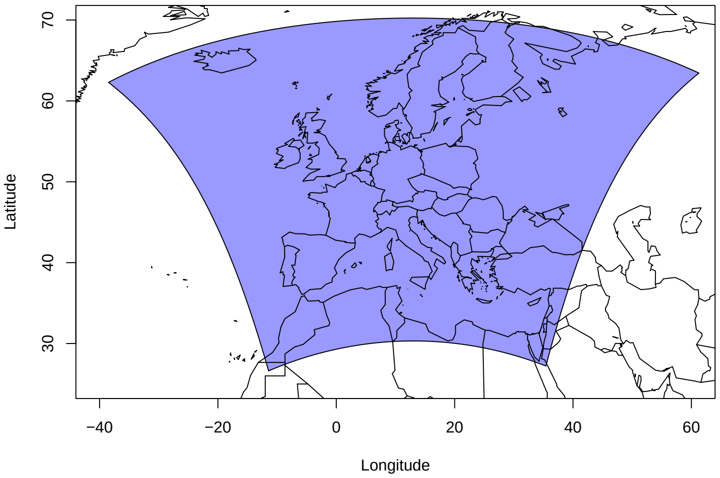


Figure 5.2.1: The region in which the catalogue of storm tracks cross at some point in their lifetime.

a corridor across the North Atlantic from southwest to northeast (see Figure 5.2.2). Cyclolysis regions, where these storms terminate, tend to be located more towards the eastern Atlantic and Europe. Figure 5.2.2 also shows the spatial density of all storm track locations; it identifies distinct regions of storm activity in the mid-Atlantic and in the region between Greenland and Iceland.

We extract components of the storm track in order to explore further the variables that determine storm movement and intensity. We denote the storm location at time t by $\boldsymbol{X}_t = (X_t, Y_t)$, denoting longitude and latitude coordinates at every 3-hourly interval t respectively. Assuming the Earth is spherical, we derive the speed using distance corresponding to the shortest path between two points along the surface of the sphere, commonly known as the "great-circle" distance. We denote the speed of the track between $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t+1}$ by V_t. We choose to model track direction as the initial bearing between $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t+1}$, denoted by $\Theta_t \in [-\pi, \pi]$ measured relative to north. We denote the vorticity at location $\boldsymbol{X}_t$ by Ω_t. This variable structure is shown conceptually in Figure 5.2.3.

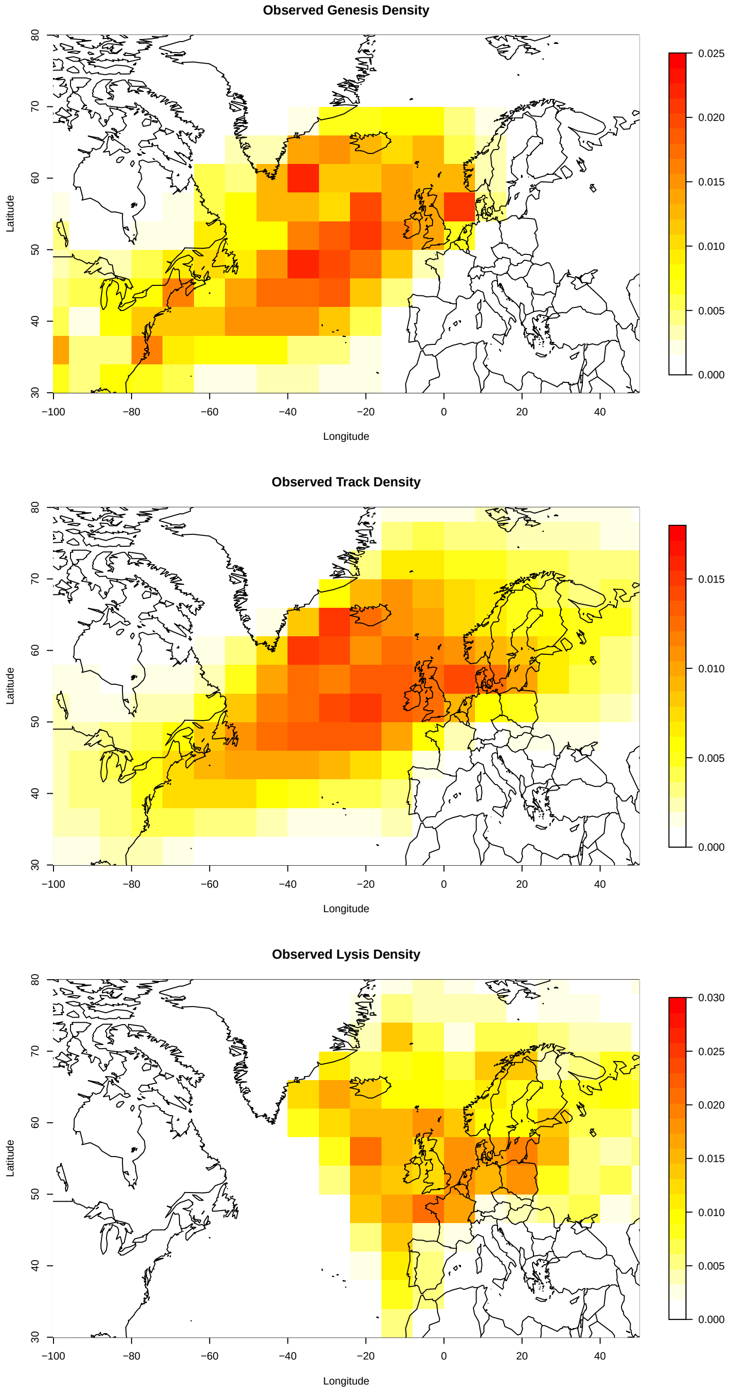


Figure 5.2.2: Spatial densities of genesis, lysis and overall storm track locations in the analysed dataset.

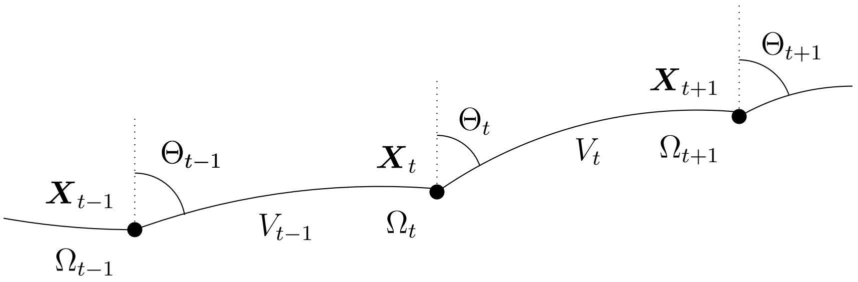


Figure 5.2.3: Conceptual diagram of the speed, direction and vorticity variables extracted from the storm track data.

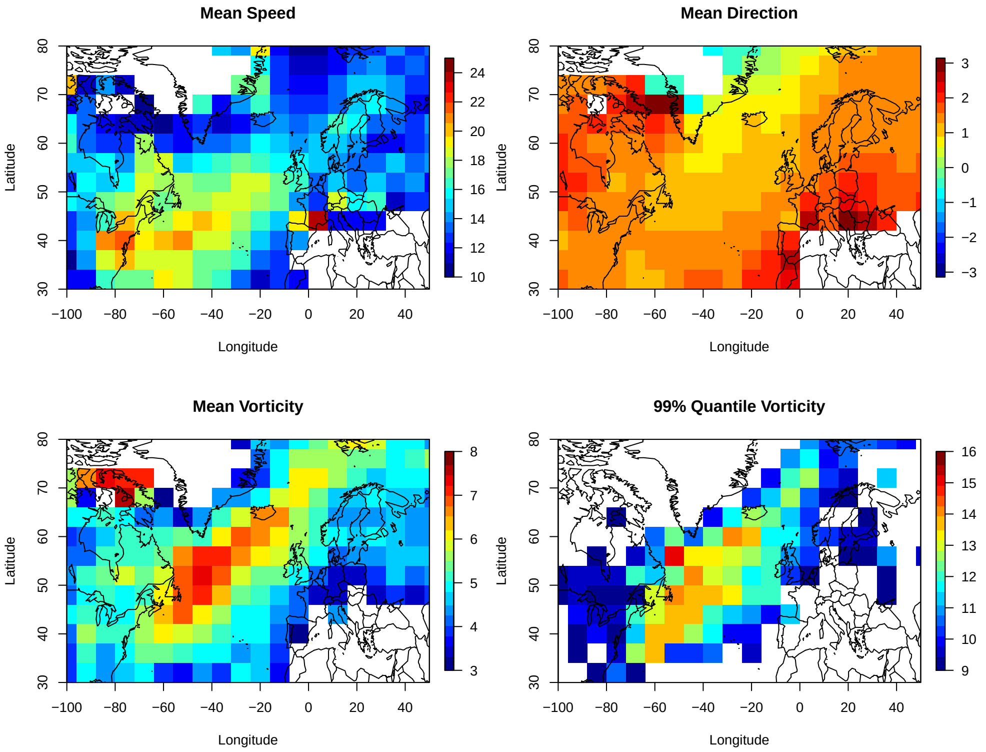


Figure 5.2.4: Mean speed (top left, in m/s), mean direction (top right, in radians), mean vorticity (bottom left, in $10^{-5}s^{-1}$) and 99% quantile of vorticity (bottom right, in $10^{-5}s^{-1}$) in $8° \times 4°$ grid cells over the spatial domain. For illustration purposes, only 99% vorticity quantiles of above $9 \times 10^{-5}s^{-1}$ are shown.

These storm variables have a distinct spatial structure (Figure 5.2.4). The storm tracks tend to begin with an easterly trajectory which becomes more northeasterly as storms move east and to higher latitudes. Speeds, ranging from 0.07 to 51.46 m/s, tend to be highest in an approximate corridor between the eastern coast of the USA and the

United Kingdom, with speeds decreasing smoothly as one moves away from this path. The maximum observed vorticity is $15.89 \times 10^{-5} s^{-1}$ with higher vorticities along a similar southwest northeast corridor albeit further north, identified by Bengtsson et al. (2006) as the highest mean intensity storm regions. Figure 5.2.4 also shows that the strongest of these storms tend to occur in the West Atlantic off the coast of North America.

We investigate the degree of temporal dependence within each variable by examining the partial autocorrelation (PACF) functions for each variable. We identify the order of temporal dependence by the maximum lag at which the PACF is significantly different from 0. However, because of the size of our dataset, the lowest PACF value that we deem significant is very small, so we interpret the PACF plot by eye to identify a practically relevant order. The PACF plots for V_t, Θ_t and Ω_t individually are shown in Figure 5.2.5; they provide evidence that a third-order relationship for speed, direction and vorticity will capture most of the structure in the data.

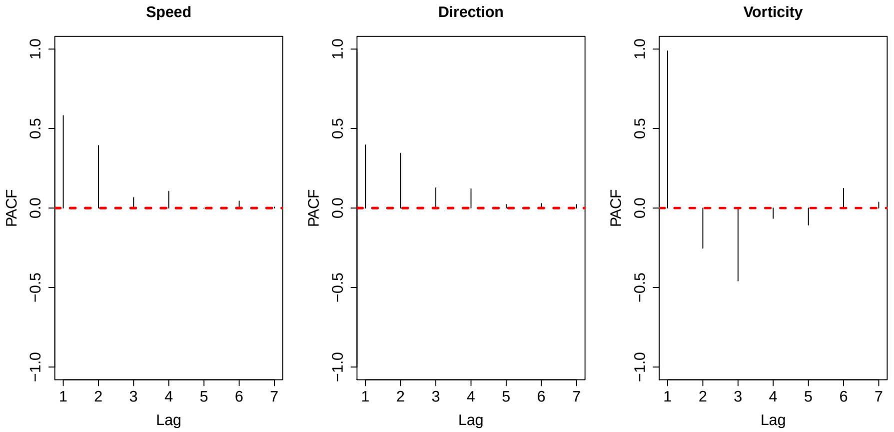


Figure 5.2.5: Partial autocorrelation plots of speed (left), direction (centre) and vorticity (right).

We explore the possibility that storm intensity and storm movement are interlinked, in other words, that speed, direction and vorticity are dependent. Figure 5.2.6 shows that quickly moving and intense storms are more associated with northeasterly/easterly trajectories. It is also clear that a wider range of trajectories are possible when the storm is moving slowly. Figure 5.2.6 also suggests that storms will move more slowly when vorticity is large.

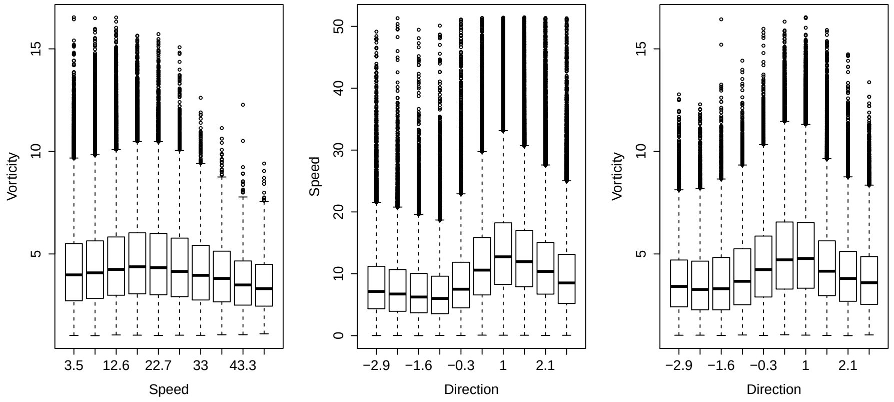


Figure 5.2.6: The mutual dependence of speed, direction and vorticity shown through boxplots defined by intervals of equal length of the variable on the x-axis.

The feature tracking algorithm of Hoskins and Hodges (2002) is designed to assign a smooth path to local vorticity maxima above a threshold of $1.0 \times 10^{-5}s^{-1}$. However, genesis and lysis vorticities are generally above this value (see Figure 5.2.7), which implies either that the storm weakens at a much higher rate or that the tracking algorithm loses the path of the storm. We note that the data may not be representative of the physical termination of a storm, but in the absence of extra information, we design our statistical model to reflect the characteristics of the data. In the context of storm termination, this requires a stochastic mechanism to account for the evident uncertainty in the data. An examination of how the proportion of termination occurrences vary (not shown here) gives evidence to suggest that storms are more likely to terminate if vorticity is low or if the storm is older. Other indicators of storm termination include sharp decreases between consecutive vorticities and the location in space. For example, Figure 5.2.2 shows that storms are more likely to terminate over western Europe than over Scandinavia.

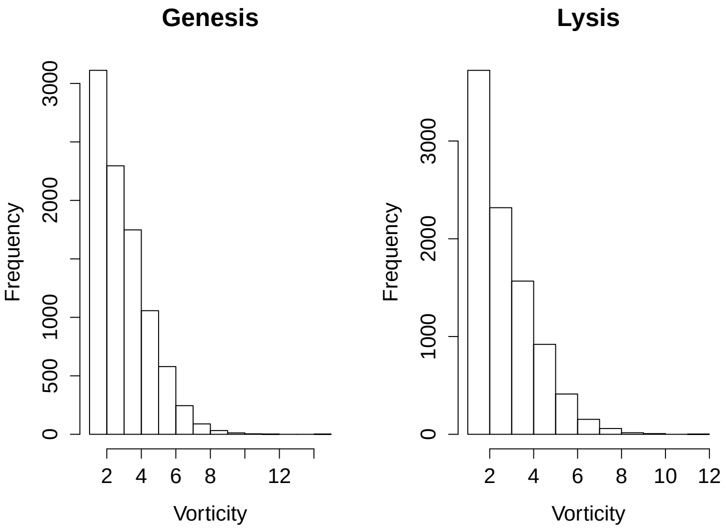


Figure 5.2.7: The distribution of genesis (left) and lysis (right) vorticities.

5.3 Methodology

5.3.1 Introduction

It is evident from the exploratory data analysis in Section 5.2 that extratropical storm tracks are complex systems with many components to be modelled. We use our findings from that analysis to inform a model that represents well the principal physical properties of the storm tracks, in particular, evolution, movement and intensity. We wish to build a model that reflects these processes on a large scale, but also retains properties unique to their spatial location (see Figure 5.2.4). We would like the main features of the track to vary smoothly in time. We would also like to extrapolate in order to derive from our model more intense storms than those observed in the data, but to do so requires a rigorous analysis of tail vorticities. We adopt a simulation-based approach to modelling these storm systems, combining sub-models for genesis, propagation and lysis to produce synthetic storm tracks that have the same statistical characteristics as those observed in the data.

As our approach aims to propagate the storm in time, it is natural to exploit the time series structure of $\{V_t\}$, $\{\Theta_t\}$ and $\{\Omega_t\}$, which control the movement and intensity of a storm. Supported by the exploratory data analysis in Section 5.2, we assume that the multivariate time series $\{(V_t, \Theta_t, \Omega_t)\}$ jointly follows a stationary kth order Markov process. By the Markov property, the distribution of the current value of a process is affected only by the previous k time steps of the process. We define an arbitrary d-dimensional multivariate time series $\boldsymbol{Z}^J_{1:n} = \{Z_{ij} : i = 1, \ldots, n; j = 1, \ldots, d\}$, where Z_{ij} denotes the i-th component of the j-th dimension and n is the length of the time series. We use the notation $\boldsymbol{Z}^J_t$ to denote the tuple at time t. We can write the joint density of $\boldsymbol{Z}^J_{1:n}$ as

$$f_{1:n}(\boldsymbol{z}^J_{1:n}) = f_{1:k}(\boldsymbol{z}^J_{1:k}) \prod_{t=1}^{n-k} f_{k+1|1:k}(\boldsymbol{z}^J_{t+k} \mid \boldsymbol{z}^J_{t:t+k-1}),$$

where $\boldsymbol{z}^J_{p:q} = \{z_{ij} : i = p, \ldots, q; j = 1, \ldots, d\}$, $f_{p:q}$ denotes the joint density function of $\boldsymbol{Z}^J_{p:q}$ and $f_{k+1|1:k}(\cdot \mid \cdot)$ is the conditional density function of $\boldsymbol{Z}^J_{k+1} \mid \boldsymbol{Z}^J_{1:k}$. This assumption simplifies the modelling process as it becomes only necessary to model the joint distribution of $\boldsymbol{Z}^J_{t:t+k}$, which is determined by its marginal distributions and its copula. In some applications, interest lies in estimating probabilities of events beyond the range of the data, for which we draw on methods from extreme value theory.

5.3.2 Marginal modelling

We denote an arbitrary marginal time series component of $\{\boldsymbol{Z}^J_t\}$ by $\{Z_t\}$. Under the assumption of stationarity of $\{\boldsymbol{Z}^J_t\}$, observations in marginal time series $\{Z_t\}_{t=1}^n$ are identically distributed with marginal density function f. A simple choice is to model f

nonparametrically using the kernel smoothed density function $\hat{f}$, such that

$$\hat{f}(z) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right), \tag{5.3.1}$$

where K denotes the kernel function, often chosen to be the standard Gaussian density function, and h is the bandwidth. However, this is known to produce biased estimates in the tails. Instead, for marginal features where the upper tail extremal behaviour is of interest, such as vorticity, we specify a mixture model where the model for the upper tail is motivated through the framework of extreme value analysis. For the remainder of this chapter, we denote the quantity $\boldsymbol{Z}_{i:j} = \{Z_l : l = i, \ldots, j\}$.

Extreme value analysis is often used to model rare occurrences with the aim of estimating probabilities of events beyond the range of available data. Asymptotic limit models are used in practice as finite-sample approximations for estimating the extreme behaviour of a process. The most widely-used approach is to consider excesses above a suitably high threshold. Under weak conditions on Z_t, the distribution of scaled excesses of a threshold by Z_t converges to the generalised Pareto distribution (GPD) (Pickands, 1975; Davison and Smith, 1990) as the threshold tends to the upper endpoint z_F. This model assumes that the limiting result holds exactly for a large enough threshold u. The GPD takes the form

$$\Pr(Z_t - u > z | Z_t > u) = \left(1 + \frac{\xi z}{\psi_u}\right)_+^{-1/\xi}, \ z > 0 \tag{5.3.2}$$

where $c_+ = \max(c, 0)$ and where $\psi_u > 0$ and $\xi \in \mathbb{R}$ denote the scale and shape parameters respectively. The scale parameter ψ_u is threshold-dependent. A negative shape parameter means that the distribution of excesses has a finite endpoint while values of $\xi = 0$ and $\xi > 0$ correspond to exponential- and heavy-tailed distributions respectively. The threshold u is determined using selection diagnostics such as mean residual life plots and checking for threshold stability of ξ (Coles, 2001). For observations of Z_t larger than a chosen threshold u, we replace the kernel estimate defined in (5.3.1) with the GPD model. The marginal model can thus be summarised by

$$F(z) = \begin{cases} \hat{F}(z) & z \leq u \\ 1 - \lambda_u \left(1 + \dfrac{\xi(z-u)}{\psi_u}\right)_+^{-1/\xi} & z > u \end{cases}, \tag{5.3.3}$$

where $\hat{F}(z) = \int_{-\infty}^{z} \hat{f}(\gamma)\mathrm{d}\gamma$, where $\hat{f}$ is defined in (5.3.1), and $\lambda_u = 1 - \hat{F}(u)$ is the rate of exceedance. A censored maximum likelihood approach is used to obtain estimates of the marginal parameters. For more details on inference for the GPD model, see Coles (2001).

5.3.3 Temporal dependence

Under the Markov assumption, the joint distribution of a time series can be determined by a product of conditional distributions determined by the order of the Markov process. This provides a natural mechanism for propagating a storm in time and incorporating the history of the process. A simple choice of model for $\Pr(Z_{t+k} \leq z \mid \boldsymbol{Z}_{t:t+k-1} = \boldsymbol{z}_{t:t+k-1})$, where k is the order of the Markov process, would be the kernel estimate of the conditional distribution function, the formulation of which can be found in Appendix C.1. However, like in the marginal model, this approach poorly captures the temporal dependence structure in the extremes, which is critical when modelling chains of vorticity. This requires an approach for modelling $Z_{t+k} \mid \boldsymbol{Z}_{t:t+k-1}$ in the context of an extreme event, that is, when some functional of $\boldsymbol{Z}_{t:t+k-1}$ exceeds a high threshold u.

Under the assumption of a stationary kth order Markov process, we can model the extremal behaviour of $\{Z_{t+k}\}$ using the joint distribution of $\boldsymbol{Z}_{t:t+k}$. We can use multivariate extreme value analysis to assess the characteristics of joint tail behaviour with separate models for the marginal and dependence structures. Extremal dependence can be summarised by two broad classes determined by the value of χ_τ, where

$$\chi_\tau = \lim_{z \to z^F} \Pr(Z_{t+\tau} > z | Z_t > z), \tag{5.3.4}$$

where $\tau \in \mathbb{Z}^+$ and z^F is the upper limit of the support of the common marginal distribution. For alternative measures of extremal dependence in a time series context, see Davis and Mikosch (2009) and Ledford and Tawn (2003). A value of $\chi_\tau > 0$ refers to the case of asymptotic dependence, where parametric models have been developed with this intrinsic property (Coles et al., 1999). Asymptotically independent models, corresponding to the case when $\chi_\tau = 0$, include contributions by Ledford and Tawn (1996) and Bortot and Tawn (1998). Distinguishing between the two classes is crucial as, for example, applying asymptotically dependent models to asymptotically independent data leads to conservative probability estimates of extreme joint events (Coles et al., 1999). However, in practice, diagnostics for choosing between the two cases are often highly uncertain. The conditional multivariate extreme value approach of Heffernan and Tawn (2004) is more flexible than standard multivariate models as it covers both cases of asymptotic dependence and asymptotic independence. However, this model gives a limiting representation only for $\boldsymbol{Z}_{t+1:t+k} \mid Z_t > u$. To enable a sequential simulation of extremes in time, we draw on methods proposed by Winter and Tawn (2017) that model the extremal temporal dependence structure in $\boldsymbol{Z}_{t+1:t+k}$ provided that $Z_t > u$; this approach is described later in this subsection. All vector calculations in this section are to be interpreted componentwise.

After estimation of the marginal model in equation (5.3.3), it is necessary to transform $\{Z_t\}$ onto common margins to assess extremal dependence after accounting for the marginal structure. Following Keef et al. (2013), we transform onto Laplace margins such that

$$S_t = \begin{cases} \log\{2F(Z_t)\} & \text{if } Z_t < F^{-1}(0.5), \\ -\log[2\{1 - F(Z_t)\}] & \text{if } Z_t \geq F^{-1}(0.5), \end{cases} \tag{5.3.5}$$

where $\{S_t\}$ denotes the standardised series and F is defined in (5.3.3). To explore the conditional distribution $\Pr(\boldsymbol{S}_{t+1:t+m} \leq \boldsymbol{s} \mid S_t > u)$ for large u and integer $m > 0$, we use an asymptotically justified form for this distribution as $u \to \infty$. However, $\boldsymbol{S}_{t+1:t+m}$ requires normalisation so that the limiting conditional distribution is non-degenerate as $u \to \infty$. Heffernan and Tawn (2004) assume that there exist functions $a : \mathbb{R} \to \mathbb{R}^m$ and $b : \mathbb{R} \to \mathbb{R}^m_+$ such that for $s > 0$

$$\Pr\left(\frac{\boldsymbol{S}_{t+1:t+m} - a(S_t)}{b(S_t)} < \boldsymbol{e}_{1:m}, S_t - u > s \,\middle|\, S_t > u\right) \to G_{1:m}(\boldsymbol{e}_{1:m}) \exp(-s), \tag{5.3.6}$$

as $u \to \infty$ with $\boldsymbol{e}_{1:m} \in \mathbb{R}^m$, where $G_{1:m}$ is a joint distribution function that is non-degenerate in each margin. Under weak assumptions on the joint distribution of $\boldsymbol{S}_{t:t+m}$, Heffernan and Resnick (2007) show that componentwise a and b must be regularly varying functions satisfying certain constraints, which for Laplace margins corresponds to each of the components of a (respectively b) being regularly varying functions of index 1 (respectively less than 1). It was found that normalising functions of the simple form

$$a(S_t) = \boldsymbol{\alpha}_{1:m} S_t, \quad b(S_t) = S_t^{\boldsymbol{\beta}_{1:m}},$$

where $\boldsymbol{\alpha}_{1:m} \in [-1, 1]^m$ and $\boldsymbol{\beta}_{1:m} \in [0, 1)^m$, hold for a very broad range of copulas representing a class of functions which enables parsimonious yet flexible modelling. Winter and Tawn (2017) claim that the stationary kth order Markov behaviour of $\{S_t\}$ does not impose any constraints on $\boldsymbol{\alpha}_{1:k}$, $\boldsymbol{\beta}_{1:k}$ and $G_{1:k}$, where k is the order of the Markov process, for $k \leq m$. However, $\boldsymbol{\alpha}_{k+1:m}$, $\boldsymbol{\beta}_{k+1:m}$ and $G_{k+1:m}$, for any $m \geq k+1$, are determined entirely by $\boldsymbol{\alpha}_{1:k}$, $\boldsymbol{\beta}_{1:k}$ and $G_{1:k}$ as a result of the stationary Markov behaviour; specific details of this when $k = 1$ are described in Papastathopoulos et al. (2017).

The parameters $\boldsymbol{\alpha}_{1:m}$ and $\boldsymbol{\beta}_{1:m}$ can be used to identify different types of extremal dependence structure. The case of asymptotic dependence between S_t and S_{t+j} corresponds to the case when $\alpha_j = 1$ and $\beta_j = 0$ for $1 \leq j \leq k$, while the case of asymptotic independence arises when $\alpha_j < 1$. Within the asymptotic independence case, positive dependence occurs with $0 < \alpha_j < 1$ or $\alpha_j = 0$ and $\beta_j > 0$; independence when $\alpha_j = \beta_j = 0$ and negative dependence when $-1 \leq \alpha_j < 0$.

