Extreme value modelling for protecting and maintaining critical infrastructure from natural hazards

Eleanor D'Arcy, B.Sc.(Hons.), M.Res

Submitted for the degree of Doctor

of Philosophy at Lancaster University.

May 2024

STOR-i

excellence with impact

Abstract

Safeguarding critical infrastructure is paramount as we face escalating natural hazards. This thesis uses extreme value modelling to enhance our understanding and estimation of extreme sea levels (ESLs) and river flows. We focus on the intrinsic complexities of non-stationarity and dependence inherent in such phenomena and propose novel statistical methodologies to address these challenges.

Coastal and fluvial flooding are some of the most widespread natural disasters today. Estimates of ESLs and river flows can be useful for guiding design criteria of flood defences. We illustrate how falsely assuming stationarity and independence affects estimation. We find that current methods over- or under-estimate extreme events, depending on the location of interest, which, if used for defence design, could lead to a waste of resources or put communities at risk, respectively.

We develop a novel methodology for ESLs that accounts for non-stationarity in skew surges and peak tides, the dependence between these components and their temporal dependence. We discuss how to simulate skew surge time series that replicate their seasonality, temporal dependence and extreme values; these are useful for coastal erosion maintenance planning and predicting surge barrier closure rates. Lastly, we present a novel model for temporal dependence in extreme river flows based on a max-autoregressive moving average process. We derive new extremal dependence features and show how this process can capture the unique features of river flow extremes. Additional methods are discussed relevant to the Extreme Value Analysis conference data

challenges.

Ultimately, this thesis contributes to advancing extreme value modelling techniques tailored specifically to protecting and maintaining critical infrastructure against natural hazards. By enhancing our ability to prepare for extreme events, our findings can inform policy-making and infrastructure planning to foster greater resilience in the face of escalating natural hazards.

Acknowledgements

Firstly, I would like to acknowledge my academic supervisor Jonathan Tawn, whose unwavering support and expert guidance have been instrumental throughout my PhD. Your mentorship has extended far beyond the academic realm, providing me with the skills, confidence and resilience needed in my future career. I am in awe of your dedication to excellence and professional achievements, I feel privileged to have had the opportunity to learn from you and be guided by your expertise. I could not have asked for a better supervisor and mentor, thank you.

I was lucky enough to complete my PhD at the EPSRC funded STOR-i Cente for Doctoral Training, and I would like to thank the STOR-i staff and students. Namely, I would like to thank Rachel McCrea for mentoring me and Daniel Grose for your support in building my R package. My PhD experience would not have been as enjoyable without the extremes research group; I am profoundly grateful for the supportive and fun research environment we created together. I would also like to acknowledge my closest friends in the office: Carla, Lídia, Matt and Holly. Your kindness, humour and shared experiences have enriched my years at STOR-i.

I would also like to acknowledge EDF Group for sponsoring my PhD. I was reassured knowing that my research had a clear motivation from the very beginning. Namely, I would like to thank my industrial supervisor Dafni Sifnioti (EDF Energy R&D UK) whose guidance and knowledge were crucial in preparing this thesis. You have taught me invaluable networking skills and encouraged me to get my name out there, leading

to many exciting opportunities. Additionally, I would like to thank Hugo Winter and Amélie Joly for your supervision in the early stages.

I have been fortunate to collaborate with some incredible researchers that I would also like to acknowledge. I am particularly grateful to Ivan Haigh (University of Southampton) and Jenny Sansom (Environment Agency) who have helped shape the impact of my research. I hope that we can continue to collaborate in the future.

During my PhD I sought professional development opportunities, so there are several organisations I would like to thank for their support. Firstly, the Parliamentary and Scientific Committee for recognising my work in the STEM for Britain final, this encouraged me to get my research into policy making. I would also like to thank the Smith Institute who selected me as the winner of the TakeAIM competition; this was a huge confidence boost and opened the door to further opportunities. Lastly, I would like to thank the Royal Statistical Society for appointing me as a William Guy lecturer, this has given me an invaluable opportunity to inspire the next generation of mathematicians. All of these endeavours have shaped me into the researcher I am today.

Last but certainly not least, I would like to thank my family and friends. To Mum and Dad, for always believing in me, especially when I couldn't believe in myself. I have admired your dedication and work ethic my entire life; I feel very fortunate that these skills were passed down to me! Your belief in my potential and your commitment to my success have been a constant source of motivation. I am also deeply grateful for the foundational values instilled in me by the strong women who surrounded me growing up: my Mum, Nana and Aunt. Your strength in the face of adversity inspires me daily. I am very lucky to have the best of friends in Emma, Heather and Mollie. Thank you for keeping me motivated (and sane) through the tougher times. And finally, to Jack, I am beyond thankful you moved to Lancaster to support me in my final year. Your willingness to listen, offer perspective and provide a shoulder to lean on has been a constant source of comfort and reassurance in these final stages.

Declaration

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

Chapter 3 has been published as D'Arcy, E., Tawn, J. A., Joly, A., and Sifnioti, D. E. (2023). Accounting for seasonality in extreme sea level estimation. The Annals of Applied Statistics, 17(4), 3500-3525. https://doi.org/10.1214/23-AOAS1773

Chapter 4 has been published as D'Arcy, E., Tawn, J. A., and Sifnioti, D. E. (2022). Accounting for climate change in extreme sea level estimation. Water, 14(19), 2956. https://doi.org/10.3390/w14192956

Chapter 5 is linked to work with Prof. Ivan Haigh (University of Southampton) which is not included in the thesis but will be submitted for publication as Haigh, I. D., D'Arcy, E., Brand, J., Inayatillah, A., Trace-Kleeberg, S., Walraven, M., and Batchelor, A. (2024). Rapid acceleration in the number of closures of storm surge barriers in the future.

Chapter 6 has been submitted for publication as D'Arcy, E. and Tawn, J. A. (2024). Extremal properties of max-autoregressive moving average processes for modelling extreme river flows.

Chapter 7 is a result of the STOR-i contribution for a competition as part of the 2021 Extreme Value Analysis conference at the University of Edinburgh. This has been published as D'Arcy, E., Murphy-Barltrop, C. J. R., Shooter, R. and Simpson, E. S. (2023). A marginal modelling approach for predicting wildfire extremes across

the contiguous United States. Extremes, 26(2), 381–398. https://doi.org/10.1007/s10687-023-00469-7. All authors contributed equally to this work.

Chapter 8 is a result of the Lancaster and Maynooth University contribution for a competition as part of the 2023 Extreme Value Analysis conference at the University of Bocconi, Italy. This has been submitted for publication as André, L. M., Campbell, R., D'Arcy, E., Farrell, A., Healy, D., Kakampakou, L., Murphy, C., Murphy-Barltrop, C. J. R. and Speers, M. (2024). Extreme value methods for estimating rare events in Utopia. My primary contributions are in Sections 8.1-8.3.

The word count for this thesis is approximately 65,163.

Eleanor D'Arcy

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In the face of escalating environmental challenges resulting from anthropogenic global warming, it has become more crucial than ever to accurately model, forecast and predict extreme environmental events. These events range from severe storms and wildfires to droughts and rising sea levels. At the time of writing, Storm Isha brings 99mph winds and heavy rain to the UK disrupting road, rail and air travel, leaving 56,000 homes without power and two fatalities. This follows just one week after record-breaking low temperatures of $-15°$C were recorded in Scotland. On the other hand (and the other side of the pond), California has faced increasingly severe wildfires in the past decade due to prolonged droughts and changing climate patterns, leading to the loss of lives and homes. As the global population increases and more infrastructure is built, the impacts felt by extreme environmental events become more significant (Seneviratne et al., 2012, 2021).

We are particularly interested in rare and potentially destructive phenomena called natural hazards; examples include floods, droughts, earthquakes and wildfires. Climate change is impacting the frequency and intensity of natural hazards (Seneviratne

et al., 2021). These events can have significant and wide-ranging impacts on society, depending on the type and magnitude of the hazard itself, but also the vulnerability of the population and the level of preparedness/resilience in place. Determining the likelihood of such events enhances our disaster preparedness and helps to implement effective policies so that the impacts on society are minimised.

In this thesis, we focus on flooding as one of the most widespread natural disasters, affecting millions of people globally each year. Specifically, we are interested in coastal and fluvial flooding. Coastal flooding poses significant risks to human communities and the natural environment, especially in the UK due to its extensive coastline that is exposed to various weather patterns. Several factors contribute to the risks associated with coastal flooding, including mean sea level rise, storm surges, high tides and large waves. Similarly, fluvial (or riverine) flooding poses an increasing risk to communities as we observe heavier rainfall events and urbanisation in many catchment areas. Consequences of flooding include infrastructure damage, loss of life and injury, displacement of people and increased erosion. Accelerating rates of erosion pose significant risks to the environment (e.g., habitat destruction) and to society (e.g., undermining of infrastructure such as roads, buildings and utility lines). As sea levels continue to rise, the importance of proactive measures to reduce vulnerability and enhance resilience becomes increasingly critical for natural hazards.

Flooding poses unique risks to the nuclear industry, for example, impacts to the cooler system (leading to potential overheating of the reactor core), loss of electrical power (affecting the ability to control and operate safety systems) and ingress of water into critical areas (e.g., the reactor buildings and auxiliary structures) (Office for Nuclear Regulation, 2021). This was demonstrated by the Fukushima nuclear accident in 2011, where, following an earthquake, a tsunami resulted in electrical grid failure and damaged the power plant's backup energy sources so that the reactors could not be cooled; this resulted in the release of radioactive contaminants into the surrounding

areas (International Atomic Energy Agency, 2015). EDF Energy (the industrial partner of this PhD project, funded by EDF Group) manages eight nuclear power stations across the UK, five of which are generating zero-carbon electricity and three that are in the defueling phase (the first stage of decommissioning). EDF Energy's nuclear fleet is located in the UK coastal zone so is susceptible to flooding and erosion.

For coastal infrastructure, the inference of such extreme events helps guide design criteria for flood defences. Flood defences, such as a sea wall or flood barrier (located at an estuary) play a critical role in safeguarding lives and protecting infrastructure. Design criteria require that defences are built to withstand rare events but with limited resources, these cannot be built too conservatively. For example, the Office for Nuclear Regulation (the UK's independent nuclear regulator for safety and security) set the design basis criterion for natural hazards at a frequency of exceedance of 10^{-4} per year (Office for Nuclear Regulation, 2014). Similarly, flood defences at other locations, such as the Thames Barrier, are often built to withstand events that are likely to occur within their projected T-year lifespan (Environment Agency, 2021). With ever-changing environmental conditions, flood defences must also be maintained; as the frequency of freak weather events increases as a result of climate change, the life expectancy of defences may decrease. It is also fundamental to understand the intensity and duration of extreme sea levels and river flows for erosion maintenance planning.

Statistical modelling of natural hazards represents a challenge since, by definition, rare events typically do not occur within the observational range so records are sparse. Extreme value theory is a powerful tool that can be used to estimate events that are expected to occur, or be exceeded, at least once in a T-year period. Therefore, these estimates are fundamental for guiding the design criteria of flood defences. Standard statistical methods are likely to perform poorly as the T-year period of interest is typically longer than the historical record of data and we may be interested in events that are more extreme than any previously observed. The underpinning asymptotic, or limit,

arguments for extreme value theory facilitate a framework which allows extrapolation beyond the maxima of data to estimate such events.

Extreme value theory has been effectively used to guide policies surrounding coastal and fluvial hazards, including for guidance surrounding nuclear safety from natural hazards (Office for Nuclear Regulation, 2021). However, these existing methods make several unjustified and restrictive assumptions about the data that we challenge in this thesis. We account for non-stationarity and temporal dependence in both sea levels and river flow, although these considerations extend beyond the applications discussed here. For our extreme sea level analysis, we decompose the sea levels into surge and tide. We account for within- and across-year non-stationarity in both components owing to seasonality and anthropogenic climate change. We also model the dependence between the two variables and their temporal dependence, which results from extreme weather events spanning multiple days. As river flows exhibit a unique temporal dependence structure, we develop the theory for modelling this based on an extension of time series models by deriving its extremal properties. By understanding the temporal dependence of both variables, we learn about prolonged periods in extreme states that can accelerate erosion rates.

Our results demonstrate an improvement on those used in practice. Our model for extreme sea levels is being implemented by the Environment Agency (the Government organisation responsible for environmental protection in the UK) for their updated coastal flood boundary report (Environment Agency, 2018); this periodically assesses and updates flood risk information, and forms the basis for design decisions on future coastal flood defence projects. In combination with our extreme river flow analysis, our work is also of interest to the Thames Barrier in determining the life expectancy of the barrier as we face unprecedented environmental challenges.

1.2 Overview of thesis

This thesis aims to develop improved methods for estimating extreme sea levels and river flows, that can be used for policy making and protecting livelihoods/infrastructure. The structure of the thesis is as follows:

Chapter 2 provides an overview of the existing methods for modelling univariate extreme values. We begin by deriving the extreme value models under strict assumptions about the data, that are not tenable in reality. Then we detail how to relax these assumptions so that environmental data can be reliably modelled and list the existing approaches for doing so.

Chapter 3 details our novel methodology for estimating extreme sea levels that is the first to capture seasonality, interannual variations and longer term changes. We use a joint probabilities method, with skew surge and peak tide as two sea level components. The tidal regime is predictable but skew surges are stochastic. We present a statistical model for skew surges, where the main body of the distribution is modelled empirically whilst a non-stationary generalised Pareto distribution (GPD) is used for the upper tail. We capture within-year seasonality by introducing a daily covariate to the GPD model and allowing the distribution of peak tides to change over months and years. Skew surge-peak tide dependence is accounted for via a tidal covariate in the GPD model and we adjust for skew surge temporal dependence through the subasymptotic extremal index. We incorporate spatial prior information in our GPD model to reduce the uncertainty associated with the highest return level estimates. Our results are an improvement on current return level estimates, with previous methods typically underestimating. We illustrate our method at four UK tide gauges. This chapter is published as D'Arcy et al. (2023b). We have also developed an accompanying R package to implement our methodology, this is presented in Appendix B.

Chapter 4 extends the methodology of Chapter 3 to account for the effects of climate change on extreme sea level estimates. As the global climate changes, rising sea levels

combined with increases in storm intensity and frequency pose an increasing risk to coastline communities. We present an updated method for estimating extreme sea levels that accounts for the effects of climate change on extreme events that are not accounted for by mean sea level trends. We model extreme skew surges using a non-stationary GPD with covariates accounting for climate change, seasonality and skew surge–peak tide interaction. We develop methods to efficiently test for extreme skew surge trends across different coastlines and seasons. We illustrate our methods using data from four UK tide gauges and estimate sea level return levels when accounting for these long term trends. This chapter is published as D'Arcy et al. (2022).

Chapter 5 presents a procedure for simulating skew surges that focuses on capturing their temporal dependence structure, as well as their seasonal behaviour and extreme values using the ideas from Chapter 3. We use a copula framework to model the pairwise dependence of values separated by different lags, assuming the data follows a Markov process. Since the strength of temporal dependence varies with the time of year, we allow the copula parameters to vary with time using harmonics. Understanding the temporal dependence of skew surges is fundamental to coastal maintenance planning; severe storms accelerate rates of coastal erosion which is further exacerbated when storm events last multiple days with prolonged levels of high skew surge. Additionally, simulated skew surge time series are useful for surge barrier management, such as the Thames Barrier, UK. These simulations can be used for estimating future closure rates and for planning barrier maintenance, which is restricted due to time constraints. We provide simulation results at four UK tide gauges, including Sheerness which is located at the Thames Estuary. Here, we compare our results with simulations that do not account for temporal dependence and demonstrate the implications this has on estimating surge barrier closure rates.

Chapter 6 derives the extremal properties of max-autoregressive moving average (Max-ARMA) models which are powerful tools for modelling time series data with

heavy-tailed behaviour. These are a non-linear version of the popular auto-regressive models. River flow data typically have features of heavy tails and non-linearity, as large precipitation events cause sudden spikes in the data that then exponentially decay. Therefore, stationary Max-ARMA models are a suitable candidate for capturing the unique temporal dependence structure exhibited by river flows. This paper contributes to advancing our understanding of extremal properties of stationary Max-ARMA processes. We detail the first approach for deriving the extremal index, the lagged asymptotic dependence coefficient, and an efficient simulation for a general Max-ARMA process. We use the extremal properties, coupled with the belief that Max-ARMA processes provide only an approximation to extreme river flow, to fit a Max-ARMA model which broadly captures river flow behaviour over a high threshold. We make our inference under a reparametrisation which gives a simpler parameter space that excludes cases where any parameter is non-identifiable. We illustrate results for river flow data from the UK River Thames.

Chapter 7 details a methodology proposed for the Extreme Value Analysis 2021 conference data challenge. We aim to predict the number and size of wildfires over the contiguous United States between 1993 and 2015, with more importance placed on extreme events. In the data set provided, over 14% of both wildfire count and burnt area observations are missing; the objective of the data challenge was to estimate a range of marginal probabilities from the distribution functions of these missing observations. To enable this prediction, we assume that the marginal distribution of a missing observation can be informed using non-missing data from neighbouring locations. In our method, we select spatial neighbourhoods for each missing observation and fit marginal models to non-missing observations in these regions. For the wildfire counts, we assume the compiled data sets follow a zero-inflated negative binomial distribution, while for burnt area values, we model the bulk and tail of each compiled data set using non-parametric and parametric techniques, respectively. Cross validation is used to select

tuning parameters, and the resulting predictions are shown to significantly outperform the benchmark method proposed in the challenge outline. We conclude with a discussion of our modelling framework and evaluate ways in which it could be extended. This chapter is published as D'Arcy et al. (2023a)

Chapter 8 introduces a variety of methods to capture the extremal behaviour and model tail behaviour of complex environmental phenomena in practice, which were used by the Lancopula Utopiversity team to tackle the data challenge of the Extreme Value Analysis 2023 conference. This data challenge was split into four sections, labelled C1-C4. Challenges C1 and C2 comprise univariate problems, where the goal is to estimate extreme quantiles for a non-stationary time series exhibiting several complex features. We propose a flexible modelling technique, based on generalised additive models, with diagnostics indicating generally good performance for the observed data. Challenges C3 and C4 concern multivariate problems where the focus is on estimating joint extremal probabilities. For challenge C3, we propose an extension of available models in the multivariate literature and use this framework to estimate extreme probabilities in the presence of non-stationary dependence. Finally, for challenge C4, which concerns a 50-dimensional random vector, we employ a clustering technique to achieve dimension reduction and use a conditional modelling approach to estimate extremal probabilities across independent groups of variables.

Chapter 9 concludes by summarising the contributions of this thesis and discussing potential avenues for further work.

Chapter 2

Literature review

2.1 Introduction

This chapter details extreme value methods that are relevant to all aspects of the thesis. Extreme value theory is a rapidly emerging area of statistics, where interest lies in modelling the tails of a distribution where events are rare. Applications exist across many sectors including environmental science, energy, finance and sports. Standard statistical techniques that rely on large samples are unsuitable since data are scarce in the tail. This motivates the necessity for extreme value models. Extreme value theory provides a statistically rigorous framework for modelling extreme events, using asymptotically justified theory, that can extrapolate beyond the range of observed data.

We begin by describing two approaches for modelling univariate extreme values in Section 2.2. These approaches were initially derived under the assumption of independent and identically distributed (IID) random variables, which is typically not tenable in reality. In Section 2.3, we present existing methods used to extend these models for sequences that are dependent but stationary. Lastly, in Section 2.4, we review approaches for dealing with non-stationarity in extreme value models.

The ideas developed in this thesis lie within the univariate setting, where we are

interested in observations of a single variable at a specific location. We acknowledge there is ongoing research in both multivariate and spatial extreme value analysis; many of our approaches could be extended to these situations. We refer the reader to Coles (2001), Heffernan and Tawn (2004) and Huser and Wadsworth (2022), although this list is certainly not exhaustive of the vast body of available literature.

2.2 Extreme value models

In this section, we discuss the two most widely used approaches for modelling extreme values in the univariate setting: the generalised extreme value distribution for modelling block maxima and the generalised Pareto distribution for modelling threshold exceedances. We detail these in Sections 2.2.1 and 2.2.2, respectively. Coles (2001) gives a thorough overview of these models. We restrict our attention to the upper tail, although both methods can be easily adapted for the lower tail.

2.2.1 Generalised extreme value distribution

Let $X_1, \ldots, X_n$ be a sequence of $n \in \mathbb{N}$ IID random variables with common distribution function F. Here, we are concerned with the behaviour of the random variable $M_n = \max\{X_1, \ldots, X_n\}$, denoting the maximum of the sequence. Understanding the distribution of M_n provides insight into the extremal behaviour of the process. Theory for the minima can be obtained analogously since,

$$\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}.$$

The distribution of M_n can be derived exactly, for all $x \in \mathbb{R}$, as

$$\Pr(M_n \leq x) = \Pr(X_1 \leq x, \ldots, X_n \leq x) = \Pr(X_1 \leq x) \cdots \Pr(X_n \leq x) = \big[F(x)\big]^n.$$

However, this is not useful in practice for two reasons. Firstly, F is typically unknown. Secondly for any $x < x^F$, where x^F is the upper endpoint of F, $F^n(x) \to 0$ as $n \to \infty$ so the asymptotic distribution of M_n is degenerate with a point mass at x^F.

Instead, we use the following result of Leadbetter et al. (1983). The Extremal Types Theorem says that if there exist sequences of constants $a_n > 0$ and b_n such that for all $x \in \mathbb{R}$,

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \left[F(a_n x + b_n)\right]^n \to G(x) \quad \text{as } n \to \infty, \tag{2.2.1}$$

where G is a non-degenerate distribution function, then G is a member of one of the following families:

$$\text{Gumbel:} \quad G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\} \quad \text{for } -\infty < x < \infty,$$

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

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

with parameters $a > 0$, $b \in \mathbb{R}$ and $\alpha > 0$. A parameterisation to unify these three distributions is given by

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{2.2.2}$$

where $y_+ = \max\{y, 0\}$ and $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$. This is known as the generalised extreme value distribution (GEV) with parameters μ, σ and ξ representing the location, scale and shape, respectively. For $\xi > 0$, this corresponds to the Fréchet distribution

with a heavy upper tail, $\xi < 0$ the reversed Weibull with a finite upper limit and $\xi = 0$ the Gumbel with an exponential tail; $\xi = 0$ should be interpreted as the limit as $\xi \to 0$. Leadbetter et al. (1983) details the proof of this theorem. In practice, we assume that for large n and all $x \in \mathbb{R}$,

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}.$$

Estimating the unknown normalising constants a_n and b_n is not required because they are absorbed into the location and scale parameters of the GEV, since if the above equality holds, then for all $x \in \mathbb{R}$,

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

so in practice we model

$$\Pr(M_n \leq x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}.$$

Parameters of the GEV are commonly estimated using likelihood inference (Coles, 2001), and this is the approach we take throughout the thesis. Other techniques, such as Bayesian methods (Coles and Tawn, 1996) or probability-weighted moments estimates (Hosking et al., 1985), can be adopted.

The GEV is the only class of distributions to satisfy the max-stability property. That is, a distribution G is max-stable if, for any $m \in \mathbb{N}$, there are constants $\alpha_m > 0$ and β_m such that for all $x \in \mathbb{R}$,

$$G^m(\alpha_m x + \beta_m) = G(x). \tag{2.2.3}$$

That is, raising a max-stable distribution to the power of $m \in \mathbb{N}$ results in a distribution from the same family, apart from a change of location and scale. So taking the maximum

of IID random variables with marginal distribution G gives a random variable with the same distributional form i.e., the GEV with the same shape parameter (but different location and scale parameters), following from equality (2.2.3). This motivates the block maxima approach, where data of length k can be split into m blocks of equal length n, where $k = mn$, and the m block maxima are treated as realisations from the GEV. This max-stability property is analogous to the sum-stability property linked to the well-studied central limit theorem, that motivates a Gaussian model for the distribution of finite sample means.

The choice of block size n is critical as it induces a bias-variance trade-off. Small blocks lead to a poor approximation by the limiting Extremal Types Theorem resulting in bias, whilst large blocks mean there are fewer maxima as $m = k/n$, thus a greater estimation variance. With IID data, this is an easier choice than in practice, as care needs to be taken to ensure the identically distributed assumption. For example, this could be invalidated by choosing seasonal blocks for environmental data. Letting n be the number of observations in a year is common in environmental extreme value analysis since this avoids these issues of non-stationary block maxima. We discuss these issues further in Section 2.4.

Extreme quantiles x_p, where $G(x_p) = 1 - p$ for $p \in (0, 1)$ with p near zero, can be estimated by inverting the form of the GEV in equation (2.2.2), to give

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

and substituting in the parameter estimates. For G the annual maxima distribution, the value x_p is termed the return level with return period $1/p$ years; this is the level we expect the annual maximum to exceed every $1/p$ years, on average. Equivalently, this corresponds to a value exceeded once per year, with probability p. Return levels are a common metric for assessing risk, especially in the environmental sector, where

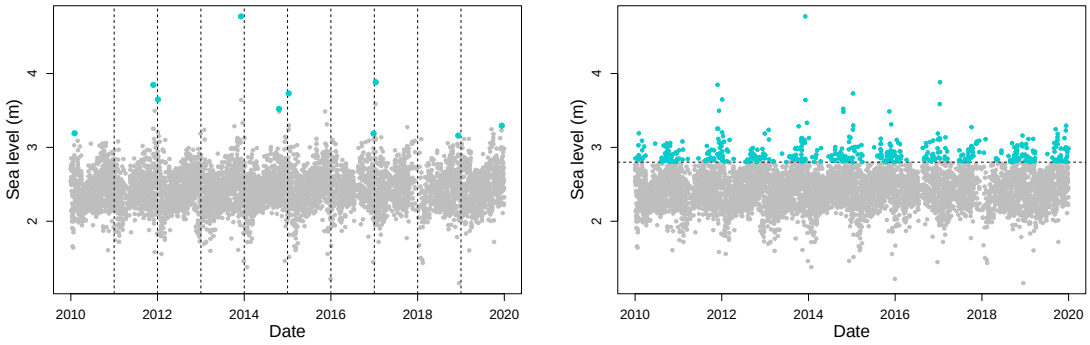


Figure 2.2.1: An illustration of the block maxima (left; with blocks corresponding to years and indicated by vertical dashed lines) and threshold exceedance (right; with a threshold of 2.8m, indicated by a horizontal dashed line) approach for defining extreme values. Both approaches are applied to approximate twice daily sea level observations from 2010-2020 at the Lowestoft tide gauge (BODC, 2020) shown in grey with extreme observations shown in blue.

small values of p are typically of interest to estimate rare events. For example, in the nuclear industry, regulators require accurate return level estimates for $p = 10^{-4}$ to ensure that power stations are robust to extremely rare natural hazards (Office for Nuclear Regulation, 2014). Accurate estimation of return levels is crucial for protection against natural hazards, such as those discussed in this thesis.

2.2.2 Generalised Pareto distribution

A key limitation of the block maxima approach is that it is often wasteful of data; observations may be recorded hourly, yet only the annual maximum is modelled. An alternative, more widely used method, is to define extreme values as exceedances of a high threshold. Figure 2.2.1 demonstrates how the peaks over threshold approach typically considers more data than the block maxima approach. Within this framework, all large events are included in the model (rather than one per year) and unusually small annual maxima are excluded from the analysis.

For an arbitrary term X in an IID sequence of random variables with distribution function F and some high threshold u, the stochastic behaviour of an extreme event is

defined by,

$$\Pr(X > u + y \mid X > u) = \frac{1 - F(u+y)}{1 - F(u)} \quad \text{for } y > 0.$$

If the block maxima have limiting distribution G (as in distribution (2.2.2)), then the threshold exceedances have a corresponding limiting distribution known as the generalised Pareto distribution (GPD). For large n and a level u, the distribution of the threshold exceedance $X - u_n(u)$ of the threshold $u_n(u)$, given that $X > u_n(u)$ where $u_n(u) = b_n + a_n u$, is asymptotically,

$$\Pr\big(X > u_n(x) \mid X > u_n(u)\big) \to \bar{H}_u(x) \quad \text{where} \quad \bar{H}_u(x) = \left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi},$$

for $x > u$ as $n \to \infty$, with a_n, b_n as in limit (2.2.1) and $(\sigma_u, \xi) \in \mathbb{R}_+ \times \mathbb{R}$ the scale and shape parameters, respectively. The shape parameter is the same as that for the GEV, whereas the scale parameter is threshold dependent since $\sigma_u = \sigma + \xi(u - \mu)$ for μ and σ the GEV parameters.

Assuming that the above limit holds in practice for a high threshold $u_n(u) = u$ and $u_n(x) = x$, then exceedances of this threshold have limiting GPD tail model,

$$\Pr(X > x) = \lambda_u \left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}, \tag{2.2.4}$$

for $x > u$ where $\lambda_u = \Pr(X > u)$. For $\xi > 0$, this corresponds to $X - u | X > u$ being Pareto distributed with heavy tails, for $\xi < 0$ the GPD is light-tailed with a finite upper endpoint and for $\xi = 0$ it corresponds to $X - u | X > u$ being exponentially distributed. Smith (1989) details the relationship between the GEV and GPD, and Coles (2001) provides a formal justification for this model in describing threshold exceedances.

For inference, an appropriate threshold must be selected. This choice is analogous to choosing a block size as in Section 2.2.1. A low threshold is likely to violate the asymptotic basis of the model which will invoke bias, whilst a high threshold results

in an increased variance of estimates due to a paucity of exceedances available for inference. Diagnostics such as mean residual life plots and parameter stability plots are common approaches for threshold selection (Coles, 2001). Parameter stability plots are motivated by the threshold stability property: Davison and Smith (1990) state that if $X - u|X > u \sim \text{GPD}(\sigma_u, \xi)$, then for any $v > u$ such that $\Pr(X > v) > 0$, $X - v|X > v \sim \text{GPD}(\sigma_u + \xi(v-u), \xi)$. Scarrott and MacDonald (2012) give an overview of various methods for threshold selection, although more recent methods have been proposed by Wadsworth (2016), Northrop et al. (2017) and Murphy et al. (2023), with these diagnostics exploiting the threshold stability property of the GPD.

As for the GEV, likelihood inference is typically used for parameter estimation, although other approaches can be used. GPD parameters are estimated by maximising the following log-likelihood function, where n_u denotes the number of exceedances of the threshold u, where exceedances of u are denoted by $y_1, \ldots, y_{n_u}$,

$$\ell(\sigma_u, \xi) = \begin{cases} -n_u \log \sigma_u - (1 + 1/\xi) \sum_{i=1}^{n_u} \log(1 + \xi y_i/\sigma_u) & \text{if } \xi \neq 0 \\ -n_u \log \sigma_u - (1/\sigma_u) \sum_{i=1}^{n_u} y_i & \text{if } \xi = 0. \end{cases} \tag{2.2.5}$$

Return levels are obtained similarly to the block maxima approach, by inverting the GPD tail model given by expression (2.2.4). This gives us the value exceeded once every m observations, so alterations need to be made to obtain return levels corresponding to maxima over specific time periods, as opposed to counts of observations. Coles (2001) details this procedure.

2.3 Extremes of dependent sequences

So far, the extreme value models have been derived assuming that the underlying process is IID. However, in practice these assumptions are unrealistic. In this section, we focus on relaxing the temporal independence assumption and instead focus on a

stationary sequence of random variables $\{X_t;\ t \in \mathbb{Z}\}$.

Stationarity is a more realistic assumption than IID for many environmental processes. This corresponds to a series whose variables may be mutually dependent but whose stochastic properties are homogeneous through time (Coles, 2001), i.e., they have the same marginal distribution across time. Formally, a time series of random variables $\{X_t;\ t \in \mathbb{Z}\}$ is stationary if the joint probability density function f of any set of values in the series is the same as if they were all shifted in time by the same lag $\tau \in \mathbb{Z}$, so that $f_{X_{j_1},\ldots,X_{j_n}}(x_{j_1},\ldots,x_{j_n}) = f_{X_{j_1+\tau},\ldots,X_{j_n+\tau}}(x_{j_1},\ldots,x_{j_n})$ for all $n \in \mathbb{N}$ and $j_1,\ldots,j_n \in \mathbb{N}$, where $j_1 \leq \ldots \leq j_n$ and all $(x_{j_1},\ldots,x_{j_n}) \in \mathbb{R}^n$ (Chatfield, 2013).

Dependence takes many forms, so we cannot generalise extremal behaviour without imposing some constraints. Often, a condition is assumed that limits the extent of long range dependence at extreme levels; this allows us to focus on the effects of short range dependence (Leadbetter et al., 1983). A stationary series $\{X_t;\ t \in \mathbb{Z}\}$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$, where $i_1,\ldots,i_p,j_1,\ldots,j_q \in \mathbb{Z}$ with $j_1 - i_p > l$,

$$\begin{aligned}\big|\Pr(X_{i_1} \leq u_n,\ldots,X_{i_p} \leq u_n, X_{j_1} \leq u_n,\ldots,X_{j_q} \leq u_n) - \\ \Pr(X_{i_1} \leq u_n,\ldots,X_{i_p} \leq u_n)\Pr(X_{j_1} \leq u_n,\ldots,X_{j_q} \leq u_n)\big| \leq \alpha(n,l),\end{aligned}$$

where $\alpha(n,l_n) \to 0$ for some sequence l_n, such that $l_n/n \to 0$ as $n \to \infty$. This condition ensures extreme events are near independent if they are sufficiently distant from each other in time. Provided this condition is satisfied for an appropriate sequence u_n i.e., $u_n = a_n z + b_n$ where a_n, b_n are as in limit (2.2.1), the limiting extreme value models of Section 2.2 remain appropriate. Leadbetter et al. (1983) show that this condition holds for all univariate, stationary Gaussian processes with a correlation function that decays to zero geometrically, so for all stationary autoregressive moving-average, ARMA(p,q), models.

Under the $D(u_n)$ assumption, the block maxima approach is still expected to be asymptotically valid since temporal dependence in block maxima vanishes. Therefore, as the block size tends to infinity, the block maxima can be considered independent of each other. We discuss this further in Section 2.3.1. Similarly, the GPD remains appropriate for modelling the marginal distribution of threshold exceedances of a stationary process $\{X_t;\ t \in \mathbb{Z} \text{ with } X_t > u\}$ consisting of dependent realisations where, for example, X_1 and X_2 could both be exceedances and be dependent. However, the likelihood inference procedure with log-likelihood function (2.2.5) would not be appropriate for modelling joint exceedances due to this dependence. We discuss methods for dealing with dependent exceedances in Section 2.3.2.

Temporal dependence is a common theme throughout this thesis, as in the wider extreme value theory context. Diagnostics for temporal dependence, including the coefficient of asymptotic dependence and the extremogram, are detailed in Section 2.3.3. We also discuss how to use copulas for modelling temporal dependence of Markov processes in Section 2.3.4.

2.3.1 Block maxima and stationarity

If for $u_n = a_n z + b_n$, with a_n and b_n defined by limit (2.2.1), the $D(u_n)$ condition holds then the maxima of a stationary series will converge to a GEV distribution. Consider a stationary process $\{X_t;\ t \in \mathbb{Z}\}$ with $M_n = \max\{X_1, \ldots, X_n\}$ and a sequence of IID variables $\{X_t^*; t \in \mathbb{Z}\}$ with $M_n^* = \max\{X_1^*, \ldots, X_n^*\}$ with the marginal distributions of X_i and X_i^* equal for all $i \in \mathbb{Z}$. As $n \to \infty$,

$$\Pr\left(\frac{M_n^* - b_n}{a_n} \leq x\right) \to G_1(x) \quad \text{if and only if} \quad \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G_2(x),$$

where G_1 is a non-degenerate distribution function and $G_2(x) = G_1^\theta(x)$, for a constant $0 < \theta \leq 1$. Specifically, if the distribution of the IID maxima is GEV with parameters

(μ, σ, ξ), then the distribution of the stationary maxima is also GEV with parameters (μ^*, σ^*, ξ) where $\mu^* = \mu - \sigma(1 - \theta^{-\xi})/\xi$ and $\sigma^* = \sigma\theta^{\xi}$.

The quantity θ is termed the extremal index; this captures the effect of the temporal dependence on the distribution of the maxima. For an independent series, $\theta = 1$ however, the converse is not true; there are many stationary series with $\theta = 1$ that are not independent, for example, all Gaussian processes satisfying the $D(u_n)$ condition. Let $N_{p_n}(u_n)$ denote the number of observations of $\{X_t, \ldots, X_{t+p_n}\}$ to exceed the threshold u_n, where $p_n = o(n)$. If $N_{p_n}(u_n) \geq 1$, clusters are defined as the set $\{X_{t+i} > u_n : i = 1, \ldots, p_n\}$. Then $N_{p_n}(u_n)$ denotes the cluster size, which has limiting probability mass function π defined by

$$\pi(m) = \lim_{n\to\infty} \Pr(N_{p_n}(u_n) = m | N_{p_n}(u_n) \geq 1),$$

for $m \in \mathbb{N}$ (Hsing et al., 1988). Leadbetter et al. (1983) characterise the extremal index as the reciprocal of the limiting mean cluster size, where limiting refers to increasing the threshold that defines exceedances to the upper endpoint. So $\theta^{-1} = \sum_{m=1}^{\infty} m\pi(m)$. Whereas O'Brien (1987) defines it in terms of the number of down-crossings at the end of a cluster of threshold exceedances, i.e.,

$$\theta = \lim_{n\to\infty} \Pr(X_2 \leq u_n, \ldots, X_{p_n} \leq u_n | X_1 > u_n). \tag{2.3.1}$$

It follows that, provided we can reasonably assume long range independence (large enough block sizes will impose this), using the GEV family to model block maxima remains appropriate. Therefore, when modelling block maxima data, dependence can be ignored and we proceed as in Section 2.2.1 but with different parameters to those of the corresponding independent series. However, as dependence levels increase, the accuracy of the GEV as a limiting approximation is likely to diminish so care is needed when modelling such data.

2.3.2 Peaks over threshold and stationarity

The GPD remains appropriate for modelling the marginal distribution of exceedances of a sequence containing dependent realisations. However, dependence in the series invalidates the likelihood inference for jointly modelling exceedances, since this is based on an IID assumption (see expression (2.2.5)). Alternative inference procedures can be adopted, including the quasi-maximum likelihood estimators (Lee and Hansen, 1994).

A common approach for dealing with dependent threshold exceedances is declustering. This involves filtering dependent observations to yield a set of approximately independent threshold exceedances. Under a long range dependence assumption, essentially very similar to the $D(u_n)$ condition, separate clusters of exceedances may be considered independent in the limit. Then values from different clusters are taken to be independent as the threshold gets sufficiently large. However, values within a cluster still exhibit dependence even at high thresholds. A common approach is to identify these independent clusters above a high threshold and evaluate the characteristic of interest for each cluster, then use these for inference as taking only one value from each cluster (e.g., the maximum) gives an IID sample (Eastoe and Tawn, 2012).

There are different methods used to identify clusters. Coles (2001) discusses the runs approach of Smith and Weissman (1994), this assumes exceedances belong to the same cluster if they are separated by fewer than a specified number of values consecutively below the threshold. Formally, let W_i be 1 if $X_i > u$ and 0 otherwise, and let $N_n(u)$ be the number of exceedances of a high threshold u in the sequence $\{X_t, \ldots, X_{t+n}\}$ for $t \in \mathbb{Z}$ and $n \in \mathbb{N}$. Then $Z_n \coloneqq \sum_{i=1}^{n} \left[W_i \prod_{j=1} r(1 - W_{i+j})\right]$ is the total number of times the final exceedance of a cluster is followed by $r \in \mathbb{N}$ non-exceedances. Then the runs estimate of the extremal index is given by

$$\hat{\theta} = Z_n / N_n(u). \tag{2.3.2}$$

The value r is termed the run length and represents the number of values below the threshold between clusters, this must be specified before estimation and the choice of r is complex. This decision is the main pitfall with the runs estimate of θ or way of identifying clusters, as discussed by Smith and Weissman (1994). The runs estimate is strongly related to definition (2.3.1) of O'Brien (1987) where clusters are defined as those separated by p_n non-exceedances, so that p_n corresponds to the run length.

Ferro and Segers (2003) propose an automatic declustering scheme to avoid the subjective choice of clusters by automatically choosing a run length. They use the fact that the non-degenerate limiting distribution of the normalised times between exceedances of a threshold (inter-exceedance times) is exponential for IID random variables. For the general stationary process case, with extremal index $\theta \in [0, 1]$, the distribution is a mixture of an exponential and a degenerate probability distribution at 0. The normalised inter-arrival time distribution has the non-degenerate mixture form of $(1 - \theta)\epsilon_0 + \theta\mu_\theta$, where ϵ_0 is a point mass at 0 and μ_θ is the exponential distribution with mean θ^{-1}, with mixture weights $1 - \theta$ and θ, respectively. The extremal index represents the proportion of non-zero inter-exceedance times, as well as the reciprocal of the mean of these non-zero times. In other words, the extremal index is the proportion of times between exceedances that can be regarded as times between clusters (inter-cluster times). This provides a limiting argument for the automatic declustering scheme, where theoretical moments of the limiting distribution are equated to their empirical counterparts.

Ferro and Segers (2003) define another estimator for the extremal index, the intervals estimator, that is based on the limiting distribution of inter-exceedance times discussed above. For a sample of size n, with n_u observations exceeding a threshold u, and observed inter-exceedance times T_i for $i = 1, \ldots, (n_u - 1)$, the intervals estimator is

given by,

$$\tilde{\theta}_n(u) = \begin{cases} \min\{1,\ \hat{\theta}_n(u)\} & \text{if } \max\{T_i : 1 \leq i \leq (n_u - 1)\} \leq 2 \\ \min\{1,\ \hat{\theta}^*_n(u)\} & \text{if } \max\{T_i : 1 \leq i \leq (n_u - 1)\} > 2, \end{cases}$$

where

$$\hat{\theta}_n(u) = \frac{2\left(\sum_{i=1}^{n_u-1} T_i\right)^2}{(n_u - 1)\sum_{i=1}^{n_u-1} T_i^2} \quad \text{and} \quad \hat{\theta}^*_n(u) = \frac{2\left\{\sum_{i=1}^{n_u-1} (T_i - 1)\right\}^2}{(n_u - 1)\sum_{i=1}^{n_u-1} (T_i - 1)(T_i - 2)}.$$

Ferro and Segers (2003) make a direct comparison of the runs and intervals estimators via a simulation study using data from a first-order stationary max-autoregressive moving average, Max-ARMA(1,0), process with differing extremal indices. Such a process is defined by $X_t = \max\{(1 - \theta)X_{t-1}, W_t\}$ for $t \in \mathbb{Z}$ where $\theta \in (0, 1]$ is the extremal index and $\{W_t; t \in \mathbb{Z}\}$ is an IID series of unit Fréchet random variables. Max-ARMA processes are an extension of the commonly used ARMA models in time series, that are useful for data with heavy-tailed behaviour; we discuss Max-ARMA(p, q) processes extensively in Chapter 6. The results of Ferro and Segers (2003) emphasize the sensitivity of the runs estimator to parameter choices (the run length and extremal index), whilst the intervals estimate avoids this sensitivity. An interesting result is that the runs estimate (2.3.2) appears superior to the intervals estimator for all Max-ARMA processes when there are fewer exceedances, based on comparisons of root mean square error.

2.3.3 Diagnostics for dependence

When studying dependence in extreme value theory, we are concerned with two different types of extremal dependence: asymptotic dependence and asymptotic independence, which are formally defined later. These are typically described in a bivariate

setting (Joe, 1997) but we derive them here for a stationary univariate time series to describe the extremal dependence in a stationary series $\{X_t; t \in \mathbb{Z}\}$ for values separated by time lag $\tau \in \mathbb{Z}$, i.e., for $(X_t, X_{t+\tau})$. In this section, we detail three measures of extremal dependence: the coefficient of asymptotic dependence χ, the coefficient of asymptotic independence η and the extremogram ρ.

Coefficient of asymptotic dependence

When considering extremal dependence in a stationary univariate time series context, it is natural to consider the probability

$$\chi_\tau = \lim_{x \to x^F} \Pr(X_{t+\tau} > x | X_t > x),$$

where $\tau \in \mathbb{N}$ and x^F is the upper end-point of the common marginal distribution of X_t and $X_{t+\tau}$ (Coles et al., 1999). We say that X_t and $X_{t+\tau}$ are asymptotically dependent, or exhibit extremal dependence, when $\chi_\tau > 0$. When $\chi_\tau = 0$, we say that X_t and $X_{t+\tau}$ are asymptotically independent, whilst perfect extremal dependence corresponds to $\chi_\tau = 1$. Considering χ_τ for all $\tau \in \mathbb{N}$ gives insight into the asymptotic dependence structure across time lags and note that $\chi_\tau = \chi_{-\tau}$ for all τ, so we only consider $\tau \in \mathbb{N}$.

Asymptotic dependence and asymptotic independence cannot be compared with generic dependence; a series can be asymptotically independent but exhibit strong correlation at the same time, and vice versa. For example, take two normal random variables (X_1, X_2) with $X_1 \sim \mathrm{N}(0, 1)$ and

$$X_2 = \begin{cases} X_1 & \text{with probability } 1/2 \\ -X_1 & \text{with probability } 1/2, \end{cases}$$

so $X_2 \sim \mathrm{N}(0, 1)$. Then, Kendall's τ measure of correlation is zero, which is used for measuring dependence over all of the data, but $\chi = 1/2$ so these variables are asymptot-

ically dependent. Note (X_1, X_2) do not have a bivariate normal distribution although they have normal margins. Alternatively, all variables from the standard bivariate normal family (with correlation parameter $\rho < 1$) are asymptotically independent but can be strongly correlated in the body if ρ is close to 1. In this case, the coefficient of asymptotic dependence $\chi_\tau = 0$ does not reflect the relative strength of the dependence in extremes at lag τ. Therefore alternative measures of extremal dependence have been developed that we detail next.

Identifying a process as asymptotically independent or asymptotically dependent is fundamental since in an asymptotically dependent sequence, a large value is likely to be followed by another large value, so we expect to observe clusters forming at extreme levels. As discussed in Section 2.3.2, identifying limiting clusters is a key step in understanding the dependence structure so that we can apply standard extreme value models.

Coefficient of asymptotic independence

Let the stationary process $\{X_t;\ t \in \mathbb{Z}\}$ have unit Fréchet margins, without loss of generality. Then, for a slowly varying function $\mathcal{L}_\tau(\cdot)$, and $\eta_\tau \in (0, 1]$ the coefficient of asymptotic independence at lag $\tau \in \mathbb{N}$, Ledford and Tawn (1996) argue that

$$\Pr(X_{t+\tau} > x \mid X_t > x) \sim \mathcal{L}_\tau(x) x^{1-1/\eta_\tau} \quad \text{as } x \to \infty.$$

When $\eta_\tau = 1$, then $\Pr(X_{t+\tau} > x | X_t > x) = \mathcal{L}_\tau(x) \to \chi_\tau > 0$ as $x \to \infty$ (provided $\mathcal{L}_\tau(x) \not\to 0$). This means there is a non-zero probability of $X_{t+\tau}$ being large when X_t is large at all extreme levels, corresponding to asymptotic dependence with $\chi_\tau > 0$. When $\eta_\tau < 1$, $\Pr(X_{t+\tau} > x | X_t > x) \to 0$ as $x \to \infty$ irrespective of the form of $\mathcal{L}_\tau(x)$; this corresponds to asymptotic independence, i.e. $\chi_\tau = 0$, since the probability that $X_{t+\tau}$ is large given X_t is large converges to 0 as more extreme levels are considered (i.e., as $x \to \infty$). Since $\eta_\tau = \eta_{-\tau}$ for all τ, we only consider $\tau \in \mathbb{N}$.

To summarise extremal dependence at lag τ, the pair of measures (χ_τ, η_τ) is needed. Under asymptotic dependence, when $\chi_\tau > 0$ and $\eta_\tau = 1$, we consider χ_τ to be a measure of the strength of dependence. Conversely, when we have asymptotic independence, where $\chi_\tau = 0$ and $\eta_\tau < 1$, we use η_τ to signify the strength of dependence between variables.

Coles et al. (1999) also introduce $\bar{\chi}_\tau$ as a measure of extremal dependence that is analogous to η_τ, but on a more useful and interpretable scale. Ledford and Tawn (2003) also introduce a measure of serial dependence in extreme values separated by lag $\tau \in \mathbb{N}$ as

$$\Lambda_\tau = 2\eta_\tau - 1 \in (-1, 1],$$

where this is interpreted similarly to the autocorrelation function in time series analysis. When $\Lambda_\tau = 1$, then $\eta_\tau = 1$, so provided $\mathcal{L}_\tau(s) \not\to 0$ as $s \to \infty$, this indicates asymptotic dependence. Whilst $\Lambda_\tau < 1$ corresponds to $\eta_\tau < 1$, so X_t and $X_{t+\tau}$ are asymptotically independent, and X_t and $X_{t+\tau}$ are independent when $\eta_\tau = 1/2$ and $\Lambda_\tau = 0$. Additionally $0 < \Lambda_\tau < 1$, $\Lambda_\tau = 0$ and $\Lambda_\tau < 0$ correspond to positive extremal association, 'near' independence and negative extremal association, respectively. As $\Lambda_\tau = \Lambda_{-\tau}$ for all τ, we only consider $\tau \in \mathbb{N}$.

The extremogram

Davis and Mikosch (2009) define an alternative measure of serial dependence: the extremogram. Like χ_τ, this is an extreme value analogue of the autocorrelation function of a stationary process. For a regularly varying marginal stationary series $\{X_t;\ t \in \mathbb{Z}\}$, the extremogram is defined for two sets A and B, that are bounded away from zero, by the limit

$$\rho_{AB}(\tau) = \lim_{n\to\infty} \Pr(a_n^{-1} X_t \in A \mid a_n^{-1} X_{t+\tau} \in B),$$

for $a_n \to \infty$ such that $\Pr(|X_t| > a_n) \sim n^{-1}$ as $n \to \infty$. If, for example, $\{X_t\}$ has

unit Fréchet margins, then $a_n = n$. By defining A and B as bounded away from zero, the events $\{a_n^{-1} X_t \in A\}$ and $\{a_n^{-1} X_{t+\tau} \in B\}$ become extreme in the limit. For $A = B = (1, \infty)$, the extremogram becomes the coefficient of extremal dependence between X_t and $X_{t+\tau}$, i.e., $\rho_{AB}(\tau) = \chi_\tau$.

2.3.4 Copulas

Copula functions provide an appropriate model for the dependence structure between variables and are classically applied in the multivariate setting. We refer the reader to Joe (2014) and Nelsen (2006) for a detailed review of copulas. We describe copulas first in this multivariate setting, but later demonstrate how they can be used for modelling temporal dependence in a univariate series under a kth order Markov model assumption (Winter and Tawn, 2017).

Multivariate case

Consider $d \geq 2$ random variables $(X_1, \ldots, X_d)$ with joint distribution function F and marginals with distribution functions $(X_1 \sim F_1, \ldots, X_d \sim F_d)$. The dependence structure among these variables can be modelled using a distribution $C : [0,1]^d \to [0,1]$ with Uniform(0,1) margins, known as a copula. If $(X_1, \ldots, X_d) \sim F$, then Sklar's theorem (Sklar, 1959) tells us that the multivariate distribution function can be written as a composition of a copula C and the marginal distributions, as follows

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

where $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Copulas are advantageous because they are invariant to marginal transformations, so can describe the association between the variables in a way that is independent of the marginal distributions (Coles et al., 1999). This copula C is unique for continuous random variables. To find the copula form, the

probability integral transform (PIT) is used to transform between margins. For any continuous random variable Y, with distribution function F_Y, we can apply the PIT to transform between margins by obtaining a Uniform(0,1) random variable via $U = F_Y(Y)$ with $U \sim$ Uniform(0,1). Similarly from a Uniform(0,1) random variable U we can obtain another continuous random variable T with distribution function F_T and inverse distribution function F_T^{-1} using $T = F_T^{-1}(U)$. Then we can obtain the copula as follows,

$$C(u_1, \ldots, u_d) = F\big(F_{X_1}^{-1}(u_1), \ldots, F_{X_d}^{-1}(u_d)\big),$$

for $(u_1, \ldots, u_d) \in (0,1)^d$.

There are many established parametric forms that copulas can take (Joe, 2014), here we present the Gaussian and multivariate logistic extreme value distribution (subsequently referred to as the logistic) copulas. These both model the dependence of variables via a single parameter.

Let U and V be uniform random variables in (0,1). The bivariate Gaussian copula with correlation parameter $\rho \in (-1, 1)$ is given by

$$C(u, v, \rho) = \mathbf{\Phi}_2\big(\Phi^{-1}(u), \Phi^{-1}(v); \rho\big), \quad u, v \in (0,1),$$

where $\mathbf{\Phi}_2(\cdot, \cdot)$ is the bivariate standard normal distribution and $\Phi^{-1}(\cdot)$ the inverse of the univariate standard normal distribution function. Then the Gaussian copula density can be written as

$$c(u, v; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2 x^2 + \rho^2 y^2 - 2\rho xy}{2(1-\rho^2)}\right\}, \quad u,\, v \in (0,1),$$

where $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$. For the Gaussian copula, the extremal dependence measures are $\chi = 0$ and $\Lambda = \rho$ so this copula is suitable for asymptotically independent variables with the sign of extremal association determined by the value of ρ (Heffernan,

2000).

The logistic copula, introduced by Émile and Gumbel (1960), with parameter $0 < \alpha \leq 1$ is given by

$$C(u, v; \alpha) = \exp\left\{ -\left[(-\log u)^{1/\alpha} + (-\log v)^{1/\alpha}\right]^{\alpha}\right\}, \quad u, v \in (0, 1).$$

When $\alpha = 1$, U and V are clearly independent, with $C(u, v; 1) = uv$. The logistic density is written as

$$c(u, v; \alpha) = \frac{C(u, v; \alpha)}{uv}(xy)^{1/\alpha - 1}\left(x^{1/\alpha} + y^{1/\alpha}\right)^{\alpha - 2}\left[\left(x^{1/\alpha} + y^{1/\alpha}\right)^{\alpha} + \alpha^{-1} - 1\right],$$

for $x = -\log(u)$ and $y = -\log(v)$. For this copula, $\chi = 2 - 2^{\alpha}$ and $\Lambda = 1$ when $0 \leq \alpha < 1$, with dependence strengthening as α decreases to the limit; $\chi = 1$ when $\alpha = 0$, so this is suitable for asymptotically dependent variables. Whereas when $\alpha = 1$, $\chi = 0$ and $\Lambda = 0$, so as stated above, U and V are independent (Heffernan, 2000).

Markov case

Copulas can be used in the univariate setting to model temporal dependence between observations separated by time lag $\tau \in \mathbb{N}$. Winter and Tawn (2017) demonstrate this through an application to heatwaves. Under the assumption that the stationary series $\{X_t;\ t \in \mathbb{Z}\}$, with marginal distribution function F_X, follows a kth order Markov chain, the joint density of $\boldsymbol{X}_{1:n} = (X_1, \ldots, X_n)$, denoted $f_{1:n}$ for $n > k$, can be written as

$$f_{1:n}(\boldsymbol{x}_{1:n}) = f_{1:k}(\boldsymbol{x}_{1:k}) \prod_{t=1}^{n-k} f_{k+1|1:k}(x_{t+k} \mid \boldsymbol{x}_{t:(t+k-1)}) = f_{1:k}(\boldsymbol{x}_{1:k}) \prod_{t=1}^{n-k} \frac{f_{1:k+1}(\boldsymbol{x}_{t:t+k})}{f_{k+1}(x_{t+k})}, \tag{2.3.3}$$

where $f_{k+1|1:k}$ is the conditional density of X_{k+1} given $\boldsymbol{X}_{1:k}$, and the subscript $i : j$ corresponds to variable indices $(i, i+1, \ldots, j)$ for $i \leq j$. We use the law of conditional probability to obtain the second equality above, exploiting the stationarity of

the process so that the joint marginal distribution satisfies

$$f_{i:j}(\boldsymbol{x}_{i:j}) = f_{i+\tau:j+\tau}(\boldsymbol{x}_{i:j}), \tag{2.3.4}$$

for $i \leq j$, any $\tau \in \mathbb{N}$ and all $\boldsymbol{x}_{i:j} \in \mathbb{R}^{j-i+1}$. Copula models can then be used for modelling the joint distribution $f_{1:k+1}$, with the corresponding joint distribution function $F_{1:k+1}$ satisfying

$$F_{1:k+1}(\boldsymbol{x}_{1:k+1}) = C_{1:k+1}\{F_X(x_1), \ldots, F_X(x_{k+1})\},$$

where $C_{1:k+1}$ is the copula for $(k+1)$ successive variables in the Markov chain. The Markov process inherits the stationary condition of equality (2.3.4) if the copula $C_{1:k+1}$ has the property that it's m-dimensional marginal distribution, for all $m < k+1$, satisfies

$$C_{i_1,\ldots,i_m}(\boldsymbol{x}_{1:m}) = C_{i_1+\tau,\ldots,i_m+\tau}(\boldsymbol{x}_{1:m}),$$

for $\tau \in \mathbb{N}$, $i_j \in \mathbb{N}$ for $j = 1, \ldots, m$ with $1 \leq i_1 < \ldots < i_m + \tau \leq k+1$, and $\boldsymbol{x}_{1:m} \in \mathbb{R}^m$. Then, the joint density $f_{1:k+1}$ of (2.3.3) can be rewritten as the density of the copula, so that we can obtain the joint marginal density $f_{1:n}$ in terms of a product of copula densities (Winter and Tawn, 2017). Therefore, under a kth order Markov assumption, multivariable copula theory can be easily extended to a stationary univariate series.

2.4 Extremes of non-stationary sequences

Whilst stationarity is a more realistic assumption than IID, it is common to have a univariate time series exhibit non-stationarity, where the characteristics of the process change systemically with covariate values. Environmental variables typically exhibit non-stationarity across time as it is common for different seasons to have different climate patterns and to observe long term trends owing to anthropogenic climate change (Seneviratne et al., 2021). Accounting for these non-stationary features is a

common theme throughout this thesis.

Standard extreme value methods discussed so far are not applicable to non-stationary processes, instead additional steps are required to capture the underlying covariate effects. There is no general theory that can be applied to a non-stationary process since the forms of non-stationarity vary significantly. We account for these features using a variety of statistical modelling techniques and use these models to enhance the extreme value models discussed in Sections 2.2.1 and 2.2.2.

Carter and Challenor (1981) present one of the earliest approaches for capturing seasonality in extreme value models for environmental variables. They suggest splitting the year into months and modelling each month's data using separate extreme value distributions so that a particular month has the same distribution across years. This assumes stationarity within each month across years but implies a discontinuity in behaviour from one month to another, and the approach is not parsimonious. More sophisticated methods for non-stationary extremes have been developed since this early approach (Chavez-Demoulin and Davison, 2005; Eastoe and Tawn, 2009), including those developed in this thesis (see Chapter 3).

In this section, we focus on the GPD for modelling non-stationary process extremes as this is more commonly used throughout the thesis and in practice. Approaches are similar for the GEV modelling in these contexts but are more complicated if covariates vary across a block. We refer the reader to Parey et al. (2013) for an example of incorporating non-stationarity in the GEV. We discuss two main approaches to capture non-stationarity. Firstly, we introduce the non-stationary GPD in Section 2.4.1 where we allow the model parameters to vary with covariates; we present parametric and non-parametric approaches for doing so and discuss extensions of the GPD model initially proposed by Davison and Smith (1990). An alternative approach is the preprocessing model of Eastoe and Tawn (2009) where covariate effects are modelled and removed before the potentially non-stationary extreme value analyses of Davison and Smith

(1990). The latter non-stationary modelling is an attempt to explain covariate effects in the tail of the distribution that are different from those in the data; we discuss this method in Section 2.4.2. Lastly, in Section 2.4.3, we demonstrate how to estimate the unconditional model for extremes from a covariate-dependent model to give results for marginal extremes.

2.4.1 Non-stationary GPD

Let $\{X_t;\ t \in \mathbb{Z}\}$ now denote a non-stationary sequence with associated covariate vector $\boldsymbol{Z}_t \in \mathbb{R}^d$ for $d \in \mathbb{N}$ and $t \in \mathbb{Z}$. We are interested in modelling the extremes of the sequence $X_t \mid \boldsymbol{Z}_t = \boldsymbol{z}$ where $\boldsymbol{z} \in \mathbb{R}^d$ for covariate realisation $\boldsymbol{z}$. We can do so by modelling the extremes of X_t using the GEV or GPD as usual, but allowing the parameters to be functions of the covariates to capture the non-stationarity. Formally, Davison and Smith (1990) define the non-stationarity GPD model as,

$$X_t - u \mid (X_t > u, \boldsymbol{Z}_t = \boldsymbol{z}) \sim \text{GPD}\big(\sigma_u(\boldsymbol{z}), \xi(\boldsymbol{z})\big),$$

where each parameter of the GPD, i.e., say $\theta \in \{\sigma_u, \xi\}$, has the form $\theta(\boldsymbol{z}) = h(\boldsymbol{z}^T\boldsymbol{\beta})$, where $\boldsymbol{z}^T$ is the transpose of the covariate vector, $\boldsymbol{\beta} \in \mathbb{R}^d$ is the vector of coefficients, and $h : \mathbb{R} \to \Theta$ is a specified function, termed the inverse link function, that transforms the linear term on $\mathbb{R}$ into the space Θ, where Θ is the feasible parameter space of θ. Link functions are commonly used to constrain a parameter to its domain, for example, the log-link function is typically used for the scale parameter function $\sigma_u(\boldsymbol{z})$ to ensure positivity for all values of $\boldsymbol{z}$, i.e., $\log \sigma_u(\boldsymbol{z}) = \boldsymbol{z}^T\boldsymbol{\beta}$ with $\Theta = (0, \infty)$. The parameters of these regression coefficients are estimated at the same time as extreme value model fitting.

When defining extremes as exceedances of a fixed threshold u, we can also model the rate of exceedance in terms of the covariates, i.e., $\lambda_u(\boldsymbol{z}) = \Pr(X_t > u | \boldsymbol{Z}_t = \boldsymbol{z})$. Here

the link function is typically the logit-link function since $\lambda_u(\boldsymbol{z}) \in [0,1] = \Theta$. Whilst any of the GPD parameters and $\lambda_u(\boldsymbol{z})$ can be influenced by covariates, it is uncommon to allow the shape parameter to vary since it is difficult to estimate; keeping the shape parameter constant across covariates avoids introducing additional uncertainty into the model (Chavez-Demoulin and Davison, 2005). Although, one should check a constant shape parameter is a reasonable assumption.

For parametric regression, the form of covariate dependence is explained through regression parameters and is specified before fitting (i.e., the choice of linearity and/or covariate selection). Choosing a parametric form is somewhat subjective and benefits from understanding the process of interest. Standard statistical model selection and comparison tools can be used to help determine the appropriate form of the covariate models. In contrast, for non-parametric regression, the relationship is not specified before model fitting. Instead, the covariate effects are estimated based on an assumption of smoothness of the function $\theta(\boldsymbol{z})$, as opposed to a particular parametric form, so that $\theta(\boldsymbol{z}) = h(\boldsymbol{z})$ where h is now d-dimensional so that $h : \mathbb{R}^d \to \Theta$ (Northrop et al., 2016). We discuss these parametric and non-parametric approaches further in the following subsections.

Parametric covariate functions

Smith (1989) and Davison and Smith (1990) were the first to use parametric regression techniques within extreme value models. Parametric functions of covariates can take a range of functional forms with differing complexities. Perhaps the simplest form for a parameter is a linear model; these can be extended to quadratic or cubic functions. Another simple parametric form uses piecewise constant functions, where stationarity is assumed over consecutive subsets of the covariate samples. Ross et al. (2018) adopt this approach for modelling extreme surges in the North Sea. Here the domain of the storm direction covariate is partitioned into subsets so that corresponding surge

observations can be reasonably assumed as stationary. Coles et al. (1994) use harmonics to model seasonal variability in their extreme value model for low temperature data. Using continuous-time periodic models avoids arbitrarily choosing seasonal blocks and allows the parameters to vary smoothly across time, or an alternative covariate, giving a parsimonious model.

Parametric covariate functions can be easily incorporated into inference by replacing the constant parameters with their functional form within the likelihood function. However, parametric regression models require that $h(\boldsymbol{z}^T\boldsymbol{\beta})$ takes a predefined functional form, which can often be restrictive to the forms for $\sigma_u(\boldsymbol{z})$ and $\lambda_u(\boldsymbol{z})$. Additionally, whilst the estimated forms for $\sigma_u(\boldsymbol{z})$ and $\lambda_u(\boldsymbol{z})$ observed within the data can be suitably modelled with a parametric regression model, it may be unlikely that this pattern will apply for $\boldsymbol{z}$ outside of the observed range. For example, long term climate change trends, say with t as time as a covariate, can often be adequately modelled by a linear function in the range of observed data. However, under different scenarios, this linear trend could change in the future as the relationship is likely with different covariates representing the presence of greenhouse gases (see Chapter 4).

Generalised additive models

More recently, generalised additive models (GAMs) have been used for capturing covariate trends using smooth functions within extreme value models (Chavez-Demoulin and Davison, 2005). GAMs are extensions of generalised linear models with a linear predictor involving a sum of smooth functions of covariates; see Wood (2017) for an overview of smooth functional forms. GAMs are a type of semi-parametric model that are less restrictive than parametric models because the functional forms between covariates and model parameters have more degrees of freedom than given by parametric relationships.

Generalised additive extreme value models (EVGAMs) were proposed by Youngman

(2019) and are widely implemented via the `evgam` R package (Youngman, 2022). Note that `evgam` captures non-stationarity through the scale and shape parameters, but not the rate parameter; in the next subsection we discuss how `evgam` is used for modelling non-stationarity in the GPD threshold. The rising popularity of EVGAM is due to its flexible functional forms that enable multiple covariate interactions for covariates that are either discrete or continuous, so that complex covariate relationships can be modelled. However, this flexibility of modelling covariates in the data range makes extrapolation for learning about future behaviours more difficult as these functions revert to a simplistic formulation when moving well beyond the observed covariate data range.

Non-stationary thresholds and quantile regression

Another extension of the Davison and Smith (1990) framework is allowing the threshold to be covariate dependent (Kyselý et al., 2010). This makes sense in practice because what is a suitably high threshold for one covariate may be too low a threshold for another covariate. Additionally, to precisely model covariate effects it makes sense to have exceedances spread across the observed covariate values (Northrop et al., 2016). Northrop and Jonathan (2011) use quantile regression to set a threshold so that the probability of exceeding this threshold is approximately constant across covariates. This procedure can be easily implemented using the `evgam` package (Youngman, 2022). Northrop and Jonathan (2011) argue that it is more logical to allow the threshold to vary with covariates than to keep a constant threshold and then model how the rate parameter λ_u varies with covariates. This makes sense when the covariate effects are large, however, when the covariate effects are small in comparison to the variability of the data, there are different perspectives on whether to model these effects through the model parameters or the threshold.

2.4.2 Pre-processing

Eastoe and Tawn (2009) propose an alternative method to those discussed in Section 2.4.1, based on the common approach for handling non-stationarity in time series analysis, known as the preprocessing model. The main difference is that non-stationarity is first modelled for the entire data set and then removed. Let $\tilde{X}_t$ denote the preprocessed series corresponding to the non-stationary series X_t, so that the main body of $\tilde{X}_t$ is now stationary. Covariate effects on the extremes of X_t may differ from those in the body of the distribution of X_t, so the tails of the preprocessed series $\tilde{X}_t$ are not guaranteed to be like that of a stationary process. Therefore, non-stationary extreme value models (such as those discussed in Section 2.4.1) are used for the extreme value analysis of $\tilde{X}_t$. The crux of the idea is that all, or at least the most complex, non-stationarity will be removed by the preprocessing step, making the subsequent non-stationary extreme value modelling simpler.

Eastoe and Tawn (2009) argue that their preprocessing approach is an improvement to the standard method of Davison and Smith (1990) since the reasons for non-stationarity are often associated with the mechanisms that generate the process. If the covariate effects observed in the body of the data are the same as those in the tail, it is inefficient to estimate these effects only using extreme values.

As with the approach of Section 2.4.1, modelling the covariate effects (here for the entire distribution of X_t) is a fundamental step. Eastoe and Tawn (2009) propose the Box-Cox location-scale model for preprocessing; this is given by

$$\frac{X_t^{\phi(\boldsymbol{z}_t)} - 1}{\phi(\boldsymbol{z}_t)} = \mu(\boldsymbol{z}_t) + \psi(\boldsymbol{z}_t)\tilde{X}_t, \tag{2.4.1}$$

for all $t \in \mathbb{Z}$ where ϕ, μ and $\log \psi$ are linear functions of the covariates $\boldsymbol{z}_t \in \mathbb{R}^d$. However, this may not fully capture all the covariate effects and more established methods could be implemented if expert knowledge of the problem exists (e.g., by using the parametric

regression functions of Section 2.4.1). Diagnostic tools for checking stationarity can be used to detect model misspecification. In situations where misspecification is small and undetected, Eastoe and Tawn (2009) show that their approach remains superior to that of Davison and Smith (1990).

Once an appropriate extreme value model is chosen for $\tilde{X}_t \mid \boldsymbol{Z}_t = \boldsymbol{z}_t$, there is a need to find the equivalent model for the original series X_t. Say the Box-Cox location-scale model of (2.4.1) is used for preprocessing, then we obtain the conditional distribution for $X_t \mid \boldsymbol{Z}_t = \boldsymbol{z}_t$, for suitably large x, as follows

$$\begin{aligned}\Pr(X_t > x \mid \boldsymbol{Z}_t = \boldsymbol{z}_t) &= \Pr\left(\mu(\boldsymbol{z}_t) + \psi(\boldsymbol{z}_t)\tilde{X}_t > \frac{x^{\phi(\boldsymbol{z}_t)} - 1}{\phi(\boldsymbol{z}_t)} \middle| \boldsymbol{Z}_t = \boldsymbol{z}_t\right) \\ &= \Pr\left(\tilde{X}_t > \frac{\frac{1}{\phi(\boldsymbol{z}_t)}[x^{\phi(\boldsymbol{z}_t)} - 1] - \mu(\boldsymbol{z}_t)}{\psi(\boldsymbol{z}_t)} \middle| \boldsymbol{Z}_t = \boldsymbol{z}_t\right) \\ &= \tilde{\lambda}_{\tilde{u}}(\boldsymbol{z}_t)\left\{1 + \tilde{\xi}(\boldsymbol{z}_t)\left[\frac{\frac{1}{\psi(\boldsymbol{z}_t)}\left(\frac{1}{\phi(\boldsymbol{z}_t)}[x^{\phi(\boldsymbol{z}_t)} - 1] - \mu(\boldsymbol{z}_t)\right) - \tilde{u}}{\tilde{\sigma}_{\tilde{u}}(\boldsymbol{z}_t)}\right]\right\}^{-1/\tilde{\xi}(\boldsymbol{z}_t)},\end{aligned}$$

where $\tilde{u}$ is the threshold for defining exceedances for $\tilde{X}_t$ and $\tilde{\sigma}_{\tilde{u}} > 0$, $\tilde{\xi}$ and $\tilde{\lambda}_{\tilde{u}}$ are the scale, shape and rate parameters, respectively, of the potentially non-stationary GPD. We detail how to obtain the marginal distribution of the original series $\{X_t\}$ in the next section.

2.4.3 Marginal quantile estimation

By capturing non-stationarity via the approaches outlined so far in this section, we obtain the conditional distribution $X_t \mid \boldsymbol{Z_t} = \boldsymbol{z}_t$ and can summarise the conditional tail characteristics. However, it is often of interest to estimate the marginal distribution of X_t, for an arbitrary $t \in \mathbb{Z}$, so that we can quote return levels that are independent of covariates. The marginal distribution can be estimated using Monte-Carlo techniques (Eastoe and Tawn, 2009). Specifically, provided we have a large sample size n, it is reasonable to assume that the observed covariate sample $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_d)$ is a

representative sample from $\boldsymbol{Z}$. Then we can approximate the marginal distribution as follows using the conditional distribution $F_{X|\boldsymbol{Z}}$, as in Section 2.4.2, and covariate density function $f_{\boldsymbol{Z}}$, as follows

$$\hat{F}_X(x) = \int_{\boldsymbol{Z}} F_{X|\boldsymbol{Z}}(x \mid \boldsymbol{z}) f_{\boldsymbol{Z}}(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \approx \frac{1}{n} \sum_{t=1}^{n} F_{X|\boldsymbol{Z}}(x \mid \boldsymbol{z}_t).$$

Chapter 3

Accounting for seasonality in extreme sea level estimation

3.1 Introduction

Extreme sea levels pose an increasing risk to coastline communities. In the absence of any mitigation, the impacts can be severe: fatality, infrastructure damage and habitat destruction. Estimates of sea level return levels are fundamental for various purposes, including coastal flood defence design and flood risk assessments. A return level is the value we expect the annual maximum sea level to exceed with probability p. For a stationary series, this corresponds to a value exceeded once every $1/p$ years, on average. We are particularly interested in rare events, where $p \in [10^{-4}, 10^{-1}]$ to cover sea levels that are important to a range of industries affected by coastal flooding. For example, we consider data from the Heysham tide gauge; Heysham is a coastal town in north-west England that is home to two nuclear power stations. Nuclear regulators require accurate return level estimates for $p = 10^{-4}$. We also consider a further three gauges at Lowestoft, Newlyn and Sheerness; Sheerness is particularly important since it is located on the River Thames estuary where extreme sea levels can propagate down the river

towards London. The annual return level, or equivalently its exceedance probability, changes with year if the series is non-stationary and discussions of appropriate design levels involve the intended lifespan of the defence (Rootzén and Katz, 2013). We focus on a fixed p and recognise that the return level varies across years.

Since the UK is regularly subject to coastal flooding, estimating extreme sea levels is crucial. The 1953 North Sea flood is the worst on record for the 20$^{\text{th}}$ century in the UK. Coastal defences were breached in 1,200 places leading to the evacuation of 30,000 people, damage to 24,000 properties and a death toll of 307 in England alone. The damage was estimated at £1.2 billion in 2014 (Wadey et al., 2015). Following this event, coastal flood defences were upgraded around most of the UK. Recently, there has been growing concern regarding anthropogenic sea level rise due to climate change. Rises in the mean sea level, coupled with changes in storm frequency and size, can increase the likelihood of coastal flooding. Therefore, it is increasingly important to accurately estimate sea level return levels so that coastline communities are protected against events such as that in 1953. The latest best estimates for extreme sea levels can be found in Environment Agency (2018).

Pugh and Woodworth (2014) give a comprehensive overview of sea level processes. Sea levels are a combination of mean sea level, tide, surge and waves. We consider still water level (with waves filtered out), for simplicity we refer to this as sea level. The mean sea level trend is removed so that tide and surge are the only components. Tides are the regular and predictable changes in sea levels driven astronomically; these are well understood and perfectly forecast (Egbert and Ray, 2017). Surges define any departure from the predicted tidal regime, often resulting from meteorological forces such as storms and are hence sometimes called storm surges or non-tidal residual (as in Figure 3.1.1); these are stochastic. However, surges can also include gauge recording errors and tidal prediction errors, but these errors are typically negligible at well maintained gauges, such as those on the UK National Tide Gauge Network (see Sec-

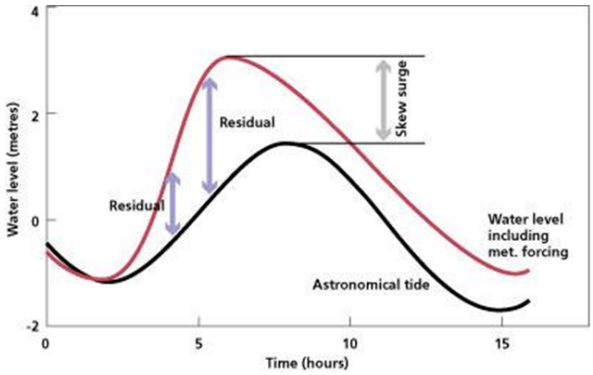


Figure 3.1.1: Sea levels during the passage of a surge for a single tidal cycle (NOC, 2021).

tion 3.3.1). Surges can also be influenced by tide-surge interaction, this is the change in the distribution of surge that is dependent on the tidal level, because the surge is essentially a wave which is influenced by the water depth. This interaction is more prominent in shallow water areas, with the largest surges typically occurring mid-tide on the rising tide and the smallest at high tide (Prandle and Wolf, 1978; Tawn, 1992).

Early methods to estimate extreme sea levels modelled the observed sea levels directly, ignoring the known tidal component. More recent approaches focus on the convolution of the surge and tide distributions. Due to the complex dependence between surge and tide, we instead consider skew surge and peak tide; skew surge is the difference between the maximum observed sea level and the maximum predicted tide (peak tide) within a tidal cycle, regardless of their timing (see Figure 3.1.1). There is a single skew surge value associated with each peak tide value (every 12 hours 26 minutes), as opposed to hourly or 15 minute interval observations of surge and tide. Even though there are fewer skew surge and peak tide observations, these components are preferred because they have much weaker dependence; Williams et al. (2016) demonstrate that it is reasonable to assume these are independent at most UK sites.

We build on the skew surge joint probabilities method (SSJPM) of Batstone et al. (2013), which assumes peak tide and skew surge are independent and that they are stationary processes within and across years, with extreme skew surges modelled by a

generalised Pareto distribution (GPD). Our novel approach is able to reflect the realism of the sea level processes by correcting the simplifying and false assumptions made in the SSJPM. We account for seasonality in the skew surges by adding a daily covariate to the rate of threshold exceedance and scale parameter of the GPD model. We also introduce a tidal covariate to capture skew surge-peak dependence. Skew surges also exhibit temporal dependence; we account for this using the extremal index (Tawn, 1992) but in a different way to previous analyses. Since tides are deterministic, we choose our tidal samples so that they account for monthly and interannual variations. We estimate return levels by deriving distributions for the annual maxima sea levels. We obtain the estimates using our proposed method and for the SSJPM, as well as some intermediate methods, at four sites on the UK coastline. We find that the SSJPM tends to underestimate return levels.

It is fundamental to recognise the uncertainty associated with return level estimates of the highest order to add value to the point estimates. We construct confidence intervals for our estimates using a stationary bootstrap procedure (Politis and Romano, 1994). This preserves the realism of the sea level processes. Our approach is the first to use a bootstrap procedure for uncertainty quantification on extreme sea level estimates. We incorporate prior information for the GPD shape parameter based on related spatial information (Environment Agency, 2018); we show that for the shape parameters and the sea-level return level estimates, this information substantially reduces the lengths of confidence intervals, without significantly changing point estimates.

We discuss the relevant extreme value theory and existing methodology for extreme sea level estimation in Section 3.2. In Section 3.3 we introduce the data and explore the seasonality of each component as well as the dependence between them. Section 3.4 describes the methodology for deriving the annual maxima distribution, starting with an idealised solution derived under simplified assumptions that are then relaxed. Our return level estimates are compared with previous methods in Section 3.5. Additional

supplementary material is presented in D'Arcy et al. (2023c) and Appendix A.

3.2 Background to method

3.2.1 Relevant extreme value methods

When deriving a model for extremes, it is natural to first consider the maximum M_n of a sequence of independent and identically distributed (IID) continuous random variables $Z_1, \ldots, Z_n$, i.e., $M_n = \max\{Z_1, \ldots, Z_n\}$. This sequence has marginal distribution function F and upper end point z^F. If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, so that the rescaled block maximum $(M_n - b_n)/a_n$ has a nondegenerate limiting distribution as $n \to \infty$, then the cumulative distribution function of this limit must have the form

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{3.2.1}$$

where $x_+ = \max\{x, 0\}$ for parameters $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ representing the location, scale and shape, respectively (Coles, 2001). This is the generalised extreme value distribution (GEV) that encompasses three families: for $\xi > 0$ this corresponds to the Fréchet distribution, $\xi < 0$ the Weibull and $\xi = 0$ the Gumbel. Note $\xi = 0$ should be interpreted as the limit as $\xi \to 0$. This result provides asymptotic motivation for using the GEV as a parametric model for observed block maxima and thus a basis to estimate and extrapolate to high return levels. However, this assumes an underlying IID process, which is unrealistic.

Now we relax the independence assumption, so that $Z_1, \ldots, Z_n$ is a stationary sequence with the same marginal distribution function F. This corresponds to a series whose variables may be mutually dependent, but whose statistical properties are homogeneous through time. The limiting distribution of the rescaled block maxima of a stationary process satisfying a long range dependence condition, which ensures events

long apart in time are near independent, is $G^{\theta}(z)$ with $G(z)$ as in equation (3.2.1) and $\theta \in (0, 1]$ the extremal index (Leadbetter et al., 1983). For an independent series $\theta = 1$, but the converse is not true.

When a process exhibits extremal dependence, groups of extreme events form above high thresholds; these groups are called clusters. We define different clusters as those separated by some number of non-extreme values (i.e., below a high threshold z), this is called the run length r. Within a cluster, extreme events are considered as dependent whilst exceedances in different clusters are assumed to be independent. The extremal index tells us about clusters since it can be estimated empirically as the reciprocal of the mean cluster size of exceedances of z. This is known as the runs method (Smith and Weissman, 1994). Clusters are identified using an arbitrary choice of run length; this is the main pitfall with the approach. Ferro and Segers (2003) propose the intervals estimator, based on the limiting distribution of normalised times between exceedances of z. This distribution is exponential for independent random variables, otherwise it is a mixture distribution of an exponential with mean θ^{-1} and a degenerate probability distribution at zero, with probabilities θ and $1 - \theta$, respectively. Ferro and Segers (2003) also propose an automatic declustering scheme using the intervals estimate. Both methods only estimate θ in the observed range of z. In practice, the runs and intervals estimators of θ are actually estimators of the subasymptotic extremal index, for threshold level z and run length r, defined by Ledford and Tawn (2003) as

$$\theta(z, r) = \Pr(\max\{Z_2, \ldots, Z_r\} < z | Z_1 > z). \tag{3.2.2}$$

Then the extremal index is the limit of expression (3.2.2) as $z \to z^F$ and $r \to \infty$, with z and r tending to their respective limits in a related fashion.

We can also define extremes as exceedances of a high threshold u. If expression

(3.2.1) holds, then for an arbitrary term Z in the sequence $Z_1, \ldots, Z_n$,

$$\Pr(Z > b_n + a_n z \mid Z > b_n + a_n u) \to H_u(z) \quad \text{where} \quad H_u(z) = \left[1 + \xi\left(\frac{z-u}{\sigma_u}\right)\right]_+^{-1/\xi},$$

for $z > u$ as $n \to \infty$, with a_n, b_n as previously and $(\sigma_u, \xi) \in \mathbb{R}_+ \times \mathbb{R}$ the scale and shape parameters, respectively (Coles, 2001). The shape parameter is the same as that for the GEV. Whereas the scale is threshold dependent since $\sigma_u = \sigma + \xi(u - \mu)$ for μ and σ the GEV parameters. This is the generalised Pareto distribution (GPD). If $Z_1, \ldots, Z_n$ are IID, then exceedances of a high threshold u are IID and have limiting GPD tail model

$$\Pr(Z > z) = \lambda_u \left[1 + \xi\left(\frac{z-u}{\sigma_u}\right)\right]_+^{-1/\xi}, \tag{3.2.3}$$

for $z > u$ where $\lambda_u = \Pr(Z > u)$. Again, if $Z_1, \ldots, Z_n$ are stationary, a common approach is to identify clusters and decluster them to yield an approximately independent sequence of cluster maxima for which the GPD remains a valid model (Fawcett and Walshaw, 2007).

For non-stationary processes, it is common to allow the parameters of a stationary statistical model to vary with time or another covariate. In a block maxima framework, observations in a block are assumed to be IID so covariates in the GEV parameters cannot change within a block. In contrast, the GPD allows the covariates to vary uncontrolled over consecutive observations. A range of methods can be adopted to incorporate covariates in the model parameters λ_u, σ_u, ξ, including via harmonics (Coles et al., 1994), splines (Jonathan et al., 2014) or generalised additive models (Chavez-Demoulin and Davison, 2005).

3.2.2 Existing methodology

The earliest methods to estimate sea level return levels fit a GEV to the annual maxima (Graff, 1978; Coles and Tawn, 1990) or the annual r-largest observed sea lev-

els (Tawn, 1988b). These direct approaches ignore the known tidal component even though it induces non-stationarity into the sea level series, so that the GEV limit may not be a good approximation. Dixon and Tawn (1999) show that these direct methods underestimate extreme sea levels, especially at longer return levels for tidally dominant sites.

Pugh and Vassie (1978) were the first to exploit the decomposition of sea levels into tide and surge for extreme sea level estimation. This method is known as the joint probabilities method (JPM), where the probability distribution of extreme sea levels is derived by convolution of the distributions of these two components. The JPM gives consideration to all surge values in the data regardless of when they occurred relative to high tide. This approach forms the basis of the subsequent methods, but has some restrictive and unrealistic assumptions. Since the empirical surge distribution is used in the JPM, return level estimates are constrained by the sum of the highest predicted tide and the highest observed surge. Hourly surge observations are assumed to be independent, this is unrealistic as surge exhibits strong temporal dependence (Tawn and Vassie, 1989). Pugh and Vassie (1978) do account for dependence between tide and surge by dividing the tidal range into ordered bands of equal probability and estimating the surge distribution for each band, but this gives results which are sensitive to the choice of the number of bands and their boundary levels.

The revised joint probabilities method (RJPM) of Tawn (1992) attempts to address these limitations. For the upper tail of surges, an extreme value distribution is used to allow extrapolation beyond what has been observed, hence improving return level estimates. To account for temporal dependence, the extremal index is used. Tawn (1992) and Dixon et al. (1998) model the surge-tide dependence by allowing parameters of the GEV to be functions of the tidal level. This is a difficult task because the relationship is complex (Prandle and Wolf, 1978). Dixon and Tawn (1994) and Haigh et al. (2010) find that the RJPM is the best performing of these methods for observed

and simulated data.

Batstone et al. (2013) propose the skew surge joint probabilities method (SSJPM) to overcome the requirement of modelling the dependence of surge on tide. Their approach assumes that skew surge and peak tide are independent. Williams et al. (2016) demonstrate this is a good approximation both physically and empirically, for many sites in Europe and the USA. The SSJPM fits a GPD to the upper tail of skew surges and the empirical distribution is used for the main body of the distribution. The extremal index is used to measure dependence for sea levels in adjacent tidal peaks. They find $\theta \approx 1$ for all sites, suggesting no evidence of dependence in the upper tail. Baranes et al. (2020) adapt the SSJPM to account for interannual variations in the tidal regime, considering summer and winter separately and assuming peak tides and skew surge are stationary within each season.

All the methods discussed so far estimate extreme sea levels at a single site, but this gives uncertain return level estimates for sites with shorter observation periods. Spatial pooling can improve estimates at sites with limited or no data. Bernardara et al. (2011) use regional frequency analysis to estimate extreme surges. This involves grouping statistically similar sites into homogeneous regions, then fitting an extreme value model with a constant shape parameter over all sites (Hosking and Wallis, 1997). More recent approaches account for uncertainty in the region selection (Asadi et al., 2018; Rohrbeck and Tawn, 2021). Batstone et al. (2013) use hindcast sea level data to interpolate SSJPM estimates along the UK coast. Other methods allow parameters of the extreme value distribution to vary with spatial covariates. Coles and Tawn (1990) do this, allowing the location parameter of the GEV for sea level annual maxima to depend on the harmonic tidal constituents. Whereas, Dixon et al. (1998) spatially smooth parameters of the extreme value model for surges used in the RJPM, using predictable tidal variations along the UK coastline.

3.3 Exploratory analysis

3.3.1 Data

We use data from the UK National Tide Gauge Network obtained from the British Oceanographic Data Centre; these undergo rigorous quality control before release. Sea level elevations are recorded at 44 sites along the UK coastline. We consider skew surge and peak tide observations, with one record every 12 hours and 26 minutes (i.e., one tidal cycle), at four sites: Heysham, Lowestoft, Newlyn and Sheerness. Heysham is located on the west coast of England and has records from 1964-2016, with 17% missing. Lowestoft is on the east coast of England, with data available from 1964-2020 (4% missing). Sheerness is at the Thames Estuary, also on the east coast, with data available from 1980-2016 and 9% missing. Newlyn is located on the south coast of England and has records from 1915-2020, with 17% missing. The data are in metres above chart datum.

To compare the relative importance of skew surge and peak tide across sites, we define a surge-tide index as the observed range of skew surges divided by the range of peak tides. Heysham and Newlyn are tidally dominant with indices 0.65 and 0.70, respectively. Lowestoft and Sheerness are surge dominant with indices 3.36 and 1.66, respectively. We choose to study these sites because they are typically affected by different storms and all have a long observational duration. Heysham and Lowestoft are of particular interest because of their widely different surge-tide indices. Newlyn has the longest study period in the tide gauge network. Howard and Williams (2021) use physical model simulations to conclude that skew surge is dependent on peak tide at Sheerness, so we also study this site. The highest astronomical tide (HAT) observed is 10.72m, 2.92m, 6.10m and 6.26m for Heysham, Lowestoft, Newlyn and Sheerness, respectively. The data were preprocessed in an attempt to remove the linear non-stationary effect of sea level rise on skew surges caused by climate change and isostatic

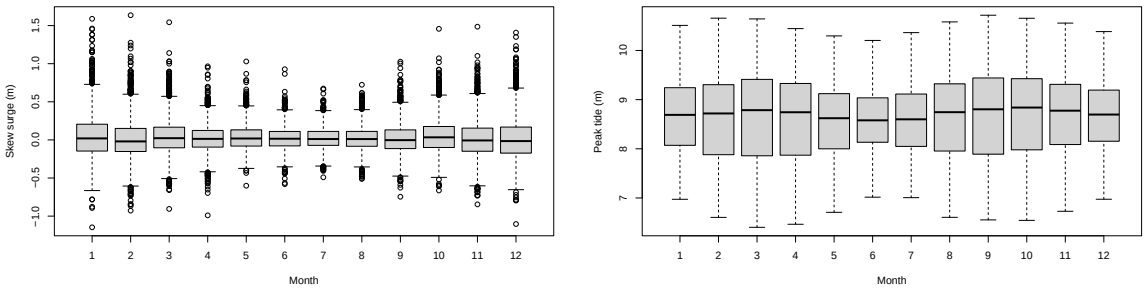


Figure 3.3.1: Monthly boxplots of skew surge (left) and peak tide (right) at Heysham.

rebound, see Environment Agency (2018) for details. Therefore all results are presented relative to the mean sea level in 2017.

3.3.2 Seasonality and temporal dependence of skew surge and peak tide

Figures 3.3.1 and A.9.1 show monthly boxplots of skew surge and peak tide observations at Heysham and the remaining sites, respectively. The median of the monthly skew surges remain relatively constant across months at Heysham, Newlyn and Sheerness but there is some variation at Lowestoft. The skew surge range varies across the year, exhibiting clear seasonality. Extreme skew surges and extreme peak tides typically occur in winter and at the equinoxes, respectively. Tides also exhibit interannual variability, including the 18.6 year lunar nodal cycle and the 8.85 year cycle of lunar perigee (Pugh and Woodworth, 2014).

To further assess seasonality, we estimate the probability that a randomly selected peak tide X is from month j where $j = 1 - 12$, given it is higher than some value x_q, the qth quantile of the distribution of peak tides, with $q \in [0, 1]$. This is defined as

$$\tilde{P}_X(j; x_q) = \hat{\Pr}(m(X) = j \mid X > x_q),$$

where $m(X)$ denotes the month of occurrence of X and $\hat{\Pr}(\cdot)$ is calculated empirically. If

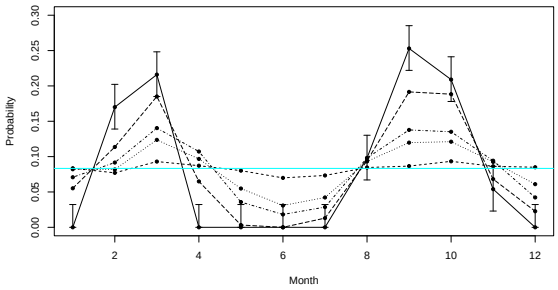


Figure 3.3.2: Estimates $\tilde{P}_X(j; x_q)$ for $j = 1 - 12$ and $q = 0.5$ (dashed), 0.9 (dotted), 0.95 (dot-dashed), 0.99 (long-dashed), 0.999 (solid) at Heysham, with 95% confidence intervals for $q = 0.999$. The horizontal cyan line indicates a probability of 1/12 for all months.

peak tides are identically distributed over the year and all months have equal duration, then $\tilde{P}_X(j; x_q) = 1/12$ for all j and q. Of course, months vary in length, but a significant departure from $\tilde{P}_X(j; x_q) = 1/12$ indicates that peak tides are not identically distributed over a year. Months with the largest peak tides will have higher values for $\tilde{P}_X(j; x_q)$ for large q.

Figure 3.3.2 shows the estimates $\tilde{P}_X(j; x_q)$ for a range of q at Heysham (see Figure A.9.2 for the remaining sites). These show that $\tilde{P}_X(j; x_{0.5}) \approx 1/12$ for all $j = 1-12$ at Heysham, Newlyn and Sheerness. At Lowestoft this varies, agreeing with the boxplot in Figure A.9.2. For all $q \geq 0.9$, there is clear evidence that $\tilde{P}_X(j; x_q)$ is largest in months nearest the equinoxes at all sites. At Heysham, $\tilde{P}_X(j; x_{0.999}) > 0$ for $j = 2, 3, 8, 9, 10$; we find the same results at Newlyn but $\tilde{P}_X(1; x_{0.999}) > 0$ also. However, at Lowestoft and Sheerness, $\tilde{P}_X(j; x_{0.999}) = 0$ in months close to the spring equinox. Confidence intervals on these estimates are constructed by exploiting the property that the number of times that an event $X > x_q$ occurs in month j follows a multinomial distribution, with probabilities $\tilde{P}_X(j; x_q)$. Figures 3.3.2 and A.9.2 also show 95% confidence intervals for $\tilde{P}_X(j; x_{0.999})$ at each site. These indicate that the differences in $\tilde{P}_X(j; x_{0.999})$ discussed above for each month are statistically significant.

Skew surge and peak tide both exhibit temporal dependence; we are only interested

in modelling the former since peak tides are deterministic. Figure A.9.5 shows the autocorrelation function (acf) plots for skew surge at each site. The correlation is stronger at lower lags, as expected, and tends to zero at higher lags; although this doesn't always reach zero due to seasonality. At Heysham and Newlyn, the correlation is significantly higher for all lags above 1, than at Lowestoft and Sheerness. In Section A.4 of Appendix A we explore temporal dependence in the skew surges further by looking at the two mean measures of extremal dependence χ and $\bar{\chi}$ (Coles et al., 1999) at different high thresholds and lags. We find similar patterns over lags, but weaker dependence, for the extreme values.

3.3.3 Skew surge and peak tide dependence

Prandle and Wolf (1978) examine the tide-surge interaction, both empirically and physically, for sites in the southern North Sea. They find that the most extreme surges occur on rising tides. This complex dependence structure motivates the use of skew surge and peak tide in the SSJPM (Batstone et al., 2013). Williams et al. (2016) conclude that the assumption of skew surge-peak tide independence is broadly well supported empirically and based on valid physical reasoning. However, Environment Agency (2018) and Williams et al. (2016) identify a weak correlation at Sheerness. Howard and Williams (2021) use oceanographic numerical model simulations to show that the highest skew surges tend to occur on lower peak tides at Sheerness. These findings raise questions about the validity of the independence assumption in the SSJPM. Ignoring the observed dependence would result in overestimates of return levels. We use a range of exploratory data analyses to investigate these claims in Section A.3 of Appendix A. We find that it is reasonable to assume skew surge-peak tide independence at Heysham, Lowestoft and Newlyn, but not at Sheerness. Our results suggest that this dependence structure is changing throughout the year, with the strongest dependence found in the summer months.

3.4 Novel methodology

3.4.1 Introduction

We begin by deriving the distribution of both the sea level monthly maxima M_j for $j = 1 - 12$ and annual maxima M under several simplifying assumptions that we subsequently relax. We use the fact that the peak sea level Z_i in tidal cycle i can be written as the sum of skew surge Y_i and peak tide X_i in that tidal cycle, for all $i = 1, \ldots, T$ where T is the total number of tidal cycles. In Section 3.4.2 we develop a simplified model, assuming skew surges are IID and independent of peak tide. We relax these assumptions to develop a novel method that reflects the realism of the sea level process; accounting for skew surge seasonality (Section 3.4.3), skew surge-peak tide dependence (Section 3.4.4) and temporal skew surge dependence (Section 3.4.5). Unless stated otherwise, inference is conducted in a likelihood framework. We provide 95% confidence intervals for model parameter estimates based on the Hessian; this assumes extreme skew surges are temporally independent and we have shown otherwise. Stationary block bootstrap confidence intervals are preferable, but we do not use these for model selection as they are computationally demanding. However, we use this bootstrap approach for the confidence intervals of return level estimates (Section 3.5.4) to ensure that estimates that are used in practice have optimal confidence intervals.

Extreme sea levels up to the ~ 20 year return period can occur with various combinations of skew surge and peak tides, e.g., a typical skew surge value combined with the largest peak tide, through to an extreme skew surge with a typical peak tide. Since peak tides are bounded above by HAT, return periods over 20 years can only be achieved with extreme skew surges. The very largest return periods require skew surges bigger than already observed. We are interested in return levels corresponding to return periods of 1 year and above, so we require an estimate for the distribution of all possible skew surges, though we are particularly interested in modelling the upper tail for inference

at high return periods. Thus in each variant of our method, we develop a model for the whole skew surge distribution under the associated assumptions about the skew surge process.

3.4.2 Idealised case

Under the assumption that the skew surges are identically distributed, we estimate the distribution of skew surges below a threshold u using the empirical distribution $\tilde{F}_Y$. This is adequate because tide gauges on the UK National Tide Gauge Network typically have long observation lengths (>20 years) so the empirical distribution describes the main body of the data well. To enable extrapolation in the tail we use the GPD model (3.2.3), with constant parameters. Then, our model is

$$F_Y(y) = \begin{cases} \tilde{F}_Y(y) & \text{if } y \leq u \\ \\ 1 - \lambda_u \left[1 + \xi\left(\frac{y-u}{\sigma_u}\right)\right]_+^{-1/\xi} & \text{if } y > u, \end{cases} \tag{3.4.1}$$

where $\lambda_u = 1 - \tilde{F}_Y(u)$ and σ_u, ξ are the parameters of the GPD. We take $\lambda_u = 0.05$. To simplify notation, we subsequently drop the u subscript on the scale σ and rate λ parameters.

Let T_j denote the number of tidal cycles in month j. We use sequential monthly peak tide samples $\{X_{j_i}; j = 1 - 12, i = 1, \ldots, T_j\}$ where j_i denotes the ith peak tide in month j. Then, if skew surges are independent and peak tides repeat exactly on an annual cycle, the distribution of month j maximum sea level, M_j, is

$$\Pr(M_j \leq z) = \prod_{i=1}^{T_j} \Pr(Y_i \leq z - X_{j_i}) = \prod_{i=1}^{T_j} F_Y(z - X_{j_i}). \tag{3.4.2}$$

Then the annual maxima skew surge distribution is given by,

$$\Pr(M \leq z) = \prod_{j=1}^{12} \prod_{i=1}^{T_j} F_Y(z - X_{j_i}). \tag{3.4.3}$$

Since peak tides form a deterministic sequence, exhibiting strong pairwise sample correlations, the peak tidal samples X_{j_i} that are used in expressions (3.4.2) and (3.4.3) must be from contiguous peak tides across the year.

Although cycles of periods up to a year dominate, peak tides have longer term periodicities. Previous methods, such as Tawn (1992), assumed peak tides are stationary within and across years so that the annual maxima distribution is given by

$$\Pr(M \leq z) = \left(\prod_{i=1}^{T} F_Y(z - X_i) \right)^{1/K}, \tag{3.4.4}$$

for T the total number of observations, K the number of years of observation and F_Y the stationary skew surge distribution (3.4.1). We incorporate interannual peak tide variations by taking the average of the yearly patterns over K years, where $K \geq 19$ to address all nodal cycle variations. We denote peak tide on the ith tidal cycle in month j of year k by $X_{j_i}^{(k)}$ for $j = 1 - 12,\ i = 1, \ldots, T_j^{(k)}$ and $k = 1, \ldots, K$. Here $T_j^{(k)}$ is the number of tidal cycles in month j and year k, which varies over k because the timing of the first cycle in the month can vary annually. Then we propose the following monthly and annual maxima sea level distributions

$$\Pr(M_j \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{i=1}^{T_j^{(k)}} F_Y(z - X_{j_i}^{(k)}), \text{ and } \Pr(M \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{j=1}^{12} \prod_{i=1}^{T_j^{(k)}} F_Y(z - X_{j_i}^{(k)}), \tag{3.4.5}$$

respectively. This no longer has the property that the distribution of the annual maxima is the product of the monthly maxima distribution, as the monthly maxima are now dependent due to associations in peak tides in different months across years.

3.4.3 Skew surge seasonality

As discussed in Section 3.3.2, skew surges are not identically distributed across a year. Here we describe how we capture seasonality in the skew surge distribution in three ways: for values above the threshold, below the threshold and in the exceedance probability. We define extreme skew surges using a month-specific threshold u_j, defined as a quantile of each monthly distribution. We use the 0.95 monthly empirical quantiles; in Section 3.6 we discuss the merits of this approach relative to taking a threshold that varies on a daily scale. We consider a range of models for how the parameters of the GPD and exceedance probability separately vary over time. These models are fit independently of each other and to different data. To distinguish between the two types of model, we precede the model number by S (for the scale and shape parameter) and R (for the rate parameter).

Firstly we look at exceedances of the monthly thresholds and model these using the GPD. There are enough data within each month, over the different years, to justify the adoption of a GPD model for exceedances of a high threshold and to assess its fit. To account for seasonality we allow the parameters to change over time in a periodic fashion, as discussed in Section 3.2.1. We consider four models to describe how these parameters change with time, expressed in terms of a daily or monthly covariate. As a first approach to account for seasonality in extremes, Carter and Challenor (1981) suggested allowing each month to have a separate distribution. To do so we fit a GPD with a monthly covariate on the scale σ_j and shape ξ_j parameters; we refer to this as Model $S0$. We use this basic approach for comparison only. To obtain a more parsimonious model, we fix the shape parameter to be the same value over all months but keep a monthly covariate on the scale parameter; we call this Model $S1$. These models give discontinuities at transitions between months, they assume skew surges are identically distributed within a month and have a large number of parameters. Therefore we introduce a covariate $d = 1 - 365$ that denotes the day in year to capture

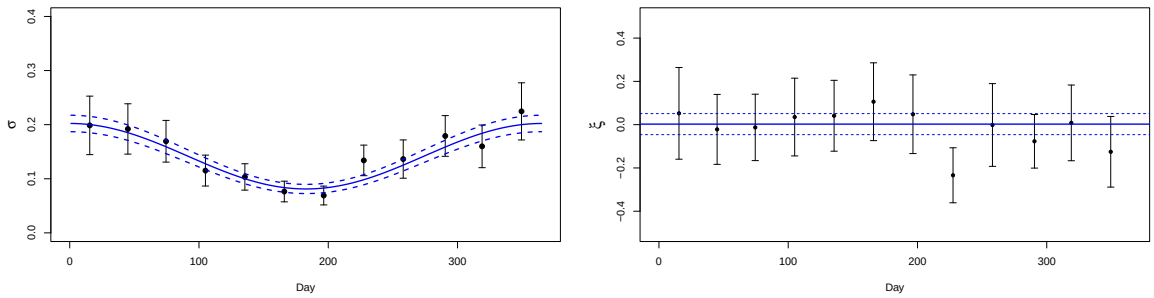


Figure 3.4.1: Scale (left) and shape (right) parameter estimates for Model $S0$ (black) and Model $S2$ (blue) at Heysham. 95% confidence intervals are added to Model $S0$ (error bars) and to Model $S2$ parameter estimates (dashed).

within-year variations smoothly. For notational simplicity, we assume d is entirely defined using the tidal cycle index i and month j, so that there exists a function h where $d = h(j, i)$. We consider two harmonic parameterisations of the scale parameter with a daily covariate. Model $S2$ uses a single harmonic defined by

$$\sigma_d = \alpha_\sigma + \beta_\sigma \sin\left(\frac{2\pi}{f}(d - \phi_\sigma)\right), \tag{3.4.6}$$

for parameters $\alpha_\sigma > \beta_\sigma > 0$, $\phi_\sigma \in [0, 365)$ and periodicity $f = 365$. Model $S3$ uses two harmonics with periodicities of f and $f/2$. The shape parameter is equal across months for both of these models, so that Models $S2$ and $S3$ have 4 and 6 parameters, respectively.

We now fit the models outlined above to skew surge threshold exceedances at each site. Figures 3.4.1 and A.9.4 show Model $S0$ parameter estimates for Heysham and the remaining sites, respectively. Since the 95% confidence intervals for monthly shape parameter estimates have considerable overlap at all sites, it is reasonable to have a common value across months as in Models $S1 - 3$. Table 3.4.1 reports the AIC and BIC scores for each model at all sites. These scores suggest that restricting the shape parameter to be an unknown constant, over months, improves model fit compared to Model $S0$ in all cases. Model $S2$ is selected as our final model as the evidence across all

Table 3.4.1: AIC and BIC scores at each site for each skew surge model relative to Model $S0$ scores, except Model $S4$ which is measured relative to Model $S2$.

		Heysham		Lowestoft		Newlyn		Sheerness	
Model	No. Parameters	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
$S1$	13	-66.66	-119.52	-57.40	-118.72	-77.36	-143.83	-67.51	-123.52
$S2$	4	-69.5	-176.61	-60.47	-171.93	-88.89	-209.74	-76.18	-178.03
$S3$	6	-65.65	-161.98	-56.53	-156.85	-86.06	-194.82	-75.06	-166.73
$S4$	5	2.99	8.42	1.82	7.39	5.77	11.82	-181.57	3.53

sites shows this is reasonable. Table 3.4.2 gives the parameter estimates for Model $S2$ at each site. Notice that the shape parameter estimates have 95% confidence intervals that encompass zero at each site. Figures 3.4.1 and S12 show the scale and shape parameter estimates graphically; $\hat{\sigma}_d$ is higher in the winter and lower in the summer, as expected. The estimated scale parameters of Model $S2$, when averaged over each month, do not differ significantly from the estimates of Model $S0$ but do capture a smooth transition within each month. We use asymptotic normality of maximum likelihood estimators and the Delta method to add 95% confidence intervals to the scale parameter estimates.

We also considered using a log link function for the scale parameter instead of the identity link function for Models $S2$ and $S3$ but found that the identity link yields a better model fit. Furthermore, the identity link function has the advantage of preserving the threshold stability property when covariates are included (Eastoe and Tawn, 2009).

Another fundamental part of capturing within year seasonality is through the rate parameter λ. So far, we have assumed this is constant as the threshold has been set at the same quantile across months. Here, we add a daily covariate d to capture smooth changes in λ throughout each month, whilst recognising that the average exceedance rate across a month is equal for all months. Let V_d be a binary random variable representing whether a skew surge value exceeds its month-specific threshold u_j or not, so that $V_d \sim \text{Bernoulli}(\lambda_d)$. Therefore, we use a logit link function $g(\cdot)$ to model λ_d and capture daily changes using a generalised linear model (GLM). To account for within

month variations, we relate $g(\lambda_d)$ to the day in month $d_j \in [1, 31]$, standardised by the monthly mean day $\bar{d}_j$, so that $(d_j - \bar{d}_j) \in [-15, 15]$ approximately. We parametrise the gradient using the day in year d to account for different gradients across the year using a harmonic with periodicity $f = 365$. We refer to this model as Model $R0$ and this is given by

$$g(\lambda_d) = g(\lambda) + (d_j - \bar{d}_j)\beta_\lambda \sin\left(\frac{2\pi}{f}(d - \phi_\lambda)\right), \tag{3.4.7}$$

for $\beta_\lambda > 0$, $\phi_\lambda \in [0, 365)$ which are parameters to be estimated and λ the exceedance rate in a month (here $\lambda = 0.05$). Fitting this model to our data demonstrates there is more variation in exceedance probabilities during spring and autumn compared to summer and winter. At all sites, the greatest range in λ_d was ~ 0.03 in April and October. This agrees with our scale parameter model where the steepest gradient is found in spring and autumn (see Figures 3.4.1 and A.9.4). At all sites, the fitted model for λ_d has a negative gradient in months at the beginning of the year so that the exceedance probability is higher earlier in the month (closer to winter). The slope is positive later in the year.

Finally, we use a month-specific empirical distribution $\tilde{F}_Y^{(j)}$ for skew surges below the threshold. Bringing this together with the parameterisations of σ_d and λ_d (expressions (3.4.6) and (3.4.7), respectively) for the upper tail model, the final full skew surge distribution F_Y is dependent on month j and day of the year d, and is given by

$$F_Y^{(d,j)}(y) = \begin{cases} \tilde{F}_Y^{(j)}(y) & \text{if } y \leq u_j \\ \\ 1 - \lambda_d\left[1 + \xi\left(\frac{y-u_j}{\sigma_d}\right)\right]_+^{-1/\xi} & \text{if } y > u_j. \end{cases} \tag{3.4.8}$$

As $F_Y^{(d,j)}(y)$ changes with day d and $d = h(j, i)$, it also changes for every tidal cycle i and month j. Consequently, the estimated monthly and annual maxima distributions

Table 3.4.2: Parameter estimates for the scale parameter for Models $S2$ and $S4$, and the rate parameter for Model $R1$ with 95% confidence intervals, at each site.

	Heysham	Lowestoft	Newlyn	Sheerness
Model $S2$				
α_σ	0.14 (0.13, 0.15)	0.15 (0.14, 0.16)	0.076 (0.073, 0.080)	0.11 (0.10, 0.12)
β_σ	0.060 (0.050, 0.070)	0.080 (0.070, 0.090)	0.024 (0.020, 0.028)	0.052 (0.043, 0.061)
ϕ_σ	271.51 (262.77, 280.23)	266.32 (260.01, 272.63)	273.58 (265.06, 282.10)	272.11 (262.66, 281.56)
ξ	0.002 (-0.042, 0.051)	0.024 (-0.023, 0.071)	-0.040 (-0.074, 0.006)	0.037 (-0.029, 0.10)
Model $S4$				
$\gamma_\sigma^{(x)}$	0.002 (-0.005, 0.009)	-0.0051 (-0.040, 0.030)	0.0048 (-0.00048, 0.010)	-0.012 (-0.026, 0.0011)
Model $R1$				
β_λ	0.0087 (0.0004, 0.017)	0.022 (0.015, 0.030)	0.024 (0.018, 0.030)	0.022 (0.014,0.032)
ϕ_λ	155.66 (100.74, 210.59)	175.16 (155.86, 194.46)	209.50 (195.48, 223.52)	184.31 (160.94,207.69)
$\alpha_\lambda^{(x)}$	-0.13 (-0.18, -0.079)	-0.055 (-0.101, 0.009)	-0.063 (-0.099, 0.108)	-0.32 (-0.37,-0.26)
$\beta_\lambda^{(x)}$	0.14 (0.068, 0.21)	-0.016 (-0.084, 0.051)	0.061 (0.014, 0.108)	0.23 (0.14, 0.31)
$\phi_\lambda^{(x)}$	311.86 (281.78, 341.94)	359.95 (265.77, 454.15)	352.38 (299.32, 405.44)	278.54 (260.44, 293.63)

of sea levels are

$$\Pr(M_j \leq z) = \frac{1}{K}\sum_{k=1}^{K}\prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j)}(z - X_{j_i}^{(k)}), \tag{3.4.9}$$

$$\Pr(M \leq z) = \frac{1}{K}\sum_{k=1}^{K}\prod_{j=1}^{12}\prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j)}(z - X_{j_i}^{(k)}). \tag{3.4.10}$$

3.4.4 Skew surge dependence on peak tide

In Section 3.3.3 we conclude that skew surge-peak tide independence is a reasonable assumption at Heysham, Lowestoft and Newlyn, but not at Sheerness. Here, we describe how we account for this dependence in our skew surge model. We do this for the upper tail by adding a tidal covariate to the GPD scale and rate parameters, but not to the shape parameter ξ to avoid additional uncertainty. For values below the threshold, we use different empirical distributions of skew surges depending on their associated peak tide band. This is a similar approach to Pugh and Vassie (1978).

We first consider how the threshold exceedance probability varies with peak tide. Here, we extend the GLM parametrisation of expression (3.4.7) by adding a peak tidal covariate x to the rate parameter λ_d. We linearly standardise peak tide via $(x - \bar{x})/s_x$, where $\bar{x}$ is the mean and s_x is the standard deviation of all peak tide observations at each site. We parametrise the gradient in the same way as expression (3.4.7), using the day in year d and a harmonic with periodicity $f = 365$ to capture smooth daily changes within a year. This captures the time varying dependence structure between skew surge and peak tide (see Section A.2 of the Appendix A). We use the notation $\lambda_{d,x}$ and the following model, denoted $R1$,

$$g(\lambda_{d,x}) = g(\lambda_d) + \left(\frac{x - \bar{x}}{s_x}\right)\left[\alpha_\lambda^{(x)} + \beta_\lambda^{(x)} \sin\left(\frac{2\pi}{f}(d - \phi_\lambda^{(x)})\right)\right], \tag{3.4.11}$$

for $g(\cdot)$ the logit link function, $g(\lambda_d)$ defined by expression (3.4.7) and $\alpha_\lambda^{(x)} \in \mathbb{R}$, $\beta_\lambda^{(x)} > 0$ and $\phi_\lambda^{(x)} \in [0, 365)$ being parameters to be estimated. The full model for $\lambda_{d,x}$ has 5 parameters.

We fit this GLM to model the exceedance probabilities at each site and give parameter estimates in Table 3.4.2. Figure 3.4.2 shows these results graphically at Sheerness for March, June, September and December (see Figure A.9.10 for the other sites). As expected the estimated exceedance probability is lower at higher tides, this result is more significant in months where we found skew surge-peak tide dependence to be stronger (April - September). As in Section 3.4.3, $\lambda_{d,x}$ changes most with day in spring and autumn corresponding to the greatest range of skew surges within a month. Figure A.9.9 shows the gradient of the tidal covariate for each month, when averaged over days. At Heysham, Newlyn and Sheerness, the gradients are greatest near the equinoxes; at Lowestoft there is little variation across months.

We compare the Model $R1$ (3.4.11) with Model $R0$ (3.4.7) using AIC and BIC scores. Model $R1$ minimises the AIC score at Heysham, Newlyn and Sheerness; this is most notable at Sheerness with a reduction of 139. We obtain almost identical AIC

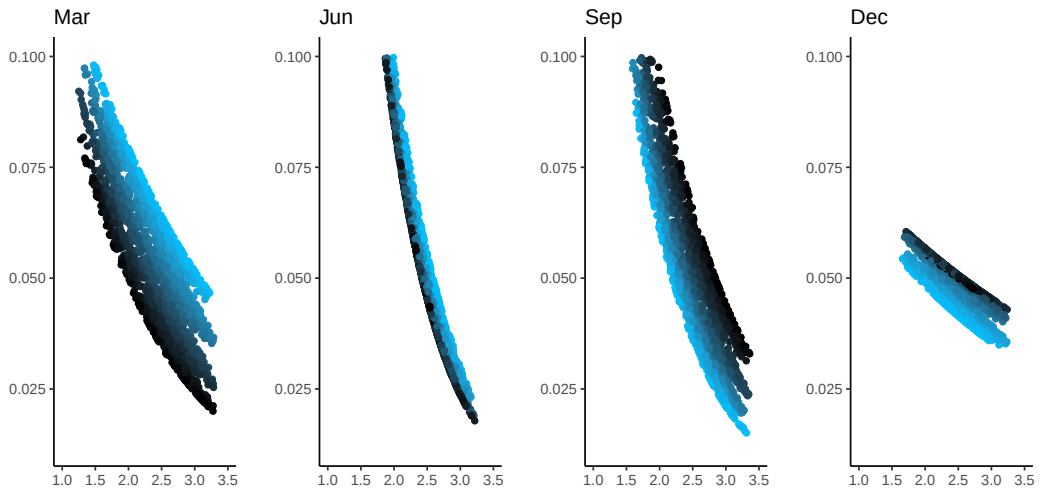


Figure 3.4.2: Estimated exceedance probability $\lambda_{d,x}$ (y-axis) under Model $R1$ in March, June, September and December with respect to x being peak tide (x-axis in metres) and d_j being day in month at Sheerness. Darker (lighter) points represent days later (earlier) in the month.

scores at Lowestoft. BIC is also minimised by Model $R1$ at Heysham and Sheerness, but not at Lowestoft and Newlyn. We also compare Model $R1$ with the model in expression (3.4.11) but with $\beta_\lambda^{(x)} = 0$, so that the gradient term for tide does not vary with day. We found that AIC favoured Model $R1$ at all sites indicating that the dependence structure varies seasonally.

We also investigate adding a tidal covariate x to the scale parameter of the GPD for extreme skew surges, building on the existing parametrisation of Model $S2$ (3.4.6) to give Model $S4$,

$$\sigma_{d,x} = \alpha_\sigma + \beta_\sigma \sin\left(\frac{2\pi}{f}(d - \phi_\sigma)\right) + \gamma_\sigma^{(x)} x, \tag{3.4.12}$$

for $\alpha_\sigma > \beta_\sigma > 0$, $\phi_\sigma \in [0, 365)$, $\gamma_\sigma^{(x)} \in \mathbb{R}$ parameters to be estimated, $f = 365$ the periodicity and d the day in year. Since we found evidence that the dependence of skew surge on peak tide changes within a year, we also considered parameterisations that allow the tidal effect to vary with day or month but found no improvement in fit relative to the loss of parsimony.

Table 3.4.2 gives estimates of $\hat{\gamma}_\sigma^{(x)}$ at each site; this tells us how $\sigma_{d,x}$ changes with tide

(see Table A.9.1 for all Model $S4$ parameter estimates). At Lowestoft and Sheerness $\hat{\gamma}_\sigma^{(x)} < 0$ which agrees with our results of Section 3.3.3 that extreme skew surges tend to occur on lower tides, although their 95% confidence intervals contain 0. To formally compare the fit of Model $S4$ with $S2$ we use AIC and BIC scores (see Table 3.4.1). Model $S4$ is only favoured by the AIC at Sheerness, this suggests the simpler model without a tidal covariate is preferred elsewhere. Given that our findings of skew surge-peak tide dependence were different at Sheerness in comparison to the other sites, we examined this further to check it wasn't an artefact of the data measurement or tide extraction processes. We explored the fits of Models $R1$ and $S4$ on a 483 year data set from a hydrodynamical model driven by a regional climate model (Howard and Williams, 2021). We find that adding peak tide covariates to the scale and rate parameters improves fit, with similar estimates to the observations. See Section A.3 of Appendix A for details. Therefore, we proceed with Model $S4$ for the scale parameter at all sites.

Below the threshold, we found it sufficient to capture the dependence by splitting the empirical distribution of skew surges into three associated peak tide bands, i.e.,

$$\tilde{F}_{j,x}(y) = \tilde{F}_j^{(1)}(y) \text{ if } x \leq x_{0.33}^{(j)}, \text{ or } \tilde{F}_j^{(2)}(y) \text{ if } x_{0.33}^{(j)} < x \leq x_{0.67}^{(j)}, \text{ or } \tilde{F}_j^{(3)}(y) \text{ if } x > x_{0.67}^{(j)}, \tag{3.4.13}$$

where $x_q^{(j)}$ is the q quantile of the peak tide distribution in month j and $\tilde{F}_j^{(l)}$ for $l = 1, 2, 3$ is the empirical distribution of skew surges in month j which are associated with the lowest ($l = 1$), medium ($l = 2$) and highest ($l = 3$) band of peak tides. The choice of 3 tidal bands is somewhat arbitrary, but appears sufficient given the weak dependence on peak tide. We could have used a kernel smoother to ensure continuity across bands, as in Bashtannyk and Hyndman (2001). Since our interest lies with the extreme values, we did not explore this further as this would not have made a practical difference to our return level estimates.

So our skew surge model, that is dependent on peak tide, is given by

$$F_Y^{(d,j,x)}(y) = \begin{cases} \tilde{F}_{j,x}(y) & \text{if } y \leq u_j \\ \\ 1 - \lambda_{d,x}\left[1 + \xi\left(\frac{y-u_j}{\sigma_{d,x}}\right)\right]_+^{-1/\xi} & \text{if } y > u_j, \end{cases} \tag{3.4.14}$$

with $\tilde{F}_{j,x}(\cdot)$, $\sigma_{d,x}$ and $\lambda_{d,x}$ defined in expressions (3.4.13), (3.4.12) and (3.4.11), respectively.

3.4.5 Skew surge temporal dependence

So far, we have assumed that skew surges are independent. We now describe how we account for their temporal dependence across tidal cycles. As discussed in Section 3.2.1, temporal dependence causes clusters of events above high thresholds; if this is ignored, sea level return levels of annual maxima will be overestimated. The best measure of dependence for the extreme values of a stationary sequence is the extremal index θ. Tawn (1992) and Batstone et al. (2013) use the extremal index of the highly non-stationary sea level series. We model the sub-asymptotic extremal index $\theta(y, r)$ (defined by equation (3.2.2)) of skew surges using varying thresholds (levels) y with a fixed run length r. Then the distribution of the monthly and annual maxima sea levels are given by

$$\begin{aligned} \Pr(M_j \leq z) &= \frac{1}{K}\sum_{k=1}^{K}\prod_{i=1}^{T_j^{(k)}}\left[F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})\right]^{\hat{\theta}(z-X_{j_i}^{(k)},r)}, \\ \Pr(M \leq z) &= \frac{1}{K}\sum_{k=1}^{K}\prod_{j=1}^{12}\prod_{i=1}^{T_j^{(k)}}\left[F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})\right]^{\hat{\theta}(z-X_{j_i}^{(k)},r)}. \end{aligned} \tag{3.4.15}$$

The empirical estimates $\tilde{\theta}(y, r)$ of the sub-asymptotic extremal index are shown in Figures 3.4.3 and A.9.8 for Heysham and the remaining sites, respectively. These

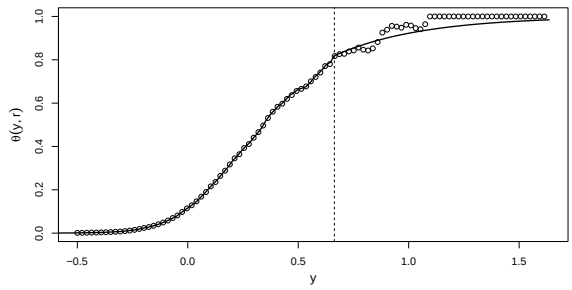


Figure 3.4.3: Estimates of the subasymptotic extremal index $\theta(y, r)$ for different skew surge levels y and run length $r = 2$ at Heysham using the runs estimate (points) and our model estimate (line) of expression (3.4.16). The dashed black line is the 0.99 skew surge quantile.

demonstrate substantial variation with y. We are interested in $\theta(y, r)$ across all ranges of skew surge, but particularly in values greater than those observed to allow extrapolation for return level estimation.

We estimate $\theta(y, r)$ in two stages. For $y \leq v$ where v is a high threshold, we use the empirical runs estimate $\tilde{\theta}(y, r)$ as this is smooth over y in this range. For computational efficiency purposes, we evaluate $\tilde{\theta}(y, r)$ on a regular grid of y values in the range $[y_F, v]$ where y_F is the minimum observed skew surge. Then we linearly interpolate these for any $y < v$ of interest. For $y > v$, the empirical estimate is increasingly variable so we adopt a parametric model for $\theta(y, r)$. We propose

$$\hat{\theta}(y, r) = \begin{cases} \tilde{\theta}(y, r) & \text{if } y \leq v \\ \theta - [\theta - \tilde{\theta}(v, r)] \exp\left[-\frac{(y-v)}{\psi}\right] & \text{if } y > v, \end{cases} \tag{3.4.16}$$

where $\psi > 0$ and $\tilde{\theta}(v, r) \leq \theta \leq 1$ are parameters to be estimated. This parametric form ensures the estimate asymptotes to the extremal index θ and is continuous at v.

The parameters ψ and θ are estimated using a weighted least squares approach with weight $w(y) = \sqrt{c(y, r) - 1}$ where $c(y, r)$ is the number of clusters above y separated by run length r. This gives a greater weight when there are more clusters, and a weight of

zero for a single cluster. We choose v to be the 0.99 quantile of all skew surges but this choice is subjective. The run length r represents the approximate duration of a storm across a single site. We take $r = 2, 10, 1$ and 10 for Heysham, Lowestoft, Newlyn and Sheerness, respectively. These choices are supported by estimating the run length using the intervals estimator for each season, where we expect the stationary assumption to be reasonably justified (Ferro and Segers, 2003). At Heysham, $\hat{\theta} = 1$ and $\hat{\psi} = 0.40$; the estimates for the other sites are given in Section A.4 of Appendix A. Figures 3.4.3 and A.9.8 show this model fit for Heysham and the remaining sites, respectively, compared with the entirely empirical estimates.

3.5 Results

3.5.1 Introduction

Using the distribution of the monthly and annual maxima sea levels derived in Section 3.4, we estimate return levels by solving

$$\Pr(M_j \leq z_{j,p}) = 1 - p \quad \text{and} \quad \Pr(M \leq z_p) = 1 - p,$$

respectively for $p \in [0, 1]$. We are interested in return levels up to the 10,000 year level, corresponding to annual exceedance probability $p = 10^{-4}$. In the monthly case, this is the level we expect the monthly maxima M_j to exceed, on average, every $1/p$ of that particular month (for example, every 10,000 Januarys). Whilst in the annual case, this is the level we expect the annual maxima to exceed every $1/p$ years, on average.

To assess the importance of each of the novel modelling steps in Section 3.4, we derive return level estimates from each stage; accounting for non-stationarity in each component, the dependence between them and skew surge temporal dependence. Each stage is detailed below in a nested list numbered (i)-(vii), so that each model below

builds on the previous method, except (iv) from (iii). We compare these to the current approach used in practice and a baseline approach, both derived under simplifying and false assumptions. We subsequently refer to each model by name, given below, using italics. The notation follows from Section 3.4.

(i) *Current*: The methodology currently used in practice, where skew surges are assumed to be IID and peak tides are stationary, with annual maxima distribution (3.4.4).

(ii) *Baseline*: As in (i) but we recognise interannual variations in the tide by averaging over yearly tidal samples. The annual maxima distribution is given by

$$\Pr(M \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{i=1}^{T^{(k)}} F_Y(z - X_i^{(k)}), \tag{3.5.1}$$

where $X_i^{(k)}$, $i = 1, \ldots, T^{(k)}$ represents an annual tidal sample for year $k = 1, \ldots, K$, where we choose $K = 100$ arbitrary but contiguous samples.

(iii) *Seasonal surge*: As in (ii) but we account for the within-year seasonality of skew surges using the skew surge distribution (3.4.8).

(iv) *Seasonal tide*: As in (ii), but conversely to (iii) we account for within-year seasonality in peak tide and not in skew surge, so the skew surge model of expression (3.4.1) is used in the monthly and annual maxima distributions (3.4.5).

(v) *Full seasonal*: As in (ii) but we account for within-year seasonality of both components, with monthly and annual maxima distributions (3.4.9) and (3.4.10), respectively.

(vi) *(Skew surge-peak tide) interaction*: As in (v) but with skew surge-peak tide dependence captured, so that the skew surge distribution (3.4.14) is used.

(vii) *(Skew surge) temporal dependence*: As in (vi) but accounting for temporal dependence in the skew surge series, with annual maxima distribution (3.4.15).

We compare our model estimates to empirical estimates. These are restricted to return periods within the range of data, but are useful for checking whether the model is capturing the distributional properties of observed sea levels. The empirical estimates act as a guide to the truth and we do not expect our model to fit these perfectly since each empirical estimate is specific to a particular annual tidal regime, whereas in our approaches (ii)-(vii) we account for tidal variations by averaging over different samples. The empirical estimates are also sensitive to missing data and so can be biased. In Sections 3.5.2 and 3.5.3, we make comparisons using point estimates to assess sensitivity to the model choice; in each case our results identify statistically significant differences between these models. Once the model choice is finalised we present measures of uncertainty and assess model fit in Section 3.5.4. In Section 3.5.5 we use our model to find the probability that an extreme sea level occurs in a specific month, given it is higher than some level with a given return period of interest.

3.5.2 Return levels: accounting for non-stationarity

We investigate how accounting for skew surge and peak tide seasonality influences sea level return level estimates by comparing monthly return level estimates from the *baseline*, *seasonal surge*, *seasonal tide* and *full seasonal* models. We are mainly interested in differences between the *baseline* and *full seasonal* estimates. We do not expect the *full seasonal* model to match the empirical estimates since we have not accounted for skew surge-peak tide dependence and temporal dependence yet. The intermediate solutions (*seasonal surge* and *seasonal tide*) allow us to understand which components' non-stationarity is influencing the return levels most in different months.

Figures 3.5.1 and A.9.11 show monthly return level estimates at Lowestoft and the other sites, respectively. At all sites and in all months, the empirical estimates lie

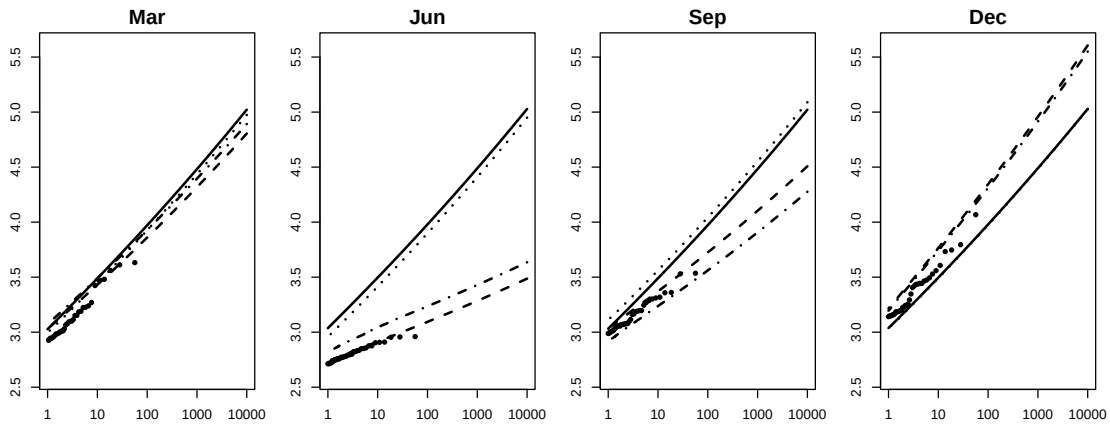


Figure 3.5.1: Monthly maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Lowestoft in March, June, September and December (from left to right), estimated using the *baseline* (black solid), *seasonal surge* (dot-dashed), *seasonal tide* (dotted) and *full seasonal* (dashed) models. Empirical estimates are shown by black points.

closer to *full seasonal* model than the *baseline*. The most noticeable feature is the difference between the *baseline* and *full seasonal* model in June, this reaches 1.54m at the 10,000 year level at Lowestoft. In June skew surges are particularly low relative to the rest of the year, so ignoring their seasonality leads to significant overestimation. This overestimation increases with return period. On the other hand, in December when skew surges are relatively high compared to the other months, ignoring seasonality results in underestimates of return levels. Since the tidal range at Lowestoft is narrow compared to the range of skew surges, differences between *seasonal tide* and *full seasonal* estimates are small relative to the differences between those from the *seasonal surge* and *full seasonal* models. For example, when the tidal range is largest at the autumn equinox in September, we observe a difference of 7cm at the 10,000 year level. Whereas at Heysham, where the tidal range is large relative to other sites, we observe a difference of 26cm at the 10,000 year level in September.

Figure 3.5.2 shows annual return level estimates from the *current*, *baseline* and *full sesaonal* models, compared to empirical estimates at all sites, up to the 100 year level (Figure A.9.12 shows these up to the 10,000 year level). The *current* method gives

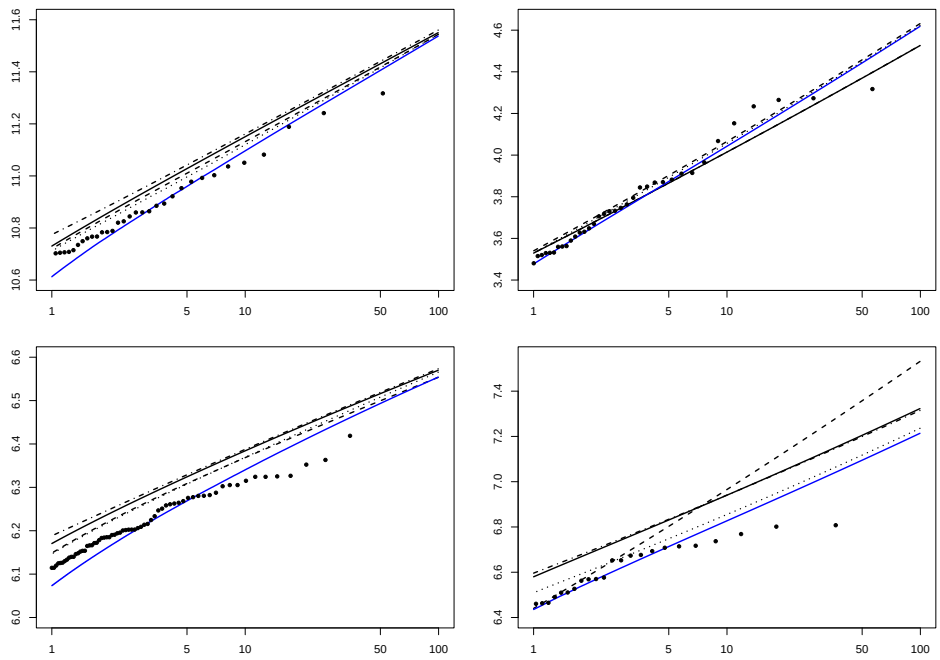


Figure 3.5.2: Annual maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right), up to the 100 year level, estimated using the *current* (dot-dashed), *baseline* (black solid), *full seasonal* (dashed), *interaction* (dotted) and *temporal dependence* (blue solid) models. Empirical estimates are shown by the black points.

similar results to the *baseline* approach for all return periods at all sites. This suggests assuming each year is identically distributed is not unreasonable. Even at Heysham, where year-to-year variations are large, the difference between the *baseline* and *current* method is small, with the *current* method giving a 1 year return level 4cm higher than the *baseline*. Since the *full seasonal* model lies closer to the empirical estimates than models (i)-(iv) at all sites, this highlights the importance of accounting for both forms of seasonality. At the 10,000 year level, the *baseline* gives a return level 6cm, 23cm and 58cm lower than the *full seasonal* method at Heysham, Lowestoft, and Sheerness, respectively, whereas it is 1cm higher at Newlyn.

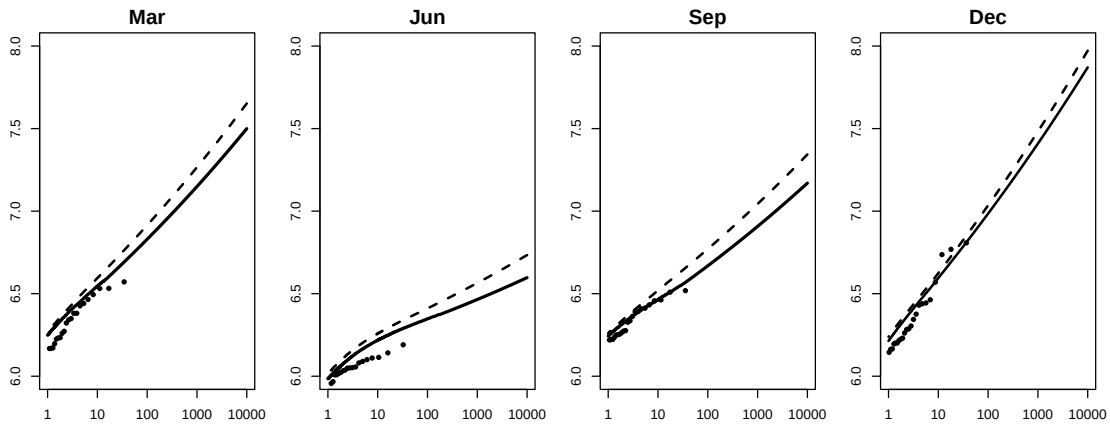


Figure 3.5.3: Monthly return level estimates (y-axis in metres) for different return periods (x-axis in years) at Sheerness in March, June, September and December, estimated using the *full seasonal* (dashed) and *interaction* (solid) models when we account for and ignore skew surge-peak tide dependence, respectively. Empirical estimates are shown by the black points.

3.5.3 Return levels: accounting for dependence

Here, we build on the *full seasonal* model to capture skew surge-peak tide dependence (*interaction* model) and skew surge temporal dependence (*temporal dependence* model). We compare these with the empirical and *full seasonal* estimates. We present monthly maxima return levels of the *interaction* model to demonstrate the changing skew surge-peak tide dependence structure within a year, discussed in Section 3.3.3. Then we compare annual maxima return level estimates from both models.

Figure 3.5.3 shows monthly return level estimates in March, June, September and December for the *full seasonal* and *interaction* models, as well as the empirical estimates at Sheerness. We illustrate the results for Sheerness because this dependence is stronger compared to the other sites, as discussed in Section 3.3.3. Ignoring skew surge-peak tide dependence overestimates return levels compared to the *full seasonal* model for return periods greater than 10 years. At the 10,000 year level, these are overestimated by 15cm, 14cm, 17cm and 10cm in March, June, September and December, respectively, with similar values down to the 1 year level. This is slightly surprising as in Section A.2 of Appendix A we found evidence that it is reasonable to assume skew surge and peak tide

are independent in March, September and December. For Sheerness the results from the *interaction* model tend to lie closer to the empirical estimates than those from the *full seasonal* model (see Figure 3.5.3); this suggests that accounting for this dependence is important and better reflects the process properties. At the other sites, the return level estimates from the *full seasonal* and *interaction* model are almost indistinguishable; this echoes that adding peak tide covariates into the skew surge model when independence is a reasonable assumption does not alter results significantly.

As skew surge temporal dependence is not yet accounted for in the *interaction* model, its return level estimates are anticipated to slightly overestimate the empiricals. Figure 3.5.2 and A.9.12 compares annual return level curves for the *temporal dependence* and *interaction* models. These are very close for high return periods (>100 years). For lower return periods (<10 years) the *temporal dependence* estimates lie closer to the empiricals at all sites. This is a natural consequence of the extremal index model given by expression (3.4.16) for temporal dependence, since the estimated extremal index is closer to zero for lower skew surges, corresponding to shorter return periods. But for the highest skew surges, and for those greater than observed, $\hat{\theta}(y, r) \to \hat{\theta} \approx 1$ and $y \to y^F$, so the largest return level estimates remain unchanged. Accounting for temporal skew surge dependence has the greatest influence at Heysham, where the acf estimates are higher compared to the other sites (see Figure A.9.5). Here, the *temporal dependence* model gives return values 10cm, 4cm and 2cm lower than the *interaction* model at the 1, 5 and 10 year return periods to better match the empirical estimates.

3.5.4 Assessment of fit and uncertainty

Figure 3.5.4 shows annual maxima return level estimates from the *temporal dependence* model, judged our best model, at Heysham and Lowestoft (see Figure A.9.13 for the other sites). We also add the estimated maximum and minimum year-specific return levels, showing the effect of interannual peak tide variations, as given by (3.4.15) with-

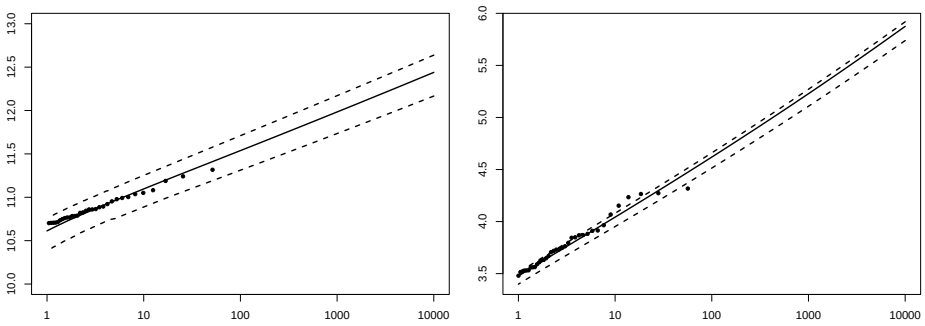


Figure 3.5.4: Return level estimates (y-axis in metres) for different return periods (x-axis in years) from the final model (*temporal dependence*) (solid line), with the maximum and minimum year-specific return level estimates (dashed line) and empirical estimates (points) at Heysham (left) and Lowestoft (right).

out averaging over K years. Figure A.9.14 shows yearly estimates at each site. The range of yearly estimates for Lowestoft is very narrow since the tidal range is small, compared to Heysham which has the largest tidal range of the four sites. Interestingly, year specific return levels at Newlyn appear to increase as the year also increases. Empirical estimates may deviate away from the return level point estimates if that specific value occurs in a year with particularly low or high tides compared to the average. Figures 3.5.4 and A.9.13 show that the empirical estimates lie within the bounds for each site, at most return periods. Enríquez et al. (2022) demonstrate that the 4.4 year perigean cycle has a greater effect on modulations in return levels than the nodal cycle in the UK. Our results are less clear (see Figure A.9.15), with Sheerness year-specific return levels having an ~ 20 year cycle.

We also assess goodness-of-fit by transforming the observed annual maxima to a uniform distribution using the probability integral transform with the distribution function of the annual maxima for their respective year. We do this for three cases; the *baseline* approach (3.5.1), our final model (3.4.15) and the year-specific final model. If the model fits well we expect the transformed annual maxima to be Uniform(0,1) distributed. Figure 3.5.5 shows PP plots for Sheerness, and the remaining sites are shown in Figure A.9.19; a good fit is indicated by the empirical and model probabilities being

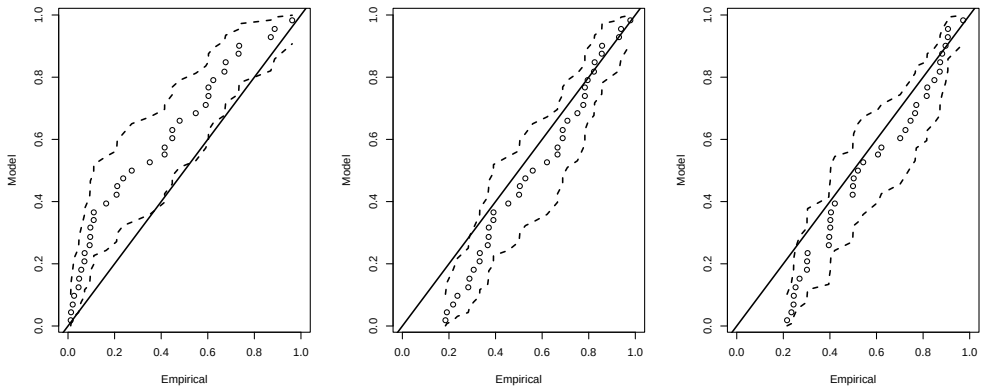


Figure 3.5.5: PP plots for the transformed annual maxima sea levels at Sheerness through the *baseline* (left), final (*temporal dependence*) (centre) and *year specific final* (right) distribution functions. The solid line shows the line of equality, $y = x$. 95% tolerance bounds are shown by dashed lines (found by bootstrapping).

equal so that the line of equality lies between the tolerance bounds. Figure 3.5.5 demonstrates the improvement in fit across the three models. We formally test this using the Kolmogorov-Smirnov test at each site for the final model and obtain p values 0.0066, 0.51, 0.0044 and 0.10 at Heysham, Lowestoft, Newlyn and Sheerness, respectively. At Heysham and Newlyn the p value is significant at the 5% level so we cannot conclude a good fit, but it is much better than for the baseline fits.

We also assess the model fit uncertainty by estimating 95% confidence intervals on the 1, 10, 100, 1000 and 10,000 year return levels using a stationary bootstrap procedure. We first transform the skew surges using our final model with their corresponding seasonal and tidal covariates, call this series $\{U_i^Y\}$ for $i = 1 \ldots, T$ for T the total number of observations, where $U_i^Y = F_Y^{(d,j,x)}(Y_i)$. This gives the basis for another approach to assess model fit. If the model is 'correct', these transformed observations will be Uniform(0,1) and we check this formally using the Kolmogorov-Smirnov test over all years, and for each year. We discuss this further in Section A.8 of Appendix A. To account for temporal dependence, we sample $\{U_i^Y\}$ using a stationary bootstrap with block length $L \sim$ Geometric(1/10), so the mean block length corresponds to 5 days (the maximum duration of a storm), and the total length of each sample is equal

to that of the original data (Politis and Romano, 1994). Call this bootstrapped series of uniform variables $\{U_i^B\}$. Then we transform the bootstrap sample back to the original scale using our final model, but preserving the original covariate information of $\{Y_i\}$, i.e., $\{X_{j_i}, d, j\}$, call this series $\{Y_i^B\}$. Then we fit our tail model to the $\{Y^B\}$ series to re-estimate all of the parameters $\alpha_\sigma, \beta_\sigma, \phi_\sigma, \gamma_\sigma^{(x)}, \xi, \beta_\lambda, \phi_\lambda, \alpha_\lambda^{(x)}, \beta_\lambda^{(x)}, \phi_\lambda^{(x)}, \theta$ and ψ, as well as the monthly empirical distributions and thresholds. These estimates are then used to derive the annual maxima distribution of sea levels and estimate return levels.

We take 200 bootstrap samples and present 95% confidence intervals in Figures 3.5.6 and A.9.18 at Sheerness and the other sites, respectively. The uncertainty associated with each return level increases with return period, as expected, since it becomes uncertain as we extrapolate. Uncertainty at the largest return periods can be attributed to the difficulty in accurately and precisely estimating the shape parameter of the GPD using data. This arises from using data at only a single site. It is a well-known limitation so a number of approaches have been developed to reduce the variation in estimates without introducing bias, with the most prominent being spatial pooling methods (see Section 3.6). We adopt a variant of this approach using the same strategy as Environment Agency (2018). They use a Normal$(0.0119, 0.0343^2)$ prior, derived from the sample of site specific shape parameter estimates for skew surges from over the entire UK network of coastal sites. We incorporate this prior information into our model, using a penalised likelihood framework for parameter estimation. Our shape parameter estimates are now -0.019 (-0.021, 0.059), 0.014 (-0.025, 0.052), -0.027 (-0.058, 0.004) and 0.008 (-0.039, 0.054) for Heysham, Lowestoft, Newlyn and Sheerness, respectively; 95% confidence intervals are given in parentheses. First note that this additional information is consistent with the data at each site since there is a close agreement in the point estimates of the shape parameters before and after including this information, with the point estimates all slightly closer to zero, whilst the scale and rate parameter estimates remain unchanged. More critically, we now have narrower confidence intervals for the

shape parameters. By incorporating this extra spatial information about skew surges, we have reduced the uncertainty associated with shape parameter estimation and hence in the return level estimates.

Figure 3.5.6 demonstrates how adding the penalty (prior) information reduces uncertainty at Sheerness, since the distribution of $\hat{\xi}$ over 200 bootstrap samples is much narrower. Figure 3.5.6 also shows the corresponding return level estimates and 95% bootstrap confidence intervals for this updated model; the return level curve is only affected at large return levels, as expected, where the estimates are less bounded. There is a dramatic reduction in the uncertainty associated with high return levels; at the 10,000 year level, adding prior information has reduced the range of confidence interval by 2.5m. Clearly, adding prior information on the shape parameter is important. However, we didn't do this sooner because it is important to allow the data to speak for itself when trying to reflect the realism of the sea level process in other aspects of our methodology. See Figure A.9.18 for the updated return level estimates and confidence intervals at Heysham, Lowestoft and Newlyn.

In Section A.5 of Appendix A we compare return level estimates for the sea level process from our final model across different quantiles used for the skew surge model to define extreme values. We find that the uncertainty associated with threshold selection is small compared to the other sources of uncertainty captured in the 95% confidence intervals, both with and without the prior information on the shape parameter.

So far, assessing sea level return level estimates derived across different statistical models has been restricted when comparing their agreement with empirical estimates, which are noisy, subject to bias from missing data and the tidal window over which they are observed, and the comparisons are limited to return periods up to the length of observation series. To aid with comparison to more reliable empirical estimates for short return periods and to enable comparison at longer return periods, we have also used the 483 year numerical model data set for Sheerness developed by Howard and

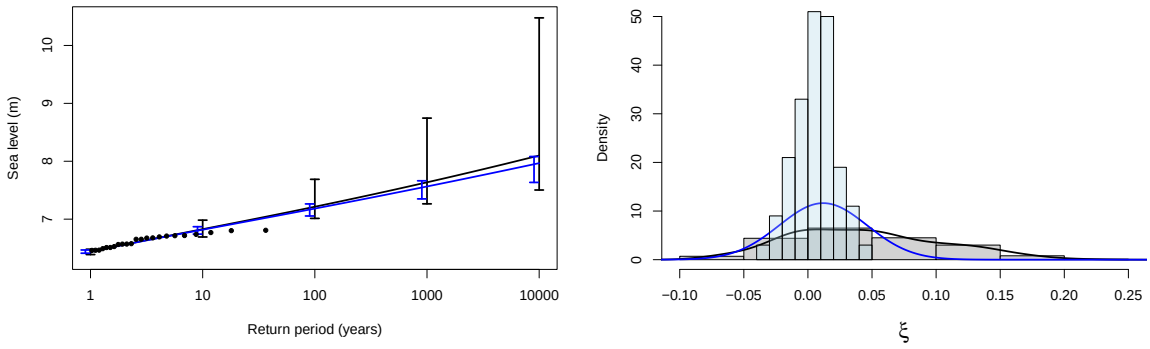


Figure 3.5.6: 95% bootstrap confidence intervals on the final (*temporal dependence*) return level estimates at Sheerness (left) before (black) and after (blue) adding a prior distribution to the shape parameter. Empirical estimates are shown by black points. Histograms of the shape parameter estimates and their densities (right) for these two models in their corresponding colours.

Williams (2021). Specifically, inference is based on repeated samples which are identical in length to the observed data at Sheerness and are assessed relative to the inference using the full 483 years of data. We find that our approach gives smaller RMSE values than the SSJPM for all return periods.

3.5.5 Sea level seasonality

It is helpful to understand when in the year the most extreme sea levels might occur, such as for coastal defence maintenance planning. We use our final model (*temporal dependence*) to investigate the seasonality of extreme sea levels. We are interested in the probability that a randomly selected sea level annual maxima M is from month j given it is equal to some level z. We consider z_p, a sea level with an associated annual exceedance probability $p \in [0, 1]$, so that z_p is derived from expression (3.4.15). Then the probability of interest is defined by,

$$\hat{P}_M(j; z) = \hat{\mathbb{P}}(m(M) = j | M = z), \tag{3.5.2}$$

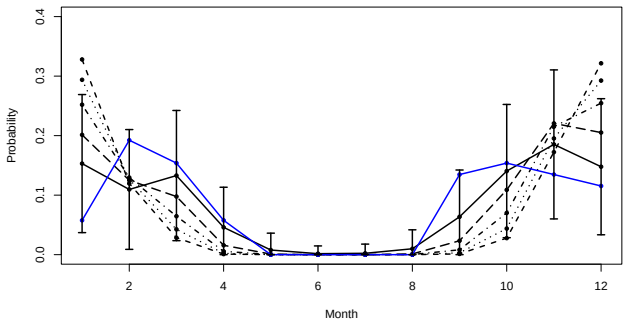


Figure 3.5.7: Estimates of $\hat{P}_M(j;z)$ for months $j = 1-12$ at Sheerness, for $p = 1$ (black solid), 0.1 (long-dashed), 0.01 (dot-dashed), 0.001 (dotted), and 0.0001 (dashed). The blue line is the empirical estimate $\tilde{P}_M(j;z_1)$, 95% confidence intervals are for $\hat{P}_M(j;z_1)$.

where $m(M)$ denotes the month of occurrence of M and $\hat{\mathbb{P}}(\cdot)$ is under our final model. This is a similar metric to that used for assessing peak tide seasonality (Section 3.3.2), however, since we have a model for the sea levels we condition on $M = z$ because we can obtain the form of the probability density function. Increasing z to rare return levels uncovers the seasonal variations in extreme sea levels. For months where the most extreme sea levels occur, $\hat{P}_M(j;z)$ takes its largest values, with $\hat{P}_M(j;z) \to 0$ for months with the least extreme sea levels. We derive the form of $\hat{P}_M(j;z)$ in Section A.7 of Appendix A.

Figure 3.5.7 shows the estimates $\hat{P}_M(j;z)$ to demonstrate sea level seasonality of extreme levels at Sheerness, as well as the empirical estimate for sea levels equal to the 1 year return level (see Figure A.9.20 for the remaining sites). These empirical estimates tend to lie within the 95% confidence intervals for the corresponding model estimates, at the 1 year return level. This demonstrates that our model for extreme sea levels is capturing the seasonality well. Our model estimates allow us to extrapolate beyond the observed data. At Sheerness, the empirical and model estimates at the 1 year return level are influenced by the tidal seasonality since they peak near the equinoxes. However, once we extrapolate to higher return levels, the sea level seasonality is almost entirely influenced by skew surge seasonality. This is the case for all sites at high return

levels, since tides are bounded above by HAT but this is less than the 10 year sea level return level for all four sites. Therefore, extremely large skew surges are required to exceed the highest return levels, which are most likely to occur in winter.

3.6 Discussion

We have developed a novel methodology for estimating extreme sea levels by accounting for seasonality in skew surge and peak tide, the dependence between them, and temporal dependence in skew surges. Our results show a significant improvement on current methods, which ignored these features of the sea level processes and instead made several simplifying assumptions. Our model also allows us to study the seasonality of sea levels exceeding levels previously unobserved; this can be useful for coastal defence maintenance planning. The return levels estimated from our model present a more accurate representation of future extreme levels, and should be used for future coastal defence upgrades. Our methodology can be applied to all UK National Tide Gauge Network sites with a sufficient data record length, and is not limited to UK locations. Our method could also be applied to estimate extreme sea currents distributions where a similar style framework for dealing with tide and surge components is used (Robinson and Tawn, 1997); this would provide substantial value for designing offshore structures to ensure required levels of safety. Seasonal variations are a common feature of many environmental variables and therefore should be accounted for when estimating return levels; the methodology discussed in Section 3.4.3 is applicable to a range of variables. We conclude by addressing some of the limitations with our model and suggest avenues for further exploration.

Our method assumes a steady state climate, since the existing mean sea level rise trend was removed before analysis, and we have not included any longer-term trends in our model. Note the mean sea level trend must be added back onto the return level

estimates presented in this paper when used in practice. It is fundamental that the effects of anthropogenic induced climate change and sea level rise are accounted for. Whilst most of this is captured by mean sea level rise, it has been suggested there is a small, but different, trend in the extreme sea levels (Haigh et al., 2010; Menéndez and Woodworth, 2010). There are two approaches to do this; the preprocessing model of Eastoe and Tawn (2009) where trends are modelled and removed before analysis, then covariates can be added to the extreme value model to explain trends in the tail of the distribution that are different from that in the body. Alternatively, covariates can be added to model parameters straight away without an attempt to remove any trends from the data (Davison and Smith, 1990). D'Arcy et al. (2022) (see Chapter 4) adopt this approach to study changes in extreme sea level relative to the much easier estimated changes in mean sea levels. They extend the methodology developed here to incorporate a global mean temperature (GMT) anomaly covariate. They find changes with GMT increases are primarily experienced through an increased frequency of extreme skew surges.

We have shown that skew surges clearly exhibit within-year seasonality. We used a monthly threshold, defined as a quantile of the monthly skew surge distribution, to define extreme values and develop our non-stationary model. Using a quantile ensures there are a similar number of exceedances to model per month. This approach is similar to that of Carter and Challenor (1981), where we first assume stationarity within months and then build in the seasonal variation on a shorter temporal resolution through covariates in the GPD parameters. This meant that we were able to capture most of the non-stationarity, as well as skew surge-peak tide dependence, at the same stage of the modelling process. However, we could have considered a smoother threshold choice by using quantile regression (Northrop et al., 2016) but we did not try this as our monthly threshold appeared sufficient.

Spatial pooling provides a promising framework to capture longer-term trends due

to climate change; single site trends are subtle but sharing information across sites could give more significant results. Spatial pooling also enables inference at locations with limited or no data, where our current singe site model would not perform well. In Section 3.5.4 we add a prior penalty for the shape parameter in our model, based on spatial information; this is a first approach to spatial pooling and drastically reduced the uncertainty associated with longer term return levels. Table A.9.1 shows that the 95% confidence intervals of $\hat{\xi}$ for Model $S4$ (no penalty) overlap at all sites, therefore it may be reasonable to fix the shape parameter across sites, as an alternative approach for borrowing information. Fixing the shape parameter in homogeneous regions is a crucial step in regional frequency analysis, originally introduced by Hosking and Wallis (1997). This is also an assumption in the Bayesian hierarchical modelling framework presented by Sharkey and Winter (2019) and a common assumption for spatial extreme value theory (see Davison et al. (2019)). We refer the reader to Batstone et al. (2013), Bernardara et al. (2011) and Haigh et al. (2010) for different approaches to spatial pooling for extreme sea level estimation. Whilst spatial pooling is clearly an important aspect of extreme sea level estimation, it is also fundamental that the marginal site estimates are accurate; therefore our approach for capturing the realism of the sea level process should be adopted within a spatial framework.

Chapter 4

Accounting for climate change in extreme sea level estimation

4.1 Introduction

The UK coastline is one of the largest in Europe at approximately 8000km for mainland Britain and is regularly subject to coastal flooding (Zsamboky et al., 2011). Coastal flooding is defined as a natural phenomenon where coastal land is inundated by sea levels above the normal tidal conditions. This has the potential to devastate coastal towns, damage infrastructure and destroy habitats. In extreme cases, coastal flooding has led to the loss of human life. The likelihood of coastal flooding is increasing with anthropogenic induced climate change (see Figure 4.1.1 and Seneviratne et al. (2012, 2021)). Therefore it is increasingly important to protect coastline communities, or at least have a well-founded scientific basis for the proposal for a managed retreat. Coastal flood defences, such as a sea wall, protect against these consequences if they are designed and built to withstand the most extreme sea levels. Estimates of sea level return levels provide crucial information for this design process; a return level is the value we expect the annual maximum sea level to exceed with probability p, i.e.,

once every $1/p$ years, on average, for a stationary series. We estimate these levels for $p \in [10^{-4}, 10^{-1}]$ to cover a range of industry interests, from agricultural preservation to nuclear power plant protection.

Coastal flooding is driven by a combination of tide, surge and waves. We are interested in the still water level, i.e., the sea level with waves filtered out, but for simplicity we refer to this as sea level. Therefore, tide and surge are the only components of sea levels that we consider. Tides are the regular and predictable changes in sea levels driven astronomically; these changes are well understood and perfectly forecast (Egbert and Ray, 2017). High tides generally occur once every 12 hours and 25 minutes, although variations are possible. We refer to the maximum tide in this cycle as the peak tide. Surges are stochastic, transient changes in sea levels often caused by strong winds and low atmospheric pressure due to a storm, hence are often referred to as storm surges. Surges are sometimes called the non-tidal residual as they define any departure from the predicted tidal regime so can also include gauge recording errors, tidal prediction errors and effects of the tide-surge interaction. These are often available at hourly or 15 minute intervals on the UK National Tide Gauge Network. We refer the reader to Pugh and Woodworth (2014) for a complete overview of sea level processes.

An alternative decomposition of sea levels is to consider the maximum level in a tidal cycle that can be written as the sum of skew surge and peak tide. Skew surge is the difference between the maximum observed sea level and the peak tide in a tidal cycle, regardless of their timing. In this case, we have less data since observations are available once every tidal cycle. However, skew surge and peak tide exhibit a much weaker dependence than surge and tide (which has a complex dependence structure), so they are often preferred. Williams et al. (2016) show that it is reasonable to assume skew surge and peak tide are independent at most sites on the UK National Tide Gauge Network; though there is physical evidence that this is not always true (Howard and Williams, 2021).

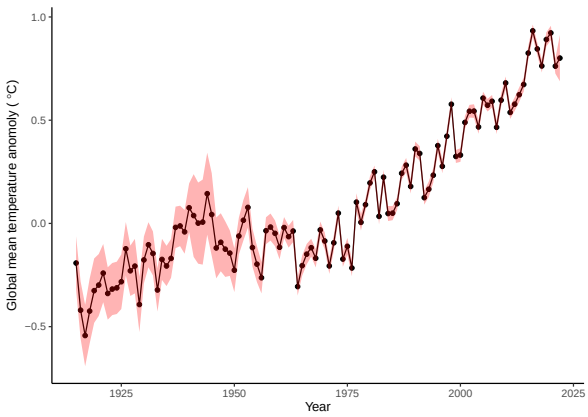


Figure 4.1.1: Global mean temperature anomalies from the HadCRUT5 dataset in 1915-2020, relative to the period 1961-1990, with associated uncertainties in red (Morice et al., 2021).

Long term changes in mean sea level have been widely studied (Woodworth and Player, 2003; Wahl et al., 2013) via empirical assessments and using hydrodynamic models linked to climate models. Typically linear models are fit to estimate these trends. Similar statistical methods have been used for extreme sea levels using regression of annual or monthly maxima data on either sea levels or skew surges. Interestingly these methods find no significant evidence for the trend in extreme sea levels to differ from that for mean sea levels (Calafat et al., 2022; Weiss and Bernardara, 2013; Wong et al., 2022; Woodworth et al., 2011). Complications to these methods are the large interannual variability, the presence of seasonality and the inefficient usage of extreme event data (through the use of maxima rather than all large values). The difference between extreme and mean sea level trends is likely to be of a smaller order than for mean sea level trends, hence they are more difficult to estimate. Furthermore, only trends in average extreme values are looked for, not changes in their variability over time. As a consequence, inference for these properties at a single site is likely to be overloaded by uncertainty, resulting in the hypothesis of identical trends in extreme and mean sea levels not being rejected.

We propose a different approach which is integrated into sea level return level estimation; this accounts for short term variations in skew surges such as seasonality, uses

all extreme skew surges above a high month specific quantile, allows for the distribution of the extreme skew surges, and enables pooling of information about the trend across sites. Critically, we separately assess changes in the rate of which extreme skew surge events occur and changes in the distribution (e.g., the mean) of these extreme events once they have occurred.

The earliest methods for estimation of sea level return levels modelled the sea level data directly, whilst the later approaches used a joint probabilities method to consider surge and tide components. More recent approaches use skew surge and peak tides. Section 4.2.3 gives an overview of the history of methodology developments. We extend the most recent method, that of D'Arcy et al. (2023b) (see Chapter 3), to account for the effects of climate change on extreme sea level estimation. They model skew surges and combine this with the known peak tide regime. D'Arcy et al. (2023b) particularly focus on the tail of the skew surge distribution, using a generalised Pareto distribution (GPD) to model exceedances of a high threshold (Coles, 2001). Covariates are added to this model representing day of the year and peak tide, to account for seasonality and skew surge-peak tide dependence, respectively, as well as capturing the temporal dependence of extreme skew surges. Their results demonstrate a considerable improvement on previous approaches since the realism of the sea level processes are captured, and significant improvements in goodness-of-fit are achieved. However, the model of D'Arcy et al. (2023b) assumed that skew surges were identically distributed across years, after a linear mean sea level trend was removed. If climate change impacts the within year skew surge variance, or even its distribution in a more subtle way than simply changing its mean value, then the extreme sea levels relative to the mean sea level will also change. Therefore we need a methodology that can incorporate such changes through to the return level estimation. Here we develop methods to account for non-stationarity in this mean adjusted skew surge data to help quantify any remaining non-stationarity in extreme skew surges.

In Sections 4.2.1 and 4.2.2 we introduce the data and relevant extreme value theory, respectively. Section 4.2.3 reviews the existing methods used for extreme sea level estimation, with a particular focus on that from D'Arcy et al. (2023b). In Section 4.2.4 we propose methods for investigating long term trends in extreme skew surges, with respect to time and global mean temperature anomaly (GMT), at a single site. We consider pooling information about the trends in the extreme value model across sites in Section 4.2.5, and suggest methods for exploring pairwise extremal dependence in skew surges across sites. We present the results for the single site model in Section 4.3.2 and estimate sea level return levels from the proposed model in Section 4.3.3. The results for the pooled method are given in Section 4.3.4. Section 4.4 concludes this paper with a summary of our findings and suggestions for future work.

4.2 Materials and methods

4.2.1 Data

Sea level observations are taken from the UK National Tide Gauge Network maintained by the Environment Agency. The data undergo rigorous quality control and can be obtained from the British Oceanographic Data Centre (BODC). This network is part of the National Tidal and Sea Level Facility where tidal elevations are recorded at 44 sites along the UK coastline. We consider data from Heysham, Lowestoft, Newlyn and Sheerness, located on the west, east, south and east (at the Thames Estuary) coast of England, respectively. Table 4.2.1 summarises information about each site. Each site has missing data, but the amount of complete data is sufficient given the model we introduce in Section 4.2.3 to account for smooth changes throughout the year.

We study these sites because they have different characteristics, are typically affected by different storms and all have long observational periods. Heysham has the second largest tidal range on the network and is tidally dominant, whereas Lowestoft is surge

Table 4.2.1: Location (latitude and longitude), observation period, percentage of missing data, highest astronomical tide (HAT) in metres and estimated mean sea level (MSL) trend in mm/yr for Heysham, Lowestoft, Newlyn and Sheerness.

	Location	Observation period	% missing	HAT (m)	MSL trend (mm/yr)
Heysham	54.03°N, 2.92°W	1964-2016	17	10.72	1.52
Lowestoft	2.47°N, 1.75°E	1964-2020	4	2.92	2.27
Newlyn	50.10°N, 5.54°W	1915-2016	17	6.10	1.73
Sheerness	51.45°N, 0.74°E	1980-2016	19	6.26	1.81

dominant. Sheerness is the only site we study where it is unreasonable to assume skew surge and peak tide are independent (D'Arcy et al., 2023b). A linear mean sea level trend was removed from the data at each site therefore all of our results are presented relative to the mean sea level in the year 2017. Environment Agency (2018) details this preprocessing stage and we report the estimated linear trend in Table 4.2.1 for each site. Of course, these trends incorporate land level changes as well as climate caused sea level changes, and also are based on different time periods as they correspond to the sample record at each site.

4.2.2 Extreme value inference

Within extreme value inference, it is natural to first consider modelling the maximum of a sequence $M_n = \max\{Z_1, \ldots, Z_n\}$. We first assume this sequence is independent and identically distributed (IID) with marginal distribution F and upper end point x^F. If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ so that the rescaled block maximum $(M_n - b_n)/a_n$ has a nondegenerate limiting distribution as $n \to \infty$, then the distribution function G of the maximum must be of the form

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{4.2.1}$$

where $x_+ = \max\{x, 0\}$ whatever the distribution function F. This distributional model G has three parameters $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ and $\xi \in \mathbb{R}$ representing the location, scale and shape, respectively (Coles, 2001). This is known as the generalised extreme value distribution (GEV). For $\xi > 0$ this corresponds to the Fréchet distribution, $\xi < 0$ the Weibull and $\xi = 0$ the Gumbel (although $\xi = 0$ should be interpreted as the limit as $\xi \to 0$). This result, often referred to as the *extremal types theorem*, gives an asymptotic justification to use the GEV as a model for block maxima, often taken to be annual maxima in environmental applications. However, in these settings, an IID assumption is usually unrealistic. A more commonly accepted assumption is stationarity, where the series can exhibit mutual dependence, but the statistical properties are homogeneous through time. If we now assume that $Z_1, \ldots, Z_n$ are from a stationary series that satisfies a long range dependence condition, so that events far enough apart in time are near independent. Under these conditions, this limiting distribution must be of the form $G^\theta(z)$ with $G(z)$ in expression (4.2.1) and $\theta \in (0, 1]$ the extremal index (Leadbetter et al., 1983).

If a process exhibits dependence, values above a high threshold z form clusters, for example during a storm that spans multiple days we might observe several extreme skew surge values consecutively. We identify clusters as those separated by a run length r defined as the number of consecutive observations below the high threshold z, i.e., *non-extreme* values. Choosing this run length can be subjective, though Ferro and Segers (2003) propose an automated selection procedure based on the distribution of all times between consecutive exceedances of z. We can reasonably assume that observations in different clusters are independent, but this is not the case for observations in the same cluster. The extremal index θ provides information about clusters because it can be empirically estimated (known as the runs estimate) as the reciprocal of the mean cluster size (Smith and Weissman, 1994). These are both actually estimates of the

subasymptotic extremal index

$$\theta(z, r) = \Pr(\max\{Z_2, \ldots, Z_r\} < z | Z_1 > z). \tag{4.2.2}$$

Then the extremal index is defined as the limit of expression (4.2.2) as $z \to z^F$ and $r \to \infty$ in a related fashion (Ledford and Tawn, 2003).

An alternative, and more popular, approach to defining extreme values is as exceedances of a high threshold u. If the extremal types theorem holds, then for an arbitrary term Z in the series $Z_1, \ldots, Z_n$,

$$\Pr(Z > b_n + a_n z \mid Z > a_n + b_n u) \to H_u(z) \quad \text{where} \quad H_u(z) = \left[1 + \xi\left(\frac{z-u}{\sigma_u}\right)\right]_+^{-1/\xi},$$

for $z > u$ as $n \to \infty$, with $\{a_n > 0\}$ and $\{b_n\}$ sequences of constants and H_u is non-degenerate. This is known as the generalised Pareto distribution (GPD) and has two parameters $\sigma_u \in \mathbb{R}_+$ and $\xi \in \mathbb{R}$ representing the scale and shape, respectively (Coles, 2001). Notice the scale parameter is threshold dependent since $\sigma_u = \sigma + \xi(u - \mu)$ for μ and σ the GEV parameters; the shape parameter is the same as that for the GEV. Assuming $Z_1, \ldots, Z_n$ are IID, exceedances of a high threshold u are also IID and have limiting GPD tail model

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

for $z > u$ where $\lambda_u = \Pr(Z > u)$. We can write the mean of excesses of the threshold u as

$$\mathbb{E}(Z - u | Z > u) = \frac{\sigma_u}{1 - \xi}. \tag{4.2.3}$$

However, if $Z_1, \ldots, Z_n$ are a dependent stationary series, a common approach is to identify clusters and decluster them (e.g., by considering cluster maxima only) to yield an approximately independent sequence so that the asymptotic justification for the GPD

remains valid (Smith et al. (1997), Fawcett and Walshaw (2007)). We subsequently drop the u subscript on the scale σ and rate λ parameters.

4.2.3 Existing methodology

The earliest methods directly modelled sea levels, but this ignores the known tidal component (Graff, 1978; Coles et al., 1999; Tawn, 1988b). Dixon and Tawn (1999) demonstrate that these approaches underestimate return levels. Pugh and Vassie (1978) were the first to exploit the components of sea levels in their joint probabilities method (JPM) using surge and tide. Tawn (1992) presents the revised joint probabilities method (RJPM) to address limitations associated with the JPM; mainly, they use an extreme value distribution to model the upper tail of surges to allow extrapolation beyond the range of observed values and, through a parametric model, attempt to account for tide-surge dependence. The main pitfall with both of these approaches is that surge and tide have a complex joint distribution which is difficult to model effectively. Batstone et al. (2013) proposed the skew surge joint probabilities method (SSJPM) to avoid this complexity. This uses skew surge and peak tide as two components of sea levels, since they have a much weaker dependence and can be reasonably assumed to be independent at most sites on the UK National Tide Gauge Network (Williams et al., 2016). Baranes et al. (2020) build on this by accounting for interannual tidal variations and considering separate distributions for summer and winter skew surges; this is the quasi non-stationary skew surge joint probabilities method (qn-SSJPM).

We build on the sea level model presented by D'Arcy et al. (2023b) that uses skew surge and peak tide as two components of sea levels in a joint probabilities framework. This was the first approach to capture within year seasonality of each component and the dependence between them by adding covariates to the model parameters. They also account for skew surge temporal dependence which addresses previous issues of overestimation at short return periods. We describe their model for the annual maxima

sea level M. For a tidal cycle i, the peak sea level Z_i can be written as the sum of the deterministic peak tide X_i and stochastic skew surge Y_i. We first present their skew surge model, and then describe how this is combined with the known peak tides to derive a sea level distribution. Lastly, we detail their model for the extremal index of skew surges used to derive the annual maxima distribution.

Since extreme sea levels can occur with various combinations of skew surge and peak tide, it is important to have a model for the entire skew surge distribution. To split the distribution into the body and tail, D'Arcy et al. (2023b) use a monthly threshold u_j for $j = 1, \ldots, 12$ to account for seasonality, with u_j being a quantile, for a fixed percentile, of month j's skew surge distribution. This choice ensures a similar number of exceedances for each month. They use the 0.95 quantile, this is chosen based on monthly parameter stability plots (Coles, 2001). Skew surges below these thresholds are modelled using the monthly empirical distribution $\tilde{F}_{j,x}$ to capture within year non-stationarity, that is also dependent on peak tide x to account for skew surge-peak tide dependence. This empirical distribution is split into three associated peak tide bands:

$$\tilde{F}_{j,x}(y) = \begin{cases} \tilde{F}_j^{(1)}(y) & \text{if } x \leq x_{0.33}^{(j)} \\ \tilde{F}_j^{(2)}(y) & \text{if } x_{0.33}^{(j)} < x \leq x_{0.67}^{(j)} \\ \tilde{F}_j^{(3)}(y) & \text{if } x > x_{0.67}^{(j)}, \end{cases}$$

where $x_q^{(j)}$ denotes the q quantile of the peak tide distribution for month j and $\tilde{F}_j^{(l)}$ for $l = 1, 2, 3$ is the empirical distribution of skew surges in month j which are associated with the lowest ($l = 1$), medium ($l = 2$) and highest ($l = 3$) band of peak tides. Since tide gauges on the UK National Tide Gauge Network usually have long observational periods, this can reliably model the main body of the data. For exceedances of the

monthly threshold u_j, they use a non-stationary GPD dependent on day in year $d = 1, \ldots, 365$, month j and peak tide x. Therefore, the full skew surge model is given by

$$F_Y^{(d,j,x)}(y) = \begin{cases} \tilde{F}_{j,x}(y) & \text{if } y \leq u_j \\ \\ 1 - \lambda_{d,x}\left[1 + \xi\left(\frac{y-u_j}{\sigma_{d,x}}\right)\right]_+^{-1/\xi} & \text{if } y > u_j, \end{cases} \tag{4.2.4}$$

where $\lambda_{d,x}$, $\sigma_{d,x}$ and ξ are parametric functions to be estimated. Notice that the shape parameter ξ does not vary with any covariate; this is kept fixed to avoid introducing additional uncertainty associated with estimating this parameter. The rate and scale parameters both depend on day and peak tide. They model the scale parameter using a harmonic for seasonal variations and a linear trend in terms of tide,

$$\sigma_{d,x} = \alpha_\sigma + \beta_\sigma \sin\left(\frac{2\pi}{f}(d - \phi_\sigma)\right) + \gamma_\sigma x, \tag{4.2.5}$$

for $\alpha_\sigma > \beta_\sigma > 0$, $\phi_\sigma \in [0, 365)$, $\gamma_\sigma \in \mathbb{R}$ parameters to be estimated and $f = 365$ the periodicity. The rate parameter is modelled similarly, using a generalised linear model with logit link function and a harmonic to capture seasonal variations. They also use a harmonic to capture how skew surge-peak tide dependence changes with time; D'Arcy et al. (2023b) show that this relationship varies throughout the year at Sheerness, with the strongest dependence occurring in May. This parameterisation is given by

$$\begin{aligned} g(\lambda_{d,x}) = g(\lambda) &+ (d_j - \bar{d}_j)\beta_\lambda^{(d)} \sin\left(\frac{2\pi}{f}(d - \phi_\lambda^{(d)})\right) \\ &+ \left(\frac{x - \bar{x}}{s_x}\right)\left[\alpha_\lambda^{(x)} + \beta_\lambda^{(x)} \sin\left(\frac{2\pi}{f}(d - \phi_\lambda^{(x)})\right)\right], \end{aligned} \tag{4.2.6}$$

for $g(\cdot)$ the logit link function (selected to help our modelling of probabilities with linear models), λ the constant exceedance probability in a month, $d_j \in [1, 31]$ the day in month (standardised by the monthly mean day $\bar{d}_j$), $\bar{x}$ is the mean and s_x the standard

deviation of peak tides, and $\alpha_\lambda^{(x)} \in \mathbb{R}$, $\beta_\lambda^{(d)}, \beta_\lambda^{(x)} > 0$, $\phi_\lambda^{(d)}, \phi_\lambda^{(x)} \in [0, 365)$ are parameters to be estimated.

To derive a distribution for the sea levels, D'Arcy et al. (2023b) use a joint probabilities method and the fact that peak tides are deterministic. So that

$$\Pr(Z_i \leq z) = \Pr(X_i + Y_i \leq z) = \Pr(Y_i \leq z - X_i) = F_Y(z - X_i), \text{ for } -\infty < z < \infty.$$

Let $T_j^{(k)}$ denote the number of tidal cycles in month j and year k. They capture within and across year peak tide non-stationarity by using sequential monthly and yearly peak tide samples $X_{j_i}^{(k)}$, so that j_i denotes the ith peak tide in month j, where $i = 1, \ldots, T_j^{(k)}$ and $k = 1, \ldots, K$ represents the year. Since peak tides are temporally dependent, the samples $\{X_{j_i}^{(k)}\}$ are from contiguous peak tides. Then the distribution of the annual maxima sea level M is

$$\Pr(M \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{j=1}^{12} \prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})^{\theta(z - X_{j_i}^{(k)}, r)}, \tag{4.2.7}$$

where $F_Y^{(d,j,x)}$ is the skew surge model (4.2.4) and $\theta(z - X_{j_i}^{(k)}, r)$ is a model for the extremal index, dependent on skew surge level $y = z - X_{j_i}^{(k)}$ and run length r, to capture temporal dependence of skew surges. This model is given by

$$\hat{\theta}(y, r) = \begin{cases} \tilde{\theta}(y, r) & \text{if } y \leq v \\ \theta - [\theta - \tilde{\theta}(v, r)] \exp\left(-\frac{y - v}{\psi}\right) & \text{if } y > v, \end{cases}$$

where v is a high threshold (D'Arcy et al. (2023b) take the 0.99 quantile), $\psi > 0$ and $\tilde{\theta}(v, r) \leq \theta \leq 1$ are parameters to be estimated and $\tilde{\theta}(y, r)$ is the empirical runs estimate. The run length reflects the approximate duration of a storm at each site, these were selected by estimating the run length using the intervals estimator of Ferro and Segers (2003) for each season, where we expect the stationary assumption to be

reasonably justified.

4.2.4 Incorporating interannual variations in the skew surge distribution

We provide a framework to explore long term trends in extreme skew surges that may result from an increase in storm frequency and intensity. After removing the mean sea level trend, we follow the approach of Eastoe and Tawn (2009) by adding yearly and global mean temperature anomaly (GMT) covariates into the scale and rate parameters to the GPD model for extreme skew surges of D'Arcy et al. (2023b). We do not consider adding covariates to the shape parameter or the empirical distribution used for non-extreme skew surges. Another option would be to add covariates to the threshold choice, but it is difficult to account for uncertainty in threshold selection in extreme value inference (Northrop et al., 2017; Wadsworth, 2016). Since we are interested in temporal changes of extreme events, it seems problematic to allow the threshold to also vary with time.

The model of D'Arcy et al. (2023b) already accounts for short term variations in the threshold exceedance rate and the GPD scale parameter. So here we are focusing on the additional long term changes in these two features, knowing that estimates of these are not contaminated by short term features. Trends in the two features tell us about different aspects of the occurrence of extreme skew surge events. An increase in the threshold exceedance rate tells us simply that more extreme events are occurring over time or with GMT increases. In contrast, increases in the scale parameter (when $\xi \approx 0$) inform us that the nature of the extreme events are changing, in that their average size is getting larger. So it is of interest to explore both aspects. Our proposed models for both parameters build on those presented in D'Arcy et al. (2023b) using additional additive components in terms of year k and GMT anomaly in year k, denoted m_k and measured in $^{\circ}C$. GMT is a potential causal covariate for climate change effects, whilst

year is a non-causal covariate but may capture long term changes over time.

First, we consider a model for the threshold exceedance probability to understand how the frequency of extreme skew surges is changing in response to climate change. We refer to the model for $\lambda_{d,x}$ introduced by D'Arcy et al. (2023b) as $R0$, given by (4.2.6). We propose four model extensions of $R0$ to account for how the threshold exceedance rate also changes with k (Models $R1$ and $R2$) and with m_k (Models $R3$ and $R4$); with the odd numbered models having a single trend across the year and the even numbered models having a different linear trend per season, with seasons $\{\mathcal{S}_s,\ s = 1, 2, 3, 4\}$ denoting winter (December, January, February), spring (March, April, May), summer (June, July, August) and autumn (September, October, November), respectively. These models are parametrised as follows,

$$
\begin{aligned}
&\text{Model } R1: \quad g(\lambda_{d,x,\tilde{k}}) = g(\lambda_{d,x}) + \delta_{\lambda}^{(\tilde{k})}\tilde{k}, \qquad (4.2.8)\\
&\text{Model } R2: \quad g(\lambda_{d,x,\tilde{k}}) = g(\lambda_{d,x}) + \sum_{s=1}^{4} \delta_{\lambda,s}^{(\tilde{k})}\tilde{k}\mathbf{1}_{\{d\in\mathcal{S}_s\}},\\
&\text{Model } R3: \quad g(\lambda_{d,x,m_k}) = g(\lambda_{d,x}) + \delta_{\lambda}^{(m)} m_k,\\
&\text{Model } R4: \quad g(\lambda_{d,x,m_k}) = g(\lambda_{d,x}) + \sum_{s=1}^{4} \delta_{\lambda,s}^{(m)} m_k \mathbf{1}_{\{d\in\mathcal{S}_s\}}, \qquad (4.2.9)
\end{aligned}
$$

where $g(\cdot)$ is the logit link function, $\delta_{\lambda}^{(\tilde{k})}, \delta_{\lambda,s}^{(\tilde{k})}, \delta_{\lambda}^{(m)}, \delta_{\lambda,s}^{(m)} \in \mathbb{R}$ are parameters to be estimated and $\tilde{k} \in \mathbb{R}$ denotes the standardised year, defined as $\tilde{k} = (k - 1968)/53$ where k is the year of observation. This standardisation uses information for Newlyn since it has the longest observation period where 1968 is the midpoint and 53 is half of the range, but is used across sites so that parameter estimates are easily comparable. For our study period, the covariates take values $\tilde{k} \in [-1, 1]$ and $m_k \in (-0.56, 0.94)$. Recall GMT is an anomaly centred at the temperature in the period 1961-1990, so it has been somewhat standardised.

We consider these four models to investigate whether time or GMT is the best linear

predictor of extreme skew surge non-stationarity over our observation period, and to explore if the long term trends are non-stationary within a year, for example, if extreme skew surges are becoming more frequent in the winter but less so in summer. For Model $R1$, we are particularly interested in the change in threshold exceedance probability over the period 1920-2020 (100 years), this is given by $\Delta_\lambda^{(\tilde{k})} = \lambda_{d,x,b} - \lambda_{d,x,a}$, for $a = -0.91$ (1920) and $b = 1$ (2020). Similarly for Model $R3$, we define the change in exceedance probability with an increase in GMT of 1°C as $\Delta_\lambda^{(m)} = \lambda_{d,x,1} - \lambda_{d,x,0} = \lambda_{d,x,1} - \lambda_{d,x}$.

Next, we investigate how the GPD scale parameter changes with year and GMT to understand if the magnitude of extreme events is changing due to climate change. We extend the $\sigma_{d,x}$ parameterisation (4.2.5) of D'Arcy et al. (2023b) (call this Model $S0$) and consider four models to capture changes with year, GMT and season as we did for the threshold exceedance rate,

$$\begin{aligned}
&\text{Model } S1: && \sigma_{d,x,k} = \sigma_{d,x} + \delta_\sigma^{(\tilde{k})}\tilde{k}, && (4.2.10)\\
&\text{Model } S2: && \sigma_{d,x,k} = \sigma_{d,x} + \sum_{s=1}^{4} \delta_{\sigma,s}^{(\tilde{k})}\tilde{k}\mathbf{1}_{\{d\in\mathcal{S}_s\}}, &\\
&\text{Model } S3: && \sigma_{d,x,m_k} = \sigma_{d,x} + \delta_\sigma^{(m)} m_k, &\\
&\text{Model } S4: && \sigma_{d,x,m_k} = \sigma_{d,x} + \sum_{s=1}^{4} \delta_{\sigma,s}^{(m)} m_k \mathbf{1}_{\{d\in\mathcal{S}_s\}}, && (4.2.11)
\end{aligned}$$

with parameters $\delta_\sigma^{(\tilde{k})}, \delta_{\sigma,s}^{(\tilde{k})}, \delta_\sigma^{(m)}, \delta_{\sigma,s}^{(m)} \in \mathbb{R}$ to be estimated and $\tilde{k}$, m_k, $\mathcal{S}_s$ as in (4.2.8)-(4.2.9).