Our model for the conditional distribution of $\boldsymbol{S}_{t+1:t+k}$ given $S_t > u$ is motivated by the limiting form of the conditional distribution (5.3.6), which we assume is valid for a sufficiently high threshold u and $m = k$. Assuming that $\boldsymbol{S}_{t:t+k}$ has a density, we have that

$$\boldsymbol{S}_{t+1:t+k} | S_t > u = \boldsymbol{\alpha}_{1:k} S_t + {S_t}^{\boldsymbol{\beta}_{1:k}} \boldsymbol{E}_{1:k}, \tag{5.3.7}$$

where $\boldsymbol{E}_{1:k}$ is a random variable, independent of t and S_t, with joint distribution function $G_{1:k}$ and joint density $g_{1:k}$. Winter and Tawn (2017) propose an asymptotically motivated heuristic approach to model $S_{t+k} | \boldsymbol{S}_{t:t+k-1}$ when $S_t > u$. Under the assumption that model (5.3.7) holds for $S_t = s_t > u$, it follows that

$$S_{t+k} | (\boldsymbol{S}_{t:t+k-1} = \boldsymbol{s}_{t:t+k-1}) = \alpha_k S_t + {S_t}^{\beta_k} E_{k|1:k-1}, \tag{5.3.8}$$

where $E_{k|1:k-1}$ is a random variable with the same distribution as the conditional distribution of E_k given that

$$\boldsymbol{E}_{1:k-1} = \frac{\boldsymbol{s}_{t:t+k-1} - \boldsymbol{\alpha}_{1:k} s_t}{{s_t}^{\boldsymbol{\beta}_{1:k}}} := \boldsymbol{e}_{1:k-1}.$$

It follows that $S_{t+k+j} | (\boldsymbol{S}_{t:t+k+j-1} = \boldsymbol{s}_{t:t+k+j-1})$, for $j = 1, \ldots$ is also given by equation (5.3.8), provided $S_{t+j} > u$. We adopt this approach to simulate sequential realisations of an extremal kth order Markov process when $S_t > u$. Series generated under this process have negative drifts that ensure the process returns from an extreme state to the body of the distribution, upon which values are generated using the conditional kernel approach outlined in Appendix C.1. Dependence parameters $\boldsymbol{\alpha}_{1:k}$ and $\boldsymbol{\beta}_{1:k}$ are estimated using maximum likelihood under the working assumption that $\boldsymbol{E}_{1:k}$ follows a Gaussian distribution. The distribution function $G_{1:k}$ is estimated using the kernel smoothed distribution function of the values of $\boldsymbol{e}_{1:k}$, which are found by inversion of equation (5.3.8) under the fitted model. For more details on the inference procedure, see Heffernan and Tawn (2004) and Winter and Tawn (2017).

5.4 Simulation model

5.4.1 Cyclogenesis

We construct a model for cyclogenesis conditions using the data observed at the beginning of a storm. In doing so, we would like to model the joint distribution of genesis speed V_0, direction Θ_0 and vorticity Ω_0. The spatial variability of V_t, Θ_t and Ω_t as shown in Figure 5.2.4 is also reflected in the genesis conditions, and thus $(V_0, \Theta_0, \Omega_0)$ should be simulated with respect to genesis location $\boldsymbol{X}_0$. For this reason, we impose an artificial grid on the spatial domain, where each grid cell has dimensions of $8^\circ \times 4^\circ$. This was chosen to be small enough to be able to capture the properties as locally as possible and large enough so that there are enough data to estimate the joint distribution of these properties with sufficient accuracy. We denote Δ_t as the grid cell of the location of the storm track at time t.

We simulate the genesis location from the kernel joint density estimate $\hat{f}(\boldsymbol{x}_0)$, defined in Appendix C.1, where the initial locations $\boldsymbol{x}_0$ are the locations from the observed set of storm tracks discussed in Section 5.1. We then use the conditional kernel approach described in Appendix C.1 to simulate $(v_0, \theta_0, \omega_0)$ jointly from $(V_0, \Theta_0, \Omega_0) \mid \boldsymbol{X}_0 = \boldsymbol{x}_0 \in \Delta_0$. We use a Gaussian density kernel function in all cases. We considered using a von Mises kernel for Θ_0 to ensure continuity of the density function over $[-\pi, \pi]$. However, we used a non-cyclic Gaussian kernel with repeated shifts of 2π in the data which produced similar results. We use a correlated kernel for V_0, X_0 and Ω_0 and an independent kernel for Θ_0, as we believed that a correlated kernel could not sufficiently capture the correlation structure of a cyclic variable. Figure 5.4.1 shows the density of genesis locations from storm tracks simulated from the model and indicates that our genesis model captures the large scale features quite well whilst having some smaller scale differences such as the southern flank extending a little too far and having a maxima over the UK rather than the North Sea (see Figure 5.2.2).

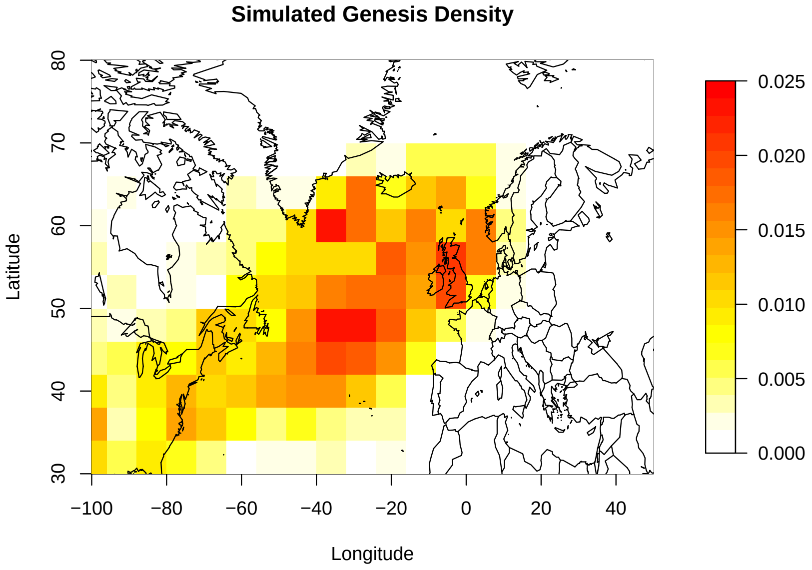


Figure 5.4.1: Density of the genesis locations of a set of synthetic storm tracks simulated from the model.

5.4.2 Propagation

As discussed in Section 5.3, our exploratory analysis supports the assumption that the storm variables determining movement and intensity jointly follow a kth order Markov process. As well as temporal dependence, we would like to incorporate any dependence between variables into the propagation scheme in order to represent the joint properties accurately. The simulated storm track should also reflect the local properties of observed tracks as it moves through space, which the artificial grid introduced in Section 5.4.1 is designed to induce.

Combining these approaches allows us to construct a joint distribution for V_t and Θ_t, which determine the movement of the storm track, conditional on previous states of the variable, states of other variables, and the grid cell of the storm track location. For all times $1 \leq j \leq k$, where k is the order of the Markov process jointly for $\{(V_t, \Theta_t, \Omega_t)\}$, we simulate

$$\begin{aligned} \theta_j &\sim \Theta_j \mid \Theta_{0:j-1} = \theta_{0:j-1}, \boldsymbol{x}_j \in \Delta_j \\ v_j &\sim V_j \mid V_{0:j-1} = v_{0:j-1}, \Theta_j = \theta_j, \boldsymbol{x}_j \in \Delta_j \end{aligned} \tag{5.4.1}$$

When $j > k$, we simulate:

$$\begin{aligned} \theta_j &\sim \Theta_j \mid \Theta_{j-k:j-1} = \theta_{j-k:j-1}, \boldsymbol{x}_j \in \Delta_j \\ v_j &\sim V_j \mid V_{j-k:j-1} = v_{j-k:j-1}, \Theta_j = \theta_j, \boldsymbol{x}_j \in \Delta_j. \end{aligned} \tag{5.4.2}$$

Simulated values v_j and θ_j are obtained from the kernel estimate of the conditional distribution in (5.4.1) and (5.4.2) as discussed in Section 5.3.3 and formulated in Appendix C.1. The exploratory analysis in Section 5.2 suggests that $k = 3$ is an appropriate choice. The dependence between speed and direction is induced through conditioning V_j on Θ_j. As in Section 5.4.1, we used a Gaussian kernel function in both cases. When simulating θ_j, we again considered a von-Mises kernel to ensure continuity of the density function over $[-\pi, \pi]$, but a non-cyclic Gaussian kernel with repeated shifts of 2π in the data produced similar results. We use the simulated values to propagate the storm. Longitude and latitude coordinates $\boldsymbol{x}_{j+1} = (x_{j+1}, y_{j+1})$ are calculated using the formula:

$$\begin{aligned} y_{j+1} &= \sin^{-1}\left(\sin(y_j)\cos\left(\frac{v_j \nabla_j}{R}\right) + \cos(y_j)\sin\left(\frac{v_j \nabla_j}{R}\right)\cos(\theta_j)\right); \\ x_{j+1} &= x_j + \mathrm{Tan}^{-1}\left(\sin(\theta_j)\sin\left(\frac{v_j \nabla_j}{R}\right)\cos(y_j)\cos\left(\frac{v_j \nabla_j}{R}\right) - \sin(y_j)\sin(y_{j+1})\right), \end{aligned}$$

where (x_j, y_j) denote the longitude and latitude coordinates at time j, R denotes the radius of the Earth, taken to be 6371 km, ∇_j denotes the time difference in seconds between time j and $j+1$ and Tan^{-1} denotes the four-quadrant inverse tangent function. If at time $j+1$, a simulated track enters a region such that there has been no observed storm activity in the data in Δ_{j+1} , the track is reverted to time j, giving the algorithm 10 opportunities to find a trajectory towards a grid cell that has observed storm activity. If no such trajectory is found, the storm is terminated. The QQ plots in Figure 5.4.2 show that the simulation model replicates exceptionally well the observed marginal distributions of speed and direction. Figure 5.4.2 also shows the model captures the tendency of storm tracks to move more quickly in a northeasterly direction (see Figure 5.2.6).

5.4.3 Vorticity modelling

The relationship between the vorticity of a storm track and its influence on the weather is complex. Data analysis (not shown here) demonstrates that vorticity is weakly correlated with the maximum wind speed observed in the vicinity of the track, as well as showing evidence that large spatial wind speed events are linked with large vorticities.

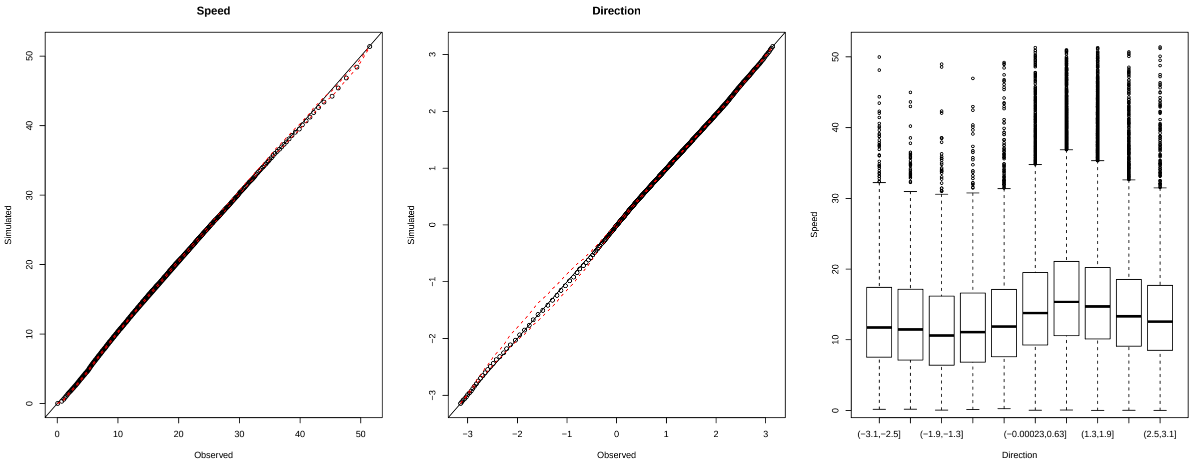


Figure 5.4.2: QQ plots comparing the observed and simulated marginal distributions of speed (left) and direction (middle) each with 95% tolerance intervals and the dependence between simulated speed and direction shown through a boxplot (right). The simulated data is based on one set of storm tracks drawn from the model of the same number as in the observed set.

Therefore it is critical to model carefully the spatial and temporal characteristics of this variable as well as its behaviour in the extremes. While the kernel approach is useful at simulating realistic chains of V_t and Θ_t, the marginal distribution and dependence structure is estimated using the entire series and may therefore lead to bias in the extremes. In practice, we would like to estimate probabilities of observing storms with a vorticity not yet observed. We use asymptotically justified limit models from extreme value theory to estimate these probabilities using the observed extreme events. Our simulation method combines the kernel approach used in (5.4.1) and (5.4.2) with techniques outlined in Sections 5.3.2 and 5.3.3 for tail models with extremal temporal dependence structure.

The exploratory analysis in Section 5.2 shows how the upper tail of vorticity varies with respect to the movement and location of the track. To account for this in an extreme value model, we first use the preprocessing method of Eastoe and Tawn (2009) to transform the data to approximate stationarity. Specifically, we use a Box-Cox-location-scale model of the form

$$\frac{\Omega_t^{\lambda} - 1}{\lambda} = \mu(\boldsymbol{\nu_t}) + \sigma(\boldsymbol{\nu_t}) W_t, \tag{5.4.3}$$

where W_t is assumed to be approximately stationary, λ denotes the Box-Cox parameter and σ and μ are functions of a vector of covariates $\boldsymbol{\nu}_t$. For the purpose of inference on λ, μ and σ, W_t is assumed to be $\mathcal{N}(0, 1)$. Parameter estimates are obtained using maximum likelihood. We fit this model to data within the longitude range $(-60°, 20°)$ and latitude range $(40°, 80°)$ as our interest lies in the extremal behaviour of vorticity in this region. A number of combinations of covariates were considered and the model fit was assessed using likelihood ratio testing. The best-fitting model features functions of latitude, longitude, direction and speed in both μ and σ. This ensures that the variation in large vorticities over space is captured (see Figure 5.2.4) while also ensuring the dependence structure shown in Figure 5.2.6 holds for large vorticity values.

We model the excesses of W_t above some suitably high threshold u using the GPD tail model as discussed in Section 5.3.2. A threshold of $u = 1.5$ is selected, corresponding to the 98.13% quantile of W_t. The maximum likelihood estimates are $\hat{\psi}_u = 0.449$ and $\hat{\xi} = -0.246$. Note that the negative shape parameter estimate implies a physical upper limit to the vorticity distribution. This is consistent with the extremal analysis of mean surface level pressure in Economou et al. (2014), as vorticity and mean surface level pressure tend to behave similarly in the context of extratropical storms (Hoskins and Hodges, 2002). Desired tail quantiles of Ω_t are determined by back-transformation of (5.4.3).

The temporal propagation of vorticity is set out as follows. We describe separately the cases for simulating realisations of Ω_j, for some arbitrary time j, given that the k previous observations of W_j (a function of Ω_j) are in non-extreme and extreme states respectively. First, consider the process when Ω_j is such that the k previous observations of W_j are in a non-extreme state. In particular, we consider two cases. For all times $1 \leq j \leq k = 3$ and $\boldsymbol{\Omega}_{0:j-1} = \boldsymbol{\omega}_{0:j-1}$ such that $\max\{\boldsymbol{W}_{0:j-1}\} < u$, we simulate

$$\omega_j \sim \Omega_j \mid \boldsymbol{\Omega}_{0:j-1} = \boldsymbol{\omega}_{0:j-1}, \Theta_{j-1} = \theta_{j-1}, \boldsymbol{x}_j \in \Delta_j.$$

Next, consider when $j > k$ and $\boldsymbol{\Omega}_{j-k:j-1} = \boldsymbol{\omega}_{j-k:j-1}$ such that $\max\{\boldsymbol{W}_{j-k:j-1}\} < u$. In this case, we simulate

$$\omega_j \sim \Omega_j \mid \boldsymbol{\Omega}_{j-k:j-1} = \boldsymbol{\omega}_{j-k:j-1}, \Theta_{j-1} = \theta_{j-1}, \boldsymbol{x}_j \in \Delta_j.$$

Empirical evidence suggests that vorticity and track speed are approximately independent conditional on the bearing, and since speed is simulated with this conditioning in (5.4.1) and (5.4.2), we believe that simulating ω_j conditional on θ_{j-1} is sufficient to represent the dependence between storm movement and intensity. The conditional distribution $\Omega_j \mid \cdot$ is estimated using the kernel approach described in Appendix C.1.

Next, when Ω_j is such that at least one of the previous k observations of W_j are in an extreme state, we adopt the model of Winter and Tawn (2017) for simulating tail chains under the assumption of an extremal kth order Markov process. In particular, we transform the preprocessed series W_t onto Laplace margins as in (5.3.5), denoting the transformed quantity by S_t. Provided at least one of the last k observations previous to W_j is in an extreme state, we simulate realisations of S_j using the tail chain approach before backtransforming to obtain a realisation of Ω_j. To be precise, let l be the number of consecutive excesses of $\{W_t\}$ above u previous to time j, such that

$$l_j = \max\{i \in \{1, \ldots, k\} : \min\{\boldsymbol{W}_{j-i:j-1}\} > u\}.$$

For example, if S_{j-3} and S_{j-2} are less than u but $S_{j-1} > u$, we use a first order structure to simulate S_j. Therefore l_j represents the order to be used in the simulation of S_j. After determining the order $l = l_j$, we then simulate:

$$S_j = \hat{\alpha}_l S_{j-l} + (S_{j-l})^{\hat{\beta}_l} e_{j|j-l+1:j-1},$$

where $(\hat{\alpha}_l, \hat{\beta}_l)$ denote the maximum likelihood estimates of the dependence parameters and $e_{j|j-l+1:j-1}$ is sampled independently from $\hat{G}_{j|j-l+1:j-1}$. The value S_j is transformed to obtain the preprocessed W_j by inverting equation (5.3.5). Vorticity $\Omega_j = \omega_j$ is then obtained by inverting equation (5.4.3). The QQ plot in Figure 5.4.3 shows that the model captures well both the body of the vorticity distribution and its extremes.

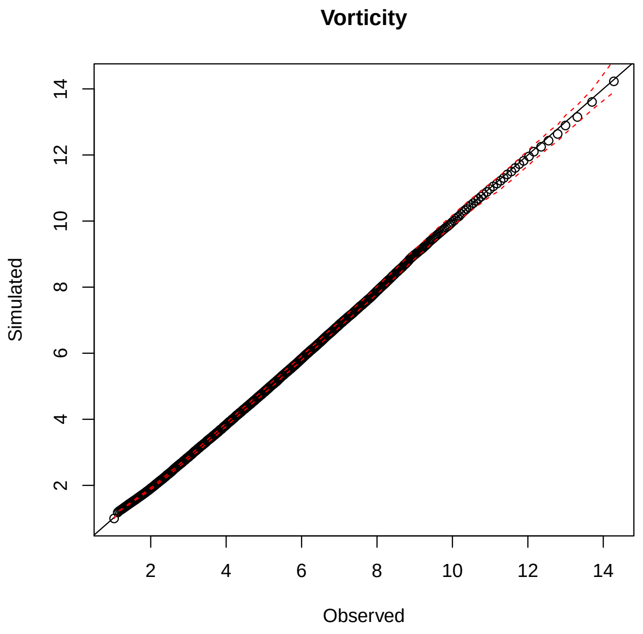


Figure 5.4.3: QQ plot (with 95% tolerance intervals) comparing observed marginal distributions to vorticities from a set of simulated storm tracks of the same number as those in the observed set.

5.4.4 Cyclolysis

The termination of a storm and its track, termed cyclolysis, is not well defined in the observed track dataset, as discussed in Section 5.2. There are a number of instances where a storm terminates at a value of vorticity that is significantly larger than the critical threshold defined by the tracking algorithm, suggesting storms can fade quickly. This motivates the need for a stochastic termination mechanism to be applied to the simulated storm, as discussed in Section 5.1. The exploratory analysis in Section 5.2 suggests several factors influence the risk of termination, including vorticity, age and location. We model these covariate effects using a logistic generalised additive model (Wood, 2006). We estimate a probability of termination at each simulated 3-hourly time step of the storm for $t \geq 8$ in order to replicate the constraint of the tracking algorithm only to consider storms that last for at least 24 hours. The termination mechanism is implemented after the storm track enters the region shown in Figure 5.2.1. Let T_t be a Bernoulli random variable such that:

$$T_t = \begin{cases} 1 & \text{when the storm terminates at time } t \\ 0 & \text{otherwise} \end{cases}$$

So $T_t \sim \text{Bernoulli}(p_t)$, where

$$p_t = \begin{cases} 0 & t < 8 \\ \dfrac{\exp\left\{\sum_{i=1}^{q} s_i(\nu_{i,t})\right\}}{1 + \exp\left\{\sum_{i=1}^{q} s_i(\nu_{i,t})\right\}} & t \geq 8 \end{cases}$$

where s_i is a smooth non-linear function of covariate ν_i with $i \in (1, \ldots, q)$, and q is the number of covariates. The smooth functions are represented by penalised regression splines, where the smoothing parameter is determined using generalised cross validation (GCV) and the model is fitted using a penalised maximum likelihood formulation. For more details on additive models, see Wood (2006).

The effect of the covariates on the model fit was assessed using AIC. The best-fitting model under this criterion consisted of functions of several variables including vorticity, age, longitude and latitude. As hypothesised based on the exploratory data analysis, the fitted model shows that a track is more likely to terminate if the vorticity is low or the storm has experienced a large sudden reduction in vorticity. A track is also more likely to terminate if it is older. The QQ plot in Figure 5.4.4 shows that storm lifetimes are being well captured by the model, while the spatial density of lysis locations compares well with the observed (see Figure 5.2.2). In both cases, storms tend to terminate over the northeast Atlantic and northwest Europe.

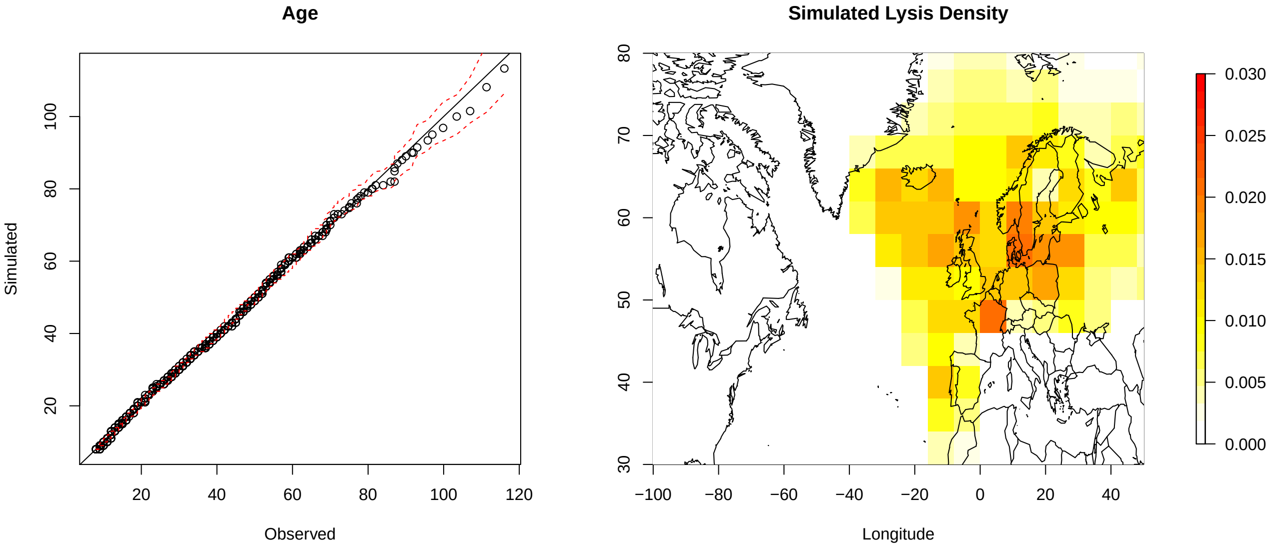


Figure 5.4.4: QQ plot (with 95% tolerance intervals) comparing observed storm lifetimes with lifetimes of storms simulated from the model (left). One unit of age is defined as one 3-hourly interval. Spatial density of storm lysis locations (right) based on a set of storm tracks simulated from the model of the same number at those in the observed set.

5.4.5 Risk analysis

As discussed in Section 5.4.3, the extreme weather impact caused by extratropical storms is complex and warrants further investigation. As the vorticity of a storm is known to be correlated with characteristics of large wind speed events, it is useful for practitioners to be aware of the rate and size of extreme vorticity events in different regions. We can estimate the probability of such events through Monte Carlo simulation. For illustration purposes, we fix our region of interest to be a longitude and latitude range containing the UK; in particular, we define the region $\Gamma = \{(x, y) : x \in (-11^\circ, 2^\circ); y \in (50^\circ, 60^\circ)\}$. To estimate the probability of exceeding a vorticity ω in this region, we calculate

$$\hat{\Pr}(\Omega_t > \omega \mid \boldsymbol{X}_t \in \Gamma) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \mathbb{I}\{\omega_{ij} > \omega, \boldsymbol{x}_{ij} \in \Gamma\}}{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \mathbb{I}\{\boldsymbol{x}_{ij} \in \Gamma\}}, \tag{5.4.4}$$

where $\mathbb{I}$ is the indicator function and $\boldsymbol{x}_{ij}$ and ω_{ij} denote the location and vorticity respectively at the jth time step of the ith storm, n_i denotes the time length of storm i and N denotes the number of simulated storms. One could alternatively characterise a risk measure in terms of the maximum vorticity of a storm, denoted $\omega_{\max}$, such that

$$\hat{\Pr}(\Omega^{\max} > \omega_{\max} \mid \boldsymbol{X}_t \in \Gamma) = \frac{\sum_{i=1}^{N} \mathbb{I}\{\max_j \omega_{ij} > \omega_{\max}, \boldsymbol{x}_{ij} \in \Gamma\}}{\sum_{i=1}^{N} \mathbb{I}\{\boldsymbol{x}_{ij} \in \Gamma\}},$$

which would remove the possibility that multiple excesses could be observed in the same storm. For the purpose of illustration, however, we continue with the characterisation in (5.4.4).

We simulate $N = 84,000$ synthetic storms from our model, which represents approximately $1,000$ years worth of storms, assuming the same average number of storms per year as observed in the data. Figure 5.4.5 shows the spatial density of the synthetic storms; it is clear that the model is capturing the spatial extent of the observed tracks as

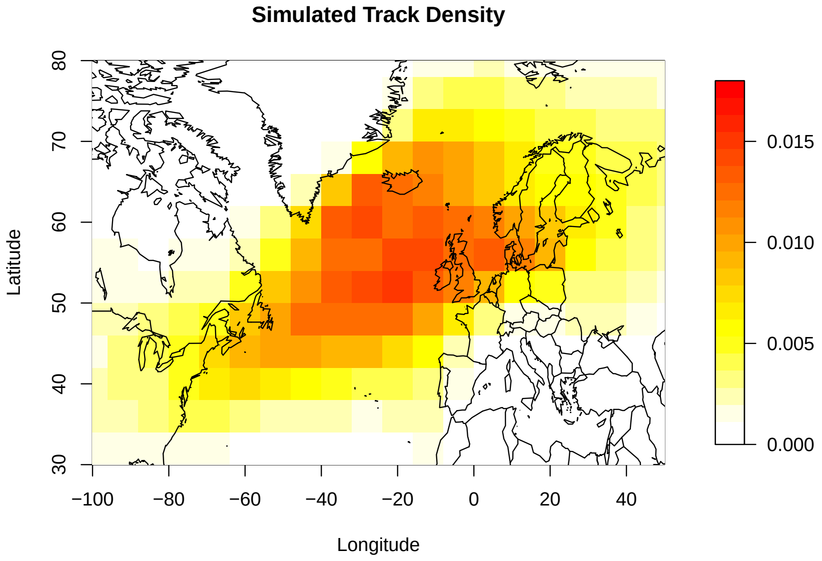


Figure 5.4.5: Spatial density of a set of synthetic storm tracks simulated from the model.

shown in Figure 5.2.2. We assess the model fit by constructing 95% confidence intervals of the spatial density in each cell using a nonparametric bootstrap. Model estimates of the density were within these intervals more than 99% of the time, which we deem sufficient evidence to suggest that the model is performing well.