4.2.5 Spatial pooling

Improved Inference by Pooling

So far we have described the modelling of extreme skew surges at a single site. However, this approach can be very inefficient, particularly for sites with short records or where the physical processes exhibit similar characteristics over the sites, e.g., the same storm

events effect all of the different sites. In such cases, we anticipate certain parameters of the extreme surge skew distribution to be similar, or even identical, in value across sites. By imposing this feature into the inference and carrying out joint inferences across sites, known as pooling, this can lead to large improvements in parameter estimation, by effectively sharing information about extreme events across sites, which in turn reduces estimation uncertainty resulting in narrower confidence intervals.

In the model of D'Arcy et al. (2023b) the benefits of pooling were illustrated for the shape parameter. This parameter is known to be very difficult to estimate with much precision, with the variability in its estimator being the primary source of uncertainty in return level estimation. This parameter has been recognised across a wide spectrum of problems as being very similar for a given process over large spatial regions, e.g., for rainfall (Davison et al., 2012), sea levels (Dixon et al., 1998) and air temperature (Huser and Genton, 2016) with different values for the shape parameter for plains and mountains. D'Arcy et al. (2023b) use information from Environment Agency (2018) that the shape parameter estimates, estimated separately from each site over the UK, followed a normal distribution with mean 0.0119 and variance 0.0343^2. They account for this in the likelihood inference, using this distribution as a prior penalty function. D'Arcy et al. (2023b) obtained shape parameter estimates which were more similar over sites with much reduced uncertainty, thus resulting in uncertainty reduction of high return level estimates. For example, for the 10,000 year return level at Sheerness, the 95% confidence interval was reduced by 2.5m, corresponding to a factor of 6.

In our context, the difficult parameters to estimate are those of the long term trends in expressions (4.2.8)-(4.2.9) for the threshold exceedance rate and (4.2.10)-(4.2.11) for the GPD scale parameter. Here we also want to share spatial information through pooling. Given that we do not know if these trend parameters are identical over sites, and we only are illustrating the method at four sites, we undertake a formal likelihood testing method to assess the evidence to see if we can treat these trend parameters

as constant over sites, without reducing the quality of the fit relative to the improved parsimony.

The pooled inference procedure involves a combined likelihood function $L(\boldsymbol{\theta})$ which combines the likelihood functions $L_\ell(\boldsymbol{\theta}_\ell)$ from each of the $\ell = 1, \ldots, 4$ sites through

$$L(\boldsymbol{\theta}) = \prod_{\ell=1}^{4} L_\ell(\boldsymbol{\theta}_\ell),$$

where $\boldsymbol{\theta}_\ell$ are the parameters for site ℓ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_4)$. This likelihood enables hypothesis testing to be carried out, to assess the evidence for whether certain parameters are the same at all, or a subset of, the sites, i.e., is the time trend gradient parameter the same at all sites. The joint likelihood function then enables the sharing of information about this common parameter across sites whilst allowing the other parameters to vary over sites. The choice of this joint likelihood function has potential restrictions; since it is a product over sites, this implicitly implies that extreme skew surges are being assumed to be independent across the sites. In cases where this assumption is unreasonable, the point estimates of the parameters will still be good (asymptotically consistent) but the variance of the estimates and the confidence intervals for the parameters will be underestimated. The degree of underestimation is dependent on the level of ignored true dependence between skew surges at the different sites. Therefore before exploiting the pooling strategy it is important to check that the independence assumption, for the extreme values of skew surge, is a reasonable approximation.

Spatial Independence Diagnostics

We discuss how to check for pairwise dependence between skew surges at different sites. Kendall's τ correlation coefficient can be used to check for dependence skew surge observations; this is a measure of rank correlation so is robust to outliers but it is a measure across all values of the variables. However, since our interest lies with the

dependence of the extreme values, it is natural to also study the two main measures of extremal dependence χ and $\bar{\chi}$ (Coles et al., 1999), as described next.

Let Y^A and Y^B denote skew surge random variables at two different sites A and B, in the same tidal cycle with marginal distribution functions F_A and F_B respectively. The simplest measure of dependence is to see how the joint probability of Y^A and Y^B both being above their respective $(1-p)$th marginal quantiles, compares to p (the value of this probability under perfect dependence of Y^A and Y^B) and relative to p^2 (the value under independence of Y^A and Y^B). Under positive dependence we would expect that

$$p^2 < \Pr\{Y^A > F_A^{-1}(1-p), Y^B > F_B^{-1}(1-p)\} < p. \tag{4.2.12}$$

Coles et al. (1999) formalise this intuition to define the measure of extremal dependence as $p \to 0$, i.e., as we look above increasing quantiles. Specifically, they take

$$\chi = \lim_{p \to 0} \Pr\{Y^B > F_B^{-1}(1-p) \ \mid \ Y^A > F_A^{-1}(1-p)\},$$

where $\chi \in [0, 1]$. Increasing values of χ correspond to stronger extremal dependence, and $\chi = 1$ corresponds to perfect dependence between Y^A and Y^B. Thus χ is the limiting probability of one variable being extreme given that the other is equally extreme. If $\chi \in (0, 1]$, we say Y^A and Y^B are asymptotically dependent, with there being a non-zero probability of Y^B being large when Y^A is large at extreme levels. Though the class of extremal dependence where $\chi > 0$ is widely studied, this only corresponds to cases where the joint probability in (4.2.12) is of $O(p)$, i.e., decays as a multiple of p as $p \to 0$. We find that $\chi = 0$ in all other dependence cases as well as when Y^A and Y^B are actually independent, this class of variables is known as being asymptotically independent, and χ doesn't give us any information on the level of asymptotic independence. We need a more refined measure of extremal dependence than χ to enable us to separate between when there is some dependence of large values and independence of

Y^A and Y^B. Coles et al. (1999) also define

$$\bar{\chi} = \lim_{p \to 0} \frac{2 \log \Pr\{Y^A > F_A^{-1}(1-p)\}}{\log \Pr\{Y^B > F_B^{-1}(1-p), Y^A > F_A^{-1}(1-p)\}} - 1.$$

where $\bar{\chi} \in (-1, 1]$. Here $\bar{\chi} = 1$ and $-1 < \bar{\chi} < 1$ correspond to asymptotic dependence and asymptotic independence, respectively. When $\bar{\chi} = 0$ this shows there is no dependence in the tails of (Y^A, Y^B) as it arises when Y^A and Y^B are independent, with $0 < \bar{\chi} < 1$ and $-1 < \bar{\chi} < 0$ indicating positive and negative dependence in the joint tails of Y^A and Y^B, respectively.

To assess inter-site dependence in extreme skew surges we evaluate these dependence measures using empirical estimates of the associated probabilities using the `texmex` R package (Southworth et al., 2020). Specifically, we use skew surge daily maxima for each pairwise combination of the four study sites, using data on the same day and with lags of ± 1 day to account for time lags between the peak of surge reaching each site, when events last multiple days. Here we have lags $t = 1$ and $t = -1$ so that site A is one day ahead or behind site B, respectively. Since the variables are not identically distributed, due to seasonality for example, this can affect the evaluation of χ and $\bar{\chi}$. We address this potential concern by also using the marginal distributional model of D'Arcy et al. (2023b) $F_Y^{(d,j,k)}$, given by expression (4.2.4), to account for this through a transformation of the variables to identical uniform margins and then re-evaluate these measures. These results are discussed in Section 4.3.4.

4.3 Results

4.3.1 Introduction

We now present the results from applying the extreme skew surge models discussed in Sections 4.2.4 and 4.2.5, in Sections 4.3.2 and 4.3.4, respectively to data from our four

study sites. In Section 4.3.3, we estimate sea level return levels using the best fitting model from Section 4.2.4 under a single-site analysis using the annual maxima distribution in expression (4.2.7). Here we define extreme skew surges as being exceedances of the monthly 0.95 quantile, as in D'Arcy et al. (2023b). All models are fit in a likelihood framework, with 95% confidence intervals provided for parameter estimates based on the hessian, i.e., using asymptotic normality of maximum likelihood estimators. The likelihood is constructed under the assumption that extreme skew surges are temporally independent for single site inference, but also that observations at different sites are independent for spatial pooling. These are not unreasonable assumptions for model selection, the former being found as a reasonable approximation in D'Arcy et al. (2023b) as the extremal index is near one for large levels and the validity of the latter being assessed before we apply any spatial pooling. We compare models using Akaike and Bayesian information criteria (AIC and BIC, respectively) scores; these are commonly used measures of the quality of a statistical model for a particular data set relative to the parsimony of the model. The chosen best fitting model should minimise these scores. It should be noted that all estimates presented here are after the mean sea level trends have been removed. An estimated change here means that the change is additional to the mean sea level, so negative trend estimates correspond to the extreme sea levels not rising as fast as the mean level at the site.

4.3.2 Single-site analysis

We fit the models of Section 4.2.4 to the GPD rate and scale parameters for extreme skew surges individually at each site. We start with the threshold exceedance probability parameter, λ, fitting Models $R0-4$. AIC/BIC scores and estimates of $\delta_{\lambda}^{(\tilde{k})}$, $\delta_{\lambda,s}^{(\tilde{k})}$, $\delta_{\lambda,s}^{(m)}$ and $\delta_{\lambda}^{(m)}$ are given in Table 4.3.1. Since the parameter estimates are not intuitive, we consider the change in exceedance probability with the particular covariate of interest for the annual trends in Models $R1$ and $R3$.

Table 4.3.1: Parameter estimates for the Models $R0 - R4$ with AIC and BIC scores for each model fit at each site. The minimum AIC and BIC scores are highlighted in red and blue, respectively, for each site. The 95% confidence intervals are given in parentheses for parameter estimates.

	Heysham	Lowestoft	Newlyn	Sheerness
Model $R0$				
AIC	12234.21	15312.08	24498.77	9286.58
BIC	12275.89	15354.88	24543.93	9326.94
Model $R1$				
$\delta_{\lambda}^{(k)}$	0.091 (-0.008, 0.191)	-0.061 (-0.150, 0.028)	0.215 (0.154,0.276)	-0.114 (-0.219, -0.010)
AIC	12232.89	15312.26	24453.12	9283.96
BIC	12282.91	15363.61	24507.31	9332.40
Model $R2$				
$\delta_{\lambda,1}^{(k)}$	0.161 (-0.033, 0.335)	0.063 (-0.106, 0.232)	0.114 (-0.011, 0.238)	-0.032 (-0.228, 0.164)
$\delta_{\lambda,2}^{(k)}$	0.034 (-0.161, 0.230)	-0.141 (-0.322, 0.040)	0.197 (0.077, 0.316)	-0.250 (-0.468, -0.032)
$\delta_{\lambda,3}^{(k)}$	0.207 (0.013, 0.400)	-0.094 (-0.266, 0.078)	0.209 (0.089, 0.328)	-0.189 (-0.405, 0.026)
$\delta_{\lambda,4}^{(k)}$	-0.047 (-0.261, 0.167)	-0.081 (-0.264, 0.102)	0.338 (0.217, 0.460)	-0.021 (-0.221, 0.178)
AIC	12235.30	15315.29	24452.52	9286.48
BIC	12310.32	15392.33	24533.80	9359.14
Model $R3$				
$\delta_{\lambda}^{(m)}$	0.204 (0.074, 0.334)	-0.012 (-0.12, 0.427)	0.336 (0.245, 0.427)	-0.164 (-0.304, -0.024)
AIC	12227.14	15314.04	24451.34	9283.20
BIC	12277.16	15365.39	24505.53	9331.64
Model $R4$				
$\delta_{\lambda,1}^{(m)}$	0.256 (-0.002, 0.514)	0.103 (-0.107, 0.312)	0.135 (-0.058, 0.329)	-0.079 (-0.340, 0.181)
$\delta_{\lambda,2}^{(m)}$	0.111 (-0.143, 0.365)	-0.067 (-0.282, 0.149)	0.322 (0.144, 0.501)	-0.374 (-0.672, -0.076)
$\delta_{\lambda,3}^{(m)}$	0.416 (0.167, 0.665)	-0.048 (-0.259, 0.163)	0.393 (0.212, 0.574)	-0.273 (-0.568, 0.022)
$\delta_{\lambda,4}^{(m)}$	0.010 (-0.274, 0.293)	-0.040 (-0.262, 0.182)	0.478 (0.300, 0.655)	0.008 (-0.253, 0.269)
AIC	12227.92	15318.49	24450.27	9284.69
BIC	12302.94	15395.53	24531.56	9357.34

We find that Model $R3$ minimises AIC at Heysham, Lowestoft and Sheerness, whilst at Newlyn Model $R4$ is preferable. The BIC is minimised by Model $R3$ at Newlyn, but elsewhere Model $R0$ is favourable. This suggests that if any long term trends are included in the model to capture changes in the rate of extreme events (relative to mean sea level), GMT should be used as a covariate as opposed to the year.

We look at Models $R1$ and $R3$ in more detail, these have a fixed trend parameter within the year with respect to year and GMT, respectively. At Newlyn we find a significant increasing trend for both models, since the confidence intervals do not contain zero. We also find positive trends at Heysham, but only the GMT trend in Model $R3$ is

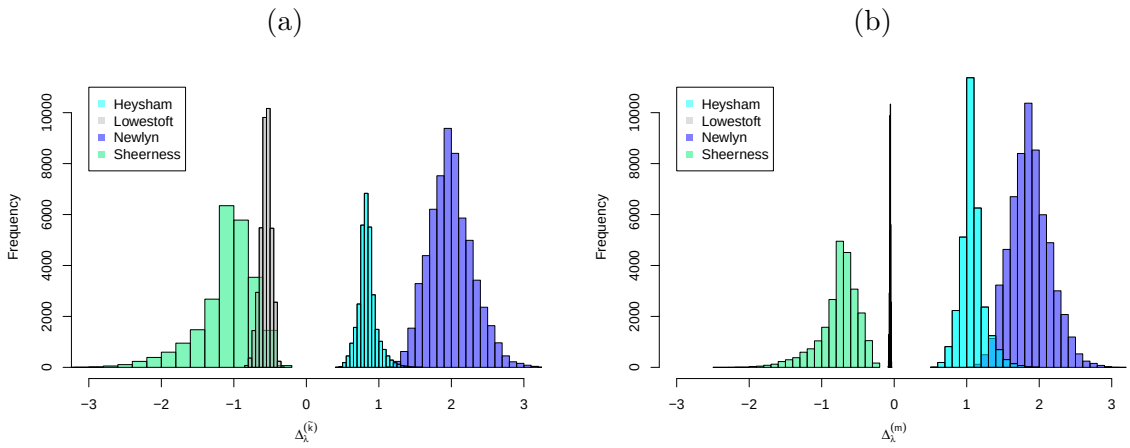


Figure 4.3.1: Histograms of (a) $\Delta_\lambda^{(\tilde{k})}$ over 100 years and (b) $\Delta_\lambda^{(m)}$ with a 1°C increase in GMT, as percentages, for all day d and peak tide x combinations at each site.

significant. Neither trend is significant at Lowestoft, but we find a significant decreasing trend for both models at Sheerness. Figure 4.3.1 shows histograms of the estimates of $\Delta_\lambda^{(\tilde{k})}$ and $\Delta_\lambda^{(m)}$ (defined in Section 4.2.4), based on all combinations of day d and peak tide x, so these do not account for uncertainty in $\delta_\lambda^{(\tilde{k})}$ or $\delta_\lambda^{(m)}$ estimates but are simply a reflection that the rate of threshold exceedance varies over the short term. For Model $R1$, we find an increase in $\lambda_{d,x,\tilde{k}}$ over 100 years at Newlyn, with $\max \Delta_\lambda^{(\tilde{k})} = 3\%$, so that the exceedance probability almost doubles from 3.5% to 6.5% in 1920-2020. However, we observe decreases in exceedance probability at Sheerness. For Model $R3$, we also find an increase in exceedance probability with a 1°C increase in GMT at Newlyn, where $\max \Delta_\lambda^{(m)} = 3\%$, but a negative trend at Sheerness. If the trends were statistically significant at Heysham and Lowestoft, the exceedance probability would increase and decrease with both trend parameters, respectively.

Next we look at Models $R2$ and $R4$ with season-specific trend parameters for year and GMT, respectively. The trends at Newlyn are significant in both models, except for winter, whilst at Heysham only the increasing trends in summer are significant. None of the seasonal trends are significant at Lowestoft but we find a significant decreasing trend for spring in both models at Sheerness. As for Models $R1$ and $R3$, we obtain a

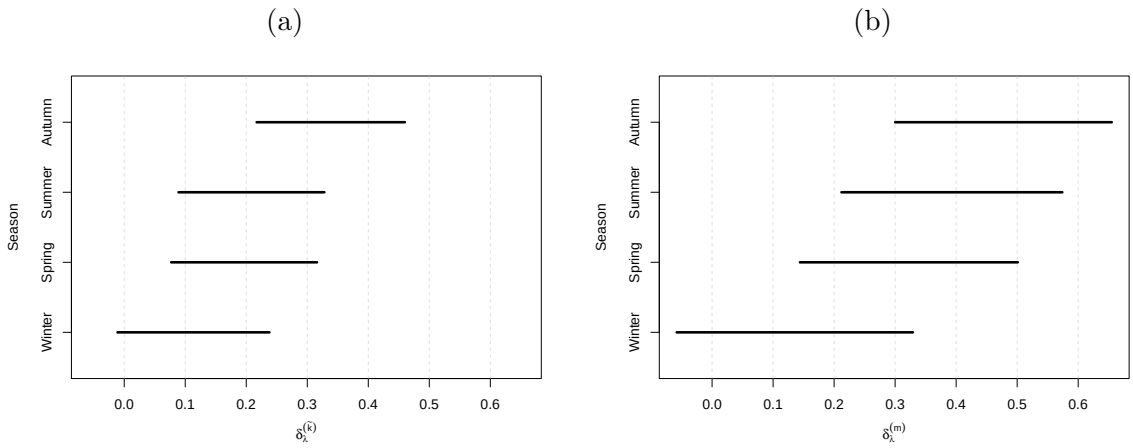


Figure 4.3.2: Confidence intervals for parameter estimates (a) $\hat{\delta}^{(\tilde{k})}_{\lambda,s}$ and (b) $\hat{\delta}^{(m)}_{\lambda,s}$ at Newlyn for $s = 1, 2, 3, 4$ denoting winter, spring, summer and autumn, respectively.

variety of results across sites; Newlyn has an increasing exceedance probability with year and GMT in all seasons. However, for Lowestoft and Sheerness we obtain a mixture of positive and negative parameters throughout the year for both models. The confidence intervals for the four parameter estimates in Models $R2$ and $R4$ at Heysham, Lowestoft and Sheerness overlap, suggesting that there isn't significant within-year variation of the long term trend parameters so that the simpler Models $R1$ and $R3$ are sufficient. Whilst at Newlyn, this overlap is small (see Figure 4.3.2). Here, we find the greatest trend in autumn, which is not too concerning for extreme sea level estimation since the most extreme sea levels tend to occur in winter (D'Arcy et al., 2023b), but using Models $R1$ and $R3$ with common trend parameters across the year could overestimate the trends in winter, hence influencing sea level return level estimation.

Next, we consider models for the scale parameter at each site individually (Models $S0 - 4$, introduced in Section 4.2.4). Table 4.3.2 shows the parameter estimates for each model, along with AIC and BIC scores. We do not report the corresponding shape parameter estimates here as there is no significant difference in these estimates across models. Models $S1$ and $S3$ have a single parameter denoting a common long term trend across the year, but neither of these is an improvement on Model $S0$ (without a long

Table 4.3.2: Parameter estimates for the Models $S0-4$ with AIC and BIC scores for each model fit at each site. The minimum AIC and BIC scores are highlighted in red and blue, respectively, for each site. The 95% confidence intervals are given in parentheses for parameter estimates.

	Heysham	Lowestoft	Newlyn	Sheerness
Model $S0$				
AIC	-3091.53	-3672.07	-10152.63	-2974.317
BIC	-3064.77	-3644.20	-10122.42	-2948.854
Model $S1$				
$\delta_\sigma^{(\tilde{k})}$	-0.009 (-0.032, 0.013)	-0.006 (-0.024, 0.011)	0.001 (-0.003, 0.005)	0.016 (-0.013, 0.044)
AIC	-3088.558	-3670.55	-10150.80	-2973.05
BIC	-3056.448	-3637.11	-10114.54	-2942.50
Model $S2$				
$\delta_1^{(\tilde{k})}$	0.022 (-0.034, 0.078)	-0.041 (-0.090, 0.007)	0.004 (-0.009, 0.016)	0.023 (-0.032, 0.078)
$\delta_2^{(\tilde{k})}$	0.022 (-0.014, 0.059)	-0.030 (-0.055, -0.006)	0.006 (-0.002, 0.014)	-0.001 (-0.036, 0.035)
$\delta_3^{(\tilde{k})}$	-0.025 (-0.051, 0.001)	0.012 (-0.010, 0.034)	-0.003 (-0.009, 0.003)	0.023 (-0.010, 0.055)
$\delta_4^{(\tilde{k})}$	-0.035 (-0.079, 0.008)	-0.015 (-0.053, 0.023)	0.001 (-0.008, 0.011)	0.008 (-0.039, 0.054)
AIC	-3095.28	-3674.12	-10146.27	-2971.19
BIC	-3047.12	-3623.96	-10091.89	-2925.36
Model $S3$				
$\delta_\sigma^{(m)}$	-0.011 (-0.033, 0.011)	-0.008 (-0.026, 0.009)	-0.001 (-0.008, 0.006)	0.006 (-0.020, 0.032)
AIC	-3088.42	-3670.90	-10149.07	-2972.43
BIC	-3056.32	-3637.46	-10112.82	-2941.87
Model $S4$				
$\delta_1^{(m)}$	0.036 (-0.027, 0.099)	-0.050 (-0.105, 0.004)	-0.0003 (-0.021, 0.020)	0.025 (-0.042, 0.091)
$\delta_2^{(m)}$	0.029 (-0.012, 0.070)	-0.037 (-0.061, -0.012)	0.005 (-0.008, 0.018)	-0.015 (-0.054, 0.023)
$\delta_3^{(m)}$	-0.027 (-0.054, -0.00009)	0.017 (-0.006, 0.039)	-0.003 (-0.013, 0.006)	0.013 (-0.018, 0.045)
$\delta_4^{(m)}$	-0.030 (-0.081, 0.021)	-0.024 (-0.066, 0.017)	-0.005 (-0.021, 0.010)	-0.006 (-0.053, 0.040)
AIC	-3093.73	-3677.95	-10145.39	-2970.79
BIC	-3045.56	-3627.79	-10091.01	-2924.96

term trend) at any site. All of the 95% confidence intervals for $\delta_\lambda^{(\tilde{k})}$ or $\delta_\lambda^{(m)}$ estimates contain zero, suggesting these trends are not significant. If we ignore this uncertainty, the point estimates suggest small changes in the scale parameter. At Heysham and Lowestoft our results show a decrease with both year and GMT, suggesting that the magnitude of extreme skew surge events are getting smaller with anthropogenic climate change effects. In the 100 year period 1920-2020 at Newlyn, the point estimate $\delta_\sigma^{(\tilde{k})}$ corresponds to an increase in mean excesses (see expression (4.2.3)) of 2mm (relative to a mean of 94mm in 1920), whilst at Sheerness in the years of observation 1980-2016 this corresponds to an increase of 10mm relative to a mean of 125mm in 1980. Notice there

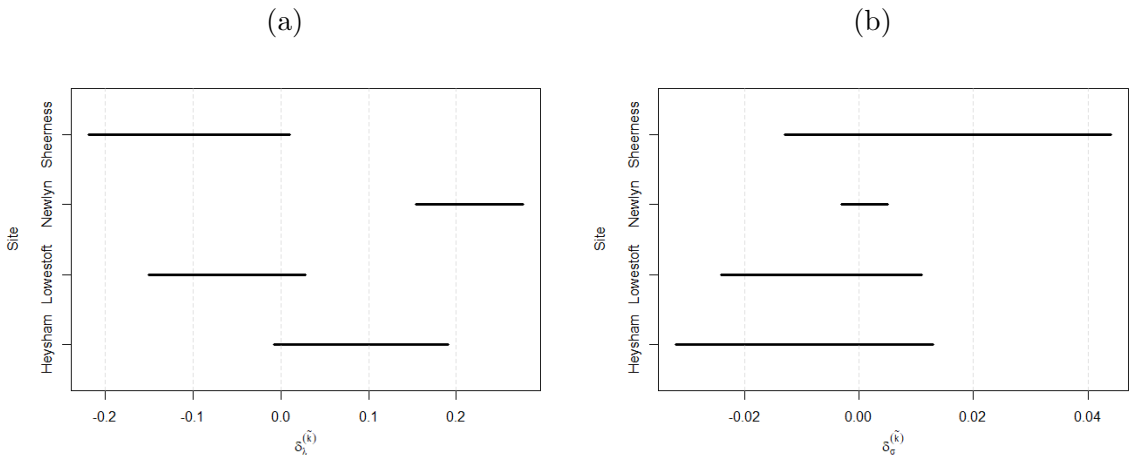


Figure 4.3.3: Confidence intervals for parameter estimates (a) $\hat{\delta}_{\lambda}^{(\tilde{k})}$ and (b) $\hat{\delta}_{\sigma}^{(\tilde{k})}$ for all sites.

is overlap in the parameter estimates for $\delta_{\sigma}^{(\tilde{k})}$ and $\delta_{\sigma}^{(m)}$ across sites; in Section 4.3.4 we fit similar model with these trend parameters common across sites (see Figure 4.3.3).

Models $S2$ and $S4$ have four additional parameters relative to Model $S0$, these denote a separate trend for each season with respect to year and GMT. AIC and BIC are still minimised by Model $S0$, except AIC scores for Heysham and Lowestoft, which favour Models $S2$ and $S4$, respectively. However, the four confidence intervals overlap at each site, suggesting a fixed trend within a year is sufficient. At Heysham, the overlap across all seasons is small but there is considerable overlap between winter and spring, with positive trend parameters, and likewise for summer and autumn with negative trend parameters for both models. If these trends were statistically significant it would suggest that the magnitude of extreme skew surges is increasing with increases in GMT in December-April, but decreasing for the rest of the year. Given the timing of the most extreme events, this could be important if statistically significant.

4.3.3 Return level estimation

Using Models $S0$ (4.2.5) and $R4$ (4.2.9) for the scale and rate parameter, respectively, we estimate sea level return levels from the annual maxima distribution in expres-

sion (4.2.7). Recall Model $R4$ for the GPD rate parameter has a linear seasonal trend with respect to GMT in year k, denoted m_k. Solving

$$\Pr(M \leq z | m_k = m) = 1 - p,$$

for $p \in [0, 1]$ gives us the level we expect the annual maxima M to exceed once every $1/p$ years, on average, when the GMT covariate is fixed at some value m. We estimate return levels for temperatures in 1915, 2020 and for a year when the GMT anomaly value is 1°C higher than that in 2020; these correspond to anomalies of -0.19, 0.92 and 1.92°C, respectively.

Table 4.3.3 gives the sea level return level estimates for the 1, 100 and 10,000 year level at each site. These are relative to the mean sea level in 2017 since the linear mean sea level trend was removed when preprocessing the data, so these trends are in excess to those already observed in the mean or will occur as GMT increases. Return level estimates increase with temperature anomaly for all sites at all return periods. The 1 year level increases similarly (∼3-4cm) over the four sites, with the greatest difference of 10cm observed at Heysham for the 10,000 year return level; this is a significant difference for coastal flood defence design. Return levels will be underestimated if the long-term trends in extreme skew surge occurrence are not accounted for and instead only estimated changes in mean sea level are used to update return level estimates. At Lowestoft, Newlyn and Sheerness, the 10,000 return level increases by 4, 3 and 2cm, respectively, when GMT increases from -0.19 to 1.92°C. Therefore, even when some of the parameter estimates of Model $R4$ for the seasonal GMT trend ($\delta^m_{\lambda,s}$ for $s = 1, 2, 3, 4$) were negative, the resulting return level estimates still increase with GMT. This outcome depends on which seasons have which trends. For annual maximum sea levels, it is only the winter and autumn trends that are influential. Although the seasonal changes seem non-homogeneous in the GPD model for extreme skew surges, our results show that, when combined with tidal information, the sea level return levels

Table 4.3.3: Estimates of the 1, 100 and 10,000 year sea level return levels (in metres), relative to the mean sea level in 2017, using Models $R4$ and $S0$ for the GPD rate and scale parameters, respectively, for skew surges with GMT as a fixed covariate equal to anomalies of -0.19°C (as in 1915), 0.92°C (as in 2020) and 1.92°C.

	Heysham			Lowestoft			Newlyn			Sheerness		
	1	100	10,000	1	100	10,000	1	100	10,000	1	100	10,000
−0.19°C	10.61	11.52	12.45	3.47	4.60	5.81	6.07	6.55	6.94	6.41	7.17	7.98
0.92°C	10.63	11.56	12.50	3.49	4.61	5.83	6.09	6.57	6.95	6.42	7.18	7.99
1.92°C	10.65	11.60	12.55	3.50	4.63	5.85	6.11	6.60	6.97	6.44	7.19	8.00

exhibit much more consistent behaviour with GMT changes across sites.

4.3.4 Spatial pooling

We present the results from pooling information across sites, for the long term trend parameters, when refitting the models of Section 4.2.4. Before pooling information, we use the dependence measures discussed in Section 4.2.5 to check if it is reasonable to assume each pair of sites are independent in their extreme skew surge values. We estimate the dependence measures for the daily maximum observed skew surges and a standardised transformation of them to remove sources of within-year non-stationarity via mapping to uniform margins through the distribution function (4.2.4). The results are shown in Table 4.3.4. For most combinations of sites at lags $t = -1, 0, 1$ (days) the dependence is weak, except for Newlyn with each of the east coast sites where Kendall's τ is near 0, whilst for Lowestoft and Sheerness this value is approximately 0.5. The effect of de-seasonalising the data (by transforming to uniform margins) has typically decreased dependence. With the exception of Lowestoft and Sheerness, it is not unreasonable to make an independence in extremes approximation for the data. We find $\bar{\chi} < 1$ for all pairs, giving evidence of asymptotic independence with weak dependence in the observed tails of the variables. The strongest dependence is found between Lowestoft and Sheerness at lag $t = 0$. This is not surprising since these sites are close in proximity, with extreme skew surges progressing south down the east coast through Lowestoft onto Sheerness. Therefore they are highly likely to be affected by

Table 4.3.4: Kendall's τ, χ and $\bar{\chi}$ measures of dependence for daily maximum skew surge observations at pairs of sites. We show the dependence over lags -1 (LHS site is 1 day behind RHS), 0 and 1 (LHS site is 1 day ahead of RHS); in bold we show the largest dependence over these lags. χ and $\bar{\chi}$ are measures of extremal dependence for exceedances of the 0.95 quantile.

	Heysham-Lowestoft			Heysham-Newlyn			Heysham-Sheerness			Lowestoft-Newlyn			Lowestoft-Sheerness			Newlyn-Sheerness		
lag	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1
Observations																		
τ	0.133	0.160	**0.309**	0.287	**0.322**	0.259	0.153	0.149	**0.298**	**0.089**	0.040	0.034	0.155	**0.510**	0.238	0.137	0.168	**0.196**
χ	0.095	0.129	**0.270**	0.127	**0.145**	0.076	0.092	0.111	**0.259**	**0.017**	0	0	0.145	**0.509**	0.200	0.054	0.077	**0.121**
$\bar{\chi}$	0.200	0.251	**0.424**	0.249	**0.276**	0.160	0.195	0.224	**0.412**	**0.040**	-0.018	-0.055	0.274	**0.645**	0.344	0.120	0.158	**0.237**
Transform to Uniform(0,1)																		
τ	0.103	0.130	**0.289**	0.285	**0.318**	0.244	0.108	0.102	**0.262**	**0.086**	0.036	0.028	0.143	**0.523**	0.228	0.139	0.173	**0.200**
χ	0.026	0.040	**0.180**	0.103	**0.122**	0.053	0.056	0.036	**0.173**	0	0	0	0.095	**0.494**	0.174	0.003	0.016	**0.050**
$\bar{\chi}$	0.069	0.100	**0.321**	0.215	**0.236**	0.114	0.123	0.084	**0.313**	**-0.012**	-0.107	-0.134	0.198	**0.634**	0.316	0.003	0.035	**0.114**

the same storms. Despite this pair of sites giving clear evidence of dependence, we continue under the belief that it is reasonable to assume skew surge daily maxima at all pairs of sites are sufficiently close to being independent for the purposes of spatial pooling.

Firstly, we focus on pooling information across sites regarding the long term trend parameters with respect to year k and GMT m_k for the rate parameter. Figure 4.3.3 shows there is considerable overlap in the confidence intervals for $\hat{\delta}_{\lambda}^{(\tilde{k})}$ at Lowestoft and Sheerness; similarly, there is some overlap for Heysham and Newlyn. Although pooling information across randomly selected subsets of sites should be discouraged, here we note that the pairs of sites with similarities are on different coastlines, so we explore pooling over sites on the east coast (Lowestoft and Sheerness) and separately on the west coast (Heysham and Newlyn). Here, we consider refitting Models $R1$ and $R3$ (i.e., a fixed trend parameter within a year) with common trend parameters $\delta_{\lambda}^{(\tilde{k})}$ and $\delta_{\lambda}^{(m)}$ between the pairs of sites. We obtain negative trends parameters $\hat{\delta}_{\lambda}^{(\tilde{k})} = -0.084\ (-0.151, -0.016)$ and $\hat{\delta}_{\lambda}^{(m)} = -0.070\ (-0.156, 0.015)$ for Sheerness and Lowestoft, whilst at Newlyn and Heysham we obtain statistically significant positive trend parameters $\hat{\delta}_{\lambda}^{(\tilde{k})} = 0.180\ (0.128, 0.231)$ and $\hat{\delta}_{\lambda}^{(m)} = 0.285\ (0.208, 0.359)$. Note 95% confidence intervals are given in parentheses here. Both of these models are an improvement on the previous results, where a separate long term trend parameter is used

for each site; the model with a yearly trend parameter reduces AIC by 48 and the BIC by 0.5, whilst the model with a GMT parameter reduces AIC and BIC by 54 and 6, respectively. This highlights the importance of sharing information spatially.

There is also information to be gained from sharing spatial information about long term trends in the scale parameter since there is considerable overlap in the confidence intervals for $\hat{\delta}_\sigma^{(\tilde{k})}$ (see Figure 4.3.3) and $\hat{\delta}_\sigma^{(m)}$ (see Table 4.3.2). We refit the models of Section 4.2.4 for the scale parameter with common long term trend parameters across sites, but neither parameter estimates are significant. We find that $\hat{\delta}_\sigma^{(\tilde{k})} = 4.8 \times 10^{-4}$, corresponding to an increase in scale parameter of 0.48mm over 1915-2020. For GMT $\hat{\delta}_\sigma^{(\tilde{k})} = -0.0024$, i.e., a 24mm decrease in scale parameter with a 1°C increase in temperature. Neither of these models improve the fit relative to having no long term trends (in addition to those in the mean sea level), although the AIC scores are close. We also fit a model similar to that of Models $S2$ and $S4$ so there is a common seasonal trend across sites, with respect to year and GMT but find that neither of these improve model fit. This agrees with our single-site results of Section 4.3.2 where we found no evidence of changes in the magnitude of extreme skew surge events with respect to year or GMT.

4.4 Discussion

We have presented a framework to investigate the effects of anthropogenic climate change on extreme skew surges as any increases in the magnitude or frequency in these events can have catastrophic consequences if not included in extreme sea level estimation for coastal flood defence design. These trends can be different to those observed in the main body of the data, such as mean sea level rise. We use year and GMT as covariates in our statistical model for extreme event occurrence, building on a model developed by D'Arcy et al. (2023b) that accounts for seasonality and skew

surge-peak tide dependence. Recall that our results are relative to the mean sea level trend in 2017 so this would need to be added onto any sea level return level estimates when used in practice. We show that there is evidence of an increase in the probability of an extreme skew surge event with GMT increases at Heysham and Newlyn, but evidence of both increases and decreases in the likelihood of these events at Lowestoft and Sheerness across the year. We do not find any significant changes in the magnitude of extreme skew surges, i.e., in the scale parameter, and hence in the mean of the skew surge excesses of the threshold. Accounting for seasonal changes in extreme skew surge occurrence with GMT in sea level return level estimation shows that return levels increase with GMT. For a 2.1°C increase in GMT, the 10,000 year return levels increased by 10, 4, 3 and 2cm at Heysham, Lowestoft, Newlyn and Sheerness, respectively. The ideas presented in this paper could be applied to more locations, but also to other environmental variables to investigate trends in extreme values.

We demonstrate the advantages of pooling information across sites, although this is only primarily illustrative since we consider just four sites here. There are 44 sites on the UK National Tide Gauge Network where this methodology could be extended. It would be interesting to apply our methodology within a spatial framework, for example in regional frequency analysis where sites in a homogeneous region not only have a common shape parameter, but also common long term trends due to anthropogenic climate change.

Skew surges are also believed to change over decadal time scales with climate indices. For example, the North Atlantic Oscillation index (NAO) describes such time scale changes in regional weather systems, so is believed to impact storm surges, and thus skew surge. Araújo and Pugh (2008) find a negative correlation between storm surge and air pressure patterns, using NAO. It would be interesting to explore how adding an NAO covariate into the GPD for extreme skew surges would change model fit.

Chapter 5

Simulating extreme skew surge for coastal management

5.1 Introduction

Coastline communities face a complex array of risks from natural hazards, exacerbated by human activities and climate change. Mainly, these communities are vulnerable to coastal flooding and erosion which are intrinsically linked (Pollard et al., 2019). Coastal erosion is a natural geological process that gradually wears away coastal landforms threatening buildings, roads and other forms of infrastructure near the coastline. Storm surges are temporary rises in sea levels caused by intense storms, that can lead to coastal flooding, causing damage to infrastructure and homes. Over 520,000 properties in England are at risk of coastal flooding and erosion, and the cost of these hazards to the UK economy is estimated to be around £1.1 billion each year (Committee on Climate Change, 2018).

Coastal erosion occurs over various timescales, from hours to millennia, but can be accelerated by human activities and environmental factors. Typically, erosion is caused by waves exerting forces on the coastline, currents and tidal movements transporting

sediment along the shoreline and severe weather events causing heightened storm surges and water levels (Masselink and Russell, 2007). Increases in storminess resulting from anthropogenic climate change have increased levels of erosion, along with sea level rise and changes in human activities at the coastline (Masselink et al., 2020). Severe storms bring extreme storm surges to the coastline that accelerate rates of coastal erosion; this is exacerbated when the storm event lasts multiple days with prolonged levels of high storm surge. Therefore, understanding the temporal structure of extreme storm surges is fundamental for effective coastal management.

Coastal flooding is also becoming increasingly complex to manage due to sea level rise and increases in storminess (Seneviratne et al., 2021). Storm surge barriers are an attractive and economical solution for flood protection in densely populated estuarine regions. These are made up of fully or partly movable gates that are closed before a storm to prevent flooding behind the barrier, then they subsequently reopen to facilitate shipping and allow the natural movement of tides (Mooyaart and Jonkman, 2017). There are more than 50 storm surge barriers globally, contributing to the protection of millions of people collectively. For example, the Thames Barrier in London (UK) protects 1.3 million people and £200 billion worth of property (Environment Agency, 2021).

With accelerating rates of sea level rise and increases in storminess resulting from anthropogenic climate change, surge barriers are closing more frequently and are closing in months when previously not required. Figure 5.1.1 shows closure rates of the Thames Barrier since it opened in 1982 (Environment Agency, 2021). From July 1982 to June 2002 the Thames Barrier in London closed 63 times to protect the city from flooding but closed 130 times over the next 20 years (July 2002 to June 2022), so closures more than doubled since the first 20 years. In 2020, the Thames Barrier closed in May for the first time; historically all other closures have occurred between September and April (Environment Agency, 2021).

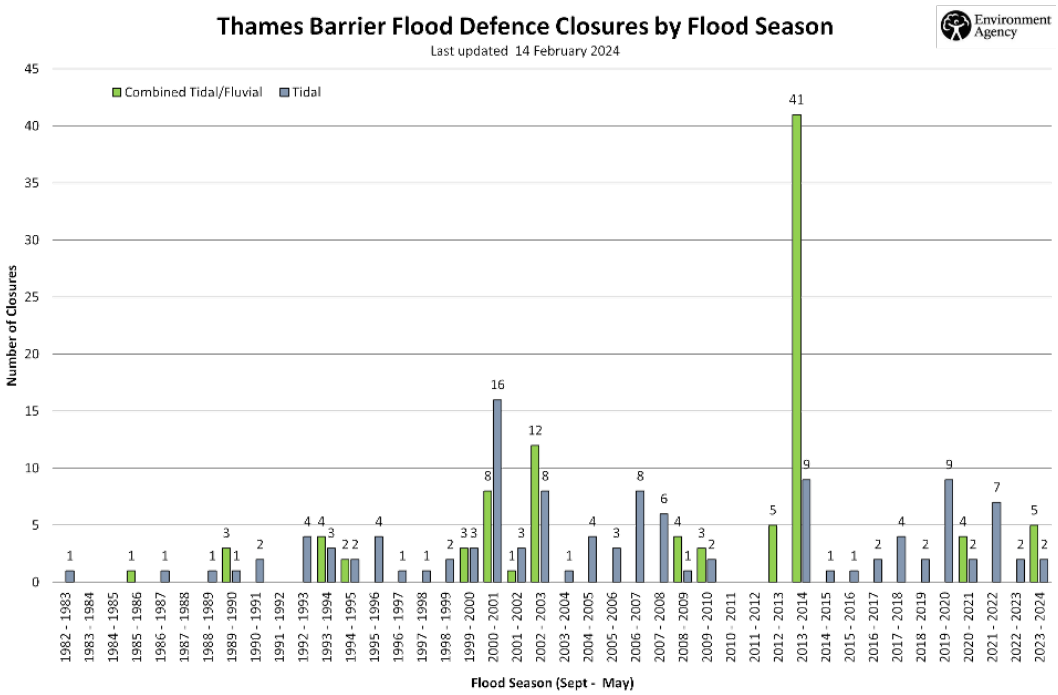


Figure 5.1.1: Thames Barrier closures by flood season (i.e., winter). This figure is taken from Environment Agency (2023b).

Increased use of barriers has critical implications for barrier management, maintenance and operation. This affects the integrity and reliability of a barrier and their projected life expectancy. Changes in the closure times (i.e., the months when the barrier closes) impact the maintenance, upgrading and testing that is typically scheduled over the low-risk period, which is crucial for ensuring the barrier can continue in operation. As barrier closure rates continue to rise, it is vital that their operators adapt the barrier management and maintenance to account for climate change effects, and plan for inevitable barrier upgrades.

We aim to simulate future time series of skew surges that reflect their temporal dependence structure and non-stationarity. Our simulations are useful for coastal flooding and erosion management, and for predicting barrier closure rates/times. We use the model of D'Arcy et al. (2023b) (outlined in Chapter 3) for modelling skew surges that accounts for their seasonality and dependence on peak tide, identified at Sheerness (Williams et al., 2016). In this model, temporal dependence of extreme events

was captured using the extremal index which summarises the clustering time of extreme values (Smith and Weissman, 1994). However, we propose a more sophisticated approach that captures temporal dependence in the entire range of data, whilst reproducing the extremal dependence similarly to the extremal index. We assume that skew surges follow a kth order Markov process and model the pairwise dependence of values separated by κ observations (where $\kappa < k$) using a Gaussian copula (Joe, 2014). We investigate the non-stationarity of temporal dependence in skew surges and capture this by incorporating a regression model on the parameters of the Gaussian copula.

We first summarise the methods for modelling extreme skew surges that were initially derived in Chapter 3; this was the first approach to account for seasonality, skew-surge peak tide dependence and temporal dependence. This was then extended in Chapter 4 to account for longer-term changes resulting from anthropogenic climate change. For simplicity of presentation, coupled with the finding that trends in extreme skew surges are minimal relative to trends in mean sea level, we do not consider climate change effects in the tail here. Here, the approach for capturing extremal dependence was extended to allow for simulation, so that the simulated data reflects the temporal dependence structure in both the body and the tail. We outline each of these existing approaches in Section 5.3 before detailing our proposed extensions in Section 5.4. We present our resulting simulations at four UK coastal locations in Section 5.5.

5.2 Background

5.2.1 Extreme value methods

We are interested in modelling extreme values of the skew surge series $\{Y_t\}$ and use techniques from extreme value statistics to do so; we detail these methods in this section. Assume, for now, that the series $\{Y_t\}$ is independent and identically distributed (IID). One way of defining extreme observations is to consider the block maximum M_n of

the sequence $Y_1, \ldots, Y_n$, i.e., $M_n = \max\{Y_1, \ldots, Y_n\}$. The sequence $\{Y_t\}$ is assumed to have a continuous marginal distribution function F and upper end point y^F. If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, so that the rescaled block maximum $(M_n - b_n)/a_n$ has a nondegenerate limiting distribution as $n \to \infty$, then the cumulative distribution function G of this limit is the generalised extreme value distribution (GEV), defined in Section 2.2.1 of Chapter 2. The GEV has parameters $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ representing the location, scale and shape, respectively (Coles, 2001). This result provides asymptotic motivation for using the GEV as a parametric model for observed block maxima.

We relax the independence assumption, so that $Y_1, \ldots, Y_n$ is a stationary sequence with the same marginal distribution function F. The limiting distribution of the rescaled block maxima of a stationary process satisfying a long-range dependence condition (see Section 2.3 of Chapter 2) is $G^\theta(y)$ with $G(y)$ as above and $\theta \in (0, 1]$ the extremal index (Leadbetter et al., 1983; O'Brien, 1987). For an independent series $\theta = 1$, but the converse is not true. Based on the limiting expression for θ of O'Brien (1987), Ledford and Tawn (2003) propose the threshold-based extremal index $\theta(y)$ as a measure of temporal dependence for exceedances of some threshold y as

$$\theta(y) = \Pr(\max\{Y_2, \ldots, Y_r\} < y | Y_1 > y), \tag{5.2.1}$$

where r is termed the run length (Smith and Weissman, 1994).

A more popular approach for modelling extremes is to define them as exceedances of a high threshold u. If $Y_1, \ldots, Y_n$ are IID, then exceedances of a high threshold u are also IID and have limiting GPD tail model

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

for $y > u$ where $\lambda_u = \Pr(Z > u)$ and $(\sigma_u, \xi) \in \mathbb{R}_+ \times \mathbb{R}$ are the scale and shape parame-

ters, respectively (Coles, 2001). If $Y_1, \ldots, Y_n$ are dependent but stationary, a common approach is to identify independent clusters above a high threshold and evaluate the cluster maxima as a method of declustering; the run length r of equation (5.2.1) is used for identifying different clusters (Fawcett and Walshaw, 2007).

The extremal index is not the only measure of temporal dependence for extremes. Coles et al. (1999) define the following measure of asymptotic dependence

$$\chi_\kappa = \lim_{y \to y^F} \Pr(Y_{t+\kappa} > y | Y_t > y),$$

where $\kappa \in \mathbb{N}$ and y^F is the upper end-point of the common marginal distribution of Y_t and $Y_{t+\kappa}$. We say that Y_t and $Y_{t+\kappa}$ are asymptotically dependent, or exhibit extremal dependence, when $\chi_\kappa > 0$. When $\chi_\kappa = 0$, we say that Y_t and $Y_{t+\kappa}$ are asymptotically independent, whilst perfect extremal dependence corresponds to $\chi_\kappa = 1$. Since χ_κ fails to signify the level of asymptotic independence, Coles et al. (1999) also define the measure $\bar{\chi}_\kappa$ as

$$\bar{\chi}_\kappa = \lim_{y \to y^F} \frac{2 \log \Pr(Y_t > y)}{\log \Pr(Y_t > y, Y_{t+\kappa} > y)} - 1,$$

where $\bar{\chi}_\kappa \in (-1, 1]$. Asymptotic dependence and asymptotic independence correspond to $\bar{\chi}_\kappa = 1$ and $\bar{\chi}_\kappa < 1$, respectively, whilst $0 < \bar{\chi}_\kappa < 1$ and $-1 < \bar{\chi}_\kappa < 0$ correspond to positive and negative association, respectively, and $\bar{\chi}_\kappa = 0$ corresponds to near independence. To summarise extremal dependence at lag κ, the pair of measures $(\chi_\kappa, \bar{\chi}_\kappa)$ is needed.

A variety of approaches exist for modelling non-stationary processes, where it is common to allow the parameters of a stationary statistical model to vary with time or another covariate. We discuss the methodology of Chapter 3 for capturing the non-stationarity of skew surges in Section 5.3.

5.2.2 Data

We use data from the UK National Tide Gauge Network at Heysham, Lowestoft, Newlyn and Sheerness (BODC, 2020). Observations of skew surge are at an approximate twice daily temporal resolution (specifically, every 12 hours and 26 minutes, i.e., one tidal cycle). Heysham is located on the west coast of England and has records from 1964-2016, with 17% missing. Lowestoft is on the east coast of England, with data available from 1964-2020 (4% missing). Sheerness is at the Thames Estuary, also on the east coast, with data available from 1980-2016 and 9% missing. Newlyn is located on the south coast of England and has records from 1915-2020, with 17% missing. The data are in metres relative to chart datum. Sheerness and Lowestoft are both surge dominant sites, whilst Heysham and Newlyn are tidally dominant.

5.2.3 Exploratory analysis

Chapter 3 conducts a thorough exploratory analysis to explore the non-stationarity of extreme skew surges at Heysham, Lowestoft, Newlyn and Sheerness. Here, we focus on analysing skew surge temporal dependence which is present because observations are driven by meteorological conditions which often span multiple tidal cycles. Therefore, we expect to observe temporal dependence in the main body and the tails. We investigate the relationship between values separated by $\kappa = 1, \ldots, 6$ time lags where a lag of 1 corresponds to a tidal cycle of ~ 12.5 hours, so our choice of κ ranges from adjacent values to those separated by ~ 3 days.

We look at the dependence in the main body of the data, using Kendall's τ measure. This is a measure of rank correlation, so is robust to outliers and any marginal distributional features, but it is a dependence measure across all values of the variables. However, since our interest lies with the dependence of the extreme values, it is natural to also study the two main measures of extremal dependence χ_κ and $\bar{\chi}_\kappa$ defined in Section 5.2.1 for values separated by lag κ.

	Heysham						Lowestoft						Newlyn						Sheerness					
κ	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Kendall's τ																								
Obs	0.51	0.42	0.35	0.31	0.25	0.24	0.31	0.12	0.10	0.12	0.08	0.06	0.62	0.50	0.43	0.39	0.33	0.31	0.28	0.13	0.11	0.13	0.08	0.07
Uniform	0.50	0.42	0.34	0.30	0.23	0.23	0.30	0.12	0.09	0.12	0.07	0.06	0.61	0.49	0.42	0.38	0.30	0.29	0.28	0.13	0.10	0.13	0.07	0.07
χ_κ																								
Obs	0.14	0.09	0.06	0.06	0.05	0.06	0.17	0.04	0.05	0.06	0.03	0.03	0.31	0.14	0.12	0.13	0.12	0.12	0.08	0.05	0.04	0.06	0.04	0.04
Uniform	0.11	0.05	0.04	0.04	0.03	0.02	0.16	0.03	0.02	0.03	0.01	0.00	0.30	0.11	0.07	0.08	0.06	0.07	0.10	0.05	0.03	0.03	0.02	0.01
$\bar{\chi}_\kappa$																								
Obs	0.41	0.32	0.26	0.24	0.23	0.24	0.45	0.20	0.23	0.25	0.16	0.15	0.60	0.41	0.37	0.39	0.38	0.38	0.31	0.22	0.18	0.25	0.20	0.20
Uniform	0.36	0.22	0.20	0.18	0.16	0.12	0.43	0.16	0.09	0.15	0.07	0.02	0.59	0.35	0.29	0.29	0.27	0.27	0.34	0.19	0.15	0.16	0.12	0.07

Table 5.2.1: Kendall's τ, χ and $\bar{\chi}$ empirical estimates for lags $\kappa = 1, \ldots, 6$ at each site, for the observed data and a transform to uniform margins. The 0.99 quantile is used for estimating χ and $\bar{\chi}$.

Table 5.2.1 shows these measures for the observed data and a standardised transformation to remove sources of within-year non-stationarity via mapping to uniform margins through the seasonal skew surge model of Chapter 3; we discuss their method further in Section 5.3.1. At all sites, the lag 1 Kendall's τ measure is stronger than for $\kappa = 2, \ldots, 6$, as expected. Newlyn exhibits the strongest dependence for all lags, followed by Heysham; the dependence for the east coast sites (Lowestoft and Sheerness) is much weaker at all lags. All values of χ_κ lie close to zero (i.e., less than 0.18) whilst $\bar{\chi}_\kappa$ lies between 0 and 1, suggesting asymptotic independence but positive association. The effect of de-seasonalising the data (by transforming to uniform margins to remove marginal seasonality) has typically decreased dependence, suggesting that some of the temporal dependence identified directly from the observations is driven by seasonal variations.

Figure 5.2.1 shows estimates of Kendall's τ, χ and $\bar{\chi}$ for each season, for the transformed series onto uniform margins at Sheerness. The seasons are defined as winter (December, January, February), spring (March, April, May), summer (June, July, August) and autumn (September, October, November). For each year, we estimate each dependence measure and show a box plot of the estimates over all years. This shows that summer has the strongest dependence in both the body and the tail, with the largest estimates of Kendall's τ and χ_1, on average. We make the same conclusions for

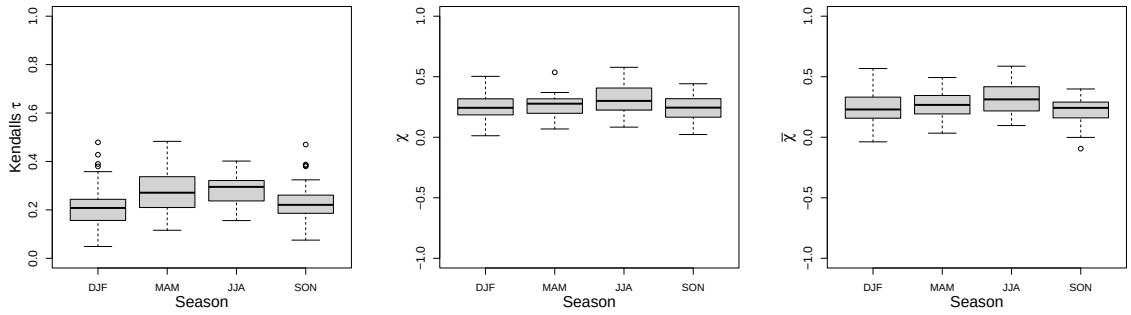


Figure 5.2.1: Seasonal Kendall's τ (left), χ (centre) and $\bar{\chi}$ (right) empirical estimates for lags $\kappa = 1$ at Sheerness for the data transformed to uniform margins, for each year. The seasons are defined as DJF (winter: December, January, February), MAM (spring: March, April, May), JJA (summer: June, July, August) and SON (autumn: September, October, November).

higher order κ, although the differences across seasons are smaller as we increase the lag κ, since they are closer to zero. We also find stronger dependence in summer for the remaining sites so make similar conclusions about the seasonal structure of temporal dependence.

5.3 Existing methods

An extreme sea level can either occur from a combination of a moderate tide and a large skew surge, or a combination of a moderate skew surge with a large tide. Therefore, it is fundamental that the extremes of both variables are accounted for within a sea level model. Since tides are deterministic, they have a known upper bound and do not require statistical modelling. However, skew surges are stochastic so we want to build a model for all the data, specifically focusing on the upper tail.

The earliest method to estimate extreme sea levels by decomposing the water level into skew surge and peak tide was the skew surge joint probabilities method (SSJPM) of Batstone et al. (2013). They assume skew surge-peak tide independence, and that both processes are IID. The generalised Pareto distribution (GPD) is used for modelling

the upper tail of skew surges, and the empirical distribution for the main body of the distribution.

In Chapter 3, we extend the SSJPM to account for seasonality in both skew surge and peak tide, as well as their dependence; we discuss this model in Section 5.3.1. In Section 5.3.2, we review our approach from Chapter 3 for capturing temporal dependence in skew surges.

5.3.1 Marginal modelling

We summarise the marginal skew surge model developed in Chapter 3. To split the distribution into the body and tail, we use a monthly threshold u_j for $j = 1, \ldots, 12$ to account for seasonality, with u_j being the 0.95 quantile of month j's skew surge distribution. Skew surges below these thresholds are modelled using the monthly empirical distribution, that is dependent on the associated peak tide x, denoted $\tilde{F}_{j,x}$. The empirical distribution is split into three associated peak tide bands corresponding to the lowest, medium and highest thirds of observed peak tides. This means that both skew surge-peak tide dependence and skew surge seasonality in the body are accounted for.

For exceedances of the monthly threshold, we use a non-stationary GPD dependent on the day in year $d = 1, \ldots, 365$, month j and associated peak tide x. The shape parameter is kept fixed whilst the rate and scale parameters both depend on day and peak tide. We model the scale parameter using a harmonic for seasonal variations, using day in year d, and a linear trend in terms of tide x,

$$\sigma_{d,x} = \alpha_\sigma + \beta_\sigma \sin\left(\frac{2\pi}{f}(d - \phi_\sigma)\right) + \gamma_\sigma x, \tag{5.3.1}$$

for $\alpha_\sigma > \beta_\sigma > 0$, $\phi_\sigma \in [0, 365)$, $\gamma_\sigma \in \mathbb{R}$ parameters to be estimated and $f = 365$ the periodicity. The rate parameter $\lambda \in [0, 1]$ is modelled similarly, using a generalised linear model with logit link function, a harmonic to capture seasonal variations and a

second harmonic to capture skew surge-peak tide dependence. This parameterisation is given by

$$g(\lambda_{d,x}) = g(\lambda)+(d_j-\bar{d}_j)\beta_\lambda^{(d)} \sin\left(\frac{2\pi}{f}(d-\phi_\lambda^{(d)})\right)+\left(\frac{x-\bar{x}}{s_x}\right)\left[\alpha_\lambda^{(x)}+\beta_\lambda^{(x)} \sin\left(\frac{2\pi}{f}(d-\phi_\lambda^{(x)})\right)\right], \tag{5.3.2}$$

for $g(\cdot)$ the logit link function, λ is the constant exceedance probability in a month, $d_j \in [1, 31]$ is day in month (standardised by the monthly mean day $\bar{d}_j$), $\bar{x}$ is the mean and s_x is the standard deviation of peak tides, and $\alpha_\lambda^{(x)} \in \mathbb{R}$, $\beta_\lambda^{(d)}, \beta_\lambda^{(x)} > 0$, $\phi_\lambda^{(d)}, \phi_\lambda^{(x)} \in [0, 365)$ parameters to be estimated.

Then, the skew surge model of Chapter 3 is given by,

$$F_Y^{(d,j,x)}(y) = \begin{cases} \tilde{F}_{j,x}(y) & \text{if } y \leq u_j \\ 1 - \lambda_{d,x}\left[1+\xi\left(\frac{y-u_j}{\sigma_{d,x}}\right)\right]_+^{-1/\xi} & \text{if } y > u_j, \end{cases} \tag{5.3.3}$$

where $\tilde{F}_{j,x}(\cdot)$ is the empirical distribution and $\sigma_{d,x}$ and $\lambda_{d,x}$ are defined by expressions (5.3.1) and (5.3.2), respectively.

Each sea level observation Z_t for $t = 1, \ldots, n$, where n is the total number of observations, can be written as the sum of skew surge Y_t and peak tide X_t. Therefore, in Chapter 3 we use the joint probabilities method to derive the distribution of Z_t. Let $T_j^{(k)}$ denote the number of tidal cycles in month j and year k. We capture within and across year peak tide non-stationarity using sequential monthly and yearly peak tide samples $Y_{j_i}^{(k)}$, so that j_i denotes the ith peak tide in month j, and $k = 1, \ldots, K$ represents the year, for K years of observation. Since peak tides are temporally dependent, the samples $\{Y_{j_i}^{(k)}\}$ are from contiguous peak tides. Then, assuming skew surges are independent, the distribution of the annual maxima sea level M is

$$\Pr(M \leq z) = \frac{1}{K}\sum_{k=1}^{K}\prod_{j=1}^{12}\prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j,x)}(z - X_{j_i}^{(k)}), \tag{5.3.4}$$

where $F_Y^{(d,j,x)}$ is the skew surge model (5.3.3).

5.3.2 Temporal dependence

To capture in skew surge temporal dependence in skew surges in Chapter 3, we use the extremal index. This is incorporated in the annual maximum sea level distribution of (5.3.4) as follows

$$\Pr(M \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{j=1}^{12} \prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})^{\theta(z - X_{j_i}^{(k)}, r)}, \tag{5.3.5}$$

where $F_Y^{(d,j,x)}$ is the skew surge model (5.3.3) and $\theta(z - X_{j_i}^{(k)}, r)$ is a model for the extremal index, dependent on skew surge level $y = z - X_{j_i}^{(k)}$ and run length r, to capture temporal dependence of skew surges. This model is given by

$$\hat{\theta}(y, r) = \begin{cases} \tilde{\theta}(y, r) & \text{if } y \leq v \\ \theta - [\theta - \tilde{\theta}(v, r)] \exp\big(-\frac{y - v}{\psi}\big) & \text{if } y > v, \end{cases}$$

where v is a high threshold (taken to be the 0.99 quantile), $\psi > 0$ and $\tilde{\theta}(v, r) \leq \theta \leq 1$ are parameters to be estimated and $\tilde{\theta}(y, r)$ is the empirical runs estimate (Smith and Weissman, 1994).

This extremal index model only accounts for temporal dependence in the tail of the distribution. However, in Section 5.2.3 we find strong temporal dependence in the main body of skew surges. Since extreme sea levels can result from a high tide and moderate skew surge observation, temporal dependence must be also captured for non-extreme events. Since coastal erosion and flooding rates are accelerated during prolonged periods of high water levels, it is fundamental to capture temporal dependence across the range of data, since a single extreme event is potentially less damaging than a series of high values that are not considered extreme (Masselink and Russell, 2007). Addi-

tionally, when simulating from the skew surge model (5.3.3), the temporal dependence is not yet accounted for as it only comes into play in the sea level maxima model, see expression (5.3.5). Therefore, in the following sections, we propose an alternative approach.

5.4 Methodology

In this section, we detail our methodology for accounting for skew surge temporal dependence using Markov models and copulas. Assuming that the observed skew surges $\{Y_t\}$, for $t = 1, \ldots, N$, where N is the total number of observations, follow a kth order Markov process, we can then model the dependence between Y_t and $Y_{t+\kappa}$ for all t and $\kappa \leq k$ using copulas. Copula functions provide an appropriate model for the dependence structure between variables and are classically applied in the multivariate setting. In Section 2.3.4 of Chapter 2 we discuss how this framework can be applied in the univariate setting, as is necessary here. We refer the reader to Joe (2014) and Nelsen (2006) for a detailed review of copulas.

Copulas can be used to model temporal dependence in a univariate setting for observations separated by time lag κ, for $\kappa \in \mathbb{N}$. For simplicity, assume that the skew surge series $\{Y_t;\ t = 1, \ldots, N\}$ is stationary, with marginal distribution function F_Y, and that it follows a kth order Markov chain. Winter and Tawn (2017) show that the joint density of $\boldsymbol{Y}_{1:n} = (Y_1, \ldots, Y_n)$, denoted $f_{1:n}$ for $n > k$, can be written as

$$f_{1:n}(\boldsymbol{y}_{1:n}) = f_{1:k}(\boldsymbol{y}_{1:k}) \prod_{t=1}^{n-k} f_{k+1|1:k}(y_{t+k} \mid \boldsymbol{y}_{t:(t+k-1)}) = f_{1:k}(\boldsymbol{y}_{1:k}) \prod_{t=1}^{n-k} \frac{f_{1:k+1}(\boldsymbol{y}_{t:t+k})}{f_{k+1}(y_{t+k})}, \tag{5.4.1}$$

where $f_{k+1|1:k}$ is the conditional density of Y_{k+1} given $\boldsymbol{Y}_{1:k}$, and the subscript $i : j$ corresponds to variable indices $(i, i+1, \ldots, j)$ for $i \leq j$. A stationarity assumption is

required so that the joint marginal distribution satisfies

$$f_{i:j}(\boldsymbol{y}_{i:j}) = f_{i+\kappa:j+\kappa}(\boldsymbol{y}_{i:j}), \tag{5.4.2}$$

for $i \leq j$, any $\kappa \in \mathbb{N}$ and all $\boldsymbol{y}_{i:j} \in \mathbb{R}^{j-i+1}$. Copula models can be used for modelling the joint distribution $f_{1:k+1}$, with joint distribution function $F_{1:k+1}$ satisfying

$$F_{1:k+1}(\boldsymbol{y}_{1:k+1}) = C_{1:k+1}\{F_Y(y_1), \ldots, F_Y(y_{k+1})\},$$

where $C_{1:k+1}$ is the copula for $(k+1)$ successive variables in the Markov chain. The Markov process inherits the stationary condition (5.4.2) if the copula $C_{1:k+1}$ has the property that it's m-dimensional marginal distribution, for all $m < k+1$, satisfies

$$C_{i_1,\ldots,i_m}(\boldsymbol{y}_{1:m}) = C_{i_1+\kappa,\ldots,i_m+\kappa}(\boldsymbol{y}_{1:m}),$$

for $\kappa \in \mathbb{N}$, $i_j \in \mathbb{N}$ for $j = 1, \ldots, m$ with $1 \leq i_1 < \ldots < i_m + \kappa \leq k+1$, and $\boldsymbol{y}_{1:m} \in \mathbb{R}^m$. The joint density $f_{1:k+1}$ of (5.4.1) can be rewritten as the density of the copula, so that we can obtain the joint marginal density $f_{1:n}$ in terms of a product of copula densities.

In Section 5.4.1 we test two different copula models with contrasting extremal dependence structures: asymptotic independence and asymptotic dependence (Coles et al., 1999). Since the strength of temporal dependence varies across the year, as discussed in Section 5.2.3, we investigate a non-stationary dependence model by allowing the copula parameters to vary with time in Section 5.4.2.

5.4.1 Stationary dependence model

We consider the Gaussian and bivariate logistic extreme value distribution (subsequently referred to as the logistic) copulas as these capture contrasting extremal dependence structures through a single parameter. The Gaussian copula is suitable for

asymptotically independent data whilst the logistic copula is useful for asymptotically dependent data.

Let U and V be uniform random variables in (0,1), i.e., Y_t and $Y_{t+\kappa}$ transformed through the marginal model of Section 5.3.1. The bivariate Gaussian copula with correlation parameter $\rho \in (-1, 1)$ is given by

$$C(u, v, \rho) = \mathbf{\Phi}_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho), \quad u, v \in (0, 1),$$

where $\mathbf{\Phi}_2(\cdot, \cdot)$ is the bivariate standard normal distribution and $\Phi^{-1}(\cdot)$ the inverse of the univariate standard normal distribution function. Then the Gaussian copula density can be written as

$$c(u, v; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2 x^2 + \rho^2 y^2 - 2\rho xy}{2(1-\rho^2)}\right\}, \quad u,\, v \in (0, 1),$$

where $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$. For the Gaussian copula, $\chi = 0$ so this copula is suitable for asymptotically independent variables (Heffernan, 2000).

The logistic copula, introduced by Émile and Gumbel (1960), with parameter $0 < \alpha \leq 1$ is given by

$$C(u, v; \alpha) = \exp\big\{ - \left[(-\log u)^{1/\alpha} + (-\log v)^{1/\alpha}\right]^{\alpha}\big\}, \quad u, v \in (0, 1).$$

When $\alpha = 1$, U and V are clearly independent, with $C(u, v; 1) = uv$. The logistic density is written as

$$c(u, v; \alpha) = \frac{C(u, v; \alpha)}{uv}(xy)^{1/\alpha - 1}\left(x^{1/\alpha} + y^{1/\alpha}\right)^{\alpha - 2}\left[\left(x^{1/\alpha} + y^{1/\alpha}\right)^{\alpha} + \alpha^{-1} - 1\right],$$

for $x = -\log(u)$ and $y = -\log(v)$. For this copula, $\chi = 2 - 2^{\alpha}$ with dependence strengthening as α decreases to the limit so that $\chi = 1$ when $\alpha = 0$, therefore this is suitable for asymptotically dependent variables. Whereas when $\alpha = 1$, $\chi = 0$ so as

	Heysham		Lowestoft		Newlyn		Sheerness	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Gaussian	64704	64712	96049	96057	102161	102169	62608	62616
Logistic	75699	75707	106078	106086	122039	122048	69198	69206

Table 5.4.1: AIC and BIC scores (rounded to the nearest integer) for the Gaussian and logistic copula model for lag $\kappa = 1$, assuming a 6th order Markov process.

stated above, U and V are independent (Heffernan, 2000).

Table 5.4.1 compares the fit of these bivariate Gaussian and logistic copula to the series Y_t and $Y_{t+\kappa}$ for $\kappa = 1$ at all sites. A 6th order Markov assumption is used to allow comparison with models defined later, since $\kappa = 6$ is the highest value we consider. Model comparison is based on Akaike and Bayesian information criteria (AIC and BIC, respectively). The Gaussian copula performs best at each site (i.e., minimises AIC and BIC). This is unsurprising since our exploratory analysis of Section 5.2.3 suggests that the data separated by lag $\kappa = 1$ are asymptotically independent. Therefore, since we find that observations separated by lags $\kappa = 2, \ldots, 6$ are also asymptotically independent, it follows only to consider the Gaussian copula when developing a temporal dependence model for skew surges.

Table 5.4.2 compares values of κ, i.e., how many bivariate Gaussian copulas we consider at each site for sequences Y_t and $Y_{t+\kappa}$. We denote the best value of κ by $\kappa_S \in \mathbb{N}$ for each site $S = \{HEY, LOW, NEW, SHE\}$ denoting Heysham, Lowestoft, Newlyn and Sheerness, respectively. Again, each model is fit under a 6th order Markov assumption to allow comparison using AIC and BIC scores, which are used to choose the best value κ_S at each location. Our results suggest that $\kappa_{HEY} = 6$ and $\kappa_{NEW} = 5$. At Heysham, we tested $\kappa > 6$ also but found that $\kappa_{HEY} = 6$ still minimised AIC and BIC. The choice of κ_S is less obvious for Lowestoft and Sheerness since AIC and BIC scores suggest differing values; AIC suggests $\kappa_{LOW} = 5$ and $\kappa_{SHE} = 6$ whilst BIC says that $\kappa_{LOW} = \kappa_{SHE} = 4$.

5.4.2 Non-stationary dependence model

We focus on the bivariate Gaussian copula for modelling temporal dependence in skew surges separated by lag $\kappa = 1, \ldots, 6$ where we assume that skew surges follow a 6th order Markov process. This means that we can compare model fit using AIC and BIC scores as all models are fit to the same data as in Section 5.4.1. We allow the copula dependence parameter ρ to vary with day in year d, so that the temporal dependence model is non-stationary within a year. We investigate several ways to do this, and compare these approaches against the stationary model of Section 5.4.1.

Under a 6th order Markov process assumption, we consider the following models for $\kappa = 1, \ldots, 6$, which are subsequently referred to as the name in italics, as follows:

- *Model 1*: This is the standard Gaussian copula model, where the dependence parameters $\{\rho_\kappa\}$ are fixed across time of year,

- *Models 1W* and *1S*: The Gaussian copula with a different dependence parameter $\{\rho_\kappa\}$ for winter (*1W*) and summer (*1S*) data, defined as October - March and April - September, respectively. These parameters are denoted $\{\rho_\kappa^{(W)}\}$ and $\{\rho_\kappa^{(S)}\}$ for winter and summer, respectively. This allows the two seasons to have different order Markov process assumptions and copula models. By assuming a 6th order Markov process at this stage for both seasons, the AIC and BIC scores for these models are constructed so that different κ can be used for each season but we can sum the scores to be comparable with the remaining models,

- *Model 2*: The Gaussian copula dependence parameters $\{\rho_\kappa(d)\}$ are time dependent on the day of the year $d \in [1, 365]$, characterised by a harmonic trend, so that $\rho_\kappa(d) = \alpha_k + \beta_k \sin\left(\frac{2\pi}{365}\,(d - \phi_k)\right)$ where we estimate parameters $\alpha_\kappa, \beta_\kappa, \phi_\kappa$,

- *Model 3*: The Gaussian copula dependence parameters $\{\rho_\kappa\}$ are time dependent for $\kappa = 1, 2$ but are fixed across time, as in *Model 1*, for $\kappa > 2$. The harmonic of *Model 2* is used for $\rho_1(d)$ and $\rho_2(d)$.

Our exploratory analysis shows that empirical estimates of χ_1 and χ_2 vary the most with day in year d, compared with χ_κ for $\kappa > 2$. This motivates our modelling choice for *Model 3* but of course, other parameters could be fixed or allowed to vary with day. We discuss this further in Section 5.6.

Table 5.4.2 shows AIC and BIC scores for each model above, under a 6th order Markov assumption, for each site. This suggests that, at least, $\rho_1, \ldots, \rho_4$ are required to describe the temporal dependence structure for all sites, so that a 4th order Markov assumption is needed at a minimum. At Newlyn, the AIC and BIC scores are in agreement that *Model 2* is the best fit with $\kappa_{NEW} = 6$. However, the scores suggest different models for Heysham, Lowestoft and Sheerness. AIC suggests *Model 2* for these sites with $\kappa_{HEY} = 6$, $\kappa_{LOW} = 5$ and $\kappa_{SHE} = 6$, whilst BIC suggests *Model 3* with $\kappa_{HEY} = 6$, $\kappa_{LOW} = 4$ and $\kappa_{SHE} = 6$. Where $\kappa_S = 6$, we test $\kappa > 6$ (and adjust the Markov assumption accordingly) for these sites but find that the information criteria were still minimised for $\kappa_S = 6$; the details are omitted. The choices of κ_S under BIC agree with our exploratory analysis of Section 5.2.3 where we find strong dependence at all lags $\kappa = 1, \ldots, 5$ for Heysham and Newlyn, but the dependence diminishes for $\kappa > 4$ at Lowestoft and Sheerness. We use BIC to indicate our best performing model going forward.

5.4.3 Formulation of conditional parameters

In Sections 5.4.1 and 5.4.2 we follow the approach of Winter and Tawn (2017) to fit the Gaussian copula models, outlined in Section 5.4. Before model fitting, the data are transformed to standard Gaussian margins, using the probability integral transform (PIT): since Y is a continuous random variable with distribution function $F_Y^{(d,j,x)}$ as in equality (5.3.3), we can apply the PIT to transform between margins by obtaining a Uniform(0,1) random variable via $U = F_Y^{(d,j,x)}(Y)$ with $U \sim$ Uniform(0,1). Then, from a Uniform(0,1) random variable U, we can obtain a standard Gaussian random variable

	Model 1		*Model 1S*		*Model 1W*		*Model 2*		*Model 3*	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Heysham										
κ =1	65006	65015	32753	32760	32229	32236	64951	64976	64951	64976
κ =2	63571	63588	32079	32094	31465	31482	63530	63579	63530	63579
κ =3	63470	63495	31927	31949	31443	31465	63329	63404	63326	63383
κ =4	63304	63338	31896	31926	31295	**31325**	63155	63254	63143	63201
κ =5	63299	63340	31866	31904	31312	31350	63126	63250	63162	63237
κ =6	**63220**	**63270**	**31824**	**31870**	**31289**	31331	**63064**	**63213**	**63084**	**63167**
Lowestoft										
κ =1	96224	96233	48524	48531	47578	47586	96059	96085	96059	96085
κ =2	96097	96114	48387	48403	47554	47570	95884	95935	95884	95935
κ =3	95837	95863	48203	48227	47462	47485	95621	95698	95621	95681
κ =4	95423	**95456**	**48009**	**48040**	47238	47268	95211	**95313**	**95206**	**95266**
κ =5	**95419**	95462	48010	48049	47236	47275	**95193**	95321	95208	95285
κ =6	95429	95480	48068	48115	**47180**	**47227**	95202	95355	95216	95301
Newlyn										
κ =1	102334	102343	53032	53041	49027	49035	101860	101887	101860	101887
κ =2	101944	101962	52761	52777	48856	48872	101404	101458	101404	101458
κ =3	101369	101396	52018	52043	48830	48854	100622	100703	100747	100810
κ =4	100967	101003	51724	51757	**48679**	**48712**	100208	100315	**100249**	**100311**
κ =5	**100917**	**100962**	51633	51674	48695	48736	100166	100300	100253	100333
κ =6	100931	100984	**51620**	**51670**	48737	48787	**100093**	**100253**	100271	100361
Sheerness										
κ =1	62568	62576	33014	33021	29377	29385	62321	62346	62321	62346
κ =2	62551	62567	33017	33032	29344	29359	62290	62338	62290	62338
κ =3	62447	62472	32983	33005	29289	29311	62218	62290	62229	62286
κ =4	62210	**62242**	32866	**32896**	29169	29198	61977	**62074**	61989	62045
κ =5	62228	62268	32877	32914	29172	29209	61988	62109	62002	62075
κ =6	**62194**	62243	**32865**	32909	**29144**	**29188**	**61952**	62096	**61964**	**62045**

Table 5.4.2: AIC and BIC scores (rounded to the nearest integer) for the seasonal Gaussian copula model, assuming a 6th Markov process, with dependence parameters ρ_κ where $\kappa = 1, \ldots, 6$ for skew surges at Heysham, Lowestoft, Newlyn and Sheerness. The bold values show the minimised score for each model (i.e., the best value of κ) and the bold red/blue indicates the minimum AIC/BIC scores across all models.

T with standard Gaussian distribution function Φ and inverse distribution function Φ^{-1} using $T = \Phi^{-1}(U)$.

Under the Markov assumption, we use the joint distribution of observations as in expression (5.4.1) for model fitting. This means the $\{\hat{\rho}_{\kappa_S}\}$ parameter estimates are interpreted as the correlation of values separated by some lag κ_S. However, the estimate $\hat{\rho}_\kappa$ could be influenced by the correlation at shorter lags, i.e., $(\hat{\rho}_{\kappa-1}, \ldots, \hat{\rho}_1)$. Therefore, for interpretability, we instead focus on the conditional correlation parameters defined by

$$\rho_\kappa^{(C)} = \text{corr}(T_i, T_{i-\kappa} | T_{i-1}, \ldots, T_{i-(\kappa-1)}),$$

for $i = 1, \ldots, N$ and $\kappa = 1, \ldots, \kappa_S$. We find the forms of these conditional parameters using the expectation of the conditional distribution $T_i | T_{i-1}, \ldots, T_{i-\kappa_S}$, which can be written as a linear combination of the conditioning terms,

$$\mathbb{E}(T_i | T_{i-1} = t_{i-1}, \ldots, T_{i-\kappa_S} = t_{i-\kappa_S}) = \sum_{\kappa=1}^{\kappa_S} \rho_\kappa^{(C)} t_{i-\kappa},$$

where $\rho_\kappa^{(C)} = \phi_\kappa(\rho_1, \ldots, \rho_\kappa)$ for some function ϕ_κ. Here, we show how to get the exact form of the conditional distributions for the kth order Markov process of interest.

Let $(T_1, \ldots, T_N)$ be the standard normal marginal variables, assumed to follow a kth order Markov process with Gaussian copula and dependence parameters $(\rho_1, \ldots, \rho_{\kappa_S})$ for $\kappa_S \leq k$. So that the joint distribution is given by $T_i, \ldots, T_{i-\kappa_S} \sim \text{MVN}_{\kappa_S+1}(\mathbf{0}_{\kappa_S+1}, \Sigma_{\kappa_S+1})$ where $\mathbf{0}_{\kappa_S+1} = (0, \cdots, 0) \in \mathbb{R}^{(\kappa_S+1)}$ and

$$\Sigma_{\kappa_S+1} = \begin{pmatrix} 1 & \rho_1 & \cdots & & \rho_{\kappa_S} \\ \rho_1 & 1 & & & \vdots \\ \vdots & & \ddots & & \\ & & & 1 & \rho_1 \\ \rho_{\kappa_S} & \cdots & & \rho_1 & 1 \end{pmatrix} \in \mathbb{R}^{(\kappa_S+1)\times(\kappa_S+1)}. \tag{5.4.3}$$

Then $T_i|(T_{i-1} = y_1, \ldots, T_{i-\kappa_S} = y_{\kappa_S}) \sim N(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$, where

$$\bar{\boldsymbol{\mu}} = \Sigma_{1\kappa_S}\Sigma^{-1}_{\kappa_S\kappa_S}\boldsymbol{y} \quad \text{with} \quad \boldsymbol{y} = (y_1, \ldots, y_{\kappa_S}) \quad \text{and} \quad \bar{\Sigma} = \Sigma_{1\kappa_S}\Sigma^{-1}_{\kappa_S\kappa_S}\Sigma_{\kappa_S 1},$$

where the variance terms are defined as

$$\begin{aligned}
\Sigma_{1\kappa_S} &= \big(\mathrm{cov}(T_i, T_{i-1}),\ \cdots,\ \mathrm{cov}(T_i, T_{i-\kappa_S})\big) = (\rho_1, \ldots, \rho_{\kappa_S}), \\
\Sigma_{\kappa_S 1} &= \big(\mathrm{cov}(T_i, T_{i-1}),\ \cdots,\ \mathrm{cov}(T_i, T_{i-\kappa_S})\big)^T = (\rho_1, \ldots, \rho_{\kappa_S})^T,
\end{aligned}$$

where the power T denotes the transpose of the vector, and

$$\begin{aligned}
\Sigma_{\kappa_S\kappa_S} &= \begin{pmatrix}
\mathrm{var}(T_i) & \mathrm{cov}(T_i, T_{i-1}) & \cdots & \mathrm{cov}(T_i, T_{i-\kappa_S-1}) \\
\mathrm{cov}(T_i, T_{i-1}) & \mathrm{var}(T_{i-1}) & & \vdots \\
\vdots & & \mathrm{var}(T_{i-(\kappa_S-1)}) & \mathrm{cov}(T_i, T_{i-1}) \\
\mathrm{cov}(T_i, T_{i-\kappa_S}) & \cdots & \mathrm{cov}(T_i, T_{i-1}) & \mathrm{var}(T_{i-\kappa_S})
\end{pmatrix} \\
&= \begin{pmatrix}
1 & \rho_1 & \cdots & \rho_{\kappa_S-1} \\
\rho_1 & 1 & & \vdots \\
\vdots & & 1 & \rho_1 \\
\rho_{\kappa_S-1} & \cdots & \rho_1 & 1
\end{pmatrix}.
\end{aligned}$$

For example, consider a 2nd order Markov process and $\kappa_S = 2$, so with parameters (ρ_1, ρ_2), on Gaussian margins. Then the conditional distribution is given by

$$T_i|(T_{i-1} = t_{i-1}, T_{i-2} = t_{i-2}) \sim N\bigg(\frac{(\rho_1 - \rho_1\rho_2)t_{i-1} + (\rho_2 - \rho_1^2)t_{i-2}}{1 - \rho_1^2}, \\ 1 - \frac{1}{1 - \rho_1^2}(\rho_1^2 - 2\rho_1^2\rho_2 + \rho_2^2)\bigg),$$

so that the conditional correlation parameters are given by $\rho_1^{(C)} = (\rho_1 - \rho_1\rho_2)/(1 - \rho_1^2)$ and $\rho_2^{(C)} = (\rho_2 - \rho_1^2)/(1 - \rho_1^2)$.

Figures 5.4.1 and 5.4.2 show the estimates of the parameters $\hat{\rho}_\kappa$ and $\hat{\rho}_\kappa{}^{(C)}$, respectively, for $\kappa = 1, \ldots, \kappa_S$ for the best performing Gaussian copula models outlined in Section 5.4.2 for each site S. We illustrate how the parameters change with day of year d. Whilst the parameters for $\kappa > 2$ are fixed for *Model 3*, the corresponding conditional parameters will vary with d as they also depend on time-varying, lower order correlation parameters. Here the series is assumed to follow a kth order Markov process where $k = \kappa_S$ at each site S, rather than $k = 6$ everywhere as in Sections 5.4.1 and 5.4.2.

Figure 5.4.1 shows that $\hat{\rho}_1 > \max_{\kappa=2,\ldots,6}\{\hat{\rho}_\kappa\}$, as expected, so that the strongest dependence exists for adjacent observations. The parameter estimates typically decrease as the lag κ increases; when some parameters vary with day d, we find that $\max_{d\in[1,365]}\{\hat{\rho}_\kappa(d)\} < \max_{d\in[1,365]}\{\hat{\rho}_{\kappa\prime}(d)\}$ for any $\kappa < \kappa\prime$. Parameter estimates that vary with day d tend to peak in the central 6 months of the year, rather than the first/final 3 months; this agrees with parameter estimates from *Models 1W/S* and our exploratory analysis of Section 5.2.3. Newlyn is an exception to this since $\hat{\rho}_1(d)$ and $\hat{\rho}_2(d)$ peak in spring whilst $\hat{\rho}_\kappa(d)$ for $\kappa = 3, \ldots, 6$ peak in winter. Also at Newlyn, $\hat{\rho}_6(d)$ varies the most with day and this variability decreases as we decrease κ from 6. These results suggest fixing $\hat{\rho}_1(d)$ across time since there is little variability with d, however, when doing so we did not find an improvement in fit according to AIC and BIC scores.

Figure 5.4.2 shows that $\hat{\rho}_1^{(C)}(d) > \max_\kappa\{\hat{\rho}_\kappa^{(C)}(d)\}$ for all d and across all sites; $\hat{\rho}_1^{(C)}(d)$ also takes the greatest range of values compared to the remaining parameter estimates. For $\kappa > 3$, there is little variation with day d at all sites, except for Newlyn where there is still substantial variation for $\kappa = 4$ and 5.

5.5 Simulations

In this section, we simulate from the copula models identified as the best for each site. The procedure for doing so is outlined in Section 5.5.1. We first replicate patterns in

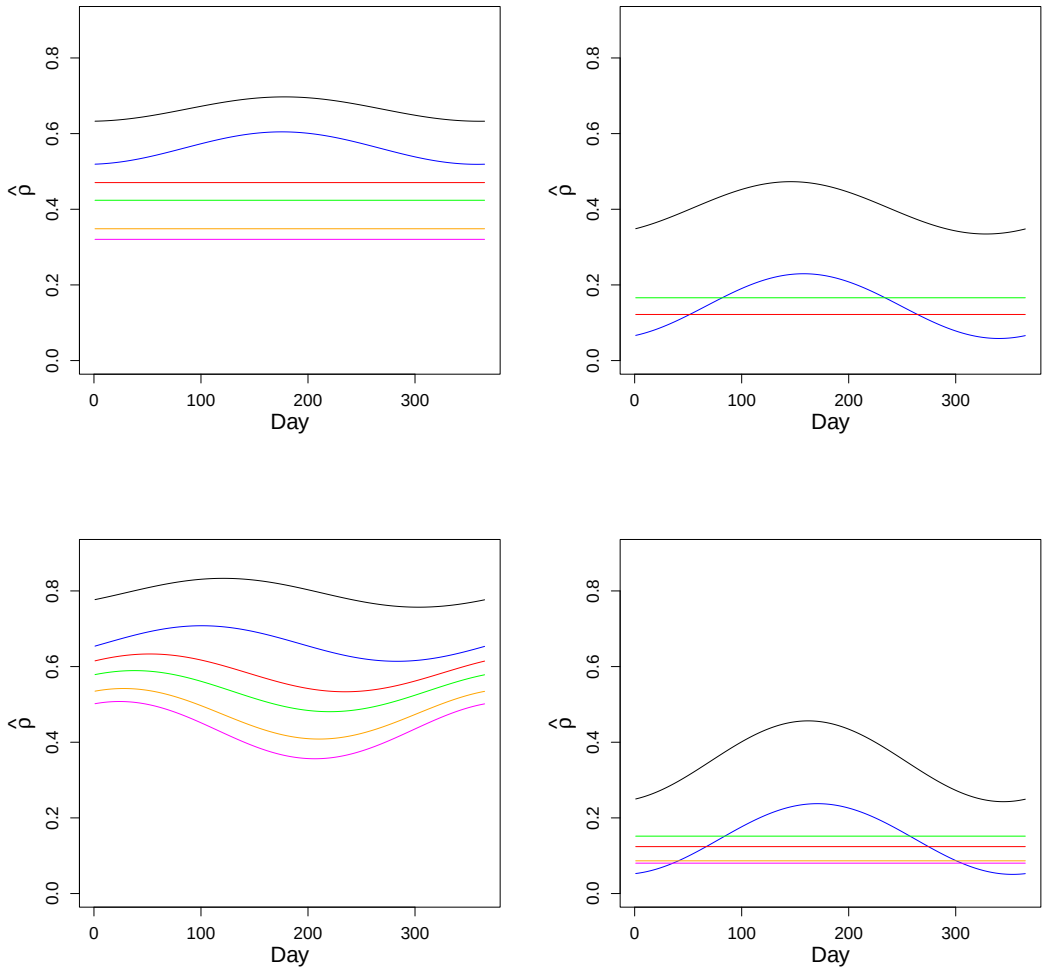


Figure 5.4.1: Dependence parameter estimates (y-axis) changing with day (x-axis) for the best fitting model at Heysham (top left), Lowestoft (top right), Newlyn (bottom left), Sheerness (bottom right): $\hat{\rho}_1$ (black), $\hat{\rho}_2$ (blue), $\hat{\rho}_3$ (red), $\hat{\rho}_4$ (green) and $\hat{\rho}_5$ (orange) and $\hat{\rho}_6$ (magenta).

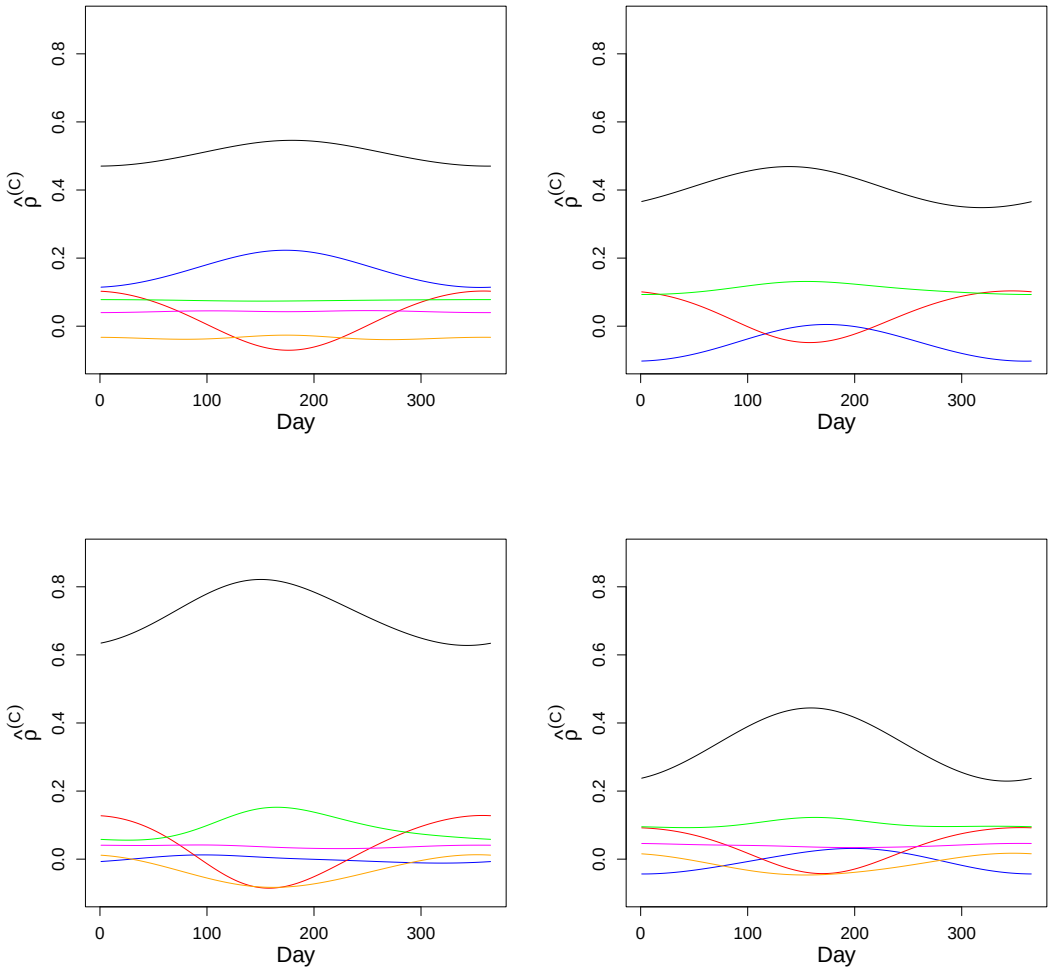


Figure 5.4.2: Conditional dependence parameter estimates (y-axis) changing with day (x-axis) for the best fitting model at Heysham (top left), Lowestoft (top right), Newlyn (bottom left), Sheerness (bottom right): $\hat{\rho}_1^{(C)}$ (black), $\hat{\rho}_2^{(C)}$ (blue), $\hat{\rho}_3^{(C)}$ (red), $\hat{\rho}_4^{(C)}$ (green) and $\hat{\rho}_5^{(C)}$ (orange) and $\hat{\rho}_6^{(C)}$ (magenta).

the historic data, so that the covariates previously observed (i.e., the observed tidal regime) are used for simulations; this allows comparison with the true, observed data. These comparisons are given in Section 5.5.2.

5.5.1 Simulation procedure

We describe the procedure for simulating a series of length $n \in \mathbb{N}$ (equal to the length of the observed data), denote this sequence by $\{\tilde{Y}_t\}_{t=1}^n$. Let $\hat{\rho}_1, \ldots, \hat{\rho}_{\kappa_S}$ denote the fitted dependence parameters of the best fitting Gaussian copula model of Section 5.4.2 for any site S; note that some of these may depend on day in year d but we omit this in the notation for simplicity. We first aim to simulate a stationary Markov process with order $k = \kappa_S$ on standard Gaussian margins. To do so, we simulate $(T_1, \ldots, T_k)$ from the joint (multivariate) normal distribution given by expression (5.4.3) using parameter estimates $(\hat{\rho}_1, \ldots, \hat{\rho}_{\kappa_S})$. Then for each T_t where $t \geq \kappa_S$, we simulate from the conditional (univariate) normal distribution detailed in Section 5.4.3, using conditional parameter estimates $(\hat{\rho}_1^{(C)}, \ldots, \hat{\rho}_{\kappa_S}^{(C)})$.

Then we have a sequence of stationary standard Gaussian random variables $\{T_t\}_{t=1}^n$ that follow a κ_Sth order Markov process with a Gaussian copula to describe the temporal dependence. Next, we use the PIT to transform the simulations from Gaussian to uniform margins and then to the margins of the skew surge model in Chapter 3 to reflect non-stationarity and peak tide dependence; we refer to this as *skew surge* margins. We use normal distribution function Φ to obtain Uniform(0,1) random variables $U_t = \Phi(T_t) \sim$ Uniform(0,1) for all $t = 1, \ldots, n$. Then to obtain the simulations on *skew surge* margins, we invert the distribution function of equation (5.3.3) and denote this $Q_Y^{(d,j,x)}$, i.e., the quantile function, so that $\tilde{Y}_t = Q_Y^{(d,j,x)}(U_t)$ is our final simulated series.

5.5.2 Simulating historic data

We simulate $B = 200$ realisations of length n using the historic tidal regime and denote each series as $\{\tilde{Y}_t^{(b)}\}_{t=1}^n$ for $b = 1, \ldots, B$. We compare these simulations with the observed data, in terms of seasonality and temporal dependence in the body and tail of the data. We are particularly interested in the extremal dependence of our simulations, which we assess by estimating the extremal index θ (via the runs estimate, with run length 10) and measures of extremal dependence, χ_τ and $\bar{\chi}_\tau$ for $\tau = 1, \ldots, 5$; all measures are defined in Section 5.2.1.

Figure 5.5.1 compares a traceplot of observations, on a day-in-year temporal resolution, with simulations from four samples for a single year. This shows that the simulated seasonal and dependence structure is similar to the observed series, but we investigate this further. To study the seasonality in more detail, we compare monthly boxplots of the observations and simulations; Figure 5.5.2 shows this for one simulation sample at Heysham. This demonstrates that we have captured the seasonal structure well, and have simulated observations bigger than those observed within the data. At the remaining sites, we are also able to reproduce the seasonal structure but omit the supporting figures as this is not the main focus of this chapter.

Next, we investigate temporal dependence in our simulations. Figure 5.5.3 compares autocorrelation function (acf) values at all sites for the observations and one realisation of simulations; although the results are similar across our 200 samples. Some dependence will be attributable to seasonal variations since we have not transformed to uniform margins here, as we did in Section 5.2.3. At all sites, we capture the site-specific dependence behaviour across lags and our simulations have almost identical acf values at Heysham, Lowestoft and Sheerness, especially at the earlier lags. However, at Newlyn, acf values for the simulations are higher than the observed data at almost all lags; this difference is most noticeable for lags 10-20.

To study extremal dependence, we first compare empirical runs estimates of the

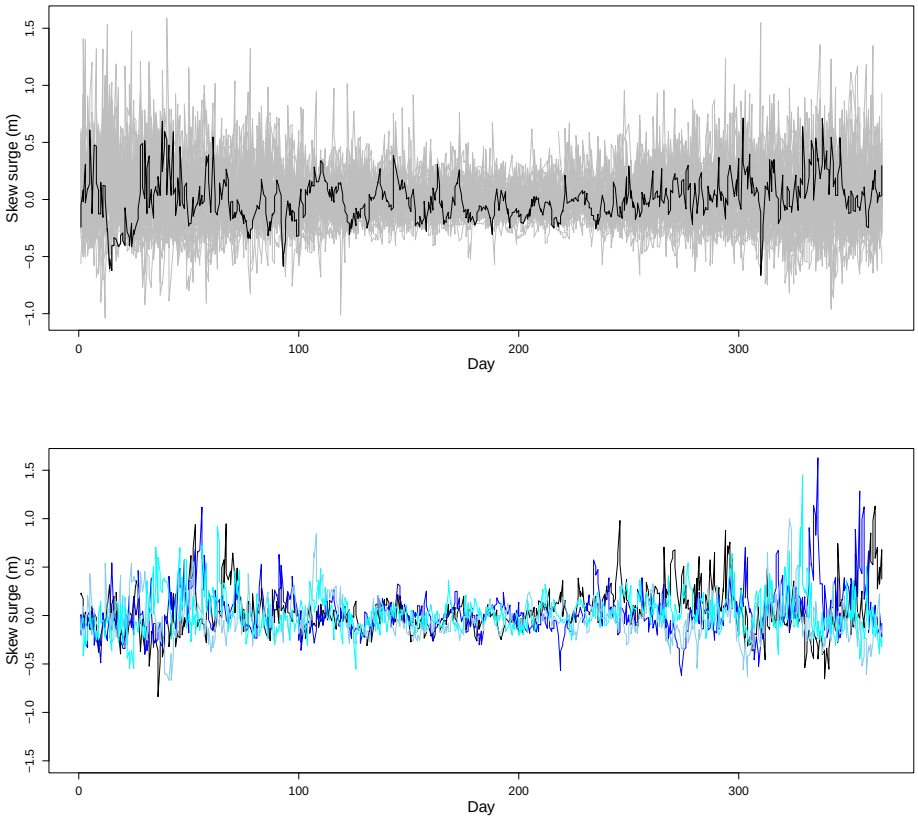


Figure 5.5.1: Top: Traceplots of observed skew surges (y-axis in metres) across the year (x-axis, in terms of day in year d) for a given year (black) and all year years (grey) at Heysham. Bottom: Traceplots of a single year (x-axis) from four different simulated series (indicated by four different colours) of skew surges (y-axis in metres) at Heysham.

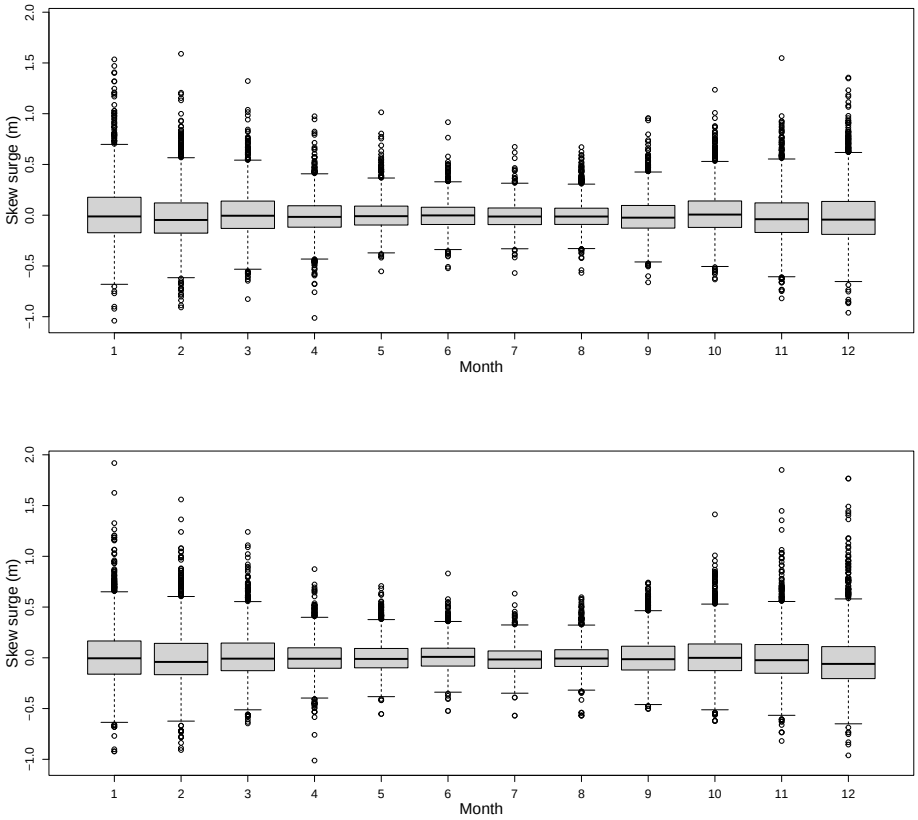


Figure 5.5.2: Monthly boxplots of skew surge observations from observations (top) and one realisation of simulations (bottom) at Heysham.

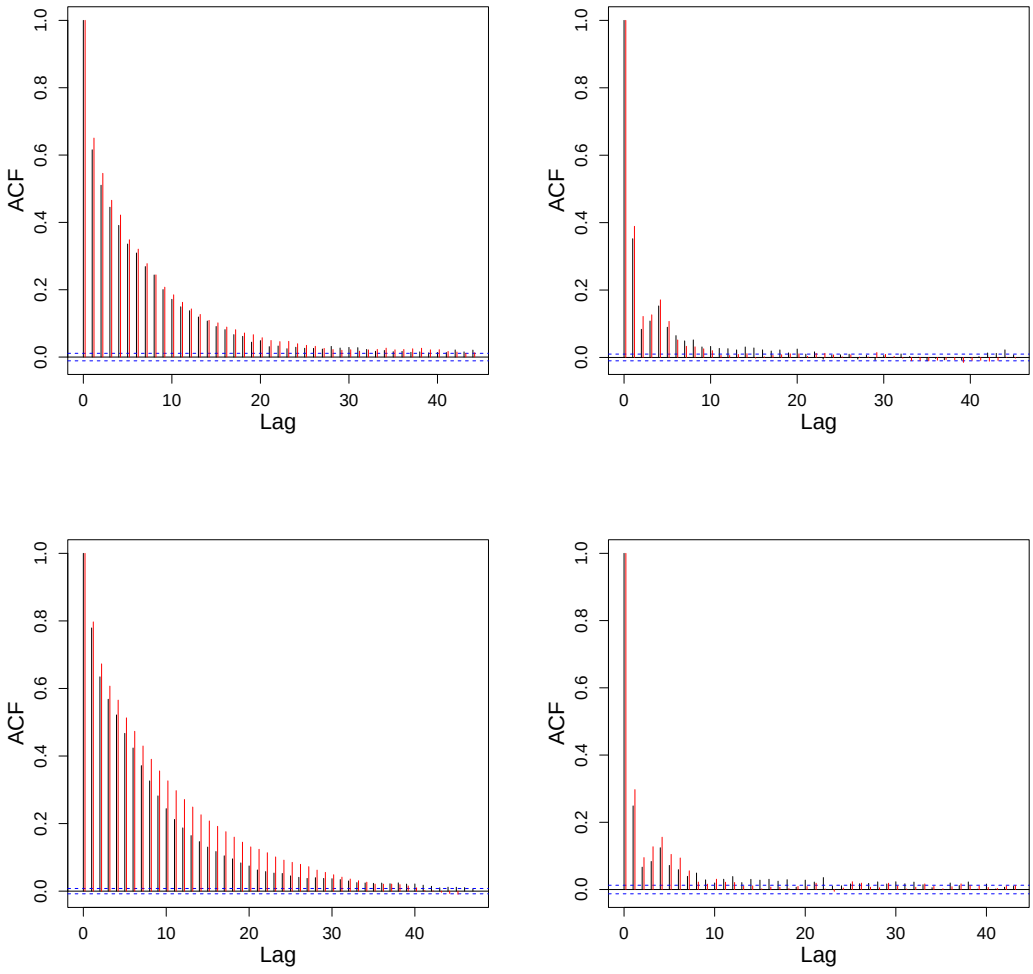


Figure 5.5.3: Autocorrelation function (acf) plots of skew surge observations (black) and one realisation of our simulations (red) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right) on the original data scale.

extremal index θ across different quantiles in Figure 5.5.4. At Heysham, Newlyn and Sheerness, we obtain good agreement between estimates for the observed series and over our 200 simulations, with the observed estimates falling within the range of the simulated ones. Although for the higher quantiles, we obtain a slight underestimate in the simulated series at Heysham and Newlyn, but a slight overestimate at Sheerness. The boxplots at these three sites lie within the 95% confidence intervals for the empirical estimates, found using a bootstrap procedure. At Lowestoft, the overestimation of θ for our simulations is more significant, with no overlap between the confidence intervals on the empirical estimates and the range on the boxplots for the simulated estimates.

Figures 5.5.5 and 5.5.6 compare empirical estimates of χ_1 and χ_5 across different thresholds, for the observed data, with 95% confidence intervals found via the delta method, and over all 200 simulated series for each site. Overall, estimates of χ_5 are better matched across the data sets than those for χ_1. Estimates of both χ_1 and χ_5 are most similar for observations and simulations at Heysham and Newlyn. Whilst at Lowestoft and Sheerness, our simulations give lower estimates of χ_1 and χ_5 than the observed data, especially for the lower quantiles. Although, at the higher quantiles, there is an overlap in the confidence intervals on the empirical estimates and the range on the boxplots for the simulated estimates. This means our simulations will have weaker extremal dependence than the observed data. We also show the same results for estimates of $\bar{\chi}_5$ in Figure 5.5.7, and make the same conclusions; there are more similarities with the estimates at Heysham and Newlyn, but the simulated series have lower estimates at Lowestoft and Sheerness. Overall, we capture the extremal dependence structure well at all sites but discuss areas for improvement in Section 5.6.

Since Sheerness is located close to the Thames Barrier, we use these simulations to estimate historic closure rates of the barrier and compare these to closure rates from the observed data. The decision to close the barrier is guided by a matrix based on forecast water levels at Southend (a tide gauge not on the UK National Tide Gauge

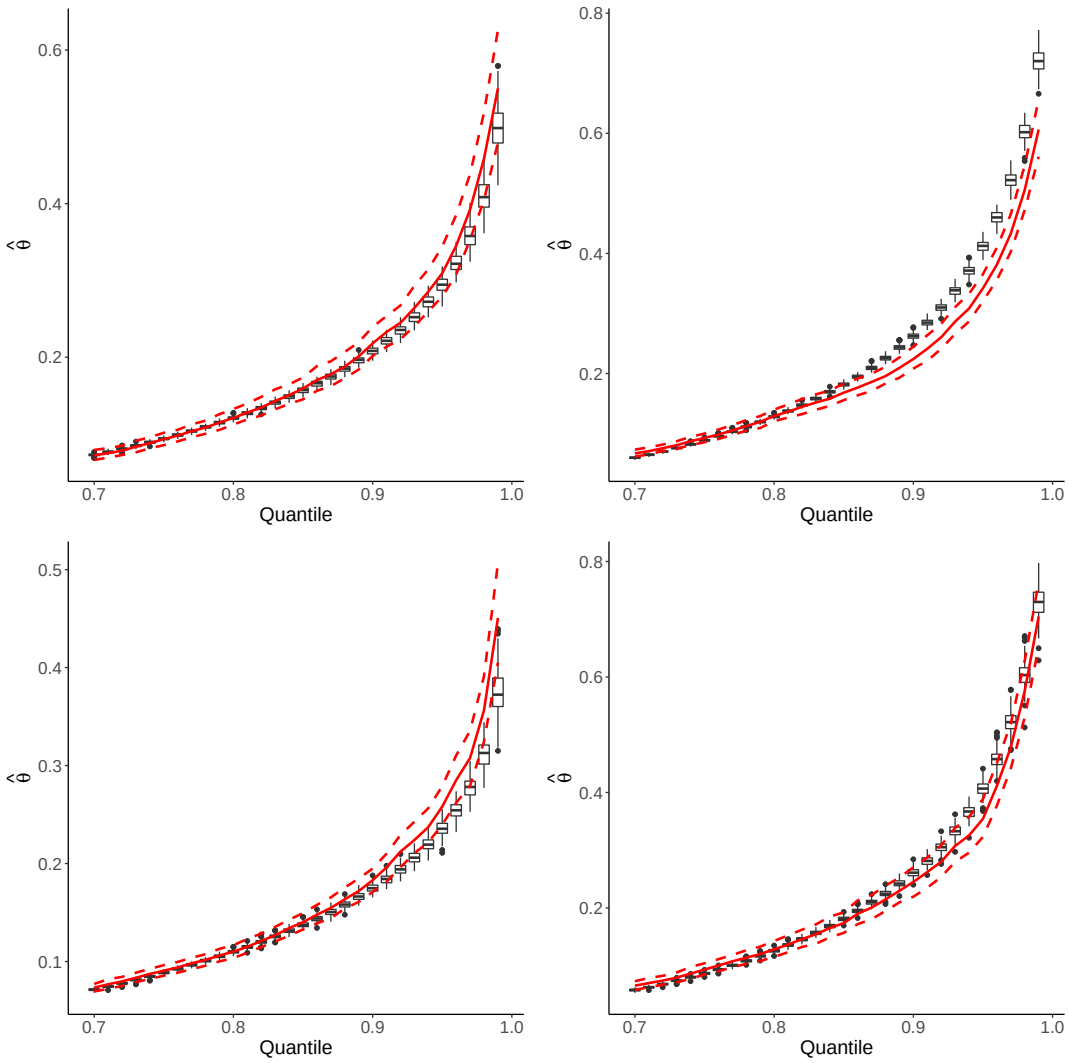


Figure 5.5.4: Empirical estimates of the extremal index θ with the threshold set as different quantiles (x-axis) of the data set of interest, for the observed data (red line), with 95% confidence intervals (red-dashed lines), and over 200 simulations (box plots) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

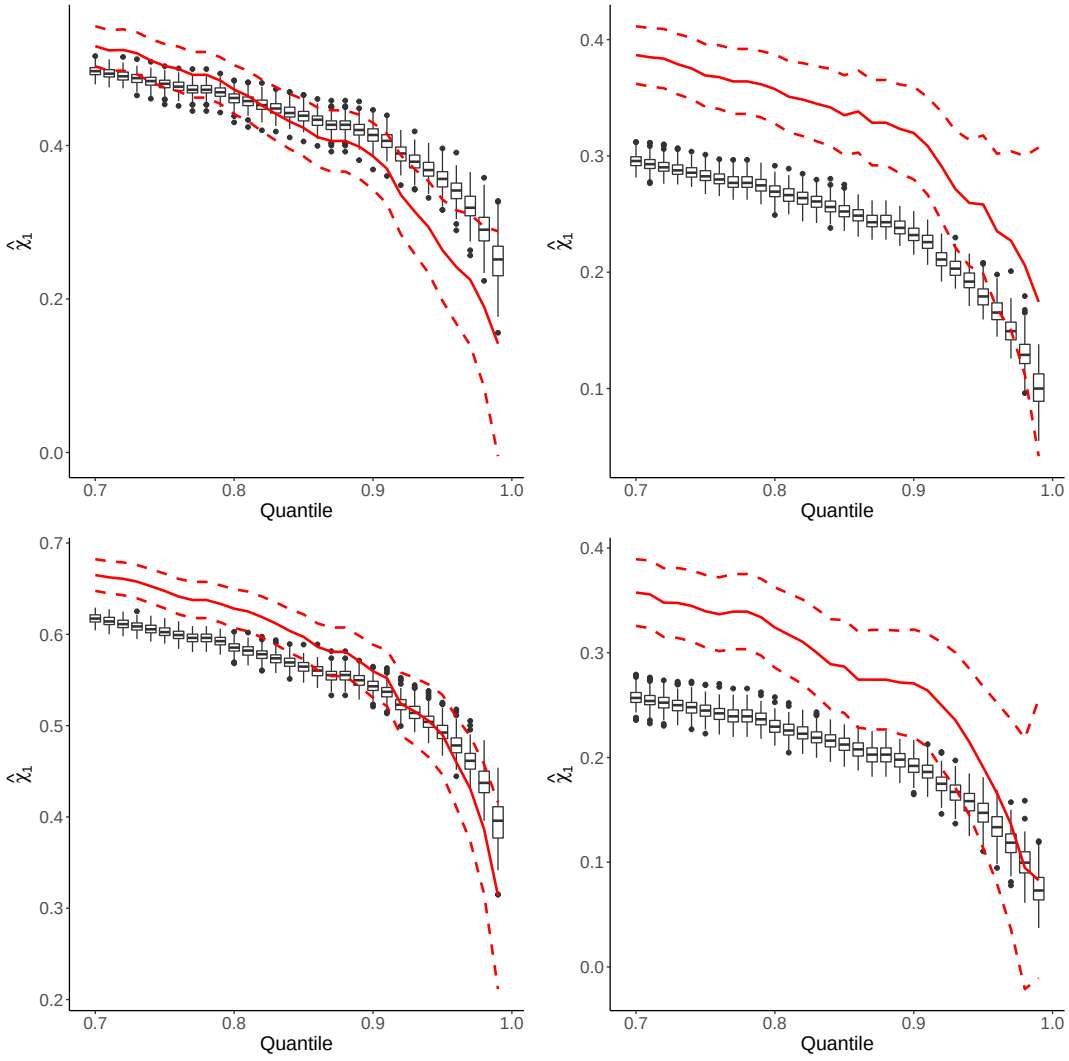


Figure 5.5.5: Empirical estimates of χ_1 with the threshold set as different quantiles (x-axis) of the data set of interest, for the observed data (red line), with 95% confidence intervals (red-dashed lines), and over 200 simulations (box plots) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

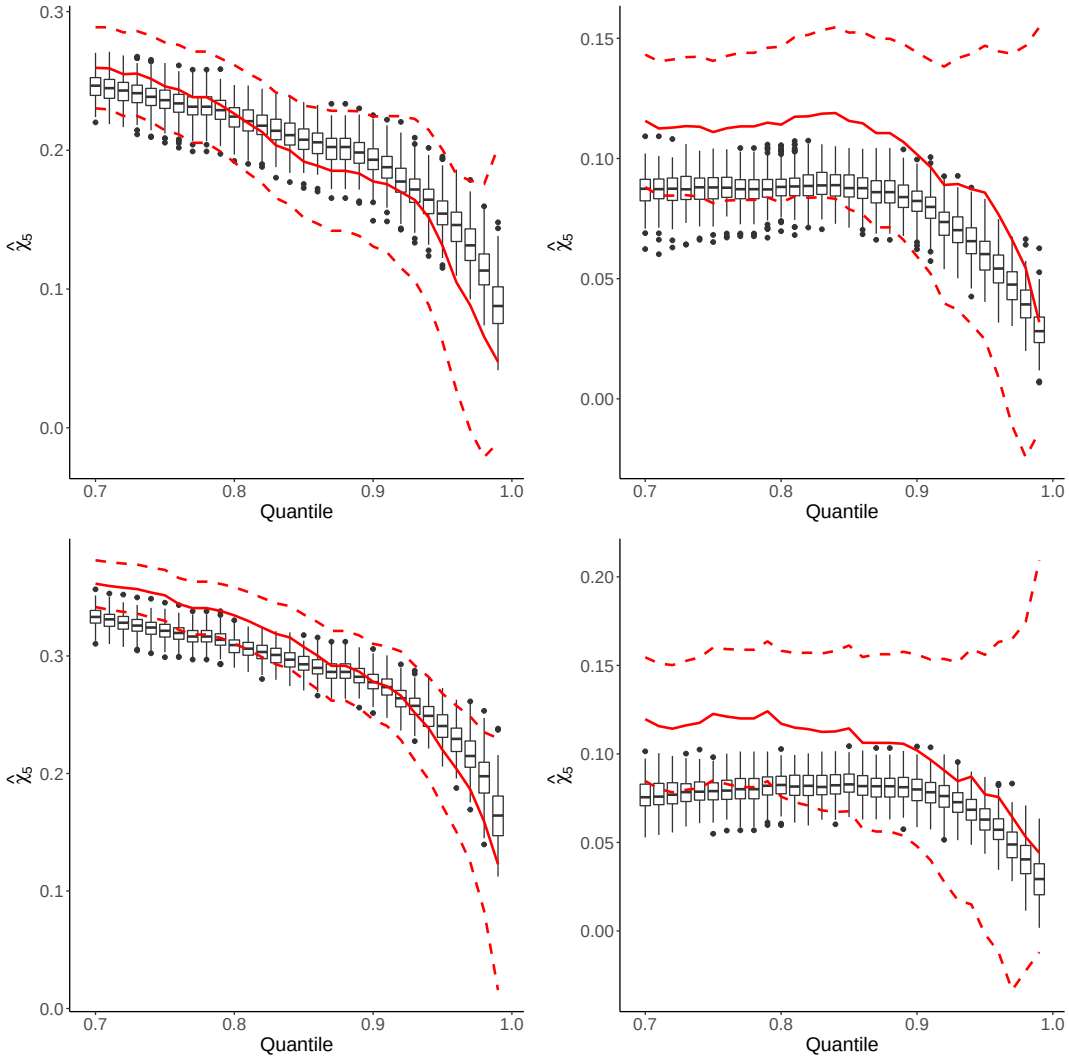


Figure 5.5.6: Empirical estimates of χ_5 with the threshold set as different quantiles (x-axis) of the data set of interest, for the observed data (red line), with 95% confidence intervals (red-dashed lines), and over 200 simulations (box plots) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

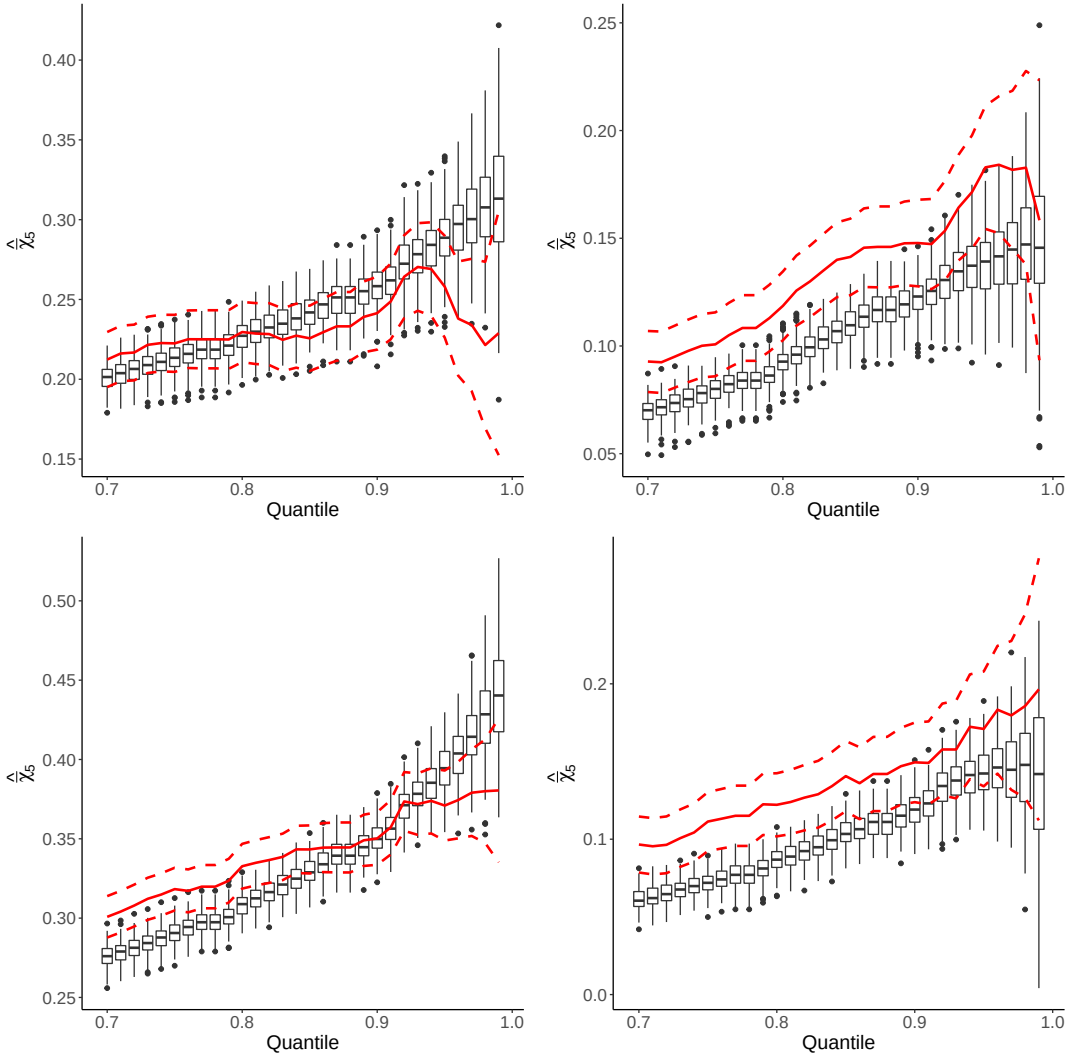


Figure 5.5.7: Empirical estimates of $\bar{\chi}_5$ with the threshold set as different quantiles (x-axis) of the data set of interest, for the observed data (red line), with 95% confidence intervals (red-dashed lines), and over 200 simulations (box plots) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

Network, located ∼6 miles across the estuary from Sheerness) and river flow measured at Teddington Weir. The closure matrix is confidential, so the true values are unknown to us. Here, we use data from Sheerness as opposed to Southend and focus on closures from water levels. We can add the known, historic tidal regime onto our simulated series to obtain synthetic water level series for the period 1980-2022, without trends in mean sea level which would be included in practice.

We compare the closure rates using simulations from our proposed method, where temporal dependence is accounted for in the entire range of data, with those from the marginal skew surge model of Section 5.3.1 where temporal dependence is ignored. Figure 5.5.8 shows the estimated number of barrier closures, if the threshold for closure is 3.5m (relative to the mean sea level), from the observed data and 500 replicates of simulated series from both models (i.e., with and without temporal dependence). We say that exceedances of this threshold separated by less than 6 tidal cycles (since $\kappa_{SHE} = 6$) correspond to a single barrier closure. We find that the true number of closures over the observation period under this threshold is 49. Our proposed model replicates the closure rates well across our simulations, with the estimates centred around the true value. However, simulating from the model of Section 5.3.1, where temporal dependence is not accounted for, tends to overestimate compared to the observed data. Overestimation means that closures will be planned more regularly than required, taking up valuable time needed for barrier maintenance. Our proposed model estimates more than 60 closures just 5% of the time, whilst ignoring temporal dependence does so 19% of the time. Additionally, Figure 5.5.8 shows that our proposed method estimates the correct closure rates more often than when temporal dependence is ignored.

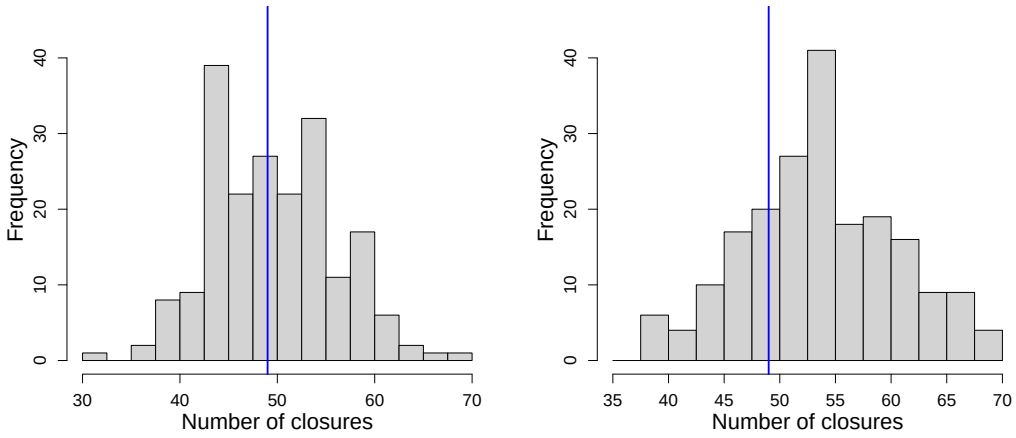


Figure 5.5.8: Histogram of the number of barrier closures, for a closure threshold of 3.5m, across the simulated samples at Sheerness, with the corresponding number of closures in the observed data indicated by the blue line, for our proposed model (left) and using the marginal model of Section 5.3.1 where temporal dependence is ignored (right).

5.6 Discussion

We have presented a procedure for simulating skew surges that replicates their seasonality, temporal dependence and extreme values, and provide simulations at four UK National Tide Gauge Network sites. These simulations are useful for coastal management; where understanding the temporal dependence is fundamental as prolonged storm events that span multiple tidal cycles accelerate erosion rates. Our simulations can also be useful for predicting the life expectancy of a surge barrier, such as the Thames Barrier, as future simulations provide information about future barrier closure rates. The novelty of our approach lies with modelling the temporal dependence of skew surges using a copula framework, assuming the data follows a Markov process. We found that the temporal dependence structure is not constant throughout the year, with stronger dependence found in the summer months, and we captured this using harmonics for the copula dependence parameters.

We used the Gaussian copula for modelling the temporal dependence as this can

capture the asymptotic independence structure exhibited by skew surges. We chose this copula as it has a single dependence parameter in the bivariate case, however, other choices are available: the Frank, Clayton and Inverted Logistic copulas capture asymptotic independence through a single parameter also (Nelsen, 2006; Joe, 2014). Further work could investigate these copulas and compare them against the Gaussian copula, to investigate if they improve our simulation results.

When incorporating non-stationarity into the dependence parameter of the Gaussian copula, we used harmonics as these provide an easily interpretable and parsimonious way to capture smooth changes throughout the year. However, alternative approaches could be taken such as using generalised additive models (Chavez-Demoulin and Davison, 2005) or splines (Youngman, 2019).

For choosing the order k of the Markov process we used AIC and BIC, however, we found that these information criteria disagree in some instances. An alternative approach could be used to select the order k. For example, using the partial autocorrelation function (pacf) to identify the largest lag at which the pacf differs from zero, however, this focuses on dependence in the body and not the tail. Winter and Tawn (2017) present a diagnostic approach similar to the threshold stability plots for threshold selection in univariate extreme value analysis (Coles, 2001), but using their derived cluster functionals.

Uncertainty quantification is a crucial part of statistical modelling of any kind. By replicating simulations over 200 samples, we obtained measures of uncertainty for our simulations that will be useful in practice. However, it would be insightful to have confidence intervals for the dependence parameters given in Figures 5.4.1 and 5.4.2. This can be done using a block bootstrap procedure to preserve the temporal dependence (Politis and Romano, 1994), however, care would need to be taken in choosing the block length as if this is too small, the dependence in the series would be disrupted. Alternatively, since likelihood inference is used, the asymptotic normality of the maxi-

mum likelihood estimates could be exploited to obtain confidence intervals based on the Hessian. Obtaining confidence intervals would be a necessary step in the next stages of this work; identifying when confidence intervals overlap or contain zero could help simplify the model to achieve parsimony by sharing information across ρ_κ for $\kappa = 1, \ldots, \kappa_S$ or setting $\rho_\kappa = 0$, respectively.

In Chapter 4, we shared information across sites regarding the parameter estimates associated with increases in global mean temperature anomaly (GMT). A similar approach could be taken here to pool information across sites that share temporal dependence characteristics. For example, Sheerness and Lowestoft are both located on the east coast and are likely to be affected by similar storms, therefore we expect them to have a similar dependence structure. Our exploratory analysis of Section 5.2.3 demonstrated this, as similar estimates of Kendall's τ, χ_κ and $\bar{\chi}_\kappa$ were found. When fitting the Gaussian copula models in Section 5.4.2, we obtain different values of κ_S, however the parameter estimates shown in Figure 5.4.1 are very similar for $\left\{\rho_\kappa : \kappa = 1, \ldots, \min\{\kappa_{LOW}, \kappa_{SHE}\}\right\}$. This suggests that sharing spatial information would be sensible, and this would reduce the uncertainty associated with dependence parameter estimation. However, we only study four sites here so continuing with the single site analysis for a larger set of sites located nearby would be a necessary first step.

In Section 5.5.2 we simulated data corresponding to a historic period but in practice, simulations of future records are more interesting. Using the predicted tidal regime, it is straightforward to use the procedure outlined in Section 5.5.1 however, our marginal model of Section 5.3.1 assumes a steady state climate which is, of course, unrealistic. Increasing trends in mean sea level can be added back onto the simulations for historic simulations, and for future simulations, predicted trends under different climate scenarios can be used. As identified in Chapter 4, there exist longer term trends in extreme skew surges that differ from those in the mean; we capture these by adding a GMT

covariate into the GPD model of Section 5.3.1. Then, simulations can be obtained under different scenarios of future GMT increases.

Chapter 6

Extremal properties of max-autoregressive moving average processes for modelling extreme river flows

6.1 Introduction

Any stationary time series that exhibits large peaks or sudden bursts of extreme observations is a candidate for modelling by a max-autoregressive moving average, Max-ARMA(p, q), process for $p \in \mathbb{N}$ and $q \in \{\mathbb{N} \cup 0\}$ (Davis and Resnick, 1989). This process is a non-linear version of the well-known ARMA models (Box and Jenkins, 1970) where a maxima replaces the summation. Max-ARMA(p, q) models are suitable for data with shock noise behaviour, where the process broadly descends exponentially from each spike and then fluctuates around small values until the next major spike.

River flow is an example where a Max-ARMA process is a suitable candidate for modelling its behaviour, particularly in extreme states. A large precipitation event

can cause river levels with sizeable catchments to remain high for days after, as the spatially distributed large volume of water takes time to propagate down the river and travel through underground systems. By adjusting the order and parameter values of the process, the frequency, width and detailed structure of the spikes can be altered. Figure 6.1.1 (left) shows daily maximum river flow trace plots from the River Thames (gauge at Kingston-upon-Thames) during four different winter seasons. Two of the seasons shown correspond to the worst flood events on record for the Thames: November 1894 and February 1928. The plot shows that river flow has features of heavy tails and non-linearity, with rapid rises and slower falls around the peaks. Figure 6.1.1 (right) shows estimates of the asymptotic dependence coefficients χ_κ (see Section 6.3.2) at different lags $\kappa \in [1, 100]$, which provides a similar type of information as for an autocorrelation function, but here with a metric to measure extremal dependence across time. The estimated χ_κ values demonstrate strong temporal dependence until approximately lag 14, corresponding to 2 weeks, after which the dependence steadily decays. Estimates of χ_κ vary with the threshold used to define an extreme event, with illustrations of estimates using three possible thresholds (see Section 6.4 for further discussion of these estimates). Figure 6.1.1 (right) also shows Pearson's correlation coefficient over time lags κ to illustrate the differing dependence structure in the body and tail of the data; dependence is stronger in the body until approximately lag 90. Similar to the estimates of χ_κ, there is a change in the rate that the dependence measure decays beyond lag 14, after which the correlation coefficient decays at a steadier rate than for $\kappa \leq 14$, although the decay is much quicker than for χ_κ across all lags.

Modelling sea levels and river flow is fundamental for forecasting future closures of flood barriers, such as the Thames Barrier, in both the short- and long-term. Short-term predictions are fundamental for anticipating closure times and durations to ensure preparedness (Dale et al., 2014; Environment Agency, 2021). We are interested in longer-term forecasts to explore the impacts of climate change on the number of barrier

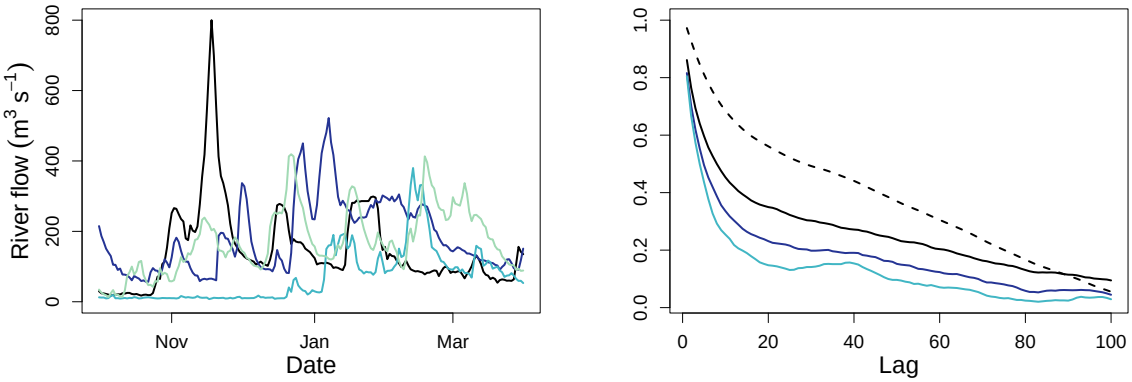


Figure 6.1.1: Left: River flow time series of the River Thames for the winter season (October - March) in 1894/95 (black), 1927/28 (dark blue), 1973/74 (light blue) and 2019/2020 (green) (Environment Agency, 2023a). Right: Pearson's correlation coefficient (dashed line) and empirical estimates of $\chi_\kappa(u)$ (solid lines; see Section 6.3.2) for the Thames data over different lags, κ (in days). Quantiles of $u = 0.9$ (black), 0.95 (dark blue) and 0.975 (light blue) are used for estimating $\chi_\kappa(u)$.

closures. Increases in barrier closures, in turn, require an increase in the number of maintenance and safety checks. This affects the reliability of the barrier because during these checks, the barrier cannot be used and, due to time constraints, the number of checks per year is restricted. Under climate change conditions, the number of times the barrier can close each year is likely to be exceeded soon (Environment Agency, 2021). Further measures must be taken to ensure the barrier can remain in operation, but if this is not possible, to plan for barrier updates or replacement. Therefore, by forecasting changes in future closure rates in the long term, one can provide critical information to barrier operators about barrier management and provide insight into its life expectancy (Trace-Kleeberg et al., 2023).

The barrier closes when sea levels and/or river flows exceed pre-determined, confidential, thresholds. Therefore, it is fundamental that we focus on modelling the extremes of both variables as these values are more likely to contribute to barrier closures. A novel method for estimating extreme sea levels was recently developed by D'Arcy et al. (2023b) for tide gauge data at Sheerness, located at the mouth of the River

Thames. Using a Max-ARMA(p, q) process, our aim is to develop a similar method for estimating extreme river flow. Simulations for river flow and sea levels can then be combined to determine closure rates, however, one would need to investigate if these are, in fact, independent (Svensson and Jones, 2004; Hendry et al., 2019). Of course, the Max-ARMA(p, q) model for river flow is not restricted to flood barrier and estuary locations, but can be applied to all river flow data in no-drought areas.

Processes with Max-ARMA behaviour exhibit unique temporal dependence structures that we aim to explore at extreme levels. Davis and Resnick (1989) derive conditions on the parameters of a Max-ARMA(p, q) process for it to be stationary and under which each of the parameters of the model is identifiable. We derive further conditions for all parameters to be identifiable and present a simplified parameter space that is useful for inference. We are interested in deriving the well-known extremal index θ (Leadbetter et al., 1983) and the coefficient of asymptotic dependence χ_κ at lags $\kappa \in \mathbb{Z}$; the latter is a special case of the extremogram defined by Davis and Mikosch (2009). The Max-ARMA(1,0) process has been studied previously: Robinson and Tawn (2000) and Ferreira (2011) derived the form of the θ and χ_κ, respectively. However, to the best of our knowledge, these measures have not yet been derived for a general Max-ARMA(p, q) process. We illustrate how these measures can be used for moments-based inference using the River Thames data as an example.

Max-ARMA(p, q) processes are defined on specific margins, for example, we use unit Fréchet margins. This can be thought of as in copula theory where the margins can be changed using the probability integral transform (Nelsen, 2006). Therefore we need an additional transformation from the observed series $\{Y_t\}$ to the Max-ARMA(p, q) series $\{X_t\}$, via $X_t = T(Y_t)$, for all $t \in \mathbb{Z}$, where T is defined in Section 6.5.3. This transformation allows us to compare the observed series in their extreme states, such as those in Figure 6.1.1, with simulated Max-ARMA(p, q) realisations.

In Section 6.2 we formally define the Max-ARMA(p, q) process and derive conditions

on its parameters for this to be an identifiable process. In Section 6.3 we detail the form of the extremal index and the coefficient of asymptotic dependence. In Section 6.4 we explain how to simulate from a Max-ARMA process and provide examples of such simulations. Davis and Resnick (1989) find super-efficient estimators of the parameters for known (p, q); in Section 6.5 we take a different inference approach, as we believe that the Max-ARMA(p, q) process is only an approximation to the true generating process (of river flows) and when they are in an extreme state. We propose a generalised moments-based inference procedure for the joint behaviour of exceedances of a high threshold. In Section 6.6 we illustrate this inference method on data from the River Thames. All proofs are given in Section 6.7. We conclude with a discussion in Section 6.8.

6.2 Model definition and parameter constraints

Davis and Resnick (1989) define a discrete-time stochastic process $\{X_t\}$ for $-\infty < t < \infty$ as following a Max-ARMA(p, q) model if,

$$X_t = \max\{\alpha_1 X_{t-1}, \ldots, \alpha_p X_{t-p}, \beta_0 Z_t, \beta_1 Z_{t-1}, \ldots, \beta_q Z_{t-q}\}, \tag{6.2.1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_q)$ are parameters of the model, for $p \in \mathbb{N}$ and $q \in \{0, \mathbb{N}\}$, and $\{Z_t\}$ is an independent and identically distributed (IID) innovation process on Fréchet(γ) margins, so that $F_Z(z) = \exp(-\gamma/z)$ for $z > 0$ and $\gamma > 0$. For the model to be well-defined the parameters must satisfy the constraints $\alpha_i \geq 0$ for $i = 1, \ldots, p-1$ and $\alpha_p > 0$, $\beta_0 = 1$, $\beta_j \geq 0$ for $j = 1, \ldots, q-1$ and $\beta_q > 0$. Despite the restriction on β_0, we use the more general β_0 notation throughout for mathematical convenience for expressing results. By construction (6.2.1), X_t and Z_s are independent for all $s > t$.

We make further restrictions on the parameters α_i for $i = 1, \ldots, p$ to ensure the process is stationary and derive the form of γ, the scale parameter for the innovation

process, such that $\{X_t\}$ has a specific marginal form, in the remark below.

Remark 6.2.1. *A Max-ARMA(p, q) process is a stationary process with Fréchet(σ) marginal distribution if and only if* $0 \leq \alpha_i < 1$ *for* $i = 1, \ldots, p-1$ *and* $0 < \alpha_p < 1$*, and the scale parameter* $\gamma > 0$ *of the Fréchet distribution for the innovation process* $\{Z_t\}$ *is given by*

$$\gamma := \sigma \left(\sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau} \left\{ \beta_j \prod_{\substack{i=1,\ldots,p: \\ \alpha_i > 0}} \alpha_i^{a_i} \right\} \right)^{-1} < \infty, \tag{6.2.2}$$

for $0 < \sigma < \infty$*, where*

$$\mathcal{S}_\tau = \left\{ (i, j, a_i) \in \{1, \ldots, p\} \times \{0, \ldots, q\} \times \{0, 1, \ldots, \tau\} : \sum_{i=1}^{p} i a_i + j = \tau \right\}. \tag{6.2.3}$$

Remark 6.2.1 is a reformulation of the result in Davis and Resnick (1989) into a simpler expression, and we explicitly state the marginal distribution for $\{X_t\}$. The formulation of Davis and Resnick (1989) means that the Max-ARMA parameters determine both marginal and dependence parameters, whereas our standardisation of the margins, through the choice of γ, makes the dependence parameters independent of the marginal distribution, which is particularly useful for copula style inferences, which is required in Section 6.5.2. Stationarity requires no further restrictions on $\boldsymbol{\beta}$. From this point onwards in our theoretical developments, we take $\sigma = 1$.

For a Max-ARMA(1,0) process, defined by $X_t = \max\{\alpha_1 X_{t-1}, Z_t\}$, to have unit Fréchet margins with stationary constraints given in Remark 6.2.1, i.e., $0 < \alpha_1 < 1$, we obtain that $0 < \gamma < 1$, with

$$\gamma = \left(\sum_{\tau=0}^{\infty} \alpha_1^\tau \right)^{-1} = 1 - \alpha_1. \tag{6.2.4}$$

Next, we impose further restrictions on $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, to ensure that the parameters are identifiable, i.e., in the sense that they have some effect on the feasible sample paths of the $\{X_t\}$ process. Remark 6.2.2 presents a result by Davis and Resnick (1989), with a

minor adaptation to allow for elements of $\boldsymbol{\alpha}$ to be zero.

Remark 6.2.2. *For a stationary Max-ARMA(p, q) process with $p \geq 2$, α_k is identifiable only if*

$$\max_{\mathcal{R}_k} \left\{ \prod_{\substack{i=1,\ldots,(k-1): \\ \alpha_i > 0}} \alpha_i^{a_i} \right\} < \alpha_k < 1,$$

for $k = 2, \ldots, p$, where

$$\mathcal{R}_k = \{ \, (i, a_i) \in \{1, \ldots, (k-1)\} \times \{0, 1, \ldots, k\} : \sum_{i=1}^{k-1} i a_i = k \}.$$

Since $\beta_0 = 1$, it is always identifiable. For $j \in 1, \ldots, \min\{p, q\}$, β_j is identifiable if $\beta_j > \alpha_j$; and when $q > p$ if $\beta_j > 0$ it is identifiable if $j \in \{p+1, \ldots, q\}$.

Note that $\mathcal{R}_k$ is defined similarly to $\mathcal{S}_\tau$ in expression (6.2.3) for a Max-ARMA$(k-1, 0)$ process, so that there are no $\boldsymbol{\beta}$ parameters. In the following proposition, we extend Remark 6.2.2, which gives a condition for the marginal identifiability of α_i, to derive a condition under which all α_i for $i = 1, \ldots, p$ are identifiable. We prove this result in Section 6.7.

Proposition 6.2.3. *For a stationary Max-ARMA(p, q) process with $p \geq 2$, if α_i for all $i = 1, \ldots, (k-1) < p$ are identifiable, then α_k is identifiable if*

$$\max_{i=1,\ldots,\lfloor k/2 \rfloor} \{\alpha_i \alpha_{k-i}\} \;\; < \;\; \alpha_k,$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Therefore, all α_i for $i = 1, \ldots, p$ are identifiable if the above holds for all i.

Remark 6.2.4. *All parameters $\alpha_2, \ldots, \alpha_p$ are not identifiable if $\alpha_i \leq \alpha_1^i$ for $i = 2, \ldots, p$.*

Remark 6.2.5. *If $\alpha_1 = 0$ and $p \geq 3$, then by Proposition 6.2.3, if $\min\{\alpha_2, \alpha_3\} > 0$, both α_2 and α_3 are identifiable. However, for all remaining $\boldsymbol{\alpha}$ parameters to be identifiable,*

Proposition 6.2.3 must hold, with this simplifying, for all $i = 4, \ldots, p$, *to*

$$\max_{j=2,\ldots,\lfloor i/2 \rfloor} \{\alpha_j \alpha_{i-j}\} \; < \; \alpha_i.$$

In Section 6.6 we discuss the implications of imposing these stationarity and identifiability constraints on parameter inference for the Max-ARMA process.

6.3 Estimating extremal measures

6.3.1 Extremal index

Consider the maximum M_n of a sequence of IID unit Fréchet random variables $X_1, \ldots, X_n$, i.e., $M_n = \max\{X_1, \ldots, X_n\}$. If there exist sequences of constants $\{c_n > 0\}$ and $\{d_n\}$, so that the rescaled block maximum $(M_n - d_n)/c_n$ has a nondegenerate limiting distribution G as $n \to \infty$, then G is a member of the generalised extreme value distribution (GEV) family (Leadbetter et al., 1983). In this case, for unit Fréchet random variables, using $c_n = n$ and $d_n = 0$, G is also unit Fréchet so that $G(x) = \exp(-1/x)$ for $x > 0$; this is a member of the GEV family.

Now consider $X_1, \ldots, X_n$ as a stationary sequence, still with unit Fréchet margins, but satisfying the long-range dependence condition, $D(u_n)$ of Leadbetter et al. (1983) with $u_n = c_n x + d_n$ for any x in the domain of G, which ensures events long apart in time are near independent but does not impose any bounds on the local dependence conditions. The limiting distribution of $(M_n - d_n)/c_n$ of this stationary process when it exists, with c_n and d_n as for the IID case, is $G^\theta(x)$ with $G(x) = \exp(-1/x)$ for $x > 0$ and $\theta \in (0, 1]$ the extremal index. For an independent series $\theta = 1$, but the converse is not true, so θ captures the effect of the temporal dependence on the distribution of the maximum.

For dependent sequences, clusters of extreme events form above high thresholds.

Extreme observations in different clusters are assumed to be independent, whilst events within the same cluster exhibit dependence. Let $N_{p_n}(u_n)$ denote the number of observations of $X_1, \ldots, X_{p_n}$ to exceed the threshold u_n, where $p_n = \mathcal{O}(n)$. Then a cluster occurs when $N_{p_n}(u_n) > 0$, so the cluster size distribution π is defined by

$$\pi(m) = \lim_{n\to\infty} \Pr(N_{p_n}(u_n) = m | N_{p_n}(u_n) > 0),$$

for $m \in \mathbb{N}$ (Hsing et al., 1988). The extremal index, defined by Leadbetter et al. (1983), is characterised as the reciprocal of the limiting mean cluster size, where *limiting* refers to exceedances of an increasing threshold tending to the upper endpoint. O'Brien (1987) gives an alternative definition of the extremal index, based on the number of down-crossings in clusters of threshold exceedances, i.e.,

$$\theta = \lim_{n\to\infty} \Pr(X_2 \leq u_n, \ldots, X_{p_n} \leq u_n | X_1 > u_n), \tag{6.3.1}$$

where $\{p_n\}$ is an increasing sequence with $p_n = \mathcal{O}(n)$. They give an alternative long-range dependence condition to that of Leadbetter et al. (1983), known as the asymptotic independence of maxima condition. The well-known runs method of estimation for the extremal index (Smith and Weissman, 1994) is strongly related O'Brien's definition, where clusters are defined as those separated by p_n non-extreme values, where p_n is termed the run length.

For a Max-ARMA(1,0) process, Robinson and Tawn (2000) show that $\theta = 1 - \alpha_1$. They also show that for this particular process, the cluster sizes are geometrically distributed, with distribution function $\pi(m) = \alpha_1^{m-1}(1 - \alpha_1)$ for $m \in \mathbb{N}$. This is a suitable model as the number of exceedances determines a cluster, whilst a non-exceedance would terminate a cluster. To the best of our knowledge, the extremal index of a Max-ARMA(p, q) process has not been derived before. We prove the following result in Section 6.7.

Proposition 6.3.1. *The extremal index of a stationary Max-ARMA(p, q) process is*

$$\theta = \gamma \max\{\beta_0, \beta_1, \ldots, \beta_q\}, \tag{6.3.2}$$

where $\gamma > 0$ is the scale parameter of the innovation Z_t distribution, defined by expression (6.2.2).

As $\beta_0 = 1$, then $\max\{\beta_0, \ldots, \beta_q\} \geq 1 > \max\{\alpha_1, \ldots, \alpha_p\}$ due to stationarity. Hence if $\beta_J = \max\{\beta_0, \ldots, \beta_q\}$, then $\beta_J > \alpha_J$ and hence it is identifiable by Remark 6.2.2, so the result in Proposition 6.3.1 is well defined. According to the above proposition, a Max-ARMA(1,0) process with $\beta_j = 0$ for all j and $\gamma = 1 - \alpha_1$, as identified by expression (6.2.4), has extremal index $\theta = 1 - \alpha_1$. This agrees with the results of Robinson and Tawn (2000).

6.3.2 Extremal dependence measure

An alternative measure of extreme temporal dependence to the extremal index is the coefficient of extremal dependence χ_κ for specific lags κ (Coles et al., 1999; Heffernan et al., 2007). When used for a stationary time series context, this is a similar type of pairwise dependence measure to the auto-correlation function used for classic ARMA models. This is defined as

$$\chi_\kappa = \lim_{u \to x^F} \Pr(X_{t+\kappa} > u | X_t > u), \tag{6.3.3}$$

for $k \in \mathbb{Z}$ and a threshold u, where x^F is the upper endpoint of the marginal distribution function F_X, so for a Max-ARMA process with Fréchet margins $x^F = +\infty$. If $\chi_\kappa \in (0, 1]$, we say that X_t and $X_{t+\kappa}$ are asymptotically dependent; this means there is a non-zero probability of $X_{t+\kappa}$ being large when X_t is large at all extreme levels. Whereas $\chi_\kappa = 1$ and $\chi_\kappa = 0$ correspond to perfect dependence in the extremes and asymptotic independence, respectively. Ferreira (2011) derives χ_κ for the Max-ARMA(1,0) process

with Pareto(ξ_P) margins, as $\chi_\kappa = \alpha_1^{|\kappa/\xi_P|}$ where $\xi_P > 0$ is the tail index of the Pareto distribution. Here, we derive χ_κ for a general Max-ARMA(p, q) process which has not been done before, to the best of our knowledge.

A similar measure of serial extremal dependence is the extremogram, defined by Davis and Mikosch (2009). For a regularly varying stationary series $\{X_t\}$, the extremogram is defined for two sets A and B, that are bounded away from zero, by the limit

$$\rho_{AB}(\kappa) = \lim_{n\to\infty} \mathbb{P}(c_n^{-1} X_t \in A \mid c_n^{-1} X_{t+\kappa} \in B),$$

for $c_n \to \infty$ such that $\mathbb{P}(|X_t| > c_n) \sim n^{-1}$. By defining A and B as bounded away from zero, the events $\{c_n^{-1} X_t \in A\}$ and $\{c_n^{-1} X_{t+\kappa} \in B\}$ become extreme in the limit. The extremogram becomes the extremal dependence measure above when $A = B = (1, \infty)$ and $c_n = n$, with the expression of c_n given by $\{X_t\}$ having Fréchet margins.

For $\tau \in \mathbb{N}$ we define

$$\gamma_\tau := \max_{\mathcal{S}_\tau} \left\{ \beta_j \prod_{\substack{i=1,\ldots,p: \\ \alpha_i > 0}} \alpha_i^{a_i} \right\}, \tag{6.3.4}$$

where $\mathcal{S}_\tau$ is defined by expression (6.2.3). Then $0 \leq \gamma_\tau \leq \max\{\beta_0, \beta_1, \ldots, \beta_q\}$. If $\max\{\beta_1, \ldots, \beta_q\} > 1$, then $\gamma_\tau > 1$ for some τ but $\gamma_\tau \to 0$ always as $\tau \to \infty$ for a stationary Max-ARMA process, since $\max\{\alpha_1, \ldots, \alpha_p\} < 1$ and for each τ there exist is at least one $i = 1, \ldots, p$ such that $a_i \to \infty$ as $\tau \to \infty$. This gives us the final notation required for the following proposition about χ_k for a Max-ARMA(p, q) process, which we prove in Section 6.7.

Proposition 6.3.2. *The extremal dependence measure χ_κ for observations separated by $\kappa \in \mathbb{N}$ time lags, for a stationary Max-ARMA(p, q) process $\{X_t\}$, is*

$$\chi_\kappa = \gamma \sum_{\delta=0}^{\infty} \min\{\gamma_\delta, \gamma_{\delta+\kappa}\}, \tag{6.3.5}$$

where γ_δ is defined by expression (6.3.4), and $\chi_{-\kappa} = \chi_\kappa$.

Remark 6.3.3. *It follows from the formulation of* γ_τ *in expression* (6.3.4) *that* $\gamma_{(q+1)\tau+j}$ *decays at a geometric rate for each* $j = 1, \ldots, q$ *each converging to zero as* $\tau \to \infty$. *So* $\lim_{\kappa\to\infty} \chi_\kappa = 0$.

Remark 6.3.4. *When* $1 > \alpha_1 > \ldots > \alpha_p > 0$ *and* $\beta_0 = 1 > \beta_1 > \ldots > \beta_q > 0$, γ_τ *is a strictly decreasing function for all* $\tau \in \mathbb{N}$. *Then the form of* χ_κ *for* $\kappa \geq 0$ *simplifies as follows,*

$$\chi_\kappa = \gamma \sum_{\tau=0}^{\infty} \gamma_{\tau+\kappa} = \gamma \sum_{\tau=\kappa}^{\infty} \gamma_\tau = \gamma \left(\sum_{\tau=0}^{\infty} \gamma_\tau - \sum_{\tau=0}^{\kappa-1} \gamma_\tau \right) = \gamma \left(\frac{1}{\gamma} - \sum_{\tau=0}^{\kappa-1} \gamma_\tau \right) = 1 - \gamma \sum_{\tau=0}^{\kappa-1} \gamma_\tau,$$

where the infinite sum is equal to $1/\gamma$ *by Remark 6.2.1.*

For the stationary Max-ARMA(1,0) process, $\gamma = 1 - \alpha_1$ from expression (6.2.2), and as $\beta_0 = 1$, $\gamma_\tau = \alpha_1^\tau$ from expression (6.3.4) so, for $\kappa \geq 0$, we obtain that

$$\chi_\kappa = 1 - (1 - \alpha_1) \sum_{\delta=0}^{\kappa-1} \alpha_1^\delta = \alpha_1^\kappa,$$

which agrees with the result of Robinson and Tawn (2000).

6.4 Simulation of Max-ARMA(p, q) processes

Davis and Resnick (1989) simulate from a stationary Max-ARMA(1,0) process, which is straight-forward as X_1 can be generated from the marginal distribution. and then expression (6.2.1) can be used recursively for $t > 2$. However, they are not specific about how to generate X_1, but calculating γ in this case is straightforward using expression (6.2.4). Extending this method to a higher order, particularly with $q \geq 1$ is more complex; we illustrate how to do so in this section.

We aim to simulate a series of length $n \in \mathbb{N}$, denoted by $\{\tilde{X}_t\}_{t=1}^n$, from a stationary Max-ARMA(p, q) process. Since the Max-ARMA process is driven by an IID innovation process $Z_t \sim$ Fréchet(γ), we first simulate $m + n$ values from this distribution, denoted

by $\{\tilde{Z}_t\}_{t=-m}^{n}$, with scale parameter $\gamma > 0$ determined by the parameters of the Max-ARMA process, as stated in Remark 6.2.1, to ensure stationarity. Here, we have m values linked to the requirement for a burn-in period; see below for the reasons why.

The first step is to derive the value of γ, which is given by an infinite sum (6.2.2) of terms, where the τth term itself is the maximum over a complex set $\mathcal{S}_\tau$. Given parameter values $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we approximate the infinite sum by a partial sum up to the Nth term, for large N. We find that $N = 100$ gives a suitable approximation for a wide range of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, but larger values may be necessary when $\max\{\alpha_1, \ldots, \alpha_p\}$ is very close to 1.

For each step $\tau = 0, \ldots, N$ there is a need to find a solution set that satisfies the conditions of $\mathcal{S}_\tau$. We achieve this by trying all possible combinations and keeping only the feasible set. Specifically, we check for all possible combinations of $a_i \in \{0, \ldots, \tau : \alpha_i > 0\}$ for $i = 1, \ldots, p$ and $j \in \{0, 1, \ldots, q\}$ that satisfy the properties of $\mathcal{S}_\tau$ outlined in (6.2.3), for each iteration of τ. We then find which element of this feasible solution set maximises $\beta_j \prod_{i=1,\ldots,p\,:\,\alpha_i>0} \alpha_i^{a_i}$.

First, consider how to simulate forward given the process is already in a stationary state. At time $t \in \mathbb{N}$, with $p \leq t \leq n$, we can easily simulate $\tilde{X}_t$ using expression (6.2.1) since the *past values* $(\tilde{X}_{t-1}, \ldots, \tilde{X}_{t-p})$ and $(\tilde{Z}_{t-1}, \ldots, \tilde{Z}_{t-q})$ are available. However, for the earliest terms in the simulated $\tilde{X}_t$ sequence where $t < p$, the *past values* are unknown so the joint distribution for $\tilde{X}_t$ is complex. Instead, we simulate these early terms $\tilde{X}_1, \ldots, \tilde{X}_p$ from a unit Fréchet distribution. These observations have the correct marginal distribution but are independent of one another. Therefore a burn-in period of length $m \geq q - \min\{p, q\}$ is required to ensure that the simulated series has the correct stationary dependence structure.

We illustrate the extremal properties of a stationary Max-ARMA(p, q) process through four examples, labelled series 1-4, where a burn-in period of length $m = 1000$ is discarded. Series 1 and 2 are Max-ARMA(3,0) processes whilst series 3 and 4 are

	Series				Extremal properties			
	(p, q)	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	γ	θ	χ_1	χ_2	χ_3
1	$(3, 0)$	$(0.85, 0.77, 0.7)$	$(0, 0, 0)$	0.11	0.11 (0.11)	0.88 (0.88)	0.79 (0.80)	0.70 (0.71)
2	$(3, 0)$	$(0.3, 0, 0.1)$	$(0, 0, 0)$	0.65	0.65 (0.58)	0.35 (0.36)	0.16 (0.19)	0.1 (0.14)
3	$(3, 3)$	$(0.85, 0.77, 0.7)$	$(2, 1, 0.9)$	0.05	0.11 (0.10)	0.89 (0.88)	0.8 (0.79)	0.72 (0.72)
4	$(3, 3)$	$(0.85, 0.77, 0.7)$	$(50, 10, 5)$	0.002	0.11 (0.11)	0.89 (0.87)	0.79 (0.78)	0.70 (0.70)

Table 6.4.1: Values of γ, θ and χ_κ for $\kappa = 1, 2, 3$ of different Max-ARMA processes derived from Remark 6.2.1 and Propositions 6.3.1 and 6.3.2, respectively. Empirical estimates of $\theta(u)$ and $\chi_\kappa(u)$, derived using a simulation of length 10^6 and the 0.95 marginal quantile as the threshold u, are given in parentheses. All values are given to 2 decimal places, where appropriate.

Max-ARMA(3,3) processes. The values of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are given in Table 6.4.1 and are chosen to ensure identifiability, except for series 2 where we explore the effect of $\alpha_2 = 0$ with α_1, α_3 identifiable. This table also presents the values of γ, θ and χ_κ $(\kappa = 1, 2, 3)$ for each series, where these are evaluated using the limiting theoretical expressions given by Remark 6.2.1 and Propositions 6.3.1 and 6.3.2, respectively. Series 2 exhibits less extremal dependence than the remaining series due to its smaller $\boldsymbol{\alpha}$ parameters. Series 1, 3 and 4 have similar values for θ and χ_κ; this shows that adding $\boldsymbol{\beta}$ parameters to series 1 hasn't changed the extremal dependence structure and neither has increasing the magnitude of the $\boldsymbol{\beta}$ parameters, as in series 4. The values of θ indicate a cluster size of 9 for processes 1, 3 and 4. The values of χ_κ decay with increasing lag κ in all cases.

To gain further insight into temporal trajectories of series 1-4, we simulated realisations from each series, with Figure 6.4.1 showing a segment of 1000 consecutive values from each of these processes (after burn-in). To avoid the largest events dominating the image, each series presented in Figure 6.4.1 is on a standard Gumbel marginal scale, i.e., we show the time series of $\log X_t$. For all four series the dominating feature of the plots is the repeated spikes - sudden jumps up in values - followed by an approximate linear decay (an exponential decay on unit Fréchet margins becomes linear on this Gumbel marginal scale). Despite all four series having identical choices of p and, for

series 1, 3 and 4, similar values of θ, their trajectories differ: series 2 has more sporadic behaviour with frequent spikes that instantly decay, whereas series 4 has larger jumps from a typical value to an extreme event due to the $\boldsymbol{\beta}$ parameters being much larger than for the other series.

Simulation from these Max-ARMA processes also provides an independent assessment of the theoretically derived properties of θ and χ_κ. To obtain Monte Carlo estimates with limited noise we use simulations of length 10^6 and estimate $\theta(u)$ and $\chi_\kappa(u)$ using empirical values based on expressions (6.3.1) and (6.3.3). These values are given in Table 6.4.1. For estimates of $\theta(u)$, we have $u_n = u$ equal to the 0.95-marginal quantile and $p_n = 1$ for all series. Empirical estimates of $\chi_\kappa(u)$ decrease as κ increases so the choice of $p_n = 1$ is sufficient. Table 6.4.1 shows an excellent agreement between these Monte Carlo estimates and the true values for all four series, despite our threshold not being very extreme.

6.5 Inference

6.5.1 Model parameterisation

In Section 6.2 we show that the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of a Max-ARMA(p, q) process need to satisfy the conditions of Remarks 6.2.1 and 6.2.2 to achieve stationarity and identifiability, respectively, with the conditions of Remark 6.2.2 on the $\boldsymbol{\alpha}$ parameters expressed more parsimoniously in Proposition 6.2.3. In particular, if for any $k = 2, \ldots, p$, we have that $\alpha_1, \ldots, \alpha_{k-1}$ all satisfy the identifiability conditions of Remark 6.2.2, then if α_k is less than or equal to the lower bound of the constraint (6.2.5) then α_k plays no role in determining the sample path of the Max-ARMA process, so it is equivalent to setting $\alpha_k = 0$. Similarly, if for any $j = 1, \ldots, q$ we have $\beta_j \leq \alpha_j$ then β_j also does not affect the sample paths, so has no difference from taking $\beta_j = 0$. However, for a Max-ARMA(p, q) process to be well-defined we need both the α_p and β_q terms

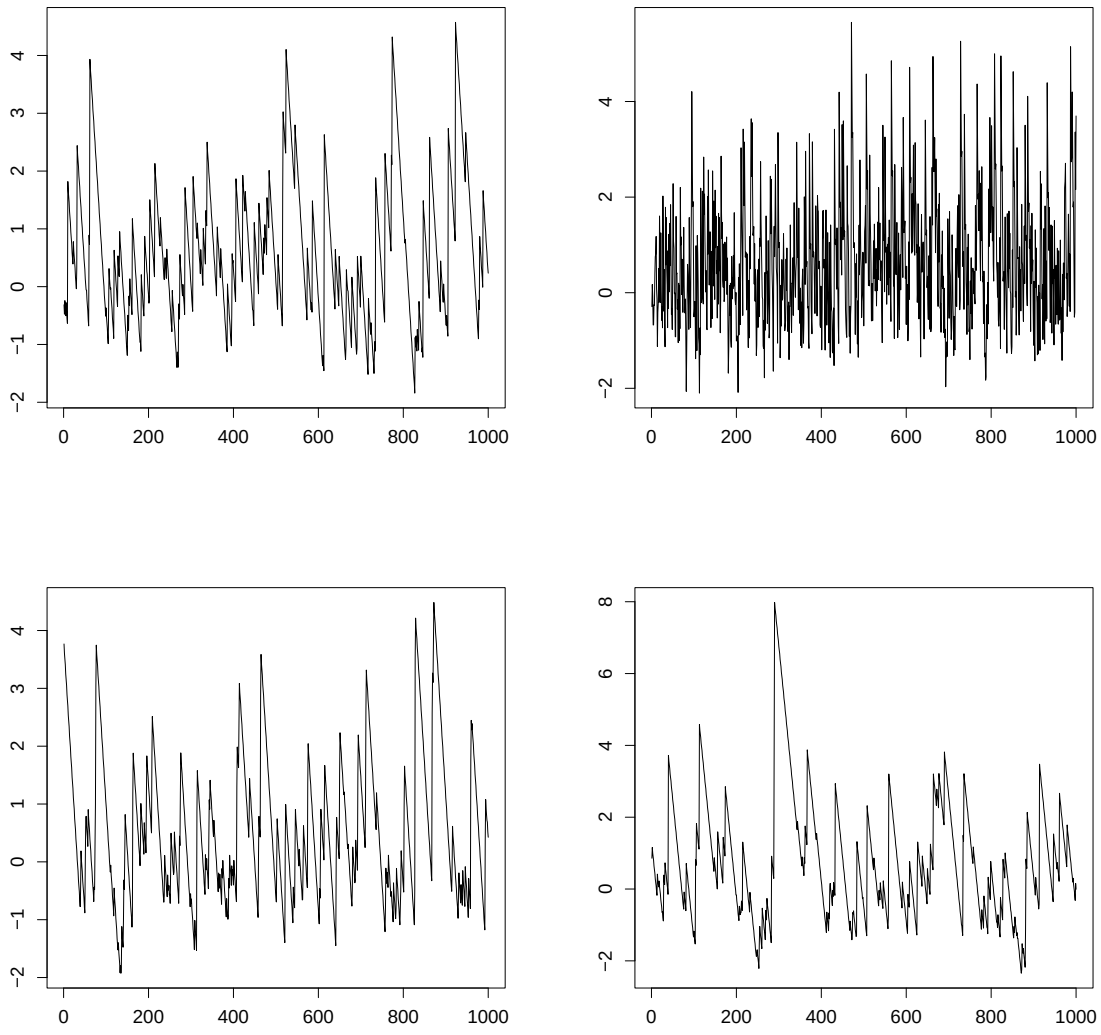


Figure 6.4.1: Simulations from stationary Max-ARMA(p, q) processes $\{X_t\}$, presented on Gumbel margins, i.e., for $\log X_t$, with sample sizes $n = 1000$: $(p, q) = (3, 0)$ (top row) and $(p, q) = (3, 3)$ (bottom row) with parameters $\boldsymbol{\alpha} = (0.85, 0.77, 0.7)$ (top left), $\boldsymbol{\alpha} = (0.3, 0, 0.1)$ (top right), $\boldsymbol{\alpha} = (0.85, 0.77, 0.7)$ and $\boldsymbol{\beta} = (2, 1, 0.9)$ (bottom left), and $\boldsymbol{\alpha} = (0.85, 0.77, 0.7)$ and $\boldsymbol{\beta} = (50, 10, 5)$ (bottom right).

to be identifiable, and for statistical inference we cannot allow for multiple points in the parameter space to give processes with identical sample paths if they have identical realisations of the innovation sequence $\{Z_t\}$. Thus, for stationary processes, without any redundancy for lack of identifiability, and for (p, q) to both be uniquely defined, we must further restrict the parameter space, identified in Section 6.2, to the parameter space $\Theta_{p,q}$ given by

$$\begin{aligned}\Theta_{p,q} = \Bigg\{ & (\boldsymbol{\alpha}, \boldsymbol{\beta}) : 0 \leq \alpha_1 < 1, \max_{j=1,\ldots,\lfloor i/2 \rfloor} \{\alpha_j \alpha_{i-j}\} \ \leq \ \alpha_i < \ 1 \ \forall \ i = 2, \ldots, (p-1), \\ & \max_{j=1,\ldots,\lfloor p/2 \rfloor} \{\alpha_j \alpha_{p-j}\} \ < \ \alpha_p < \ 1, \beta_0 = 1, \ \beta_j \geq \alpha_j \ \forall j = 1, \ldots, \min\{p, q-1\}, \\ & \text{and for } q > p, \ \beta_j \ \geq \ 0 \text{ for } j = (p+1), \ldots, (q-1), \ \beta_q \ > \ 0, \\ & \text{whilst for } p \ \geq \ q, \ \beta_q \ > \ \max\{\alpha_q, 0\} \Bigg\}.\end{aligned}$$

This novel formulation for $\Theta_{p,q}$ has the benefit of the parameter space being continuous whilst allowing for any combination of the parameters $\boldsymbol{\alpha}_{-p}$ for $p \geq 2$ (since α_1 is always identifiable) and $\boldsymbol{\beta}_{-(0,q)}$ (since $\beta_0 = 1$ is always identifiable) not to affect the sample paths; this is achieved when they satisfy the equal conditions in their respective bounds in the specification of $\Theta_{p,q}$. Imposing the parameter space $\Theta_{p,q}$ on the MaxARMA(p, q) process has no effect on either γ, θ or χ_τ in expressions (6.2.2), (6.3.2) and (6.3.5) respectively, and is key for inferences in Sections 6.5.2 and 6.6.

Due to the complexity of the parameter space defined above, we reparameterise $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to achieve a more orthogonal parameter space so that inference is easier. Consider the parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_q)$ defined by

$$\delta_i = \begin{cases} \alpha_1 & \text{for } i = 1 \\ \alpha_i - \max_{j=1,\ldots,\lfloor i/2 \rfloor} \{\alpha_j \alpha_{i-j}\} & \text{for } i = 2, \ldots, p, \end{cases} \tag{6.5.1}$$

and

$$\epsilon_j = \begin{cases} \beta_j - \alpha_j & \text{for } j = 1, \ldots, \min\{p, q\} \\ \beta_j & \text{for } j = (p+1), \ldots, q \text{ if } q > p. \end{cases} \tag{6.5.2}$$

Thus, when $\delta_i = 0$, for any $i = 1, \ldots, p-1$ (or when $\epsilon_j = 0$ for any $j = 1, \ldots, q-1$) then α_i (or β_j) has no impact on the sample path of the Max-ARMA(p, q) process. With this transformation, the new parameter space becomes

$$\begin{aligned} \tilde{\Theta}_{p,q} = \big\{ & 0 \leq \delta_i < \Delta_i \text{ for } i = 1, \ldots, (p-1),\ 0 < \delta_p < \Delta_i, \\ & \epsilon_j \geq 0 \text{ for } j = 1, \ldots, (q-1),\ \epsilon_q > 0 \big\}, \end{aligned} \tag{6.5.3}$$

where Δ_i expresses the bound $\alpha_i < 1 - \max_{j=1,\ldots,\lfloor i/2 \rfloor}\{\alpha_j \alpha_{i-j}\}$ in terms of $(\delta_1, \ldots, \delta_{i-1})$. Although Δ_i is complex for a general i, the condition is easily checked after transforming $\boldsymbol{\delta}$ into $\boldsymbol{\alpha}$ and is simple when p is small, e.g., $\Delta_1 = 1, \Delta_2 = 1 - \delta_1^2$. Though at first sight $\tilde{\Theta}_{p,q}$ may not seem much simpler than $\Theta_{p,q}$, in practice it is much easier to use in optimisations such as in Section 6.5.2. This is because the constraints between the ith components of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have been removed for $i = 1, \ldots, \min\{p, q\}$, as has the complex inequality $\max_{j=1,\ldots,\lfloor i/2 \rfloor}\{\alpha_j \alpha_{i-j}\} < \alpha_i$ for all $i = 2, \ldots, p$. The new constraint with upper bound Δ_i is typically satisfied for most trial values of $\boldsymbol{\delta}$ in an optimisation as it would be unlikely for the true parameters to be very close to the upper limits on α_i, i.e. corresponding to non-stationarity. Thus we found that using $(\boldsymbol{\delta}, \boldsymbol{\epsilon}) \in \tilde{\Theta}_{p,q}$ is a major simplification to using $\Theta_{p,q}$.

6.5.2 Inference strategy

Davis and Resnick (1989) considered inference issues in theory for a Max-ARMA(p, q) process. They find super-efficient estimators of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters when (p, q) are known. To illustrate this, it can easily be shown that for a Max-ARMA(1,0) process $\{X_t\}$, if $X_t > u$ for large u and any $t \in \mathbb{N}$, then $X_{t+1} = \alpha X_t$ for some $\alpha \in (0, 1)$ with

probability tending to 1 as $u \to \infty$. More generally, for a Max-ARMA(p, q) process, they show that if α_i is identifiable, then $\Pr(X_t = \alpha_i X_{t-i}) > 0$ and that $X_t/X_{t-i} \geq \alpha_i$ for all $i = 1, \ldots, p$; and they find some similar features involving the $\boldsymbol{\beta}$ parameters. So, for a sample of size n, they proposed the following estimator, for all $i = 1 \ldots, p$,

$$\hat{\alpha}_i = \min_{t=i+1,\ldots,n} \left\{ \frac{X_t}{X_{t-i}} \right\},$$

with a positive probability that $\hat{\alpha}_i = \alpha_i$, and this probability tending to 1, with a geometric rate, as $n \to \infty$.

The pseudo-deterministic behaviour of the Max-ARMA(p, q) process will not be observed exactly in practice for any real-world system. So, our inference strategy differs from that of Davis and Resnick (1989) as we consider the Max-ARMA(p, q) process to be only an approximation to the actual process generating the data (e.g., river flow data) and we believe this model is useful only when the process is in an extreme state. So, we only assume the Max-ARMA model provides an approximate formulation of the process for extreme observations, i.e., when $\max\{X_{t-1}, \ldots, X_{t-p}\} > u$, for some high threshold u. Given this perspective, likelihood inference is not suitable for fitting a Max-ARMA(p, q) process of specified orders p and q to observational data as we do not view that the data actually comes from this precise model. Furthermore, we treat (p, q) as unknown, and therefore they also need estimating.

Instead of likelihood inference, we take a moments-based inference approach for fitting the model to the extremes of observational data. Our strategy is motivated by the approach proposed by Rodriguez-Iturbe et al. (1988) for rainfall models and continued through a series of extensive work, with a recent example being Kaczmarska et al. (2015). In this approach, key properties of the process are derived in closed form and then the model parameters are estimated using a method of generalised moments; this minimises a weighted sum of squared differences between empirical properties of interest and the parametric estimates under the model. In the rainfall context, the

generalised moments include features of the body of the process, such as correlations, mean length of dry periods and means of aggregated rainfall over different time windows. Here, as extremes are of most interest, we follow a similar approach but use extremal properties of the process to ensure a good fit, such as θ and χ_k for $k \in \mathbb{N}$.

To estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for a Max-ARMA(p, q) process we use $p+q+2$ moments. Let $\hat{M}_m$ and M_m denote the empirical and true extremal dependence measures, respectively, for $m = 1, \ldots, (p + q + 2)$. Then the moments are formally defined, for the empirical measures, as

$$\hat{M}_m = \begin{cases} \hat{\theta}(u) & \text{for } m = 1 \\ \hat{\chi}_1(u) & \text{for } m = 2 \\ \hat{\chi}_{T_m}(u) & \text{for } 2 < m < (p + q + 2) \\ \hat{\chi}_T(u) & \text{for } m = p + q + 2, \end{cases} \tag{6.5.4}$$

where $\hat{\theta}(u)$ and $\hat{\chi}_\kappa(u)$ are estimates of the sub-asymptotic empirical estimates for a threshold u, given in Sections 6.3.1 and 6.3.2, respectively. Furthermore, $T \geq p + q$, is chosen as the value κ for which the rate of decay of $\hat{\chi}_\kappa(u)$ first changes (i.e., decelerates), and then $T_m = \lfloor T(m-2)/(p+q) \rfloor$ ensures that the lags of $\hat{\chi}_{T_m}(u)$ used are equally spaced for $2 \leq m \leq p + q + 2$. In this way, we use values of χ_κ with not necessarily consecutive lags and spread over all lags with the strongest temporal dependence. The true measures M_m are defined analogously, with the parametric forms for extremal dependence measures θ and χ_κ for a Max-ARMA(p, q) process derived in Propositions 6.3.1 and 6.3.2, respectively.

As there is a strong dependence between the estimates of χ_κ for different κ values, we found it necessary to add additional features to improve model fit. We combine the simple estimators for α_i $(i = 1, \ldots, p)$ of Davis and Resnick (1989) with our inference strategy, exploiting this information into our generalised moment structure as an additional moment to M_m $(m = 1, \ldots, p+q+2)$, but using only extreme observation pairs,

i.e., when $\min\{X_t, X_{t-i}\} > u$. By combining it with the other generalised moments, we can let our estimate of α_i be the same as the Davis and Resnick (1989) estimator, if that were restricted to a fit on only the largest values, but only if the fit to the other generalised moments is not compromised.

As discussed in Section 6.5.1, we reparametrise $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in terms of $(\boldsymbol{\delta}, \boldsymbol{\epsilon})$ to simplify the parameter space we are working with. However, the extremal dependence measures derived in Propositions 6.3.1 and 6.3.2, that are required for the method of moments, are defined in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. It is straightforward to obtain expressions for these in terms of the new parameters $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$ using the inverse of expressions (6.5.1) and (6.5.2), respectively. We minimise the objective function,

$$\mathcal{M}(\boldsymbol{\delta}, \boldsymbol{\epsilon}; p, q) = \frac{\omega}{p+q+2} \sum_{m=1}^{p+q+2} (\hat{M}_m - M_m)^2 + \frac{1-\omega}{p} \sum_{i=1}^{p} \min_{t \in T(u,i)} \left\{ \left(\frac{X_t}{X_{t-i}} - \alpha_i \right)^2 \right\}, \tag{6.5.5}$$

where $T(u, i) = \{t = 1, \ldots, n : \min\{X_t, X_{t-i}\} > u\}$, over $(\boldsymbol{\delta}, \boldsymbol{\epsilon}) \in \tilde{\Theta}_{p,q}$ for fixed (p, q), where $\tilde{\Theta}_{p,q}$ is defined by the set (6.5.3). Here we have weighting terms $\omega/(p+q+2)$ and $(1-\omega)/p$ for $0 < \omega < 1$, so that as (p, q) are changed $\mathcal{M}(\boldsymbol{\delta}, \boldsymbol{\epsilon}; p, q)$ should be reasonably stable as the two different terms are averaged over their values for (m, i) respectively, and that greater importance is given to the extremal generalised moments as ω is increased.

For a given pair (p, q), minimising $\mathcal{M}(\boldsymbol{\delta}, \boldsymbol{\epsilon}; p, q)$ gives our estimated values $(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q})$ of $(\boldsymbol{\delta}, \boldsymbol{\epsilon})$ and equivalent of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. However, we cannot minimise $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ over (p, q) to find the best values for these indices. By our construction of $\mathcal{M}(\boldsymbol{\delta}, \boldsymbol{\epsilon}; p, q)$ to use averages in the objective function, over $(p+q+2)$- and p-terms respectively, we should see a form of stability in the values of $\mathcal{M}(\boldsymbol{\delta}, \boldsymbol{\epsilon}; p, q)$ once $p \geq p_0$ and $q \geq q_0$ for true values (p_0, q_0), but for smaller values of (p, q), increasing either should result in $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ decreasing. This suggests a form of elbow plot for order selection, as used in many areas, such as in determining k in a k-mean clustering algorithm (Syakur

et al., 2018).

6.5.3 Marginal inference

Since we define Max-ARMA(p, q) processes on unit Fréchet margins, we require an additional transformation from the observed stationary process $\{Y_t\}$ to the stationary Max-ARMA(p, q) process $\{X_t\}$ for inference. We assume that the marginal upper tail of $\{Y_t\}$ has heavy tailed margins, since we believe that any process for which the Max-ARMA process will be suitable to model its dependence structure will have heavy tails, and this is the case for the River Thames data that we analyse in Section 6.6.

We model the upper tail of the distribution $\{Y_t\}$, above a marginal threshold u_M, using a Pareto distribution where u_M is within the sample of observed $\{Y_t\}$ values. Specifically, we have that the survival function is modelled by

$$\Pr(Y_t > y) = d\left(\frac{u_M}{y}\right)^c \text{ for } y \geq u_M, \tag{6.5.6}$$

for $c \in \mathbb{R}_+$ and $d \in (0, 1)$. To estimate (c, d) we use maximum likelihood methods, making the working assumption that the observations $\{Y_t > u_M\}$ are independent, which gives the estimates

$$\hat{c} = \left(\frac{1}{n_u}\sum_{j=1}^{n_u} \log\frac{y_j}{u_M}\right)^{-1} \quad \text{and} \quad \hat{d} = \frac{n_u}{n},$$

where $y_1, \ldots, y_{n_u}$ are the n_u observations of Y_t that exceed u_M. Here $\hat{c}$ is identical to the Hill estimator (Hill, 1975), corresponding to the reciprocal of the mean of the exceedances of u_M on a log scale, and $\hat{d}$ is the sample proportion above u_M. Below the threshold u_M we have no theoretical basis for the form of the distribution of $\{Y_t\}$, so we estimate the distribution function function $F_Y(y)$, for $y < u_M$, using the empirical distribution of $\{Y_t\}$. As defined, the two components of the estimated distribution for $\{Y_t\}$, denoted as a combined function $\hat{F}_Y$, are continuous at u_M.

We use the probability integral transform to map from $\{Y_t\}$ to $\{X_t\}$. Specifically, we use the transformation $X_t = T(Y_t)$ for all t, where $T(y) = -1/\log(\hat{F}_Y(y))$. Likewise, after simulating realisations $\{X_t^*\}$ of the fitted Max-ARMA model, we back-transform to give simulated realisations on the original margins, denoted by $\{Y_t^*\}$, where $Y_t^* = \hat{F}_Y^{-1}(\exp(-1/X_t^*))$. Here $\hat{F}_Y^{-1}(v)$ is well-defined for $d \leq v < 1$ from inverting expression (6.5.6), whereas for $0 < v < d$ we linearly interpolate $\hat{F}_Y$ between jumps to obtain a one-to-one function with a unique inverse.

6.6 Illustrative analysis of River Thames extreme flows

We illustrate the inference procedure set out in Section 6.5 for fitting a Max-ARMA(p, q) process and selecting (p, q) for the UK River Thames daily maximum river flow data, for which four winter segments of the series are shown in Figure 6.1.1. We apply a greater weight to the generalised extremal moments component than that of Davis and Resnick (1989) parameter estimation by setting $\omega = (p + q + 2)/(2p + q + 2)$ so that when rescaled by the number of moments in each component, the weights are equal; see expression (6.5.5). The extremal moments θ and χ_κ, derived in Propositions 6.3.1 and 6.3.2, respectively, are defined in terms of an infinite sum which we find is reasonably approximated by a sum up to 100, as discussed in Section 6.4.

Before fitting the Max-ARMA model, which has unit Fréchet margins, the observed river flow series $\{Y_t\}$ must be transformed componentwise to also have unit Fréchet margins. We do so using the Hill estimator (Hill, 1975) and probability integral transform, as outlined in Section 6.5.3, giving the series $\{X_t\}$. We use a threshold $u_M = 270.4$ (equivalent to the 0.98 quantile) and obtain a Hill estimate of $\hat{c} = 5.1$, so the data has much lighter tails than a unit Fréchet series where $c = 1$. Figure 6.6.1 shows a QQ plot for the data exceedances of u_M, after using the fitted Pareto tail model to trans-

form the data to Gumbel margins, i.e., the same as $\log X_t$. The plot shows evidence that our marginal model fits reasonably well in the upper tail, especially for the less extreme values but with a slight overestimation for the largest observations; changing the threshold u_M could improve the fit. The log transformation of the series is also shown in Figure 6.6.1 (i.e., on Gumbel margins) for the same winter seasons as given in Figure 6.1.1, so that the two largest winter events and two randomly selected winter periods are shown. The data are now on the same marginal scale as our simulations for various Max-ARMA(p, q) processes shown in Figure 6.4.1, where the data and simulations have a broadly similar structure in terms of spikes and decays. The largest increase from a typical event to an extreme event is similar to that of simulated series 4, suggesting a large β_j parameter, for some $1 \leq j \leq q$, might be required to capture a change of this magnitude. The observed data in Figure 6.6.1 has stronger dependence than these simulated series which can be seen by the longer decays following a spike. Additionally, the spikes in our simulations occur instantly, going from a typical value to an extreme in 1 or 2 time lags, whilst the observed data takes much longer to reach a spike due to the large catchment size of the River Thames of $\sim$16,200km^2.

To obtain the moments of expression (6.5.4) for inference of the river flow series we must find the value of T (the maximum lag κ for χ_κ that we use) where the estimate of $\chi_T(u)$ is small and the rate of decay of $\hat{\chi}_\kappa(u)$ slows down for $\kappa > T$. Figure 6.1.1 shows that $T = 14$ (i.e., 2 weeks) for both the 0.9 and 0.95 quantile. So, for example, for fitting a Max-ARMA(2,0) we would be interested in extremal moments of $\theta, \chi_1, \chi_7, \chi_{14}$.