Using the set of synthetic storms, we can estimate ω_r, the r-year return level, that is, the vorticity value we expect to exceed once every r years in Γ, which we obtain using order statistics of the simulated data. Figure 5.4.6 shows estimates of the r-year return level for Γ from the observed data and using the Monte Carlo simulations from the model, along with 95% confidence interval corresponding to the model estimate, which we obtained using a parametric bootstrap. The model replicates well the tail behaviour of observed vorticities in this region, although estimates of tail risk are slightly higher under our model. Our approach allows us to estimate return levels corresponding to events beyond the range of the data, meaning we can estimate the vorticity corresponding to a 100-year or 1000-year event, for example.

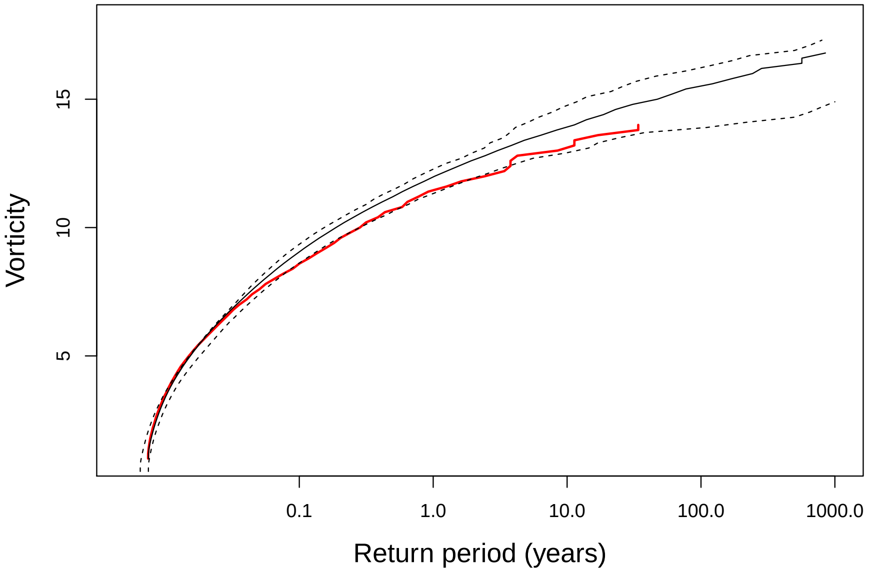


Figure 5.4.6: The r-year return level curve for Γ estimated empirically for a range of r from the observed data (red) and from the Monte Carlo simulation from the model (black). The black dashed lines represent 95% confidence intervals derived using a parametric bootstrap.

One of the most destructive events to impact the UK in the last 35 years was Storm Herta, which caused approximately \$1.5 billion worth of damage after hitting northern Europe in February 1990 and had a maximum observed vorticity of $13.36 \times 10^{-5} s^{-1}$ (Roberts et al., 2014). Through Monte Carlo simulation, we can estimate the return period of this event and assess the relative risks of extreme vorticity events over space. Figure 5.4.7 shows the return period corresponding to an observation of $\omega = 13.36 \times 10^{-5} s^{-1}$ over different grid cells in a region containing the UK on a $4° \times 3°$ grid. Storm Herta reached its maximum vorticity at $\boldsymbol{x} = (1.89°, 55.63°)$. The estimated return period of this event in the cell containing this location (under this particular discretisation of space) is approximately 107 years, with 95% confidence intervals $(38.49, 230.98)$ obtained using a parametric bootstrap. This event is much less rare in the North

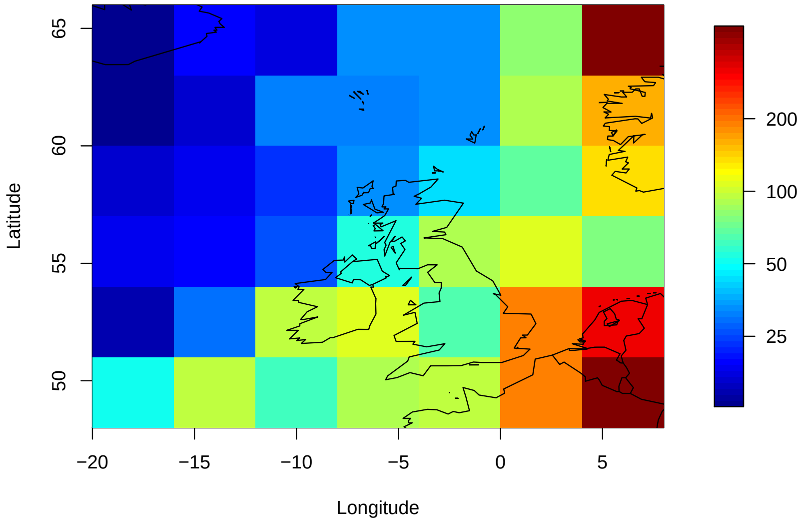


Figure 5.4.7: Return period corresponding to a vorticity value of $\omega = 13.36 \times 10^{-5}s^{-1}$ over space, estimated from the model through Monte Carlo simulation.

Atlantic, where the estimated return period at one location is 18 years $(7.94, 46.19)$. This illustrates the strength of our model in assessing the relative risk of such extreme vorticity events over the spatial domain. Similarly, Figure 5.4.8 shows the 100-year and $1,000$-year return levels estimated from the model under the same discretisation of space, again highlighting the increased probability of observing extreme vorticities in the North Atlantic compared to the UK and mainland Europe.

5.5 Discussion

We have developed a novel approach for simulating extratropical cyclone tracks for the winter half year in the North Atlantic and European domain based on a stochastic model that captures the evolution and structure of observed storm systems. The storm track model is constructed by exploiting the spatio-temporal structure within observed storm tracks to initialise, propagate and terminate an individual storm, producing a

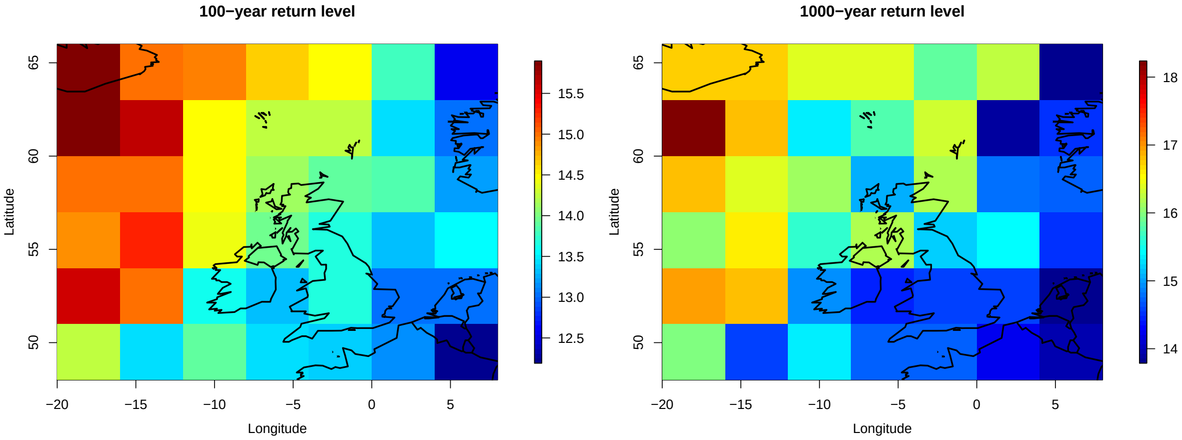


Figure 5.4.8: The 100-year (left) and 1,000-year (right) return level over space, estimated from the model using Monte Carlo simulation.

synthetic track which reflects the key physical characteristics of these weather systems. Track climatologies derived from very large numbers of simulated tracks generated from the model reproduce well the observed spatial variation of the vorticity, lifecycle and tracks of storms.

For practitioners, our model is useful for improving risk assessment related to extreme weather driven by extratropical cyclone activity. The limited observed record means that risk assessments based on empirical evidence are highly uncertain and restricted to observed intensities, with no extrapolation possible beyond the range of the data. By supplementing the observed data with synthetic tracks from our physically-motivated model, probabilities of rare events can be calculated with increased confidence, including events of intensity not yet seen, which can assist in the design of defensive infrastructures. The robust validation of the simulated storm track climatology supports such an approach. We find that the return period of storms with the same vorticity as storm Herta in February 1990, which caused approximately \$1.5 billion worth of damage in northern Europe, is relatively frequent at $10 - 25$ years over the North Atlantic, reducing to ~ 80 years for Scotland and to ~ 200 years for central France.

The effect of a changing climate on the climatology of the North Atlantic storm track and the potential future risk from extreme storms is of pressing concern but one which is also challenged by the sampling issues addressed in this chapter. Earlier studies have indicated a poleward shift in the storm track with decreasing frequency and increasing intensity (McCabe et al., 2001; Bengtsson et al., 2006) whilst more recent studies have indicated that the future response is regionally and seasonally dependent (Zappa et al., 2013) with uncertainty arising from competing physical processes and large internal variability in the climate system (Shaw et al., 2016). Our storm track model provides another tool to assess such changes and future risk through its application to storm track data from future climate simulations.

Improvements to the storm track model, both in terms of its physical realism and utility, include capturing the annual cycle of storm track characteristics. During the summer the preferred path of extratropical cyclones migrates northwards and returns southwards for winter, a feature the track model does not currently capture. Furthermore, large scale modes of atmospheric variability, such as the North Atlantic Oscillation (NAO) are known to influence the path and frequency of extratropical storm tracks and the subsequent risk of extreme rainfall (Brown, 2017). The NAO, an anomalous dipole pressure pattern between the Icelandic low and the Azores high, significantly modifies the strength of the large scale westerly flow and location of the storm tracks. A positive NAO is associated with increased cyclonic activity in Northern Europe, while southern Europe is typically susceptible to more storm events during a negative NAO phase (Mailier et al., 2006). Exploring the annual cycle of track behaviour and their dependence an NAO represent interesting avenues for future work, enabling simulation of synthetic tracks specific to season and NAO phase.

Chapter 6

Modelling the spatial extent and severity of extreme European windstorms

6.1 Introduction

While the winter climate of the United Kingdom and northern Europe is typically associated with mild, wet weather that poses little infrastructual or societal risk, there has been an increased focus in recent years on the impact of windstorms in this part of the world. These events are often the consequence of extratropical cyclones, and are directly linked to the occurrence of flooding, transport chaos and considerable damage to infrastructure. Roberts et al. (2014) describe a comprehensive catalogue of European windstorms in the period 1979-2012 that contains extensive information related to the meteorology and monetary impact of each storm. Storm Daria, which occurred in January 1990, is believed to be the most destructive windstorm in this period, with an estimated insured loss of $8.2bn.

Windstorms are often a consequence of the passage of extratropical cyclones. Extratropical cyclones are synoptic-scale weather systems associated with low pressure that generally originate in the North Atlantic and progress northeasterly towards northern Europe. These systems can be characterised by paths of local vorticity maxima, which we refer to as tracks. Cyclones are typically formed as a result of horizontal temperature gradients and evolve according to a particular lifecycle with associated frontal systems (Shapiro and Keyser, 1990). Windstorms tend to occur along the boundary where cold and warm air masses converge, commonly referred to as a weather front (Hewson and Neu, 2015). A large body of research exists on cyclone identification, storm tracking and feature extraction in reanalysis datasets (Murray and Simmonds, 1991; Hodges, 1995), which produce good approximations of how a track develops in space and time. However, methods don't currently exist to track the evolution of windstorms relative to the cyclone centre, and how to quantify these in a robust way.

The data record is relatively short with regard to storm tracks, and even more so with regard to windstorms, which motivates the need for a statistical model to provide extra information about the possible extreme, long-term characteristics of windstorms that are generated by the extratropical cyclone. In particular, we would like to assess the likelihood of observing more severe storms than those observed, where these might occur, and how large the spatial extent of the event might be. We would also like to assess the joint risk of multiple locations experiencing the same windstorm event. This is particularly difficult to model as the sites experiencing the event are largely determined by the position of the track relative to these sites. We therefore require a method that accounts for the spatial variability of the storm track in quantifying the risk associated with windstorms at multiple sites. This chapter describes an approach to simulate synthetic windstorms that statistically represents the winds within the cyclone, which can be used in practice to assess the marginal and joint risk of these weather systems over Europe while accounting for the varying probability of storm tracks over the region.

A common approach to statistical modelling of extreme wind speeds is to use techniques from extreme value analysis, which use models built on asymptotic arguments to estimate probabilities of events beyond the range of the data. In meteorological applications, such probabilities are commonly used by practitioners to design infrastructure appropriately to defend against the natural hazard being studied. The most widely-used approach in extreme value analysis is to consider excesses above a suitably high threshold. Consider a sequence of independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$. Under weak conditions on the X_i, the unique, non-degenerate distribution that the scaled excesses of a threshold by X_i converge to, as the threshold tends to the upper limit x_F of X_i, is the generalised Pareto distribution (GPD) (Pickands, 1975; Davison and Smith, 1990). We make the assumption that this limiting result holds for a large enough threshold u. The GPD takes the form

$$\Pr(X_i - u > x | X_i > u) = \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi}, \; x > 0 \tag{6.1.1}$$

where $c_+ = \max(c, 0)$ and where $\sigma_u > 0$ and $\xi \in \mathbb{R}$ denote the scale and shape parameters respectively. The shape parameter is invariant to the choice of threshold but the scale parameter is threshold-dependent. The threshold u is typically determined using some standard selection diagnostics (Coles, 2001) such as ensuring that the parameters are stable with respect to the threshold for all threshold choices larger than u.

There have been numerous studies using extreme value models to estimate extreme wind speeds (Coles and Walshaw, 1994; Fawcett and Walshaw, 2006; Ribatet, 2013). However, these models have no consideration of the physical processes generating the extremes. Some recent studies have, however, modelled extreme winds in the context of an extratropical cyclone. Della-Marta and Pinto (2009) use a GPD model to assess changes in extreme wind intensity under climate change scenarios, which led to results showing that that the frequency of intense wind events in Europe is predicted to increase. Sienz et al. (2010) extended this approach to model the effect of the North Atlantic Oscillation (NAO) index. Bonazzi et al. (2012) modelled the tail dependence of wind speeds between locations over Europe using a bivariate extreme value copula and found dependence to be greater in the west-east direction, which is consistent with the passage of extratropical cyclone tracks over Europe. More recently, Youngman and Stephenson (2016) use extreme value analysis coupled with a geostatistical model to capture the spatio-temporal development of windstorms over Europe, but again direct influence of the storm track is not accounted for.

The approaches described above share a common philosophy in that windstorms are modelled in an Eulerian frame of reference. In fluid mechanics, this refers to the scenario whereby an observer measures observations of a process at a fixed location while the process, e.g., a windstorm, passes over. This is the most common approach to statistical modelling, with the advantage that one can build large spatial data sets with time series at each location being observed. However, if one's concern is focused more on modelling the evolution and influence of the process itself, a Lagrangian frame of reference is required. Historically, Eulerian approaches have been used to model observational data from sources such as weather gauges. However, recent advances in climate modelling have resulted in the increased availability of high-resolution datasets that are spatially and temporally complete, in which large-scale processes can be modelled in a Lagrangian framework. Previous climate studies on extratropical cyclones in a Lagrangian framework include Catto et al. (2010), Rudeva and Gulev (2011) and Dacre et al. (2012). This approach to modelling requires following the process and collecting observations as it moves through space and time. This is a natural framework on which to build a model for windstorms, as it allows us to explore and model the behaviour of extreme winds relative to the storm track as it moves across the North Atlantic. This is especially useful as many sites have not experienced events of a large spatial extent or magnitude during the observational record, possibly as a consequence of the path storms have taken.

In Chapter 5, a model is described such that synthetic storm tracks are generated that replicate the climatology of extratropical cyclones in the North Atlantic. This model also allows the generation of larger extreme storms, with regard to vorticity, than previously observed. In this chapter, we present an extension to this work that allows for the simulation of synthetic wind events relative to the storm tracks generated by the model of Chapter 5. In particular, we are interested in modelling in a Lagrangian framework the most damaging wind events, which we refer to as windstorms, relative to the centre of an extratropical cyclone. We first describe a model for the area affected by strong winds in the vicinity of the storm centre, which we refer to as a footprint. We represent the footprint as an ellipse and model the evolution of its characteristics through time relative to the storm centre. Next, we describe an approach for modelling the magnitude and spatial distribution of the extreme winds within the footprint. These approaches allow us to generate a series of footprints for multiple windstorms associated with the synthetic storm tracks of Chapter 5, providing a method for estimating the risk associated with extreme windstorms over the North Atlantic and Europe.

The chapter is structured as follows. In Section 6.2, we introduce the data and our methods for extracting the features of the windstorm from the data. We also detail an exploratory analysis of these features in this section. We discuss our modelling strategy in Section 6.3, introducing our approaches to modelling the evolution of the windstorm footprints and the extreme winds within the footprints. We examine some key results in Section 6.4, before concluding in Section 6.5 with some discussion.

6.2 Windstorm definition and exploratory analysis

6.2.1 Data description

As in Chapter 5, our work uses storm track data covering the North Atlantic and European domain. In particular, our dataset consists of storm track locations at 3-hourly intervals with an associated vorticity measure representing the strength of the storm. Storms are identified and tracked over the period 1979-2014 from the ERA-Interim reanalysis dataset (Dee et al., 2011) using a feature extraction approach outlined in Hoskins and Hodges (2002) based on the tracking algorithm introduced in Hodges (1995). We restrict our attention to the set of storm tracks produced during an extended winter period (October-March), when storms are widely regarded to be most intense. We exclude Mediterranean storms as these often arise as a result of convective behaviour in the atmosphere and are not captured well by reanalysis data (Akhtar et al., 2014). We denote the longitude and latitude coordinates of the storm track at time t by Lon_t and Lat_t respectively. The vorticity associated with the track at $(\mathrm{Lon}_t, \mathrm{Lat}_t)$ is denoted by Ω_t.

Our model is based on wind speed data from the EURO4 numerical weather prediction model (Standen et al., 2017), which is a downscaled version of the ERA-Interim reanalysis dataset. Data are available on a 4 km spatial resolution over Europe and part of the North Atlantic, amounting to $1,100,000$ cells (see Figure 6.2.1). Values are obtained at hourly intervals over the period 1979-2014. We linearly interpolate the storm track locations and vorticity within each 3-hourly interval to match the hourly temporal resolution of the wind speed data. We select only the wind speed fields at times corresponding to the set of storm tracks. In particular, as we are looking to model the effect of the storm track on the spatio-temporal evolution of wind speeds in the vicinity of the track, we isolate the field of interest as a square-shaped region centred at the

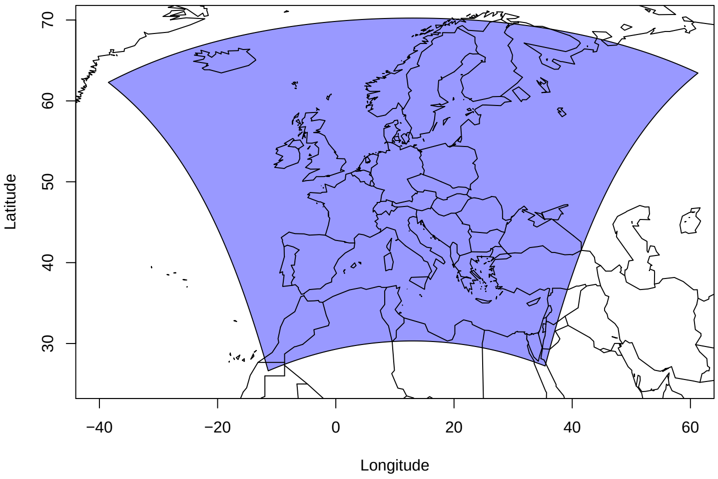


Figure 6.2.1: The region corresponding to the availability of data from the EURO4 numerical weather prediction model.

storm centre with sides of approximately $1,600$ km in length (see Figure 6.2.2, which in the left panel shows such a region at a time step when storm Daria was located over the UK). We believe that this field is large enough so that the extreme winds generated by a windstorm are sufficiently captured.

6.2.2 Marginal model

Initial investigation of the data confirms, as expected, that winds over the sea are markedly stronger than those over land (see Figure 6.2.2, left panel). This is largely due to open water exerting significantly less drag on the atmosphere in contrast with the land surface, orography and man-made structures that impede strong winds. The contrast in scale over land and sea, and to a lesser extent, over low-lying and high-lying land, motivates a standardisation of wind speeds in each cell to have a common marginal distribution. Let $X(s,t)$ be a random variable denoting the wind speed in cell s at time t, for $s = 1, \dots, 1,100,000$. We propose a marginal model for $X(s,t)$ of the form

$$F_s(x) = \begin{cases} \hat{F}_s(x) & x \leq u_s \\ 1 - \lambda_{u_s}\left(1 + \xi_s \dfrac{x - u_s}{\sigma_{u_s}}\right)^{-1/\xi_s} & x > u_s, \end{cases}, \tag{6.2.1}$$

where $\hat{F}_s$ denotes the empirical distribution function for realisations of $X(s,t)$. For realisations above u_s, the GPD in (6.1.1) is used as a conditional model for excesses above u_s, with cell-specific parameters (σ_{u_s}, ξ_s). To undo this conditioning, a third parameter λ_{u_s}, denoting the probability of an exceedance of u_s, must be specified. Parameter stability plots were checked at a number of cells, which indicated that a threshold corresponding to the 98% quantile would be a good choice for all cells. We therefore choose this quantile to be the cell-specific threshold in each cell. Parameter estimates are obtained using maximum likelihood techniques. We note that we do not attempt to impose spatial smoothness on the form of F_s.

Figure 6.2.3 shows the parameter estimates of the GPD corresponding to each cell in the region over Europe as shown in Figure 6.2.1 along with the threshold, which corresponds to the 98% quantile in each cell. This shows explicitly the contrast in wind speed magnitudes between locations on land and sea. This contrast is also reflected in the estimation of the scale parameter, but the shape parameter exhibits no such contrast between land and sea, with most estimates occurring in the region $(-0.2, 0)$, indicating that the distribution of wind speeds has a finite endpoint in general. The numerical maximisation algorithm used to obtain the parameter estimates mostly converges, however there are certain regions that exhibit unusual behaviour. For example, the 98% quantile corresponding to the threshold is a lot higher in areas over Iceland than other land locations, while the Italian Alps see unusually high estimates of the shape parameter. As weather variables are not represented well by reanalysis data in regions of high orography, as evidenced by these two particular locations, we exclude regions with orography $>$ 500m in our analysis. Marginal bias of this type has been

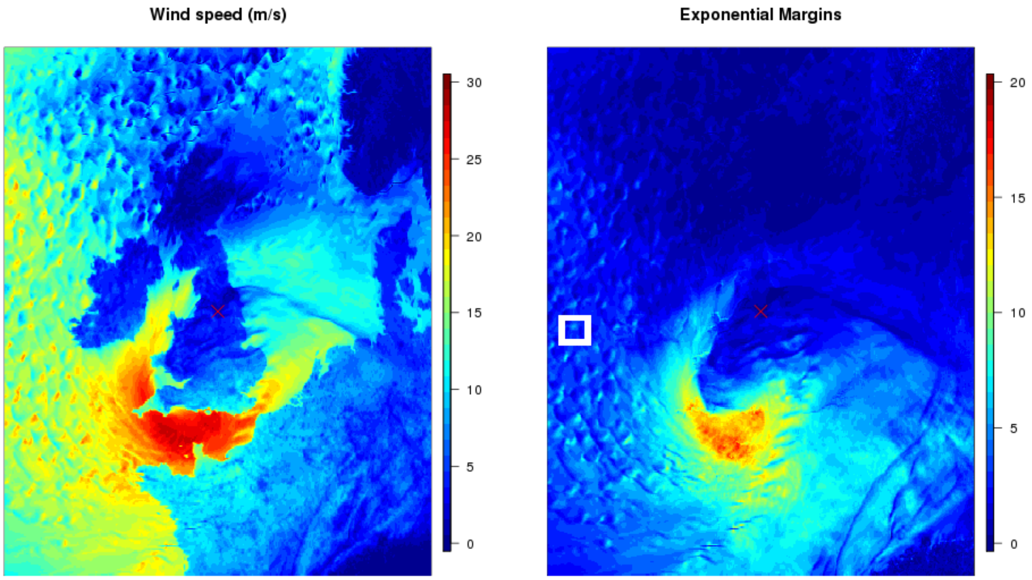


Figure 6.2.2: Wind speeds, in m/s, at 3pm on January 25th, 1990 in the vicinity of Storm Daria (left) and standardised onto Exp(1) margins (right). The storm centre is represented by the cross. The white box contains an example of a localised convective event. Land/sea borders are not explicitly shown on in the left panel, but can be seen due to the contrast in magnitude between winds over land and sea.

studied, with recalibration methods proposed in Howard and Clark (2007), which offers a possible extension of our method.

We propose a marginal standardisation to unit exponential Exp(1) margins using a probability integral transform using the marginal model (6.2.1). We define $X^E(s,t)$ to be the relative wind speed on Exp(1) margins in cell s, such that for all cells and all times

$$X^E(s,t) = -\log\{1 - F_s(X(s,t))\},$$

where F_s is defined as in (6.2.1). We use the term relative wind speeds to describe $X^E(s,t)$ as this quantity defines wind speeds relative to the marginal characteristics of cell s. Figure 6.2.2 (right panel) shows the effect of the standardisation on one time step of Daria. In particular, we see spatially correlated values of high relative wind

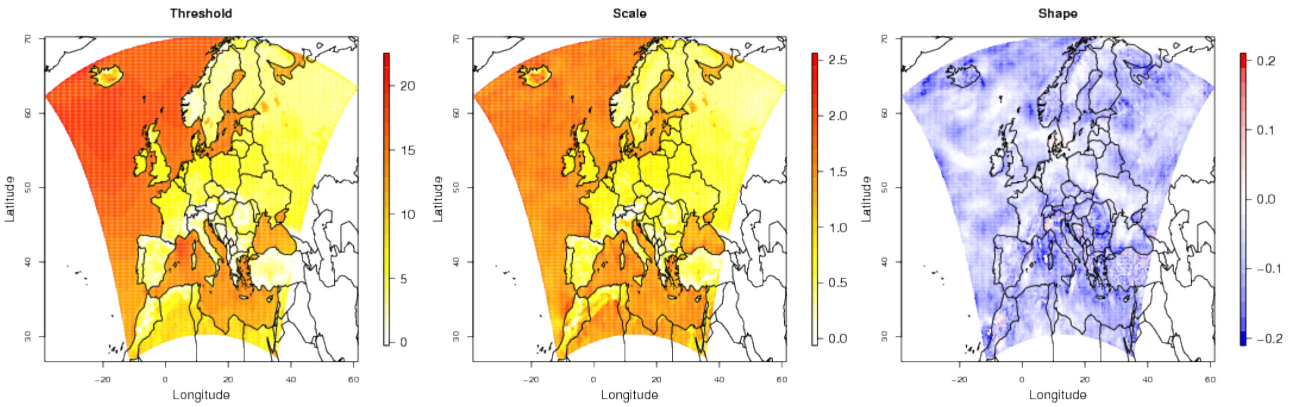


Figure 6.2.3: The extreme value threshold (left) used in the specification of the GPD model, along with parameter estimates of the scale (centre) and shape parameters (right) for each cell.

speeds over both land and sea as a result of the transform and the land/sea contrast is almost entirely removed. This approach provides some intuition regarding the shape of the windstorm event without the influence of the marginal characteristics at each location, which should allow for a simpler approach to modelling the spatial extent of the event. In this sense, our approach is based on a copula modelling strategy over space (Joe, 1997).

6.2.3 Feature extraction

Figure 6.2.2 shows that, as well as the large band of strong relative winds clustered near the storm centre, small localised fragments of high relative wind speeds are visible on the western edge of the footprint. Such events are due to localised convection and not due to the larger scale dynamics of the storm. Since we believe them not to be directly linked with the extratropical storm system, we are not concerned with modelling these features. Our work is focused on modelling the features of an extratropical cyclone that occur on larger spatial scales and have the potential to produce much larger impacts than localised convective events. This motivates developing some methodology to extract the main features of interest from the standardised fields, in particular, the larger cluster of high relative wind speeds. We apply a spatio-temporal Gaussian filter (Nixon and Aguado, 2012) to each field, which removes the effect of the small-scale convective events. From the locations for which the filtered data exceed some arbitrary threshold v, we extract the relative wind speeds. Relative winds corresponding to the other locations are set to zero, thus masking convective wind events and other non-extreme winds.

In many situations, this filtering step masks the entire field, which we interpret as there being no significant windstorm activity, that is, the windstorm is in an inactive phase. When this isn't the case, the filtered footprint consists of at least one cluster of cells corresponding to non-zero relative wind speeds. Since our objective is to model the primary effects of the storm on the surrounding winds, we extract the largest cluster of cells in terms of size and use this to define the features of the windstorm event. We do this using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm (Ester et al., 1996), which recursively groups cells into distinct clusters based on adjacency to neighbouring cells. Unlike other standard clustering approaches, this can be achieved without having to specify the number of clusters in advance. We then extract the largest cluster of cells in size as defined by the algorithm, an example of which is shown for storm Daria in Figure 6.2.4 (left panel).

Since we are interested in modelling how the storm influences the magnitude and spatial extent of wind speeds, we construct a set of variables to adequately represent and summarise these characteristics. We therefore impose the shape of the region obtained as a result of the filtering and clustering steps to take the form of an ellipse. We believe an ellipse is a suitably flexible shape to choose as it can describe well the position (by the centre of the ellipse), size and shape (by the area defined by the semi-major and semi-minor axes) of an event. To define this ellipse, we use Khachiyan's optimisation algorithm for fitting the minimum-area ellipse enclosing a set of cells (Moshtagh, 2005;

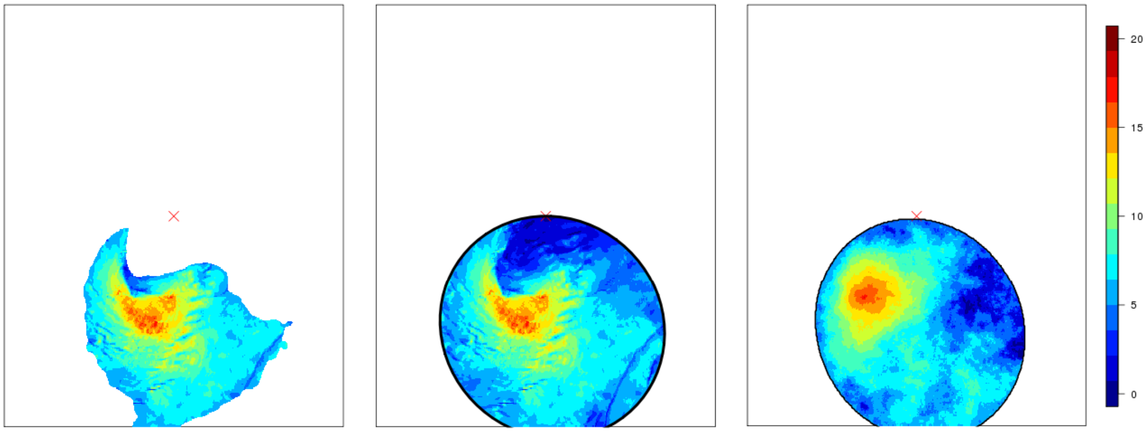


Figure 6.2.4: The largest cluster of relative winds (left) after applying the spatio-temporal filter and DBSCAN clustering to the field in Figure 6.2.2, the ellipse-shaped footprint determined by Khachiyan's algorithm (centre) and an example of simulating wind speeds from our model within this ellipse (right).

Todd and Yıldırım, 2007). Specifically, consider a set S containing m locations in 2-dimensional space such that $S = \{s_1, \ldots, s_m\} \in \mathbb{R}^2$. We can define an ellipse by the set $\mathcal{E}$ such that

$$\mathcal{E} = \{s \in \mathbb{R}^2 : (s - c)^T E(s - c) \leq 1\},$$

where c is the centre of the ellipse and E is a positive-definite matrix. The area of $\mathcal{E}$ is $\{\det E^{-1}\}^{1/2}$. Therefore, to determine the ellipse of minimum area containing the cells in S, we must find a vector c and positive definite matrix E which minimises $\det E^{-1}$ subject to $(s_i - c)^T E(s_i - c) \leq 1$ for $i = 1, \ldots, m$. This optimisation problem is solved using conditional gradient ascent methods, the details of which can be found in Moshtagh (2005). We remove the masking of wind speeds below v for the empty regions of the ellipse so that a full ellipse of relative winds is defined. Figure 6.2.4 (centre panel) shows the ellipse fitted to the largest cluster of cells for the same time step of Daria as in Figure 6.2.2. For the remainder of the chapter, we refer to the ellipse of high relative wind speeds associated with an extratropical cyclone as a footprint. In the scenario whereby a series of footprints are associated with a single cyclone, we refer to this as a

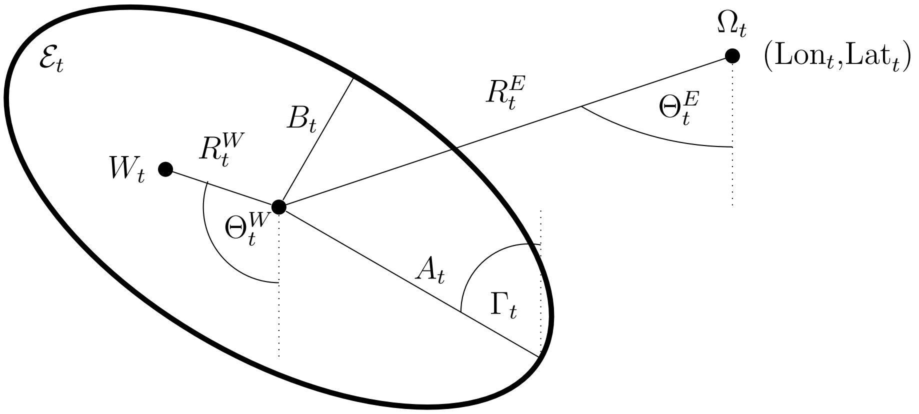


Figure 6.2.5: Graphical representation of the model variables derived from the ellipse shape used to summarise the windstorm footprint.

windstorm. If this series has gaps without footprints between periods with footprints, we refer to these as inactive and active phases of the windstorm.

Figure 6.2.5 shows the set of random variables we extract from the set of footprints that summarise the shape and magnitude of windstorms at a given time step relative to the storm centre $(\text{Lon}_t, \text{Lat}_t)$ with associated vorticity Ω_t. We use the notation $\mathcal{E}_t$ to denote the subset of cells in the spatial domain in Figure 6.2.1 contained within the footprint. We define A_t and B_t to be the semi-major and semi-minor axes of the ellipse respectively at time t. The quantity R_t^E represents the grid cell distance between the storm centre and the centre of the ellipse at time t, while $\Theta_t^E \in [-\pi, \pi]$ denotes the bearing of the centre of the ellipse relative to due south at time t. We define $\Gamma_t \in [-\pi/2, \pi/2]$ to be the orientation of the ellipse relative to due north, while W_t denotes the maximum relative wind speed observed in the footprint at time t. We define R_t^W as the grid cell distance between the centre of the ellipse and the cell corresponding of maximum relative wind speed at time t, while $\Theta_t^W \in [-\pi, \pi]$ denotes the bearing between the two cells relative to due south.

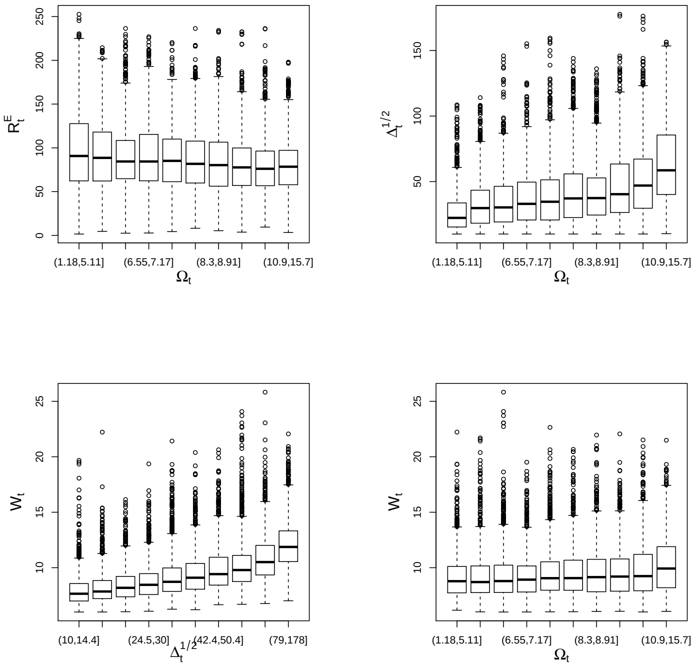


Figure 6.2.6: Boxplots showing the dependence between Ω_t and R_t^E, Ω_t and $\Delta_t^{1/2}$, $\Delta_t^{1/2}$ and W_t, and Ω_t and W_t.

6.2.4 Exploratory analysis

We undertake an extensive exploratory analysis on a number of aspects driving and influencing the evolution of windstorm activity, of which some findings are reported here. First, we investigate the dependence structure of characteristics of the ellipse representing the windstorm footprint, as shown in Figure 6.2.5. We assess the factors influencing the activation and termination of windstorm events. We also look at some quantities representing the spatial distribution of wind speeds relative to the storm centre, and how these compare with previous studies. Finally, we explore how the distribution of relative wind speeds within the footprint varies with respect to characteristics of the storm track and footprint.

Figure 6.2.6 shows boxplots illustrating some key dependencies between variables of a footprint. The area of the ellipse, which is proportional to $\Delta_t = A_t \times B_t$, tends to increase as vorticity and maximum relative wind speed increases, indicating that the strongest events tend to occur on a larger spatial scale. The radius R_t^E tends to decrease as Ω_t increases, although the effect is small, suggesting that footprints tend to occur closer to the storm centre when a large vorticity is observed. Maximum relative wind speed W_t tends to increase as Ω_t increases, though this dependence is weak. We also examine partial autocorrelation plots (not shown) to determine how individual components of the ellipse depend on their lags. This shows evidence of a second-order temporal dependence structure in most components reflecting the smooth evolution of footprints through the windstorm.

A consequence of our extraction of footprints is the loss of a large number of wind fields that have been removed for being non-extreme. Thus, if we are to construct our model that simulates footprints in time, we need a model that explains the factors influencing whether a windstorm is active or not. With regards to windstorm activation, we investigate the components of the track that may trigger an event. We find that the probability of windstorm activations tends to increase as vorticity increases, signalling a direct link between the intensity of the storm track and the occurrence of windstorm events. With regard to an inactive phase caused by termination of an active phase, we use the variables of the windstorm footprint in addition to information from the track. We find that a storm is more likely to terminate if it is associated with small values of relative maximum wind speed, vorticity and area. This indicates a termination is more likely if the windstorm weakens both in terms of its magnitude and its spatial scale.

This exploratory analysis also shows some spatial variation in the occurrence of footprints. Figure 6.2.7 (top panel) shows the density of storm track locations when windstorms are in an active phase. This shows that windstorms tend to occur over the North

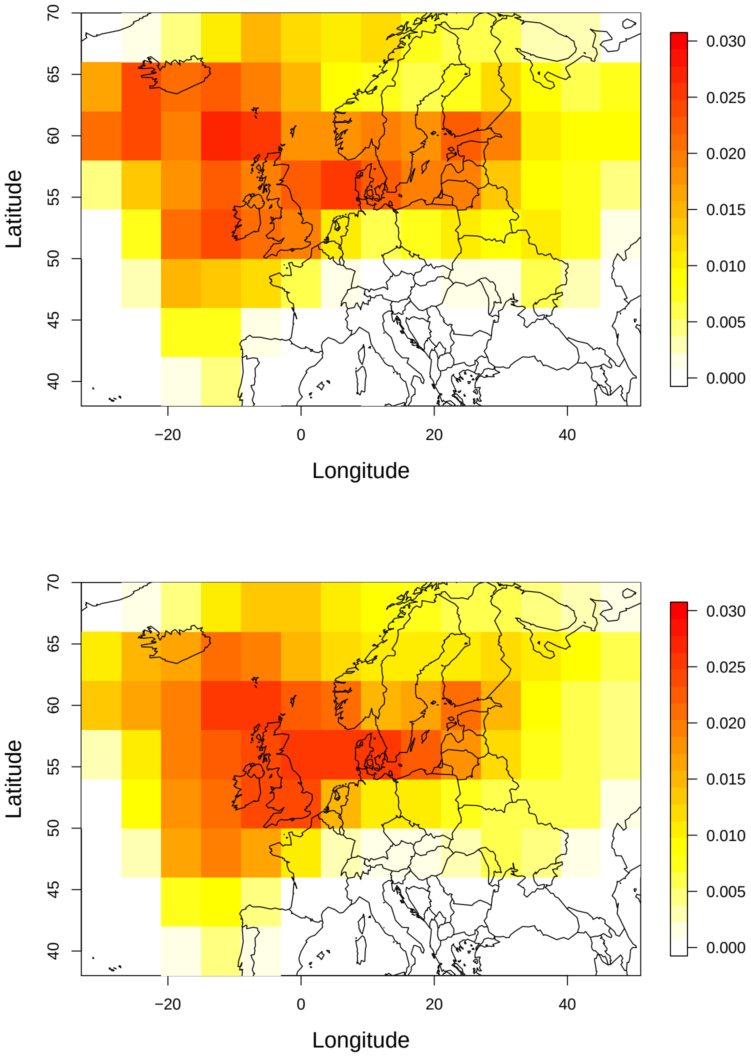


Figure 6.2.7: The spatial density of track locations that are associated with observed (top) and simulated (bottom) windstorms.

Atlantic and western Europe, with the density decaying as one moves to the edges of the domain. The reductions in density on the western boundary are not considered to be physical but the result of both the filtering of footprints to avoid their truncation by the EURO4 boundary and the effect of the boundary within the EURO4 simulation of weather. Most extratropical cyclones enter the EURO4 domain through this western boundary but their intensity is diminished due to the coarser resolution of the driving GCM. Their subsequent intensification takes a number of time steps which results in

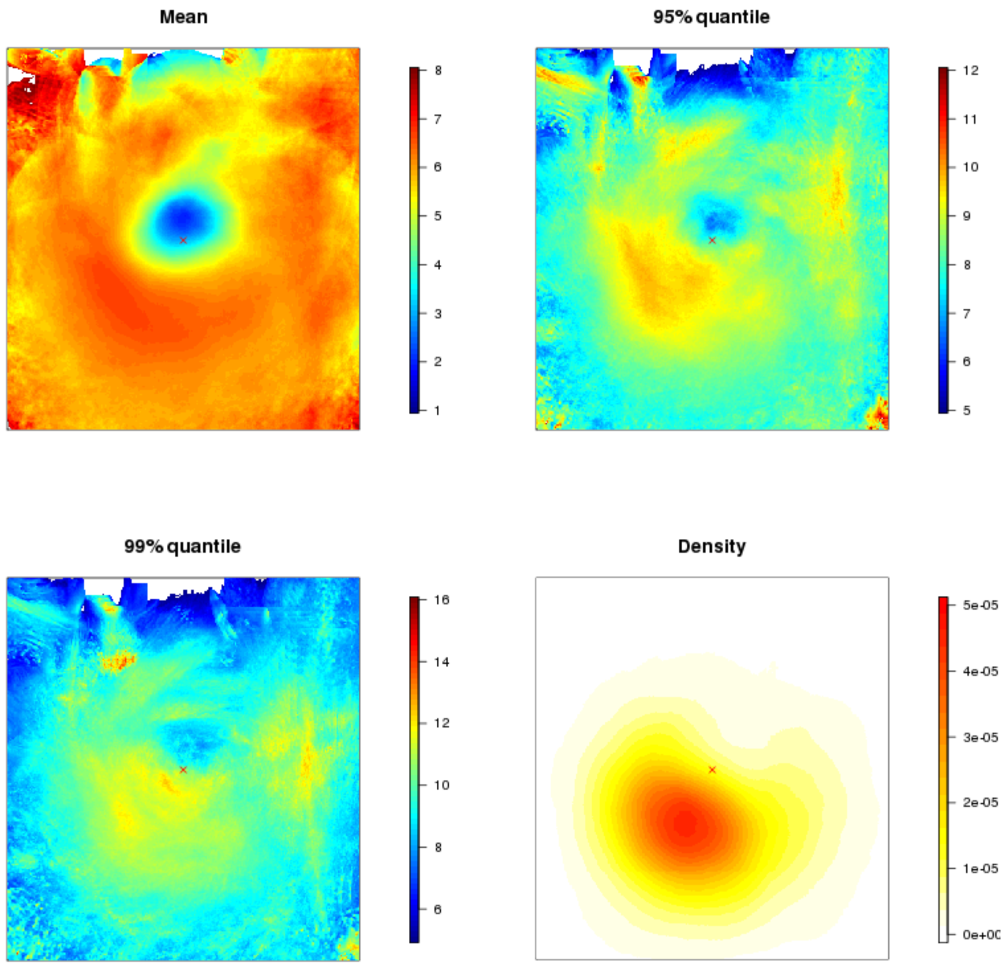


Figure 6.2.8: The mean, 95% quantile, 99% quantile and density of the occurrence of footprints relative to the storm centre (represented here as a cross) over all observed footprints.

smaller footprints and lower winds towards this boundary (see Figure 6.2.3).

We collect footprints over all fields centred at a point on the storm track, like in Figure 6.2.2, and assess the spatial distribution of these fields relative to the storm centre. We find that some events are picked up spuriously by the feature extraction algorithm that may not be generated as a result of the extratropical cyclone. With this in mind, we exclude ellipses whose centre occurs on the outer edges of the footprint as well as events that are sufficiently small. Figure 6.2.8 shows the mean, 95% quantile and 99%

quantile of wind speeds over the remaining fields after this filtering is implemented. This shows that relative wind speeds tend to be larger in regions southwest of the storm centre, which arise with the passage of cold fronts. This figure also shows the density of events over all footprints, illustrating that windstorms are most likely to occur southwest of the storm centre, with very few events occurring in the northern half of the field in comparison. These are consistent with the observations of Catto et al. (2010) and Rudeva and Gulev (2011).

Unusual events with large magnitudes are detected on the northwest and southeast edges of the domain. Despite our filtering, these events may not be generated by the extratropical cyclone; however, they are very rare and their impact minimal. The mean behaviour shows a local minimum occurring close to the storm centre, which likely arises as a result of low wind speeds occurring at pressure minima (Catto et al., 2010), which is known to be spatially adjacent to where the maximum vorticity occurs (Hoskins and Hodges, 2002).

We investigate the factors influencing the distribution of relative wind speeds within each footprint through analysis of their mean and standard deviation. This reveals, intuitively, that the mean and standard deviation of winds tend to increase as the maximum relative wind increases. In contrast, an increase in vorticity tends to slightly reduce the mean relative wind, while increasing the standard deviation. Additionally, the standard deviation of wind increases as the area of the ellipse increases, which one would expect given the increased chances of observing smaller and larger values of wind speed. By construction the strongest relative winds tend to occur near the location of the maximum, determined by R_t^W and Ω_t^W, while weaker relative winds are more likely closer to the perimeter of the ellipse. Additionally, the relative winds tend to exhibit anisotropic properties, which tends to manifest in a ‘stretching’ of the band of strongest winds oriented in the direction perpendicular to Θ_t. As the windstorm evolves, this gives the effect of the winds 'bending' around the storm centre.

6.3 Windstorm modelling

6.3.1 Introduction

We propose an approach for simulating windstorms relative to an extratropical cyclone track. In particular, we describe methods, motivated by our findings in Section 6.2.4, for simulating a footprint at each time step of a windstorm along with the spatial distribution of wind speeds within a footprint. The times at which a windstorm is in an active phase with respect to a cyclone track are denoted by $\{t_S, \ldots, t_A, \ldots, t_T\}$, where t_S and t_T denote the times corresponding to the chronological beginning and end of an active phase respectively and t_A denotes the time at which the windstorm is activated within the simulation. If the windstorm simulation is activated at time t_A, forwards and/or backwards propagation routines, outlined in Section 6.3.2, are used to model the forward and backwards evolution of the footprint through the windstorm until times t_T and/or t_S are reached, which we refer to as the termination of the active phase. In practice, it can be more subtle as a windstorm can have repeated phases of activity and inactivity along a cyclone track. Details of this feature are discussed in Section 6.3.3. Stochastic models for determining t_A, t_S and t_T are also described in Section 6.3.3, along with a more detailed description of the simulation procedure. Our approach for modelling the wind speed fields within the footprint is outlined in Section 6.3.4.

6.3.2 Footprint modelling

Our approach aims to model the spatio-temporal evolution of footprints using the ellipse structure of Section 6.2.3. As this structure describes the spatial extent of a windstorm at a given time step, we can model the temporal evolution of the footprint by exploiting the time series structure of the footprint variables shown in Figure 6.2.5, which determine the position, size and magnitude of a footprint relative to the storm centre. Supported by the exploratory data analysis in Section 6.2.4, we assume that the multivariate time series $\boldsymbol{Z}_t = \{A_t, B_t, W_t, R_t^E, \Theta_t^E, R_t^W, \Theta_t^W, \Gamma_t\}$ jointly follows a stationary kth order Markov process during active phases along the cyclone track and are conditionally independent given the track information over tracks and on different active phases along a track. By the Markov property, the distribution of the current value of a process is affected only by the previous k time steps of the process. We define an arbitrary d-dimensional kth order stationary multivariate Markov process $\boldsymbol{Z}_{1:n} = \{Z_{tj} : t = 1, \ldots, n; j = 1, \ldots, d\}$, where Z_{tj} denotes the t-th time step of the j-th component and n is the length of the time series. Consequently, it is only necessary to model the joint distribution of $\boldsymbol{Z}_{t:t+k}$, from which the conditional density function of $\boldsymbol{Z}_{k+1} \mid \boldsymbol{Z}_{1:k}$ can be derived. This joint distribution is determined by its marginal distributions and its copula. This structure also allows us to capture dependence between the different components of the time series, e.g., vorticity and distance from the storm centre.

Under the assumption of stationarity of $\boldsymbol{Z}_{t:t+k}$, observations of this $d \times (k+1)$ vector are identically distributed over t and with joint density function f. A simple choice is to model f nonparametrically using a multivariate kernel smoothed density function $\hat{f}$, such that

$$\hat{f}(\boldsymbol{z}) = \frac{1}{n} \sum_{i=1}^{n} K_H \left(\boldsymbol{z} - \boldsymbol{z}_i \right), \tag{6.3.1}$$

where K_H denotes the kernel function, defined with respect to a $(k+1) \times (k+1)$ bandwidth matrix H. The function K_H is chosen to be the multivariate Gaussian density function with variance H. Our exploratory analysis in Section 6.2.4 helps to identify which components of $\boldsymbol{Z}_{t:t+k}$ are independent or conditionally independent, which simplifies the form of H and helps to identify k. For example, we found $k = 2$ and a weakly dependent relationship between Ω_t and W_t with Ω_t and W_t conditionally independent given A_t and B_t. This allows us to simplify our model for $\Pr(\boldsymbol{Z}_{t+k} \leq z \mid \boldsymbol{Z}_{t:t+k-1} = \boldsymbol{z}_{t:t+k-1})$, for which we use the kernel estimate of the conditional distribution function, the formulation of which can be found in Appendix C.1.

Our approach is designed to construct the joint distribution for $\{\boldsymbol{Z}_{t_S}, \ldots, \boldsymbol{Z}_{t_A}, \ldots, \boldsymbol{Z}_{t_T}\}$, where $\{t_S, \ldots, t_A, \ldots, t_T\}$ are the consecutive time steps at which the windstorm is active, as discussed in Section 6.3.1. The characteristics of the footprint are influenced by the storm track components $(\text{Lon}_t, \text{Lat}_t, \Omega_t)$ so our footprint propagation routine needs to reflect this. To account for the spatial variation of the footprint characteristics, we simulate realisations of $\boldsymbol{Z}_t$ only using conditional kernel density estimates based on footprints associated with values of $(\text{Lon}_t, \text{Lat}_t)$ in ∇_t, where ∇_t is a region of size $10^\circ \times 6^\circ$ centred at $(\text{Lon}_t, \text{Lat}_t)$. First, consider forward propagation from a given value of $\boldsymbol{Z}_{t_A}$ whilst the windstorm remains in an active phase. For all times $t_A+1 \leq j \leq t_A+2$. We simulate forwards realisations of $\boldsymbol{Z}_j$ from a conditional kernel density derived from $\hat{f}(z)$, as described in Appendix C.1, such that

$$\begin{aligned}
r_j^E &\sim R_j^E \mid R_{t_A:j-1}^E = r_{t_A:j-1}^E, \Omega_j = \omega_j, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
\theta_j^E &\sim \Theta_j^E \mid \Theta_{t_A:j-1}^E = \theta_{t_A:j-1}^E, R_j^E = r_j^E, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
a_j &\sim A_j \mid A_{t_A:j-1} = a_{t_A:j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
b_j &\sim B_j \mid B_{t_A:j-1} = b_{t_A:j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
w_j &\sim W_j \mid W_{t_A:j-1} = w_{t_A:j-1}, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, A_j = a_j, B_j = b_j, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
\gamma_j &\sim \Gamma_j \mid \Gamma_{t_A:j-1} = \gamma_{t_A:j-1}, \Theta_j^E = \theta_j^E, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
r_j^W &\sim R_j^W \mid R_{t_A:j-1}^W = r_{t_A:j-1}^W, A_j = a_j, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j \\
\theta_j^W &\sim \Theta_j^W \mid \Theta_{t_A:j-1}^W = \theta_{t_A:j-1}^W, (\text{Lon}_j, \text{Lat}_j) \in \nabla_j.
\end{aligned}$$

When $j > t_A + 2$, we simulate from the conditional kernel model

$$\begin{aligned}
r_j^E &\sim R_j^E \mid R_{j-k:j-1}^E = r_{j-k:j-1}^E, \Omega_j = \omega_j, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
\theta_j^E &\sim \Theta_j^E \mid \Theta_{j-k:j-1}^E = \theta_{j-k:j-1}^E, R_j^E = r_j^E, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
a_j &\sim A_j \mid A_{j-k:j-1} = a_{j-k:j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
b_j &\sim B_j \mid B_{j-k:j-1} = b_{j-k:j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
w_j &\sim W_j \mid W_{j-k:j-1} = w_{j-k:j-1}, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, A_j = a_j, B_j = b_j, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
\gamma_j &\sim \Gamma_j \mid \Gamma_{j-k:j-1} = \gamma_{j-k:j-1}, \Theta_j^E = \theta_j^E, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
r_j^W &\sim R_j^W \mid R_{j-k:j-1}^W = r_{j-k:j-1}^W, A_j = a_j, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j \\
\theta_j^W &\sim \Theta_j^W \mid \Theta_{j-k:j-1}^W = \theta_{j-k:j-1}^W, (\mathrm{Lon}_j, \mathrm{Lat}_j) \in \nabla_j.
\end{aligned}$$

Backwards simulation is implemented similarly, but with substituting $\boldsymbol{Z}_j$ for $\boldsymbol{Z}_{t_A-j}$. The details of whether a forward or backward routine should be used are found in Section 6.3.3. Realisations of R_j^W and Θ_j^W are rejected if the simulated position of the maximum occurs outside of the simulated ellipse. For marginal components of $\boldsymbol{Z}_t$ where the extremal behaviour is of interest, e.g., W_t, one could follow the approach of Chapter 5 by specifying a model for the extremal marginal and temporal dependence structure; this is discussed in more detail in Section 6.5.