We fit Max-ARMA(p, q) models for all combinations of $p = 1, 2, 3$ and $q = 0, \ldots, 4$ to two sets of the River Thames data: exceedances of the 0.9 and 0.95 quantiles, so that the threshold u in expression (6.5.5) is set equal to these quantiles. By using these high quantiles, most of the observations we use for inference will be from the winter months when the largest events occur due to the seasonality of river flows. We tested thresholds higher than those shown here but omit the details as the results were noisy so

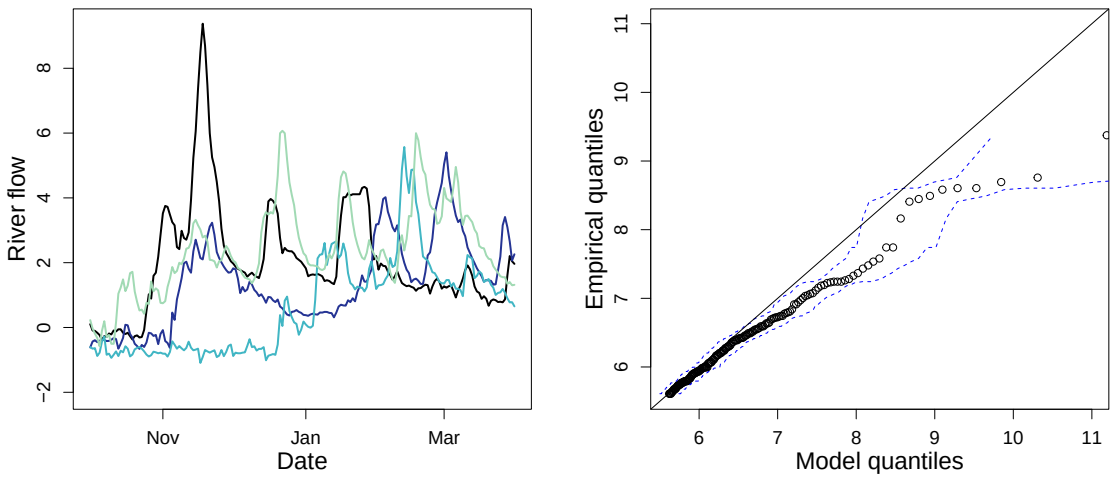


Figure 6.6.1: Left: River flow trace plot of the River Thames on Gumbel margins for the winter season (October - March) in 1894/95 (black), 1927/28 (dark blue), 1973/74 (light blue) and 2019/2020 (green). Right: QQ plot of the marginal Pareto tail model fitted to the River Thames exceedances of u_M on Gumbel margins with 95% tolerance bounds.

it was difficult to choose p and q in these cases. Thresholds lower than the 0.9 quantile were not considered as we are only interested in extreme events. We compare fits using the minimised objective function values $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ of expression (6.5.5) for each (p, q) combination, as well as assessing how close empirical estimates of the extremal dependence measures θ and χ_κ from the data are to the model based estimates for different orders (p, q).

Figure 6.6.2 shows the minimised $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ of expression (6.5.5) for different orders p and q, when the model is fitted to exceedances of the 0.9 and 0.95 quantile separately. As mentioned in Section 6.5.2, we cannot find the best model fit over (p, q) by choosing this to be the (p, q) combination where $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ is minimised, instead we look for (p_0, q_0) such that there is stability in the $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ values for all $p \geq p_0$ and $q \geq q_0$. For the model fits to exceedances of the 0.9 quantile, we observe stability in the minimised $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ for $(p, q) \geq (2, 3)$. For both choices of threshold u, $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ is relatively stable across all values of q for a fixed

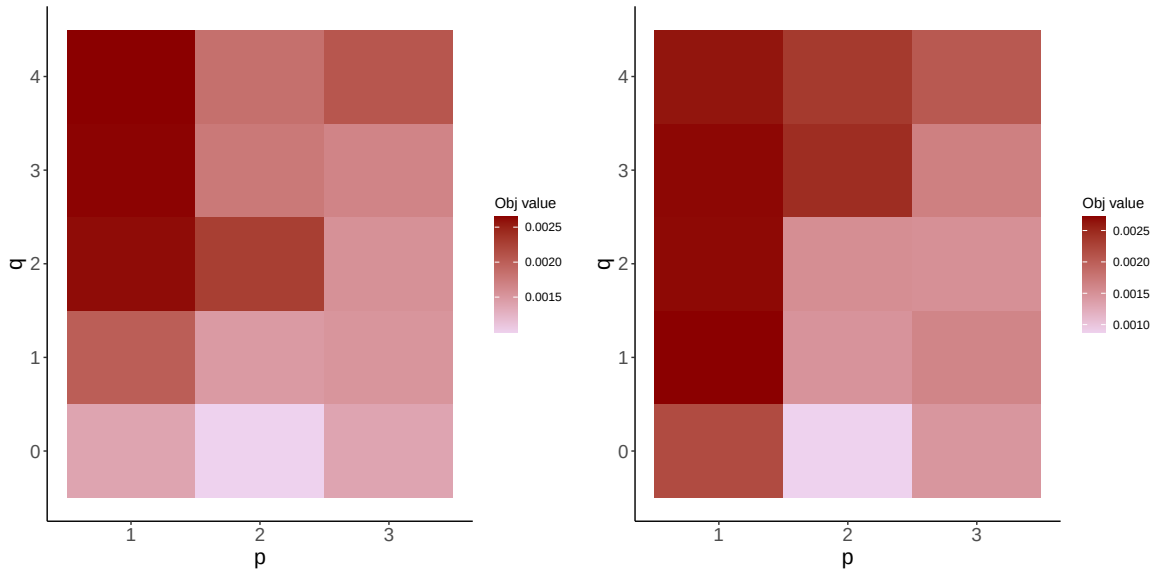


Figure 6.6.2: The minimised objective function value for our moments-based inference of expression (6.5.5) for Max-ARMA fits of different orders $p = 1, 2, 3$ (x-axis) and $q = 0, \ldots, 4$ (y-axis) to the River Thames data using a threshold u of the 0.9 (left) and 0.95 (right) quantiles. Darker (lighter) red indicates a higher (lower) objective function value.

p when $p = 3$. For exceedances of the 0.95 quantile, $\mathcal{M}(\hat{\boldsymbol{\delta}}_{p,q}, \hat{\boldsymbol{\epsilon}}_{p,q}; p, q)$ appears to be stable for $q \geq 3$ across all $p = 1, 2, 3$, suggesting a Max-ARMA$(3, 3)$ is the best fitting model based on this criteria, but we make further comparisons below for choosing our selected values of (p, q) here.

In Figures 6.6.3 and 6.6.4 we study estimates of the extremal dependence measures from a given Max-ARMA model across models with different orders (p, q). We use simulation methods to evaluate our model based estimates rather than using the theoretical limits derived in Section 6.3 that we used in the objective function (6.5.5). This is because we are interested in the sub-asymptotic extremal dependence measures $\theta(u)$ and $\chi_\kappa(u)$ so that we can compare these with the corresponding estimates from the observed series. To find these model based estimates, we simulate a large sample and obtain empirical estimates from this sample from the fitted Max-ARMA model using Monte Carlo methods, where we limit the Monte Carlo noise by taking the simulated sample to be of length 10^6.

First, we compare empirical estimates of extremal dependence measures $\theta(u)$ and $\chi_\kappa(u)$ for $\kappa = 1, 7, 14$ from the observed data and model-based estimates, when fit to exceedances of the 0.95 quantile. These are shown in Figure 6.6.3. The Max-ARMA models with orders $(2,0)$, $(2,1)$, $(3,0)$, $(3,2)$ and $(3,3)$ have estimates of $\theta(u)$ that lie within the 95% confidence intervals (found via bootstrapping) of the corresponding empirical estimate for the data. These same models, as well as when $p = q = 2$, also give model estimates of $\chi_1(u)$ within the confidence intervals (based on the Binomial sampling distribution) of the associated empirical estimates. Models with $p = 1$ always underestimate $\theta(u)$ compared to the empirical estimates. All models overestimate $\chi_7(u)$, except for $p = 3$ and $q = 4$, whilst most models (except when (p, q) is $(1,0)$, $(2,4)$ or $(3,4)$) give $\chi_{14}(u)$ estimates that lie within the 95% confidence intervals from the empirical estimates. Therefore, combining the results from Figures 6.6.2 and 6.6.3, we conclude that a Max-ARMA$(3,3)$ is the best fitting model for river flow exceedances of the 0.95 quantile at the River Thames. For this model we obtain parameter estimates $\hat{\boldsymbol{\alpha}} = (0.69, 0.78, 0.54)$ and $\hat{\boldsymbol{\beta}} = (3.15, 2.16, 0.99)$. Since all $\beta_j > \alpha_j$ for $j = 1, 2, 3$, all $\boldsymbol{\beta}$ are identifiable and affect the sample path of simulations from this model. Additionally, $\alpha_2 > \alpha_1^2$ whilst $\alpha_3 - \alpha_1\alpha_2 > 0$, but small, so primarily α_1 and α_2 influence the sample paths of the simulations from this model.

Figure 6.6.4 shows estimates of $\chi_\kappa(u)$ for $\kappa = 1, \ldots, 14$ from the data, with 95% confidence intervals, and from the Max-ARMA(3,3) model based estimates. For $\kappa = 1$ and $\kappa \geq 11$, we capture the structure of data estimate of $\hat{\chi}_\kappa(u)$ very well, but for $2 \leq \kappa \leq 10$ the model based estimates slightly overestimate this asymptotic dependence measure. For the other choices of (p, q) we tried, we did not obtain model based estimates of $\chi_\kappa(u)$ within the confidence intervals of the data estimates for lags in the range $2 \leq \kappa \leq 10$.

Figure 6.6.4 shows four simulated time series plots over 183 days (corresponding to the length of the winter seasons shown in Figure 6.1.1) using our fitted Max-ARMA(3,3)

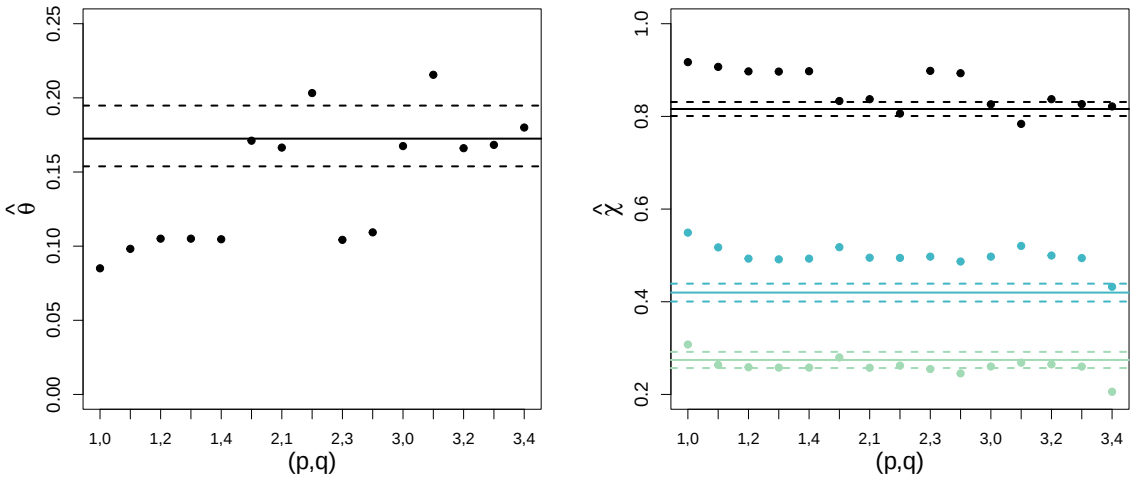


Figure 6.6.3: Estimates of $\theta(u)$ (left) and $\chi_\kappa(u)$ (right) for $k = 1, 7$ and 14 (black, blue and green, respectively) using a threshold u of the 0.95 quantile for the River Thames with empirical estimates (solid lines) and estimates using fitted Max-ARMA models (points) with varying orders (p, q) (x-axis). Horizontal dashed lines show 95% confidence intervals for the empirical estimates of $\theta(u)$ and $\chi_\kappa(u)$ for the data.

model, with the simulated series transformed onto the original data scale using the probability integral transform, as discussed in Section 6.5.3. We show time series from the two largest simulated events (where the simulated time series is the same length as the total River Thames winter data, so these events should be comparable in size to the largest observed events) as well as two randomly selected time series segments to illustrate *typical* behaviour. All four simulated time series segments exhibit a saw-tooth structure because $\hat{\alpha}_2 > \hat{\alpha}_1$ and we find this behaviour in all model fits with $p > 1$. Our simulated realisations of time series exhibit similar behaviour to the observed data in the decay, following a major spike, so our model is capturing the features of decay of the original series well. The simulations of the typical level segments of the time series also exhibit a similar number of smaller peaks to the observed data in Figure 6.1.1. A noticeable feature of the observed data that our simulations fail to capture is that the major spikes can take several days (or even up to a month, as for the flood event of winter 1894; see black time series segment in Figure 6.1.1) to rise to their peak, however,

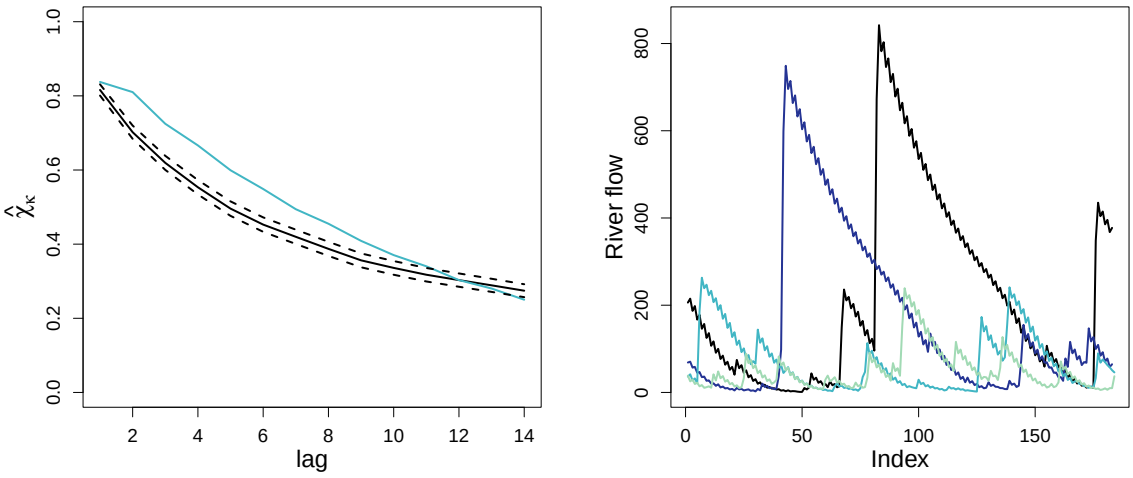


Figure 6.6.4: Left: Empirical estimates of $\chi_\kappa(u)$ for $\kappa = 1, \ldots, 14$, with u equal to the 0.95 quantile, for the data (black with 95% confidence intervals in dashed lines) and estimates from a Max-ARMA(3,3) model fit to the River Thames data (blue). Right: Four time series plots of length 183 (corresponding to the length of the winter season October-March) from the fitted Max-ARMA(3,3) to exceedances of the 0.95 quantile of the River Thames data, transformed to the original data scale. The time series plots correspond to the two largest events (black and dark blue) and two randomly selected series (light blue and green).

in the simulated realisations of the fitted models the spikes occur almost instantly. So this is a limitation of the best fitting model for the data that we have considered.

Our results are primarily illustrative for the inference procedure, rather than to demonstrate a definitive approach for choosing (p, q). We consider this to be of potential interest for future work, but for our purposes the choices are sufficient for barrier closure assessment, see the discussion below. However, to get a better fit of the model for the spike segments of the time series we could have imposed a much larger value for q, with the β_j increasing with j so that for a large innovation given by Z_t say, this would lead to a rising spike in the series through successive values with $\{X_{t+j} = \beta_j Z_t\}$ for $j = 0, \ldots, q$, which would grow towards a spike at time q. The instant spike behaviour we have estimated with $q = 3$ is typical of river flows with a smaller area and non-chalk based catchments areas, where these types of rivers are known in the hydrology

area as *flashy*, see Stewart et al. (2008). So it would appear that, with a more careful choice of q, we have the flexibility with the Max-ARMA model to describe well the extremal features and key hydrology aspects of the time series profiles of rivers in different catchment sizes. Further work could investigate an approach for selecting the optimal (p, q) orders for these data or explore applying our inference procedure to rivers with smaller catchments.

From the perspective of the problem of barrier closure assessment for the River Thames, as the set out in Section 6.1, we have demonstrated the capability of a Max-ARMA model coupled with a Pareto marginal tail model to capture the core aspects of extremal dependence structure in river flows of the River Thames, i.e., the marginal tail decay, θ and χ_κ which primarily describe the magnitude and duration of the events more than the actual profile of events. It is these dependence aspects that are crucial for forecasting future closure rates of the Thames Barrier; as barrier closures are determined by peak sea levels per the tidal cycle (i.e., peak tides plus skew surge) and/or river flows in extreme states. Specifically, as extreme skew surge events tend to last 1-3 days yet extreme river levels can last up to 14-20 days with a mean cluster length of 5 days, it is possible that more than one extreme skew surge event can occur during the duration of an extreme river flow event. Accounting for this possibility is vital for determining barrier closure rates and their clustering in time. So long term simulations from these models provide good approximations to the observed series for considering barrier closure rate properties, as well as allowing for more extreme events than have been observed both in terms of their marginal sizes and their temporal duration. Given that river flow is independent of skew surge (Svensson and Jones, 2004), future work can look at combining simulations of both variables (using the skew surge model of (D'Arcy et al., 2023b), the deterministic predicted peak tide series, and the Max-ARMA model for river flow) to estimate future closure rates, and how they cluster in time, and ultimately understand the reliability of the barrier as we face unprecedented challenges

resulting from anthropogenic climate change.

6.7 Proofs

6.7.1 Proof of Proposition 6.2.3

Proof. The proof works by induction. Let P_k be the statement: All $\alpha_i > 0$ for $i = 1, \ldots, k$ where $k \leq p$ are identifiable if $\max_{i=1,\ldots,\lfloor k/2 \rfloor}\{\alpha_i \alpha_{k-i}\} \leq \alpha_k$ for all $k \leq p$.

For $k \geq 2$, assume that all statements $\{P_i\}_{i=2}^{k}$ hold. Then consider statement P_{k+1}. For α_{k+1} to be identifiable, we know from Remark 6.2.2 that

$$\max_{\mathcal{R}_{k+1}} \left\{ \prod_{\substack{i=1,\ldots,k: \\ \alpha_i > 0}} \alpha_i^{a_i} \right\} < \alpha_{k+1}. \tag{6.7.1}$$

This can be written as the maximum of the following components,

$$\max \left\{ \alpha_1 \max_{\mathcal{R}_k} \left\{ \prod_{\substack{i=1,\ldots,(k-1): \\ \alpha_i > 0}} \alpha_i^{a_i} \right\}, \max_{\mathcal{R}_{k+1}} \left\{ \prod_{\substack{i=2,\ldots,k: \\ \alpha_i > 0}} \alpha_i^{a_i} \right\} \right\} < \alpha_{k+1}, \tag{6.7.2}$$

where the first term above is any elements in the maxima set of expression (6.7.1) that contain α_1, whilst the second term is those without α_1 terms.

Since α_k is identifiable, we know from Remark 6.2.2 that,

$$\max_{\mathcal{R}_k} \left\{ \prod_{\substack{i=1,\ldots,(k-1): \\ \alpha_i > 0}} \alpha_i^{a_i} \right\} < \alpha_k.$$

Then inequality (6.7.2) becomes

$$\max \left\{ \alpha_1 \alpha_k, \max_{\mathcal{R}_{k+1}} \left\{ \prod_{\substack{i=2,\ldots,k: \\ \alpha_i > 0}} \alpha_i^{a_i} \right\} \right\} < \alpha_{k+1}.$$

Similarly, we rewrite the second term so that the above inequality becomes

$$\max\left\{\alpha_1\alpha_k, \alpha_2 \max_{\mathcal{R}_{k-1}}\left\{\prod_{\substack{i=1,\ldots,(k-2):\\ \alpha_i>0}} \alpha_i^{a_i}\right\}, \max_{\mathcal{R}_{k+1}}\left\{\prod_{\substack{i=3,\ldots,k:\\ \alpha_i>0}} \alpha_i^{a_i}\right\}\right\} < \alpha_{k+1},$$

where the second term is any elements containing α_2 and the third term is those without α_2. Since α_{k-1} is identifiable, we can rewrite this using Remark 6.2.2 as

$$\max\left\{\max\{\alpha_1\alpha_k, \alpha_2\alpha_{k-1}\}, \max_{\mathcal{R}_{k+1}}\left\{\prod_{\substack{i=3,\ldots,k:\\ \alpha_i>0}} \alpha_i^{a_i}\right\}\right\} < \alpha_{k+1}.$$

Continuing in this way for a further $(\delta-2)$ iterations where $\delta \in \mathbb{N}$ and $\delta > 2$, we obtain

$$\max\left\{\max\{\alpha_1\alpha_k, \alpha_2\alpha_{k-1}, \ldots, \alpha_\delta\alpha_{k-\delta}\}, \max_{\mathcal{R}_{k+1}}\left\{\prod_{\substack{i=(\delta+1),\ldots,k:\\ \alpha_i>0}} \alpha_i^{a_i}\right\}\right\} < \alpha_{k+1}. \tag{6.7.3}$$

Consider when $\delta = \lfloor k/2 \rfloor$, then

$$\max_{\mathcal{R}_{k+1}}\left\{\prod_{\substack{i=\lfloor k/2\rfloor+1,\ldots,k:\\ \alpha_i>0}} \alpha_i^{a_i}\right\} = \emptyset,$$

because $i = 1, \ldots, \lfloor k/2 \rfloor - 1$ are required to meet the conditions of the set $\mathcal{R}_{k+1}$, but we only consider $i = (\delta+1), \ldots, k$ and for these values, there are no solutions in $a_i \in (0, \ldots, k+1)$ such that $\sum_{i=\lfloor k/2\rfloor}^{p} i a_i = k+1$. For $\delta > \lfloor k/2 \rfloor$, this is always the case and so the above set is always empty. This is the first time when the second term in expression (6.7.3) becomes the empty set, because when $\delta = \lfloor k/2 \rfloor - 1$, this term becomes

$$\max_{\mathcal{R}_{k+1}}\left\{\prod_{\substack{i=\lfloor k/2\rfloor,\ldots,k:\\ \alpha_i>0}} \alpha_i^{a_i}\right\} = \alpha_{\lfloor k/2\rfloor}\alpha_{(k+1)-\lfloor k/2\rfloor}.$$

Therefore, $\delta = \lfloor k/2 \rfloor$, giving the required result, $\max\{\alpha_1\alpha_k, \alpha_2\alpha_{k-1}, \ldots, \alpha_{\lfloor k/2\rfloor}\alpha_{k-\lfloor k/2\rfloor}\} \leq$

α_{k+1}. □

6.7.2 Proof of Proposition 6.3.1

Proof. Let $\{X_t\}$ be a stationary Max-ARMA(p, q) process, defined in Section 6.2, under the conditions of Davis and Resnick (1989) (see Remark 6.2.1). Consider the maximum of the process $\{X_t\}$, first under an IID variables assumption (denote this sequence by $\{\hat{X}_t\}$) with an identical marginal distribution to that of $\{X_t\}$. Let $\hat{M}_n = \max\{\hat{X}_1, \ldots, \hat{X}_n\}$ be the maximum and the limiting non-degenerate distribution of the normalised $\hat{M}_n$ is denoted by $\hat{G}$. We find the distribution of the scaled maximum by exploiting this independence assumption,

$$\Pr(\hat{M}_n/n \leq x) = \exp(-n/nx) = \exp(-1/x) := \hat{G}(x) \quad \text{for } x > 0.$$

Next, we derive the same limiting distribution for the Max-ARMA process, i.e., without the independence assumption; denote this by $G(x)$. We begin by considering the case when $n = 3$ before considering the $n > q$ case, as the former reveals the key steps in the latter.

Case $n = 3$: By the definition of a Max-ARMA(p, q) we can write

$$\begin{aligned} M_3 &= \max\{X_1, \alpha_1 X_1, \beta_0 Z_2, \beta_1 Z_1, \alpha_1 X_2, \alpha_2 X_1, Z_3, \beta_1\beta_0 Z_2, \beta_2 Z_1\} \\ &= \max\{X_1, \alpha_1 X_1, \beta_0 Z_2, \beta_1 Z_1, \alpha_1^2 X_1, \alpha_1\beta_0 Z_2, \alpha_1\beta_1 Z_1, \alpha_2 X_1, \beta_0 Z_3, \beta_1 Z_2, \beta_2 Z_1\} \\ &= \max\big\{ \max\{1, \alpha_1, \alpha_2\} X_1, \max\{\beta_1, \alpha_1\beta_1, \beta_2\} Z_1, \max\{\alpha_1\beta_0, \beta_0, \beta_1\} Z_2, \beta_0 Z_3 \big\}. \end{aligned}$$

As $0 < \alpha_i < 1$ for all $i = 1, \ldots, p$ under stationarity and $\beta_0 = 1$, from Remark 6.2.1, we have

$$\Pr(M_3 \leq nx) = \Pr(X_1 \leq nx, \max\{\beta_1, \beta_2\} Z_1 \leq nx, \max\{\beta_0, \beta_1\} Z_2 \leq nx, Z_3 \leq nx).$$

Since Z_2 and Z_3 are independent of X_1 and Z_1, this joint probability can be factorised to give

$$\Pr(M_3 \leq nx) = \Pr(X_1 \leq nx, \max\{\beta_1, \beta_2\}Z_1 \leq nx) \Pr(\max\{\beta_0, \beta_1\}Z_2 \leq nx) \Pr(Z_3 \leq nx).$$

We simplify the joint probability of X_1 and Z_1 by

$$\begin{aligned}\Pr(X_1 \leq nx, \max\{\beta_1, \beta_2\}Z_1 \leq nx) &= \Pr(X_1 \leq nx) \Pr(\max\{\beta_1, \beta_2\}Z_1 \leq nx | X_1 \leq nx) \\ &= \Pr(X_1 \leq nx) \Pr(\max\{\beta_1, \beta_2\}Z_1 \leq nx | Z_1 \leq nx),\end{aligned}$$

where the conditioning event $\{X_1 \leq nx\}$ changes to $\{Z_1 \leq nx\}$ as this is the only information from the event involving X_1 that is relevant to $\{\max\{\beta_1, \beta_2\}Z_1 \leq nx\}$. It follows that,

$$\begin{aligned}\Pr(M_3 \leq nx) = &\Pr(X_1 \leq nx) \Pr(Z_3 \leq nx) \Pr(Z_2 \leq nx / \max\{\beta_0, \beta_1\}) \\ &\times \min\{\exp(-\gamma(\max\{\beta_1, \beta_2\} - 1)/(nx)), 1\},\end{aligned}$$

because if $\max\{\beta_1, \beta_2\} \leq 1$, then $\Pr(\max\{\beta_1, \beta_2\}Z_1 \leq nx | Z_1 \leq nx) = 1$ but otherwise, this conditional probability would need to be evaluated. This final term can be written more simply in our last expression for M_3 as

$$\begin{aligned}\Pr(M_3 \leq nx) = &\Pr(X_1 \leq nx) \Pr(Z_3 \leq nx) \Pr(Z_2 \leq nx / \max\{\beta_0, \beta_1\}) \\ &\times \exp\Big(- \frac{\gamma}{nx}(\max\{\beta_0, \beta_1, \beta_2\} - 1)\Big).\end{aligned}$$

Case $n > q$: Following the same logic as for the $n = 3$ case, using the notation $\beta_M = \max\{\beta_0, \ldots, \beta_q\}$ which, by definition of the Max-ARMA process, satisfies $\beta_M \geq 1$ and $\beta_{M:j} = \max\{\beta_0, \beta_1, \ldots, \beta_j\}$ so that $1 \leq \beta_{M:j} \leq \beta_M$ for all $j = 0, 1, \ldots, q$ and

$\beta_{M:q} = \beta_M$, the distribution of the rescaled maxima is

$$\begin{aligned}\Pr(M_n/n \leq x) = &\Pr(X_1 \leq nx) \prod_{j=0}^{q-1} \Pr\left(Z_{n-j} \leq nx/\beta_{M:j}\right) \prod_{i=2}^{n-q} \Pr\left(Z_i \leq nx/\beta_M\right) \\ &\times \exp\left(-\frac{\gamma}{nx}[\beta_M - 1]\right).\end{aligned}$$

As $X_t \sim$ Fréchet(1) and $Z_t \sim$ Fréchet(γ) with $0 < \gamma < \infty$ defined in Remark 6.2.1, we have

$$\begin{aligned}\Pr(M_n/n \leq x) = &\exp\left(-\frac{1}{nx}\right)\left\{\prod_{j=0}^{q-1} \exp\left(-\frac{\gamma\beta_{M:j}}{nx}\right)\right\} \\ &\times \exp\left(-\frac{\gamma\beta_M}{x}\left[1 - \frac{q-2}{n}\right]\right) \exp\left(-\frac{\gamma}{nx}(\beta_M - 1)\right) \\ = &\exp\left(-\frac{1}{nx}\left[1 + \gamma \sum_{j=0}^{q-1} \beta_{M:j}\right]\right) \exp\left(-\frac{\gamma\beta_M}{x}\left[1 - \frac{q-2}{n}\right]\right) \\ &\times \exp\left(-\frac{\gamma}{nx}(\beta_M - 1)\right) \\ \to &\left[\exp(-1/x)\right]^{\gamma\beta_M} = \left[\hat{G}(x)\right]^{\gamma\beta_M} \quad \text{as } n \to \infty.\end{aligned}$$

This limit follows as the first and last terms tend to 1 and in the second term $(q-2)/n \to 0$. From Section 6.3.1 the extremal index θ is defined via $G(x) = [\hat{G}(x)]^\theta$, hence $\theta = \gamma\beta_M = \gamma \max\{\beta_0, \beta_1, \ldots, \beta_q\}$.

□

6.7.3 Proof of Proposition 6.3.2

Proof. The definition of χ_κ can be rewritten for a stationary Max-ARMA process $\{X_t\}$ with unit Fréchet margins (using conditional probability and a Taylor expansion) as

$\chi_\kappa = \lim_{x\to\infty} \chi_\kappa(x)$ with, as $x \to \infty$,

$$\begin{aligned}\chi_\kappa(x) &= \frac{\Pr(X_{t+\kappa} > x, X_t > x)}{\Pr(X_t > x)} = \frac{\Pr(X_{t+\kappa} > x, X_t > x)}{1 - \exp(-1/x)} \\ &= \frac{\Pr(X_{t+\kappa} > x, X_t > x)}{1 - (1 - 1/x - \mathcal{O}(x^{-2}))} = x \Pr(X_{t+\kappa} > x, X_t > x)[1 + o(1)],\end{aligned}$$

where the calculations for the denominator follow as $X_t \sim$Frećhet(1). So we need to find the probability of the event $\mathcal{J}^X(x,\kappa) := \{X_{t+\kappa} > x, X_t > x\}$ for different lags κ. In this proof, we consider $\kappa \in \mathbb{N}$ only, but $\chi_\kappa = \chi_{-\kappa}$ holds, due to the symmetry of the above expression, so our results hold for $\kappa \in \mathbb{Z}$.

To determine the asymptotic behaviour of the joint probability in $\chi_\kappa(x)$ we make the following partition, with $\beta_M = \max\{\beta_0, \ldots, \beta_q\} \geq 1$ and $M^Z_{s:t} := \max\{Z_s, \ldots, Z_t\}$ for $s \leq t$,

$$\begin{aligned}\Pr(\mathcal{J}^X(x,\kappa)) &= \Pr(\mathcal{J}^X(x,\kappa), M^Z_{t+1:t+\kappa} < x/\beta_M) + \Pr(\mathcal{J}^X(x,\kappa), M^Z_{t+1:t+\kappa} \geq x/\beta_M) \\ &= \Pr(\mathcal{J}^X(x,\kappa), M^Z_{t+1:t+\kappa} < x/\beta_M)[1 + o(1)],\end{aligned}$$

as $x \to \infty$. The reason that the probability of the second term in the partition is smaller order than the probability of the first partition term is because it requires the occurrence of an extreme Z_j value from at least one of only κ values for j, which has a probability that is an order of magnitude smaller than not requiring this event.

First, we identify some conditions that control which type of extreme event the process exhibits. For $t \in \mathbb{Z}$ and $\delta \in \mathbb{N} \cup \{0\}$, we define the events

$$\begin{aligned}\mathcal{H}^Z(x,\delta,\kappa) &= \{M^Z_{t-q-\delta:t-1-\delta} < x/\beta_M, \beta_M Z_{t-\delta} > x, M^Z_{t-\delta+1:t+\kappa} < x/\beta_M\}, \\ \mathcal{H}^X(x,\delta) &= \{M^X_{t-p-\delta:t-\delta-1} < x/\alpha_M\},\end{aligned}$$

where $M^X_{s:t} := \max\{X_s, \ldots, X_t\}$ for $s \leq t$ and $\alpha_M = \max\{\alpha_1, \ldots, \alpha_p\}$ which, by definition of the Max-ARMA process, satisfies $\alpha_M > 0$. The event $\mathcal{H}^Z(x,\delta,\kappa)$ gives that there

is only one extreme value of the innovation process in the $(q+\delta+\kappa)$ values prior to and including at time $(t+\kappa)$. Specifically, at time $(t-\delta)$ the innovation exceeds x but all the other innovations are less than x as $x/\beta_M \leq x$. The event $\mathcal{H}^X(x,\delta)$ requires that prior to $(t-\delta)$, all p previous values of the Max-ARMA process are sufficiently small that they cannot produce an extreme value. So if event $\mathcal{H}(x,\delta,\kappa) := \mathcal{H}^Z(x,\delta,\kappa) \cap \mathcal{H}^X(x,\delta)$ occurs then the only feature of the Max-ARMA and innovation process up to $(t-\delta)$ that can produce a value of the Max-ARMA process that exceeds x at time $(t-\delta)$ and subsequently is $Z_{t-\delta}$.

Then we can rewrite the joint probability of interest by partitioning over the time when the large innovation $Z_{t-\delta}$ occurs before time t, i.e., for $\delta = \{0, \mathbb{N}\}$. Then we can express the probability as

$$\Pr(\mathcal{J}^X(x,\kappa), M^Z_{t+1:t+\kappa} < x/\beta_M) = \sum_{\delta=0}^{\infty} \Pr(\mathcal{J}^X(x,\kappa), \mathcal{H}(x,\delta,\kappa))[1+o(1)], \qquad (6.7.4)$$

as $x \to \infty$, where the additional inclusion of additional terms of $\mathcal{H}(x,\delta,\kappa)$ on the right hand side only change the probability by a little $o(1)$ term in x. We now focus on finding the asymptotic behaviour of a generic term in the sum in expression (6.7.4). Specifically,

$$\Pr(\mathcal{J}^X(x,\kappa), \mathcal{H}(x,\delta,\kappa)) \;=\; \Pr(\mathcal{H}(x,\delta,\kappa)) \Pr(\mathcal{J}^X(x,\kappa) \mid \mathcal{H}(x,\delta,\kappa)).$$

The marginal probability here can be written asymptotically as $\beta_M\gamma/x$ as $x \to \infty$ as the probability of the event $\{\beta_M Z_{t-\delta} > x\}$ has this limiting behaviour and all the rest of the finite events in $\mathcal{H}(x,\delta,\kappa)$ have probabilities tending to one. If event $\mathcal{H}(x,\delta,\kappa)$ occurred for large enough x, and all α_i and β_j coefficients were non-zero, then $X_{t-\delta} = Z_{t-\delta}$, $X_{t-\delta+1} = \max\{\alpha_1, \beta_1\} Z_{t-\delta} = \gamma_1 Z_{t-\delta}$, $X_{t-\delta+\tau} = \gamma_\tau Z_{t-\delta}$ for all $\tau \geq 1$, where γ_τ is defined by expression (6.3.4). Thus, under this conditioning, the process is deterministic given the value of $\beta_M Z_{t-\delta} > x$. When any of the Max-ARMA coefficients are zero there is

a potential for $\gamma_\tau = 0$ for some τ, and in these cases the associated $X_{t-\delta+\tau} = o_p(x)$ as $x \to \infty$. So we then have

$$\begin{aligned}
\Pr(\mathcal{J}^X(x,\kappa) \mid \mathcal{H}(x,\delta,\kappa)) &= \Pr(\gamma_\delta Z_{t-\delta} > x, \gamma_{\delta+\kappa} Z_{t-\delta} > x | \beta_M Z_{t-\delta} > x)[1+o(1)] \\
&= \Pr(\min\{\gamma_\delta, \gamma_{\delta+\kappa}\} Z_{t-\delta} > x | \beta_M Z_{t-\delta} > x)[1+o(1)] \\
&= \Pr(Z_{t-\delta} > x / \min\{\gamma_\delta, \gamma_{\delta+\kappa}\}) / \Pr(\beta_M Z_{t-\delta} > x)[1+o(1)] \\
&\to \min\{\gamma_\delta, \gamma_{\delta+\kappa}\},
\end{aligned}$$

as $x \to \infty$. As expression (6.3.4) shows, $\max_{\delta \in \mathbb{N}}(\gamma_\delta) \leq \beta_M$, so in the third equality the event of interest is a subset of the conditioning event. Also, the final limit exploits the property that $\Pr(Z_{t-\delta} > y) \sim \beta_M \gamma / y$ as $y \to \infty$.

Combining these results together we have that

$$\chi_\kappa = \lim_{x \to \infty} \chi_\kappa(x) = \lim_{x \to \infty} \Pr(\mathcal{J}^X(x,\kappa)) / \Pr(X_t > x) = \gamma \sum_{\delta=0}^{\infty} \min\{\gamma_\delta, \gamma_{\delta+\kappa}\}.$$

This sum will always converge because $\gamma_\tau \to 0$ as $\tau \to \infty$ so for large τ $\min\{\gamma_\delta, \gamma_{\delta+\kappa}\} < 1$ and each γ_τ geometrically decays for large τ.

□

6.8 Discussion

A key feature of all the stationary Max-ARMA(p,q) processes is that they are asymptotically dependent at all lags, i.e., $\chi_\tau > 0$, with this term defined by limt (6.3.3). Thus this class of processes excludes the possibility of asymptotic independence at different lags, i.e., $\chi_\tau = 0$ for at least some $\tau \geq 1$, see Ledford and Tawn (2003), so extensions of Max-ARMA(p,q) processes that allow for asymptotic independence are of interest. One such example is the power max-autoregressive (pMax-ARMA) process proposed by Ferreira (2011). This process takes the form $X_t = \max\{X_{t-1}^{\alpha_1}, Z_t\}$ with $0 < \alpha_1 < 1$

and Z_t an IID sequence of random variables with real positive support. They demonstrate that this process is asymptotically independent for all lags. However, they do not extend this for a general pMax-ARMA(p, q) process so the identifiability and stationarity constraints, that we address for the Max-ARMA process, are not relevant there; deriving conditions under which these properties hold would be an interesting avenue for further work. It would also be very interesting to develop a class of models that joins well between these two formulations, so that either asymptotic dependence or asymptotic independence can occur at different lags. Perhaps a way to achieve this is to change this innovation series to have a more rapidly decaying tail. An alternative is to change the process in expression (6.2.1) from combining terms using the L_∞-norm to instead being the L_r-norm, for some $1 < r < \infty$, as that provides a natural link between the Max-ARMA process and the L_1-norm case, corresponding to the standard ARMA(p, q) processes, with issues like this have been considered by Schlather (2001) in a different context.

Chapter 7

A marginal modelling approach for predicting wildfire extremes across the contiguous United States

7.1 Introduction

7.1.1 Motivation and data description

This paper details an approach to the data challenge organised for the EVA 2021 conference. The subject of the challenge was wildfire modelling, and two important sub-challenges were proposed within this setting. In particular, teams were asked to develop methods for predicting the number of fires (i.e., individual fires that are separated in space), as well as the amount of burnt land resulting from these fires, over different months for gridded locations across the continental United States (US).

In the absence of mitigation, wildfires can have devastating consequences, including loss of life and damage to property. The northern California wildfire in October 2017 burned approximately 150,000 acres of land, resulting in 7,000 damaged structures and 100,000 evacuations (Wong et al., 2020). Recent increases in both the number and

severity of wildfires can be linked to climate change, and in particular to anthropogenic warming (Jones et al., 2020). Focusing specifically on the western US, Zhuang et al. (2021) demonstrate that a high proportion of the observed increases in weather events leading to wildfires may be attributed to this aspect of climate change. Extreme events in wildfire modelling are especially important; the more individual wildfires that occur, the greater the potential destruction, and the impact of large wildfires (in terms of the amount of land area burnt) can be particularly devastating. It is therefore of interest to develop models for wildfires, and in particular wildfire extremes.

The challenge data set consists of monthly wildfire count (CNT) and burnt area (BA) observations from 1993 to 2015 at 3,503 grid cell locations spanning the contiguous US. There are 35 auxiliary variables also recorded relating to land cover types, climate and altitude. Observation locations are arranged on a $0.5^\circ \times 0.5^\circ$ (approximately 55 km $\times$ 55 km) regular grid of longitude and latitude coordinates, with observations recorded from March to September; further details are provided by Opitz (2023).

In order to compare the predictions produced by the teams participating in the data challenge, several observations were removed from the data to act as a validation set; this contained 80,000 observations for each of CNT and BA. The selection of these validation points was not done completely at random, so there is some spatio-temporal dependence between them. This will be discussed further in Section 7.3.4, with a pictorial example given in Figure 7.3.1. Let CNT_i and BA_i, $i = 1, \ldots, N$, denote the i-th observation of the wildfire CNT or BA data, respectively, where $N = 563,983$ is the total number of observations across the training and validation sets for each variable over all sites, months and years. We denote the set of observation indices in the validation sets for CNT and BA by $CNT^{val}, BA^{val} \subset \{1, \ldots, N\}$, respectively, with $|CNT^{val}| = |BA^{val}| = 80,000$. An important feature is that the validation indices are not identical for the CNT and BA data, but there is a reasonable overlap, i.e., $CNT^{val} \neq BA^{val}$ but $CNT^{val} \cap BA^{val} \neq \emptyset$. We discuss ways to exploit this aspect in

Section 7.2.2.

The objective of the challenge was to predict cumulative probability values for both CNT and BA at the times and locations in their respective validation sets. The resulting estimates were then ranked using a score computed from the true observed values, with lower scores corresponding to more accurate probability predictions. These scores were weighted so that more importance is placed on the estimation of the extremes; see Opitz (2023). Statistical techniques that do not explicitly model the tail are therefore unlikely to produce the best scores.

7.1.2 Data exploration

In this section, we give an overview of the features of the data set that motivate our modelling approach. We consider the relationship between CNT and BA, as well as the temporal non-stationarity of each variable separately; we also investigate how these features vary over the spatial domain.

We begin by exploring the dependence between CNT and BA; for the bulk of the data, we consider Kendall's τ measure of rank correlation, whilst for the extremes we consider the widely-used measures χ and $\bar{\chi}$. Consider a random vector (X, Y) with marginal distribution functions F_X and F_Y, respectively. Coles et al. (1999) define $\chi = \lim_{u \to 1} \chi(u)$, where $\chi(u) = \Pr(F_Y(Y) > u \mid F_X(X) > u) \in [0, 1]$, as a measure of asymptotic dependence. If $\chi \in (0, 1]$, X and Y are said to be asymptotically dependent, with $\chi = 1$ corresponding to perfect dependence. Asymptotic independence between X and Y is present only when $\chi = 0$, meaning that χ fails to signify the level of asymptotic independence. To account for this, Coles et al. (1999) define a further measure that provides additional detail in this case, namely $\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u) \in (-1, 1]$ where

$$\bar{\chi}(u) = \frac{2 \log \Pr(F_Y(Y) > u)}{\log \Pr(F_Y(Y) > u, F_X(X) > u)} - 1.$$

Under asymptotic dependence, $\bar{\chi} = 1$, and for asymptotic independence, $\bar{\chi} < 1$; the further sub-cases $\bar{\chi} \in (0, 1)$ and $\bar{\chi} \in (-1, 0)$ correspond to positive and negative association, respectively, while $\bar{\chi} = 0$ indicates independence.

We estimate these measures separately for subsections of the US to investigate spatial variability in the dependence structure between CNT and BA. We start by splitting the spatial domain into quadrants corresponding to the north east (NE; $>$ 37.5°N, $<$ 100°W), south east (SE; $\leq$ 37.5°N, $<$ 100°W), south west (SW; $\leq$ 37.5°N, $\geq$ 100°W) and north west (NW; $>$ 37.5°N, $\geq$ 100°W). Kendall's τ measure suggests strong overall correlation between CNT and BA, with estimates of 0.926 $(0.925, 0.927)$, 0.827 $(0.825, 0.829)$, 0.858 $(0.855, 0.860)$ and 0.868 $(0.867, 0.870)$ for the NE, SE, SW and NW respectively, with the values in brackets denoting 95% confidence intervals obtained via bootstrapping. However, estimates of $\chi(u)$ and $\bar{\chi}(u)$ suggest this dependence diminishes in the extremes, leading to asymptotic independence. We obtain estimates (and 95% bootstrap confidence intervals) of $\chi(0.999) = 0.071$ $(0.043, 0.126)$, 0.038 $(0.017, 0.072)$, 0.012 $(0, 0.024)$ and 0.043 $(0.024, 0.077)$, and $\bar{\chi}(0.999) = 0.438$ $(0.343, 0.521)$, 0.282 $(0.191, 0.392)$, 0.092 $(-0.05, 0.179)$ and 0.317 $(0.253, 0.413)$, for the NE, SE, SW and NW regions respectively. The NE region exhibits the strongest dependence between CNT and BA in the bulk of the data, as well as the strongest extremal dependence. We extended this analysis to look at smaller spatial domains, but our conclusions did not change.

Figure 7.1.1 shows the spatial distribution of average CNT and BA values in two different time groupings: in the summer months (May, June, July and August; MJJA), when wildfires are more likely to occur, and in the remaining cooler months (March, April and September; MAS). The highest average CNT values are observed in the east for MAS and the west for MJJA. The highest average BA values typically occur in the west of the US during MJJA whilst the majority of the eastern US locations have relatively low average BA values in both time groups, with the exception of Florida.

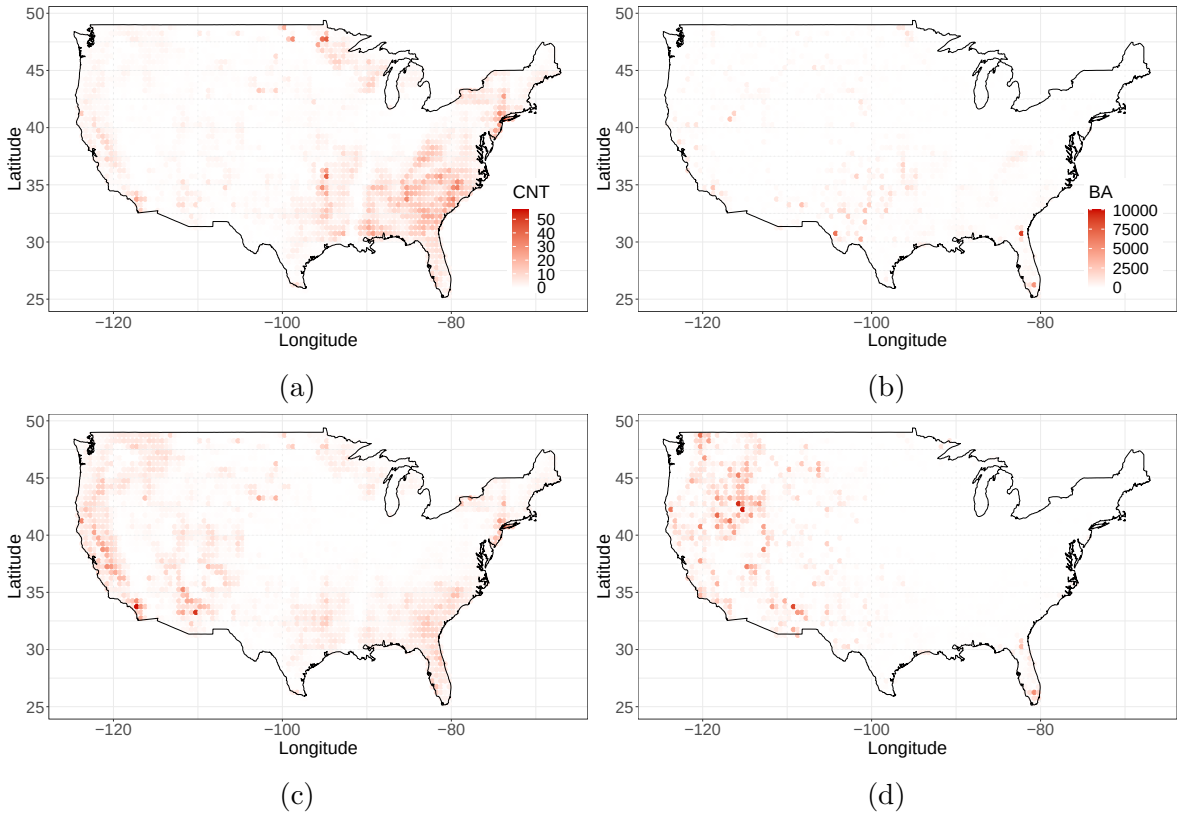
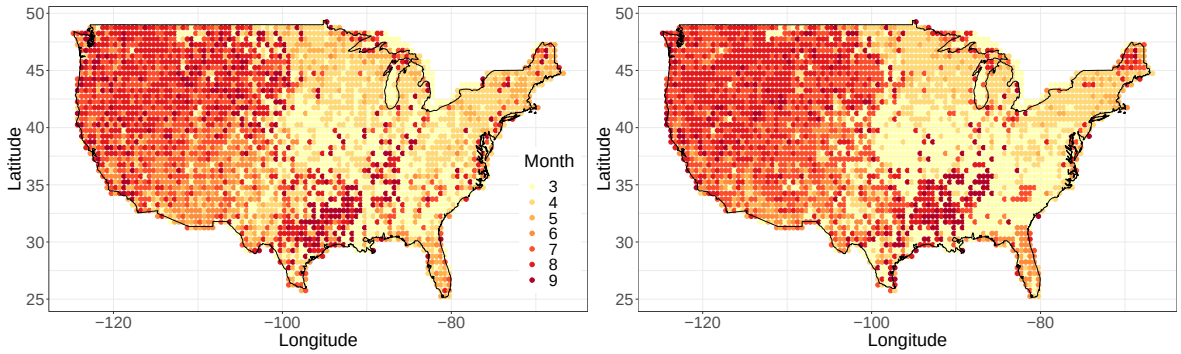


Figure 7.1.1: Average CNT (a & c) and BA (b & d) across all years for each grid cell, for MAS (a & b) and for MJJA (c & d).

This demonstrates that there is both spatial and temporal variability in the wildfire observations.

To further demonstrate this spatio-temporal variability, Figure 7.1.2 illustrates the months when the maximum CNT and BA observations occur for each grid cell. In the eastern US, the maxima of each variable tend to occur in July or August (shown by red points) whilst in the west, the maxima typically occur in March and April (illustrated by lighter yellow points). As global temperatures rise with anthropogenic climate change, the frequency and intensity of wildfires are generally expected to increase (Preisler et al., 2004; Wuebbles et al., 2017). To investigate this, we fit a linear model between year and annual mean CNT and BA separately, assuming independence across annual means. We find significant trends for both CNT and BA. Therefore, assuming stationarity across the entire spatial domain over the observation period would be unreasonable.

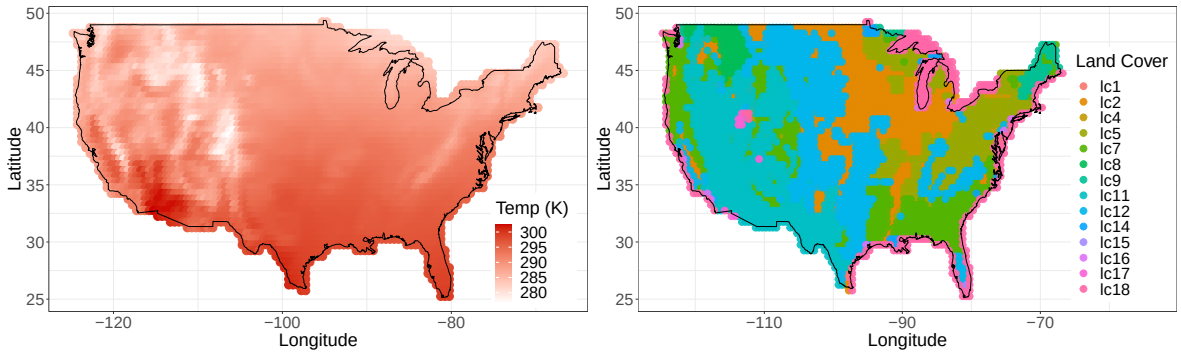


(a) (b)

Figure 7.1.2: Month where the maximum CNT (a) and BA (b) across all years occurs for each grid cell.

Due to the nature of wildfires, we expect to observe relationships between both CNT and BA observations and certain climate variables. For example, high temperature coupled with low rainfall and low wind speed are the ideal conditions for wildfires to ignite and spread (Holden et al., 2018; Son et al., 2021). No significant linear relationships exist for either wildfire variable with any of the climate covariates, suggesting such relationships are complex in nature. Figure 7.1.3(a) shows the average temperature for each grid cell; temperature is non-stationary across the US but there is some spatial dependence, with nearby locations exhibiting similar values. Some form of spatial dependence exists for all climate variables. Since these variables are given as monthly averages, it is difficult to associate these covariates directly with the wildfire observations, which are also given as monthly aggregates.

Another factor likely to alter wildfire behaviour across the US is the type of land cover. For example, locations with large proportions of water or urban areas are typically not conducive to wildfires, whilst those with forest areas probably are. Eighteen land cover variables, given as proportions of each grid cell, are provided in the challenge data set; these are denoted $lc(j)$ for $j = 1, \ldots, 18$ and defined in Opitz (2023). Figure 7.1.3(b) illustrates the maximum land cover variable for each location. Spatial heterogeneity can be observed over different regions. For example, a large portion



(a) (b)

Figure 7.1.3: Mean temperature in Kelvin (a) and the most common land cover variable (b) for each location.

of the western US is taken up by shrubland ($lc(11)$), whereas the eastern region is dominated by cropland ($lc(j)$ for $j = 1, 2, 3$) and tree-based land cover types ($lc(j)$ for $j = 5, 6, 7, 8$). Unsurprisingly, many coastal locations are predominantly covered by water ($lc(18)$), and regions containing national forests (such as Kootenai and Stanislaus) are easily identifiable, since they are mostly made up of tree categories.

7.1.3 Existing methods

Various methods exist for modelling and predicting wildfire frequency and intensity. For example, generalised additive models (GAMs) with climatic, anthropogenic and/or spatial covariates are commonly used; see, e.g., Krawchuk et al. (2009) or Sá et al. (2018). The latter captures covariate information via a fire index; many such indices have been proposed within the literature (Ziel et al., 2020). Each index is typically developed with country-specific considerations in mind, such as land cover types and climate factors, and are often used by government bodies to assess risks and prioritise fire responses. In the US, the National Fire-Danger Rating System is the primary tool used for wildfire management (Cohen and Deeming, 1985). There have been attempts in the literature to use fire indices as a means to model extreme wildfire events (Koh et al., 2023). However, several approaches have found that certain fire indices are poor

predictors of wildfires. For example, Sharples et al. (2009) show that the Forest Fire Danger Index, typically used in Australia, is inadequate for predicting the behaviour of moderate to high-intensity wildfires.

Machine learning techniques have also been adopted for wildfire modelling: Richards and Huser (2022) and Ivek and Vlah (2023) use deep learning techniques; Cisneros et al. (2023) present a four-stage process including a random forest algorithm; and Koh (2023) develops a gradient boosting model trained with loss functions appropriate for predicting extreme values. We take a simpler, marginal-based approach.

The remainder of this paper is structured as follows. In Section 7.2, we illustrate how certain properties of the training data can be exploited to infer a subset of probability estimates for observations in the validation set. In Section 7.3, we introduce our marginal modelling techniques for both CNT and BA. We also discuss our technique for estimating spatial neighbourhoods and corresponding tuning parameters. We conclude with a discussion of our approach in Section 7.4.

7.2 Exploiting properties of the training data set

7.2.1 Re-scaling burnt area values

In this section, we discuss various properties of the wildfire data set, and how these can be exploited to improve the estimation of the predictive distributions for missing observations.

To begin, observe that BA is an absolute measurement; this results in varying measurement scales across different locations. To better understand this, consider that some grid cells in the data set do not lie completely inside the continental US; this feature is captured by the 'area' variable, denoted p_i, $i = 1, \ldots, N$, which describes the proportion of each grid cell that lies in the region of interest. BA observations depend upon this variable since for grid cells with smaller area values, there is less available

land for wildfires to occur and hence lower BA values. For these reasons, the raw BA observations cannot be easily compared across locations.

To account for this, we propose re-scaling BA observations to ensure all observations are on a unified, relative scale. Recall that BA_i, $i = 1, \ldots, N$, with $N = 563,983$, denotes the i-th observation of the BA data, and that $BA^{val} \subset \{1, \ldots, N\}$ is the set of indices for missing BA observations. We consider here the i-th observation, with corresponding grid cell area $p_i \in (0, 1]$. For each $i \in \{1, \ldots, N\}$, the total surface area of the grid cell is computed by taking the corresponding longitude and latitude coordinates and applying a formula derived from Archimedes' theorem (Kelly and Šavrič, 2021). We denote these surface area values by SA_i. The surface area contained within the continental US is then computed by multiplying the total surface area by the grid cell area variable, i.e., $SA_i \times p_i$. We denote these values by SA_i^*: such values will naturally vary between locations, especially for locations lying on a borderline. Moreover, SA_i^* values naturally decrease going from South to North of the continent, since grid cells defined using longitude and latitude suffer from unequal cell sizes (Budic et al., 2016). We refer to this variable as the true surface area.

Using this variable, we derive a relative measure for BA, which we term burnt area proportion (BAP), i.e., for each i, define $BAP_i := BA_i/SA_i^* \in [0, 1]$. This value denotes the proportion of the true surface area that has been burnt for each observation. It is arguably a better indicator of the impact and/or severity of wildfire events compared to raw BA observations, since it puts the absolute magnitude in context for each location. Moreover, this proportion is a relative measure, meaning the data for all locations are on the same scale; this allows for a more straightforward comparison between neighbouring observations with different (true) surface areas.

We recall that the objective of the data challenge is to obtain probability estimates

of the form $\Pr(BA_i \leq u)$ for all $u \in \mathcal{U}_{BA}$, where

$$\mathcal{U}_{BA} = \{0, 1, 10, 20, 30, \ldots, 100, 150, 200, 250, 300, 400, 500, 1000,$$
$$1500, 2000, 5000, 10000, 20000, 30000, 40000, 50000, 100000\},$$

and $i \in BA^{val}$. This can be derived using the marginal distribution of BAP_i, since

$$\Pr(BA_i \leq u) = \Pr(BAP_i \times SA_i^* \leq u) = \Pr(BAP_i \leq u/SA_i^*).$$

Consequently, we evaluate the distribution function of BAP_i for all $u \in \mathcal{U}_{BAP}^i$, where $\mathcal{U}_{BAP}^i := \mathcal{U}_{BA}/SA_i^*$, to obtain the required predictive probabilities. We introduce our technique for estimating this distribution function in Section 7.3.3.

We can also use these proportional data to deduce information about the upper tail of the distribution for BAP_i at any $i \in BA^{val}$. Since it is impossible to observe a BA observation which exceeds the true surface area at any location, we can immediately deduce that $\Pr(BAP_i \leq u) = 1$ for any $u \in \mathcal{U}_{BAP}^i$ with $u \geq 1$. In practice, over 1% of missing BA observations satisfied the inequality $\max\{\mathcal{U}_{BAP}^i\} \geq 1$, meaning a non-negligible amount of information can be uncovered via this preliminary step.

We considered a similar re-scaling for CNT observations; however, there did not appear to be any obvious relationship between the true surface area and CNT values. Furthermore, unlike BA, no natural upper bound arises for CNT observations, so we cannot deduce properties of the upper tail distribution for missing observations.

7.2.2 Exploiting features of the missing data

Before introducing our marginal modelling procedures, we highlight how the training data can be used to provide information about the missing values we are required to estimate. This is possible since the missing values in the CNT and BA variables do not always occur at the same space-time locations, although there is some overlap in

their missingness. We show that if exactly one of the CNT or BA values is known at a particular index, we can deduce information about the other.

Recall that we are interested in estimating the predictive distribution of CNT_i for some $i \in CNT^{val}$, i.e., $\Pr(CNT_i \leq u)$ for $u \in \mathcal{U}_{CNT}$, where $u \in \mathcal{U}_{CNT} = \{0, 1, \ldots, 9, 10, 12, \ldots, 30, 40, \ldots, 100\}$. If $i \notin BA^{val}$ and $BAP_i = 0$, we can immediately deduce that $CNT_i = 0$ and $\Pr(CNT_i \leq u) = 1$ for all $u \in \mathcal{U}_{CNT}$. Moreover, if $i \notin BA^{val}$ and $BAP_i > 0$, we have that $CNT_i > 0$, implying $\Pr(CNT_i \leq 0) = 0$, though we are still required to estimate the predictive distribution for all $u \in \mathcal{U}_{CNT} \backslash \{0\}$. The values we can infer for BAP_i from CNT_i, with $i \in BA^{val}$, are analogous, so the detail is omitted here.

We find that $CNT_i = 0$ for approximately 23% of the points in the CNT validation set, and that $CNT_i > 0$ for an additional 15%. We can also deduce similar proportions for the BAP values we are required to predict. A reasonable amount of information can therefore be uncovered using this simple step.

We also found that for the non-missing CNT and BA observations, the probability of observing a zero observation exceeded 0.999 for both variables when $lc(18)_i > 0.94$, where $lc(18)_i$ denotes the proportion of each location covered by water. Therefore, for any $i \in CNT_{val}$ ($i \in BA_{val}$) with $lc(18)_i > 0.94$, we set $CNT_i = 0$ ($BAP_i = 0$), implying $\Pr(CNT_i \leq u) = 1, \forall\ u \in \mathcal{U}_{CNT}$ ($\Pr(BAP_i \leq u) = 1, \forall\ u \in \mathcal{U}^i_{BAP}$).

As well as improving estimates of the predictive distribution of some locations, the additional steps introduced in this section also increase the amount of information available. This aids the marginal estimation procedures detailed in Sections 7.3.2 and 7.3.3.

7.3 Marginal modelling of missing values

7.3.1 Neighbourhood selection

For our approach, we make the following assumption: for any observation with index $i \in \{1, \ldots, N\}$, there exists some spatial neighbourhood of indices, $\mathcal{N}_i$, where all corresponding observations come from the same marginal distribution. Through the estimation of this distribution, we can obtain predictive probabilities for missing CNT and BAP observations. In this section, we introduce our approach for selecting these neighbourhoods for CNT observations; the approach for BAP is analogous.

Consider the observation with index $i \in \{1, \ldots, N\}$, and denote the corresponding spatial location, month and year by $\boldsymbol{s}_i \in \mathbb{R}^2$, $m_i \in \{3, \ldots, 9\}$ and $y_i \in \{1993, \ldots, 2015\}$, respectively. We define the spatial neighbourhood as

$$\mathcal{N}_i := \left\{ j \in \{1, \ldots, N\} : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq k_1^{CNT}, m_j = m_i, y_j = y_i \right\}, \tag{7.3.1}$$

for some $k_1^{CNT} \geq 0$, i.e., the indices of all observations occurring in the same year and month as observation i with a spatial distance of at most k_1^{CNT} from $\boldsymbol{s}_i$. The spatial distance $\|\cdot\|$ is measured in kilometres (km) using the Haversine formula; in practice, these are calculated via the `distm` function in the R package `geosphere` (Hijmans, 2019). We treat k_1^{CNT} as a tuning parameter and introduce a cross validation technique to select it in Section 7.3.4. We denote the CNT values corresponding to neighbourhood $\mathcal{N}_i$ by $CNT^{\mathcal{N}_i} = \{CNT_j : j \in \mathcal{N}_i\}$. The definitions of k_1^{BAP} and $BAP^{\mathcal{N}_i}$ for $i \in \{1, \ldots, N\}$ are analogous.

More complex spatial neighbourhoods, which incorporated temporal and covariate-based information, were also considered but ultimately resulted in worse quality marginal estimates. This is discussed in Appendix D, where we present prediction scores for other neighbourhoods we considered.

7.3.2 A parametric approach for modelling CNT

Following Joseph et al. (2019), we assume all observations in the set $CNT^{\mathcal{N}_i}$ follow a zero-inflated negative binomial distribution for all $i \in CNT^{val}$, i.e., for any $CNT \in CNT^{\mathcal{N}_i}$, we have that

$$\Pr(CNT = j) = \begin{cases} \pi + (1-\pi)g(0) & \text{if } j = 0 \\ (1-\pi)g(j) & \text{if } j > 0, \end{cases} \tag{7.3.2}$$

where $\pi \in [0, 1]$ denotes a probability and $g(j),\ j \geq 0$, is the probability mass function of the negative binomial distribution. We estimate the parameter π and those of the negative binomial distribution using likelihood inference. We then evaluate distribution (7.3.2) for all $u \in \mathcal{U}_{CNT}$ using the estimated parameters, resulting in the predictive distribution for the missing observation CNT_i. In practice, we use the same tuning parameter, k_1^{CNT}, for all $i \in CNT^{val}$; we discuss our approach to selecting this value in the Section 7.3.4.

7.3.3 A semi-parametric approach for modelling BAP

Given any $i \in BA^{val}$, we assume all observations in the set $BAP^{\mathcal{N}_i}$ follow the semi-parametric marginal distribution given in Richards et al. (2022). This distribution was proposed for modelling precipitation data, which are similar to wildfire data in the sense that they typically contain a large number of zero observations. These data structures are referred to as mixture distributions, since they are a mix of a discrete (zero observations) and a continuous (positive BAP observations) process. Values in the bulk of the data, including zeros, are modelled empirically, while values in the upper tail are modelled using a generalised Pareto distribution (GPD). This distribution is typically referred to in the context of the 'peaks over threshold' approach (Balkema and de Haan, 1974; Pickands, 1975), whereby a GPD is fitted to independent and identically

distributed exceedances of a high threshold. This overall marginal model is given by

$$\Pr(BAP \leq x) = \begin{cases} z_i & \text{if } x = 0 \\ \frac{1-\lambda_i - z_i}{F_i^*(u_i)} F_i^*(x) + z_i & \text{if } 0 < x \leq u_i \\ 1 - \lambda_i(1 - H_{u_i}(x)) & \text{if } x > u_i, \end{cases} \tag{7.3.3}$$

for all $BAP \in BAP^{\mathcal{N}_i}$, where z_i is the probability of observing a zero, u_i is some high threshold to be chosen, $\lambda_i = \Pr(BAP > u_i)$, F_i^* is the distribution function of strictly positive observations, and $H_{u_i}(x)$ denotes the cumulative distribution function of the GPD, i.e., $H_{u_i}(x) = 1 - [1 + (\xi_i(x - u_i))/(\sigma_i)]_+^{-1/\xi_i}$, with $x_+ = \max\{x, 0\}$ and $(\sigma_i, \xi_i) \in \mathbb{R}_+ \times \mathbb{R}$. We refer to σ_i and ξ_i as the scale and shape parameters, respectively. See Coles (2001) for a more detailed discussion of the peaks over threshold approach.

We set $k_2^{BAP} := 1 - \lambda_i$ for all $i \in BA^{val}$ and treat k_2^{BAP} as another tuning parameter, which we again estimate using cross validation; see Section 7.3.4. Both u_i and z_i can be estimated empirically, alongside F_i^*. Note that this marginal model is only valid when $z_i < 1 - \lambda_i$: in such cases, the GPD scale and shape parameters are estimated using likelihood inference. We then evaluate the distribution described in equation (7.3.3) at the fitted parameters for all $u \in \mathcal{U}_{BAP}^i$, resulting in the predictive distribution for the missing observation BAP_i.

In the cases when $z_i \geq 1 - \lambda_i$ (i.e., the marginal model is not valid), we use a fully empirical distribution. Such cases occur when the estimated threshold equals zero, corresponding to neighbourhood sets containing a significant proportion of zeros, indicating a low occurrence of wildfires.

7.3.4 Tuning parameter selection

We now consider how to select the tuning parameters k_1^{CNT}, k_1^{BAP} and k_2^{BAP} used in our marginal modelling approaches. One option is to use leave-one-out cross validation

and select the tuning parameter values that minimise the score used for ranking in the data challenge: see Opitz (2023) for more information. However, the locations for the validation data are not randomly distributed across the spatial domain, and are generally clustered in space and time. We demonstrate this in Figure 7.3.1, where we show the locations of the CNT validation data for March 1994; the resulting plots have similar features for BAP, as well as for different months and years. In the case of BAP data, this implies that for a fixed value of k_1^{BAP}, there will be a larger number of missing values in the set $BAP^{\mathcal{N}_i}$ for $i \in BA^{val}$ than for $i \notin BA^{val}$, on average. The same holds when considering CNT data for a fixed value of k_1^{CNT}. This feature of the validation set means that using standard leave-one-out cross validation over all training locations could lead to selecting smaller neighbourhoods than are really appropriate.

We instead propose to carry out the parameter selection procedure using only a subset of the observations in the training data. Focusing on BAP, for each $i \in BA^{val}$ and any combination of (k_1^{BAP}, k_2^{BAP}) values, we allow the observation indexed by

$$\underset{j:m_i=m_j,y_i=y_j,j\notin BA^{val}}{\arg\min} \|\boldsymbol{s}_i - \boldsymbol{s}_j\|,$$

to contribute to the score, i.e., giving the spatially-nearest non-missing observation that occurs in the same month and year as observation i. Ties may be broken at random, or using any rule that results in only one nearest neighbour per location. Since these locations can be the nearest neighbour of more than one validation location, some of the corresponding observations are included more than once in the score calculation. An alternative would have been to use each of these observations only once to avoid duplicates, but this means some validation locations would not be represented in the score calculation.

We consider the following candidate values for the tuning parameters: $k_1^{BAP} \in \{50, 75, \ldots, 400\}$, $k_2^{BAP} \in \{0.05, 0.10, \ldots, 0.95\}$. For each combination of candidate values, we recalculate the score function proposed in Opitz (2023), summing over all

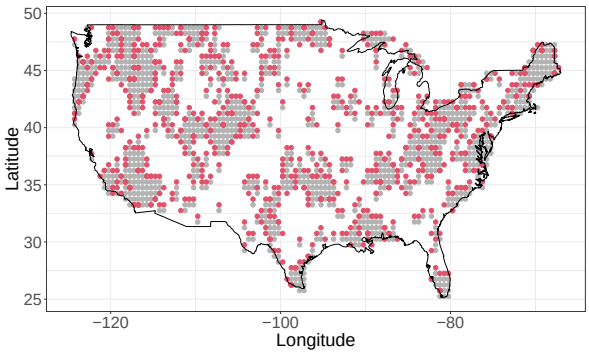


Figure 7.3.1: Locations in the set CNT^{val} for March 1994 (grey) and the corresponding locations of observations for tuning parameter selection (red).

values corresponding to our set of nearest neighbours, before finally selecting the parameter combination that minimises the score. The procedure in the CNT case is analogous, albeit without the GPD quantile parameter k_2^{BAP}. In Figure 7.3.1, we demonstrate the locations of observations that contribute to the tuning parameter selection procedure for CNT in March 1994. This results in selected tuning parameter values of $k_1^{BAP} = 175$, $k_2^{BAP} = 0.5$, and $k_1^{CNT} = 125$.

We note that our final approach has similarities with the winning entry to the 2017 EVA data challenge (Stephenson et al., 2018), where the authors combine data across locations with sufficient observations, in order to fit a generalised extreme value distribution for predicting precipitation extremes. They advocate the use of cross validation for tuning parameter selection and to compare potential modelling approaches in data challenges such as this, where the aim is to optimise some pre-determined metric.

7.4 Discussion of limitations and possible extensions

In this paper, we have discussed a marginal modelling approach for predicting wildfire events across the contiguous US. This framework was applied to obtain estimates of the cumulative distribution function at locations with missing entries for either CNT or BA. The resulting estimates were then "ranked" using a score function weighted to give

higher importance to extreme observations (Opitz, 2023). Our method produced scores of 4080.559 and 3640.92 for CNT and BA, respectively, resulting in an overall score of 7721.479; this is a significant improvement on the proposed benchmark technique.

Unlike all the techniques introduced in Section 7.1.3, our approach does not attempt to specify the relationships between the auxiliary and wildfire variables. Such relationships appear to be complex and non-linear in nature, which may be explained by a variety of hypotheses. For example, the monthly aggregated format of the wildfire variables arguably makes it more difficult to associate them with any climate covariates, which are given as monthly means. Instead, our approach relies on the assumption that wildfire observations within spatial neighbourhoods arise from the same marginal distribution. We believe that this is realistic since neighbouring locations are likely to have similar auxiliary covariates, as demonstrated in Figure 7.1.3, and a large wildfire event occurring at one location is likely to increase the probability of wildfires in neighbouring locations. Furthermore, since the values in neighbourhood sets vary over time for each missing CNT or BA observation, our approach accounts for the temporal non-stationarity discussed in Section 7.1.2. We also propose several preliminary steps in Section 7.2; these steps do not require expert knowledge of wildfires to implement. Furthermore, such steps lead to significant improvements in the predictive distributions obtained using our approach by increasing the amount of information available and bringing all BA observations onto a unified scale.

While developing our approach, we investigated the possibility of accounting for covariate influence on extreme CNT values via the use of GAMs (this option was also mentioned in Section 7.1.3), but found their predictive performance to be poor in this setting; see Wood (2017) for details on these types of models. In particular, we fitted a continuous GPD for each month, with the scale parameter having a GAM form (Youngman, 2019) comprising spatial and climatological covariates; a continuous approximation was used due to having discrete CNT values. The advantage of this method

is that covariate effects can be directly assessed by examining the smooth functions underlying the models. From the fitted GAMs, the general spatial behaviour of the CNT data was modelled fairly well in each month, but these models did appear to suffer from oversmoothing, even when using models with the lowest level of smoothness. On the other hand, we found that physically-interpretable covariate behaviour for the climatological variables was hard to capture, making model selection difficult; the precise reason for this is unclear. It is likely that the aforementioned oversmoothing combined with other issues, such as poor convergence of the underlying numerical optimisation routines and difficulties combining the fitted GAMs with models for the bulk of the CNT values, lead to poor model performance against the benchmark. Therefore, it appears that this type of approach may not be favourable in situations where the prediction of unknown values is required, and is more suited to analyses where the aim is to account for uncertainty whilst modelling complex covariate effects. Indeed, in addition to the GAM-based approach mentioned in Section 7.1.3, Sá et al. (2018), Zhang et al. (2017) and Rodríguez-Pérez et al. (2020) have also successfully applied GAM techniques for modelling wildfire data.

One possible extension of the modelling techniques proposed in Sections 7.3.2 and 7.3.3 would be to introduce weights into the marginal estimation procedures. In the current format, observations within spatial neighbourhoods are given equal weights, even though it is likely that locations with a closer proximity to a missing observation would provide more useful information than locations that are further away. Our current method could therefore be extended by introducing weights to the marginal estimation procedures, with closer observations contributing more to probabilistic estimates. We would expect different values of the tuning parameters k_1^{CNT} and k_1^{BAP}, defining the spatial range of the neighbourhoods, to be appropriate in this case, but our cross validation approach could be used analogously.

Changes to the definition of the neighbourhoods in equation (7.3.1) could lead to

improvements with our approach. One drawback with our current implementation is that the values k_1^{CNT} and k_1^{BAP} are chosen to be the same across all validation locations. Although we selected these tuning parameters carefully via cross validation, it is possible that this is an over-simplification and allowing the values to depend on covariates such as location, month or year may have been more appropriate. An extension of our approach could allow for this possibility, e.g., by separating the spatial domain into smaller sections and implementing cross validation separately in each one. It may also be reasonable to apply clustering algorithms as a preliminary step, to inform the spatial regions where setting the tuning parameters (k_1^{CNT}, k_1^{BAP}) as constant is a reasonable assumption. Allowing these parameters to vary across space also has the potential to provide insight into the behaviour of wildfires across the spatial domain. Additionally, we considered allowing the neighbourhoods themselves to depend on covariate-based clusters or to cover larger time windows, but the results presented in Appendix D suggest the simpler spatial neighbourhood approach was more successful.

While the zero-inflated negative binomial distribution proposed for CNT neighbourhoods is not motivated by extreme value theory, our analysis indicated the fitted marginal distributions performed reasonably well, including in the upper tail in the majority of cases. Several other distributions were tested, including fully empirical and discrete GPD models (Hitz et al., 2017); however, in every case, these distributions resulted in poorer prediction quality when ranked by the objective function given in Opitz (2023). This is likely due to the difficulties that arise in trying to capture behaviour in the bulk and tail simultaneously, and perhaps due to an insufficient amount of data in each of our spatial neighbourhoods for fitting the discrete GPD. In addition, alternative marginal distributions for BAP observations have the potential to further improve the predictive ability of our modelling framework.

Chapter 8

Extreme value methods for estimating rare events in Utopia

8.1 Introduction

This paper details an approach to the data challenge organised for the Extreme Value Analysis (EVA) 2023 Conference. The objective of the challenge was to estimate extremal probabilities, or their associated quantiles, for simulated environmental data sets for various locations on a fictitious planet called Utopia. The data challenge is split into 4 challenges; challenges C1 and C2 focus solely on the univariate setting, whereas challenges C3 and C4 concern multivariate data sets, i.e., we use data from single and multiple locations, respectively.

Challenge C1 requires estimation of the 0.9999-quantile of the distribution of the environmental response variable Y conditional on a covariate vector $\boldsymbol{X}$, for 100 realisations of covariates. To do so, we model the tail of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ using a generalised Pareto distribution (GPD; Pickands, 1975) and employ the extreme value generalised additive modelling (EVGAM) framework, first introduced by Youngman (2019), to account for the non-stationary data structure. We consider a variety of model formulations and

select our final model using cross validation (CV). Furthermore, 50% bootstrap confidence intervals are estimated, and the final model performance is assessed using the number of times the true conditional quantile lies in the confidence intervals (Rohrbeck et al., 2023). For challenge C2, we are required to estimate the T-year return level for the marginal distribution of Y. In other words, we are interested in the value q satisfying $\Pr(Y > q) = 1/(300T)$, such that $T = 200$.

Challenges C3 and C4 concern the estimation of probabilities for extreme multivariate regions. Such estimates require techniques for modelling and extrapolating within the joint tail. For challenge C3, three unknown non-stationary environmental variables are provided, and we are required to estimate two joint tail probabilities. To achieve this, we propose a non-stationary extension of the model introduced by Wadsworth and Tawn (2013). Lastly, for challenge C4, we wish to estimate the probability that 50 environmental variables jointly exceed prespecified extreme thresholds. To simplify the analysis, we separate the variables into five independent groups and obtain separate probabilities for each group using the conditional extremes approach of Heffernan and Tawn (2004).

The remainder of the paper is structured as follows. A suitable background on EVA is introduced in Section 8.2. Section 8.3 covers all aspects of the univariate challenges C1 and C2; we outline our exploratory data analysis (EDA) and detail our methodology based on EVGAM, introducing tools for model selection and comparison. In Section 8.4 we cover the first multivariate challenge C3. After establishing the presence of non-stationarity in the underlying data through EDA, we detail our extension of the Wadsworth and Tawn (2013) model, alongside inferential techniques for this extended framework. Given this non-stationarity, an overview of quantile regression and model fitting is presented. Challenge C4 is covered in Section 8.5. Given the high-dimensional nature of this problem, our data analysis provides the basis for clustering the variables into independent subgroups, and the conditional extremes approach is used to approx-

imate probabilities for each subgroup. The paper ends with a discussion of the results of all challenges. Supporting material can be found in Appendix E.

8.2 EVA background

8.2.1 Univariate modelling

Univariate EVA methods are concerned with capturing the tail of a distribution, allowing extreme quantities to be estimated. The most common univariate approach is the peaks-over-threshold framework. Consider a continuous, independent and identically distributed (IID) random variable Y with distribution function F and upper endpoint $y^F := \sup\{y : F(y) < 1\}$. Pickands (1975) shows that, for some high threshold $v < y^F$, the excesses $(Y - v) \mid Y > v$, after suitable rescaling, converge in distribution to a GPD as $v \to y^F$. In practice, this limit is taken to hold exactly for an appropriately chosen high threshold v such that

$$\Pr(Y > y + v \mid Y > v) = \begin{cases} (1 + \xi y/\sigma)_+^{-1/\xi} & \text{if } \xi \neq 0 \\ \exp(-y/\sigma) & \text{if } \xi = 0, \end{cases}$$

for $y > 0$, $w_+ = \max(w, 0)$, shape parameter $\xi \in \mathbb{R}$ and threshold-dependent scale parameter $\sigma > 0$. Note that the case when $\xi = 0$ is taken in the limit as $\xi \to 0$. We write $(Y - v) \mid Y > v \sim \text{GPD}(\sigma, \xi)$. For $\xi < 0$, the distribution has a finite upper end-point at $v - \sigma/\xi$ but is unbounded above for $\xi \geq 0$. Davison and Smith (1990) provide an overview of the properties of the GPD, and also propose an extension of this framework to the non-stationary setting. Given a non-stationary process Y with associated covariate(s) $\boldsymbol{X}$, they propose the following model

$$\Pr(Y > y + v \mid Y > v, \boldsymbol{X} = \boldsymbol{x}) = \left(1 + \frac{y\xi(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right)_+^{-1/\xi(\boldsymbol{x})},$$

for $y > 0$, where $\sigma(\cdot), \xi(\cdot)$ denote functions of the covariate(s). Recent extensions of the Davison and Smith (1990) framework include allowing the threshold to be covariate-dependent, i.e., $v(\boldsymbol{x})$ (Kyselý et al., 2010; Northrop and Jonathan, 2011), and capturing the covariate functions in a flexible manner using generalised additive models (GAMs; Chavez-Demoulin and Davison, 2005; Youngman, 2019).

8.2.2 Extremal dependence measures

In addition to capturing modelling marginal tail behaviours, multivariate EVA methods are concerned with quantifying the dependence between extremes of multiple observations. An important classification of this dependence, is obtained through the measure χ (Joe, 1997): given a d-dimensional random vector $\boldsymbol{X}$, with $d \geq 2$ and $X_i \sim F$ for all $i \in \{1, \ldots, d\}$, consider the probability

$$\chi(u) := \frac{\Pr(F(X_i) > u, i \in A \subseteq \{1, \ldots, d\})}{1-u}, \quad |\, A \,| \geq 2, \tag{8.2.1}$$

where $|\cdot|$ denotes set cardinality. Where the limit exists, we set $\chi := \lim_{u \to 1} \chi(u) \in [0, 1]$. When $\chi > 0$, we say that the variables exhibit asymptotic dependence, i.e., can take their largest values simultaneously, with the strength of dependence increasing as χ approaches 1. If $\chi = 0$, the variables cannot all take their largest values together. In particular, for $d = 2$, we refer to the case $\chi = 0$ as asymptotic independence.

We also consider the coefficient of tail dependence proposed by Ledford and Tawn (1996). Using the formulation given in Resnick (2002), let

$$\eta(u) := \frac{\log(1-u)}{\log P\left(F(X_i) > u, i \in A \subseteq \{1, \ldots, d\}\right)}.$$

When the limit exists, we set $\eta := \lim_{u \to 1} \eta(u) \in (0, 1]$. The cases $\eta = 1$ and $\eta < 1$, correspond to cases $\chi > 0$ and $\chi = 0$, respectively. For $\eta < 1$, this coefficient quantifies the form of dependence for random vectors that do not take their largest

values simultaneously.

Since χ and η are limiting values, they are unknown in practice and must be approximated using numerical techniques. Therefore, when quantifying extremal dependence, we approximate χ (η) using empirical estimates of $\chi(u)$ ($\eta(u)$) for some high threshold u. See Murphy-Barltrop et al. (2023) for a detailed example.

8.3 Challenges C1 and C2

This section discusses our approach for challenges C1 and C2 in the univariate setting. Both challenges concern 70 years of daily data for the capital city of Amaurot. Each year has 12 months of 25 days and two seasons (season 1 for months 1-6, and season 2 for months 6-12). Suppose Y is an unknown response variable, and $\boldsymbol{X} = (V_1, \ldots, V_8)$ is a vector of covariates, (V_1, V_2, V_3, V_4) denoting unknown environmental variables and (V_5, V_6, V_7, V_8) denoting season, wind direction (radians), wind speed (unknown scale), and atmosphere (recorded monthly), respectively.

For C1, we build a model for $Y \mid \boldsymbol{X}$ and estimate the 0.9999-quantile, with associated 50% confidence intervals, for 100 different covariate combinations denoted $\boldsymbol{x}_i$ for $i \in \{1, \ldots, 100\}$. Note $\boldsymbol{x}_i$ are not covariates observed within the data set, but new observations provided by the challenge organisers.

For C2, we estimate the marginal quantile q such that $\Pr(Y > q) = (6 \times 10)^{-4}$, which corresponds to a once in 200 year event in the IID setting. Therefore, we first estimate the marginal distribution $F_Y(y)$ using Monte-Carlo techniques; see for instance, Eastoe and Tawn (2009). Since we have a large sample size $n = 21,000$ it is reasonable to assume that the observed covariate sample is representative of $\boldsymbol{X}$. We can approximate the marginal distribution as follows,

$$\hat{F}_Y(y) = \int_{\boldsymbol{X}} F_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \approx \frac{1}{n} \sum_{t=1}^{n} F_{Y_t|\boldsymbol{X}_t}(y_t \mid \boldsymbol{x}_t). \tag{8.3.1}$$

To approximate $F_Y(y)$, we first re-estimate the GPD parameters, now using a penalised log-likelihood which incorporates the following loss function, provided by the challenge organisers,

$$\mathcal{L}(q, \hat{q}) = \begin{cases} 0.9(0.99q - \hat{q}) & \text{if } 0.99q > \hat{q} \\ 0 & \text{if } |q - \hat{q}| \leq 0.01q \\ 0.1(\hat{q} - 1.01q) & \text{if } 1.01q < \hat{q}, \end{cases} \tag{8.3.2}$$

where q and $\hat{q}$ are the true and estimated marginal quantiles, respectively. This loss function penalises under-estimation more heavily than an over-estimation.

We conduct the same EDA for both challenges given the same covariates are used; this is outlined in Section 8.3.1. In Section 8.3.2 we introduce our techniques for modelling $Y \mid \boldsymbol{X}$, which is then used for modelling Y via (8.3.1). Our approach for uncertainty quantification is outlined in Section 8.3.3, and we give our results for both challenges in Section 8.3.4.

8.3.1 Exploratory data analysis

This section details our exploratory analysis for challenges C1 and C2. We are informed that the response variable $Y_t,\ t \in \{1, \ldots, n\}$, is independent over time (Rohrbeck et al., 2023), but is affected by the covariate vector $\boldsymbol{X_t} = \{V_{1,t}, \ldots, V_{8,t}\}$. However, it is not clear which covariates affect Y, and what form these covariate-response relationships take. In what follows, we aim to explore these relationships so we can account for them in our modelling framework.

To begin, we explore the dependence between all variables to understand the relationships between covariates, as well as relationships between individual covariates and the response variable. We investigate dependence in the main body of the data using Kendall's τ measure, while for the joint tails, we use the pairwise extremal dependence coefficients χ and η defined in Section 8.2; values for all pairs are shown in Figure 8.3.1,

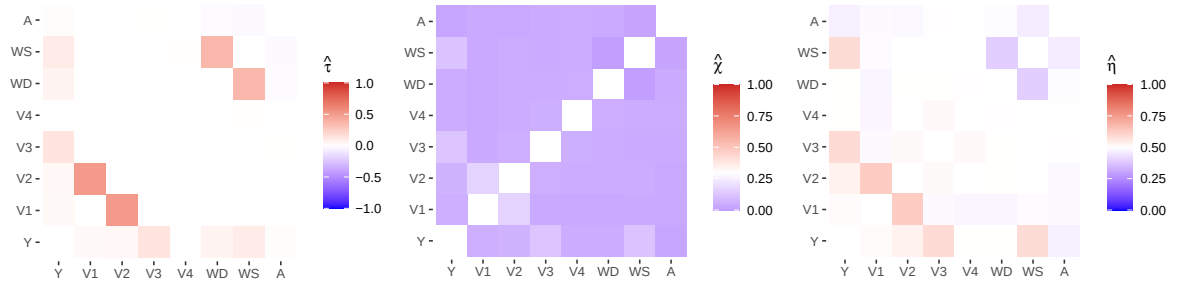


Figure 8.3.1: Heat maps for dependence measures for each pair of variables: Kendall's τ (left), χ (middle) and η (right). Note the scale in each plot varies, depending on the support of the measure, and the diagonals are left blank, where each variable is compared against itself.

with threshold u set at the empirical 0.95-quantile for the extremal measures.

The response variable Y has the strongest dependence with V_3 in the body of the distribution, followed by V_6 (wind speed) and then V_7 (wind direction). Similarly, Y has strong dependence with V_2, V_3 and V_6 in the tail. We also find strong dependence between V_6 and V_7 in the body, but evidence of weak dependence in the tail (dark blue for $\hat{\chi}$ and $\hat{\eta}$). There is also strong dependence between V_1 and V_2 in both the body and tail (see dark red for $\hat{\eta}$). We find very similar dependence relationships when the data are split into seasons. In Appendix E, we show scatter plots of each covariate against the response variable; these demonstrate a highly non-linear relationship for each explanatory variable with Y.

Since V_6 and V_7 have strong relationships with the response variable Y, we explore these variables in more detail. The most notable feature of both variables is that they have a significant shift in distribution. We use the `changepoint` package (Killick and Eckley, 2014) in R to estimate the difference in mean for both variables; just 5 observations separate the estimated changepoints, so we assume the changes occur simultaneously. Before the changepoint, winds typically occur in the southwest direction and are greater in magnitude, while after the changepoint, they occur mainly in

the northeast direction and are lower in magnitude; see Appendix E. Rohrbeck et al. (2023) state this non-stationarity was created unintentionally when designing the data challenge.

Next, we explore temporal relationships for the response variable Y. We first find temporal non-stationarity in that the distribution of Y varies significantly with V_5 (season); see Appendix E for more detail. The magnitude of Y_t is higher for season 1 than season 2, in both the main body and tail of the distribution. However, within each season, across months, there is little temporal variation in the distribution of Y. We also find that Y exhibits temporal independence at all lags, with acf values close to zero (see Appendix E).

We also explore temporal dependence in the covariates. Rohrbeck et al. (2023) states that $V_1, \ldots, V_4$ are temporally independent. We find that for V_6 and V_7, the acf remains significantly different from zero at all time lags (see Appendix E) due to the changepoint discussed earlier. Finally, V_8 has high acf values at the earliest lags that decrease rapidly until lag 25 (i.e., the length of a month) and then continue to decrease at a slower rate until ~ 50 time lags. This is because atmosphere is recorded monthly.

As noted in Rohrbeck et al. (2023), 11.7% of the observations have at least one possible predictor variable missing completely at random (MCAR). A detailed breakdown of the pattern of missing predictor observations is provided in Appendix E. Since we can assume the data are MCAR, ignoring the observations that have a missing predictor variable will not bias our inference. However, a complete case analysis is undesirable due to the amount of data loss. To mitigate against this, we attempt to impute the observations where predictors are missing but ultimately found a case analysis approach works best for our data. This results in only 4% of observations being removed for our final model.

8.3.2 Methods

In this section, we explain our model development procedure for $Y \mid \boldsymbol{X}$. As the challenges concern extreme quantile estimation, we use a non-stationary GPD model. This requires threshold selection and comparison of different covariate-dependent GPD parameterisations. Recall that we utilise the same model formulation for both C1 and C2 via (8.3.1).

When fitting a GPD, the first challenge is to select an appropriate threshold. This selection involves a bias-variance trade-off: too low a threshold is likely to violate the asymptotic basis of the model, incorporating bias in the GPD fit, whereas higher threshold choices lead to additional uncertainty due to fewer exceedances being used to fit the model. A variety of methods exist which aim to balance this trade-off; Scarrott and MacDonald (2012) provide a review and more recent developments include Northrop et al. (2017) and Schneider et al. (2021). Owing to its favourable properties for IID data, we employ the threshold selection method of Murphy et al. (2023) and extend this approach to select a threshold for non-stationary, covariate-dependent GPD models. The method selects a threshold based on minimising the expected quantile discrepancy (EQD) between the sample quantiles and fitted GPD model quantiles. When fitting a non-stationary model, the excesses will not be identically distributed across covariates. Thus, to utilise the EQD method in this case, we use the fitted non-stationary GPD parameter estimates to transform the excesses to common standard exponential margins and compare sample quantiles against theoretical quantiles from the standard exponential distribution.