6.3.3 Activation and termination

As discussed in Section 6.3.1, we require an approach for determining the times at which a windstorm is in an active phase. Whether the windstorm is active or not is modelled by a Bernoulli logistic generalised additive model (Wood, 2006) based on covariate information from the physical structure of an extratropical cyclone. Section 6.2.4 demonstrated that the windstorm is typically active at the time when the maximum vorticity is observed, which we denote by t_Ω, as this time is associated with the strongest winds over the cyclone track. However, we observe some instances where the windstorm is inactive at t_Ω, but is active at other times on the cyclone track. Similarly, the termination of a windstorm is highly uncertain, but it is shown in Section 6.2.4 to be linked to weakening events in terms of spatial scale and magnitude.

Specifically, first consider the activation of a windstorm phase. Let T_t be a Bernoulli random variable such that:

$$T_t = \begin{cases} 1 & \text{the storm is active at time } t \\ 0 & \text{otherwise.} \end{cases}$$

So $T_t \sim \text{Bernoulli}(p_t)$, where

$$p_t = \frac{\exp\left\{\sum_{i=1}^{q} \beta_i(\nu_{i,t})\right\}}{1 + \exp\left\{\sum_{i=1}^{q} \beta_i(\nu_{i,t})\right\}},$$

where β_i is a smooth non-linear function of covariate ν_i with $i \in (1, \ldots, q)$, where q is the number of covariates and $\nu_{i,t}$ denotes the realisation of the ith covariate at time t. The smooth functions are represented by penalised regression splines, where the smoothing parameter is determined using generalised cross validation (GCV) and the model is fitted using penalised maximum likelihood. For more details on additive models, see Wood (2006). An identical formulation is used for the termination of a windstorm.

For a storm track, let t denote the time from its start $t = 1$ until its end $t = l$. We first attempt to activate the windstorm simulation process at time $1 \leq t_\Omega \leq l$. If a windstorm is determined to be active at t_Ω, then $t_A = t_\Omega$ and we propagate forwards and backwards in time using the methods described in Section 6.3.2 until the active phase terminates. If a windstorm is inactive at t_Ω, we proceed successively forwards in time along the track and check whether a windstorm becomes active at times $t_\Omega < t \leq l$. If a windstorm is inactive at all these times, we proceed successively backwards along the track and check whether an activation occurs at times $1 \leq t < t_\Omega$. If a windstorm is first found to be active during the forwards procession such that $t_A > t_\Omega$, we set $t_S = t_A$ and propagate forwards using the approach in Section 6.3.2 until termination at time $t_T \leq l$. Likewise, if a windstorm is first found to be active during the backwards procession such that $t_A < t_\Omega$, we set $t_T = t_A$ and propagate backwards until termination at time $t_S \geq 1$. We allow for the possibility that multiple phases of consecutive footprints can occur on the same track. If $t_S > 1$, we proceed backwards along the track to check if the windstorm reactivates, in which the same procedure applies until $t = 1$. Similarly, if $t_T < l$, we proceed forwards along the track to check if the windstorm reactivates, in which the same procedure applies until $t = l$.

Given that a windstorm is active, some initial footprint characteristics are required for propagating the storm in time. In doing so, we would like to model the joint distribution of $\boldsymbol{Z}_{t_A} = \{A_{t_A}, B_{t_A}, W_{t_A}, R^E_{t_A}, \Theta^E_{t_A}, R^W_{t_A}, \Theta^W_{t_A}, \Gamma_{t_A}\}$, where $\boldsymbol{Z}_{t_A}$ represents the characteristics of the footprint at time t_A. Our model should account for the dependence structure between $\boldsymbol{Z}_{t_A}$ and the components of the track, specifically $(\mathrm{Lon}_{t_A}, \mathrm{Lat}_{t_A}, \Omega_{t_A})$. We therefore simulate a realisation of $\boldsymbol{Z}_{t_A}$ jointly from $\boldsymbol{Z}_{t_A} \mid (\mathrm{Lon}_{t_A}, \mathrm{Lat}_{t_A}, \Omega_{t_A})$, which we estimate using a conditional kernel density (see Appendix C.1). If $t_A = t_\Omega$, we use all observed footprint data $\boldsymbol{z}_{t_\Omega}$ in constructing our density function. Otherwise, if $t_A \neq t_\Omega$, we use $\boldsymbol{z}_\tau$ where $\tau = \{t_A : t_A \neq t_\Omega\}$, that is, footprint data corresponding to when the windstorm is activated during the forwards and backwards procession along the track.

6.3.4 Modelling wind speeds within a footprint

We also require an approach for modelling the spatial distribution of relative wind speeds within a footprint at each time step of a windstorm. A natural class of models to consider is Gaussian processes, which are widely used in geostatistics to model spatial data. A Gaussian process can be used to describe the joint distribution of random variables over a continuous domain such as space inside a footprint, while for any finite collection of locations in the space the variables follow a multivariate Gaussian distribution. For a comprehensive overview of Gaussian process modelling for spatial data, see Cressie (1993), Stein (1999) and Diggle and Ribeiro (2007).

Let $\{X^E(s,t) : s \in \mathcal{E}_t\}$ denote the field of relative wind speeds in the ellipse $\mathcal{E}_t$ at time t, where $X^E(s,t)$ is marginally Exp(1) distributed over t for each s. Let D_t be the distribution function of $X^E(s,t)$ for all $s \in \mathcal{E}_t$ at time t. We use a probability integral transform to convert to a Gaussian field, which we denote by $X^G(s,t)$, with

$$X^G(s,t) = \Phi^{-1}\left[D_t(X^E(s,t))\right], \tag{6.3.2}$$

for all $s \in \mathcal{E}_t$ and all t, where Φ denotes the standard Gaussian distribution function. We make the assumption that $\{X^G(s,t) : s \in \mathcal{E}_t\}$ follows a Gaussian process with zero mean and unit variance for each t such that

$$X^G(s,t) \sim \mathrm{GP}(0, 1, \rho((s_1, s_2)J_t\Psi_t)),$$

where $\rho(\cdot)$ denotes an isotropic correlation function, with $s_1, s_2 \in \mathcal{E}_t$ and

$$\Psi_t = \begin{bmatrix} \cos\psi_t & -\sin\psi_t \\ \sin\psi_t & \cos\psi_t \end{bmatrix} \qquad J_t = \begin{bmatrix} 1 & 0 \\ 0 & \zeta_t \end{bmatrix},$$

where ψ_t is the time-varying anisotropy angle representing the counter-clockwise rotation of the space and $\zeta_t > 1$ is the time-varying anisotropy ratio, which controls the degree of stretching along the angle where correlation decays most slowly with increasing distance. Supported by the exploratory analysis in Section 6.2.4, we fix ψ_t to be the angle perpendicular to Θ_t, while empirical evidence suggests that a good choice of ζ_t would be the ratio A_t/B_t.

The correlation function is typically chosen so that the correlation between $X^G(s_1, t)$ and $X^G(s_2, t)$ decreases as the distance $|s_2 - s_1|$ between the sites increases. A common choice of correlation function is the Matérn family, which has the form

$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\alpha_t)^{\kappa}K_{\kappa}(u/\alpha_t), \tag{6.3.3}$$

where $\kappa > 0$ is a shape parameter that determines the smoothness of the underlying process, K_κ denotes a modified Bessel function of the second kind of order κ, and $\alpha_t > 0$ is a time-varying scale parameter with dimensions of distance. The Matérn family is a generalisation of other common choices of correlation functions. For example, choosing $\kappa = 0.5$ gives the exponential correlation function, while as $\kappa \to \infty$, the form reduces to a Gaussian correlation function $\rho(u) = e^{-\alpha_t u^2}$. The parameters κ and α_t are highly non-orthogonal, but larger values of κ are typically associated with smoother fields, while larger values of α_t gives fields that are correlated at larger distances. We fix $\kappa = 0.6$, which in practice generates fields of similar smoothness to the observed fields. We estimate the parameter α_t corresponding to each footprint using variogram methods. We avoid the use of likelihood methods due to the computational difficulties that arise with large spatial datasets like ours. For more details on inference for Gaussian processes, see Diggle and Ribeiro (2007). Exploratory analysis suggests that α_t is constrained by the value of Δ_t, which is proportional to the area of the ellipse. We therefore model $\alpha_t \mid \Delta_t$ using a conditional kernel density, the formulation of which is outlined in Appendix C.1. At time t, we generate a realisation of α_t conditional on the simulated realisation of Δ_t determined by the model in Section 6.3.2.

We require an approach for determining the marginal distribution of relative wind speeds within the simulated footprint. As discussed in Section 6.2.4, the relative wind speeds within the ellipse are linked to the vorticity of the track, the size of the event and the maximum relative wind speed. We construct a weighted nonparametric estimate of the distribution function D_t, where D_t denotes the distribution function of wind speeds within the footprint at time t. Using a kernel, we weight observed relative wind speeds in $\mathcal{E}_t$ conditional on the corresponding values of W_t, Δ_t and Ω_t, from which D_t can be estimated using Monte Carlo methods.

We incorporate information about the physical structure of the footprint in determining the structure of the Gaussian field $\{\tilde{X}^G(s,t) : s \in \mathcal{E}_t\}$ using conditional simulation (Diggle and Ribeiro, 2007). We impose three conditions on the simulated fields: that the maximum relative wind speed is simulated at the position determined by R_t^W and Θ_t^W; that the lower limit of D_t occurs on the outer perimeter of the ellipse; and that the lower limit of D_t occurs everywhere in the region corresponding to the local minimum of wind speeds (see Section 6.2.4) near the storm centre when a footprint is simulated in the vicinity of this region. The first and second conditions create a two-dimensional pseudo-Brownian bridge between the position of maximum and the perimeter of the footprint. To impose the third condition, we specify a second ellipse centred at the local minimum with random dimensions for size; specifically we fix the maximum length of the semi-major and semi-minor axes of this ellipse to be 40 and 35 units of grid-cell distance respectively, with random perturbations modelled as Exp(0.05) random variables. These values were found to replicate well the average behaviour of wind speeds in the region at which the local minimum occurs.

After simulation of $\{\tilde{X}^G(s,t) : s \in \mathcal{E}_t\}$, we transform this field to obtain a field of relative wind speeds conditional on the characteristics of the footprint such that

$$\tilde{X}^E(s,t) = D_t^{-1}\left(\Phi[\tilde{X}^G(s,t)]\right).$$

An example of this is shown in Figure 6.2.4 (right panel), in which we have simulated a field of relative wind speeds conditional on the footprint of storm Daria at this particular time step. Our model captures the correlation structure of the field quite well, with the weakest winds occurring on the outer perimeter of the ellipse and the large winds occurring in similar locations to the observed footprint. For this simulated field, the fine structure in terms of the decay from the maximum relative wind speed being more isotropic than the observed field. When viewed in an Eulerian framework, this level of spatial difference is not too important as these footprints move over space with the cyclone track, so blur out this distinction. Having obtained the simulated relative wind fields, we can then transform these onto the observed margins, such that for each $s \in \mathcal{E}_t$

$$\tilde{X}(s,t) = F_s^{-1}(1 - e^{-\tilde{X}^E(s,t)}),$$

where F_s denotes the marginal model for cell s, as defined in (6.2.1). With this formulation, we are assuming the relative wind fields at consecutive time steps are conditionally independent given temporally correlated realisations of W_t and α_t. While this assumption gives good results in practice, further investigation may be necessary to assess whether performance could be improved by specifying a spatio-temporal structure in the correlation function $\rho(\cdot)$.

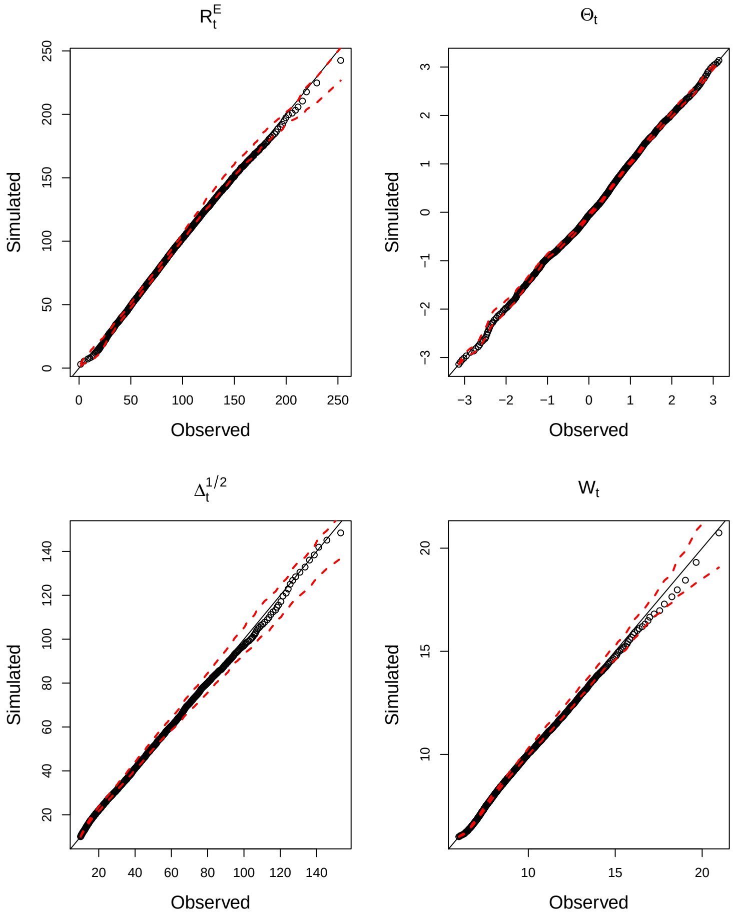


Figure 6.4.1: QQ plots, with 95% tolerance intervals, comparing the observed and simulated marginal distributions of R_t^E, Θ_t, $\Delta_t^{1/2}$ and W_t. Simulated values are based on footprints relative to 2,944 synthetic tracks from the model of Chapter 5.

6.4 Results

We examine the performance of the windstorm model in terms of simulated footprint characteristics and then the wind speeds within the footprint. The joint risk from extreme windstorms at locations in northern England and eastern Germany is then explored by combining the windstorm model presented here with the track model of Chapter 5 to produce estimates of joint event probabilities through simulation.

6.4.1 Validation of footprint model

We explore first whether the characteristics of windstorm footprints are being captured through an assessment of the marginal distributions of the individual components and their dependence structure. QQ plots based on the simulation of footprints using the model described in Section 6.3.2, applied to $2,944$ synthetic storm tracks, the same number as in the observed record, were assessed both for the observed tracks and tracks generated by the model of Chapter 5. Both showed similar positive results, so we illustrate only the latter (see Figure 6.4.1). They show that the marginal distributions of radius, bearing, proportional area and relative maximum wind speed are being captured well by the model. Figure 6.4.2 shows that, when compared with Figure 6.2.6, the dependence structure of these components is also consistent with the observations. We can thus conclude that the physical structure of the observed windstorm footprints is replicated sufficiently by the model.

We examined the components of a windstorm influencing the activation and termination models that were outlined in Section 6.3.3. In both cases, we investigate multiple combinations of covariates and compare model fit using AIC. The best fitting activation model includes functions of vorticity, longitude and latitude. Figure 6.4.3 (left panel) shows the estimated smooth function β_i associated with vorticity, which shows that β_i tends to increase approximately linearly as vorticity increases. This has the effect that the probability of activation tends to increase as vorticity increases, which reflects our findings from Section 6.2.4. The best fitting termination model includes functions of $\Delta_t^{1/2}$ and W_t. Figure 6.4.3 (centre and right panels) shows that β_i tends to decrease non-linearly as both $\Delta_t^{1/2}$ and W_t increase, which means that the probability of termination also tends to decrease. The effect of $\Delta_t^{1/2}$ tends to level off at high values, though wide confidence intervals suggest that this effect is highly uncertain. This analysis confirms our belief that weakening events in terms of spatial extent and maximum

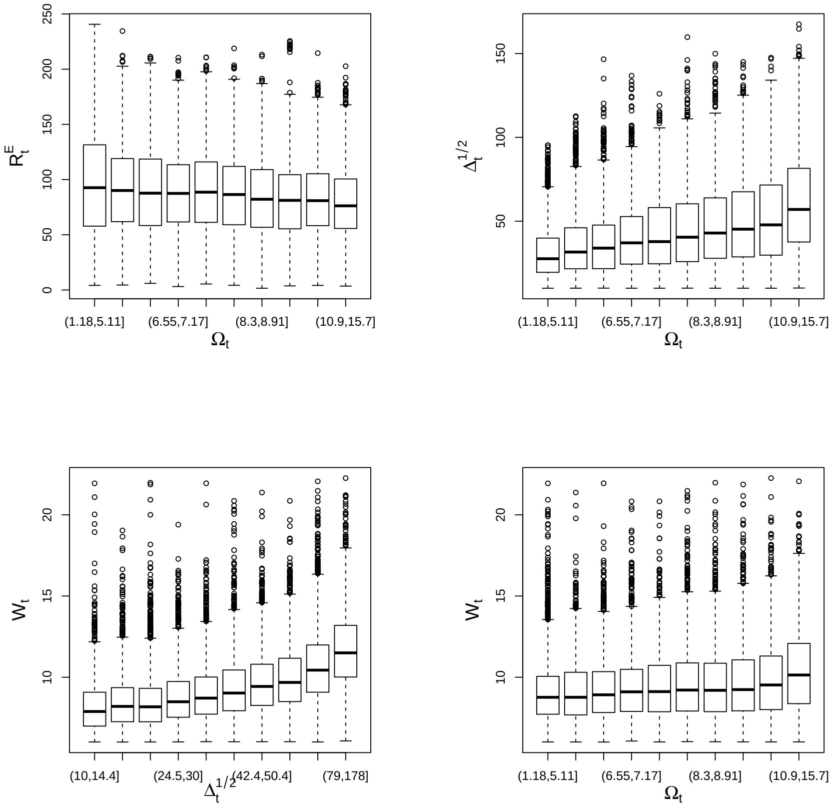


Figure 6.4.2: Boxplots showing dependence structure of some aspects of the simulated footprints. Simulated values are based on footprints relative to 2,944 synthetic tracks from the model of Chapter 5.

relative wind speed are consistent with the termination of an active phase of a windstorm. Figure 6.2.7 (bottom panel) shows the spatial density of windstorm occurrence in the simulations. We see that the large-scale spatial variation of windstorms from the model reflects that of the observations in Figure 6.2.7 (top panel).

6.4.2 Validation of model for wind speeds within a footprint

We check that the model replicates well the physical structure and location of winds relative to the storm centre. For this task we simulate the wind speeds within the footprints generated in Section 6.4.1 using the wind model of Section 6.3.4. Figure 6.4.4 shows the mean, 95% quantile, 99% quantile and spatial density of simulated winds

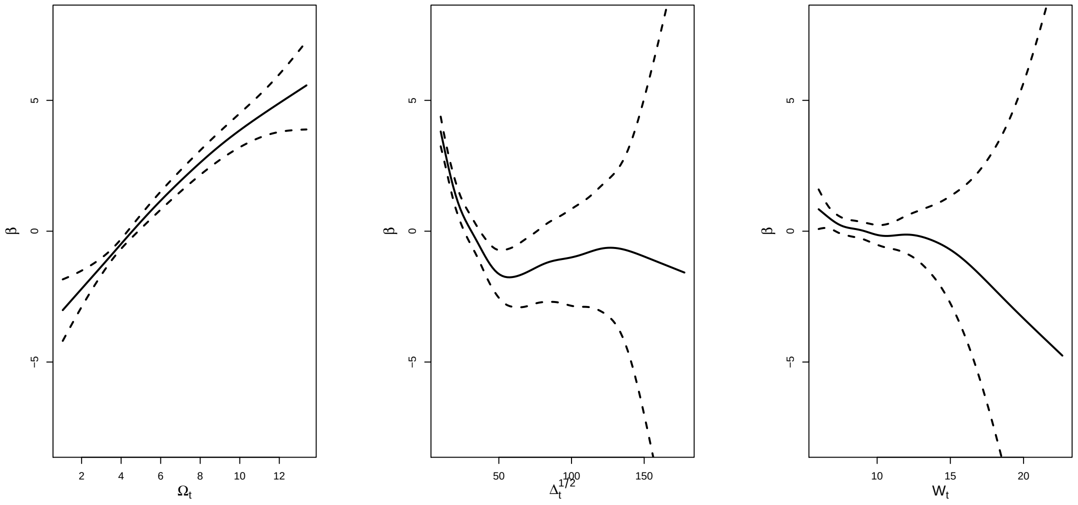


Figure 6.4.3: The smooth functions β_i (with 95% confidence intervals) of the logistic model showing the effect of Ω_t on the probability of windstorm activation (left) and of $\Delta_t^{1/2}$ (centre) and W_t (right) on the probability of termination.

from all simulated footprints relative to the centre of the storm, quantities that were previously studied for the observed data in Figure 6.2.8. The comparison of observed and simulated winds is very good; in particular it replicates the feature that the largest magnitudes tend to be found in the region southwest of the storm centre. The local minimum that occurs near the storm centre as a result of small pressure gradients is also captured. The upper quantiles of the spatial distribution are slightly more dispersed compared to the observed characteristics; however, we are satisfied that the large-scale features have been captured by the model. Figure 6.4.4 also shows that high magnitude events can be generated by the model on the outer edges of the domain. These are rare occurrences, but we believe these to be attributed to the detection of spurious events in the feature extraction algorithm. Section 6.5 discusses possible improvements to the algorithm so that the detection of spurious events is minimised.

By simulating windstorms relative to synthetic tracks, our model also allows us to perform an Eulerian analysis at different locations over the North Atlantic and Europe.

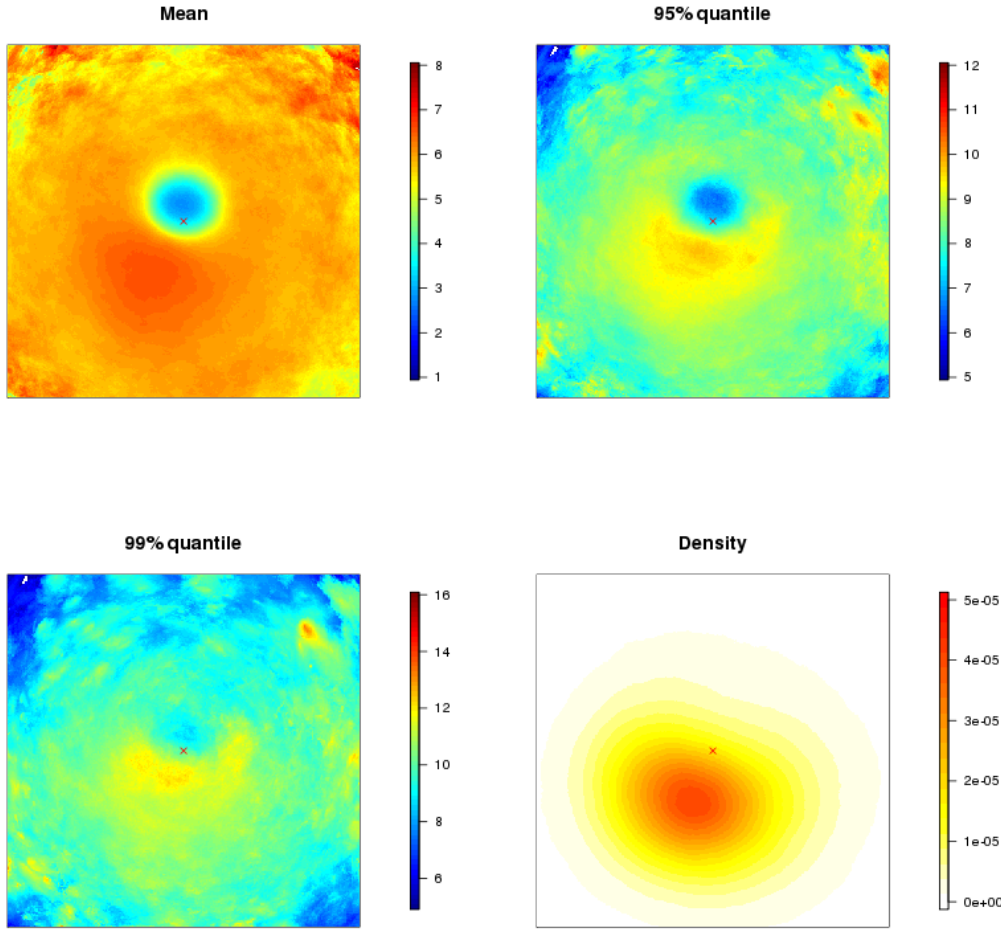


Figure 6.4.4: The mean, 95% quantile, 99% quantile wind speed and density of within-footprint winds relative to the storm centre (represented by the cross) over a set of simulated windstorms.

We assess how our model captures the distribution of extreme winds at a number of locations to examine whether our approach succeeds in generating physically realistic synthetic values at these different sites. We compare our simulated values with winds that are contained within the observed footprints at each site. QQ plots for six locations, three on land and three at sea, are shown in Figure 6.4.5. The marginal model captures the contrast in scale between observations on land and at sea, and our windstorm model largely captures the marginal extremal behaviour of wind speeds at these sites.

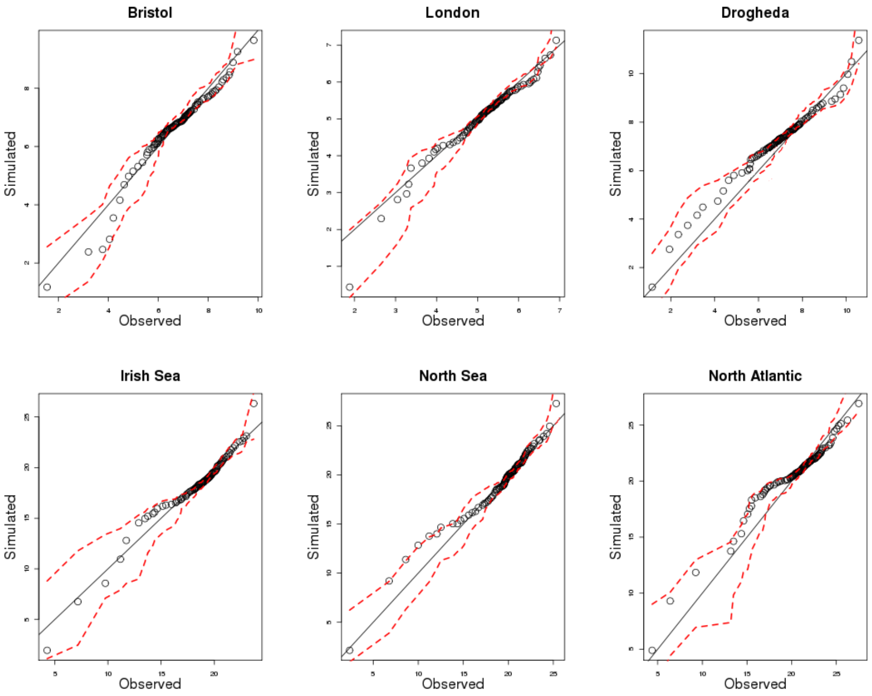


Figure 6.4.5: QQ plots, with 95% tolerance intervals, at six locations comparing the distribution of wind speeds from the observed and simulated windstorms.

6.4.3 Joint risk from windstorms

We use our approach to estimate quantities related to joint risk, that is, the probability that multiple locations are affected by extreme wind speeds simultaneously. As our model captures the spatial extent of these meteorological events, we should therefore capture the risk of multiple locations experiencing extreme wind speeds from the same storm. The results are based on 50,000 extratropical cyclone tracks generated using the track model of Chapter 5. For each track, a windstorm is simulated using the windstorm model of Section 6.3. This dataset represents approximately 600 years of data, under the assumption that there are on average 811 windstorms per year as found in the observed record.

Figure 6.4.6 (left panel) compares the joint behaviour on common Exp(1) margins between wind speeds in Lancaster and Manchester, two cities in northern England

73km apart. The joint correlation structure is largely captured by the model, which has the added benefit of being able to simulate joint events of magnitudes beyond the range of the observation record. One way of summarising the joint extremal behaviour of a process at arbitrary locations s_1 and s_2 is to estimate the quantity $\chi(q; s_1, s_2)$ (Coles et al., 1999), defined as

$$\chi(q; s_1, s_2) = \Pr(X^E(s_2, t) > x^q \mid X^E(s_1, t) > x^q),$$

where x^q is the $100q\%$ quantile of the common Exp(1) distribution. Estimates of the quantity $\chi(q; s_1, s_2)$ obtained from the observed and simulated data are shown for a range of x^q in Figure 6.4.6 (right panel), where s_1 and s_2 are chosen to be sites in Lancaster and Manchester respectively. Estimates from the data and the large simulated sample from the model are obtained as conditional proportions. Figure 6.4.6 shows that extremal dependence tends to decrease as the magnitude of the event increases. Here, 95% binomial confidence intervals are used to assess the uncertainty for the observed data and the Monte Carlo uncertainty for the simulated data. To obtain these intervals, we use an effective sample size $n\theta(x^q)$ defined in terms of the sample size n and a threshold-based extremal index $\theta(x^q)$ (Ferro and Segers, 2003; Eastoe and Tawn, 2012) to account for temporal dependence. For the model-based estimates, the confidence intervals do not represent the uncertainty due to the model parameter estimation. A fuller assessment of model uncertainty can be obtained using a parametric bootstrap, which would have the effect of widening the model-based confidence intervals. Despite not representing the full uncertainty in the model-based estimates, it is clear that there is no statistically significant difference between data and model-based estimates of $\chi(q; s_1, s_2)$ here over q within the range of the observed data. Critically though, this figure also illustrates how our model allows estimation of $\chi(q; s_1, s_2)$ beyond the range of the observational record, indicating that $\chi(q; s_1, s_2)$ continues to decay to 0 beyond the observed data.

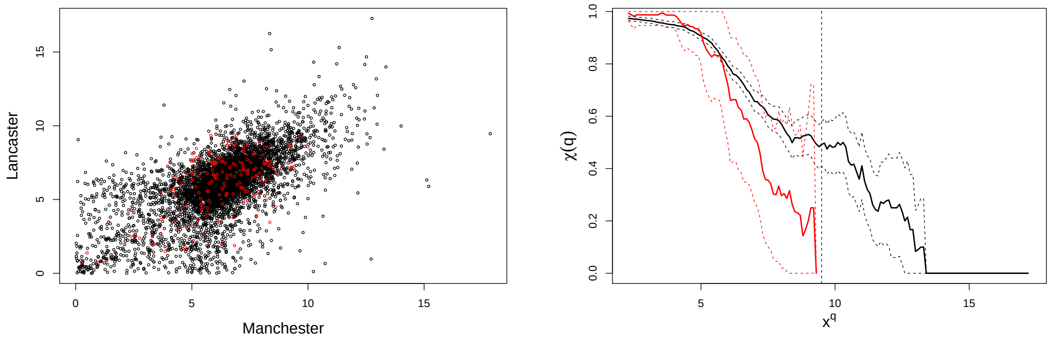


Figure 6.4.6: Scatter plot (left) showing observed (red) and simulated (black) wind speeds on an Exp(1) scale in Manchester and Lancaster. Estimates of $\chi(q; s_1, s_2)$ measuring extremal dependence between these locations as a function of x^q (right) using the observed (red) and simulated (black) data, with 95% Binomial confidence intervals using an effective sample size. The vertical line denotes the maximum observed wind speed on the Exp(1) scale.

Estimating $\chi(q; s_1, s_2)$ at a fixed critical level x^q at a set of sites s_2 can allow us to explore the spatial extent of extreme events. Figure 6.4.7 (top panels) shows $\chi(q; s_1, s_2)$ calculated across a number of locations in northern England, with s_1 being Lancaster in this instance. We explore two cases where x^q is chosen to be the 90% quantile and the 10-year return level at each site. In particular, we see that the probability surface decays more steeply as $|s_2 - s_1|$ increases for the more extreme events. We also see that Liverpool, Manchester and Leeds are more likely than not to experience an event on the 90% quantile simultaneously to Lancaster; however, this scenario is less likely for an event corresponding to the 10-year return level. Similarly, in Germany, as shown in the bottom panels in Figure 6.4.7, the probability of experiencing an extreme windstorm event simultaneously with Berlin decreases as the event becomes more extreme. The spatial extremal dependence is slightly stronger for Berlin than Lancaster, as might be expected given Berlin is more inland on a large land mass. In both regions, there is little evidence of anisotropy in the extremal dependence structure, with perhaps some indication of stronger dependence in the northwest-southeast direction centred at Berlin.

The results in Figures 6.4.6 and 6.4.7 illustrate how the spatial extent of an extreme windstorm event becomes more localised as the magnitude increases. This implies that in the limit, extreme values at each location tend not to occur simultaneously, which corresponds to the property of asymptotic independence. Models for asymptotic dependence, that is, when $\chi(q; s_1, s_2) \to c > 0$ as $q \to 1$ for $s_1 \neq s_2$, lead to extreme values tending to occur simultaneously, are well-established but tend to over-estimate the probability of extreme joint events given that the underlying process is asymptotically independent, i.e., when $c = 0$. Models that capture asymptotic independence are less well-established; see Ledford and Tawn (1996), Heffernan and Tawn (2004), Wadsworth and Tawn (2012a) and Winter et al. (2016) for some examples. We have shown that our model captures the property of asymptotic independence over space, while accounting for the complex non-stationarity of the extratropical cyclone system.

6.5 Discussion

We have presented a novel approach to modelling windstorms in a Lagrangian framework. We described two models; first, for the evolution and development of the footprint relative to the storm track, and second, for the spatial distribution of extreme winds within the footprint. The Lagrangian framework allows us to pool information regarding events over the spatial domain being studied, which allows extrapolation of the characteristics of windstorms over space. The model provides a mechanism for generating synthetic windstorm events, the analysis of which allows improved estimation of joint risk associated with extreme windstorms over Europe.

There are, however, opportunities to improve the performance of the model. While the feature extraction approach introduced in Section 6.2.3 appears to extract the main features of a windstorm, steps could be taken to improve the robustness of this proce-

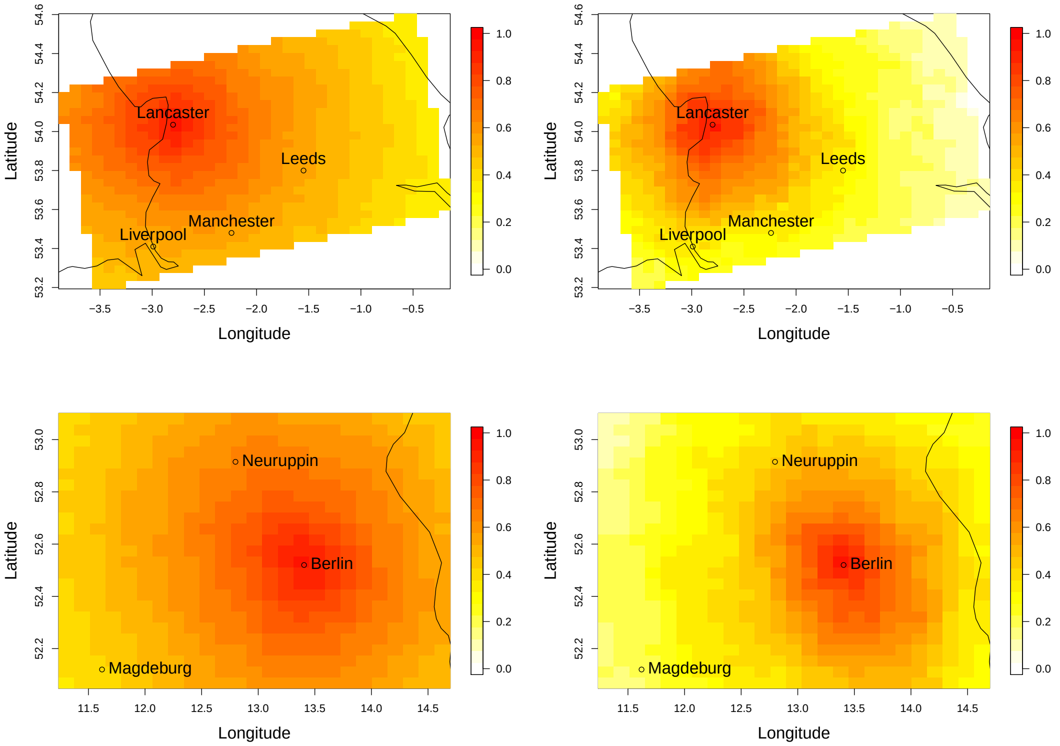


Figure 6.4.7: Estimates of $\chi(q; s_1, s_2)$ for northern England (top) and eastern Germany (bottom) conditioning on a critical value x^q, where s_1 is the cell where Lancaster (top) and Berlin (bottom) are located. In the left panels, x^q is taken to be the 90% quantile, while x^q is taken to be the 10-year return level in the right panels. Both regions are of equal size.

dure. Firstly, one could conduct a sensitivity study on the choice of threshold v, which controls the level under which the wind speed fields are masked. Ideally, one should choose a high enough value of v such that convective events and non-extreme winds are masked, but small enough so that the localised features of the windstorm aren't excluded. "Sting jets", meteorological phenomena associated with rapidly developing storms, can produce damaging winds on very small spatial scales (Baker, 2009; Hewson and Neu, 2015) and therefore it is important that the extraction algorithm should not exclude such features.

The feature extraction algorithm could also be improved to minimise the detection of spurious footprints that may not be generated by the extratropical cyclone. We see examples of this through high magnitude events occurring on the outer edges of the domains in Figure 6.2.8. After the spatio-temporal filtering step, our algorithm selects the largest cluster in size to define the footprint. One could alternatively define a score function that incorporates multiple criteria in a detection strategy, which could be motivated by physical intuition regarding the structure of a windstorm. For example, the score function could give a higher weight to clusters with a bearing relative to the storm centre closest to southwest, where most footprints seem to occur. To induce some smoothness between consecutive time steps, one could assign a higher weight to a cluster such that the Euclidean distance between the position of this cluster and the selected cluster at the previous time step is minimised. Exploration of different score functions could add valuable improvements to our feature extraction algorithm, and ultimately, the model.

Our conditional kernel strategy for modelling $\{\boldsymbol{Z}_t\}$ appears to perform well. However, we could alternatively model the extremal behaviour of W_t using the approaches described in Winter and Tawn (2017) and Chapter 5, whereby a GPD model is defined above a high threshold and the extremal temporal dependence structure is modelled using a kth order extremal Markov process. This approach stems from the conditional multivariate extreme value methodology of Heffernan and Tawn (2004). We did not implement this approach as part of this study due to the additional complexity involved and that extrapolation occurred naturally through the features of our model. In addition, the observed values of W_t correspond to upper tail values of X_i, the distribution of wind speeds in cell i. Because we have a large dataset of observations in the upper tail, we felt that standard statistical modelling approaches were sufficient in this case. However, investigation of the benefits of imposing an extremal temporal dependence structure on the upper tail of W_t represents an interesting avenue of future research.

Chapter 7

Conclusion

In this chapter, we conclude by reflecting on the contributions made by this thesis. We also present possible opportunities for building upon the models developed in this thesis.

7.1 Summary of contributions

This section summarises the main findings of this thesis. In particular, we reflect on the contributions made to improving inference for existing extreme value methodology, and also to using these methods to develop a physically-motivated Lagrangian statistical model for European windstorms. The tools presented in this thesis have their benefits for practitioners interested in answering questions related to more accurate representations of risk derived from extreme value models, and also for those interested in quantifying risk associated with extratropical cyclones.

In Chapter 3, we proposed an approach for improving estimation of the Poisson process model for threshold excesses. Unlike the generalised Pareto distribution, the Poisson process model parameters are invariant to the choice of extreme value threshold, and thus offer a more natural parameterisation for covariate modelling. However, the non-orthogonal dependence structure of the model parameters makes inference difficult. In this chapter, we proposed a reparameterisation of this model using a tuning parameter defined in the likelihood, which transforms the parameters so that they are near-orthogonal. We implemented this in the Bayesian framework, in which the mixing associated with the Markov Chain Monte Carlo routine was substantially improved. We presented some theoretical results giving an analytical expression for the optimal selection of the tuning parameter in the i.i.d. case, based on minimising the asymptotic posterior correlations of the parameters. In the non-i.i.d. case, we also described an approach for the selection of the tuning parameter which requires a numerical minimisation of the posterior correlations. We detailed results based on application to both simulated data and rainfall data in Cumbria, UK, achieving promising results. This approach increases the possibility that this extreme value modelling framework will be applied more widely in practice.

Chapter 4 described a Bayesian spatial hierarchical model for threshold excesses using a generalised Pareto distribution in the data layer, applied to precipitation data in Great Britain. These models aim to borrow strength across locations to improve estimation of marginal return levels, and to reduce uncertainty of these estimates. They assume conditional independence over time and space, thus ignoring the concept that extreme precipitation events can occur simultaneously across locations. These models are well-established, but users of this approach often ignore accounting for the dependence in the data when quantifying return level uncertainty, after the misspecification of conditional independence has been made. In this chapter, we presented an approach, based on previous studies on max-stable processes, that correctly quantifies the uncertainty in return level estimates under misspecification of the model. In particular, we described an adjustment to the likelihood that inflates the asymptotic posterior variance to adjust for spatial and temporal dependence in the data. Application of this approach to the precipitation data showed that model uncertainty increases relative to the approach where conditional independence is assumed correct. This chapter highlighted the importance of quantifying model uncertainty correctly given a model applied under false assumptions, which we hope serves as a useful guideline to practitioners when conducting risk assessments based on these models.

In Chapter 5, we introduced the first component of our Lagrangian model for European windstorms by presenting a stochastic mechanism for modelling the lifecycle and track of extratropical cyclones. As discussed in Chapter 5, extreme winds are generated in the wake of a storm track, which is defined as a path of local vorticity maxima. Motivated by findings from an exploratory data analysis, we developed a model from which synthetic tracks can be simulated by extracting the features of the track that determine its movement and intensity. In particular, we simulated initial conditions of track speed, direction, vorticity and location, before exploiting the temporal dependence structure of these variables to propagate the track forward in time and space. An extreme value approach is used to model the tail behaviour of vorticity, which allows us to simulate storms that are more intense than those observed. Monte Carlo simulation allowed us to estimate the probability of observing extreme events in a region around the UK, where we showed that the intensity of storm Herta over the UK corresponded to a return period of 107 years, whereas the return period of observing a storm of this intensity over parts of the North Atlantic is $10-25$ years. Our model serves as a useful tool for practitioners interested in quantifying risk associated with extratropical storms, in particular answering questions related to the spatial variability of storm tracks, the probability of observing extreme storms and where such storms are likely to occur.

In Chapter 6, we developed a model for European windstorms relative to synthetic extratropical cyclone tracks produced using the model in Chapter 5. We initially standardised wind speed fields over Europe and the North Atlantic onto common margins, which removed features dominated by marginal characteristics, specifically the contrast in wind speed magnitudes over land and sea. We specified some filtering and clustering methods for extracting the extreme winds generated by the extratropical cyclone system, before using an ellipse structure to summarise the position and spatial extent of the windstorm relative to the cyclone centre. We used a Markov model to propagate the characteristics of the ellipse in time, while specifying models for the activation and termination of a windstorm along the track. We also described a model for simulating physically realistic wind fields within the ellipse generated by the Markov model. We concluded with an illustration of how one could assess the joint risk of extreme windstorm events, with examples in northern England and Germany.

7.2 Further work

In this section, we present three possible ways which methods described in this thesis could be improved, along with some guidelines regarding how this could potentially be achieved.

7.2.1 Modelling the effect of climate indices on the track and intensity of extratropical cyclones

The natural variability of the climate can have a significant impact on weather patterns. This variability often manifests as an oscillation over a spatial domain, where atmospheric conditions in certain regions of a domain can offset one another. One of the first recognitions of this phenomenon was made by missionaries in Greenland in the 18th century, who noticed that severe winters there were concurrent with mild winters in Denmark, and vice-versa. These oscillations in weather patterns can arise from shifts in the climate system from a variety of sources, including sea-surface temperature and sea-level pressure. Today, these modes of climate variability are named and have been well-studied. The El Nino-Southern Oscillation (ENSO), for example, refers to variations in the temperature of the surface of the tropical eastern Pacific Ocean and in air surface pressure in the tropical western Pacific. However, in this analysis, we focus on what is widely-regarded to be the dominant mode in the North Atlantic region, the North Atlantic Oscillation (NAO).

The NAO describes the fluctuations in the difference of sea-level pressure between the Icelandic low and the Azores high. As this difference increases, corresponding to a positive NAO phase, the pressure gradient increases, leading to increased wind speed activity in the North Atlantic. Positive NAO indices are also consistent with a north-eastern shift in storm track activity, with a higher number of storms affecting Northern Europe than the Mediterranean region. In constrast, a negative NAO index tends to suppress westerly winds, causing the storm track to shift southwards towards the Mediterranean and northern Africa, resulting in cold, dry winters in northern Europe. The NAO can be described on a number of time-scales, with daily and monthly indices freely available. Figure 7.2.1 shows the difference between the density of tracks corresponding a negative NAO phase and tracks corresponding to a positive NAO phase. This suggests that the NAO could be a major predictor of extratropical cyclone activity, and the modelling approach in Chapter 5 could be improved by incorporating this phenomenon. However, NAO is difficult to forecast on short timescales, though there have been some studies showing trends in the decadal behaviour of the NAO (Hurrell, 1995).

Figure 7.2.1 clearly shows the effect of NAO on the movement of a storm track, which motivates an adjustment of our track model in Chapter 5 to incorporate this. A simple and natural approach would be to simulate a realisation of track speed V_t and direction

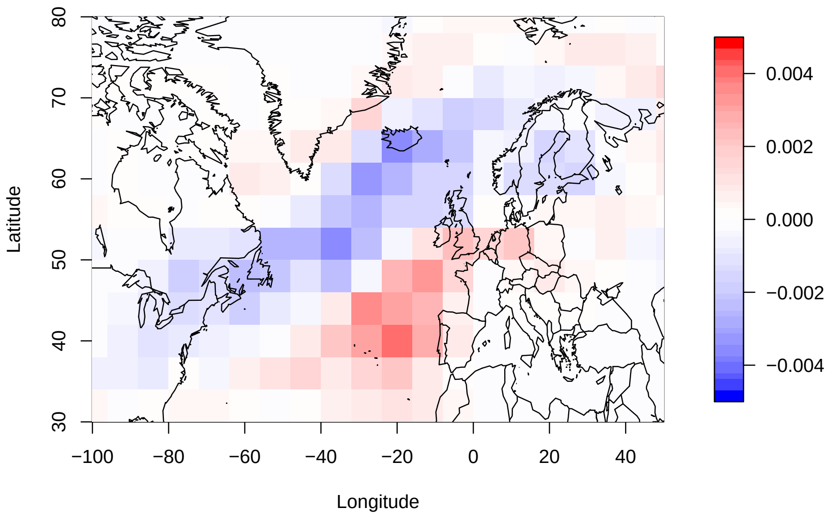


Figure 7.2.1: From analysis of the catalogue of observed storm tracks introduced in Chapter 5: the difference in track density between the set of storms corresponding to a negative NAO index and the set of storms corresponding to a positive NAO index.

Θ_t at time t conditional on the corresponding monthly NAO index. The monthly index is chosen to best represent the low-frequency variability of extratropical cyclones. Let $\boldsymbol{x}_j$ denotes the location of the storm at time j and Δ_j represent the grid cell corresponding to this location. For all times $1 \leq j \leq k$, where k is the order of the Markov process jointly for $\{(V_t, \Theta_t, \Omega_t)\}$, we adapt (5.4.1) and (5.4.2) by first simulating

$$\begin{aligned} \theta_j &\sim \Theta_j \mid \Theta_{0:j-1} = \theta_{0:j-1}, \boldsymbol{x}_j \in \Delta_j, \mathrm{NAO}_j \\ v_j &\sim V_j \mid V_{0:j-1} = v_{0:j-1}, \Theta_j = \theta_j, \boldsymbol{x}_j \in \Delta_j, \mathrm{NAO}_j. \end{aligned}$$

When $j > k$, we then simulate

$$\begin{aligned} \theta_j &\sim \Theta_j \mid \Theta_{j-k:j-1} = \theta_{j-k:j-1}, \boldsymbol{x}_j \in \Delta_j, \mathrm{NAO}_j \\ v_j &\sim V_j \mid V_{j-k:j-1} = v_{j-k:j-1}, \Theta_j = \theta_j, \boldsymbol{x}_j \in \Delta_j, \mathrm{NAO}_j. \end{aligned}$$

Figure 7.2.2 shows evidence of a weakly positive relationship between NAO index and the maximum vorticity of a storm. This reflects the northward shift of storm tracks during a positive NAO phase into regions where vorticity is highest (see Figure 5.2.4). NAO could be incorporated into simulating realisations of vorticity in the same way as with speed and direction of the track. The Monte Carlo approaches described in Chapter 5 could be used to estimate the probability of observing an extreme vorticity event given a specific phase of the NAO. Alternatively, given a long-term forecast, one could estimate the marginal risk of extreme events by averaging over all events with a specific NAO forecast.

Another interesting avenue of future work could involve the effect of NAO on the extremal behaviour of the process, in particular its effect on the marginal parameters of an extreme value model, or perhaps even the temporal dependence structure of an extreme vorticity event. One could follow the approach of Winter et al. (2016), who explored the effect of ENSO on heatwave characteristics, in our study by modelling NAO as a covariate in the dependence parameters of the conditional extremes approach. This would be useful in answering questions related to whether the NAO has an effect on the magnitude or duration of an extreme event.

7.2.2 Improved extraction of windstorm information

We described in Chapter 6 the procedure for extracting windstorm footprints from fields of wind speeds in the vicinity of the storm centre. In particular, we detailed the ellipse shape used to define the position and size of the footprint, as well as extracting other features such as its orientation and magnitude. This approach appears to perform well in practice, and in most cases the main features of the windstorm are detected. However, there is evidence to suggest that steps can be taken to increase robustness of the procedure.

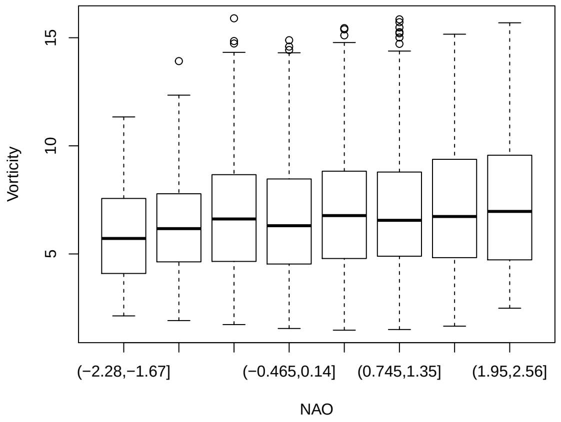


Figure 7.2.2: A boxplot defined at intervals of NAO of equal length showing the dependence between NAO and the maximum vorticity of a storm

The regions of extreme wind speeds remaining after the spatio-temporal filtering of the footprints are determined by masking below a pre-specified threshold ν. The choice of ν isn't straightforward as there is no intuition regarding what constitutes as 'extreme' for spatially filtered data. One should choose a high enough value of ν such that the convective events and ordinary winds are masked, but small enough so that genuine localised features of the windstorm aren't excluded. Sting jets, for example, are intense winds associated with downward flow originating from rapidly deepening depressions, which tend to occur on very small spatial scales. They are largely attributed to be the source of the destruction associated with the Great Storm of 1987 (Hewson and Neu, 2015). Therefore, we advise that a sensitivity study be carried out on the choice of ν such that the most damaging features of the windstorm be captured appropriately.

Empirical analysis of the extracted ellipses describing the windstorm event shows that, in some cases, the extracted ellipse does not evolve smoothly with respect to the storm centre. In some cases, the ellipse appears to 'jump' in location between consecutive time steps as a consequence of our extraction algorithm detecting another region of extreme winds. These may be a genuine feature of the windstorm, but in many cases, the algorithm detects spurious events that are not generated by the cyclone. After the spatio-temporal filtering of the footprints, the algorithm selects the largest region of wind speeds remaining, which may be an arguably simplistic criterion to use.

An alternative approach could be to define a score that incorporates multiple criteria in a detection strategy. These criteria could be defined using empirical observations or physical intuition regarding the behaviour of a windstorm with respect to the storm centre. For example, we know that the most extreme winds are generated along cold fronts, which are usually situated to the southwest of the storm centre. Let $i_j = 1, \ldots, n$ be the number of distinct clusters remaining after the spatio-temporal filtering of the footprints at time j. We could select $i_j^* \in \{1, \ldots, n\}$, the most likely cluster to be linked to a windstorm event, such that

$$i_j^* = \arg\max_i S_{i,j},$$

where $S_{i,j}$ is a user-specified score function evaluated for cluster i at time j. One example of such a score function could be

$$S_{i,j} = \omega_1 A_{i,j} + \omega_2 W_{i,j} + \omega_3 \exp\{-|\Theta_{i,j} - \pi/4|\} + \omega_4 \exp\{-|x_{i,j} - x_{i_{j-1}^*}|\},$$

where $A_{i,j}$, $W_{i,j}$, represent the area and maximum relative wind speed (both transformed to Uniform margins) respectively of cluster i and time j. The quantities $\Theta_{i,j}$ and $x_{i,j}$ represent the bearing (relative to south) with respect to the centre of the storm and position respectively, and $x_{i_{j-1}^*}$ denotes the position of the selected cluster at time $j-1$. The quantities $\omega_1, \ldots, \omega_4$ are weights defined such that $\sum_{k=1}^4 \omega_k = 1$, which are user-specified. The function $S_{i,j}$ gives high scores corresponding to clusters that have a large spatial extent and magnitude, while also giving higher scores to clusters located southeast relative to the storm centre. This function attempts to overcome the issue of windstorms ‘jumping’ in space between consecutive time steps by giving higher scores to clusters that are closer in position to the selected cluster at the previous time step, which could ensure that the cluster representing the windstorm evolves smoothly with respect to the storm centre. Figure 7.2.3 shows an example of how this score function chooses a different cluster compared to the approach we take in Chapter 6. We define the position in this example to be the centre of each ellipse. We choose $w_1 = 0.2$, $w_2 = 0.2$, $w_3 = 0.3$ and $w_4 = 0.3$. In this particular instance, the proposed score function chooses the correct cluster based on a desirable bearing with respect to the storm centre and the position of the cluster at the previous time step.