Owing to the complex covariate structure observed in the data, as described in Section 8.3.1, we employ the flexible EVGAM framework proposed in Youngman (2019) for modelling GPD tail behaviour. Under this framework, GAM formulations are used to capture non-stationarity in the threshold, scale and shape functions introduced in Section 8.2. In general, GAMs provide flexible functional forms that allow us to capture

multiple covariate interactions. Moreover, both discrete and continuous covariates can be incorporated in the GAM framework.

Without loss of generality, consider the scale function $\sigma(\boldsymbol{x})$. We assume that

$$h(\sigma(\boldsymbol{x})) = \psi_\sigma(\boldsymbol{x}), \quad \text{with} \quad \psi_\sigma(\boldsymbol{x}) = \beta_0 + \sum_{\kappa=1}^{K} \sum_{p=1}^{P_\kappa} \beta_{\kappa p} b_{\kappa p}(\boldsymbol{x}), \tag{8.3.3}$$

where $h(x) := \log(x)$ denotes the link function which ensures the correct support, with coefficients $\beta_0, \beta_{\kappa p} \in \mathbb{R}$ and basis functions $b_{\kappa p}$ for $p \in \{1, \ldots, P_\kappa\}, \kappa \in \{1, \ldots, K\}$. Analogous forms are taken for $v(\boldsymbol{x})$ and $\xi(\boldsymbol{x})$, adjusting the link function h as appropriate, with ψ_v and ψ_ξ denoting the respective formulations.

For all GAM formulations, model fitting is carried out using the `evgam` software package (Youngman, 2022), whereby restricted maximum likelihood estimation (REML) is used to approximate the GAM coefficients. In general, REML schemes avoid overfitting through penalisation of the likelihood function. Furthermore, formulation via likelihood functions avoids the use of Markov Chain Monte-Carlo methods, which can be computationally expensive; see Wood (2017) for further details.

For model performance assessment and selection, we apply k-fold CV (Hastie et al., 2001, Ch 7.). We divide the data into k groups (folds), where each fold is removed in turn and the model is fitted to the remaining data. Choosing a higher number k provides less biased CV metrics with a higher variance. A lower k is computationally cheaper, however, it may overestimate the test error rate. We explore model ranking by taking both $k = 10$ and 50, and find that both give an equivalent ranking; we present results for the latter. For each omitted fold, we compute several goodness-of-fit measures: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the continuous ranked probability score (CRPS, Gneiting and Katzfuss, 2014). AIC and BIC aid in guiding the model search, favouring parsimony by penalising model complexity, while CRPS describes the discrepancy between the predicted distribution

function and observed values without the specification of empirical quantiles. Finally, we report the average criterion over all 50 folds for each model; lower values of AIC, BIC and CRPS indicate a better model fit.

Model selection

Our analysis in Section 8.3.1 indicates that V_3, V_5 (season), and V_6 (wind speed) exhibit non-trivial dependence relationships with the response variable; we therefore assume these variables can be used as predictor variables for modelling Y, and set $\tilde{\boldsymbol{X}} := (\boldsymbol{V}_j)_{j\in\{3,5,6\}}$, with $\tilde{X}_{r,t}$ denoting the r^{th} component of $\tilde{\boldsymbol{X}}$, $r \in \{1,2,3\}$. Although V_7 (wind direction) also exhibits predictor power, we have not considered it here since it is highly correlated with wind speed so would involve adding complex interaction terms to the model formulation, and V_6 has a stronger relationship with Y compared to V_7 (see Figure 8.3.1).

There is a clear variation in the distribution of Y between seasons. Due to this distinct difference, we explore the inclusion of a stepped-threshold according to season. In particular, we set $v(\tilde{\boldsymbol{x}}_t) := \mathbb{1}(\tilde{x}_{2,t} = 1)v_1 + \mathbb{1}(\tilde{x}_{2,t} = 2)v_2$, $v_1, v_2 \in \mathbb{R}$, with corresponding rate parameter $\lambda(\tilde{\boldsymbol{x}}_t) := \mathbb{1}(\tilde{x}_{2,t} = 1)\lambda_1 + \mathbb{1}(\tilde{x}_{2,t} = 2)\lambda_2$, where $\lambda_1, \lambda_2 \in [0,1]$ denote the non-exceedance probabilities for seasons 1 and 2, respectively. This seasonal threshold significantly improves model fits; see Appendix E for further details. GAM forms for the threshold were also explored, but did not offer significant improvement. Furthermore, the smooth GAM formulation of the GPD scale parameter adequately captures any residual variation in the response arising due to covariate dependence.

Using the `evgam` package, we fit the non-stationary GPD to the tail of Y, with $\sigma(\boldsymbol{x})$ as in equation (8.3.3). We keep the shape function $\xi(\boldsymbol{x}) := \xi \in \mathbb{R}$ constant across covariates; this is common in non-stationary analyses since this parameter is difficult to estimate (Chavez-Demoulin and Davison, 2005). Even after accounting for seasonal variability in the threshold, including the seasonal variable in the scale

function formulation via an indicator function improves model fits. From Section 8.3.1, we know there is a step change in the mean for V_6; we consider both an indicator function and a spline to capture the relationship between Y and this predictor. When using splines, we are required to select a basis dimension $B \in \mathbb{N}$; this determines the number of coefficients to be estimated. Basis dimension is the most important choice within spline modelling procedures and directly corresponds with the flexibility of the framework (Wood, 2017). In practice, it is better to select a higher dimension than we would expect to be necessary since the REML scheme will adjust estimates of coefficients to avoid over-fitting. Furthermore, as long as the basis dimension is sufficiently high, the locations of knots have little impact on the resulting model fits. Note that we use thin-plate regression splines as they can smooth with respect to any number of covariates, and it is the default in the `evgam` package (Youngman, 2022). Other types of splines could be used, but we do not consider them here.

To determine B for V_6, we build a model for $Y \mid \tilde{X}_3$, allowing us to consider the effect of this predictor on the response directly. We vary the basis dimension and compare the resulting models using CV. Through this, we set $B = 3$ since this appears to offer sufficient flexibility to capture the observed dependence. For V_3, we employ a similar procedure and set $B = 4$.

To determine the best-fitting model, we use a forward selection process based on minimising the CV score using CRPS. Specifically, by comparing model fits with only a single predictor variable (models 2-4 in Table 8.3.1), we find model 4 minimises the CV score and so we add V_3 to the model. Continuing in this fashion results in our final model being model 7. Wind speed (V_6) was also modelled using an indicator function to try and capture its changepoint although this did not improve the model fit and has been omitted in Table 8.3.1. We have omitted models that included additional predictor variables and interaction terms as they do not reduce the CV score further.

Let y_t and $\tilde{\boldsymbol{x}}_t$ denote the observations of the response variable and predictive co-

Table 8.3.1: Table of selected models considered for challenge C1. $\mathbb{1}(\cdot)$ denotes an indicator function, $s_i(\cdot)$ for $i \in \{1, 2\}$ denote thin-plate regression splines, β_0, β_1 are coefficients to be estimated, and $\tilde{x}_{r,t}$ is defined as in the text. All values have been given to one decimal place.

Model	$\sigma(\tilde{\boldsymbol{x}}_t)$	CRPS	AIC	BIC
1	β_0	11.5	92,059.9	92,074.5
2	$\beta_0 + \beta_1 \mathbb{1}(\tilde{x}_{2,t} = 1)$	11.0	92,026.5	92,048.4
3	$\beta_0 + s_1(\tilde{x}_{1,t})$	10.6	91,651.4	91,695.3
4	$\beta_0 + s_2(\tilde{x}_{3,t})$	11.0	91,775.6	91,797.7
5	$\beta_0 + \beta_1 \mathbb{1}(\tilde{x}_{2,t} = 1) + s_1(\tilde{x}_{1,t})$	10.6	91,634.1	91,686.4
6	$\beta_0 + s_1(\tilde{x}_{1,t}) + s_2(\tilde{x}_{3,t})$	10.5	91,307.2	91,357.3
7	$\beta_0 + \beta_1 \mathbb{1}(\tilde{x}_{2,t} = 1) + s_1(\tilde{x}_{1,t}) + s_2(\tilde{x}_{3,t})$	**10.4**	**91,279.9**	**91,339.2**

variates, respectively. Then our model has the following form,

$$F_{Y_t|\tilde{\boldsymbol{X}}_t}(y_t|\tilde{\boldsymbol{X}}_t = \tilde{\boldsymbol{x}}_t, y_t > v(\tilde{\boldsymbol{x}}_t)) = 1 - \lambda(\tilde{\boldsymbol{x}}_t)\left[1 + \xi\left(\frac{y_t - v(\tilde{\boldsymbol{x}}_t)}{\sigma(\tilde{\boldsymbol{x}}_t)}\right)\right]_+^{-1/\xi},$$

for all $t \in \{1, \ldots, n\}$. The formulation of $\sigma(\cdot)$ is defined in Table 8.3.1 as model 7, where $\beta_0, \beta_1 \in \mathbb{R}$ denote coefficients, and s_1, s_2 are thin-plate regression splines for V_3 and V_6, respectively, with corresponding basis coefficients $\boldsymbol{\beta}_2 \in \mathbb{R}^3$ and $\boldsymbol{\beta}_3 \in \mathbb{R}^4$. For challenge C2, we are required to incorporate the loss function of equation (8.3.2) into the modelling framework. Letting $\mathcal{I}_v := \{t \in \{1, \ldots, n\} \mid y_t > v(\tilde{\boldsymbol{x}}_t)\}$ and $n_v := |\mathcal{I}_v|$, we consider the following objective function

$$S(\boldsymbol{\theta}) := -l_R(\boldsymbol{\theta}) + \sum_{i \in \mathcal{I}_v} \mathcal{L}(q_i^*, \hat{q}_i)/n_v,$$

where $l_R(\boldsymbol{\theta})$ denotes the penalised log-likelihood function of the REML approach (Wood, 2017), $\boldsymbol{\theta} := (\beta_0, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \xi)$ denotes the parameter vector associated with the GPD formulation, and $\sum_{i \in \mathcal{I}_v} \mathcal{L}(q_i^*, \hat{q}_i)/n_v$ denotes the average loss between observed and model quantiles. Specifically, if we denote the mapping between $\mathcal{I}_v$ and the order statistics

of $(y_t - v(\tilde{\boldsymbol{x}}_t))_{t\in\mathcal{I}_v}$ by π, then q_i^* is the $\pi(i)^{\text{th}}$ order statistic of $(y_t - v(\tilde{\boldsymbol{x}}_t))_{t\in\mathcal{I}_v}$ and $\hat{q}_i = \sigma(\boldsymbol{x}_{\pi(i)})[\{1 - \pi(i)/(n_v + 1)\}^{-\xi} - 1]/\xi$. Minimising $S(\boldsymbol{\theta})$ ensures the parameter estimates also account for the loss function. We use this formulation to adjust the GPD parameters for challenge C2 once a threshold is selected.

8.3.3 Uncertainty quantification

We are required to construct central 50% confidence intervals for 100 different covariate combinations provided by the challenge organisers, $\boldsymbol{x}_i$ for $i \in \{1, \ldots, 100\}$. We approximate these intervals using the stationary block bootstrapping procedure adopted by D'Arcy et al. (2023b) that preserves temporal dependence and covariate information; we outline this below.

First, the response variable Y_t is transformed to Uniform(0,1) margins; denote this sequence $U_t^Y = F_{Y_t|\tilde{\boldsymbol{X}}_t}(Y_t|\tilde{\boldsymbol{X}}_t = \tilde{x}_t)$. We then adopt the stationary bootstrap procedure of Politis and Romano (1994) to retain the temporal dependence in the response and explanatory variables. The block length L is simulated from a Geometric$(1/l)$ distribution, where the mean block length $l \in \mathbb{N}$ is carefully selected based on the autocorrelation function. We choose 50 days; the maximum lag for which the autocorrelation was significant across all variables; see Appendix E. Denote this bootstrapped sequence on uniform margins by U_t^B. We transform U_t^B back to the original scale using our fitted model, preserving the original structure of Y_t; we denote this series Y_t^B. Then we fit our model to Y_t^B to re-estimate all of the parameters and thus the quantile of interest. We repeat this procedure to obtain 200 bootstrap samples.

8.3.4 Results

For C1, we use our final model of Section 8.3.2 to estimate the 0.9999-quantile of $Y \mid \tilde{\boldsymbol{X}} = \tilde{\boldsymbol{x}}_i$, $i \in \{1, \ldots, 100\}$, for the set of 100 covariate combinations. The left panel of Figure 8.3.2 shows the quantile-quantile (QQ) plot for our model. There is general

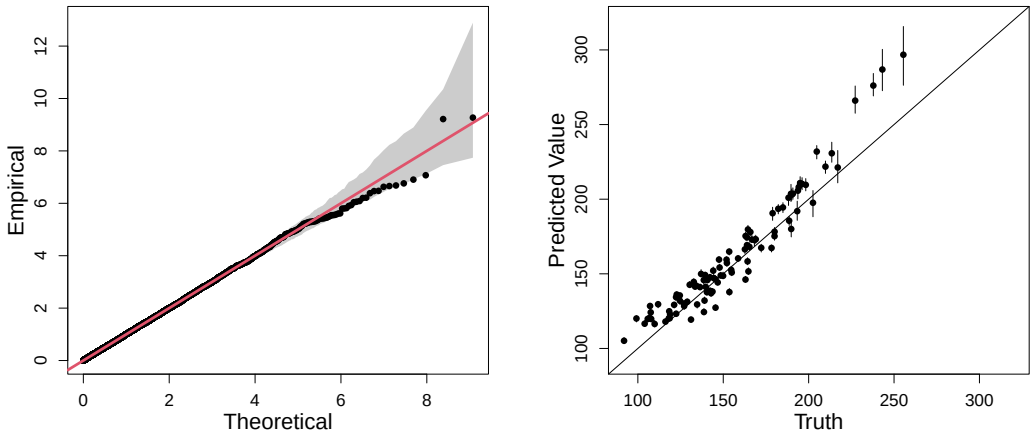


Figure 8.3.2: QQ plot for our final model, model 7 in Table 8.3.1, on exponential margins. The $y = x$ line is given in red and the grey region represents the 95% tolerance bounds (left). Predicted 0.9999−quantiles against true quantiles for the 100 covariate combinations. The points are the median predicted quantile over 200 bootstrapped samples and the vertical error bars are the corresponding 50% confidence intervals. The $y = x$ line is also shown (right).

alignment between the model and empirical quantiles; however, there is some under-estimation in the upper tail, and our 95% tolerance bounds do not contain some of the most extreme response values. The right panel of Figure 8.3.2 shows our predicted quantiles, and their association confidence intervals, compared to their true quantiles. This figure is different from the one presented by Rohrbeck et al. (2023) due to an error in our code being fixed after submission. In this scenario, our estimated confidence intervals lead to a 14% coverage of the true quantiles, which does not alter our ranking for this challenge. Our performance and model improvements are discussed in Section 8.6.

For challenge C2, we estimate the quantile of interest as $\hat{q} = 212.5913$ (208.3783, 246.0764). Due to a coding error, this value differs from the original estimate submitted for the data challenge. The updated estimate over-estimates compared to the truth. A 95% confidence interval for the estimate is given in parentheses based on the

bootstrapping procedure outlined in Section 8.3.3.

8.4 Challenge C3

8.4.1 Exploratory data analysis

For challenge C3, we are provided with 70 years of daily time series of an environmental variable for three towns on the island of Coputopia. These series are denoted by $Y_{i,t}$, $i \in \{1, 2, 3\}$, $t \in \{1, \ldots, n\}$, where i is the index of each town and t is the point in time. Each year consists of 12 months, each lasting 25 days, resulting in $n = 21,000$ observations for each location.

Alongside the time series, we are also provided with daily covariate observations $\boldsymbol{X}_t = (S_t, A_t)$, where S_t and A_t denote seasonal and atmospheric conditions, respectively. Season is a binary variable, taking values in the set $\{1, 2\}$, with each year of observations exhibiting both seasons for exactly 150 consecutive days. On the other hand, the atmospheric conditions are piecewise constant over months, with large variations in the observed values between months. A descriptive figure of both covariates is given in Appendix E.

In Rohrbeck et al. (2023), we are informed that $Y_{i,t}$ are distributed identically across all sites and over time, with a standard Gumbel distribution function. However, it is not known whether the covariates $\boldsymbol{X}_t$ influence the dependence structure of $\boldsymbol{Y}_t := (Y_{1,t}, Y_{2,t}, Y_{3,t})$. We are also informed that, conditioned on covariates, the process is independent over time, i.e., $(\boldsymbol{Y}_t \mid \boldsymbol{X}_t) \perp\!\!\!\perp (\boldsymbol{Y}_{t'} \mid \boldsymbol{X}_{t'})$ for any $t \neq t'$. In this section, we examine what influence, if any, the covariate process $\boldsymbol{X}_t$ may have on the dependence structure of $\boldsymbol{Y}_t$. If such relationships exist, they need to be accounted for when estimating joint tail probabilities.

We begin by transforming the time series to standard exponential margins using the probability integral transform, i.e., set $Z_{i,t} := -\log(1 - F(Y_{i,t}))$, $i \in \{1, 2, 3\}$,

$t \in \{1, \ldots, n\}$, where $F(y) = \exp(-\exp(-y))$, $y \in \mathbb{R}$. This transformation is common in the study of multivariate extremes and can simplify the description of extremal dependence (Keef et al., 2013). To explore the extremal dependence in the Coputopia time series, we consider all 2- and 3-dimensional subvectors of the process, i.e., $\{Z_{i,t}, i \in I, t \in \{1, \ldots, n\}\}$, $I \in \mathcal{I} := \{\{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$. This separation is important to ensure the overall dependence structure is fully understood, since intermediate scenarios can exist where a random vector exhibits $\chi = 0$, but $\chi > 0$ for some 2-dimensional subvector(s) (Simpson et al., 2020).

Furthermore, to explore the impact of covariates on the dependence structure, we partition the time series into subsets using the covariates. For the seasonal covariate, let $G^S_{I,j} := \{Z_{i,t}, i \in I, S_t = j\}$ for $j = 1, 2$. For the atmospheric covariate, we let $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ denote the permutation associated with the order statistics of A_t, defined so that ties in the data are accounted for. We then split the data into 10 equally sized subsets corresponding to the atmospheric order statistics, i.e., $G^A_{I,k} := \left\{Z_{i,t}, i \in I, t \in \Sigma^k\right\}$ for $k = 1, 2, \ldots, 10$, where $\Sigma^k := \{t \mid (k-1)n/10 + 1 \leq \pi(t) \leq kn/10\}$. We can observe that the atmospheric values associated with each subset $G^A_{I,k}$ will increase over k.

The idea behind these subsets is to examine whether altering the values of either covariate impacts the extremal dependence structure. Consequently, we set $u = 0.9$ and estimate $\chi(u)$ using the techniques outlined in Section 8.2, with uncertainty quantified through bootstrapping with 200 samples. The bootstrapped χ estimates for $G^A_{I,k}$ with $I = \{1, 2, 3\}$ are given in Figure 8.4.1. The plots for the remaining index sets in $\mathcal{I}$, along with the subsets associated with the seasonal covariate, are given in Appendix E. The estimates of χ appear to vary, in the majority of cases, across both subset types (seasonal and atmospheric), suggesting both covariates have an impact on the dependence structure. For the atmospheric process in particular, the values of χ tend to decrease for higher atmospheric values, suggesting a negative association between positive extremal

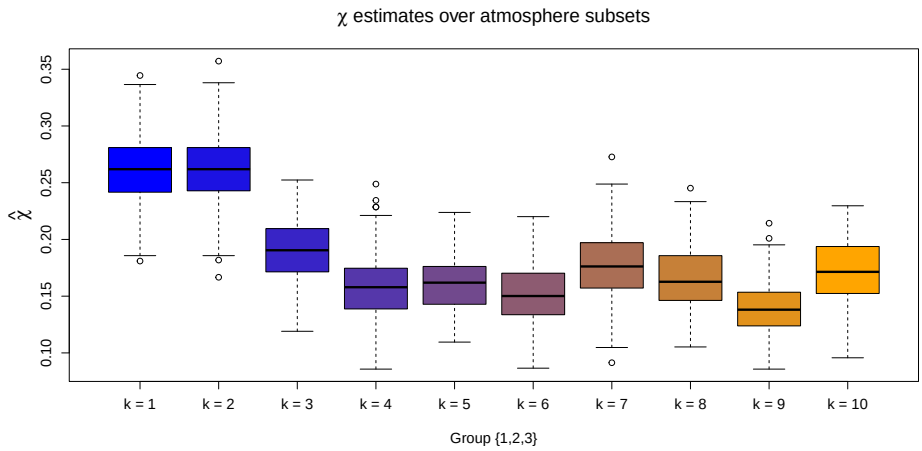


Figure 8.4.1: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

dependence and atmosphere. We also observe that across all subsets, χ appears consistently low in magnitude, suggesting the extremes of some, if not all, of the sub-vectors are unlikely to occur simultaneously. As such, for modelling the Coputopia time series, we require a framework that can capture such forms of dependence.

We also consider pointwise estimates of the function λ, as defined later in equation (8.4.2), over $G^S_{I,j}$ and $G^A_{I,k}$ for fixed simplex points; these results are given in Appendix E. Similar to χ, estimates of λ vary significantly across subsets, providing additional evidence of non-stationarity within the extremal dependence structure.

8.4.2 Modelling of joint tail probabilities under asymptotic independence

For challenge C3, we are required to estimate probabilities

$$p_1 := \Pr\left(Y_1 > y, Y_2 > y, Y_3 > y\right) \quad \text{and} \quad p_2 := \Pr\left(Y_1 > v, Y_2 > v, Y_3 < m\right),$$

with $y = 6$, $v = 7$ and $m = -\log(\log(2))$. We refer to p_1 and p_2 as parts 1 and 2 of the challenge, respectively. Note that p_1 and p_2 are independent of the covariate

process and correspond to different extremal regions in $\mathbb{R}^3$. For the remainder of this section, we will consider the transformed exponential variables (Z_1, Z_2, Z_3), omitting the subscript t for ease of notation. Observe that $F_{(-Z_3)}(z) = e^z$, for $z < 0$; setting $\tilde{Z}_3 := -\log\left(1 - \exp\{-Z_3\}\right)$, we have

$$p_2 = \Pr\left(Z_1 > \tilde{v}, Z_2 > \tilde{v}, Z_3 < \tilde{m}\right) = \Pr\left(Z_1 > \tilde{v}, Z_2 > \tilde{v}, \tilde{Z}_3 > \tilde{m}\right),$$

where $\tilde{v}$ and $\tilde{m}$ denote the values v and m transformed to the standard exponential scale, e.g., $\tilde{v} := -\log\left(1 - \exp\{-\exp\{-v\}\}\right)$. Similarly, we have $p_1 = \Pr\left(Z_1 > \tilde{y}, Z_2 > \tilde{y}, Z_3 > \tilde{y}\right)$. Consequently, both p_1 and p_2 can be considered as joint survivor probabilities.

In many applications of multivariate extremes, joint tail probabilities are estimated using models that implicitly assume random vectors are regularly varying (Tawn, 1988a; Resnick, 2002). However, this framework is unable to accurately extrapolate into the joint tail for data not exhibiting asymptotic dependence (Ledford and Tawn, 1996; Heffernan and Tawn, 2004). This has motivated the development of modelling approaches for data structures where not all extremes are observed simultaneously, such as the Coputopia data set.

The first such approach was proposed by Ledford and Tawn (1996); given a slowly varying function $\mathcal{L}$, i.e., $\lim_{u\to 1} \mathcal{L}(tu)/\mathcal{L}(u) = 1$ for all $t > 0$, they assume the joint tail is given by

$$\Pr\left(Z_1 > u, Z_2 > u, Z_3 > u\right) = \Pr\left(\min\{Z_1, Z_2, Z_3\}\right) = \mathcal{L}\left(e^u\right) e^{-u/\eta}, \tag{8.4.1}$$

as $u \to \infty$, where $\eta \in (0, 1]$ is the coefficient of tail dependence defined in Section 8.2. If $\eta = 1$ and $\lim_{u\to\infty} \mathcal{L}(u) \neq 0$, we have asymptotic dependence, while $\eta < 1$, or $\eta = 1$ and $\lim_{u\to\infty} \mathcal{L}(u) = 0$, implies the variables cannot be extreme simultaneously.

In practice, the approach of Ledford and Tawn (1996) is only applicable within regions where all variables are large. To overcome this limitation, Wadsworth and

Tawn (2013) proposed a general extension of equation (8.4.1). For any ray $\boldsymbol{\omega} \in \boldsymbol{S}^2 := \{(w_1, w_2, w_3) \in [0,1]^3 : w_1 + w_2 + w_3 = 1\}$, where $\boldsymbol{S}^2$ denotes the standard 2-dimensional simplex, the authors assume

$$\begin{aligned} \Pr\left(Z_1/w_1 > r, Z_2/w_2 > r, Z_3/w_3 > r\right) &= \Pr\left(\min\{Z_1/w_1, Z_2/w_2, Z_3/w_3\} > r\right) \\ &= \mathcal{L}(e^r; \boldsymbol{\omega}) e^{-r\lambda(\boldsymbol{\omega})}, \end{aligned} \tag{8.4.2}$$

as $r \to \infty$, where $\lambda(\boldsymbol{\omega}) \geq \max(\boldsymbol{\omega})$ is known as the angular dependence function. Asymptotic dependence occurs at the lower bound, i.e., $\lambda(\boldsymbol{\omega}) = \max(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \boldsymbol{S}^2$, and model (8.4.2) reduces to model (8.4.1) for $\boldsymbol{\omega} = (1/3, 1/3, 1/3)$, with $\eta = 1/\{3\lambda(\boldsymbol{\omega})\}$. In practice, equation (8.4.2) can be used to evaluate extreme joint survivor probabilities; in particular, probabilities p_1 and p_2 can be identified with the rays $\boldsymbol{\omega}^{(1)} := (\tilde{u}, \tilde{u}, \tilde{u})/r^{(1)}$ and $\boldsymbol{\omega}^{(2)} := (\tilde{v}, \tilde{v}, \tilde{m})/r^{(2)}$ in $\boldsymbol{S}^2$, respectively, where $r^{(1)} := \tilde{u}+\tilde{u}+\tilde{u}$ and $r^{(2)} := \tilde{v}+\tilde{v}+\tilde{m}$. See Section 8.4.4 for further details.

As a final remark, we note that there exist several additional methods capable of approximating extreme joint survivor probabilities for non-asymptotically dependent data structures (e.g., Heffernan and Tawn, 2004; Wadsworth et al., 2017). However, in our approach, we prefer to stick with the model of Wadsworth and Tawn (2013) since this framework can be easily adapted to capture non-stationary dependence.

8.4.3 Accounting for non-stationary dependence

In the stationary setting, pointwise estimates of λ can be obtained via the Hill estimator (Hill, 1975), from which tail probabilities can be approximated. However, alternative procedures are required for data exhibiting trends in dependence, such as the Coputopia data set. Existing approaches for capturing non-stationary dependence structures are sparse in the extremes literature, and most approaches are limited to asymptotically dependent data structures. For the case when data are not asymp-

totically dependent, Mhalla et al. (2019) and Murphy-Barltrop and Wadsworth (2022) propose non-stationary extensions of the Wadsworth and Tawn (2013) framework, while Jonathan et al. (2014) and Guerrero et al. (2023) propose non-stationary extensions of the Heffernan and Tawn (2004) model. See Murphy-Barltrop and Wadsworth (2022) for a detailed review.

To account for non-stationary dependence in C3, we propose an extension of the Wadsworth and Tawn (2013) framework. With $\boldsymbol{Z}_t = (Z_{1,t}, Z_{2,t}, Z_{3,t})$ and $\boldsymbol{X}_t$, defined as in Section 8.4.1, we define the structure variable $T_{\boldsymbol{\omega},t} := \min\{Z_{1,t}/w_1, Z_{2,t}/w_2, Z_{3,t}/w_3\}$, for any $\boldsymbol{\omega} \in \boldsymbol{S}^2$; we refer to $T_{\boldsymbol{\omega},t}$ as the min-projection at time t. From Section 8.4.1, we know that the joint distribution of $\boldsymbol{Z}_t$ is not identically distributed over t; this implies non-stationarity in the distribution of $T_{\boldsymbol{\omega},t}$.

To account for non-stationarity in $T_{\boldsymbol{\omega},t}$, Mhalla et al. (2019) and Murphy-Barltrop and Wadsworth (2022) assume the following model

$$\Pr\left(T_{\boldsymbol{\omega},t} > u \mid \boldsymbol{X}_t = \boldsymbol{x}_t\right) = \mathcal{L}\left(e^u \mid \boldsymbol{\omega}, \boldsymbol{x}_t\right) e^{-\lambda(\boldsymbol{\omega}|\boldsymbol{x}_t)u} \text{ as } u \to \infty, \tag{8.4.3}$$

for all t. Note that this assumption is very similar in form to equation (8.4.2), with the primary difference being the function λ is non-stationary over t. From equation (8.4.3), it is straightforward to see that

$$\Pr\left(T_{\boldsymbol{\omega},t} - u > z \mid T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\right) = e^{-\lambda(\boldsymbol{\omega}|\boldsymbol{x}_t)z} \text{ as } u \to \infty, \tag{8.4.4}$$

for $z > 0$. Consequently, equation (8.4.3) is equivalent to assuming $(T_{\boldsymbol{\omega},t} - u) \mid \{T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\} \sim \text{Exp}(\lambda\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right))$ as $u \to \infty$.

In unreported exploratory analysis, we found that equation (8.4.3) was not flexible enough to capture the tail of $T_{\boldsymbol{\omega},t}$ for the Coputopia data; see Section 8.4.3 for further discussion. Thus, we propose the following model: given any $z > 0$ and a fixed $\boldsymbol{\omega} \in \boldsymbol{S}^2$,

we assume

$$\Pr\left(T_{\boldsymbol{\omega},t} - u > z \mid T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\right) = \left(1 + \frac{\xi\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right) z}{\sigma\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right)}\right)^{-1/\xi(\boldsymbol{\omega}|\boldsymbol{x}_t)} \quad \text{as } u \to \infty. \tag{8.4.5}$$

This is equivalent to assuming $(T_{\boldsymbol{\omega},t} - u) \mid \{T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\} \sim \text{GPD}(\sigma\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right), \xi\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right))$ as $u \to \infty$, and equation (8.4.4) is recovered by taking the limit as $\xi\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right) \to 0$ for all t.

Our proposed formulation in equation (8.4.5) allows for additional flexibility within the modelling framework via the inclusion of a GPD shape parameter $\xi\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right)$, which quantifies the tail behaviour of $T_{\boldsymbol{\omega},t}$. Given the wide range of distributions in the domain of attraction of a GPD (Pickands, 1975), it is reasonable to assume that the tail of $T_{\boldsymbol{\omega},t}$ can be approximated by equation (8.4.5). For the Coputopia time series, this assumption appears valid, as demonstrated by the diagnostics in Section 8.4.3.

Model fitting

To apply equation (8.4.5), we first fix $\boldsymbol{\omega} \in \boldsymbol{S}^2$ and assume that the formulation holds approximately for some sufficiently high threshold level from the distribution of $T_{\boldsymbol{\omega},t}$; we denote the corresponding quantile level by $\tau \in (0, 1)$. For simplicity, the same quantile level is considered across all t. Further, let $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)$ denote the corresponding threshold function, i.e., $\Pr(T_{\boldsymbol{\omega},t} \leq v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t) \mid \boldsymbol{X}_t = \boldsymbol{x}_t) = \tau$ for all t. Under our assumption, we have $(T_{\boldsymbol{\omega},t} - v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)) \mid \{T_{\boldsymbol{\omega},t} > v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t), \boldsymbol{X}_t = \boldsymbol{x}_t\} \sim \text{GPD}(\sigma\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right), \xi\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right))$. We emphasise that $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)$ is not constant in t, and we would generally expect $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t) \neq v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_{t'})$ for $t \neq t'$.

As detailed in Section 8.4.2, both p_1 and p_2 can be associated with points on the simplex $\boldsymbol{S}^2$, denoted by $\boldsymbol{\omega}^{(1)}$ and $\boldsymbol{\omega}^{(2)}$, respectively. Letting $\boldsymbol{\omega} \in \{\boldsymbol{\omega}^{(1)}, \boldsymbol{\omega}^{(2)}\}$, our estimation procedure consists of two stages: estimation of threshold function $v_\tau(\boldsymbol{\omega}, \boldsymbol{z}_t)$ for a fixed $\tau \in (0, 1)$, followed by estimation of GPD parameter functions $\sigma\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right), \xi\left(\boldsymbol{\omega} \mid \boldsymbol{x}_t\right)$.

For both steps, we take a similar approach to Section 8.3.2 and use GAMs to capture these covariate relationships. To simplify our approach, we falsely assume that the atmospheric covariate A_t is continuous over t; this step allows us to utilise GAM formulations containing smooth basis functions. Given the significant variability in A_t between months, discrete formulations for this covariate would significantly increase the number of model parameters and result in higher variability.

Let $\log(v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)) = \psi_v(\boldsymbol{x}_t)$, $\log(\sigma(\boldsymbol{\omega} \mid \boldsymbol{x}_t)) = \psi_\sigma(\boldsymbol{x}_t)$ and $\xi(\boldsymbol{\omega} \mid \boldsymbol{x}_t) = \psi_\xi(\boldsymbol{x}_t)$ denote the GAM formulations of each function, where ψ_- denotes the basis representation of equation (8.3.3). Exact forms of basis functions are specified in Section 8.4.3. As in Section 8.3.2, model fitting is carried out using the `evgam` software package (Youngman, 2022), with REML procedures used to avoid over-fitting. For the first stage, $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)$ is estimated by exploiting a link between the loss function typically used for quantile regression and the asymmetric Laplace distribution (Yu and Moyeed, 2001). The spline coefficients associated with ψ_σ and ψ_ξ are estimated subsequently using the obtained threshold exceedances.

Selection of GAM formulations and diagnostics

Prior to estimation of the threshold and parameter functions, we specify a quantile level τ and formulations for each of the GAMs. To begin, we fix $\tau = 0.9$ and restrict attention to the latter problem; this in turn involves selecting basis functions and basis dimensions. A variety of formulations were considered for each GAM. By comparing metrics for model selection, namely AIC, BIC and CRPS, we found the following formulations to be sufficient

$$\psi_v(\boldsymbol{x}_t) = \beta_u + s_v(a_t) + \beta_s \mathbb{1}(s_t = 2), \quad \psi_\sigma(\boldsymbol{x}_t) = \beta_\sigma + s_\sigma(a_t) \quad \text{and} \quad \psi_\xi(\boldsymbol{x}_t) = \beta_\xi, \tag{8.4.6}$$

for parts 1 and 2, where $\beta_u, \beta_\sigma, \beta_\xi \in \mathbb{R}$ denote constant intercept terms, $\mathbb{1}$ denotes the indicator function with corresponding coefficient $\beta_s \in \mathbb{R}$, and s_u, s_σ denote cubic

regression splines of dimension $B = 10$. The shape parameter is set to constant for the reasons outlined in Section 8.3.2. Cubic basis functions are used for ψ_v and ψ_σ since they have several desirable properties, including continuity and smoothness (Wood, 2017). Setting $B = 10$ appears more than sufficient to capture the trends relating to the atmosphere variable. Alternative smooth splines and basis dimensions were tested for both parts 1 and 2, but this made little difference to the resulting model fits. Moreover, the fact the same model selection appeared suitable for both parts of C3 provides evidence of robustness for these GAM formulations.

We remark that the seasonal covariate is only present with the formulation for ψ_v. Once accounted for in the non-stationary threshold, the seasonal covariate appeared to have little influence on the fitted GPD parameters. More complex GAM formulations were tested involving interaction terms between the seasonal and atmospheric covariates. However, such formulations offered negligible improvements in model fits, and thus we prefer the simpler formulations on the basis of parsimony.

With GAM formulations selected, we now consider the quantile level $\tau \in (0, 1)$; this is analogous to the bias-variance trade-off discussed in Section 8.3.2. To assess sensitivity in our formulation, we set $\mathrm{T} := \{0.8, 0.81, \ldots, 0.99\}$ and fit the GAMs outlined in equation (8.4.6) for each $\tau \in \mathrm{T}$. Letting $\delta_{\boldsymbol{\omega},t}$ and $\mathcal{T}_\tau := \{t \in \{1, \ldots, n\} \mid \delta_{\boldsymbol{\omega},t} > v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)\}$ denote the min-projection observations and indices of threshold-exceeding observations, respectively, we expect the set

$$\mathcal{E} := \{-\log\{1 - F_{GPD}(\delta_{\boldsymbol{\omega},t} - v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t) \mid \sigma(\boldsymbol{\omega} \mid \boldsymbol{x}_t), \xi(\boldsymbol{\omega} \mid \boldsymbol{x}_t)\} \mid t \in \mathcal{T}_\tau\},$$

to be distributed according to a unit exponential distribution.

With all exceedances transformed to a unified scale, we compare the empirical and model exponential quantiles using QQ plots, through which we assess the relative performance of each $\tau \in \mathrm{T}$. We then selected τ values for which the empirical and theoretical quantiles appeared most similar in magnitude. From this analysis, we set $\tau = 0.83$ and

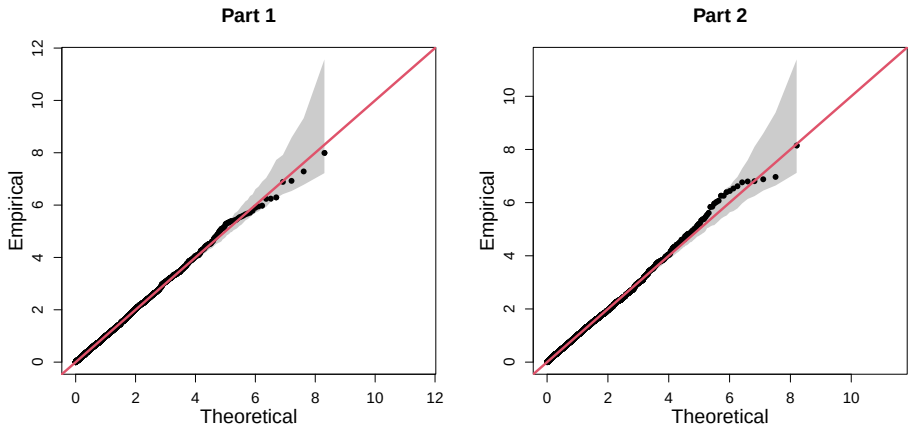


Figure 8.4.2: Final QQ plots for parts 1 (left) and 2 (right) of C3, with the $y = x$ line given in red. In both cases, the grey regions represent the 95% bootstrapped tolerance bounds.

$\tau = 0.85$ for parts 1 and 2, respectively. The corresponding QQ plots are given in Figure 8.4.2, where we observe reasonable agreement between the empirical and theoretical quantiles. However, whilst these values appeared optimal within T, we stress that adequate model fits were also obtained for other quantile levels, suggesting our modelling procedure is not especially sensitive to the exact choice of quantile. Furthermore, we also tested a range of quantile levels below the 0.8-level, but were unable to improve the quality of model fits.

Plots illustrating the estimated GPD scale parameter functions are given in Appendix E, with the resulting dependence trends in agreement with the observed trends from Section 8.4.1. We also remark that the estimated GPD shape parameters obtained for parts 1 and 2 were $0.042\,(0.01, 0.075)$ and $0.094\,(0.059, 0.128)$, respectively, where the brackets denote 95% confidence intervals obtained using posterior sampling; see Wood (2017). These estimates, which indicate slightly heavy-tailed behaviour within the min-projection variable, provide insight into why the original exponential modelling framework is not appropriate for C3.

Overall, these results suggest different extremal dependence trends exist for the two simplex points $\boldsymbol{\omega}^{(1)}$ and $\boldsymbol{\omega}^{(2)}$, illustrating the importance of the flexibility in our model.

These findings are also in agreement with empirical trends observed in Section 8.4.1, suggesting our modelling framework is successfully capturing the underlying extremal dependence structures.

8.4.4 Results

Given estimates of threshold and parameter functions, probability estimates can be obtained via Monte Carlo techniques. Taking p_1, for instance, we have

$$\begin{aligned}
p_1 &= \Pr(Z_1 > \tilde{y}, Z_2 > \tilde{y}, Z_3 > \tilde{y}) \\
&= \Pr\left(\min\left(Z_1/w_1^{(1)}, Z_2/w_2^{(1)}, Z_3/w_3^{(1)}\right) > r^{(1)}\right) \\
&= \int_{\boldsymbol{X}_t} \Pr(T_{\boldsymbol{\omega}^{(1)},\, t} > r^{(1)} \mid \boldsymbol{X}_t = \boldsymbol{x}_t) f_{\boldsymbol{X}_t}(\boldsymbol{x}_t) \mathrm{d}\boldsymbol{x}_t \\
&= (1-\tau) \int_{\boldsymbol{X}_t} \Pr(T_{\boldsymbol{\omega}^{(1)},\, t} > r^{(1)} \mid T_{\boldsymbol{\omega}^{(1)},\, t} > v_\tau(\boldsymbol{\omega}^{(1)}, \boldsymbol{x}_t), \boldsymbol{X}_t = \boldsymbol{x}_t) f_{\boldsymbol{X}_t}(\boldsymbol{x}_t) \mathrm{d}\boldsymbol{x}_t \\
&\approx \frac{1-\tau}{n} \sum_{t=1}^{n} \left(1 + \frac{\xi(\boldsymbol{\omega}^{(1)} \mid \boldsymbol{x}_t)\left(r^{(1)} - v_\tau(\boldsymbol{\omega}^{(1)}, \boldsymbol{x}_t)\right)}{\sigma\left(\boldsymbol{\omega}^{(1)} \mid \boldsymbol{x}_t\right)}\right)^{-1/\xi\left(\boldsymbol{\omega}^{(1)} \mid \boldsymbol{x}_t\right)},
\end{aligned}$$

assuming $\{\boldsymbol{x}_t : t \in \{1, \ldots, n\}\}$ is a representative sample from $\boldsymbol{X}_t$. The procedure for p_2 is analogous. We note that this estimation procedure is only valid when $r^{(1)} > v_\tau(\boldsymbol{\omega}^{(1)}, \boldsymbol{x}_t)$, or $r^{(2)} > v_\tau(\boldsymbol{\omega}^{(2)}, \boldsymbol{x}_t)$, for all t: however, for each $\tau \in \mathrm{T}$, this inequality is always satisfied, owing to the very extreme nature of the probabilities in question. Through this approximation, we obtain $\hat{p}_1 = 1.480449 \times 10^{-5}$ and $\hat{p}_2 = 2.460666 \times 10^{-5}$.

8.5 Challenge C4

8.5.1 Exploratory data analysis

Challenge C4 entails estimating survival probabilities across 50 locations on the island of Utopula. As stated in Rohrbeck et al. (2023), the Utopula island is split into two administrative areas, for which the respective regional governments 1 and 2 have collected

data concerning the variables $Y_{i,t}$, $i \in I = \{1, \ldots, 50\}$, $t \in \{1, \ldots, 10,000\}$. The index i denotes the i^{th} location, with locations $i \in \{1, \ldots, 25\}$ and $i \in \{26, \ldots, 50\}$ belonging to the administrative areas of governments 1 and 2, respectively. Index t denotes the time point in days; however, since $Y_{i,t}$ are IID for all i, we drop the subscript t for the remainder of this section.

Many multivariate extreme value models are only applicable in low-to-moderate dimensions, and we would not generally expect such techniques to scale to the 50-dimensional setting (Engelke and Ivanovs, 2021). It is therefore reasonable to consider techniques for dimensional reduction. We explore this possibility by examining the extremal dependence structure of the data. In particular, we look at pairwise $\chi(u)$ extremal dependence coefficients, introduced in equation (8.2.1), for all possible pairwise combinations of sites; the resulting estimates with $u = 0.95$ are visualised in the heat map of Figure 8.5.1.

Figure 8.5.1 suggests the existence of 5 distinct subgroups where all variables within each subgroup have similar extremal dependence characteristics, but variables in different subgroups appear to be approximately independent of each other in the extremes. It is worth noting that the same clusters are identified when examining pairwise $\eta(u)$ extremal dependence coefficients; the resulting estimates can be found in Appendix E. Moreover, examining the magnitudes of the aforementioned $\chi(\cdot)$ and $\eta(\cdot)$ estimates, assuming asymptotic dependence between variables in each group does not appear reasonable. We therefore consider models that can be applied to data structures that do not take their extreme values simultaneously. The indices of the five aforementioned subgroups are given by:

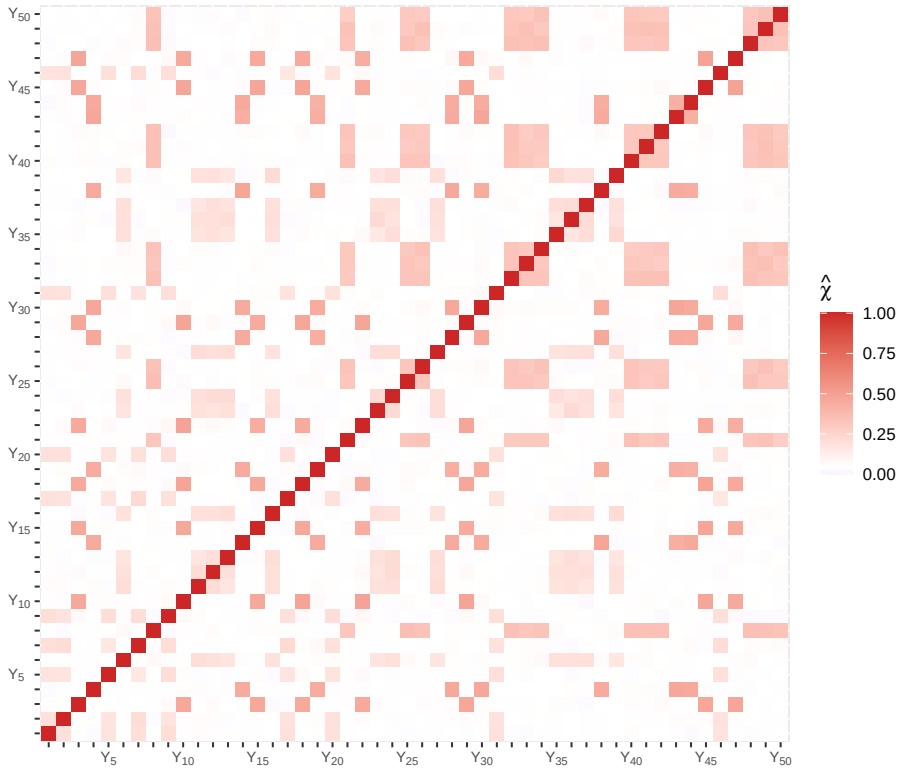


Figure 8.5.1: Heat map of estimated empirical pairwise $\chi(u)$ extremal dependence coefficients with $u = 0.95$.

$$G_1 = \{4, 14, 19, 28, 30, 38, 43, 44\},$$
$$G_2 = \{3, 10, 15, 18, 22, 29, 45, 47\},$$
$$G_3 = \{8, 21, 25, 26, 32, 33, 34, 40, 41, 42, 48, 49, 50\},$$
$$G_4 = \{1, 2, 5, 7, 9, 17, 20, 31, 46\},$$
$$G_5 = \{6, 11, 12, 13, 16, 23, 24, 27, 35, 36, 37, 39\}.$$

Groups G_1 and G_2 include the most strongly dependent variables (shown by the darkest colour blocks in Figure 8.5.1), followed by groups G_3 and G_4, while group G_5 contains the most weakly dependent variables. We henceforth assume independence between these groups of variables, i.e., $\Pr((Y_i)_{i\in G_k} \in A_k, (Y_i)_{i\in G_{k'}} \in A_{k'}) = \Pr((Y_i)_{i\in G_k} \in A_k) \Pr((Y_i)_{i\in G_{k'}} \in A_{k'})$, $A_k \subset \mathbb{R}^{|G_k|}, A_{k'} \subset \mathbb{R}^{|G_{k'}|}$, for any $k \neq k' \in \{1, \ldots, 5\}$.

Challenge C4 requires us to estimate the probabilities $p_1 = \Pr(Y_i > s_i;\, i \in I)$ and $p_2 = \Pr(Y_i > s_1;\, i \in I)$, where $s_i := \mathbb{1}(i \in \{1, 2, \ldots, 25\})s_1 + \mathbb{1}(i \in \{26, 27, \ldots, 50\})s_2$

and s_1 (s_2) denotes the marginal level exceeded once every year (month) on average. We refer to p_1 and p_2 as parts 1 and 2 of C4, respectively. Under the assumption of independence between groups, the challenge can be broken down to 5 lower-dimensional challenges involving the estimation of joint tail probabilities for each G_k, $k \in \{1, \ldots, 5\}$. These can then be multiplied together to obtain the required overall probabilities due to (assumed) between-group independence. Specifically, we have $p_1 = \prod_{k=1}^{5} \Pr(Y_i > s_i;\ i \in G_k)$ and $p_2 = \prod_{k=1}^{5} \Pr(Y_i > s_1;\ i \in G_k)$. We now consider the estimation of within group probabilities.

8.5.2 Conditional extremes

In this section, we detail the modelling approach proposed by Heffernan and Tawn (2004), commonly referred to as the conditional extremes model. The main appeal of this approach is that it provides a flexible multivariate extreme value framework, capable of capturing a range of extremal dependence forms. It has thus been applied extensively to the joint modelling of environmental data sets (Keef et al., 2013; Jonathan et al., 2014) and in the high dimensional setting (e.g., Quinn et al., 2019). For these reasons, we opt to employ this as our method for modelling extremes in challenge C4.

Consider a d-dimensional random variable $\boldsymbol{W} = (W_1, \ldots, W_d)$ on Laplace margins. The conditional extremes model approach assumes the existence of normalising functions $\boldsymbol{a}_{|i}$, $\boldsymbol{b}_{|i}$ such that

$$\lim_{w_i \to \infty} \Pr\left\{ \boldsymbol{W}_{-i} \leq \boldsymbol{a}_{|i}(w_i) + \boldsymbol{b}_{|i}(w_i)\, \boldsymbol{z}_{|i} \mid W_i = w_i \right\} = H_{|i}\left(\boldsymbol{z}_{|i}\right), \tag{8.5.1}$$

for non-degenerate distribution functions $H_{|i}$, $i \in \{1, \ldots, d\}$, where $\boldsymbol{W}_{-i}$ denotes the vector $\boldsymbol{W}$ with the i^{th} component removed. Here, $\boldsymbol{z}_{|i}$ is within the support of the

residual random vector

$$\boldsymbol{Z}_{|i} = (\boldsymbol{W}_{-i} - \boldsymbol{a}_{|i}(w_i))/\boldsymbol{b}_{|i}(w_i) \sim H_{|i}. \tag{8.5.2}$$

A direct consequence of assumption (8.5.1) is that, conditional on $W_i > u_i$, the variables $W_i - u_i$ and $\boldsymbol{Z}_{|i}$ are independent in the limit as $u_i \to \infty$, with limiting distributions as unit exponential and $H_{|i}$, respectively.

Parametric families for the normalising functions $\boldsymbol{a}_{|i}$ and $\boldsymbol{b}_{|i}$ were provided for the case of Laplace margins by Keef et al. (2013). Specifically, they set

$$\boldsymbol{a}_{|i}(w_i) = \boldsymbol{\alpha}_{-|i} w_i \quad \text{and} \quad \boldsymbol{b}_{|i}(w_i) = w_i^{\boldsymbol{\beta}_{-|i}}, \tag{8.5.3}$$

for $\boldsymbol{\alpha}_{-|i} \in [-1, 1]^{d-1}$ and $\boldsymbol{\beta}_{-|i} \in (-\infty, 1]^{d-1}$, with vector operations applied componentwise. Let $\alpha_{j|i}$ and $\beta_{j|i}$, $j \in \{1, \ldots, d\} \setminus \{i\}$, denote the elements of the respective parameter vectors associated with variable W_j.

Estimation of the $\boldsymbol{\alpha}_{-|i}$ and $\boldsymbol{\beta}_{-|i}$ can be achieved via standard maximum likelihood techniques, provided we make an additional assumption regarding the parametric form of the distribution $H_{|i}$. Accepted practice is to assume IID marginal distributions $H_{j|i} \sim N(\mu_j, \sigma_j^2)$, $\mu_j \in \mathbb{R}, \sigma_j \in \mathbb{R}_+$ for all $j \in \{1, \ldots, d\} \setminus \{i\}$, as suggested by Heffernan and Tawn (2004).

Having obtained maximum likelihood estimates for $\boldsymbol{\alpha}_{-|i}$, $\boldsymbol{\beta}_{-|i}$ for some high threshold u_i of W_i, extreme predictions in the form of $(\boldsymbol{W}_{-i}, W_i)|W_i > v_i$ can be made for any $v_i \geq u_i$, by exploiting the limiting independence of $(W_i - v_i)$ and $\boldsymbol{Z}_{|i}$. Drawing a realisation $\boldsymbol{z}_{|i}$ of $\boldsymbol{Z}_{|i}$ from the set of fitted residual values (8.5.2) and independently generating an observation from $(W_i - v_i)$ provides a new observation of $\boldsymbol{W}$ in the region $\{\boldsymbol{w} \in \mathbb{R}^d; w_i > v_i\}$. Algorithm 1 details how we employ this method when estimating extreme set probabilities.

We apply the above methodology when estimating probabilities p_1 and p_2 defined

Algorithm 1 Extremal probability estimation using conditional extremes

1: Fit the conditional extremes model as described above.

2: Select a number $N \in \mathbb{Z}^+$ of values to generate.

3: **for** $j \in \{1 \dots, N\}$ **do**

4: Simulate a realisation w_i of $(W_i - v_i)|(W_i > v_i)$ from a unit exponential distribution.

5: Sample $\mathbf{z}_{|i}$ from the set of fitted residual values (8.5.2).

6: Set $\boldsymbol{w}_{-i} = \boldsymbol{\alpha}_{-|i} w_i + \boldsymbol{\beta}_{-|i}^{w_i} \boldsymbol{z}_{|i}$.

7: Obtain a realisation $\boldsymbol{w}^j := (\boldsymbol{w}_{-i}, w_i)$ of $(\boldsymbol{W}_{-i}, W_i)|W_i > v_i$.

8: **end for**

From this we obtain realisations $\{\boldsymbol{w}^j\}_{j=1}^N$.

9: Estimate the probability $\Pr(\boldsymbol{W} \in A)$ for $A \subset \{\boldsymbol{x} \in \mathbb{R}^d;\, x_i > v_i\}$ via

$$\hat{p}_A = \left(\frac{1}{N} \sum_{j=1}^{N} \mathbb{1}(\boldsymbol{w}^j \in A) \right) \left(\frac{1}{2} \exp(-v_i) \right),$$

where the right-side term originates from the unit Laplace marginal exceedance probability.

in Section 8.5.1. The analysis of Section 8.5.1 identified five subgroups $G_1, \dots, G_5$, between which we assume independence. As is discussed in Section 8.5.1, we leverage this dependence pattern when estimating probabilities p_1 and p_2 by calculating the individual subgroup probabilities separately.

We first transform the data onto Laplace margins via $W_i := F_L(F_G^{-1}(Y_i))$, where F_G and F_L denote standard Gumbel and Laplace distribution functions, respectively. Setting $\boldsymbol{W} := (W_1, \dots, W_{50})$, we divide this random vector into the five subgroups identified in Section 8.5.1. The subgroup probabilities are each estimated using Algorithm 1 for each subvector. For this, we select the first element of each subvector as the conditioning variable and simulate 10^8 replicates from each fitted model to obtain

probability estimates. To account for uncertainty in the estimation of the normalising functions (8.5.3), we perform a parametric bootstrapping procedure with 100 samples. That is, we fit an initial conditional extremes model for a given conditioning threshold, then use this model to generate 100 predictive samples and apply Algorithm 1 to each to estimate 100 realisations of the target probabilities. Sensitivity analyses of the estimated probabilities to the choice of conditioning variable suggest no significant effect. Furthermore, we consider a range of conditioning thresholds; the corresponding estimates of subgroup probabilities defined in Section 8.5.1 appear relatively stable with respect to the conditioning threshold quantile. We ultimately select 0.85-quantiles for the conditioning thresholds of our final probability estimates.

8.5.3 Results

Figure 8.5.2 shows the bootstrapped estimated individual group and overall probabilities with respect to the conditioning threshold quantile for part 1. The results for part 2 are given in Appendix E. Our final estimates are given by $\hat{p}_1 = 1.093634 \times 10^{-26}\,(2.149591 \times 10^{-36}, 1.359469 \times 10^{-24})$ and $\hat{p}_2 = 1.075787 \times 10^{-31}\,(1.596381 \times 10^{-46}, 1.850425 \times 10^{-29})$, with 95% confidence intervals given in parentheses.

8.6 Discussion

In this paper, we have proposed a range of statistical methods for estimating extreme quantities for challenges C1-C4. For the univariate challenge C1, we estimate an extremal quantile for a response conditioned on a set of environmental covariates. Specifically we wanted to estimate the 0.9999-quantile, and the associated 50% confidence intervals, of $Y \mid \boldsymbol{X} = \boldsymbol{x_i}$, $i \in \{1, \ldots, n\}$. For challenge C2, we estimated a quantile, corresponding to a once in 200 year level, of the marginal distribution Y whilst incorporating the loss function in equation (8.3.2). Overall we ranked 6^{th} and and 4^{th} for

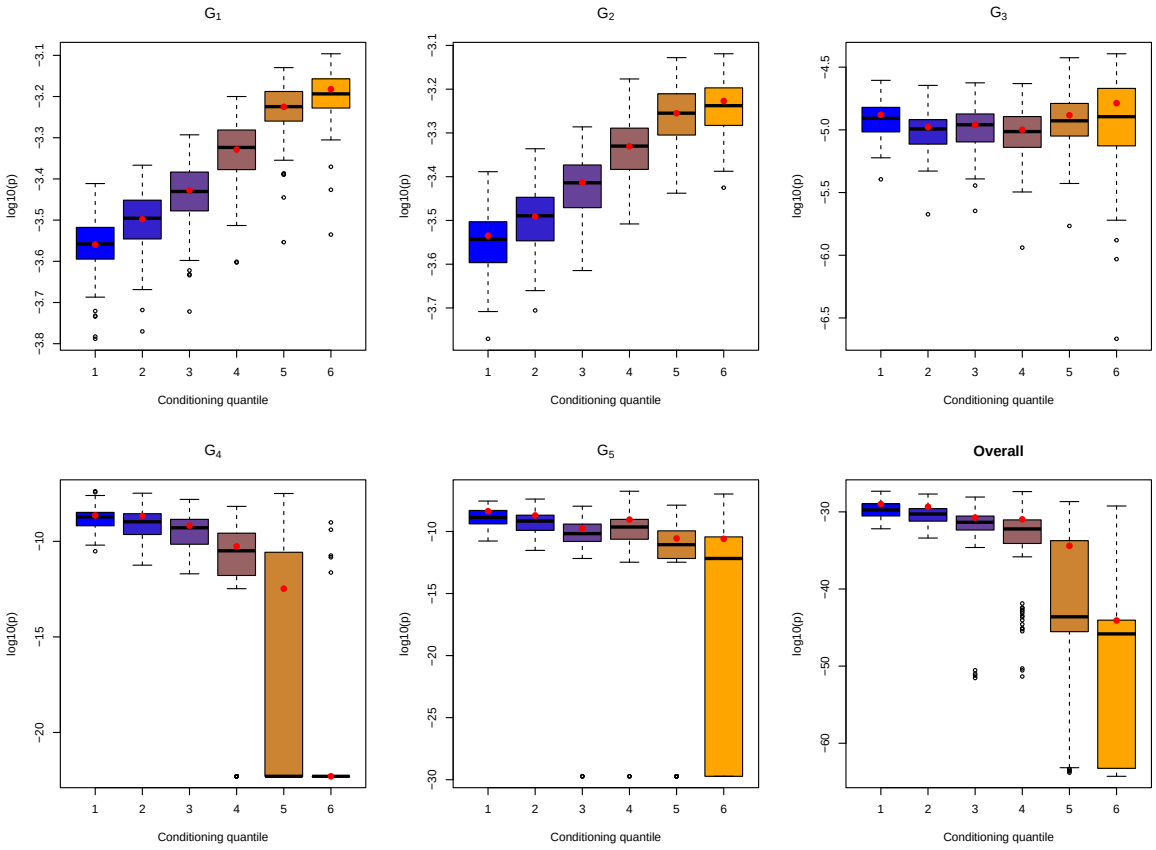


Figure 8.5.2: Part 1 subgroup and overall bootstrapped probability estimates on the log scale. The red points indicate the original sample estimates and the colouring of the boxplots indicates the choice of conditioning threshold, with the conditioning quantile indices 1-6 referring to the quantile levels $\{0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$, respectively.

challenges C1 and C2, respectively.

For challenge C1, our final model (model 7 in Table 8.3.1) was chosen to minimise the model selection criteria but the QQ plot under-estimates the most extreme values of the response (see Figure 8.3.2). Despite this under-estimation, the conditional quantiles calculated for C1 are generally well aligned with the true quantiles. If we ignore the model selection criteria and chose the model based on a visual assessment of QQ plots, we would have chosen model 5 in Table 8.3.1 and this would have covered the true quantile on fewer occasions than our chosen model. Therefore, the main issue with our results is the narrow confidence intervals.

Narrow confidence intervals are an indication of over-fitting and this could have

arisen in several places. For instance, Rohrbeck et al. (2023) suggested all the seasonality is captured in the threshold, while our model includes a seasonal threshold and a covariate for seasonality in the scale parameter of the GPD model. As well as overfitting, the model may not have been flexible enough. This could be, in part, due to our model missing covariates. For instance, the true model contained V_2 as a covariate (Rohrbeck et al., 2023) whilst our model did not. In addition, the basis dimensions for our splines are low. In practice, one should choose a higher dimension than we would expect and although we chose the dimension using a model-based approach, it may have resulted in the splines not being flexible enough to capture all of the trends in the data. Since we used the same model for challenges C1 and C2, these issues can also be attributed to our ranking in C2.

For the first multivariate challenge C3, we employ an extension of Wadsworth and Tawn (2013) to estimate the probability of three variables lying in an extremal set. Our extension to this framework accounts for non-stationarity in the data when estimating the associated coefficient which relies on GAMs to obtain the parameters. The QQ plots for the resulting fitted exceedance model suggested a reasonable fit; however, parameter estimates show the proposed modelling approach was not the most suitable. For this challenge, we ranked 5^{th} and our estimates are on the same order of magnitude as the truth (Rohrbeck et al., 2023).

We note similarities in the methodologies presented for the challenges C1, C2, and C3. Specifically, each of the proposed methods used the EVGAM framework for capturing non-stationary tail behaviour via a generalised Pareto distribution. We acknowledge that the model selection tool proposed for C1 and C2 could also be applied for C3. However, we opted not to use this tool for several reasons. Firstly, unlike the univariate setting, there is no guarantee of convergence to a GPD in the limit, and the GPD tail assumption thereby needs to be tested. Moreover, in exploratory analysis, we tested the model selection tool for C3 but found the selected models and quantiles to not be

satisfactory, particularly in the upper tail of the min-projection. We therefore select a model manually, using QQ plots to evaluate performance. Exploring threshold and model selection techniques for multivariate extremes represents an important area of research.

In the final multivariate challenge C4, we were asked to estimate very high-dimensional joint survival probabilities. To do so, we split the probability into 5 lower-dimensional asymptotically independent components, then estimated each using the conditional extremes method of Heffernan and Tawn (2004). In the final rankings of Rohrbeck et al. (2023), we ranked 3$^{\text{rd}}$ for this challenge. A more prudent method could have been implemented, as groups of variables were never truly independent. Alternatively, we could consider estimating individual group probabilities across varying thresholds and then taking an average value as our final result. Even though sensitivity analyses indicate little effect of thresholds on model parameter estimates, it may have had a more significant impact on final probability estimates.

Chapter 9

Conclusions and further work

In this final chapter, we summarise our contributions to the area of extreme value statistics that result from Chapters 3-8 of this thesis. The research presented advances to the existing methodology for modelling extremal behaviour of sea levels and river flows. Accounting for non-stationarity and temporal dependence in these univariate processes have been key themes throughout this thesis and our methods have demonstrated an improvement in existing approaches. We summarise the content of each chapter in Section 9.1, before identifying avenues for further work in Section 9.2.

9.1 Summary of contributions

In Chapter 3 we developed a novel methodology for estimating extreme sea levels by accounting for seasonality in skew surge and peak tide, the dependence between them, and temporal dependence in skew surges. Our results showed a significant improvement on current methods, which ignored these features of the sea level processes and instead made several simplifying assumptions. Our model also allowed us to study the seasonality of sea levels exceeding levels previously unobserved which can be useful for coastal defence maintenance planning. The return levels estimated from our model presented a more accurate representation of future extreme events, and will be useful

for future coastal defence upgrades. Seasonal variations are a common feature of many environmental variables and therefore should be accounted for when estimating return levels; the methodology discussed in Chapter 3 could apply to a range of variables.

We extended the work of Chapter 3 in Chapter 4 by presenting a framework to investigate the effects of anthropogenic climate change on extreme skew surges. Increases in the magnitude or frequency of these events can have catastrophic consequences if not included in extreme sea level estimation for coastal flood defence design. These trends are typically different to those observed in the main body of the data, such as mean sea level rise. We used year and global mean temperature anomaly (GMT) as covariates in our statistical model for extreme event occurrence. After accounting for mean sea level trends, we investigated changes in the probability of an extreme skew surge, i.e., the rate parameter of the generalised Pareto distribution (GPD), with GMT across the different seasons and sites on the UK National Tide Gauge Network. We showed there is evidence of an increasing trend for most seasons across sites, however, occasionally we found decreasing trends within seasons. We did not find any significant changes in the magnitude of extreme skew surges, i.e., in the GPD scale parameter, and hence in the mean of the skew surge excesses of the threshold. Accounting for seasonal changes in extreme skew surge occurrence with GMT in sea level return level estimation showed that return levels increase with GMT. The ideas presented in Chapter 4 could be applied to more locations, but also to other environmental variables to investigate trends in extreme values.

Chapter 5 presented a procedure for simulating skew surges that reproduces the stochastic behaviour of their seasonality and extreme values (using ideas from Chapter 3), as well as their temporal dependence. It is fundamental to understand the temporal dependence of skew surges for coastal erosion maintenance planning as prolonged storm events, or clustering of separate storm events that span multiple tidal cycles, accelerate erosion rates. We used a copula framework, assuming the data fol-

lows a Markov process, to model the pairwise dependence structure of values separated by different lags. Since we found evidence of asymptotic independence between such values, we used the Gaussian copula to capture this extremal behaviour. We found that the temporal dependence structure is not constant throughout the year, with stronger dependence found in the summer months, and we captured this using harmonics for the Gaussian copula dependence parameters. We provided simulations at four UK National Tide Gauge Network sites.

In Chapter 6 we introduced max-autoregressive moving average (Max-ARMA) processes as a potential candidate for modelling extreme river flows. Large precipitation events cause sudden spikes in river flow data that then exponentially decay, so the data typically have features of heavy tails and non-linearity; Max-ARMA processes are powerful tools for modelling time series of this nature, capturing their unique temporal dependence structure. We discussed conditions for a Max-ARMA process to be stationary and identifiable. For inference, we introduced a moments-based procedure using the extremal index and coefficient of asymptotic dependence that we also derived in this chapter. We illustrated this procedure on river flow data from the River Thames, London (UK).

Chapters 7 and 8 detailed contributions, as part of a wider team, for the Extreme Value Analysis conferences held in 2021 and 2023, respectively. In Chapter 7, we discussed a marginal modelling approach for predicting wildfire events across the contiguous United States. We obtained estimates of the cumulative distribution function at locations with missing entries of wildfire count or burnt area. The resulting estimates were then "ranked" using a score function weighted to give higher importance to extreme observations (Opitz, 2023). Our method produced a significant improvement on the proposed benchmark technique. Then in Chapter 8 we proposed a range of statistical methods for estimating extreme quantities for the second data challenge, split into four sub-challenges: C1-C4. For the univariate challenge C1, we estimated an

extremal quantile for a response conditioned on a set of environmental covariates, with the associated 50% confidence intervals. For challenge C2, we estimated the marginal 200-year return level of the same environmental response variable whilst incorporating a pinball loss function. For the first multivariate challenge C3, we employed an extension to the framework of Wadsworth and Tawn (2013) to estimate the probability of three variables lying in an extremal set, whilst accounting for non-stationarity in the data using generalised additive models. In the final multivariate challenge C4, we estimated 50-dimensional joint survival probabilities by splitting the probability into five lower-dimensional asymptotically independent components and estimating each using the conditional extremes method of Heffernan and Tawn (2004).

9.2 Further work

In the following sections we discuss interesting avenues for further research to extend the ideas presented in this thesis. In Section 9.2.1 we discuss ideas for further developments to our novel methodology for estimating extreme sea levels, which was introduced in Chapters 3 and 4. Then in Section 9.2.2 we discuss potential avenues for extending our work from Chapter 5 regarding skew surge simulation. In Section 9.2.3 we detail how we envision the Max-ARMA model presented in Chapter 6 being used to capture temporal dependence for river flows so that a realistic marginal model for extreme events can be developed. Lastly, in Section 9.2.4 we discuss how to combine simulated river flow and skew surge time series for surge barrier maintenance, taking care with the dependence between these variables.

9.2.1 Extreme sea level estimation in Chapters 3 and 4

A key part of our methodologies in Chapters 3 and 4 was accounting for skew surge within-year seasonality. We used a non-stationary threshold, defined as a quantile of

the monthly skew surge distribution, to define extreme values and develop our non-stationary GPD model. Using a quantile ensures there are a similar number of exceedances to model per month. This approach is similar to that of Carter and Challenor (1981) since we first assumed stationarity within months and then built in the seasonal variation on a shorter temporal resolution through covariates in the GPD parameters. This meant that we were able to capture most of the non-stationarity, as well as skew surge-peak tide dependence, at the same stage of the modelling process. However, we could have considered a smoother threshold choice by using quantile regression (Northrop et al., 2016) but we did not try this as our monthly threshold appeared sufficient. Further work could compare differing non-stationary threshold functions; the `evgam` R package introduced in Section 2.4.1 of Chapter 2 can be used for doing so via quantile regression (Youngman, 2019, 2022).

Skew surges are also believed to change over decadal time scales with climate indices. For example, the North Atlantic Oscillation index (NAO) describes such time scale changes in regional weather systems, so is believed to impact storm surges, and thus skew surges. Araújo and Pugh (2008) find a negative correlation between storm surge and air pressure patterns, using NAO. It would be interesting to explore how adding an NAO covariate into the GPD for extreme skew surges changes model fit however, this poses two challenges: NAO is difficult to forecast (Siegert et al., 2016), so using a model of this type for future prediction may be challenging. Secondly, the marginal distribution of NAO is unknown so would need estimating to obtain the marginal distribution of skew surges; this could introduce further uncertainty into the skew surge model and care might need to be taken when modelling the tails of NAO.

As demonstrated in Chapter 4, spatial pooling provides a promising framework to capture longer-term trends due to climate change; single site trends are subtle but sharing information across sites gives more significant results. Spatial pooling also enables inference at locations with limited or no data, where our current single site

model would not perform well. Our results of Chapter 4 are primarily illustrative since we considered just four sites here. There are 44 sites on the UK National Tide Gauge Network where this methodology could be extended so that trends could be shared amongst suitable spatial regions.

In Chapter 3 we introduced a prior penalty for the GPD shape parameter based on spatial information. Our results showed drastic reductions in the uncertainty associated with longer term return levels. We found overlap in confidence intervals for the shape parameter across all sites studied, so it may be reasonable to fix this parameter, to be common but unknown, across these sites as an alternative approach for borrowing information. Fixing the shape parameter to be common in homogeneous regions is a crucial step in regional frequency analysis, originally introduced by Hosking and Wallis (1997). It could be interesting to combine our findings of Chapters 3 and 4 to apply the method in a regional frequency analysis framework, where sites in a homogeneous region not only have a common shape parameter, but also common longer term trends due to anthropogenic climate change. We refer the reader to Batstone et al. (2013), Bernardara et al. (2011) and Haigh et al. (2010) for different approaches to spatial pooling for extreme sea level estimation.

9.2.2 Simulating skew surges in Chapter 5

Our skew surge simulation procedure of Chapter 5 demonstrated promising results for replicating the temporal dependence structure of observations. We compared estimates of well-known extremal dependence measures (the extremal index and the coefficients of asymptotic (in)dependence; see Chapter 2 for definitions of these) from the observed data and over 200 simulated samples. Whilst our simulations matched the observational estimates well (see Section 5.5.2 of Chapter 5), further work could investigate if we can improve our results. For example, we used the Gaussian copula for modelling the temporal dependence as this can capture the asymptotic independence structure

exhibited by the skew surge data, however, other choices could be explored, such as the Frank, Clayton and Inverted Logistic copulas (Joe, 2014). Alternatively, when incorporating non-stationarity into the dependence parameter of the Gaussian copula we used harmonics, however, other approaches could be investigated such as using generalised additive models (Chavez-Demoulin and Davison, 2005) or splines (Youngman, 2019).

As discussed in Section 9.2.1, we shared information across sites regarding the GMT covariate in Chapter 4. We could adopt a similar approach to pool information across sites that share temporal dependence characteristics. For example, Sheerness and Lowestoft are both located on the east coast and will be affected by similar storms, therefore we expect them to have a similar dependence structure. This would reduce the uncertainty associated with dependence parameter estimation and ultimately the simulations. However, we only studied four sites here so continuing with the single site analysis for a larger set of sites in close proximity would be a necessary first step.

By replicating simulations over 200 samples, we obtained measures of uncertainty for our simulations that will be useful in practice. However, it would be insightful to have confidence intervals for the Gaussian copula dependence parameters. Then, we could check for overlap in dependence parameters across different lags or confidence intervals that contain zero, where a parameter could be disregarded. The uncertainty in dependence parameter estimates should then be propagated through into the uncertainty quantification for our simulations. The block bootstrap procedure is a potential method for quantifying uncertainty that preserves temporal dependence (Politis and Romano, 1994). Alternatively, since we used likelihood inference to fit the copula models in Chapter 5 (under a Markov assumption), the asymptotic normality of the maximum likelihood estimates could be exploited to obtain confidence intervals based on the Hessian.

The simulations we presented in Chapter 5 correspond to the historic period of observation, but in practice, simulations of future records are more interesting and

important. Obtaining simulations for a future time period is straightforward using the predicted tidal regime. However, we used the marginal model of Chapter 3 which assumes a steady state climate and is unrealistic in practice. Predicted trends in mean sea level under different climate scenarios can be easily added back onto the simulations. However, there exist longer term trends in extreme skew surges that differ from those in the mean, as identified in Chapter 4. We can also capture these trends here by adding a GMT covariate into our marginal GPD model so that simulations can be obtained under different scenarios of future GMT increases.

9.2.3 Max-ARMA models for extreme river flow in Chapter 6

The theoretical results derived for Max-ARMA models in Chapter 6 were motivated by the need to find a suitable candidate for capturing the temporal dependence of river flow series, particularly in its extreme states. A large precipitation event causes river flows to spike and then remain high for days as the large volume of water propagates downstream. Max-ARMA processes are a suitable candidate for modelling this behaviour, and we introduced a moments-based inference approach using the extremal properties derived in Chapter 6 to fit such models. However, we focussed on stationary Max-ARMA processes and, like other environmental processes considered in this thesis, river flows exhibit non-stationarity. Figure 9.2.3 shows the within-year seasonality in the UK River Thames, observed at Kingston-upon-Thames, the same data used in Chapter 6 but there only winter data were considered. Future work is required to preprocess the data to remove the seasonal trend; we discuss our initial ideas below.

We suggest extracting the baseflow process to obtain a residual series without these strong seasonal trends in the main body of data. The baseflow of a river is the portion of its flow attributed to its groundwater seepage into the river channel so that it persists even during dry periods. We aim to estimate the baseflow from our series so that it can be removed. Let $\{Y_t : \ t = 1, \ldots, n\}$ denote the daily maximum river flow series over

n observations; knowing t tells us the day in year $d \in [1, 365]$, so for some function f, $d = f(t)$. We estimate baseflow by first finding the minimum observation within some time window $w \in \mathbb{N}$ around the day of interest d for a given year $k = 1, \ldots, K$ where K denotes the total number of years, formally defined as $m_d^{(k)} = \min\{Y_{d-w}^{(k)}, \ldots, Y_{d+w}^{(k)}\}$. Then, we obtain a minimal set for each day d across all years k as $m_d = \left\{m_d^{(k)};\ k = 1 \ldots, K\right\}$. Lastly, we find the quantile $q \in [0, 1]$ of observations in the minima set for day d across all years and denote this q_d. We call the series $\{q_d\}_{d=1}^{365}$ the baseflow process, and this is the same across years so it has a cyclic annual pattern. Then we obtain a residual process $\{\epsilon_t\}$ defined as $\epsilon_t = Y_t - q_{f(t)}$ that can be assumed as more stationary than $\{Y_t\}$ within the body of data, although the extremes may still exhibit seasonality.

The quantile q and time window w must be carefully chosen; increasing the time window will create a smoother baseflow but too wide will lose the specific seasonal behaviour. Increasing the quantile q will capture more temporal dependence in the baseflow, so less for the residual process, however, too high will result in removing too much information from the data. Figure 9.2.1 shows an example from Kingston-upon-Thames for differing quantiles and time windows; this shows that $q = 0.75$ yields a much smoother curve than $q = 0.2$, and having w too low (here $w = 2$) yields a baseflow with almost very little seasonal structure. Then we fix $w = 5$ and $q = 0.75$ to illustrate that the residual process has less temporal dependence (Figure 9.2.2 shows a reduction in the acf values, especially for large lags) and less seasonality in the main body of data (see Figure 9.2.3 where the median is constant across months for the residuals). Extreme residual values have a similar seasonal structure to extreme skew surges as in Chapter 3, so a similar approach using a non-stationarity version of the GPD could be used to model the tail here.

Simulations from an appropriate Max-ARMA model act as simulations of the residual process extremes. However, these will initially be on unit Fréchet margins, by definition of a Max-ARMA process, so a marginal transformation is required to transform

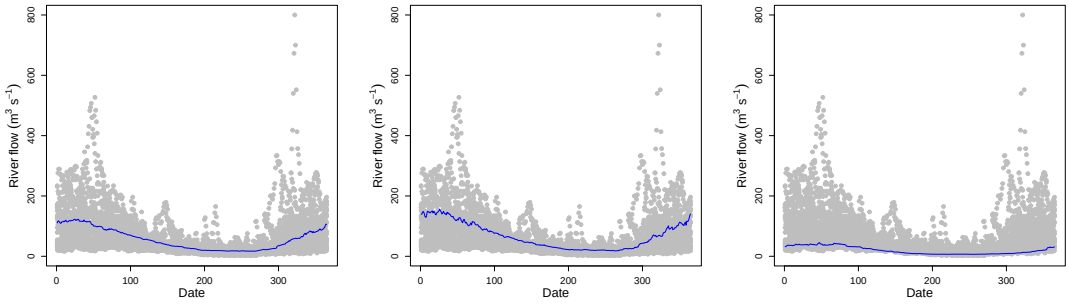


Figure 9.2.1: Baseflow process (blue) for time window $w = 5$ (left and right) and $w = 2$ (centre), and quantile $q = 0.75$ (left and centre) and $q = 0.2$ (right) plotted against the day of year d. A 20 year series of river flow observations (from Kingston-upon-Thames) are shown by grey points.

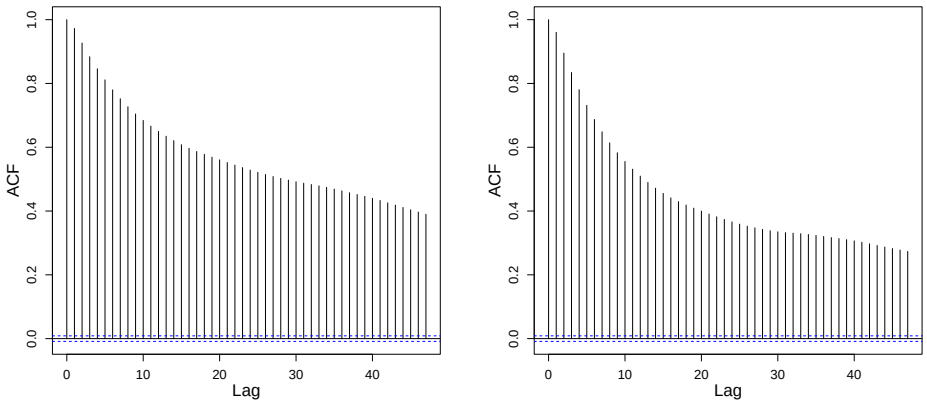


Figure 9.2.2: Autocorrelation function (acf) plots for the Kingston-upon-Thames data (left) and its corresponding residual series (right), once the baseflow is removed with $w = 5$ and $q = 0.75$.

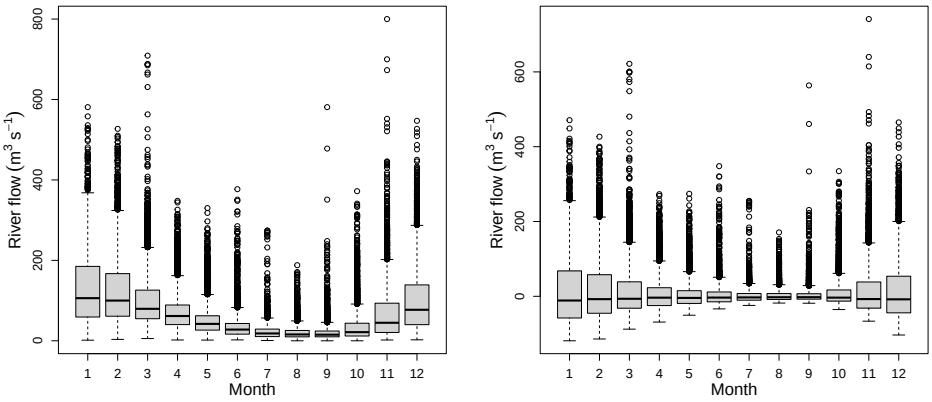


Figure 9.2.3: Boxplots of observations per month for the Kingston-upon-Thames data (left) and its corresponding residual series (right), once the baseflow is removed with $w = 5$ and $q = 0.75$.

this series to the appropriate scale; we discuss such a transformation in Section 6.5.3 of Chapter 6. These simulations can then be added to the baseflow process to obtain a simulated series of river flow extremes with the correct seasonal structure in the main body. Temporal trends in the tail will still need to be accounted for, this can be done during the marginal transformation for the residuals. However, exploring how capturing seasonality can be incorporated into the Max-ARMA framework (e.g., by including temporal covariates on the model parameters) would be an interesting avenue for further work.

9.2.4 Combining simulations from Chapters 5 and 6

One motivation driving the research presented in Chapter 5 was for surge barrier maintenance at estuary locations, such as the Thames Barrier. At these locations, barriers mitigate against coastal and fluvial flooding. In Chapter 5, we simulated time series of skew surge at Sheerness that can be used to estimate barrier closures for coastal flooding at the Thames Barrier. In Section 9.2.2, we discuss how to simulate river flow time series using the Max-ARMA model; these simulations can also be used to predict barrier closure rates resulting from fluvial flooding. However, combining these

	All	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
τ	0.16	0.20	0.19	0.21	0.18	0.17	0.08	0.21	0.14	0.09	0.15	0.05	0.17
χ	0.08	0.13	0.12	0.06	0.08	0.04	0.15	0.04	0.07	0.03	0.11	0.05	0.04
$\bar{\chi}$	0.28	0.25	0.26	0.001	0.08	0.03	0.30	0.09	0.04	-0.04	0.24	-0.002	0.13

Table 9.2.1: Empirical estimates of Kendall's τ measure, χ and $\bar{\chi}$ (both at the 0.99 quantile) for skew surge and river flow daily maximum data at Sheerness and Kingston-upon-Thames, respectively.

simulations assumes skew surge and river flow are independent, which is a reasonable assumption, but since weak dependence does exist between them, further work could investigate modelling this relationship (Hendry et al., 2019).

Table 9.2.1 illustrates initial exploratory analysis into the dependence between skew surge daily maxima at Sheerness and river flow daily maxima at Kingston-up-Thames. Kendall's τ measure suggests there is weak dependence in the body of the data. We also look at estimates on a monthly scale to assess if the relationship between skew surge and river flow is time-varying. We find the strongest dependence in the body (i.e., highest values of Kendall's τ) in March and July, followed by January and then February.

It is also important to consider dependence in the extremes since an extreme event of either variable can be destructive, but a combination of extreme events over multiple hazards is even more so. Therefore, it is fundamental to determine if extremal dependence exists between skew surge and river flow. To investigate this, we empirically estimate bivariate asymptotic dependence measures χ and $\bar{\chi}$ (see Section 2.3 of Chapter 2 for details) for exceedances of the 0.99 quantiles; see Table 9.2.1. These estimates suggest asymptotic independence across all months so that both variables are unlikely to be extreme at the same time regardless of the time of year they occur. If jointly modelled, a suitable asymptotically independent model could be implemented as there is some evidence, particularly for January, February, June and October that $\bar{\chi} > 0$ indicating a departure from independence in the occurrence of joint extremes. For example, in a copula framework, the Gaussian copula would be suitable but the

bivariate extreme value logistic copula would not; both are introduced in Section 2.3.4 of Chapter 2. It would be interesting to assess the importance of capturing the joint river flow and skew surge dependence on the annual rate of barrier closures, akin to the effect of the temporal dependence study in Chapter 5.

Appendix A

Supplementary material for Chapter 3: Accounting for seasonality in extreme sea level estimation

A.1 Introduction

This document outlines the supplementary material for Chapter 3. Firstly in Section A.2, we present the exploratory analysis to assess skew surge-peak tide dependence, discussed in Section 3.3.3 of Chapter 3; we demonstrate a time-varying relationship at Sheerness. We fit the tide dependent skew surge model presented in Section 3.4.4 of Chapter 3 to a 483 year physical model dataset in Section A.3, to illustrate the physical justification of modelling skew surge-peak tide dependence. Then in Section A.4 we investigate temporal dependence in the skew surge series, this was discussed in Section 3.3.2 of Chapter 3 and then accounted for in our methodology in Section 3.4.5. We investigate the sensitivity of our sea level return level estimates to the choice of

the threshold used in the skew surge generalised Pareto distribution (GPD) model in Section A.5. We present the results of a simulation study in Section A.6, where we estimate sea level return levels using our proposed method and that used in practice (the skew surge joint probabilities method of Batstone et al. (2013), abbreviated to SSJPM) to samples of 483 years of physical model data. In Section A.7 we derive an analytical expression for the probability that a randomly selected sea level annual maxima is from a particular month, given it is equal to a return level (see equation (3.5.2) of Chapter 3); this is evaluated in Section 3.5.5 of Chapter 3 to understand the seasonality of extreme sea levels. In Section A.8 we detail the process of transforming skew surges to uniform margins using the final model presented in Chapter 3, and we use this as a means of assessing skew surge model fit; this was used in the bootstrap procedure for uncertainty quantification on return level estimations in Section 3.5.4 of Chapter 3. Lastly, we present supplementary figures in Section A.9.

A.2 Skew surge-peak tide dependence

In this section, we present our exploratory analysis to demonstrate that it is reasonable to assume skew surge and peak tide are independent at Heysham, Lowestoft and Newlyn, but not at Sheerness. We perform various statistical tests at all sites. Firstly, we formally test if there is a relationship between extreme skew surges and their associated ranked peak tide, where extreme skew surges are defined as exceedances of different thresholds. Then we investigate if all peak tides come from the same distribution as those associated with extreme skew surges. We also do this on a monthly scale at Sheerness to understand how the dependence structure changes within a year. Lastly, we use a simple quantile regression technique to test if the quantile of skew surges associated with different ordered tidal bands varies.

We test if ranked peak tides associated with extreme skew surge (defined as ex-

Table A.2.1: Kolmogorov-Smirnov test p values for uniformity of ranked peak tides associated with extreme skew surges, defined exceedances of different quantiles of the data (0.9, 0.95, 0.975, 0.999). Average p values, after repeated bootstrapping, are shown in parentheses.

Site	Block size	0.9	0.95	0.975	0.99
HEY	19	0.0052 (0.012)	0.044 (0.083)	0.1813 (0.11)	0.32 (0.12)
LOW	5	0.63 (0.17)	0.72 (0.22)	0.78 (0.27)	0.33 (0.23)
NEW	20	0.070 (0.075)	0.091 (0.087)	0.15 (0.17)	0.12 (0.11)
SHE	6	1.5×10^{-8} (4.1×10^{-8})	1.1×10^{-4} (2.9×10^{-4})	9.9×10^{-4} (0.0011)	0.0013 (7×10−4)

ceedances of the 0.95 quantile) are uniformly distributed, using a Kolmogorov-Smirnov test (see Figure A.9.3). If the two components are independent, these will be distributed Uniform$(0, T)$ where T is the total number of tidal cycles. The standard version of the Kolmogorov-Smirnov test falsely assumes that peak tides are temporally independent; we use the bootstrap procedure of Politis and Romano (1994) to account for this. Average p values over 100 iterations are reported in Table A.2.1, with expected block sizes inferred from the site specific autocorrelation function (acf) plots (see Figure A.9.5). At Sheerness, we find strong evidence to reject the null hypothesis that ranked peak tides associated with extreme skew surges are uniformly distributed, with p value 2.89×10^{-3}. At the other three sites, we find sufficient evidence to reject this claim at the 5% level. We also explore the sensitivity of this test to the choice of the threshold used to define extreme skew surges, the p values are reported in Table A.2.1. Lower thresholds typically correspond to lower p values, suggesting we are more likely to reject the independence hypothesis when there are more exceedances.

We also test if the distribution of peak tides associated with extreme skew surges is the same as the distribution of all peak tides, using the Anderson-Darling test. Figure A.2.1 shows these distributions at all sites with their associated probability density functions, estimated using a Gaussian kernel density estimator. If peak tide and skew surge are independent, these two distributions should be identical up to sampling variation. We find insufficient evidence to reject the null hypothesis that these are from

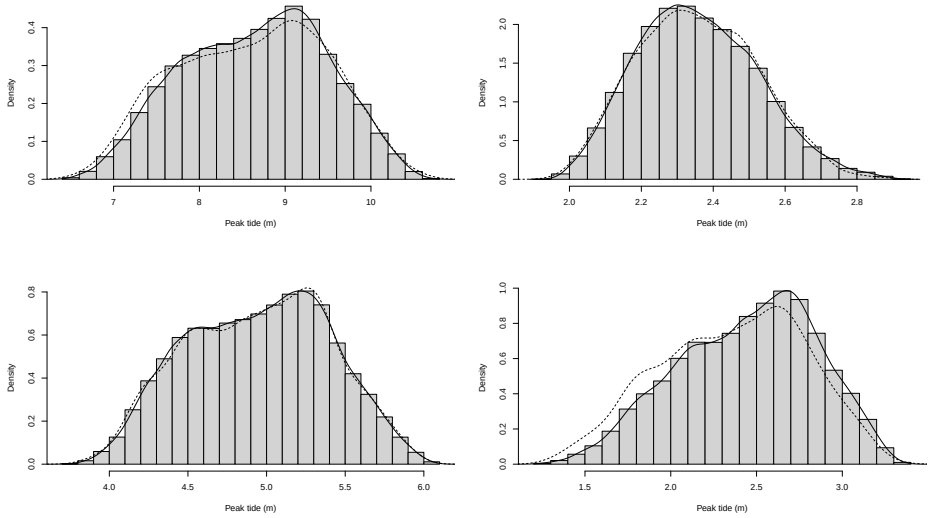


Figure A.2.1: Histogram of all peak tides at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right). Probability density function of all peak tides (solid) and peak tides associated with extreme skew surges (dashed) are interpolated onto each distribution.

the same distribution at Heysham, Lowestoft and Newlyn at the 0.01 significance level (p values are 0.014, 0.083 and 0.215, respectively). However, this is not the case at Sheerness, with p value 0.00025.