Another alternative approach to choosing the best cluster representing the windstorm is to use a multi-cluster strategy in defining a windstorm, which may offer more physical realism in representing its structure relative to the storm track. During extraction, one could specify a fixed maximum number of clusters to extract, or define some criterion on which to extract. For example, one could only extract the clusters whose size is greater than a certain threshold. When simulating windstorm events, one could model the number of clusters to simulate as a Poisson random variable. This would add additional practical complexity to the model, but is worth exploring as a potential means of improving the approach.

7.2.3 Developing a multi-hazard model

The work presented in this thesis, specifically in Chapters 5 and 6, could be the first step to developing a multi-hazard model for extratropical cyclones in Europe. This thesis has presented an approach for windstorms, the evolution of which we were able to model using an ellipse structure. Of course, other natural hazards arise as a result of

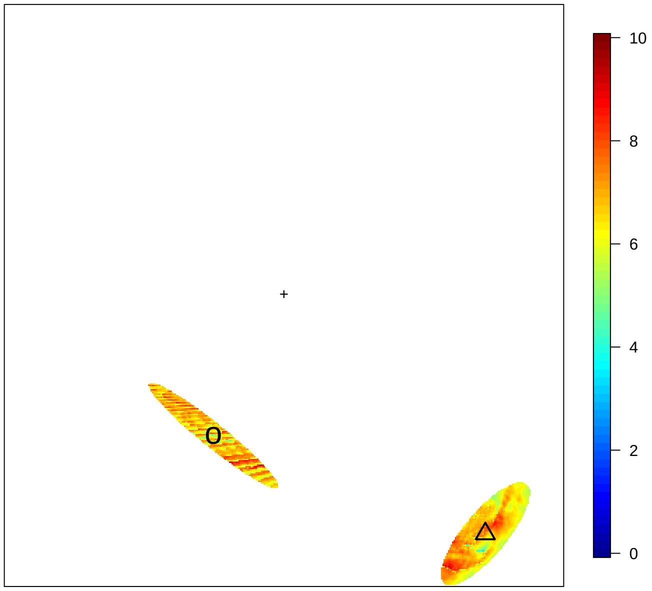


Figure 7.2.3: The clusters of relative wind speeds to choose under the feature extraction algorithm. The cluster containing the triangle was chosen using the approach described in Chapter 6, by taking the cluster of largest size. The cluster containing this circle was chosen under our proposed improved score function. The cross represents the centre of the storm.

extratropical cyclones, including precipitation and wave extremes, which pose similar infrastructural and societal risks. It represents a wise course of action to explore, in our proposed Lagrangian framework, how the spatial distribution of extreme precipitation and wave height evolves with respect to the storm centre.

We have discussed several times in this thesis how extratropical cyclones are commonly associated with the scenario when warm and cold air masses meet. The convergence of airmasses tends to result in an uplift in air, which can lead to precipitation events. The warm conveyor belt of the cyclone represents the main source of warm, moist air that feeds the cyclone. It originates in the warm sector of the cyclone and flows poleward parallel to the cold front east of the storm centre. It is typically associated with a large band of cloud, which produces extreme precipitation depending on the moisture circulation in the atmosphere. Rudeva and Gulev (2011) present empirical results showing that precipitation typically manifests east of the storm centre, and also that the magnitude of an extreme precipitation event increases as the intensity of the storm increases. This structure motivates the application of our Lagrangian statistical model to precipitation extremes relative to the centre of a storm.

One of the biggest risks to infrastructure is the threat of flooding from compound events, for example, extreme wind and rain events that occur simultaneously. Once an approach is developed for modelling extreme precipitation arising from extratropical cyclones, a natural area of further work would be to analyse some measures of joint risk. In this thesis, we have analysed the probability of multiple sites experiencing the same extreme windstorm event. Further work could aim to answer questions related to the probability of one site experiencing extreme winds and another experiencing extreme rain. For example, if location A upstream experienced high levels of rainfall and location B downstream saw strong winds, flood defences at location B could be threatened while a large volume of water is about to come downstream. The physical structure of the meteorological event is key to answering these types of questions, and it is wise to consider this information in the development of a statistical model. A model with a similar structure as in Chapter 6 could give physical basis coupled with statistical extrapolation to not only model multiple hazards simultaneously, but also the progression of the event generating these hazards in space and time.

Appendix A

Supplementary Material for Chapter 3

A.1 Proof: $\hat{\mu}_r = u$ when $m = r$

We can write the full likelihood for parameters $\boldsymbol{\theta}_r$ given a series of excesses $\{x_i\}$ above a threshold u as:

$$L(\boldsymbol{\theta}_r) = L_1 \times L_2,$$

where L_1 is the Poisson probability of r exceedances of u and L_2 is the joint density of these r exceedances, so that:

$$\begin{aligned}
L_1 &= \frac{1}{r!}\left\{ r\left[1+\xi\left(\frac{u-\mu_r}{\sigma_r}\right)\right]_+^{-1/\xi} \right\}^r \exp\left\{ -r\left[1+\xi\left(\frac{u-\mu_r}{\sigma_r}\right)\right]_+^{-1/\xi} \right\}, \\
L_2 &= \prod_{i-1}^{r} \frac{1}{\sigma_r}\left[1+\xi\left(\frac{x_i-\mu_r}{\sigma_r}\right)\right]_+^{-1/\xi-1} \left[1+\xi\left(\frac{u-\mu_r}{\sigma_r}\right)\right]_+^{1/\xi}.
\end{aligned}$$

By defining $\Lambda = \left[1+\xi\left(\frac{u-\mu_r}{\sigma_r}\right)\right]_+^{-1/\xi}$ and $\psi_u = \sigma_r + \xi(u-\mu_r)$ we can reparameterise the likelihood in terms of $\boldsymbol{\theta}^* = (\Lambda, \psi_u, \xi)$ to give:

$$\begin{aligned} L(\boldsymbol{\theta}^*) &\propto \Lambda^r \exp\{-r\Lambda\} \prod_{i=1}^{r} \frac{1}{\psi_u - \xi(u-\mu_r)} \left[\frac{\psi_u + \xi(x_i - u)}{\psi_u - \xi(u-\mu_r)}\right]_+^{-1/\xi-1} \left[\frac{\psi_u}{\psi_u - \xi(u-\mu_r)}\right]_+^{1/\xi} \\ &= \Lambda^r \exp\{-r\Lambda\} \prod_{i=1}^{r} \frac{1}{\psi_u}\left[1+\xi\left(\frac{x_i-u}{\psi_u}\right)\right]_+^{-1/\xi-1}. \end{aligned}$$

Taking the log-likelihood and maximising with respect to Λ, we get:

$$\begin{aligned} l(\boldsymbol{\theta}^*) &:= \log L(\boldsymbol{\theta}^*) = r\log\hat{\Lambda} - r\hat{\Lambda} - r\log\psi_u - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{r}\log\left[1+\xi\left(\frac{x_i-u}{\psi_u}\right)\right]_+ \\ \frac{\partial l}{\partial \Lambda} &= \frac{r}{\hat{\Lambda}} - r = 0, \end{aligned}$$

which gives $\hat{\Lambda} = 1$. Then, by the invariance property of maximum likelihood estimators, $\hat{\mu}_r = u$, and using the identity for ψ_u, we get $\hat{\sigma}_r = \hat{\psi}_u$. Because the ξ-dependent term in the log-likelihood is identical to that in a GP log-likelihood, the maximum likelihood estimators of the two models coincide.

A.2 Derivation of prior for inference on $\boldsymbol{\theta}_m$

We define a joint prior on the parameterisation of interest $\boldsymbol{\theta}_k$. However, as we are making inference for the 'optimal' parameterisation $\boldsymbol{\theta}_m$, we must derive the prior for $\boldsymbol{\theta}_m$. We can calculate the prior density of $\boldsymbol{\theta}_m$ by using the density method for one-to-one bivariate transformations. Inverting (3.1.5) to get expressions for μ_m and σ_m, i.e.

$$\begin{aligned} \mu_m &= \mu_k - \frac{\sigma_k}{\xi}\left(1-\left(\frac{m}{k}\right)^{-\xi}\right) = g_1(\mu_k, \sigma_k) \\ \sigma_m &= \sigma_k\left(\frac{m}{k}\right)^{-\xi} = g_2(\mu_k, \sigma_k), \end{aligned}$$

we can use this transformation to calculate the prior for $\boldsymbol{\theta}_m$.

$$\begin{aligned}\pi(\boldsymbol{\theta}_m) &= \pi(\mu_m, \sigma_m, \xi) \\ &= \pi(\mu_k, \sigma_k, \xi) |\det J|_{\mu_k = g_1^{-1}(\mu_m,\sigma_m), \sigma_k = g_2^{-1}(\mu_m,\sigma_m), \xi=\xi},\end{aligned}$$

where

$$\begin{aligned}\det J &= \begin{vmatrix} \frac{\partial \mu_m}{\partial \mu_k} & \frac{\partial \mu_m}{\partial \sigma_k} & \frac{\partial \mu_m}{\partial \xi} \\ \frac{\partial \sigma_m}{\partial \mu_k} & \frac{\partial \sigma_m}{\partial \sigma_k} & \frac{\partial \sigma_m}{\partial \xi} \\ \frac{\partial \xi}{\partial \mu_k} & \frac{\partial \xi}{\partial \sigma_k} & \frac{\partial \xi}{\partial \xi} \end{vmatrix} \\ &= \begin{vmatrix} \frac{\partial \mu_m}{\partial \mu_k} & \frac{\partial \mu_m}{\partial \sigma_k} & \frac{\partial \mu_m}{\partial \xi} \\ 0 & \frac{\partial \sigma_m}{\partial \sigma_k} & \frac{\partial \sigma_m}{\partial \xi} \\ 0 & 0 & \frac{\partial \xi}{\partial \xi} \end{vmatrix} \\ &= \frac{\partial \sigma_m}{\partial \sigma_k} \frac{\partial \xi}{\partial \xi} \\ &= \left(\frac{m}{k}\right)^{-\xi}.\end{aligned}$$

Therefore, $\pi(\boldsymbol{\theta}_m) = \left(\frac{m}{k}\right)^{-\xi} \pi(\boldsymbol{\theta}_k)$.

A.3 Fisher information matrix calculations for i.i.d random variables

The log-likelihood of the Poisson process model with parameterisation $\boldsymbol{\theta}_m = (\mu_m, \sigma_m, \xi)$ can be expressed as

$$l(\boldsymbol{\theta}_m) = -m \left[1 + \xi \left(\frac{u - \mu_m}{\sigma_m}\right)\right]_+^{-1/\xi} - r \log \sigma_m - \left(\frac{1}{\xi} + 1\right) \sum_{j=1}^{r} \log \left[1 + \xi \left(\frac{x_j - \mu_m}{\sigma_m}\right)\right]_+,$$

where r is the number of exceedances of X above the threshold u. For simplicity, we drop the $[\cdot]_+$ subscript in subsequent calculations. In order to produce analytic expressions for the asymptotic covariance matrix, we must evaluate the observed information matrix $\hat{I}(\boldsymbol{\theta}_m)$. For simplicity, we define $v_m = \frac{u-\mu_m}{\sigma_m}$ and $z_{j,m} = \frac{x_j-\mu_m}{\sigma_m}$.

$$\begin{aligned}
\frac{\partial^2 l}{\partial \mu_m^2} &= -\frac{m(\xi+1)}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-2} + \frac{\xi(\xi+1)}{\sigma_m^2}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-2},\\
\frac{\partial^2 l}{\partial \sigma_m^2} &= \frac{2m}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-1}v_m - \frac{m(\xi+1)}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}v_m^2 + \frac{r}{\sigma_m^2} -\\
&\quad \frac{2(\xi+1)}{\sigma_m^2}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-1} z_{j,m} + \frac{\xi(\xi+1)}{\sigma_m^2}\sum_{j=1}^{r} z_{j,m}^2\left[1+\xi z_{j,m}\right]^{-2},\\
\frac{\partial^2 l}{\partial \xi^2} &= -m[1+\xi v_m]^{-1/\xi}\left[\frac{1}{\xi}v_m^2[1+\xi v_m]^{-2} - \frac{2}{\xi^3}\log\left[1+\xi v_m\right]\right.\\
&\quad \left. + \frac{2}{\xi^2}[1+\xi v_m]^{-1}v_m + \left(\frac{1}{\xi^2}\log\left[1+\xi v_m\right] - \frac{1}{\xi}[1+\xi v_m]^{-1}v_m\right)^2\right]\\
&\quad -\frac{2}{\xi^3}\sum_{j=1}^{r}\log\left[1+\xi z_{j,m}\right] + \frac{2}{\xi^2}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-1} z_{j,m} +\\
&\quad \frac{\xi+1}{\xi}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-2} z_{j,m}^2,\\
\frac{\partial^2 l}{\partial \mu_m \partial \sigma_m} &= \frac{m}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-1} - \frac{m(\xi+1)}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}v_m\\
&\quad -\frac{\xi+1}{\sigma_m^2}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-1} + \frac{\xi(\xi+1)}{\sigma_m^2}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-2} z_{j,m},\\
\frac{\partial^2 l}{\partial \mu_m \partial \xi} &= -\frac{m}{\sigma_m}\left[\frac{1}{\xi^2}[1+\xi v_m]^{-1/\xi-1}\log\left[1+\xi v_m\right] - \frac{\xi+1}{\xi}[1+\xi v_m]^{-1/\xi-2}v_m\right] +\\
&\quad \frac{1}{\sigma_m}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-1} - \frac{\xi+1}{\sigma_m}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-2} z_{j,m},\\
\frac{\partial^2 l}{\partial \sigma_m \partial \xi} &= -\frac{m}{\sigma_m}v_m\left[\frac{1}{\xi^2}[1+\xi v_m]^{-1/\xi-1}\log\left[1+\xi v_m\right] - \frac{\xi+1}{\xi}[1+\xi v_m]^{-1/\xi-2}v_m\right] +\\
&\quad \frac{1}{\sigma_m}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-1} z_{j,m} - \frac{\xi+1}{\sigma_m}\sum_{j=1}^{r}\left[1+\xi z_{j,m}\right]^{-2} z_{j,m}^2
\end{aligned}$$

To obtain the Fisher information matrix, we take the expected value of each term in the observed information with respect to the probability density of points of a Poisson process. Let $Z = \frac{X-\mu_m}{\sigma_m}$, and R be a random variable denoting the number of excesses of X above u. The density of points in the set A_u can de defined by

$$f(x) = \frac{\lambda(x)}{\Lambda(A_u)} = \frac{[1+\xi z]^{-1/\xi-1}}{[1+\xi v_m]^{-1/\xi}},$$

where λ is a function denoting the rate of exceedance. Then, for example,

$$\begin{aligned}
\mathbb{E}_{Z,R}\left\{\sum_{j=1}^{R}[1+\xi z_{j,m}]^{-2}\right\} &= \mathbb{E}_R\mathbb{E}_{Z|R}\left\{\sum_{j=1}^{R}[1+\xi z_{j,m}]^{-2}\right\} \\
&= \mathbb{E}_R\left\{R\mathbb{E}_Z\left\{[1+\xi Z]^{-2}\right\}\right\} \\
&= \mathbb{E}_R\left\{R[1+\xi v_m]^{1/\xi}\int_{v_m}^{\infty}[1+\xi z]^{-1/\xi-3}\,\mathrm{d}z\right\} \\
&= \frac{m}{2\xi+1}[1+\xi v_m]^{-1/\xi-2}
\end{aligned}$$

Following this process, we can write the Fisher information matrix $I(\boldsymbol{\theta}_m)$ as:

$$\begin{aligned}
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m^2}\right\} &= \frac{m(\xi+1)}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-2} - \frac{m\xi(\xi+1)}{(2\xi+1)\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}, \\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\sigma_m^2}\right\} &= -\frac{2m}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-1}v_m + \frac{m(\xi+1)}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}v_m^2 - \frac{r}{\sigma_m^2} + \\
&\quad \frac{2m}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-1}\left[1+(\xi+1)v_m\right] - \\
&\quad \frac{m\xi}{(2\xi+1)\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}\left[(2\xi^2+3\xi+1)v_m^2+(4\xi+2)v_m+2\right], \\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\xi^2}\right\} &= m[1+\xi v_m]^{-1/\xi}\left[\frac{1}{\xi}v_m^2[1+\xi v_m]^{-2} - \frac{2}{\xi^3}\log[1+\xi v_m] + \right. \\
&\quad \left.\frac{2}{\xi^2}[1+\xi v_m]^{-1}v_m + \left(\frac{1}{\xi^2}\log[1+\xi v_m] - \frac{1}{\xi}[1+\xi v_m]^{-1}v_m\right)^2\right] + \\
&\quad \frac{2}{\xi^3}[1+\xi v_m]^{-1/\xi}\left[\xi+\log[1+\xi v_m]\right] - \\
&\quad \frac{2m}{(\xi+1)\xi^2}[1+\xi v_m]^{-1/\xi-1}\left[1+(\xi+1)v_m\right] -
\end{aligned}$$

$$
\begin{aligned}
& \frac{m}{\xi(2\xi+1)}[1+\xi v_m]^{-1/\xi-2}\left[(2\xi^2+3\xi+1)v_m^2+(4\xi+2)v_m+2\right], \\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m\partial\sigma_m}\right\} &= \frac{m(\xi+1)}{\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}v_m - \\
& \frac{m\xi}{(2\xi+1)\sigma_m^2}[1+\xi v_m]^{-1/\xi-2}\left[1+(2\xi+1)v_m\right], \\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m\partial\xi}\right\} &= \frac{m}{\sigma_m}\left[\frac{1}{\xi^2}[1+\xi v_m]^{-1/\xi-1}\log\left[1+\xi v_m\right]-\frac{\xi+1}{\xi}[1+\xi v_m]^{-1/\xi-2}v_m\right] - \\
& \frac{m}{\sigma_m(\xi+1)}[1+\xi v_m]^{-1/\xi-1} + \\
& \frac{m}{\sigma_m(2\xi+1)}[1+\xi v_m]^{-1/\xi-2}\left[1+(2\xi+1)v_m\right], \\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\sigma_m\partial\xi}\right\} &= \frac{m}{\sigma_m}v_m\left[\frac{1}{\xi^2}[1+\xi v_m]^{-1/\xi-1}\log\left[1+\xi v_m\right]+\frac{\xi+1}{\xi}[1+\xi v_m]^{-1/\xi-2}v_m\right] - \\
& \frac{m}{\sigma_m(\xi+1)}[1+\xi v_m]^{-1/\xi-1}\left[1+(\xi+1)v_m\right] + \\
& \frac{m}{\sigma_m(2\xi+1)}[1+\xi v_m]^{-1/\xi-2}\left[(2\xi^2+3\xi+1)v_m^2+(4\xi+2)v_m+2\right].
\end{aligned}
$$

By inverting the Fisher information matrix using a technical computing tool like Wolfram Mathematica, making the substitution $r = m[1+\xi v_m]^{-1/\xi}$, the expected number of exceedances, and using the mapping in (3.1.5), we can get expressions for asymptotic posterior covariances.

$$
\begin{aligned}
\mathrm{ACov}(\mu_m,\xi) &= \frac{1}{\xi^2 r}(\xi+1)\sigma_m\left(\frac{r}{m}\right)^{-\xi}\left(\xi(\xi+1)\left(\frac{r}{m}\right)^{\xi}\log\left(\frac{r}{m}\right)-(2\xi+1)\left(\left(\frac{r}{m}\right)^{\xi}-1\right)\right) \\
\mathrm{ACov}(\mu_m,\sigma_m) &= \frac{1}{\xi^2 r}\sigma_m^2\left(\frac{r}{m}\right)^{-\xi}\Bigg(\left(\frac{r}{m}\right)^{\xi}\Bigg((\xi+1)\log\left(\frac{r}{m}\right)\left((\xi+1)\xi\log\left(\frac{r}{m}\right)-3\xi-1\right)+ \\
& \qquad \xi(\xi(\xi+2)+3)+1\Bigg)+(\xi+1)(2\xi+1)\left(\log\left(\frac{r}{m}\right)-1\right)\Bigg) \\
\mathrm{ACov}(\sigma_m,\xi) &= \frac{1}{r}(\xi+1)\sigma_m\left((\xi+1)\log\left(\frac{r}{m}\right)-1\right)
\end{aligned}
$$

When $m = r$, $\mathrm{ACov}(\mu_m,\xi) = 0$. In addition, the m for which $\mathrm{ACov}(\mu_m,\sigma_m) = 0$ coincides with the value of m that minimises $\rho_{\boldsymbol{\theta}_m}$ as defined in (3.3.1). This root can easily be found numerically, but an analytical approximation can be calculated using a one-step Halley's method. By using $m = r$ as the initial seed, and using the formula:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n) - \frac{f(x_n)f''(x_n)}{2f'(x_n)}}$$

we get the expression (3.3.6) for $\hat{m}_2$ after one step. The quantity for $\hat{m}_1$, given by expression (3.3.5) requires two iterations of this method.

A.4 Fisher information matrix calculations corresponding to a linear trend in the location parameter

We allow for a linear covariate z in the location parameter such that $\mu_m(z) = \mu_m^{(0)} + \mu_m^{(1)} z$. As the integrated intensity requires integrating over the probability density function g of the covariate space, which is unknown, the log-likelihood of the non-stationary model becomes

$$l = -m \int_z \left[1 + \xi\left(\frac{u - \mu_m(z)}{\sigma_m}\right)\right]^{-1/\xi} g(z)\mathrm{d}z - r\log\sigma_m - \left(\frac{1}{\xi} - 1\right)\sum_{j=1}^{r}\log\left[1 + \xi\left(\frac{x_j - \mu_m(z)}{\sigma_m}\right)\right].$$

Making the substitutions $v_{m,z} = \frac{u - \mu_m(z)}{\sigma_m}$ and $w_{j,m,z} = \frac{x_j - \mu_m(z)}{\sigma_m}$, components of the observed information matrix $\hat{I}(\mu_m^{(0)}, \mu_m^{(1)}, \sigma_m, \xi)$ can be expressed as

$$\begin{aligned}
\frac{\partial^2 l}{\partial {\mu_m^{(0)}}^2} &= -\frac{m(\xi+1)}{\sigma_m^2}\int_z \left[1 + \xi v_{m,z}\right]^{-1/\xi-2} g(z)\mathrm{d}z + \frac{\xi(\xi+1)}{\sigma_m^2}\sum_{j=1}^{r}\left[1 + \xi w_{j,m,z}\right]^{-2}, \\
\frac{\partial^2 l}{\partial {\mu_m^{(1)}}^2} &= -\frac{m(\xi+1)}{\sigma_m^2}\int_z \left[1 + \xi v_{m,z}\right]^{-1/\xi-2} z^2 g(z)\mathrm{d}z + \\
&\quad \frac{\xi(\xi+1)}{\sigma_m^2}\sum_{j=1}^{r} {x_j}^2 \left[1 + \xi w_{j,m,z}\right]^{-2}, \\
\frac{\partial^2 l}{\partial \sigma_m^2} &= \frac{2m}{\sigma_m^2}\int_z \left[1 + \xi v_{m,z}\right]^{-1/\xi-1} v_{m,z} g(z)\mathrm{d}z - \frac{m(\xi+1)}{\sigma_m^2}\int_z \left[1 + \xi v_{m,z}\right]^{-1/\xi-2} v_{m,z}^2 g(z)\mathrm{d}z +
\end{aligned}$$

$$
\begin{aligned}
&\frac{r}{\sigma_m^2} - \frac{2(\xi+1)}{\sigma_m^2} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-1} w_{j,m,z} + \\
&\frac{\xi(\xi+1)}{\sigma_m^2} \sum_{j=1}^{r} w_{j,m,z}^2 \left[1+\xi w_{j,m,z}\right]^{-2}, \\
\frac{\partial^2 l}{\partial \xi^2} &= -m \int_z \Bigg\{ [1+\xi v_{m,z}]^{-1/\xi} \Bigg[\frac{1}{\xi} v_{m,z}^2 [1+\xi v_{m,z}]^{-2} - \frac{2}{\xi^3} \log\left[1+\xi v_{m,z}\right] \\
&+ \frac{2}{\xi^2} [1+\xi v_{m,z}]^{-1} v_{m,z} + \left(\frac{1}{\xi^2} \log\left[1+\xi v_{m,z}\right] - \frac{1}{\xi} [1+\xi v_{m,z}]^{-1} v_{m,z} \right)^2 \Bigg] \Bigg\} g(z) \mathrm{d}z \\
&- \frac{2}{\xi^3} \sum_{j=1}^{r} \log\left[1+\xi w_{j,m,z}\right] + \frac{2}{\xi^2} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-1} w_{j,m,z} + \\
&\frac{\xi+1}{\xi} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-2} w_{j,m,z}^2, \\
\frac{\partial^2 l}{\partial \mu_m^{(0)} \partial \mu_m^{(1)}} &= -\frac{m(\xi+1)}{\sigma_m^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-2} z g(z) \mathrm{d}z + \\
&\frac{\xi(\xi+1)}{\sigma_m^2} \sum_{j=1}^{r} x_j \left[1+\xi w_{j,m,z}\right]^{-2}, \\
\frac{\partial^2 l}{\partial \mu_m^{(0)} \partial \sigma_m} &= \frac{m}{\sigma_m^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-1} g(z) \mathrm{d}z - \frac{m(\xi+1)}{\sigma_m^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z} g(z) \mathrm{d}z \\
&- \frac{\xi+1}{\sigma_m^2} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-1} + \frac{\xi(\xi+1)}{\sigma_m^2} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-2} w_{j,m,z}, \\
\frac{\partial^2 l}{\partial \mu_m^{(1)} \partial \sigma_m} &= \frac{m}{\sigma_m^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-1} z g(z) \mathrm{d}z - \frac{m(\xi+1)}{\sigma_m^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z} z g(z) \mathrm{d}z \\
&- \frac{\xi+1}{\sigma_m^2} \sum_{j=1}^{r} x_j \left[1+\xi w_{j,m,z}\right]^{-1} + \frac{\xi(\xi+1)}{\sigma_m^2} \sum_{j=1}^{r} x_j \left[1+\xi w_{j,m,z}\right]^{-2} w_{j,m,z}, \\
\frac{\partial^2 l}{\partial \mu_m^{(0)} \partial \xi} &= -\frac{m}{\sigma_m} \Bigg[\frac{1}{\xi^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-1} \log\left[1+\xi v_{m,z}\right] g(z) \mathrm{d}z - \\
&\frac{\xi+1}{\xi} \int_z [1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z} g(z) \mathrm{d}z \Bigg] + \frac{1}{\sigma_m} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-1} - \\
&\frac{\xi+1}{\sigma_m} \sum_{j=1}^{r} \left[1+\xi w_{j,m,z}\right]^{-2} w_{j,m,z}, \\
\frac{\partial^2 l}{\partial \mu_m^{(1)} \partial \xi} &= -\frac{m}{\sigma_m} \Bigg[\frac{1}{\xi^2} \int_z [1+\xi v_{m,z}]^{-1/\xi-1} \log\left[1+\xi v_{m,z}\right] z g(z) \mathrm{d}z - \\
&\frac{\xi+1}{\xi} \int_z [1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z} g(z) z \mathrm{d}z \Bigg] + \frac{1}{\sigma_m} \sum_{j=1}^{r} x_j \left[1+\xi w_{j,m,z}\right]^{-1} -
\end{aligned}
$$

$$\frac{\xi+1}{\sigma_m}\sum_{j=1}^{r} x_j\left[1+\xi w_{j,m,z}\right]^{-2} w_{j,m,z},$$

$$\frac{\partial^2 l}{\partial\sigma_m\partial\xi} = -\frac{m}{\sigma_m}\int_z\left\{v_{m,z}\left[\frac{1}{\xi^2}[1+\xi v_{m,z}]^{-1/\xi-1}\log\left[1+\xi v_{m,z}\right]- \frac{\xi+1}{\xi}[1+\xi v_{m,z}]^{-1/\xi-2}v_{m,z}\right]\right\}g(z)\mathrm{d}z+\frac{1}{\sigma_m}\sum_{j=1}^{r}\left[1+\xi w_{j,m,z}\right]^{-1}w_{j,m,z}- \frac{\xi+1}{\sigma_m}\sum_{j=1}^{r}\left[1+\xi w_{j,m,z}\right]^{-2}w_{j,m,z}^2$$