We investigate this relationship at Sheerness further by studying the dependence on a monthly basis to understand how the skew surge-peak tide dependence changes throughout the year. Again, we compare the distributions of all peak tides and extreme skew surge-related peak tides. Figure A.2.2 shows this for February, May, August and October. In May, the mode of the distribution of peak tides associated with extreme skew surges shifted to a lower value than the distribution of all peak tides. Results from the Anderson-Darling test suggest there is significant evidence, at the 1% level, to reject the null hypothesis that peak tides and the peak tides associated with extreme skew surge come from different distributions all months except February, March, September and December. Therefore, we conclude the skew surge-peak tide independence assumption is not valid for most months. When modelling the dependence of skew surge and peak tide in Section 3.4.4 of Chapter 3, we recognise that this

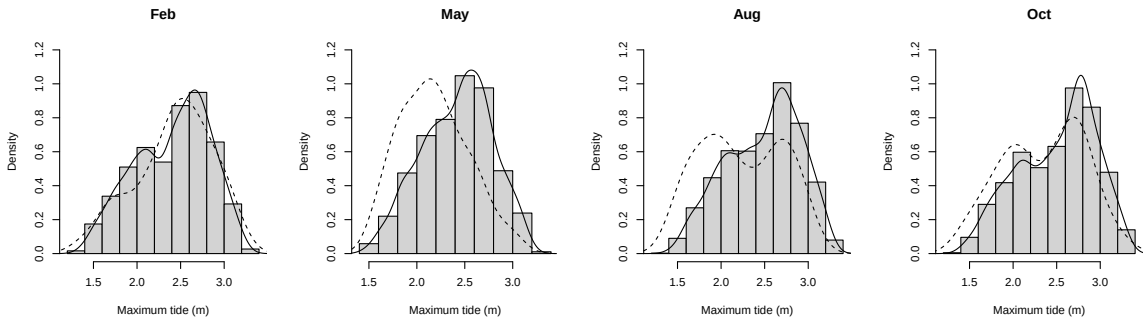


Figure A.2.2: Monthly distributions of peak tides at Sheerness in February, May, August and October. The probability density function of all peak tides (solid) and peak tides associated with extreme skew surge (dashed) are interpolated onto each distribution.

relationship is changing across the year.

To further investigate skew surge-peak tide dependence, we partition the ordered peak tide series into blocks of 100 (so that block 1 corresponds to the 100 smallest peak tide observations). For each block, we estimate the 0.95 quantile of the associated skew surges. The estimated quantiles are shown in Figure A.2.3. There is no immediate relationship between the skew surge quantile and the block number. However, when we fit a linear model, there is a significant trend at the $< 10^{-4}\%$ level at Sheerness and the 0.1% level at Newlyn, but no significant trend at Lowestoft and Heysham. Since our other tests have not found dependence at Newlyn, this relationship is likely to be a physically small relationship but statistically significant due to the length of the data series. However, at Sheerness, it is likely this is due to skew surge-peak tide dependence based on our other findings.

A.3 Physical model data

In Section A.2 here (and Section 3.3.3 of Chapter 3) we identify weak dependence between skew surge and peak tide at Sheerness; we account for this in our skew surge model. Our results show that incorporating peak tide as a covariate on the rate and scale parameter of the GPD for extreme skew surges improved the model fit. In this

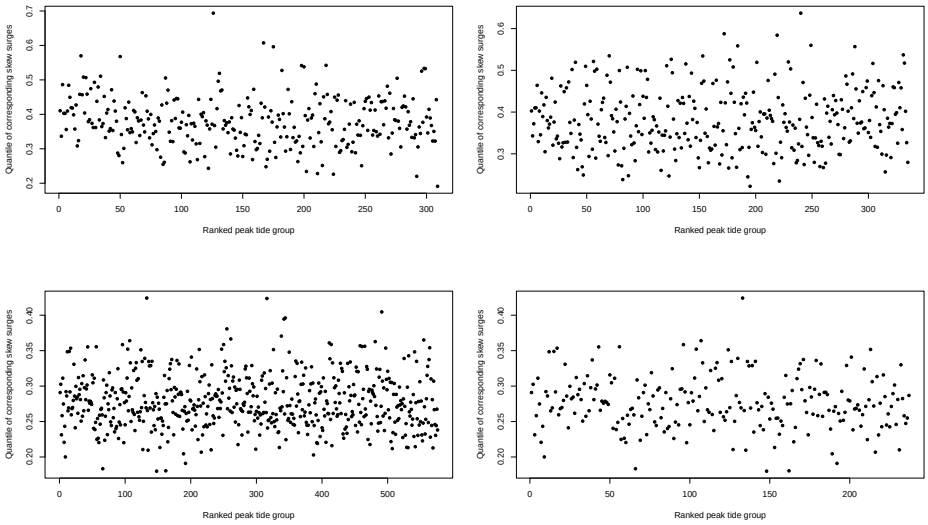


Figure A.2.3: 0.95 quantile estimates of skew surges associated with ranked peak tide groups of size 100, at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

section, we make the same comparisons on a 483 year data set from a hydrodynamical model driven by a regional climate model (HadGEM3-GC3-MM). Howard and Williams (2021) present this model to generate a dataset of 483 year present-day surges at sites on the UK National Tide Gauge Network. They use their simulations to review the skew surge-peak tide independence assumption at Sheerness, and demonstrate that extreme skew surges are more likely to occur on larger peak tides. D'Arcy and Tawn (2021) evaluate this assumption using a simplified version of the method presented in Chapter 3. For the scale parameter, we compare the models given in equations (3.4.6) and (3.4.12), and we refer to these as Models $S2$ (without a tide covariate) and $S4$ (with a tide covariate), respectively, as in Chapter 3. Models for the rate parameter are given by equations (3.4.7) and (3.4.11), we refer to these as Model $R0$ (without a tide covariate) and $R1$ (with a tide covariate), respectively.

For the physical model skew surge data, we begin by comparing models for the scale parameter only. Model $S4$ reduces both AIC and BIC by 28.75 and 21.03, respectively, relative to Model $S2$ suggesting that the extra parameter that captures variation with

peak tide tide is necessary. The likelihood ratio test strongly agrees with this, giving a significant p value of the order 10^{-8}. We estimate the tidal coefficient to be $\hat{\gamma}_\sigma = -0.013$ with 95% confidence interval (-0.018, -0.0081). This suggests that more extreme skew surges occur on lower peak tides, as found in our exploratory analysis. Since the confidence interval doesn't contain 0 the tidal coefficient is significant. This parameter estimate is reassuringly close to the corresponding estimate for the observed data of $\hat{\gamma}_\sigma = -0.012$ $(-0.026, 0.0011)$ estimated in Section 3.4.4 of Chapter 3. Whilst this suggests threshold excesses occur on lower peak tides, the confidence interval contains 0 so this result for the observed data is not statistically significant. However, this result is supported by the physical model data at Sheerness, which are based on physical reasoning with no data measurement issues and no issues with changes in the tide gauge and estuary over time.

Next, we compare models for the rate parameter on the physical model data. We find that Model $R1$ reduces AIC by 88.4 and BIC by 56.3 when compared with Model $R0$, this suggests that adding a peak tidal covariate to the rate parameter is important. Therefore, we make the same conclusions on 483 years of physical model data, as we do on 37 years of observed data. This shows the rate model parameterisation is supported empirically and physically. When we cannot reasonably assume skew surge and peak tide are independent, accounting for their dependence is important as it can lead to practical differences in the sea level return level estimates, as shown in Section 3.5.3 of Chapter 3.

A.4 Temporal skew surge dependence

Here we provide further details on skew surge temporal dependence discussed in Section 3.3.2 of Chapter 3. We study the two key measures of extremal dependence, these are a measure of asymptotic dependence χ and of asymptotic independence $\bar{\chi}$ for each

site. For random variables Y_i and $Y_{i+\tau}$ separated by lag τ from a stationary sequence, with y^F the upper endpoint of their distribution F, Coles et al. (1999) define

$$\chi_\tau = \lim_{y \to y^F} \Pr(Y_{i+\tau} > y | Y_i > y),$$

as a measure of asymptotic dependence where $\chi_\tau \in [0, 1]$. If $\chi_\tau \in (0, 1]$, we say that Y_i and $Y_{i+\tau}$ are asymptotically dependent; this means there is non-zero probability of $Y_{i+\tau}$ being large when $Y_{i+\tau}$ is large at all extreme levels. Whereas $\chi_\tau = 1$ and $\chi_\tau = 0$ correspond to perfect dependence and asymptotic independence, respectively. Therefore, χ_τ fails to signify the level of asymptotic independence, so Coles et al. (1999) also define the measure $\bar{\chi}_\tau$ as

$$\bar{\chi}_\tau = \lim_{y \to y^F} \frac{2 \log \Pr(Y_i > y)}{\log \Pr(Y_i > y, Y_{i+\tau} > y)} - 1,$$

where $\bar{\chi}_\tau \in (-1, 1]$. Asymptotic dependence and asymptotic independence correspond to $\bar{\chi} = 1$ and $\bar{\chi} < 1$, respectively, whilst $0 < \bar{\chi}_\tau < 1$ and $-1 < \bar{\chi}_\tau < 0$ correspond to positive and negative association, respectively, and $\bar{\chi}_\tau = 0$ corresponds to near independence. We evaluate both measures with y at different quantiles of the distribution in Figure A.4.1 for Heysham, and Figures A.9.6 and A.9.7 for the remaining sites.

Firstly, we use estimates of χ and $\bar{\chi}$ to choose a high threshold y to empirically estimate the extremal index θ using the runs method. We are interested in choosing a threshold where these measures tend to zero without significant noise; the 0.95 quantile is sufficient for each site. We explore the sensitivity of our estimate to this threshold choice in Figure A.4.2; as we increase the threshold, the estimate of θ increases. This is what we expect since exceedances of lower thresholds are likely to exhibit more dependence than exceedances of higher thresholds. We also explore the sensitivity to run length r in Figure A.4.2. Estimates of the extremal index are generally higher at Newlyn and Sheerness, which agrees with the acf values in Figure A.9.5. The choice of

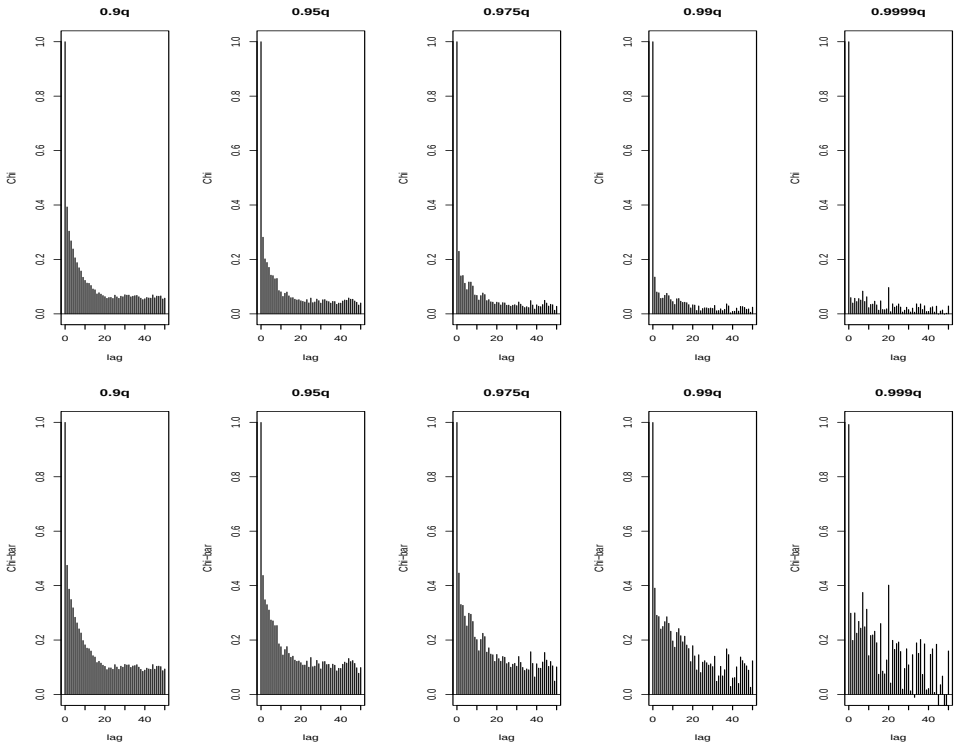


Figure A.4.1: Estimates of χ (top row) and $\bar{\chi}$ (bottom row) for Heysham for exceedances of the 0.9, 0.95, 0.975, 0.99 and 0.999 quantiles (from left to right column) at various lags.

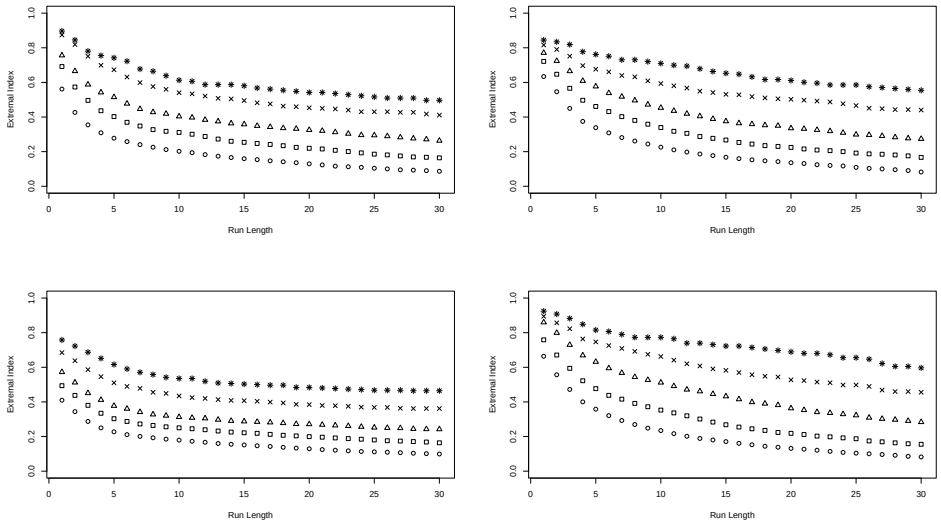


Figure A.4.2: Empirical (runs) estimates of the extremal index θ for skew surge, at various quantiles and run lengths at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right). The estimates use thresholds which are taken to be quantiles 0.9 (circle), 0.95 (square), 0.975 (triangle), 0.99 (cross) and 0.999 (star).

run length and threshold level are important because they have a significant influence on the θ estimate, so we choose these carefully.

Since our empirical estimates are sensitive to the threshold level, we develop a parametric model for the subasymptotic extremal index $\hat{\theta}(y, r)$, that is dependent on the skew surge level y; this is given in equation (3.4.16) of Chapter 3. Figure 3.4.3 of Chapter 3 shows a consistent model fit with the empiricals at Heysham, with parameter estimates $\hat{\theta} = 1$ and $\hat{\psi} = 0.33$ for $r = 2$. Figure A.9.8 shows the fit at the remaining sites and these also match closely with the empiricals. The parameter estimates are $\hat{\theta} = 1, 0.95, 0.89$ and $\hat{\psi} = 0.42, 0.17, 0.21$ for $r = 10, 1, 10$ at Lowestoft, Newlyn and Sheerness, respectively. As expected, $\hat{\theta}$ lies close to 1 in all cases, this represents the limiting extremal index and the case of independence. The estimate of ψ tells us about the speed of convergence to $\hat{\theta}$, so that Newlyn converges fastest whilst Lowestoft is the slowest.

A.5 Threshold sensitivity

In this section, we explore the sensitivity of the sea level return level estimates derived from the proposed model (3.4.15) to the choice of threshold u for the skew surge model (3.4.14); above this threshold we fit a non-stationary GPD and the empirical distribution is used below. Recall that our modelling approach uses a time-varying threshold, so we have a different threshold u_j for each month $j = 1, \ldots, 12$. In Chapter 3, we use the 0.95 quantile of monthly skew surges, here we compare these results with using the 0.9 and 0.99 quantile to define extreme skew surges.

Figure A.5.1 shows the return level estimates from the proposed model where the 0.9, 0.95 and 0.99 quantiles are used to define the extreme values at Sheerness. The 95% confidence intervals are the same as those in Chapter 3 where the 0.95 quantile is used, both before and after we add a prior distribution (see Section 3.5.4 of Chapter 3 for details). We can see that the estimates for the method using the 0.9 and 0.99 quantile of skew surges lie within the confidence intervals for the original estimates (0.95 quantile). The estimated values from these two new thresholds are much closer to the point estimate for the 0.95 quantile threshold than the endpoints of the associated 95% confidence intervals, suggesting that the impact of threshold uncertainty is small compared to other sources of uncertainty in the modelling framework. This is the case for the model with and without the prior distribution on the shape parameter of the GPD, where the uncertainty for the former is much smaller but still larger than the uncertainty attributable to threshold selection. We find similar results at the remaining sites, see Figure A.9.21.

A.6 Simulation study

So far, we have compared models when fit to observed data at the tide gauges therefore comparison with the empirical estimates are restricted to the length of observed data.

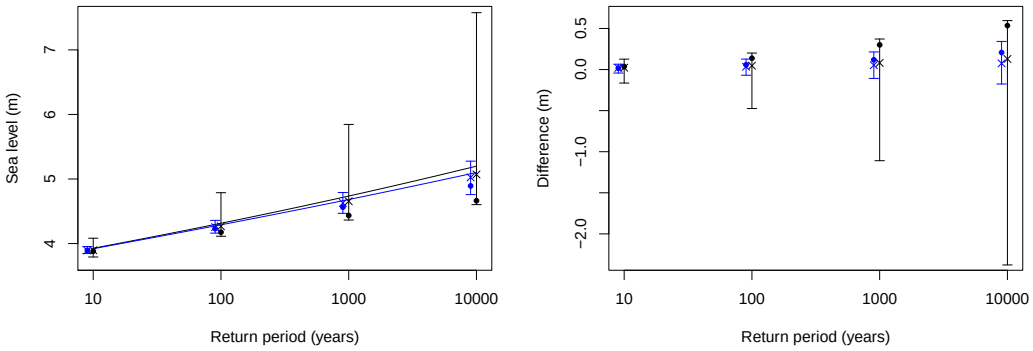


Figure A.5.1: Return level estimates from the final *(temporal dependence)* model with 95% confidence intervals before (black) and after (blue) adding a prior distribution to the shape parameter at Sheerness, where the 0.95 quantile is used for the skew surge distribution (left); estimates are also shown at the 10, 100, 1000 and 10,000 year levels when the 0.9 (crosses) and 0.99 (dots) quantiles are used in the skew surge model, both with (blue) and without (black) the prior distribution on the shape parameter of the GPD. The corresponding difference between the estimates from the original approach (0.95 quantile) with the 0.9 (crosses) and 0.99 (dots) quantiles, compared with confidence intervals (right).

Here, we use samples from the 483 year physical model data at Sheerness (introduced in Section A.3) to fit our proposed model (given in expression (3.4.15) of Chapter 3, without a prior on the shape parameter) and the SSJPM of Batstone et al. (2013), and compare these with empirical estimates. Howard and Williams (2021) introduce the data, which are generated from a widely-used physical model using a long-run hydrodynamical model driven by a long-run climate model (HadGEM3-GC3-MM) without any drivers for climate change. This gives us an entirely realistic dataset as the associated hydrodynamical model has been calibrated to give realistic skew surges and peak tides, whilst the climate model ensures features such as seasonality and temporal dependence are captured.

From the 483 year dataset, we randomly select 30 samples of 37 continuous years so that they approximately match the observed length of data at Sheerness. Note we say approximately here because the physical model data benefit from no missing data, whilst the observed tide gauge data has 9% missing at Sheerness. Additionally,

months are simplified to 12 months of 30 days each in the physical model data, so years are only 360 days long. For each sample, we fit the SSJPM and our proposed model. The latter averages over the annual maxima distribution for each specific year in the data to capture interannual tidal variations. When fitting this to a sample of physical model data, we average over all 483 years of tidal data (which can be predicted from the observed 37 years of sea level data) so that each year's annual tidal regime is captured, but the skew surge model parameters are estimated from the 37 year sample data. This ensures that longer-term tidal variations are accounted for and the results are more comparable with empirical estimates. Since the SSJPM assumes peak tides are stationary and only uses tidal information from a single nodal cycle, we only use the tides in the 37 year sample, rather than all 483 years of tides, to better represent their approach and to help show that the major improvements come from our modelling of seasonality. We then estimate sea level return levels corresponding to return periods between 1 and 100 years, and compare them to the empirical estimates; these are shown in Figure A.6.1, averaged over each sample for each return period. The average estimates from our model lie much closer to the empirical than the SSJPM, especially for lower return periods where they match almost perfectly. At the 100 year level, our estimate still lies closer but the empirical estimates are less reliable, even though they are taken from 483 years of data. At the 1, 10 and 100 year level, the RMSE over all samples for our model is 0.03m, 0.05m and 0.14m, respectively. For the SSJPM the RMSE values are greater by factors of 1.54, 1.29 and 1.26, respectively. Notice the difference between estimates from the two methods is greater for lower return periods, with our method always being superior. A summary of the 1, 10 and 100 year return level estimates is given in Table A.6.1 for both methods; the SSJPM has a greater bias and variance over the 30 samples than our method. Note that our method slightly underestimates at the 1 and 10 year return levels, whilst the SSJPM always overestimates and by a greater magnitude.

Table A.6.1: Sample standard deviation (SD), bias and RMSE (in cm) of the 1, 10 and 100 year sea level return level estimates from 30 samples of 37 years from the physical model data using our model and the SSJPM (Batstone et al., 2013). Here we take the truth as empirical estimates based on the 483 years of physical model data.

	Our model			SSJPM		
Return period	SD	Bias	RMSE	SD	Bias	RMSE
1	2.4	-2.5	3.4	2.1	4.8	5.3
10	4.8	-0.4	4.7	5.3	3.2	6.1
100	8.6	11.3	14.2	11.8	13.6	17.9

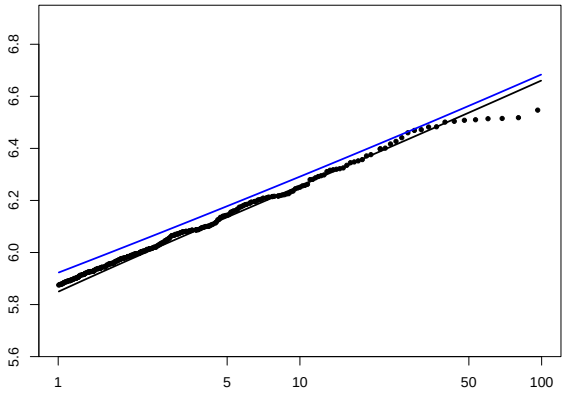


Figure A.6.1: Return level estimates (y-axis in metres) for different return periods (x-axis in years) from the final *(temporal dependence)* model (black) and the SSJPM (blue) (Batstone et al., 2013), averaged over 30 samples of 37 years from the physical model data (Howard and Williams, 2021). Empirical estimates from all 483 years of physical model data are shown by black points.

A.7 Derivation of expression for seasonal probability

We are interested in deriving an analytical expression for the probability $\hat{P}_M(j;z)$ that a randomly selected sea level (annual maxima) M is from month j given it equal to some level z, given by,

$$\hat{P}_M(j;z) = \hat{\Pr}(m(M) = j | M = z),$$

where $m(M)$ denotes the month of occurrence of the variable M and $\hat{\Pr}(\cdot)$ is under our final model of Section 3.4.5 in Chapter 3. This is probability (3.5.2) in Chapter 3 and we use this in Section 3.5.5 to evaluate extreme sea level seasonality. We are specifically interested when $z = z_p$, a level with an associated annual exceedance probability $p \in [0,1]$, so that z is a return level derived from expression (3.4.15).

As the distribution of M varies with year k due to the tidal variations, we begin by conditioning on a fixed year k, so that we only look at sea levels within a specific year and consider the probability $\hat{P}_{M^{(k)}}(j;z)$. We rewrite this in terms of the distribution and density of the month j maxima sea level, $F_{M_j^{(k)}}$ and $f_{M_j^{(k)}}$, respectively, and the density of the annual maxima sea level $f_{M^{(k)}}$, each conditional on some year $k = 1, \ldots, K$, where K is the total number of years of observation,

$$\begin{aligned}\hat{P}_{M^{(k)}}(j;z) &= \frac{f_{M_j^{(k)}}(z) \prod\limits_{\substack{J=1,\ldots,12 \\ J \neq j}} \Pr(M_J^{(k)} < z)}{f_{M^{(k)}}(z)} \\ &= \frac{1}{f_{M^{(k)}}(z)} \left[\left(\frac{f_{M_j^{(k)}}(z)}{F_{M_j^{(k)}}(z)} \right) F_{M^{(k)}}(z) \right].\end{aligned}$$

Therefore, it follows to find an expression for each of these terms to simplify the above

expression. To find the form of the density of the monthly maxima sea levels in a given year $f_{M_j^{(k)}}$, we must differentiate the corresponding distribution function,

$$F_{M_j^{(k)}}(z) = \Pr(M_j^{(k)} \leq z) = \prod_{i=1}^{T_j^{(k)}} \left[F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})\right]^{\hat{\theta}(z-X_{j_i}^{(k)},r)},$$

where $F_Y^{(d,j,x)}$ is the skew surge distribution function given by equation (3.4.14) and $\hat{\theta}(\cdot, r)$ is our extremal index model given by equation (3.4.16) for fixed run length r. Differentiating this gives

$$f_{M_j^{(k)}}(z) \approx F_{M_j^{(k)}}(z) \sum_{i=1}^{T_j^{(k)}} \frac{{f_Y}^{(d,j,x)}(z - X_{j_i}^{(k)})\hat{\theta}(z - X_{j_i}^{(k)}, r)}{F_Y^{(d,j)}(z - X_{j_i}^{(k)})}, \tag{A.7.1}$$

when we ignore smaller order terms arising from $\hat{\theta}'(\cdot, r)$. The density of skew surges $f_Y^{(d,j,x)}$ is given by

$$f_Y^{(d,j,x)}(y) = \begin{cases} \hat{f}_j(y) & \text{if } y \leq u_j \\ \frac{\lambda_{d,x}}{\sigma_{d,x}}\left[1 + \xi\left(\frac{y-u_j}{\sigma_{d,x}}\right)\right]_+^{-\frac{1}{\xi}-1} & \text{if } y > u_j, \end{cases}$$

where $\lambda_{d,x}$ and $\sigma_{d,x}$ are given by expressions (3.4.11) and (3.4.12), respectively. Below the monthly threshold u_j, we estimate the derivative of the monthly empirical distribution using a Gaussian kernel density estimator and denote this $\hat{f}_j$. We find the density of the annual maximum sea levels, for a fixed year k, by differentiating in similar way and ignoring smaller order terms, so that,

$$f_{M^{(k)}}(z) \approx F_{M^{(k)}}(z) \sum_{J=1}^{12} \sum_{i=1}^{T_J^{(k)}} \frac{f_Y^{(d,J,x)}(z - X_{J_i}^{(k)})\hat{\theta}(z - X_{J_i}^{(k)}, r)}{F_Y^{(d,J,x)}(z - X_{J_i}^{(k)})}. \tag{A.7.2}$$

Using equations (A.7.1) and (A.7.2), we can simplify $\hat{P}_{M^{(k)}}(j;z)$ to

$$\hat{P}_{M^{(k)}}(j;z) = \frac{A^{(i,j,k)}(z)}{\sum_{J=1}^{12} A^{(i,J,k)}(z)} \quad \text{where} \quad A^{(i,j,k)}(z) = \sum_{i=1}^{T_j^{(k)}} \frac{f_Y^{(d,j,x)}(z - X_{j_i}^{(k)})\hat{\theta}(z - X_{j_i}^{(k)}, r)}{F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})}.$$

Clearly $\sum_{j=1}^{12} \hat{P}_{M^{(k)}}(j;z) = 1$. Then the probability over all K years is given by $\hat{P}_M(j;z) = \frac{1}{K}\sum_{k=1}^{K} \hat{P}_{M^{(k)}}(j;z)$.

A.8 Transforming skew surges to uniform margins

In Section 3.5.4 of Chapter 3 we assess the fit of our final model for sea level annual maxima (3.4.15) by looking at year-specific distributions and bootstrap confidence intervals on the return level estimates. Here, we test the goodness-of-fit for the final skew surge model (expression (3.4.14)), using the probability integral transform; if our model fits well, transforming the observations through the fitted distribution function will give a sample of identically distributed Uniform$(0,1)$ values. This was also a step in our stationary bootstrap procedure, but here we check if these transformed values $\{U_i^Y\}$ are uniform using a Kolmogorov-Smirnov test.

Figure A.8.1 shows the transformed skew surges at Sheerness. We can immediately see these are not uniformly distributed. This is supported by results of yearly Kolmogorov-Smirnov tests for uniformity, where the p values are almost all > 0.05. Instead, there appears to be a cyclic sinusoidal trend, following the trend of the 18.6 year nodal cycle. We suspect this is a data issue, perhaps the tidal series was not correctly removed from the sea level observations when the skew surges were obtained. Figure A.8.1 also shows the annual mean skew surges at Sheerness, we can see there is a similar cyclic trend here and that the means are not centred at zero (see Figure A.9.16 for the remaining sites). To correct for this we re-centre the data at zero, by removing the corresponding annual mean from the observations. This is an ad hoc approach

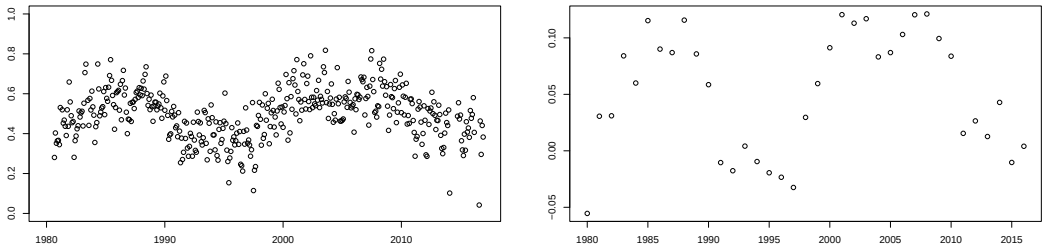


Figure A.8.1: Monthly means (y-axis in meters) of transformed skew surge observations through the final skew surge distribution function (4.14) (left) and annual mean skew surges (right), both at Sheerness, against the year of observation (x-axis).

and the data should be investigated further. We did not correct for this trend at an earlier stage in the modelling process because it does not have a significant effect on the extreme values. However, once we correct for this trend, we find that the transformed data are uniformly distributed in 32 years (out of 37) at Sheerness, indicating a good model fit for skew surges. We find similar results at the remaining sites, where a cyclic trend is first observed in $\{U_i^Y\}$ but once the annual means are removed, the transformed data can be reasonably assumed as Uniform(0,1). Figure A.9.17 shows the p values for each year at each site.

A.9 Supplementary figures

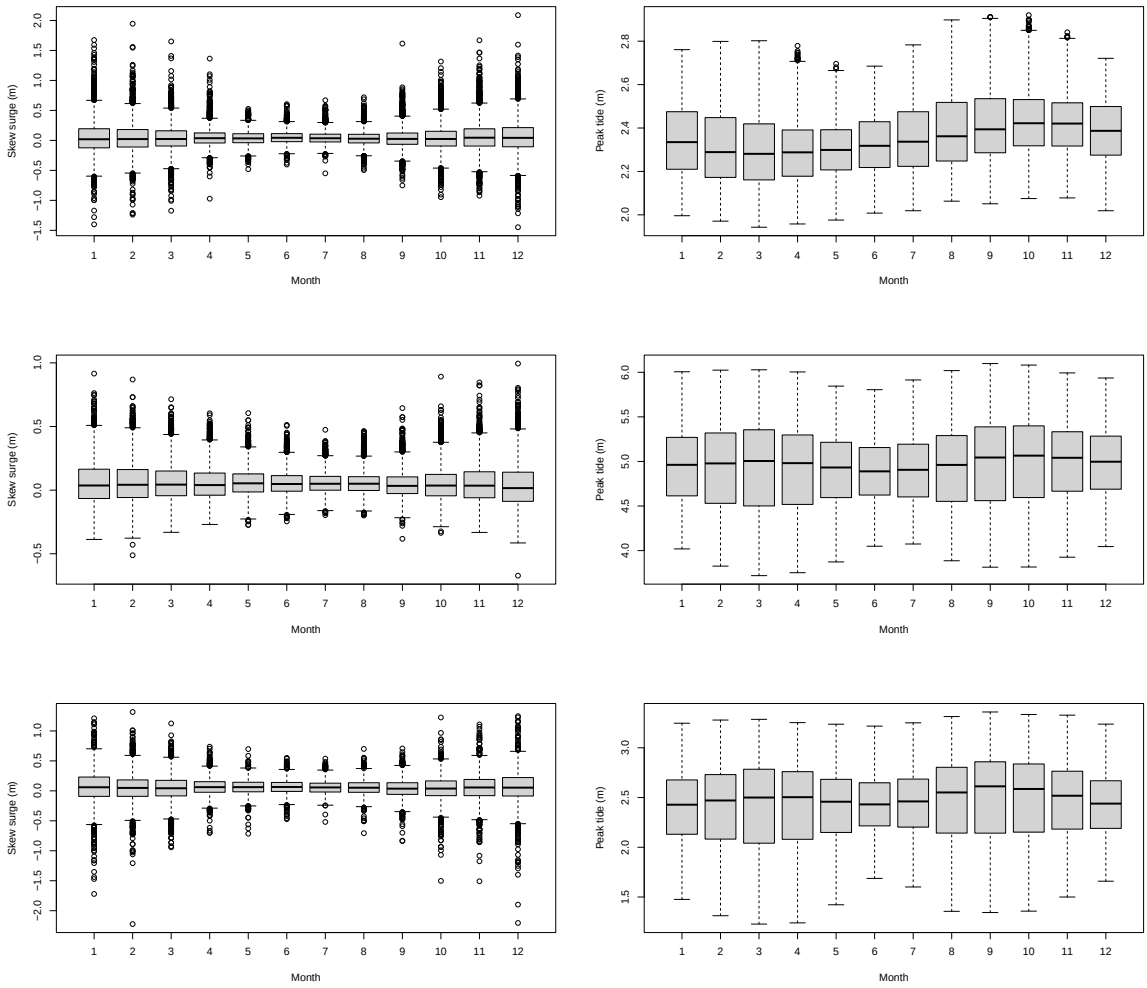


Figure A.9.1: Monthly box plots of skew surge (left column) and peak tide (right column) at Lowestoft (top row), Newlyn (middle row) and Sheerness (bottom row).

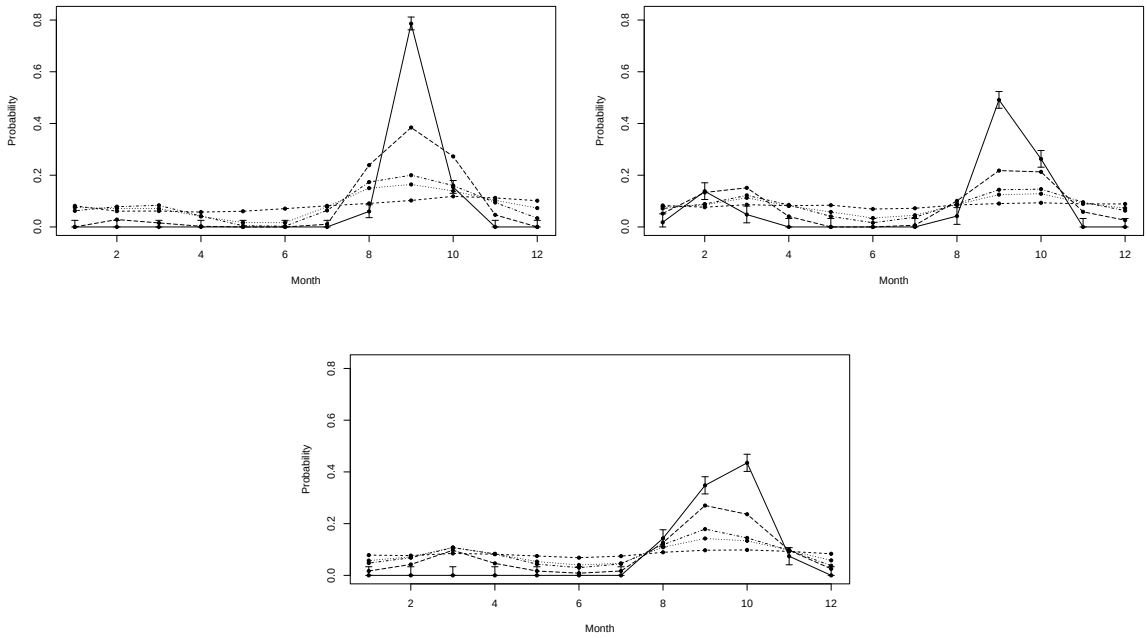


Figure A.9.2: Estimates of $\tilde{P}_X(j; x_q)$ (expression 3.3.2 of Chapter 3) for months $j = 1 - 12$ and $q = 0.5$ (dashed), 0.9 (dotted), 0.95 (dot-dashed), 0.99 (long-dashed), 0.999 (solid) at Lowestoft (top left), Newlyn (top right) and Sheerness (bottom), with 95% confidence intervals when $q = 0.99$.

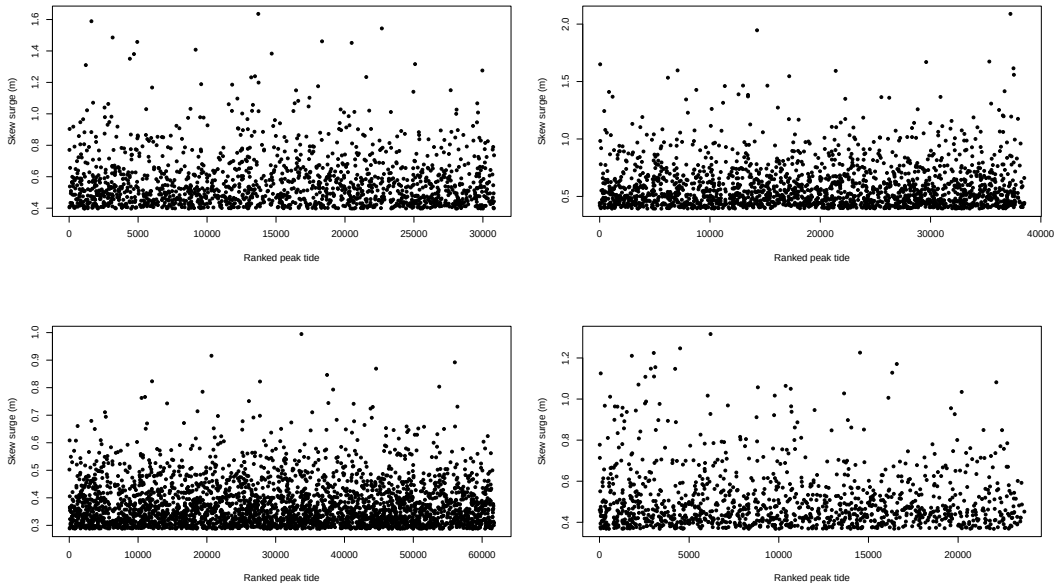


Figure A.9.3: Scatter plot of extreme skew surge observations (exceedances of 0.95 quantile) against associated ranked peak tides at Heysham (top right), Lowestoft (top left), Newlyn (bottom left) and Sheerness (bottom right).

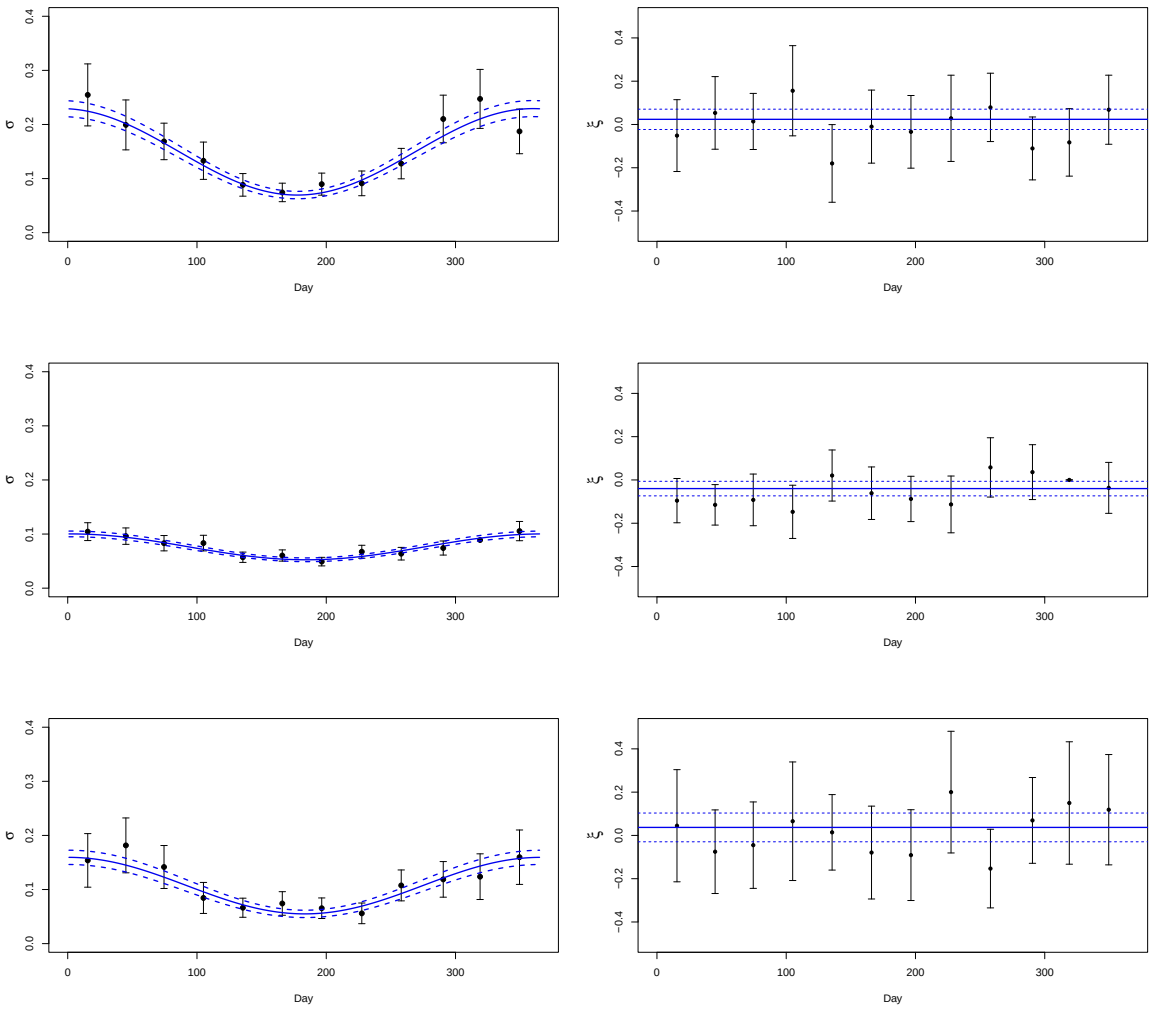


Figure A.9.4: Scale (left column) and shape (right column) parameter estimates for Model $S2$ (blue) and Model $S0$ (black) at Lowestoft (top row), Newlyn (middle row) and Sheerness (bottom row). 95% confidence intervals are added to Model $S0$ (black error bars) and to Model $S2$ parameter estimates (blue dashed lines).

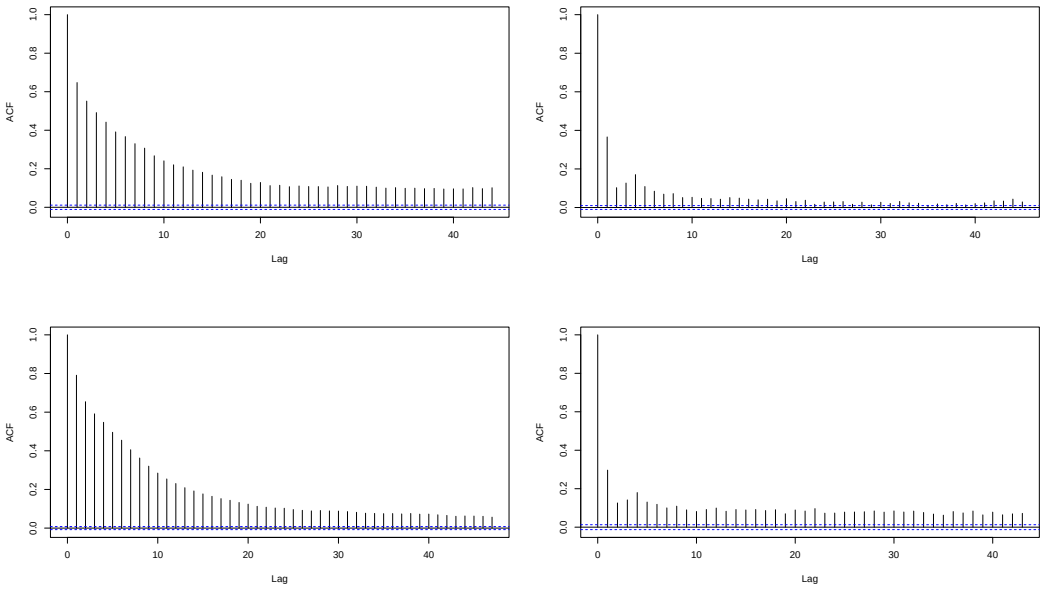


Figure A.9.5: Autocorrelation function (acf) plots for Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

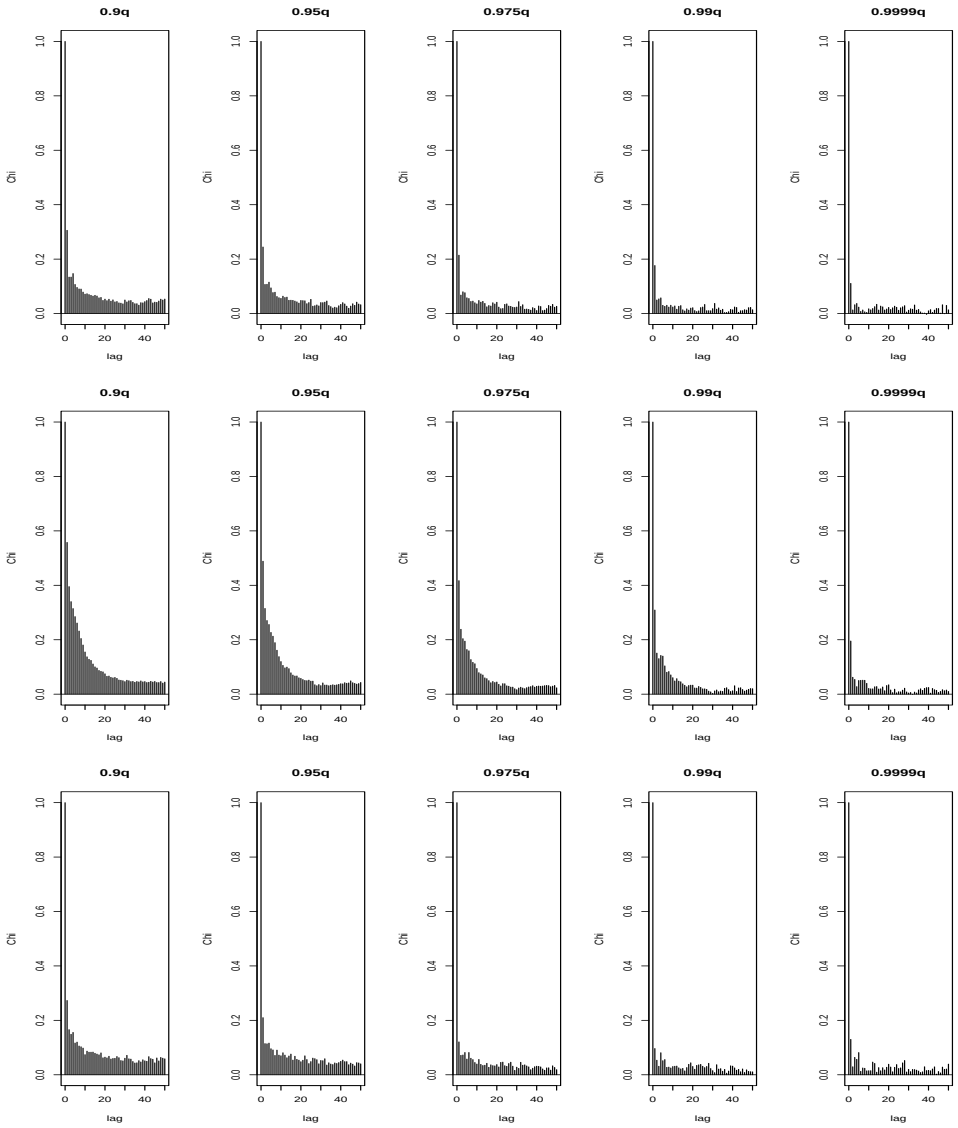


Figure A.9.6: Estimates of χ for Lowestoft, Newlyn and Sheerness (from top to bottom row) for exceedances of the 0.9, 0.95, 0.975, 0.99 and 0.999 quantiles (from left to right column) at various lags.

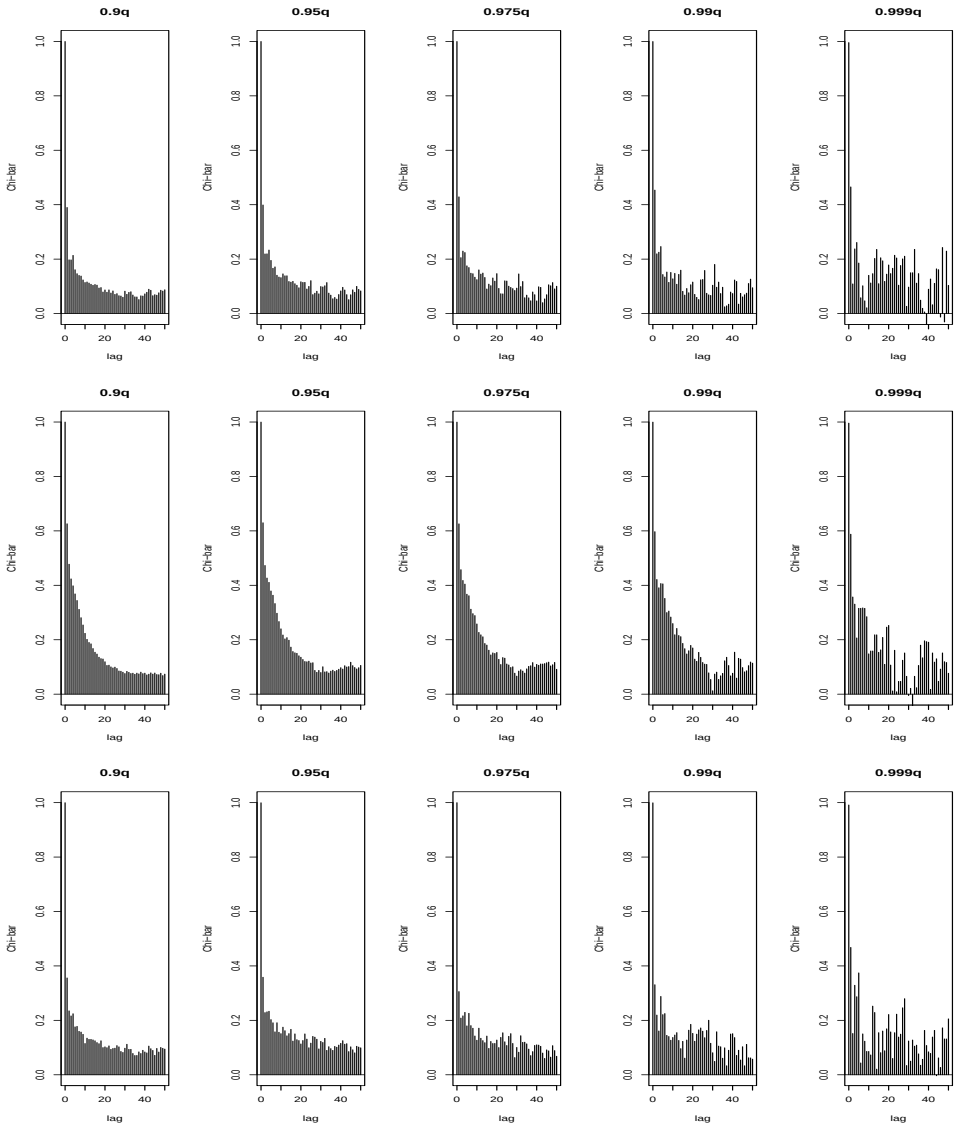


Figure A.9.7: Estimates of $\bar{\chi}$ for Lowestoft, Newlyn and Sheerness (from top to bottom row) for exceedances of the 0.9, 0.95, 0.975, 0.99 and 0.999 quantiles (from left to right column) at various lags.

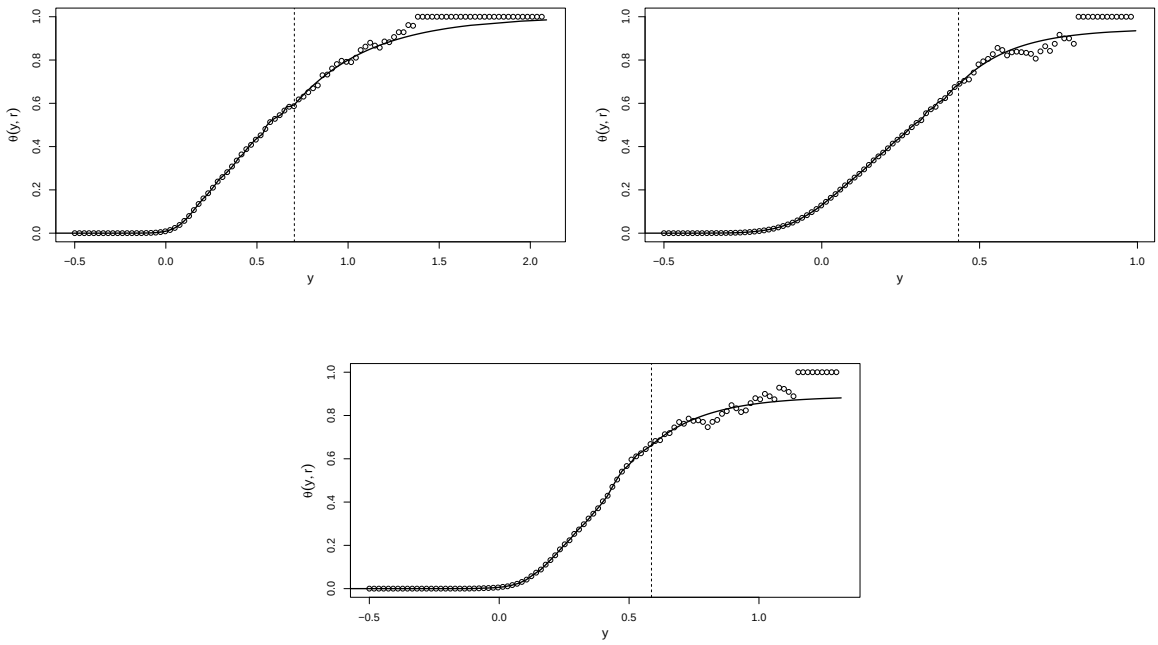


Figure A.9.8: Estimates of the subasymptotic extremal index $\theta(y, r)$ for different skew surge levels using the runs estimate (grey points) and our model estimate (black line) (expression (3.4.16) of Chapter 3) at Lowestoft (top left), Newlyn (top right) and Sheerness (bottom). Run lengths are chosen as 10, 2 and 10, respectively. The threshold v is chosen as the 0.99 skew surge quantile for all sites (black dashed line).

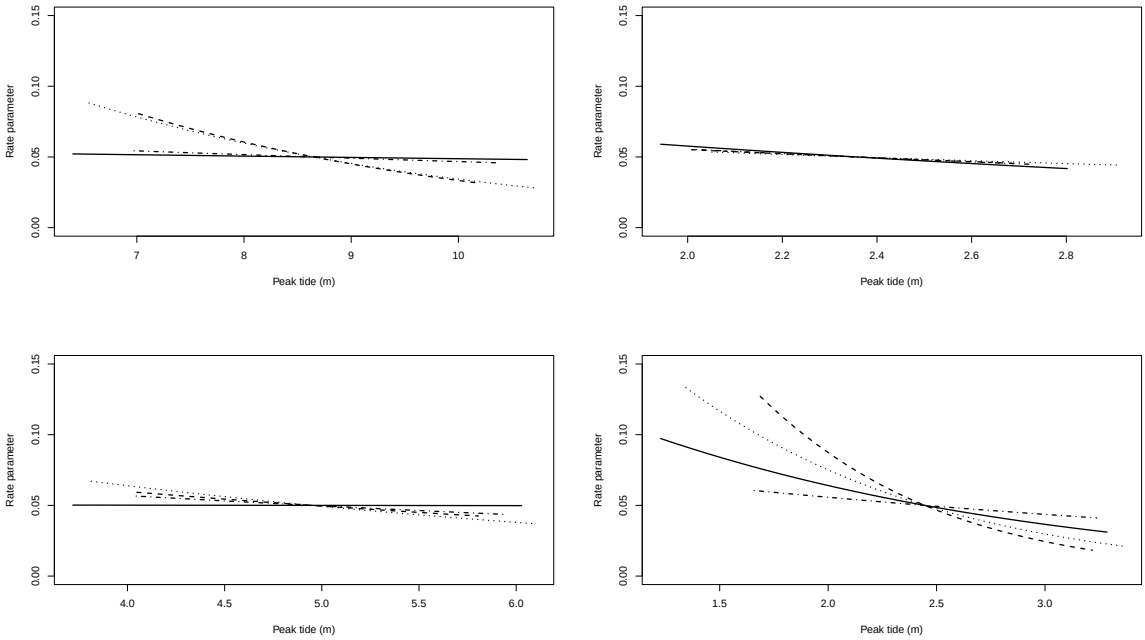


Figure A.9.9: Estimated exceedance probability $\lambda_{d,x}$ (expression (4.11)) per month with respect to peak tide x and day in month d_j, averaged over day at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right). Trends for March (solid), June (dashed), September (dotted) and December (dot-dashed) are shown here.

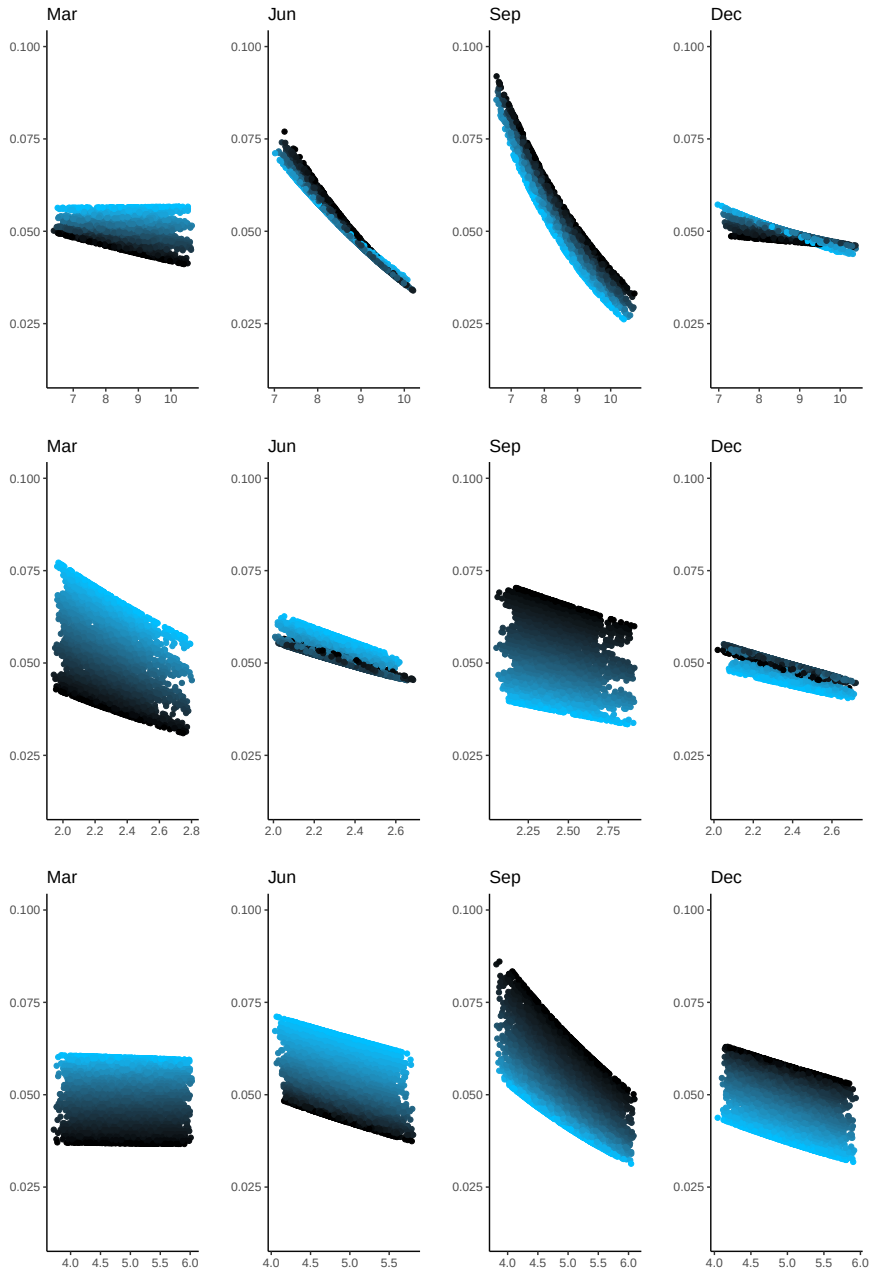


Figure A.9.10: Estimated exceedance probability $\lambda_{d,x}$ (y-axis) at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row), in March, June, September and December (from left to right by column) with respect to x being peak tide (x-axis in metres) and d_j being day in month at Sheerness. Darker (lighter) points represent days later (earlier) in the month.

Table A.9.1: Parameter estimates for the scale parameter Models $S2$ and $S4$, and the rate parameter Model $R1$ with 95% confidence intervals, at each site.

	Heysham	Lowestoft	Newlyn	Sheerness
Model $S2$				
α_σ	0.14 (0.13, 0.15)	0.15 (0.14, 0.16)	0.076 (0.073, 0.080)	0.11 (0.10, 0.12)
β_σ	0.060 (0.050, 0.070)	0.080 (0.070, 0.090)	0.024 (0.020, 0.028)	0.052 (0.043, 0.061)
ϕ_σ	271.51 (262.77, 280.23)	266.32 (260.01, 272.63)	273.58 (265.06, 282.10)	272.11 (262.66, 281.56)
ξ	0.002 (-0.042, 0.051)	0.024 (-0.023, 0.071)	-0.040 (-0.074, 0.006)	0.037 (-0.029, 0.10)
Model $S4$				
α_σ	0.13 (0.063, 0.19)	0.16 (0.078, 0.24)	0.053 (0.026, 0.79)	0.14 (0.074, 0.14)
β_σ	0.060 (0.049, 0.071)	0.080 (0.070, 0.090)	0.024 (0.020, 0.027)	0.053 (0.043, 0.061)
ϕ_σ	272.20 (263.31, 281.09)	266.10 (259.65, 272.56)	278.97 (270.32, 287.61)	271.37 (262.91, 281.32)
$\gamma_\sigma^{(x)}$	0.002 (-0.005, 0.009)	-0.0051 (-0.040, 0.030)	0.0048 (-0.00048, 0.010)	-0.012 (-0.026, 0.0011)
ξ	0.0049 (-0.044, 0.054)	0.024 (-0.023, 0.071)	-0.037 (-0.071, -0.003)	0.033 (-0.033, 0.099)
Model $S4$ with prior on shape				
ξ	0.019 (-0.021, 0.059)	0.014 (-0.024, 0.052)	-0.027 (-0.058, 0.004)	0.008 (-0.039, 0.054)
Model $R1$				
β_λ	0.0087 (0.0004, 0.017)	0.022 (0.015, 0.030)	0.024 (0.018, 0.030)	0.022 (0.014,0.032)
ϕ_λ	155.66 (100.74, 210.59)	175.16 (155.86, 194.46)	209.50 (195.48, 223.52)	184.31 (160.94,207.69)
$\alpha_\lambda^{(x)}$	-0.13 (-0.18, -0.079)	-0.055 (-0.101, 0.009)	-0.063 (-0.099, 0.108)	-0.32 (-0.37,-0.26)
$\beta_\lambda^{(x)}$	0.14 (0.068, 0.21)	-0.016 (-0.084, 0.051)	0.061 (0.014, 0.108)	0.23 (0.14, 0.31)
$\phi_\lambda^{(x)}$	311.86 (281.78, 341.94)	359.95 (265.77, 454.15)	352.38 (299.32, 405.44)	278.54 (260.44, 293.63)

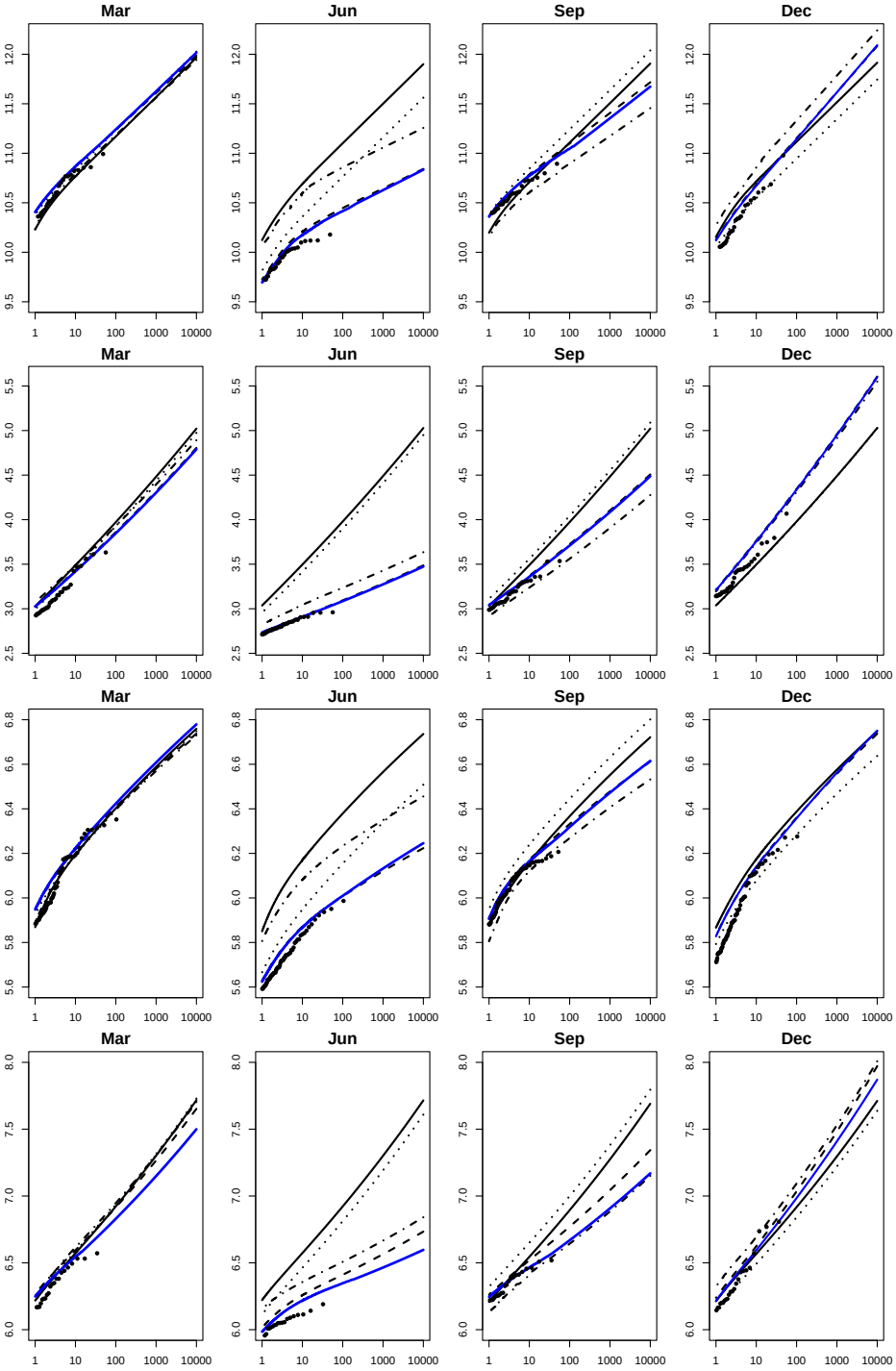


Figure A.9.11: Monthly maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham, Lowestoft, Newlyn and Sheerness (from top to bottom row) in March, June, September and December (from left to right by column) estimated using the *baseline* (black solid), *seasonal surge* (dot-dashed), *seasonal tide* (dotted), *full seasonal* (dashed) and *interaction* (blue solid) models. Empirical estimates are shown by black points.

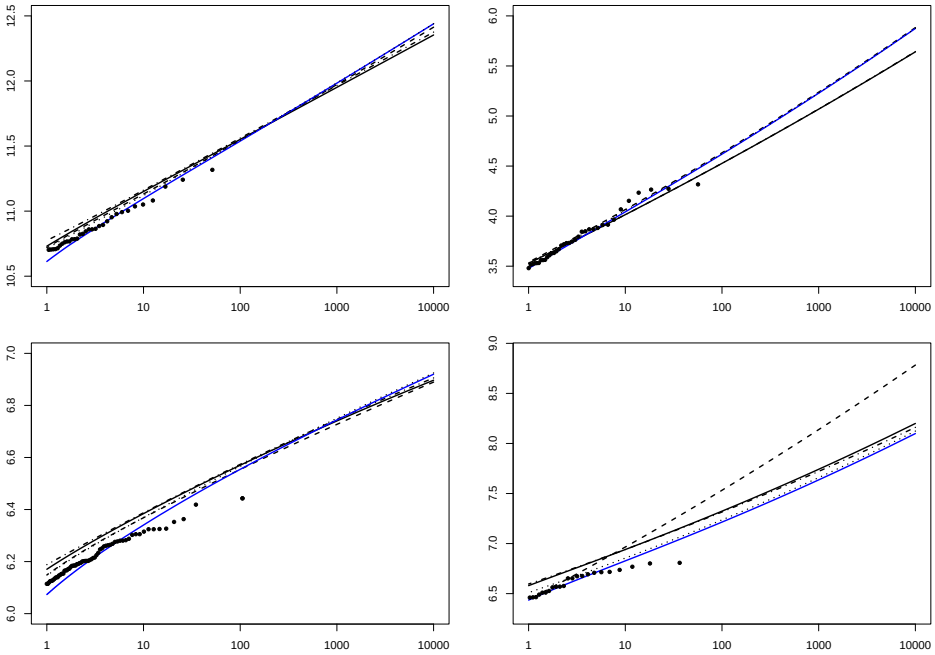


Figure A.9.12: Annual maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham (top left), Lowestoft (top right), Sheerness (bottom left) and Newlyn (bottom right), estimated using the *current* (dot-dashed), *baseline* (black solid), *full seasonal* (dashed), *interaction* (dotted) and *temporal dependence* (blue solid) methods. Empirical estimates are shown by black points.

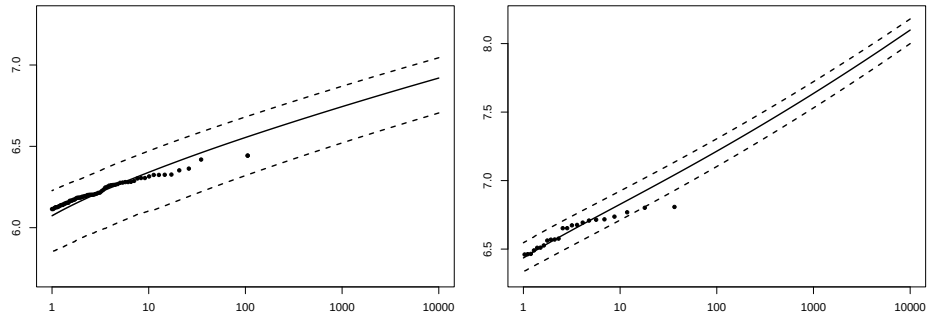


Figure A.9.13: Return level estimates (y-axis in metres) for different return periods (x-axis in years) from the final model (*temporal dependence*) (solid), with the maximum and minimum year-specific return level estimates (dashed) and empirical estimates (black) at Newlyn (left) and Sheerness (right).

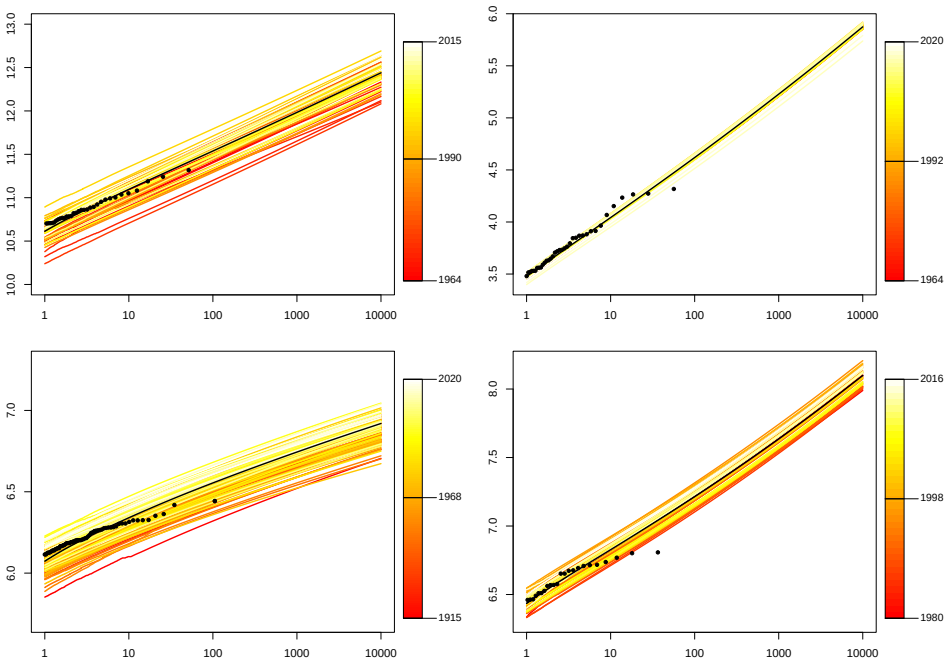


Figure A.9.14: Annual maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right) estimated using the final model (*temporal dependence*) shown by the solid black line, with year-specific return levels shown by the red-yellow lines. Empirical estimates are shown by black points.

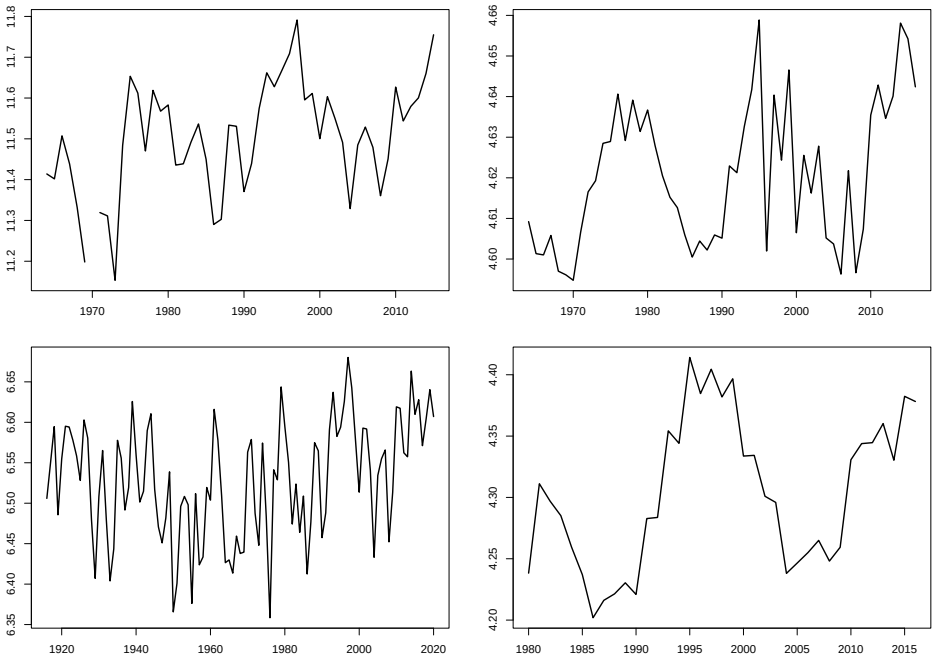


Figure A.9.15: 100 year return level estimates (y-axis in meters) for Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right) estimated using the year-specific final model over the years of observation (x-axis).

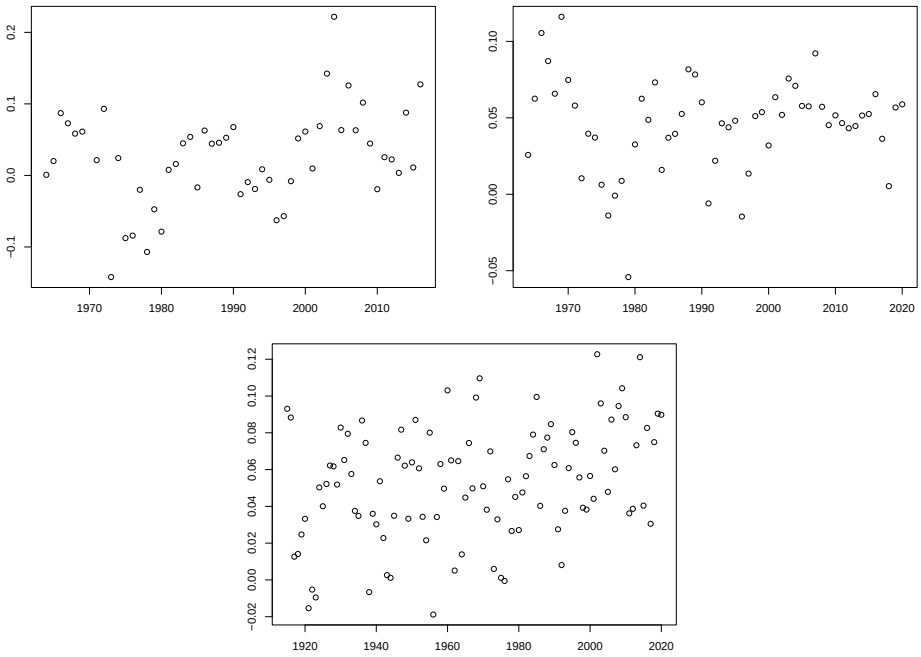


Figure A.9.16: Annual mean skew surges (y-axis in meters) against year (x-axis) at Heysham (top left), Lowestoft (top right) and Newlyn (bottom).

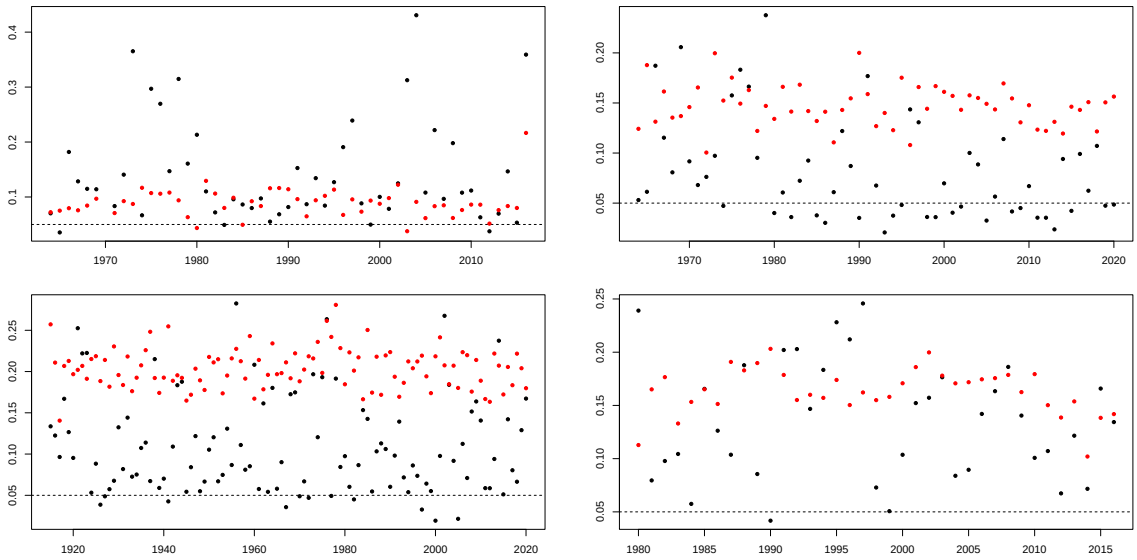


Figure A.9.17: p values from the Kolmogorov-Smirnov test for uniformity (y-axis) of yearly (x-axis) samples of the transformed skew surge observations through the final skew surge distribution function (expression (4.14)) at Heysham (top right), Lowestoft (top left), Newlyn (bottom left) and Sheerness (bottom right) before (black) and after (red) we remove the annual mean trends of the skew surge series. The 5% significance level is shown by the black dashed line.

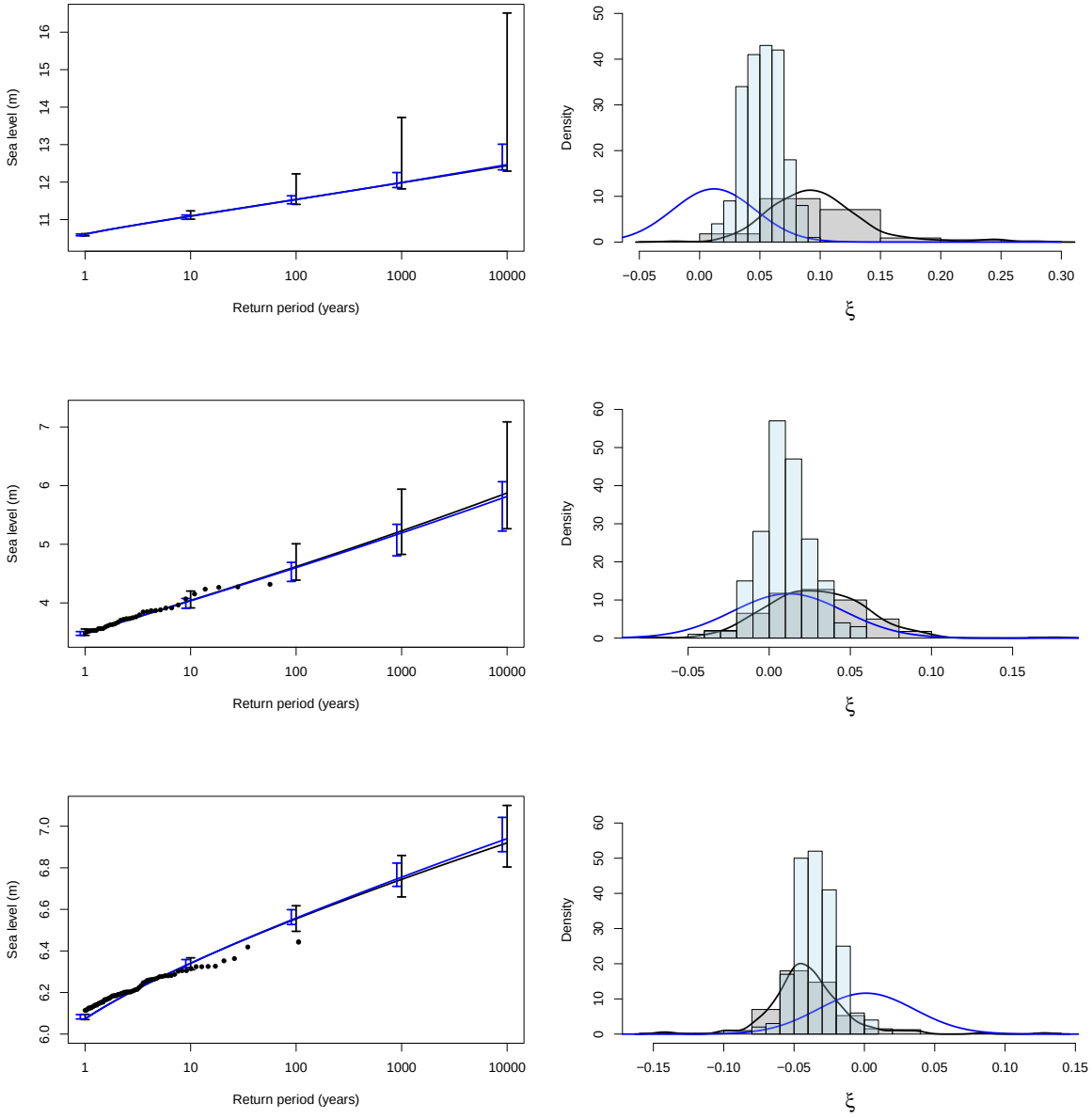


Figure A.9.18: 95% bootstrap confidence intervals on the final (*temporal dependence*) return level estimates at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row) before (black) and after (blue) adding a prior distribution to the shape parameter (left). Empirical estimates are shown by black points. Histograms of the shape parameter estimates and their densities (right) for these two models in their corresponding colours.

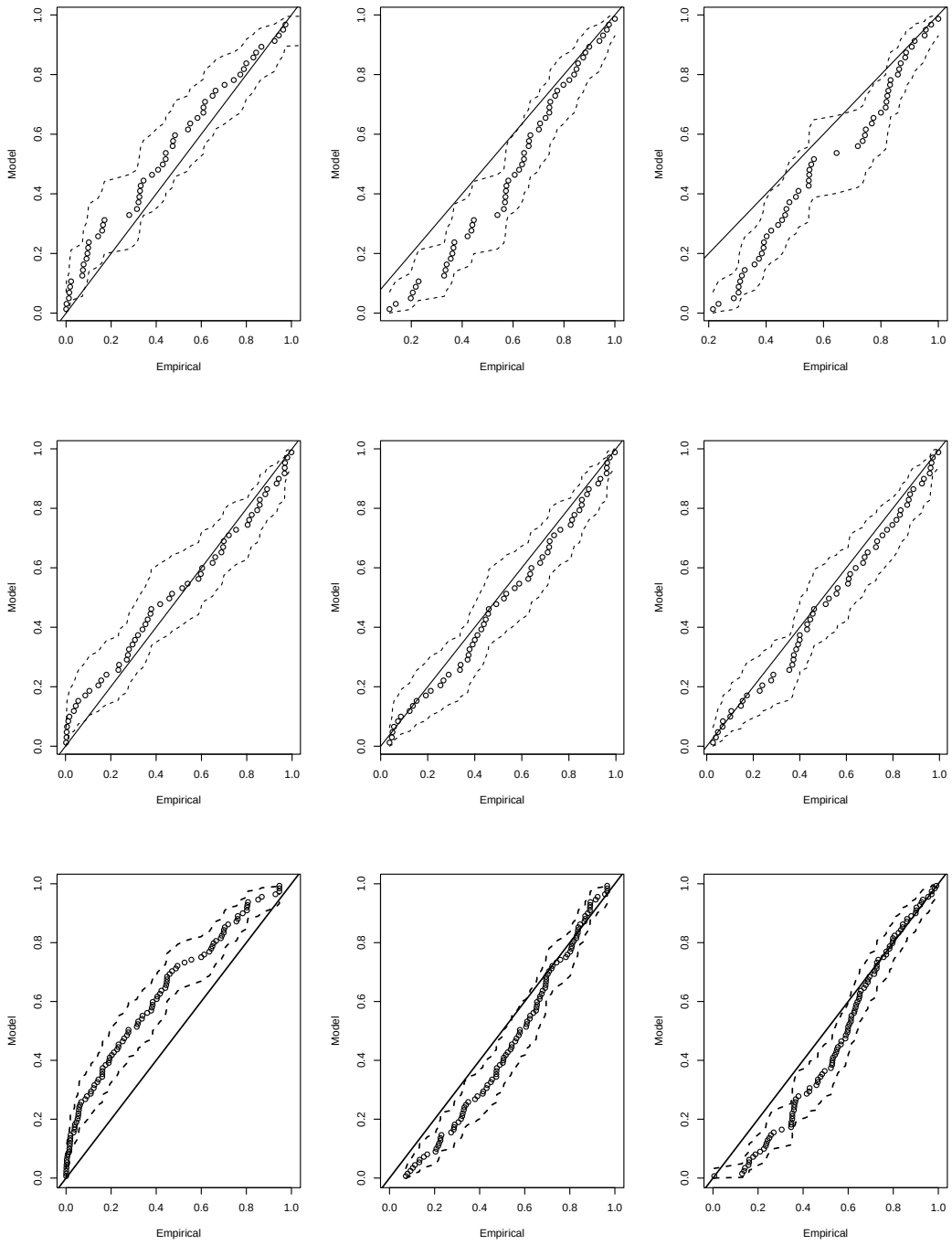


Figure A.9.19: PP plots for the transformed annual maximum sea levels to a uniform scale at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row). These are transformed using the *baseline* (left column) and final (*temporal dependence*) (central column) distribution function for the annual maxima, as well as the year specific final model (right column). The black line shows the line of equality, $y = x$.

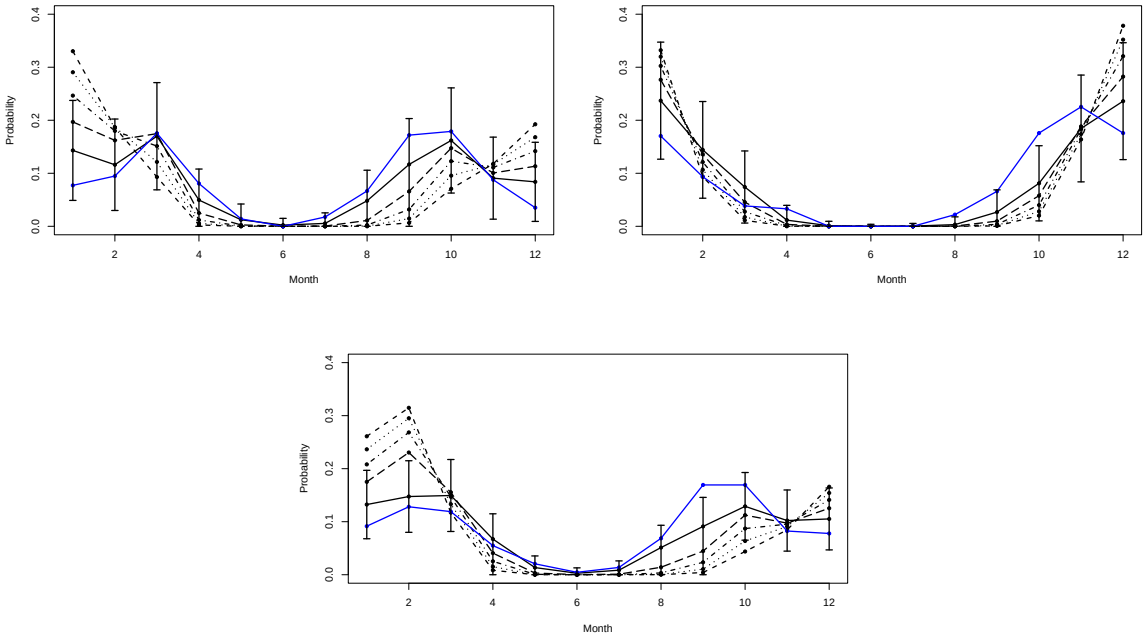


Figure A.9.20: Estimates of $\hat{P}_M(j;z)$ for months $j = 1-12$ at Heysham (top left), Lowestoft (top right) and Newlyn (bottom), for $p = 1$ (black solid), 0.1 (long-dashed), 0.01 (dot-dashed), 0.001 (dotted), and 0.0001 (dashed). The blue line is the empirical estimate $\tilde{P}_M(j;z_1)$, 95% confidence intervals are for $\hat{P}_M(j;z_1)$.