Following the same procedure as for the i.i.d. model, we can express the components of the expected information matrix as

$$\mathbb{E}\left\{-\frac{\partial^2 l}{\partial{\mu_m^{(0)}}^2}\right\} = \frac{m(\xi+1)}{\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-2}g(z)\mathrm{d}z- \frac{m\xi(\xi+1)}{(2\xi+1)\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-2}g(z)\mathrm{d}z,$$

$$\mathbb{E}\left\{-\frac{\partial^2 l}{\partial{\mu_m^{(1)}}^2}\right\} = \frac{m(\xi+1)}{\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-2}zg(z)\mathrm{d}z- \frac{m\xi(\xi+1)}{(2\xi+1)\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-2}zg(z)\mathrm{d}z,$$

$$\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\sigma_m^2}\right\} = -\frac{2m}{\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-1}v_{m,z}g(z)\mathrm{d}z+ \frac{m(\xi+1)}{\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-2}v_{m,z}^2g(z)\mathrm{d}z-\frac{r}{\sigma_m^2} +\frac{2m}{\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-1}\left[1+(\xi+1)v_{m,z}\right]g(z)\mathrm{d}z -\frac{m\xi}{(2\xi+1)\sigma_m^2}\int_z[1+\xi v_{m,z}]^{-1/\xi-2}\times \left[(2\xi^2+3\xi+1)v_{m,z}^2+(4\xi+2)v_{m,z}+2\right]g(z)\mathrm{d}z,$$

$$\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\xi^2}\right\} = m\int_z\left\{[1+\xi v_{m,z}]^{-1/\xi}\left[\frac{1}{\xi}v_{m,z}^2[1+\xi v_{m,z}]^{-2}-\frac{2}{\xi^3}\log\left[1+\xi v_{m,z}\right] +\frac{2}{\xi^2}[1+\xi v_{m,z}]^{-1}v_{m,z}+ \left(\frac{1}{\xi^2}\log\left[1+\xi v_{m,z}\right]-\frac{1}{\xi}[1+\xi v_{m,z}]^{-1}v_{m,z}\right)^2\right]\right\}g(z)\mathrm{d}z$$

$$
\begin{aligned}
&+\frac{2}{\xi^3}\int_z [1+\xi v_{m,z}]^{-1/\xi}\left[\xi+\log\left[1+\xi v_{m,z}\right]\right] g(z)\mathrm{d}z\\
&-\frac{2m}{(\xi+1)\xi^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-1}\left[1+(\xi+1)v_{m,z}\right] g(z)\mathrm{d}z-\\
&\frac{m}{\xi(2\xi+1)}\int_z [1+\xi v_{m,z}]^{-1/\xi-2}\times\\
&\left[(2\xi^2+3\xi+1)v_{m,z}^2+(4\xi+2)v_{m,z}+2\right] g(z)\mathrm{d}z,
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m^{(0)}\partial\mu_m^{(1)}}\right\} &= \frac{m(\xi+1)}{\sigma_m^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-2} z g(z)\mathrm{d}z-\\
&\frac{\xi(\xi+1)}{(2\xi+1)\sigma_m^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-2} z g(z)\mathrm{d}z,\\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m^{(0)}\partial\sigma_m}\right\} &= \frac{m(\xi+1)}{\sigma_m^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z} g(z)\mathrm{d}z-\\
&\frac{m\xi}{(2\xi+1)\sigma^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-2}\left[1+(2\xi+1)v_{m,z}\right] g(z)\mathrm{d}z,\\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m^{(1)}\partial\sigma_m}\right\} &= \frac{m(\xi+1)}{\sigma_m^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z} z g(z)\mathrm{d}z-\\
&\frac{m\xi}{(2\xi+1)\sigma_m^2}\int_z [1+\xi v_{m,z}]^{-1/\xi-2}\left[1+(2\xi+1)v_{m,z}\right] z g(z)\mathrm{d}z,\\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m^{(0)}\partial\xi}\right\} &= \frac{m}{\sigma_m}\int_z \left[\frac{1}{\xi^2}[1+\xi v_{m,z}]^{-1/\xi-1}\log\left[1+\xi v_{m,z}\right]-\right.\\
&\left.\frac{\xi+1}{\xi}[1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z}\right] g(z)\mathrm{d}z-\\
&\frac{m}{\sigma_m(\xi+1)}\int_z [1+\xi v_{m,z}]^{-1/\xi-1} g(z)\mathrm{d}z+\\
&\frac{m}{\sigma_m(2\xi+1)}\int_x [1+\xi v_{m,z}]^{-1/\xi-2}\left[1+(2\xi+1)v_{m,z}\right] g(z)\mathrm{d}z,\\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\mu_m^{(1)}\partial\xi}\right\} &= \frac{m}{\sigma_m}\int_z \left[\frac{1}{\xi^2}[1+\xi v_{m,z}]^{-1/\xi-1}\log\left[1+\xi v_{m,z}\right]-\right.\\
&\left.\frac{\xi+1}{\xi}[1+\xi v_{m,z}]^{-1/\xi-2} v_{m,z}\right] z g(z)\mathrm{d}z-\\
&\frac{m}{\sigma_m(\xi+1)}\int_z [1+\xi v_{m,z}]^{-1/\xi-1} z g(z)\mathrm{d}z+\\
&\frac{m}{\sigma_m(2\xi+1)}\int_z [1+\xi v_{m,z}]^{-1/\xi-2}\left[1+(2\xi+1)v_{m,z}\right] z g(z)\mathrm{d}z,\\
\mathbb{E}\left\{-\frac{\partial^2 l}{\partial\sigma_m\partial\xi}\right\} &= \frac{m}{\sigma_m}\int_z v_{m,z}\left[\frac{1}{\xi^2}[1+\xi v_{m,z}]^{-1/\xi-1}\log\left[1+\xi v_{m,z}\right]+\right.
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\xi+1}{\xi}[1+\xi v_{m,z}]^{-1/\xi-2}v_{m,z}\Bigg] g(z)\mathrm{d}z-\\
&\frac{m}{\sigma_m(\xi+1)}\int_z [1+\xi v_{m,z}]^{-1/\xi-1}\left[1+(\xi+1)v_{m,z}\right] g(z)\mathrm{d}z+\\
&\frac{m}{\sigma_m(2\xi+1)}\int_z [1+\xi v_{m,z}]^{-1/\xi-2}\times\\
&\left[(2\xi^2+3\xi+1)v_{m,z}^2+(4\xi+2)v_{m,z}+2\right] g(z)\mathrm{d}z.
\end{aligned}
$$

Like in the i.i.d. case, we invert the expected Fisher information matrix to obtain expressions for the asymptotic posterior covariances. Because of the integral term, there exists no analytical solution for the value of m that minimises the asymptotic posterior covariance between two given parameters. Therefore, we use a numerical solver to obtain an estimate of m that produces a near-orthogonal representation of the parameters.

Appendix B

Supplementary Material for Chapter 4

B.1 Parameters of spatial model

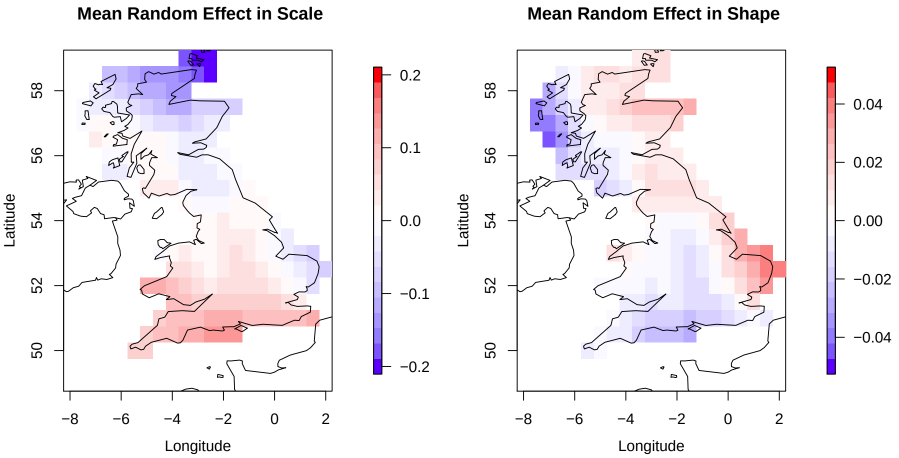
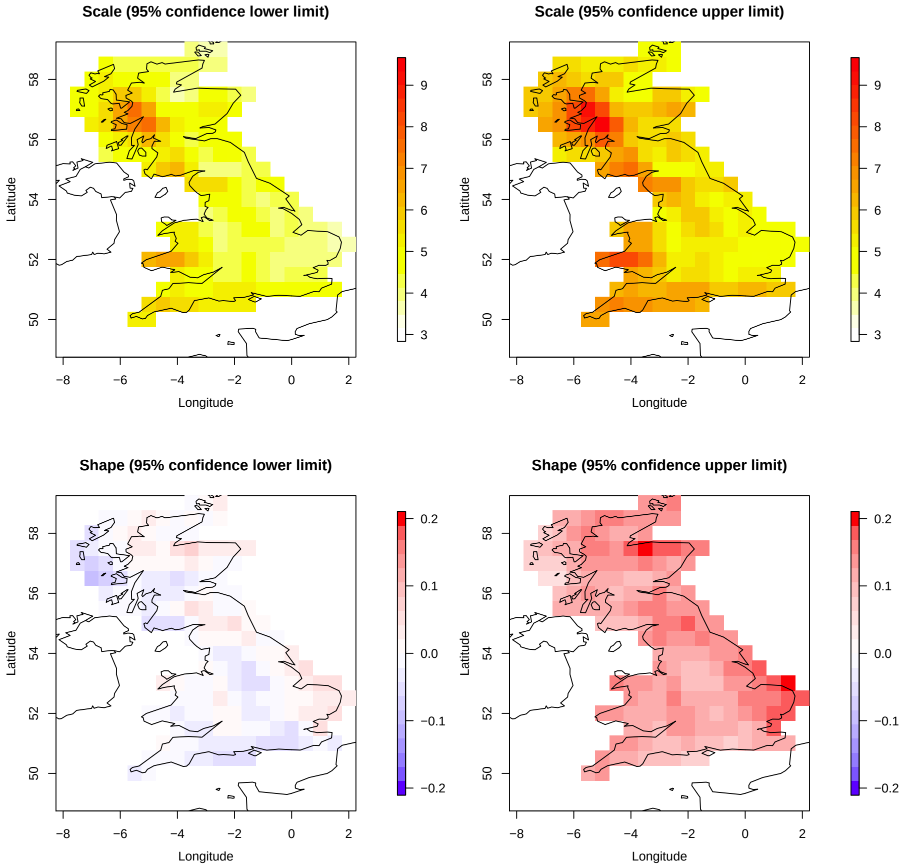


Figure B.1.1: The mean of the estimated random effect term $\boldsymbol{\phi}$ corresponding to the scale (left) and shape (right) parameters at each site.



Scale (95% confidence lower limit)

Scale (95% confidence upper limit)

Shape (95% confidence lower limit)

Shape (95% confidence upper limit)

Figure B.1.2: The estimated 95% posterior credible intervals of the scale (left) and shape (right) parameters under the spatial model.

Appendix C

Supplementary material for Chapters 5 and 6

C.1 Conditional kernel density estimation

Consider an arbitrary d-dimensional random vector $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_d)$, which is observed n times $\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \ldots, \boldsymbol{z}^{(n)}$. As a way of estimating $f(\boldsymbol{z})$, the joint probability density of $\boldsymbol{Z}$, we define the multivariate kernel density estimator as

$$\hat{f}(\boldsymbol{z}) = \frac{1}{n} \sum_{i=1}^{n} K_H \left(\boldsymbol{z} - \boldsymbol{z}^{(i)} \right), \tag{C.1.1}$$

where K_H is the kernel function and H denotes the bandwidth matrix which is symmetric and positive-definite. For our purposes, we choose K_H to be the multivariate Gaussian density function

$$K_H(\boldsymbol{z}) = (2\pi)^{-d/2} |H|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{z}^T H^{-1} \boldsymbol{z} \right\} \tag{C.1.2}$$

and the bandwidth matrix H chosen to be proportional to the rule-of-thumb selection of Scott (1992). The bandwidth matrix H can be chosen to be diagonal or oriented.

To simulate from the kernel density, we first sample uniformly a tuple $\boldsymbol{z}^{(i)}$, where $i \in \{1, \ldots, n\}$. We then simulate a vector $\tilde{\boldsymbol{z}}$, say, such that $\tilde{\boldsymbol{z}} \sim \mathrm{MVN}(\boldsymbol{z}^{(i)}, H)$.

Let $\boldsymbol{Z}$ be decomposed such that $\boldsymbol{Z} = (\boldsymbol{Z}_{-m}, \boldsymbol{Z}_m)$. Consider the case when values $\boldsymbol{Z}_{-m} = \boldsymbol{z}_{-m}$ have been observed and we wish to estimate the conditional density of $\boldsymbol{Z}_m$ given these values. We can then define the conditional kernel density estimator as

$$\hat{f}(\boldsymbol{z}_m | \boldsymbol{z}_{-m}) = \sum_{i=1}^{n} w_i(\boldsymbol{z}_{-m}) K_H \left(\boldsymbol{z}_m - \boldsymbol{z}_m^{(i)} \middle| \boldsymbol{z}_{-m} - \boldsymbol{z}_{-m}^{(i)} \right), \tag{C.1.3}$$

where

$$w_i(\boldsymbol{z}_{-m}) = \frac{K_H \left(\boldsymbol{z}_{-m} - \boldsymbol{z}_{-m}^{(i)} \right)}{\sum_{j=1}^{n} K_H \left(\boldsymbol{z}_{-m} - \boldsymbol{z}_{-m}^{(j)} \right)},$$

where $K_H(\cdot)$ is the multivariate Gaussian kernel function and $K_H(\cdot \mid \cdot)$ is the conditional Gaussian kernel function with bandwidth matrix H as defined in equation (C.1.2). Let H be partitioned such that

$$H = \begin{bmatrix} H_{m,m} & H_{m,-m} \\ H_{-m,m} & H_{-m,-m} \end{bmatrix}.$$

Conditional on having observed $\boldsymbol{z}_{-m}$, we choose a tuple $\boldsymbol{z}^{(i)}$ with probability $w_i(\boldsymbol{z}_{-m})$. Then we simulate

$$\boldsymbol{Z}_m | (\boldsymbol{Z}_{-m} = \boldsymbol{z}_{-m}) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}), \tag{C.1.4}$$

where $\bar{\boldsymbol{\mu}} = z_m^{(i)} + H_{m,-m} H_{-m,-m}^{-1} (\boldsymbol{z}_{-m} - \boldsymbol{z}_{-m}^{(i)})$ and $\bar{\boldsymbol{\Sigma}} = H_{m,m} - H_{m,-m} H_{-m,-m}^{-1} H_{-m,m}$.

Bibliography

Akhtar, N., Brauch, J., Dobler, A., Béranger, K., and Ahrens, B. (2014). Medicanes in an ocean–atmosphere coupled regional climate model. *Natural Hazards and Earth System Sciences*, 14:2189–2201.

Apputhurai, P. and Stephenson, A. G. (2013). Spatiotemporal hierarchical modelling of extreme precipitation in Western Australia using anisotropic Gaussian random fields. *Environmental and Ecological Statistics*, 20(4):667–677.

Attalides, N. (2015). *Threshold-based extreme value modelling.* PhD thesis, UCL (University College London).

Baker, L. (2009). Sting jets in severe northern European wind storms. *Weather*, 64(6):143–148.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* CRC Press.

Bengtsson, L., Hodges, K. I., and Roeckner, E. (2006). Storm tracks and climate change. *Journal of Climate*, 19(15):3518–3543.

Bjerknes, J. (1919). On the structure of moving cyclones. *Monthly Weather Review*, 47(2):95–99.

Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S. (2012). The spatial structure of

European wind storms as characterized by bivariate extreme-value copulas. *Natural Hazards and Earth System Sciences*, 12(5):1769–1782.

Bortot, P. and Tawn, J. A. (1998). Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867.

Brown, S. J. (2017). The drivers of variability in UK extreme rainfall. *International Journal of Climatology*.

Casson, E. and Coles, S. G. (2000). Simulation and extremal analysis of hurricane events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(3):227–245.

Catto, J. L., Shaffrey, L. C., and Hodges, K. I. (2010). Can climate models capture the structure of extratropical cyclones? *Journal of Climate*, 23(7):1621–1635.

Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.

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

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

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

Coles, S. G. and Powell, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, 64(1):119–136.

Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45:463–478.

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43:139–157.

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

Cooley, D. and Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental statistics*, 15(3):381–402.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1):1–39.

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

Dacre, H. F., Hawcroft, M. K., Stringer, M. A., and Hodges, K. I. (2012). An Extratropical Cyclone Atlas: A tool for illustrating cyclone structure and evolution characteristics. *Bulletin of the American Meteorological Society*, 93(10):1497–1502.

Davis, R. A. and Mikosch, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009.

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

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., and Bauer, P. (2011). The ERA-

Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Della-Marta, P. M. and Pinto, J. G. (2009). Statistical uncertainty of changes in winter storms over the North Atlantic and Europe in an ensemble of transient climate simulations. *Geophysical Research Letters*, 36(14).

Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based Geostatistics*. Springer.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45.

Eastoe, E. F. and Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55.

Economou, T., Stephenson, D. B., and Ferro, C. A. T. (2014). Spatio-temporal modelling of extreme storms. *The Annals of Applied Statistics*, 8(4):2223–2246.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65(3):457–483.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231. AAAI Press.

Fawcett, L. and Walshaw, D. (2006). A hierarchical model for extreme wind speeds. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(5):631–646.

Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, 18(2):173–188.

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

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556.

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

Gander, W. (1985). On Halley's iteration method. *American Mathematical Monthly*, 92(2):131–134.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Hall, T. M. and Jewson, S. (2007). Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus A*, 59(4):486–498.

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

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

Hewson, T. D. and Neu, U. (2015). Cyclones, windstorms and the IMILAST project. *Tellus A*, 67(1):27128.

Hills, S. E. and Smith, A. F. (1992). Parameterization issues in Bayesian inference. *Bayesian Statistics*, 4:227–246.

Hodges, K. I. (1995). Feature tracking on the unit sphere. *Monthly Weather Review*, 123(12):3458–3465.

Hosking, J. R. M. and Wallis, J. R. (2005). *Regional Frequency Analysis: an Approach Based on L-moments*. Cambridge University Press.

Hoskins, B. J. and Hodges, K. I. (2002). New perspectives on the Northern Hemisphere winter storm tracks. *Journal of the Atmospheric Sciences*, 59(6):1041–1061.

Howard, T. and Clark, P. (2007). Correction and downscaling of NWP wind speed forecasts. *Meteorological Applications*, 14(2):105–116.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation. *Science*, 269(5224):676–679.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M. (2003). An overview of the North Atlantic oscillation. *Geophysical Monograph-American Geophysical Union*, 134:1–36.

Jalbert, J., Favre, A.-C., Bélisle, C., and Angers, J.-F. (2017). A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society: Series C (Applied Statistics).*

Jiang, J. and Perrie, W. (2008). Climate change effects on North Atlantic cyclones. *Journal of Geophysical Research: Atmospheres*, 113(D9).

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Jonathan, P. and Ewans, K. (2011). Modeling the seasonality of extreme waves in the Gulf of Mexico. *Journal of Offshore Mechanics and Arctic Engineering*, 133(2):021104.

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

Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69(1):19–27.

Kereszturi, M., Tawn, J. A., and Jonathan, P. (2016). Assessing extremal dependence of North Sea storm severity. *Ocean Engineering*, 118:242–259.

Kunkel, K. E., Andsager, K., and Easterling, D. R. (1999). Long-term trends in extreme precipitation events over the conterminous United States and Canada. *Journal of Climate*, 12(8):2515–2527.

Lambert, S. J. and Fyfe, J. C. (2006). Changes in winter cyclone frequencies and strengths simulated in enhanced greenhouse warming experiments: Results from the models participating in the IPCC diagnostic exercise. *Climate Dynamics*, 26(7-8):713–728.

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

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

Lionello, P., Boldrin, U., and Giorgi, F. (2008). Future changes in cyclone climatology over Europe as inferred from a regional climate simulation. *Climate Dynamics*, 30(6):657–671.

Mailier, P. J., Stephenson, D. B., Ferro, C. A. T., and Hodges, K. I. (2006). Serial clustering of extratropical cyclones. *Monthly Weather Review*, 134(8):2224–2240.

Martins, E. S. and Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3):737–744.

McCabe, G. J., Clark, M. P., and Serreze, M. C. (2001). Trends in Northern Hemisphere surface cyclone frequency and intensity. *Journal of Climate*, 14(12):2763–2768.

Moshtagh, N. (2005). Minimum volume enclosing ellipsoid. Technical Report, University of Pennsylvania.

Murray, R. J. and Simmonds, I. (1991). A numerical scheme for tracking cyclone centres from digital data. *Australian Meteorological Magazine*, 39(3):155–166.

Nixon, M. S. and Aguado, A. S. (2012). *Feature Extraction & Image Processing for Computer Vision*. Academic Press.

Northrop, P. J. and Attalides, N. (2016). Posterior propriety in Bayesian extreme value analyses using reference priors. *Statistica Sinica*, 26(2):721–743.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):93–120.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809.

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

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

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Pinto, J. G., Ulbrich, U., Leckebusch, G., Spangehl, T., Reyers, M., and Zacharias, S. (2007). Changes in storm track and cyclone activity in three SRES ensemble experiments with the ECHAM5/MPI-OM1 GCM. *Climate Dynamics*, 29(2-3):195–210.

Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4):1430.

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique*, 154(2):156–177.

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

Robert, C. and Casella, G. (2009). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media.

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

Roberts, J. F., Champion, A. J., Dawkins, L. C., Hodges, K. I., Shaffrey, L. C., Stephenson, D. B., Stringer, M. A., Thornton, H. E., and Youngman, B. D. (2014). The XWS open access catalogue of extreme European windstorms from 1979 to 2012. *Natural Hazards and Earth System Sciences*, 14:2487–2501.

Rudeva, I. and Gulev, S. K. (2011). Composite analysis of North Atlantic extratropical cyclones in NCEP-NCAR reanalysis data. *Monthly Weather Review*, 139(5):1419–1446.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications.* CRC Press.

Rumpf, J., Weindl, H., Höppe, P., Rauch, E., and Schmidt, V. (2007). Stochastic modelling of tropical cyclone tracks. *Mathematical Methods of Operations Research*, 66(3):475–490.

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

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

Schinke, H. (1993). On the occurrence of deep cyclones over Europe and the North Atlantic in the period 1930-1991. *Contributions to Atmospheric Physics*, 66:223–237.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13(2):219–239.

Schultz, D. M., Keyser, D., and Bosart, L. F. (1998). The effect of large-scale flow on low-level frontal structure and evolution in midlatitude cyclones. *Monthly Weather Review*, 126(7):1767–1791.

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

Shadloo, M., Oger, G., and Le Touzé, D. (2016). Smoothed particle hydrodynamics method for fluid flows, towards industrial applications: Motivations, current state, and challenges. *Computers & Fluids*, 136:11–34.

Shapiro, M. A. and Keyser, D. A. (1990). *Fronts, jet streams, and the tropopause.* US Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Wave Propagation Laboratory.

Shaw, T. A., Baldwin, M., Barnes, E. A., Caballero, R., Garfinkel, C. I., Hwang, Y.-T., Li, C., O'Gorman, P. A., Riviere, G., Simpson, I. R., and Voigt, A. (2016). Storm track processes and the opposing influences of climate change. *Nature Geoscience*, 9:656–664.

Sienz, F., Schneidereit, A., Blender, R., Fraedrich, K., and Lunkeit, F. (2010). Extreme value statistics for North Atlantic cyclones. *Tellus A*, 62(4):347–360.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Smith, R. L. (1987a). Discussion of "Parameter orthogonality and approximate conditional inference" by D.R. Cox and N. Reid. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1):21–22.

Smith, R. L. (1987b). A theoretical comparison of the annual maximum and threshold approaches to extreme value analysis. Technical Report, University of Surrey.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–377.

Smith, R. L. (1990a). Max-stable processes and spatial extremes. *Unpublished manuscript.*

Smith, R. L. (1990b). Regional estimation from spatially dependent data. *Unpublished manuscript*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Standen, J., Wilson, C., Vosper, S., and Clark, P. (2017). Prediction of local wind climatology from Met Office models: Virtual Met Mast techniques. *Wind Energy*, 20(3):411–430.

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

Stephenson, A. G. (2016). Bayesian inference for extreme value modelling. *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 257–280.

Tawn, J. A. (1987). Discussion of "Parameter orthogonality and approximate conditional inference" by D.R. Cox and N. Reid. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1):33–34.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *The Annals of Applied Statistics*, 10(4):2303–2324.

Todd, M. J. and Yıldırım, E. A. (2007). On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13):1731–1744.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, 47(1/2):123–138.

Ulbrich, U., Leckebusch, G., and Pinto, J. G. (2009). Extra-tropical cyclones in the present and future climate: a review. *Theoretical and Applied Climatology*, 96(1-2):117–131.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., and Kelly, G. A. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

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

Wadsworth, J. L. and Tawn, J. A. (2012b). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):543–567.

Wadsworth, J. L., Tawn, J. A., and Jonathan, P. (2010). Accounting for choice of measurement scale in extreme value modeling. *The Annals of Applied Statistics*, 4(3):1558–1578.

Wang, Y. and So, M. K. (2016). A Bayesian hierarchical model for spatial extremes with multiple durations. *Computational Statistics & Data Analysis*, 95:39–56.

Weiss, J., Bernardara, P., and Benoit, M. (2014). Formation of homogeneous regions for regional frequency analysis of extreme significant wave heights. *Journal of Geophysical Research: Oceans*, 119(5):2906–2922.

Winter, H. C., Bernardara, P., and Clegg, G. (2017). Modelling daily rainfall extremes using regional frequency analysis. *In preparation.*

Winter, H. C. and Tawn, J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *Extremes*, 20(2):393–415.

Winter, H. C., Tawn, J. A., and Brown, S. J. (2016). Modelling the effect of the El Nino-Southern Oscillation on extreme spatial temperature events over Australia. *The Annals of Applied Statistics*, 10(4):2075–2101.

Wood, S. (2006). *Generalized Additive Models: an Introduction with R.* CRC press.

Youngman, B. D. and Stephenson, D. B. (2016). A geostatistical extreme-value framework for fast simulation of natural hazard events. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 472(2189).

Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G., and Stephenson, D. B. (2013). A multi-model assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *Journal of Climate*, 26(16):5846–5862.