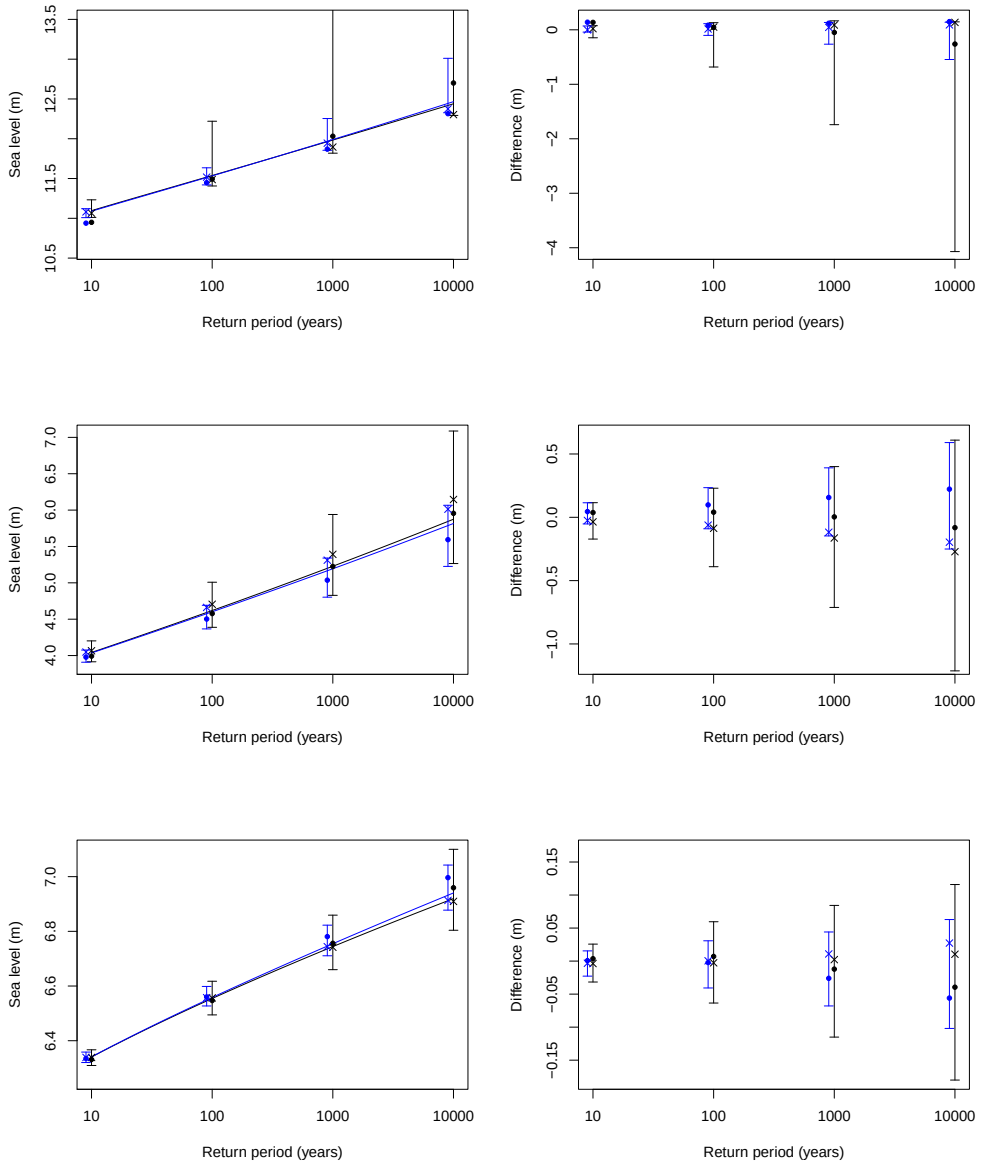


Figure A.9.21: Return level estimates from the final *(temporal dependence)* model with 95% confidence intervals before (black) and after (blue) adding a prior distribution to the shape parameter at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row), where the 0.95 quantile is used for the skew surge distribution (left column); estimates are also shown at the 10, 100, 1000 and 10,000 year levels when the 0.9 (crosses) and 0.99 (dots) quantiles are used in the skew surge model, both with (blue) and without (black) the prior distribution on the shape parameter of the GPD. The corresponding difference between the estimates from the original approach (0.95 quantile) with the 0.9 (crosses) and 0.99 (dots) quantiles, compared with confidence intervals (right column).

Appendix B

ESLestimation: An R package for estimating extreme sea levels

B.1 Introduction

We develop the `ESLestimation` R package to accompany the methodologies presented in Chapter 3. The package version is currently v1.0.4. All documentation can be found at https://github.com/eleanordarcy/ESLestimation/. The package can be installed and loaded as follows:

```
library(devtools)
install_github('eleanordarcy/ESLestimation')
library(ESLestimation)
```

The package consists of five functions and a dataset that are described in the following sections. First, we present data from Lowestoft (UK) in Section B.2 that is used as an example throughout the documentation and throughout the remainder of this chapter. We discuss the functions relevant for model fitting to skew surges in Section B.3; namely, fitting the scale and shape parameters of the generalised Pareto distribution (GPD) in Section B.3.1, fitting the regression model for the GPD rate parameter in

Section B.3.2 and fitting our parametric extremal index model in Section B.3.3. Then in Section B.4 we detail the functions for estimating extreme sea levels, that use the models from the preceding sections. We have a function for estimating return levels (Section B.4.1) and for obtaining measures of uncertainty (Section B.4.2).

B.2 Data

We provide an illustrative dataset of peak tide and skew surge observations (in metres, relative to chart datum) at Lowestoft (UK), named `Lowestoft`. This is a processed and quality-controlled version of that from BODC (2020), as part of the UK National Tide Gauge Network; this original data source is discussed in detail in Section 3.3.1 of Chapter 3. To load in the data, use the following command:

```
data(Lowestoft)
```

This loads a dataframe with 8 columns and 26848 rows. The first four columns correspond to the observation time in terms of the year, month, day (in year) and hour (in day). Observations are recorded for every tidal cycle, i.e., every 12.5 hours approximately, from 1964 to 2016 with 4% of observations missing. The fifth and sixth columns are skew surge (named `skews`) and peak tide (named `maxTide`) observations, respectively. The seventh column gives a date object for the date of observation. The eighth column is a standardised version of the peak tide column (named `stTide`), where the standardisation is done by subtracting the mean and dividing by the standard deviation of all peak tides.

B.3 Model fitting

B.3.1 GPD.fit

The `GPD.fit` function fits the non-stationary GPD model of Chapter 3 to skew surge exceedances of a monthly threshold. The function is used as follows

```
GPD.fit(data, q = 0.95, optim.method = "BFGS",
        init.par = c(0.2, 0, 100, 0.1, 0.1))
```

It takes the following arguments:

- `data`: A data frame of skew surge observations (named `skews`), along with `month`, `day` (in year) and `maxTide` observations.
- `q`: The quantile of monthly skew surges to obtain exceedances, it must be between 0 and 1. The default is 0.95 but this should be investigated using standard threshold selection techniques.
- `optim.method`: The method to be used for optimisation of the log-likelihood function for the GPD. The default is `'BFGS'` but see `optim` of R Core Team (2016) for details and alternative options.
- `init.par`: The initial values for parameters to be optimised over in `optim`. The default are `c(0.2,0,100,0.1,0.1)` but these can be altered.

As discussed in Chapter 3, this function uses maximum likelihood estimation to fit a non-stationary GPD to skew surge exceedances of a monthly threshold, corresponding to the q-quantile of the data for that month. Non-stationarity is introduced to the scale parameter of the GPD using a harmonic to capture seasonal variations and a linear trend for the tide covariate, to capture skew surge-peak tide dependence; see equation (3.4.12) of Chapter 3. The GPD log-likelihood is given within the `GPD.fit`

function, but omitted here for brevity, and is then optimised using `optim` to obtain parameter estimates.

A normal prior distribution, expressed as a penalty function, is incorporated in the log-likelihood to reduce uncertainty associated with shape parameter estimation. This prior information is based on spatial information about shape parameter estimates across the UK, as was done by Environment Agency (2018) and is discussed in Section 3.5.4 of Chapter 3.

The function returns a vector of 5 parameters for the GPD model of Section 3.4.4 in Chapter 3, namely $(\hat{\alpha}_\sigma, \hat{\beta}_\sigma, \hat{\phi}_\sigma, \hat{\gamma}_\sigma, \hat{\xi})$.

The following code demonstrates the `GPD.fit` function:

```
library(ESLestimation)


# Load in available data
data(Lowestoft)


# Check the data has the required variables that are correctly named
head(Lowestoft)


# Fit the non-stationary GPD model to extreme skew surges
# In this example we use all default arguments
GPD.fit(Lowestoft)
```

B.3.2 rateparam.fit

The `rateparam.fit` function fits a non-stationary regression model to the GPD rate parameter for skew surge exceedances of a monthly threshold. It has arguments identical to those in Section B.3.1:

```
rateparam.fit(data, q = 0.95, optim.method = "BFGS",
```

```
init.par = c(0.2, 0.2, 0.2, 0.2, 360))
```

As discussed in Section X of Chapter 3, this function uses maximum likelihood estimation to fit a harmonic-based regression model to the GPD rate parameter, used for skew surge exceedances of a monthly threshold (determined by the value of `q`). Non-stationarity is introduced to the rate parameter of the GPD using two harmonics to capture seasonal variations and skew surge-peak tide dependence; this is mathematically defined in equation (3.4.11) of Chapter 3. To fit the model, we use an indicator function to determine whether an observation exceeds its monthly quantile and assume this binary variable follows a Bernoulli distribution. The Bernoulli log-likelihood is provided within the function and is then optimised using `optim` (R Core Team, 2016).

The output of the function is a vector of parameter estimates for the rate parameter model: $(\hat{\alpha}_{\lambda}^{(x)}, \hat{\beta}_{\lambda}^{(d)}, \hat{\beta}_{\lambda}^{(x)}, \hat{\phi}_{\lambda}^{(d)}, \hat{\phi}_{\lambda}^{(x)})$.

The following code demonstrates its use:

```
library(ESLestimation)


# Load in available data
data(Lowestoft)


# Fit the rate parameter model with default arguments
rateparam.fit(Lowestoft)
```

B.3.3 extremalindex.fit

The final model fitting function is `extremalindex.fit` that fits a parametric model to empirical estimates of the extremal index (defined in Section 2.3 of Chapter 2) of skew surge observations:

```
extremalindex.fit(data, run.length = 10, thresh.quantile = 0.99)
```

The function requires the following arguments:

- `data`: A data frame of skew surge observations (variable named `skews`)
- `run.length`: This is the number of consecutive non-exceedances between two extreme observations, where we would say they belong to different clusters (as for the standard runs estimate). The default is 10 (i.e., 5 days) but it can be inferred from autocorrelation (acf) plots, i.e., the maximum lag before the acf remains close to zero.
- `thresh.quantile`: The quantile of skew surges above which the exponential decay model is required, i.e., the empirical estimates become noisy above this value. This is a single value between 0 and 1. The default is 0.99.

The `extremalindex.fit` consists of two stages, as outlined in Section 3.4.5 of Chapter 3. For values below the threshold (`thresh.quantile`), the empirical runs estimate (defined in Section 3.4.5 of Chapter 3) is used as this is smooth over skew surges in this range. For computational efficiency purposes, these empirical estimates are evaluated on a regular grid of 100 values from the minimum skew surge up to this threshold. The empirical runs estimate is found using the `extRemes` R package (Gilleland and Katz, 2016). Linear interpolation is used for values between those on the regular grid. For skew surges above the threshold, we fit a parametric exponential decay model. Parameters of this model are estimated using a weighted-least squares approach via `optim` (R Core Team, 2016). The full model and the inference procedure are outlined in Section 3.4.5 of Chapter 3.

This function outputs a vector of parameter estimates for the model of equation (3.4.16) in Chapter 3: $(\hat{\psi}, \hat{\theta})$.

We demonstrate the implementation of this function below:

```
library(ESLestimation)
```

```
# Load in available data
data(Lowestoft)

# Infer the run length from the acf plot
acf(Lowestoft$skews)
# Run length of 10 is reasonable

# Check when empirical estimates become noisy
library(evd)
plot(Lowestoft$skews, sapply(Lowestoft$skews, function(y)
                             extremalindex(Lowestoft$skews, y,
                             method='runs', run.length=10)[1]),
     xlab='Skew surge (m)', ylab=expression(tilde(theta)))
abline(v=quantile(Lowestoft$skews, 0.99, names=F))
# 0.99-quantile is reasonable

# Fit the extremal index model to skew surges using default arguments
extremalindex.fit(Lowestoft)
```

Supporting figures for the run length and threshold choice are given in Figure A.9.5 and A.9.8 of Appendix A, respectively.

B.4 Estimation

B.4.1 returnlevel.est

Now we bring together the non-stationary model for skew surges of Chapter 3 with the known tidal regime, to estimate extreme sea levels for a specified annual exceedance probability p in the `returnlevel.est` function:

```
returnlevel.est(p, data, gpd_par, rate_par, extremalindex_par,
                gpd.quantile = 0.95, extremalindex.quantile = 0.99,
                run.length = 10)
```

The function takes the following arguments:

- `p`: The annual exceedance probability (equal to the reciprocal of the return period) for the required return level estimate. This can be a single value or a vector of probabilities. This must take values between 0 and 1.
- `data`: A data frame of skew surge observations (named `skews`), along with `month`, `day` (in year) and `maxTide` observations for covariate information.
- `gpd_par`: Scale and shape parameters for the GPD fit to skew surges (found via `GPD.fit` function of Section B.3.1). This should be a vector of length 5.
- `rate_par`: Rate parameter for the GPD fit to skew surges (found via `rateparam.fit` function of Section B.3.2). This should be a vector of length 5.
- `extremalindex_par`: Parameters for the extremal index model fit (found via `extremalindex.fit` function of Section B.3.3). This should be a vector of length 2.
- `gpd.quantile`: The quantile used to define exceedances and fit the GPD model for skew surges. This is a single value between 0 and 1. The default is 0.95.
- `extremalindex.quantile`: The quantile of skew surges for the extremal index model, above which the exponential decay model is required, i.e., the empirical estimates become noisy above this value. This is a single value between 0 and 1. The default is 0.99.
- `run.length`: This is the number of consecutive non-exceedances between two extreme observations, where we would say they belong to different clusters (as for

the standard runs estimate). The default is 10 (i.e., 5 days) but can be inferred from acf plots, i.e., the maximum lag before the acf remains close to zero.

To estimate sea level return levels, we use the derived annual maxima distribution for sea levels of equation (3.4.15) in Chapter 3. This uses the fact that sea levels Z_t can be decomposed into the sum of skew surge Y_t and peak tide X_t. We summarise the methodology for estimating return levels below, but refer the reader to Chapter 3 for details.

Skew surges are modelled using a non-stationary GPD for exceedances of a monthly threshold (`gpd.quantile`), and a non-stationary monthly empirical distribution for non-exceedances. We account for non-stationarity and skew surge-peak tide dependence, using daily, monthly and peak covariates within the model. This model is fit using the `GPD.fit` function (see Section B.3.1 for details) and the output of that function (i.e., a vector of GPD parameter estimates) is used as an input to this function. The rate parameter is separately modelled, with the same covariate information, through the `rateparam.fit` function (see Section B.3.2 for details) and its output (i.e., a vector of parameter estimates) is also an input to this function. Lastly, we model the temporal dependence of skew surges through a model for the extremal index fit via `extremalindex.fit` (see Section B.3.3), the output to this function is used as an input to this function.

As tides are deterministic, we chose tidal samples to input into our skew surge distribution to derive a model for the skew surges. Then the sea level distribution is derived via a joint probabilities methodology, assuming that skew surge-peak tide dependence has been fully captured through the non-stationary skew surge model. Then the annual maxima distribution for sea levels is found by taking the product over the skew surge distribution, to the power of the extremal index model, for observations and their associated covariates within a year; see equation (3.4.10) of Chapter 3. To account for tidal seasonality, we first derive the sea level monthly maxima distribution

using month-specific tidal series; see equation (3.4.15) of Chapter 3. To account for interannual tidal variations, we do this for each year in the record, using the associated tidal sequence, and then average over all years; see equation (3.4.15) of Chapter 3.

We use the numerical solver `uniroot` (R Core Team, 2016) to find the return level z_p for the associated annual exceedance probability p. That is, to solve the equation $\Pr(M \leq z_p) = 1 - p$ for M the annual maxima. `uniroot` requires an interval to test over for values of z_p which is set as the 0.5 quantile of sea levels and the maximum sea level plus 3m for the lower and upper endpoints, respectively. If these endpoints are not wide enough, they will be increased by 0.1m recursively.

The function returns the sea level return level (in metres relative to chart datum) for the required annual exceedance probability p. If the input `p` is a vector, the output will also be a vector of corresponding return levels.

Implementation of `returnlevel.est` is as follows:

```
library(ESLestimation)


# Load in available data
data(Lowestoft)


# Estimate 10 year return level (p=0.1)
# Use model fit functions to get parameter estimates
# Default arguments are used elsewhere
returnlevel.est(p = 0.1, data = Lowestoft,
                gpd_par=GPD.fit(Lowestoft),
                rate_par=rateparam.fit(Lowestoft),
                extremalindex_par=extremalindex.fit(Lowestoft))
```

B.4.2 CI.est

The final function available to the user is `CI.est` that gives confidence intervals on the sea level return level estimates from the non-stationary model of Chaper 3:

```
CI.est(p, data, ci.prob, gpd_par, rate_par, extremalindex_par,
       block.length = 10, n.boot = 200, gpd.quantile = 0.95,
       optim.method = "BFGS", EI.run.length = 10, EI.quantile = 0.99)
```

The arguments are defined as follows

- `p`: The annual exceedance probability (equal to the reciprocal of the return period) for the required return level estimate. This can be a single value or a vector of probabilities. This must take values between 0 and 1.
- `data`: A data frame of skew surge observations (named `skews`), along with `month`, `day` (in year) and `maxTide` observations for covariate information.
- `ci.prob`: The probability for the confidence interval width. The default is 0.95. This should be a value between 0 and 1.
- `gpd_par`: Scale and shape parameters for the GPD fit to skew surges (found via `GPD.fit` function of Section B.3.1). This should be a vector of length 5.
- `rate_par`: Rate parameter for the GPD fit to skew surges (found via `rateparam.fit` function of Section B.3.2). This should be a vector of length 5.
- `extremalindex_par`: Parameters for the extremal index model fit (found via `extremalindex.fit` function of Section B.3.3). This should be a vector of length 2.
- `block.length`: The block length for the stationary bootstrap procedure, and should represent the approximate duration of a storm. The default is 10, corresponding to approximately 5 days.

- `n.boot`: The number of bootstrap samples to use to obtain confidence intervals. The default is 200.

- `gpd.quantile`: The quantile used to define the GPD threshold, we the model is fit to skew surge exceedances. This is a single value between 0 and 1. The default is 0.95.

- `optim.method`: The optimisation method when refitting the models to bootstrap samples. The default is `'BFGS'`, see R Core Team (2016) for more details on the `optim` function.

- `EI.run.length`: This is the number of consecutive non-exceedances between two extreme observations, where we would say they belong to different clusters (as for the standard runs estimate). The default is 10 (i.e., 5 days) but can be inferred from acf plots, i.e., the maximum lag before the acf remains close to zero.

- `EI.quantile`: The quantile of skew surges for the extremal index model, above which the exponential decay model is required, i.e., the empirical estimates become noisy above this value. This is a single value between 0 and 1. The default is 0.99.

This function uses the stationary bootstrap procedure to obtain confidence intervals on the sea level return level estimates given by the `returnlevel.est` function (see Section B.4.1). The block length is simulated from a Geometric distribution with mean equal to the reciprocal of the input `block.length`. This can be inferred from an acf plot.

Details of the uncertainty quantification procedure adopted here are given in Section 3.5.4 of Chaper 3. In summary, we account for uncertainty at each stage of the modelling procedure by recalculating thresholds, re-estimating model parameters and the empirical distribution for each bootstrap sample. Once the required number of

bootstrap samples is reached (`n.boot`), we find the required quantiles (indicated by `ci.prob`) for confidence intervals. A normal prior distribution is used to penalise fits of the non-stationary GPD (via `GPD.fit`) to restrict shape parameter estimates and ultimately reduce the width of confidence intervals.

The function outputs the $(1 - \texttt{ci.prob}/2)$, 0.5, and $\texttt{ci.prob} + (1 - \texttt{ci.prob}/2)$ quantiles of return level sea level estimates (in metres) over the specified number of bootstrap samples.

We demonstrate the implementation of `CI.est` below:

```
library(ESLestimation)


# Load in available data
data(Lowestoft)


# Estimate 95% CI for 10 year return level (i.e., p=0.1)
# Use model fit functions to get parameter estimates
# Default parameters are used everywhere
CI.est(0.1, data = Lowestoft, ci.prob = 0.95,
       gpd_par = GPD.fit(Lowestoft),
       rate_par = rateparam.fit(Lowestoft),
       extremalindex_par = extremalindex.fit(Lowestoft),
       block.length=10, n.boot=200, gpd.quantile=0.95,
       optim.method='BFGS', EI.run.length=10, EI.quantile = 0.99)
```

Appendix C

Additional proofs for Chapter 6: Extremal properties of max-autoregressive moving average processes for modelling extreme river flows

C.1 Introduction

In this chapter, we provide an additional proof for the stationarity property of max-autoregressive moving average, Max-ARMA(p, q), process. Recall, from Remark 6.2.1 in Chapter 6, that a Max-ARMA(p, q) process is stationary when $0 \leq \alpha_i < 1$ for $i = 1, \ldots, p-1$ and $0 < \alpha_p < 1$, and the scale parameter $\gamma > 0$ of the Fréchet

distribution for the innovation process $\{Z_t\}$ given by

$$\gamma := \sigma\left(\sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau}\left\{\beta_j \prod_{\substack{i=1,\ldots,p:\\ \alpha_i>0}} \alpha_i^{a_i}\right\}\right)^{-1} < \infty,$$

if $\{X_t\}$ has Fréchet distribution with scale parameter σ for $0 < \sigma < \infty$, where

$$\mathcal{S}_\tau = \left\{ a_i \in \{0, 1, \ldots, \tau\} \text{ for } i = 1, \ldots, p,\ j = 0, \ldots, q : \sum_{i=1}^{p} i a_i + j = \tau \right\}.$$

This result was originally proved by Davis and Resnick (1989) but we detail an alternative approach to proving this stationarity condition and obtaining the form of the scale parameter γ of the innovation process $\{Z_t\}$, and fix the margins of $\{X_t\}$ to be independent of the Max-ARMA parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, unlike Davis and Resnick (1989).

In Section C.2, we give the necessary notation for the proof. Then in Section C.3, we state a lemma that is required for the proof and prove this. Finally, in Section C.4 we prove the stationarity result of Davis and Resnick (1989), given in Remark 6.2.1 in Chapter 6.

C.2 Notation for proofs

The following notation is used across the proofs here. Let $\alpha_M = \max\{\alpha_1, \ldots, \alpha_p\}$ and $\beta_M = \max\{\beta_0, \ldots, \beta_q\}$ which, by definition of the Max-ARMA process, satisfy $\alpha_M > 0$ and $\beta_M \geq 1$. We also define $\beta_{M:j} = \max\{\beta_0, \beta_1, \ldots, \beta_j\}$ so that $1 \leq \beta_{M:j} \leq \beta_M$ for all $j = 0, 1, \ldots, q$ and $\beta_{M:q} = \beta_M$. Similarly, we let $\beta_m = \min\{\beta_j, j = 0, \ldots, q : \beta_j > 0\}$, so $0 < \beta_m \leq \beta_M$. Furthermore, we take $i_M = \arg\{i = 1, \ldots, p : \alpha_i = \alpha_M\}$, where here if multiple such indices achieve this equality, we take the largest such index.

C.3 Lemma needed for Proof of Remark 6.2.1

Lemma C.3.1. *For coefficients* $\boldsymbol{\alpha}_{>0} = \{\alpha_i > 0\}_{i=1}^{p}$ *and* $\boldsymbol{\beta}_{>0} = \{\beta_j > 0\}_{j=0}^{q}$ *satisfying the conditions of formulation* (6.2.1) *in Chapter 6, then if* $\alpha_M = \max\{\alpha_1, \ldots, \alpha_p\} < 1$, *the summation* $S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$ *is finite, where*

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) = \sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau} \left\{ \beta_j \prod_{\substack{i=1,\ldots,p: \\ \alpha_i>0}} \alpha_i^{a_i} \right\}, \tag{C.3.1}$$

with $\mathcal{S}_\tau$ *defined by expression* (6.2.3) *in Chapter 6. Otherwise,* $S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$ *is infinite.*

Proof. Assume for the moment that $\alpha_M < 1$. Then

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) \leq \beta_M \sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau} \left\{ \prod_{\substack{i=1,\ldots,p: \\ \alpha_i>0}} \alpha_i^{a_i} \right\}.$$

Then, since the power function to a non-negative power is an increasing function,

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) \leq \beta_M \sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau} \{\alpha_M^{a_1+\ldots+a_p}\}.$$

By the definition of $\mathcal{S}_\tau$, we know that

$$\sum_{i=1}^{p} a_i \geq \sum_{i=1}^{p} \frac{i}{p} a_i = \frac{\tau - j}{p} \geq \frac{\tau - q}{p}.$$

Then,

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) \leq \beta_M \sum_{\tau=0}^{\infty} \alpha_M^{(\tau-q)/p} < \infty.$$

This final sum converges as it is a geometric series with common ratio term less than one. So the sum is finite when $\alpha_M < 1$.

Now consider $\alpha_M \geq 1$. Then,

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) \geq \beta_m \sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau} \left\{ \prod_{\substack{i=1,\ldots,p: \\ \alpha_i>0}} \alpha_i^{a_i} \right\} > 0.$$

Then, we remove the maximum over $\mathcal{S}_\tau$ by focusing only on one of the terms the maximum is over, specifically the term with the highest power of α_M. Then, if $\boldsymbol{\alpha}_{-i_M}$ denotes the vector of the non-zero elements of $\boldsymbol{\alpha}_{>0}$ that excludes α_M and the function $C_\tau(\boldsymbol{\alpha}_{-i_M})$ corresponds to the remainder term after the product of the powered α_i terms is divided by $\alpha_M^{\lfloor \tau/i_M \rfloor}$, where $\lfloor \cdot \rfloor$ denotes the floor function. So we can say that,

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) \geq \beta_m \sum_{\tau=0}^{\infty} \alpha_M^{\lfloor \tau/i_M \rfloor} C_\tau(\boldsymbol{\alpha}_{-i_M}) > 0.$$

Consider only the subsequence of terms in the sum with $\tau = i_M j$, for $j = 0, 1, \ldots$ then there are no remainder terms for this particular subsequence of τ. Then,

$$S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) > \beta_m \sum_{\tau=i_M j: j=0}^{\infty} \alpha_M^{\lfloor \tau/i_M \rfloor} > \beta_m \sum_{j=0}^{\infty} 1 = \infty.$$

with the first inequality holding as the omitted terms from the sum, that are not in this subsequence, are all positive. The second inequality comes from each of the terms in this subsequence being larger than, or equal to 1, due to $\alpha_M \geq 1$ and then this resulting sum being divergent. So when $\alpha_M \geq 1$, $S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$ is infinite. □

C.4 Proof of Remark 6.2.1 in Chapter 6

Proof. Let $\{X_t\}$ be a Max-ARMA(p, q) process, defined in Section 6.2. We aim to find the value of γ, with $0 < \gamma < \infty$, the scale parameter of the Fréchet distribution of the IID innovation process $\{Z_t\}$, and the domains for the process parameters $\boldsymbol{\alpha}_{>0} = \{\alpha_i > 0\}_{i=1}^p$ and $\boldsymbol{\beta}_{>0} = \{\beta_j > 0\}_{j=0}^q$ such that $\{X_t\}$ is a stationary process.

We begin with the distribution function $F_{X_t}(x)$ and use the formulation of the Max-ARMA(p, q) process for all $-\infty < t < \infty$, exploiting the independence of X_t and Z_s for $s > t$ and the IID nature of $\{Z_t\}$ to find conditions that make F_{X_t} not a function of time t. This is as follows,

$$F_{X_t}(x) = \Pr(X_t \leq x) = \Pr(\max\{\alpha_1 X_{t-1}, \ldots, \alpha_p X_{t-p}, \beta_0 Z_t, \beta_1 Z_{t-1}, \ldots, \beta_q Z_{t-q}\} \leq x).$$

This follows from the definition of X_t in expression (6.2.1). Since $\alpha_i \geq 0$ for $i = 1, \ldots, p-1$ by definition, when $\alpha_i = 0$ we interpret $\Pr(X_t \leq x/\alpha_i)$ as $\Pr(X_t \leq \infty) = 1$ for arbitrary t, similarly for Z_t terms. Then for the maximum of the variables in the above expression to be less than x, each individual term must be less than x, so we can rewrite this above expression as a joint probability,

$$\begin{aligned} F_{X_t}(x) = &\Pr(X_{t-1} \leq x/\alpha_1, \ldots, X_{t-p} \leq x/\alpha_p, Z_{t-1} \leq x/\beta_1, \ldots, Z_{t-q} \leq x/\beta_q) \times \\ &\Pr(Z_t \leq x/\beta_0). \end{aligned}$$

We factorise out the probability for Z_t from the joint distribution since it is independent of all other terms, by definition of the process. Then, we use the same procedure to replace the highest order X term on the left hand side, in this case X_{t-1}, by its Max-ARMA definition. We also use the independence of Z_{t-1} and $(X_{t-2}, \ldots, X_{t-(p+1)}, Z_{t-2}, \ldots, Z_{t-(q+1)})$, to factorise out the Z_{t-1} term. Following from the above expression, this

gives,

$$\begin{aligned} F_{X_t}(x) = \Pr(&X_{t-2} \leq x/\max\{\alpha_1^2, \alpha_2\}, X_{t-3} \leq x/\max\{\alpha_1\alpha_2, \alpha_3\}, \ldots, \\ &X_{t-p} \leq x/\max\{\alpha_1\alpha_{p-1}, \alpha_p\}, X_{t-(p+1)} \leq x/\alpha_1\alpha_p, \\ &Z_{t-2} \leq x/\max\{\alpha_1\beta_1, \beta_2\}, \ldots, Z_{t-q} \leq x/\max\{\alpha_1\beta_{q-1}, \beta_q\}, \\ &Z_{t-(q+1)} \leq x/\alpha_1\beta_q)\times \\ &\Pr(Z_t \leq x/\beta_0)\Pr(Z_{t-1} \leq x/\max\{\alpha_1\beta_0, \beta_1\}). \end{aligned}$$

Notice in the above expression, we write the joint probability of the same random variable as a single term. For example, the joint event $\{X_{t-1} \leq x/\alpha_1^2, X_{t-1} \leq x/\alpha_2\}$ becomes $\{X_{t-1} \leq x/\max(\alpha_1^2, \alpha_2)\}$, since for X_{t-1} to be less than both x/α_1^2 and x/α_2, X_{t-1} must be less than $\min\{x/\alpha_1^2, x/\alpha_2\}$, or equivalently, $x/\max\{\alpha_1^2, \alpha_2\}$. We continue in the same way, replacing the highest order term, say X_s for $-\infty < s \leq t-2$ using the Max-ARMA definition, then factorising out the independent Z_s term from the joint probability and grouping the joint probabilities for the same random variable. We demonstrate this procedure below for two more steps, replacing X_{t-2} by its definition and factorising out Z_{t-2} firstly, then the same for X_{t-3} and Z_{t-3}.

$$
\begin{aligned}
F_{X_t}(x) = \Pr(&X_{t-3} \le x/\max\{\alpha_1^3, \alpha_1\alpha_2, \alpha_3\}, \ldots, \\
&X_{t-p} \le x/\max\{\alpha_1^2\alpha_{p-2}, \alpha_2\alpha_{p-2}, \alpha_1\alpha_{p-1}, \alpha_p\}, \\
&X_{t-(p+1)} \le x/\max\{\alpha_1^2\alpha_{p-1}, \alpha_2\alpha_{p-1}, \alpha_1\alpha_p\}, \\
&X_{t-(p+2)} \le x/\max\{\alpha_1^2\alpha_p, \alpha_2\alpha_p, \alpha_1\alpha_{p-3}, \alpha_{p-2}\}, \\
&Z_{t-3} \le x/\max\{\alpha_1^2\beta_1, \alpha_2\beta_1, \alpha_1\beta_2, \beta_3\}, \\
&Z_{t-4} \le x/\max\{\alpha_1^2\beta_2, \alpha_2\beta_2, \alpha_1\beta_3, \beta_4\}, \ldots, \\
&Z_{t-q} \le x/\max\{\alpha_1^2\beta_{q-2}, \alpha_2\beta_{q-2}, \alpha_1\beta_{q-1}, \beta_q\}, \\
&Z_{t-(q+1)} \le x/\max\{\alpha_1^2\beta_{q-1}, \alpha_2\beta_{q-1}, \alpha_1\beta_q\}, \\
&Z_{t-(q+2)} \le x/\max\{\alpha_1^2\beta_q, \alpha_2\beta_q\})\Pr(Z_t \le x/\beta_0)\times \\
&\Pr(Z_{t-1} \le x/\max\{\alpha_1\beta_0, \beta_1\})\Pr(Z_{t-2} \le x/\max\{\alpha_1^2\beta_0, \alpha_2\beta_0, \alpha_1\beta_1, \beta_2\}) \\
= \Pr(&X_{t-4} \le x/\max\{\alpha_1^4, \alpha_1^2\alpha_2, \alpha_1\alpha_3, \alpha_2^2, \alpha_4\}, \ldots, \\
&X_{t-p} \le x/\max\{\alpha_1^3\alpha_{p-3}, \alpha_1\alpha_2\alpha_{p-3}, \alpha_3\alpha_{p-3}, \alpha_1^2\alpha_{p-2}, \alpha_2\alpha_{p-2}\}, \\
&X_{t-(p+1)} \le x/\max\{\alpha_1^3\alpha_{p-2}, \alpha_1\alpha_2\alpha_{p-2}, \alpha_3\alpha_{p-2}, \alpha_1^2\alpha_{p-1}, \alpha_2\alpha_{p-1}, \alpha_1\alpha_p\}, \\
&X_{t-(p+2)} \le x/\max\{\alpha_1^3\alpha_{p-1}, \alpha_1\alpha_2\alpha_{p-1}, \alpha_2\alpha_{p-1}, \alpha_1^2\alpha_p, \alpha_2\alpha_p, \alpha_1\alpha_{p-3}, \alpha_{p-2}\}, \\
&X_{t-(p+3)} \le x/\max\{\alpha_1^3\alpha_p, \alpha_1\alpha_2\alpha_p, \alpha_3\alpha_p\}, \\
&Z_{t-4} \le x/\max\{\alpha_1^3\beta_1, \alpha_1\alpha_2\beta_1, \alpha_3\beta_1, \alpha_1^2\beta_2, \alpha_2\beta_2, \alpha_1\beta_3, \beta_3\}, \ldots, \\
&Z_{t-(q+3)} \le x/\max\{\alpha_1^3\beta_q, \alpha_1\alpha_2\beta_q, \alpha_3\beta_q\})\times \\
&\Pr(Z_t \le x/\beta_0)\Pr(Z_{t-1} \le x/\max\{\alpha_1\beta_0, \beta_1\})\times \\
&\Pr(Z_{t-2} \le x/\max\{\alpha_1^2\beta_0, \alpha_2\beta_0, \alpha_1\beta_1, \beta_2\})\times \\
&\Pr(Z_{t-3} \le x/\max\{\alpha_1^3\beta_0, \alpha_1\alpha_2\beta_0, \alpha_3\beta_0, \alpha_1^2\beta_1, \alpha_2\beta_1, \alpha_1\beta_2, \beta_3\}).
\end{aligned}
\tag{C.4.1}
$$

This recurrence continues indefinitely. Repeating this recurrence until the leading term

is for $X_{t-\upsilon}$, where $\upsilon > p$, then the expression becomes

$$F_{X_t}(x) = C_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) D_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}), \tag{C.4.2}$$

where

$$C_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) = \Pr(\cap_{i=1}^{p} \{X_{t-\upsilon-i} \leq x/A_{\upsilon,i}\}, \cap_{j=1}^{q} \{Z_{t-\upsilon-j} \leq x/B_{\upsilon,j}\}),$$

and

$$D_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) = \prod_{\tau=0}^{\upsilon-1} \Pr\left(Z_{t-\tau} \leq x/E_\tau\right),$$

where $A_{\upsilon,i}, B_{\upsilon,j} > 0$ for $i = 1, \ldots, p$, $j = 1, \ldots, q$ and $E_\tau > 0$ are constants for all υ and $\tau \in \mathbb{N}$, with these constants depending on the Max-ARMA parameters $\boldsymbol{\alpha}_{>0}$ and $\boldsymbol{\beta}_{>0}$, through expressions we will derive later and consider all three terms as $\upsilon \to \infty$ for different parameters $\boldsymbol{\alpha}_{>0}$ and $\boldsymbol{\beta}_{>0}$.

First consider the $D_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$ term. To derive a general formula for E_τ, initially consider E_3 when $p \geq 3$, $q \geq 3$, $\alpha_i > 0$ for $i = 1, 2, 3$ and $\beta_j > 0$ for $j = 1, 2, 3$. Expression (C.4.1) gives that $E_3 = \max\{\alpha_1^3, \alpha_1\alpha_2, \alpha_3, \alpha_1^2\beta_1, \alpha_2\beta_1, \alpha_1\beta_2, \beta_3\}$. Notice that for each term contributing to the maximum to give E_3 the sum of the product of the sub- and superscript of each $\alpha_i^{a_i}$ where $i = 1, \ldots, 3$ and $a_i \in \{0, \ldots, 3\}$ plus the superscript of the β_j for $j = 0, \ldots, 3$ is equal to 3. For example, looking at each term within

this maxima gives:

$$\begin{aligned}
\alpha_1^3\beta_0 = \alpha_1^3\alpha_2^0\alpha_3^0\beta_0 : (1\times 3)+(2\times 0)+(3\times 0)+0 = 3,\\
\alpha_1\alpha_2\beta_0 = \alpha_1^1\alpha_2^1\alpha_3^0\beta_0 : (1\times 2)+(2\times 1)+(3\times 0)+0 = 3,\\
\alpha_3\beta_0 = \alpha_1^0\alpha_2^0\alpha_3^1\beta_0 : (1\times 0)+(2\times 0)+(3\times 1)+0 = 3,\\
\alpha_1^2\beta_1 = \alpha_1^2\alpha_2^0\alpha_3^0\beta_1 : (1\times 2)+(2\times 0)+(3\times 0)+1 = 3,\\
\alpha_2\beta_1 = \alpha_1^0\alpha_2^1\alpha_3^0\beta_1 : (1\times 0)+(2\times 1)+(3\times 0)+1 = 3,\\
\alpha_1\beta_2 = \alpha_1^1\alpha_2^0\alpha_3^0\beta_2 : (1\times 1)+(2\times 0)+(3\times 0)+2 = 3,\\
\beta_3 = \alpha_1^0\alpha_2^0\alpha_3^0\beta_3 : (1\times 0)+(2\times 0)+(3\times 0)+3 = 3.
\end{aligned}$$

This is an exhaustive list of all possible combinations such that $(a_1+2a_2+3a_3)+j=3$. The generalisation of this maximum term, over combinations of possible components covering cases with an arbitrary set of $\{\alpha_i\}$ and $\{\beta_j\}$ coefficients being zero, is that $E_\tau = \gamma_\tau$, where γ_τ is given by expression (6.3.4). As $Z_t \sim$ Fréchet(γ), for all t, we have then that

$$D_{t,\upsilon}(x;\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0}) = \exp\left(-\frac{\gamma}{x}\sum_{\tau=0}^{\upsilon-1}\gamma_\tau\right) \to \exp\left(-\frac{\gamma}{x}S(\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0})\right),$$

as $\upsilon\to\infty$, where $S(\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0})$ is defined by summation (C.3.1), and the limit does not depend on t. From Lemma C.3.1, $0 < S(\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0})\infty$ if $\max\{\alpha_1,\ldots,\alpha_p\} < 1$, so this constraint imposes feasible parameters for obtaining a non-degenerate distribution from expression (C.4.2). In contrast, Lemma C.3.1, gives that $S(\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0})$ is infinite when $\max\{\alpha_1,\ldots,\alpha_p\} \geq 1$, in which case $D_{t,\upsilon}(x;\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0}) \to 0$ as $\upsilon\to\infty$, so in that case $F_{X_t}(x)$ must be a degenerate distribution from expression (C.4.2), as $0 \leq C_{t,\upsilon}(x;\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0}) \leq 1$ so cannot offset the zero limit for $D_{t,\upsilon}(x;\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0})$.

Now consider the $C_{t,\upsilon}(x;\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0})$ term focusing only on when $\alpha_M < 1$ as the other parameter values have been ruled out as feasible possibilities for having a stationary

Max-ARMA process. Then we have that

$$C_{t,\upsilon}(x;\boldsymbol{\alpha}_{>0},\boldsymbol{\beta}_{>0}) \geq \Pr(\cap_{i=1}^{p}\{X_{t-\upsilon-i} \leq x/A_{\upsilon,i}^{\max}, \cap_{j=1}^{q}\{Z_{t-\upsilon-j} \leq x/B_{\upsilon,j}^{\max}\}), \qquad \text{(C.4.3)}$$

where $A_{\upsilon,i}^{\max}$ is $A_{\upsilon,i}$ when we set $\alpha_i = \alpha_M$ for all $i = 1,\ldots,p$ and $B_{\upsilon,j}^{\max}$ is $B_{\upsilon,j}$ when we set $\beta_j = \beta_M$ for all $j = 0,\ldots,q$, with inequality (C.4.3) arising owing to $A_{\upsilon,i}^{\max} \geq A_{\upsilon,i}$ and $B_{\upsilon,j}^{\max} \geq B_{\upsilon,j}$ for all υ, i and j. To find the form of $A_{\upsilon,i}^{\max}$ and $B_{\upsilon,j}^{\max}$ we consider maximum Max-ARMA process, $\{X_t^*\}$ with these coefficients, i.e.,

$$X_t^* = \max\{\alpha_M X_{t-1}^*,\ldots,\alpha_M X_{t-p}^*, \beta_M Z_t, \beta_M Z_{t-1},\ldots,\beta_M Z_{t-q}\}.$$

Using the same recursive approach for X_t^*, as in expression (C.4.1) for X_{t-1}, after one step

$$X_t^* = \max\{\alpha_M X_{t-2}^*,\ldots,\alpha_M X_{t-p}^*, \alpha_M^2 X_{t-p-1}^*, \beta_M Z_{t-1},\ldots,\beta_M Z_{t-q}, \alpha_M\beta_M Z_{t-q-1} : Z_t\},$$

where Z_t is treated separately since it is independent of the remaining terms. Note that only the newly introduced terms $\{X_{t-p-1}^*, Z_{t-q-1}\}$ have a different form of coefficients from the other like terms, this is as all other new terms are multiplied by $\alpha_M < 1$ which cannot contribute to the overall maximum. Stepping back a further term, it is only the last two coefficients for the X^* and Z terms that differ from the other coefficients, each with the extra factor α_M.

Recursively stepping back until the leading term of the X^* terms is $X_{t-\upsilon-1}^*$ we get the coefficients of $(X_{t-\upsilon-1}^*,\ldots,X_{t-\upsilon-p}^*)$ are $(A_{\upsilon,1}^{\max},\ldots,A_{\upsilon,p}^{\max})$ and of $(Z_{t-\upsilon-1},\ldots,Z_{t-\upsilon-q})$ which are $(B_{\upsilon,1}^{\max},\ldots,B_{\upsilon,q}^{\max})$. To express these coefficients most effectively we need to consider how υ is linked to p and q. Specifically, for each $\upsilon \in \mathbb{N}$, we write $\upsilon = \pi_\alpha p + r_\alpha$ and $\upsilon = \pi_\beta q + r_\beta$, where $\pi_\alpha, \pi_\beta \in \mathbb{N}$ and $r_\alpha \in \{0,1,\ldots,p-1\}$ and $r_\beta \in \{0,1,\ldots,q-1\}$.

Then it follows that

$$X_t^* = \max\{\alpha_M^{\pi_\alpha+1} X_{t-\upsilon-1}^*, \alpha_M^{\pi_\alpha+1+I(r_\alpha\geq p-1)} X_{t-\upsilon-2}^*, \ldots, \alpha_M^{\pi_\alpha+I(r_\alpha\geq 1)+1} X_{t-\upsilon-p}^*,$$
$$\alpha_M^{\pi_\beta}\beta_M Z_{t-\upsilon-1}, \alpha_M^{\pi_\beta+I(r_\beta\geq q-1)}\beta_M Z_{t-\upsilon-2}, \ldots, \alpha_M^{\pi_\beta+I(r_\beta\geq 1)}\beta_M Z_{t-\upsilon-q} : T_\upsilon(\boldsymbol{Z}_{t-\upsilon:t})\},$$

where I is the indicator function and T_υ is a function of the variables $\boldsymbol{Z}_{t-\upsilon:t} = (Z_{t-\upsilon}, \ldots, Z_t)$ that covers events in the partition of terms that are covered by the probability $D_t(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$, so are not detailed here. Hence

$$A_{\upsilon,i}^{\max} = \alpha_M^{\pi_\alpha+I(r_\alpha\geq i)+1} \text{ and } B_{\upsilon,j}^{\max} = \alpha_M^{\pi_\beta+I(r_\beta\geq j)}\beta_M,$$

for $i = 0, \ldots, p-1$ and $j = 0, \ldots, q-1$.

As $\upsilon \to \infty$ it follows, from $\alpha_M < 1$, that $A_{\upsilon,i}^{\max} \to 0$ and $B_{\upsilon,j}^{\max} \to 0$, so the probability on the right hand side of inequality (C.4.3) tends to 1, and so does $C_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$ for all x and t.

So combining the limiting expressions for $C_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$ and $D_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})$, as $\upsilon \to \infty$, we have that the distribution function $F_{X_t}(x)$ as $\upsilon \to \infty$ is

$$F_{X_t}(x) = C_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) D_{t,\upsilon}(x; \boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0}) \to \exp\left(-\frac{\gamma}{x} S(\boldsymbol{\alpha}_{>0}, \boldsymbol{\beta}_{>0})\right) := F_X(x),$$

for $x > 0$, when $\alpha_M < 1$. So in this case F_{X_t} is non-degenerate for all t. As this distribution function is not a function of t, then this process is stationary with marginal distribution function $F_X(x)$. If we want $\{X_t\}$ to have Fréchet(σ) margins, with $0 < \sigma < \infty$, so that $F_X(x) = \exp(-\sigma/x)$, then rearranging gives that γ must satisfy,

$$\gamma = \sigma\left(\sum_{\tau=0}^{\infty} \max_{\mathcal{S}_\tau}\left\{\beta_j \prod_{\substack{i=1,\ldots,p: \\ \alpha_i>0}} \alpha_i^{a_i}\right\}\right)^{-1}.$$

It follows that $0 < \gamma < \infty$ in this case, giving the result for Proposition 6.2.1 when

$\alpha_M < 1$. As there is no value of $\gamma > 0$ which would make $F_{X_t}(x) > 0$ for any value of $x > 0$ when $\alpha_M \geq 1$; then there is no stationary Max-ARMA(p, q) process if $\alpha_M \geq 1$. □

Appendix D

Supplementary material for Chapter 7: A marginal modelling approach for predicting wildfire extremes across the contiguous United States

D.1 Spatio-temporal neighbourhoods

As alternatives to the spatial neighbourhoods $\mathcal{N}_i$, $i \in \{1, \ldots, N\}$, defined in equation (7.3.1) in Chapter 7, we also considered neighbourhoods of the form

$$\mathcal{N}_i^t := \big\{ j \in \{1, \ldots, N\} : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq k_1^{CNT}, m_j = m_i, \\ y_j \in \{y_i - k_y^{CNT}, y_i - k_y^{CNT} + 1, \ldots, y_i + k_y^{CNT}\} \big\},$$

and

$$\mathcal{N}_i^c := \left\{ j \in \{1, \ldots, N\} : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq k_1^{CNT}, m_j = m_i, y_j = y_i, c_j = c_i \right\},$$

for some $k_1^{CNT} \geq 0$ and $k_y^{CNT} \in \mathbb{N}$, where c_j denotes a covariate-based cluster assignment for each observation $j \in \mathcal{N}_i$. Analogous neighbourhoods were also considered for BAP.

With the observation month fixed and the spatial range defined as in $\mathcal{N}_i$, the neighbourhood $\mathcal{N}_i^t$ incorporates additional observations from neighbouring years, thus increasing the amount of data available for marginal estimation and adding a temporal element to the modelling procedure. On the other hand, the month and year are fixed for $\mathcal{N}_i^c$ so that only those data points with the same cluster assignment as observation i are considered, thus reducing the amount of information available for a fixed k_1^{CNT} or k_1^{BAP} value. However, assuming we can define clusters such that observations in the same cluster have more similar marginal tail properties, this additional step has the potential to improve marginal estimation.

Cluster assignments used within the $\mathcal{N}_i^c$ neighbourhoods were computed using divisive hierarchical clustering (Rokach and Maimon, 2005) for two different covariates: temperature and precipitation. We select these variables since they have been shown to be positively and negatively associated, respectively, to wildfire events (Duane et al., 2021; Crockett and Westerling, 2018), and therefore may allow us to group together locations with similar marginal properties for CNT and BAP.

We use hierarchical clustering since this technique has been used in practice to approximate spatial clusters with similar wildfire properties (Rodrigues et al., 2019a,b; Rahimi et al., 2020). The clustering procedure is as follows:

1. Standardise the auxiliary variable data (temperature or precipitation) for every location in $\mathcal{N}_i$.

Table D.1.1: Total scores obtained using neighbourhood $\mathcal{N}_i^t$.

Distance	$k_y = 1$	$k_y = 2$	$k_y = 3$	$k_y = 4$	$k_y = 5$	$k_y = 6$
$k_1 = 50$	8393.3	8089.3	8003.6	7991.2	7925.9	7909.3
$k_1 = 75$	8225.5	8134.2	8161.9	8187.4	8162.6	8159.6
$k_1 = 100$	8295.6	8222	8264.4	8293.9	8275.5	8274.1
$k_1 = 125$	8439.9	8401.4	8457	8487.2	8474.1	8474
$k_1 = 150$	8547	8519	8578.2	8608.6	8599.5	8600.6
$k_1 = 175$	8624.8	8604.8	8664.5	8694.1	8687.5	8689
$k_1 = 200$	8711.2	8696.7	8755.2	8784.2	8778.4	8781.5
$k_1 = 225$	8779.8	8768.9	8823.8	8852.3	8848	8851.2
$k_1 = 250$	8843.9	8838.3	8892	8919.1	8915.4	8919.4

2. Apply hierarchical clustering, using the standardised covariate data, to obtain two clusters, i.e., for each $j \in \mathcal{N}_i$, $c_j = 1$ or 2.

3. Compute the subset of locations with the same cluster assignment as location i, i.e., $\{j \in \mathcal{N}_i \mid c_j = c_i\}$.

In our analysis, we found that both the neighbourhoods $\mathcal{N}_i^t$ and $\mathcal{N}_i^c$ resulted in worse prediction scores compared to the simpler spatial neighbourhood approach outlined in Section 7.3 of Chapter 7. This is illustrated by the results in Tables D.1.1 and D.1.2, where we present the overall prediction scores, as outlined in Opitz (2023), for $\mathcal{N}_i^t$ and $\mathcal{N}_i^c$, respectively; recall that we aim to minimise this score. For these scores, we let $k_1^{CNT} = k_1^{BAP} := k_1 \in \{50, 75, \ldots, 250\}$ to incorporate a variety of spatial distances and set $k_2^{BAP} = 0.5$ to match the existing selected tuning parameter from Section 7.3.4 of Chapter 7. For $\mathcal{N}_i^t$, we let $k_y^{CNT} = k_y^{BAP} := k_y \in \{1, 2, 3, 4, 5, 6\}$, resulting in time windows of up to 13 years. Note that these scores correspond to the final prediction scores, i.e., when the missing data are known, and in practice, one would need to select the tuning parameters using the cross validation procedure outlined in Section 7.3.4 of Chapter 7.

Table D.1.2: Total scores obtained using neighbourhood $\mathcal{N}_i^c$ with clusters computed using temperature and precipitation.

Distance	Temperature clusters	Precipitation clusters
$k_1 = 50$	11149	11161
$k_1 = 75$	9608.8	9607.7
$k_1 = 100$	9245.2	9309.2
$k_1 = 125$	8651.9	8718.5
$k_1 = 150$	8522.9	8582.1
$k_1 = 175$	8431.5	8505.2
$k_1 = 200$	8460.3	8489.5
$k_1 = 225$	8470.2	8511.1
$k_1 = 250$	8493.6	8539.5

One can observe that all the prediction scores from Tables D.1.1 and D.1.2 exceed the final score obtained using the method described in Section 7.3 of Chapter 7, and in many cases, the scores obtained were significantly worse. The fact that these predictions were worse for a wide range of tuning parameter combinations gives further support to our main modelling approach.

In the case of temporal neighbourhoods, since the predictive scores do not tend to decrease with the parameter k_y, our results suggest that marginal wildfire behaviour can vary significantly over neighbouring years. Therefore, even though incorporating information from neighbouring years increases the amount of data available for model fitting, it does not appear to improve the quality of marginal estimates.

In the case of cluster-based neighbourhoods, our results indicate that observations with similar temperature and precipitation values may not be those with similar wildfire behaviour. We suspect this may occur due to the complex nature of the relationships between the wildfire and auxiliary variables described in Section 7.4 of Chapter 7. Such relationships are unlikely to be picked up by incorporating this additional clustering step. Clustering also reduces the amount of data available for model fitting, which also

appears to reduce the quality of marginal estimates.

On the whole, these results suggest that incorporating additional information, both from temporal windows and covariate-based clusters, does not improve the quality of marginal estimates for either CNT or BAP under our modelling approach. Combined with the principle of parsimony, we do not consider these alternative neighbourhoods further.

Appendix E

Supplementary material for Chapter 8: Extreme value methods for estimating rare events in Utopia

In this chapter, we present additional figures to support the work of Chapter 8 that concerns the Extreme Value Analysis (EVA) conference in 2023. All notation and terminology follow from that introduced in Chapter 8.

E.1 Additional figures for Section 8.3

In this section, we present additional figures for Section 8.3 of Chapter 8, concerned with challenges C1 and C2. Figures E.1.1-E.1.4 support the exploratory analysis for challenges C1 and C2. We explore the within-year seasonality of the response variable Y in Figure E.1.1, looking at the distribution of Y per month and across the two seasons. This shows that there is a significant difference in the distribution of Y between seasons 1 and 2, but within each season there is little difference across months.

Figure E.1.2 shows a scatter plot of Y against each covariate $V_1, \ldots, V_8$, excluding V_6 which corresponds to season. Covariates V_1, V_2 and V_8 do not seem to have a relationship

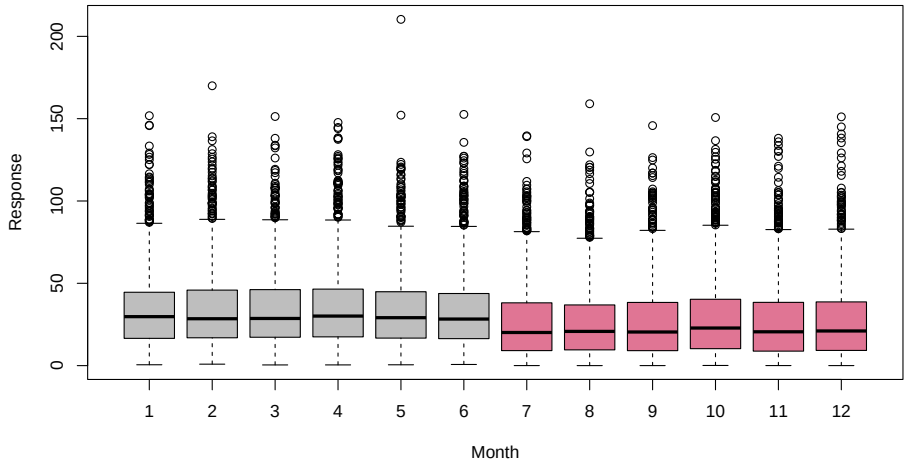


Figure E.1.1: Box plot of the response variable Y with each month and season (season 1 in grey and season 2 in red).

with Y, whilst there appears to be dependence for the remaining covariates. These observed relationships are complex and non-linear.

Figure E.1.3 shows wind rose plots before and after the changepoint we find for V_6 (wind speed) and V_7 (wind direction). These graphs demonstrate the direction of wind observed, as well as the speeds for each direction (divided into eight sectors). We find that wind speeds are greater before the changepoint, and mostly come from the south-westerly direction. Whilst after the changepoint, winds tend to come from the north-easterly direction and with lower speeds.

Figure E.1.4 details the auto-correlation function (acf) values for the response Y and explanatory variables $V_1, \ldots, V_4, V_6, \ldots, V_8$, up to lag 60. All variables have negligible acf values beyond lag 0, except V_6 (wind speed), V_7 (wind direction) and V_8 (atmosphere).

Figure E.1.5 shows the QQ-plots corresponding to a standard GPD model fitted to the excesses of Y above a constant (left) and seasonally-varying threshold (right). 95% tolerance bounds (grey) show a lack of agreement between observations and the standard GPD model above a constant threshold. The second plot demonstrates a

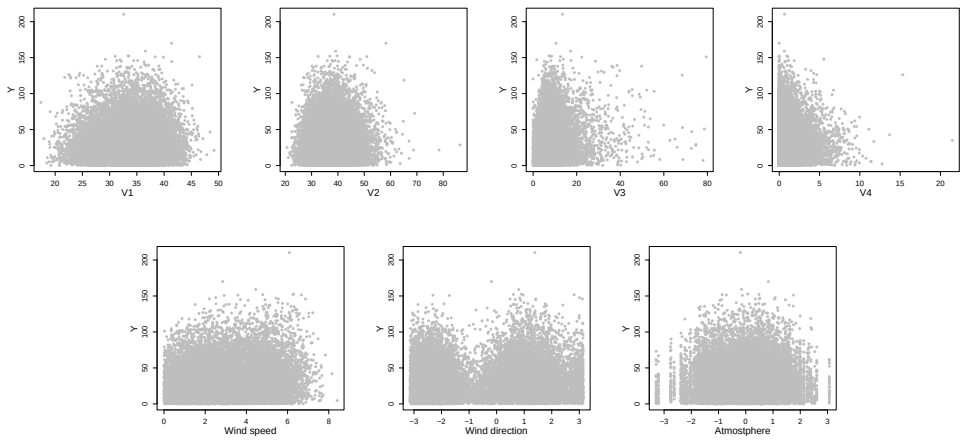


Figure E.1.2: Scatter plots of explanatory variables $V_1, \ldots, V_4$, wind speed (V_6), wind direction (V_7) and atmosphere (V_8), from top-left to bottom-right (by row), against the response variable Y.

significant improvement in model fit.

Figure E.1.6 shows a detailed summary of the pattern of missing data in the data and can be produced using the `missing_pattern` function in the `finalfit` package in R (Harrison et al., 2023). To interpret the figure note that blue and red squares represent observed and missing variables, respectively. The number on the right indicates the number of missing random variables (i.e., the number of red squares in the row), while the number on the left is the number of observations that fall into the row category. On the bottom, we have the number of observations that fall into the column category. For example, 18,545 observations are fully observed (denoted by the first row), there are 407 observations where only $V4$ is missing (denoted by the second row), there are 13 observations where both $V4$ and $V6$ are missing (denoted by the fourth row), there are 456 observations where $V4$ and at least one other predictor is missing (denoted by the last column in the table) etc. There are very few observations where more than one predictor is missing.

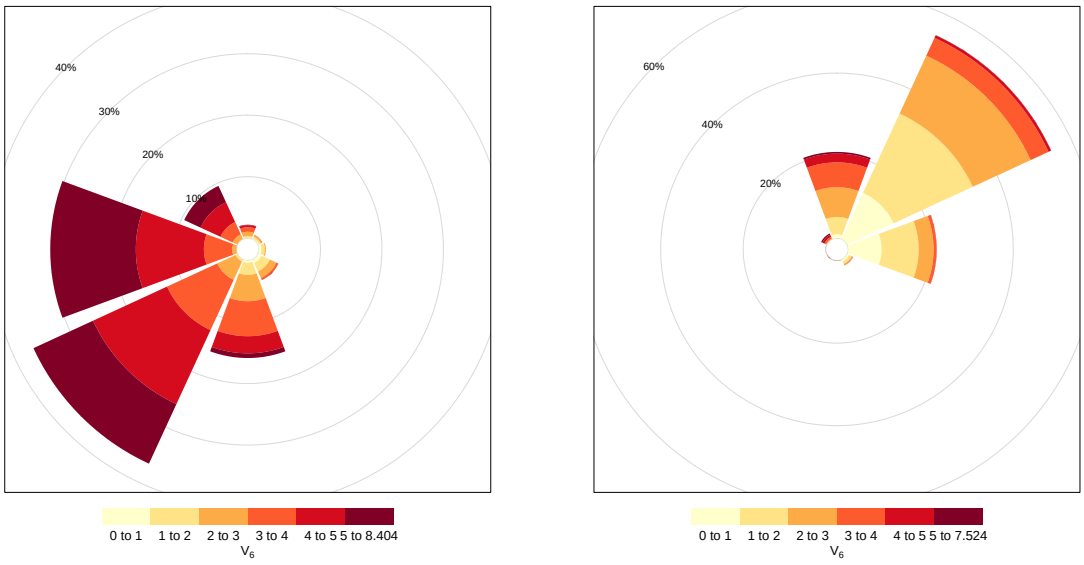


Figure E.1.3: Wind rose plot before (left) and after (right) the changepoint.

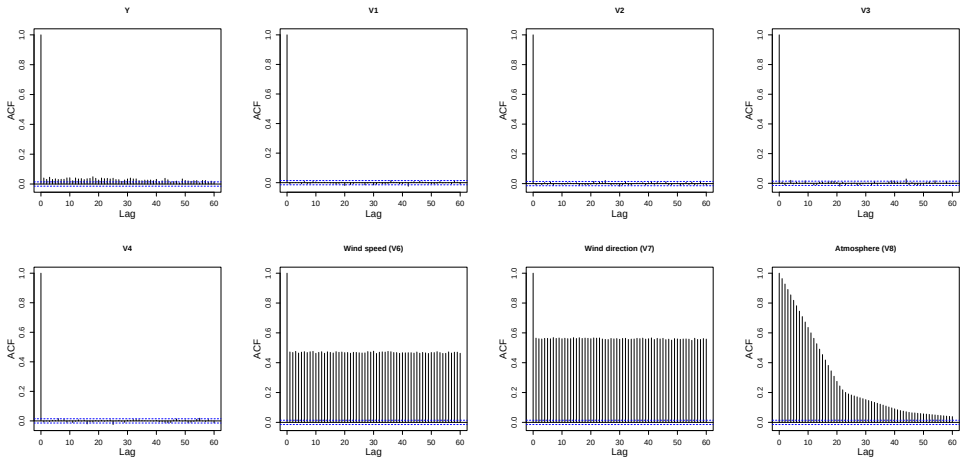


Figure E.1.4: Autocorrelation function plots for the response variable Y and explanatory variables $V1, \ldots, V4$, wind speed ($V6$), wind direction ($V7$) and atmosphere ($V8$), from top-left to bottom-right (by row), against the response variable Y.

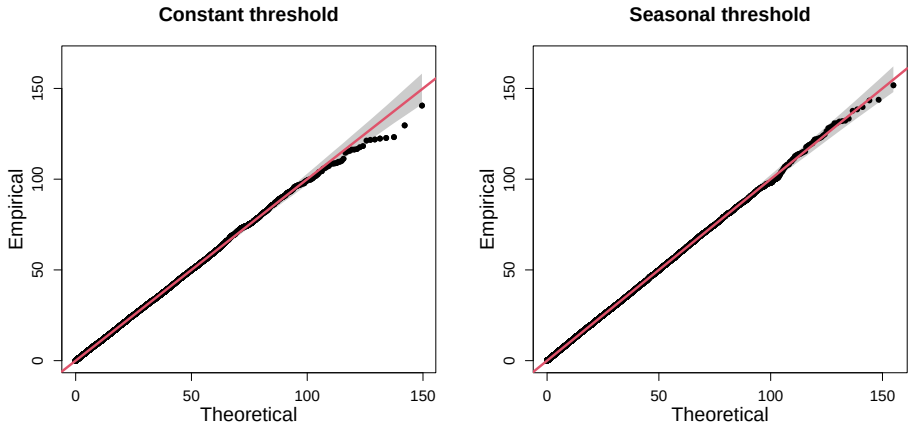


Figure E.1.5: QQ-plots showing standard GPD model fits with 95% tolerance bounds (grey) above a constant (left) and stepped-seasonal (right) threshold.

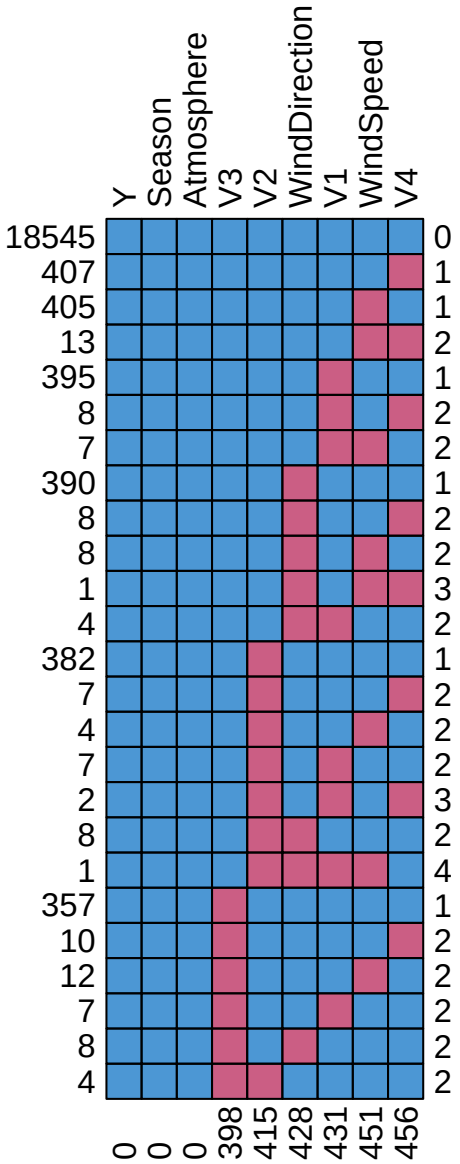


Figure E.1.6: Detailed pattern of missing predictor variables in the Amaurot data set.

E.2 Additional figures for Section 8.4

In this section, we present additional plots related to Section 8.4 of Chapter 8. Figure E.2.1 illustrates the time series of both covariates for the first 3 years of the observation period. We observe how the seasons vary periodically over each year, as well as the discrete nature of the atmospheric covariate.

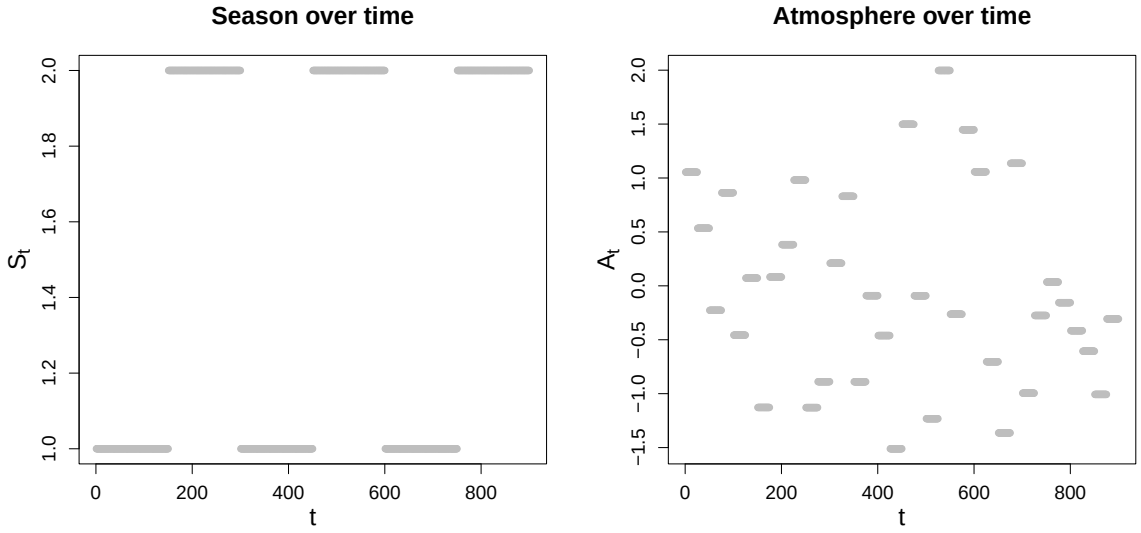


Figure E.2.1: Plots of S_t (left) and A_t (right) against t for the first 3 years of the observation period.

Bootstrapped χ estimates for the groups $G^A_{I,k}, k \in \{1, \ldots, 10\}$, $I \in \mathcal{I} \setminus \{1, 2, 3\}$ and $G^S_{I,k}, k \in \{1, 2\}$, $I \in \mathcal{I}$ are given in Figures E.2.2 - E.2.5. These estimates illustrate the impact of atmosphere on the dependence structure.

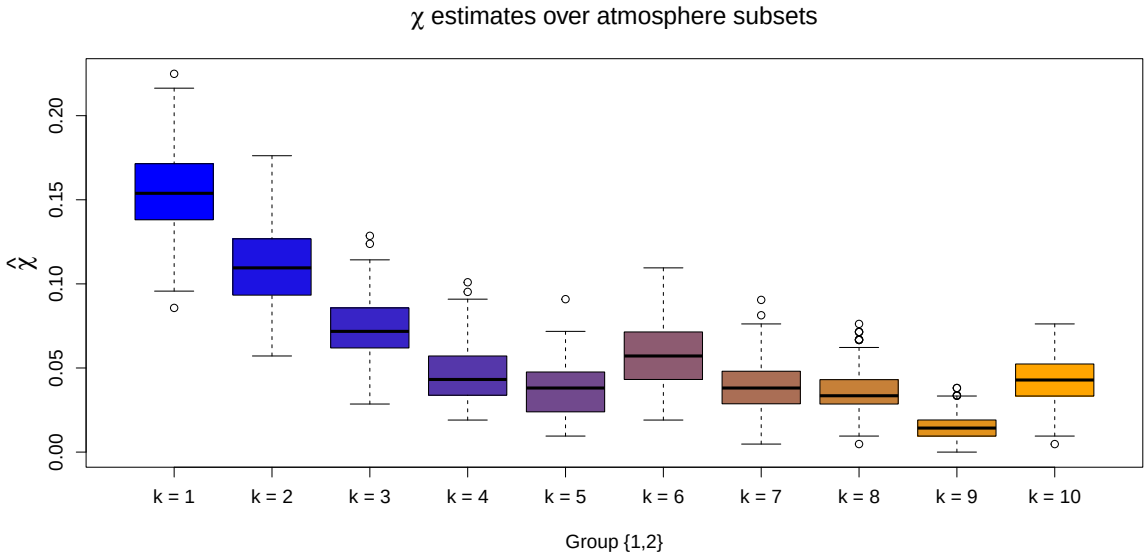


Figure E.2.2: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

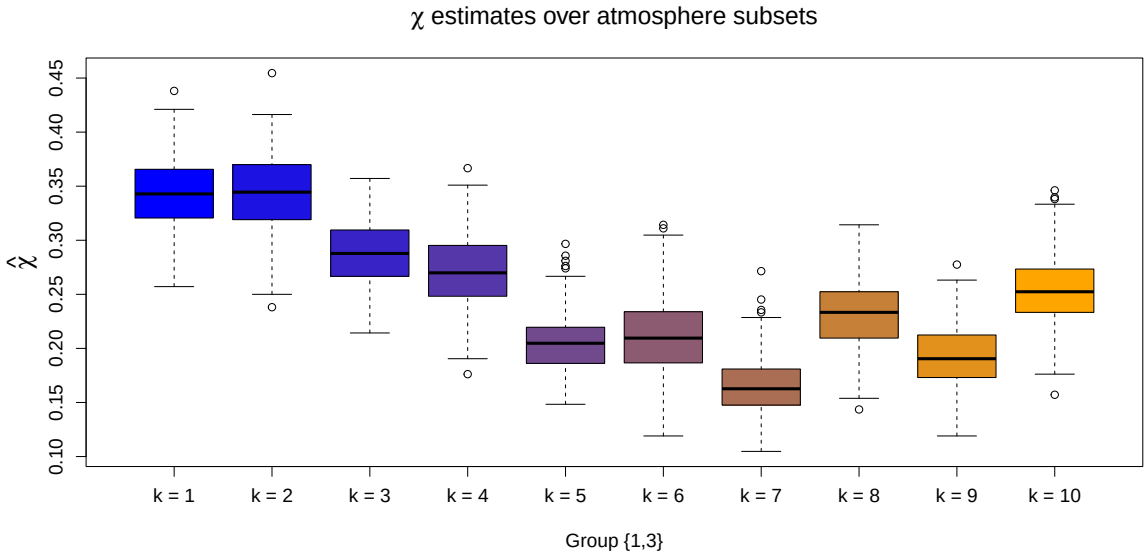


Figure E.2.3: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

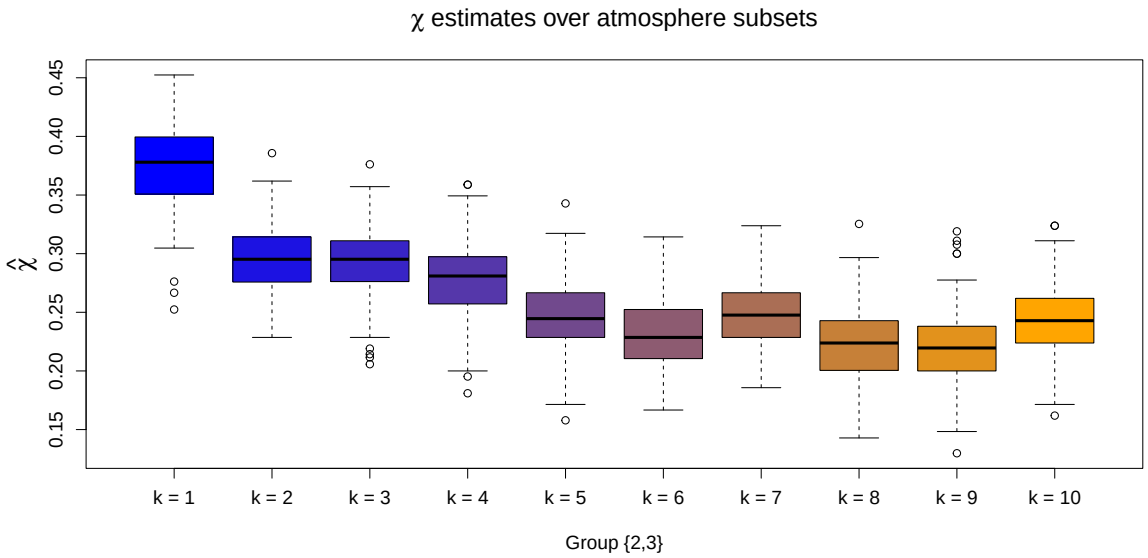


Figure E.2.4: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

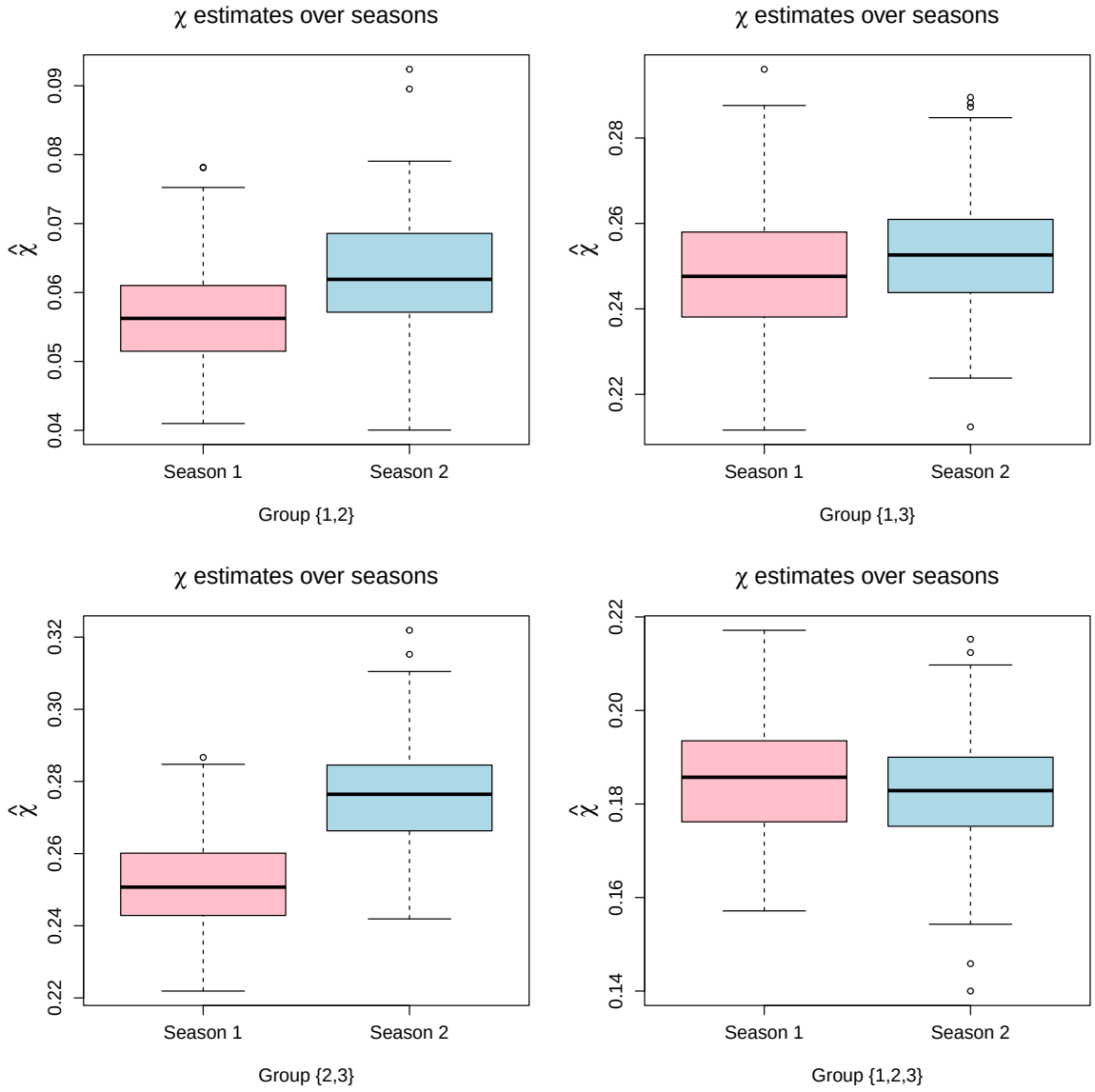


Figure E.2.5: Boxplots of empirical χ estimates obtained for the subsets $G_{I,k}^S$, with $k = 1, 2$. In each case, pink and blue colours illustrate estimates for seasons 1 and 2, respectively. From top left to bottom right: $I = \{1, 2, 3\}$, $I = \{1, 2\}$, $I = \{1, 3\}$, $I = \{2, 3\}$.

For a 3-dimensional random vector, the angular dependence function, denoted λ, is defined on the unit-simplex $\boldsymbol{S}^2$ and describes extremal dependence along different rays $\boldsymbol{\omega} \in \boldsymbol{S}^2$. As noted in Section 8.4 of the Chapter 8, we can associate each of the probabilities from C3, p_1 and p_2, with points on $\boldsymbol{S}^2$, denoted $\boldsymbol{\omega}^1$ and $\boldsymbol{\omega}^2$ respectively. With $I = \{1, 2, 3\}$, we consider $\lambda(\boldsymbol{\omega}^1)$ and $\lambda(\boldsymbol{\omega}^2)$ over the subsets $G^S_{I,k},\ k \in \{1, 2\}$ and $G^A_{I,k},\ k \in \{1, \ldots, 10\}$. We note that $\lambda(\boldsymbol{\omega}^1)$ is analogous with the coefficient of tail dependence $\eta \in (0, 1]$ (Ledford and Tawn, 1996), with $\eta = 1/3\lambda(\boldsymbol{\omega}^1)$; this corresponds with the region where all variables are simultaneously extreme. Furthermore, $\lambda(\boldsymbol{\omega}^2)$, which corresponds to a region where only two variables are extreme, is only evaluated after an additional marginal transformation of the third Coputopia time series; see Section 8.4.2 of Chapter 8.

Estimation of λ for each simplex point and subset was achieved using the Hill estimator (Hill, 1975) at the 90% level, with uncertainty subsequently quantified via bootstrapping. These results are given in Figures E.2.6 - E.2.9. These plots provide further evidence of a relationship between the extremal dependence structure and the covariates.

To illustrate the estimated trend in dependence, Figure E.2.10 illustrates the estimated scale functions, $\sigma(\boldsymbol{\omega} \mid \boldsymbol{x}_t)$, over atmosphere for parts 1 and 2. Under the assumption of asymptotic normality in the spline coefficients, 95% confidence intervals are obtained via posterior sampling; see Wood (2017) for more details. We observe that σ tends to increase and decrease over atmosphere for parts 1 and 2, respectively, although the trend is less pronounced for the latter. Under our modelling framework, we note that higher values of σ are associated with less positive extremal dependence in the direction $\boldsymbol{\omega}$ of interest; to see this, observe that the survivor function of the GPD with fixed ξ is negatively associated with σ. Considering the trend in $\sigma(\boldsymbol{\omega} \mid \boldsymbol{x}_t)$, our results indicate a decrease in dependence in the region where all variables are extreme.

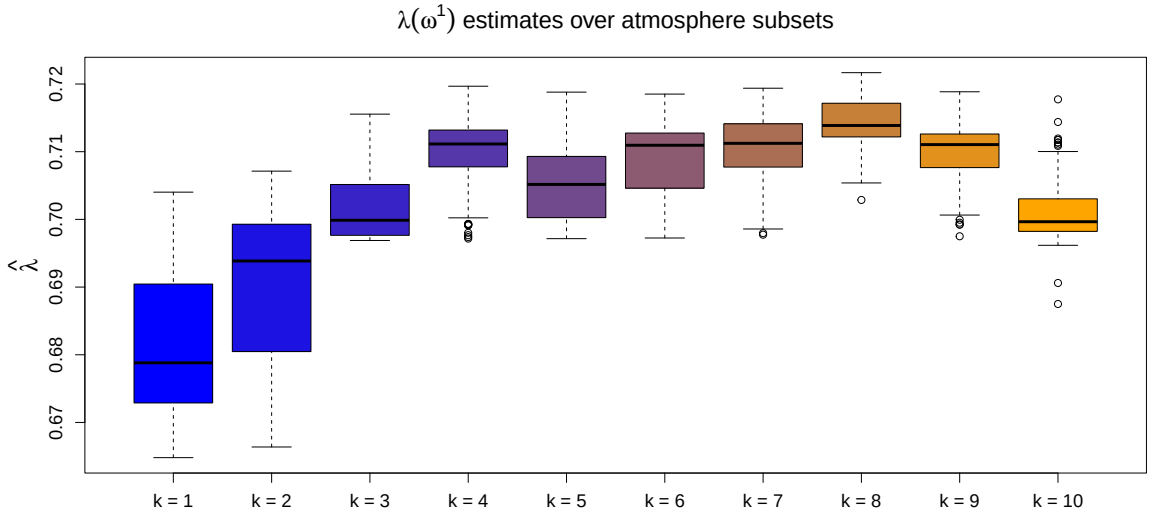


Figure E.2.6: Boxplots of empirical $\lambda(\boldsymbol{\omega}_i)$ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in λ estimates as the atmospheric values are increased.

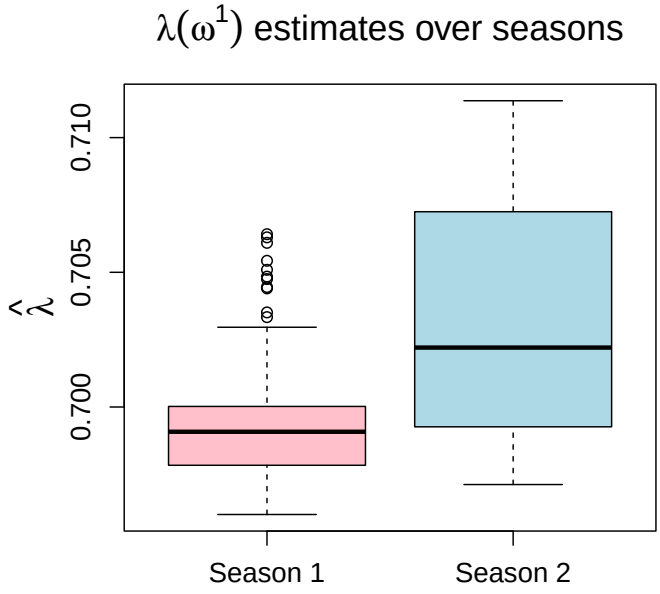


Figure E.2.7: Boxplots of empirical $\lambda(\boldsymbol{\omega}_i)$ estimates obtained for the subsets $G^S_{I,k}$, with $k = 1, 2$ and $I = \{1, 2, 3\}$. In each case, pink and blue colours illustrate estimates for seasons 1 and 2, respectively.

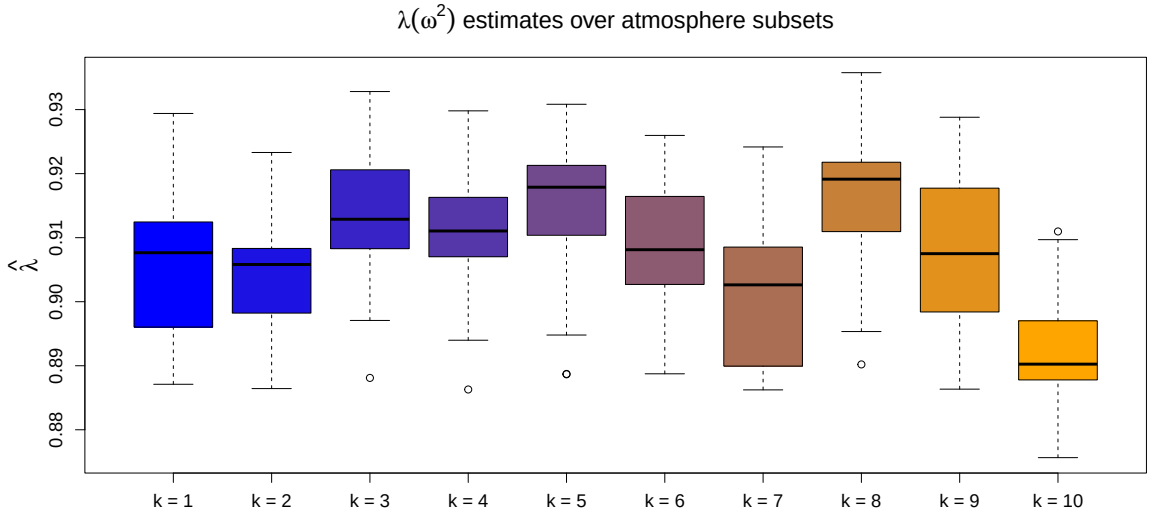


Figure E.2.8: Boxplots of empirical $\lambda(\boldsymbol{\omega}_{ii})$ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in λ estimates as the atmospheric values are increased.

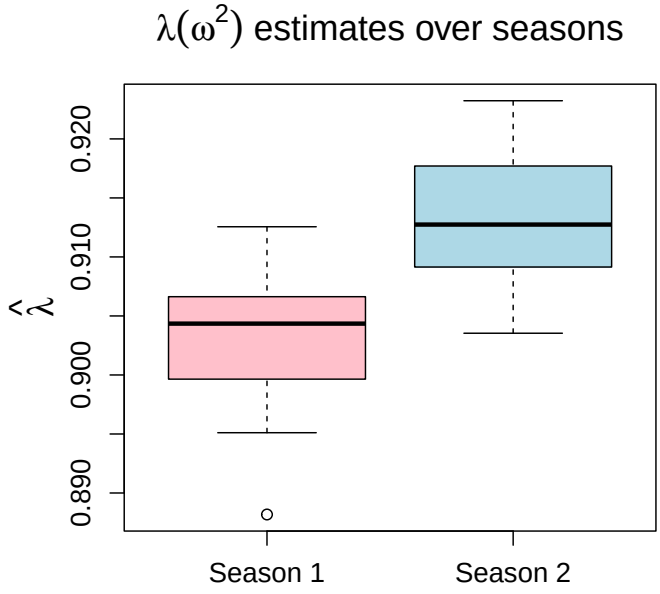


Figure E.2.9: Boxplots of empirical $\lambda(\boldsymbol{\omega}_{ii})$ estimates obtained for the subsets $G^S_{I,k}$, with $k = 1, 2$ and $I = \{1, 2, 3\}$. In each case, pink and blue colours illustrate estimates for seasons 1 and 2, respectively.

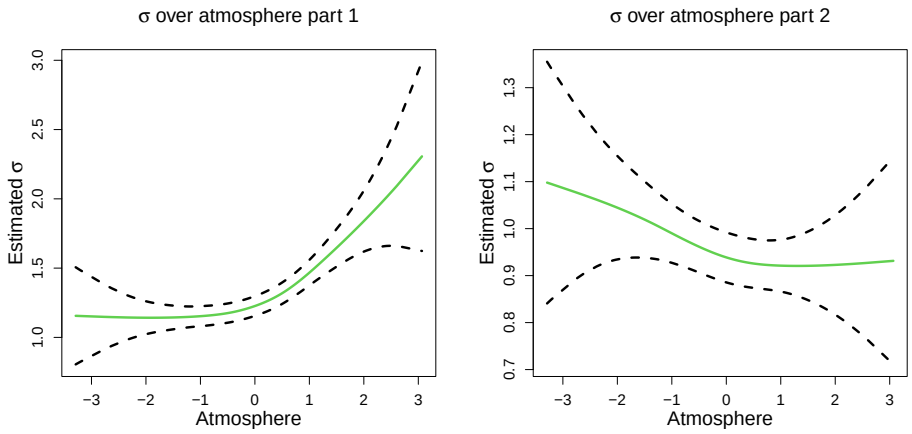


Figure E.2.10: Estimated σ functions (green) over atmosphere for part 1 (left) and 2 (right). In both cases, the regions defined by the black dotted lines represent 95% confidence intervals obtained using posterior sampling.

E.3 Additional figures for Section 8.5

Figure E.3.1 shows a heat map of empirically estimated $\eta(\cdot)$ dependence coefficients and provides further evidence of the existence of the 5 dependence subgroups identified in our exploratory analysis for challenge C4. It also suggests that between group independence as well as within group asymptotic independence – in the sense that the extremes of within group variables do not occur simultaneously – are both reasonable modelling assumptions.

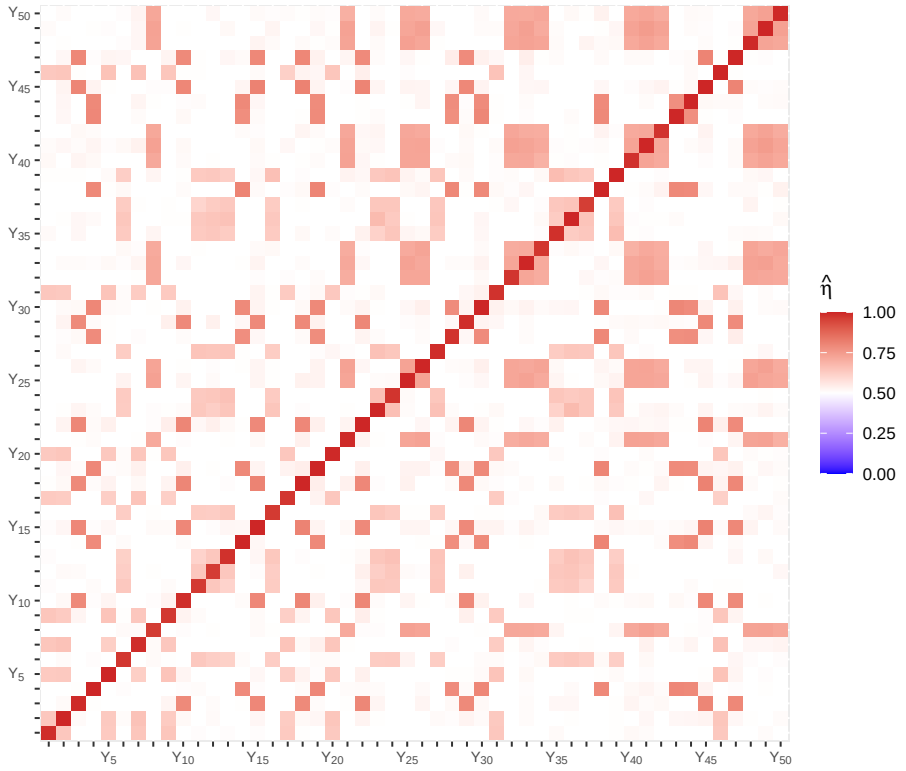


Figure E.3.1: Heat map of estimated empirical pairwise $\eta(u)$ extremal dependence coefficients with $u = 0.95$.

Figure E.3.2 shows the bootstrapped estimated individual group and overall probabilities with respect to conditioning threshold quantile for part 2 of challenge C4.

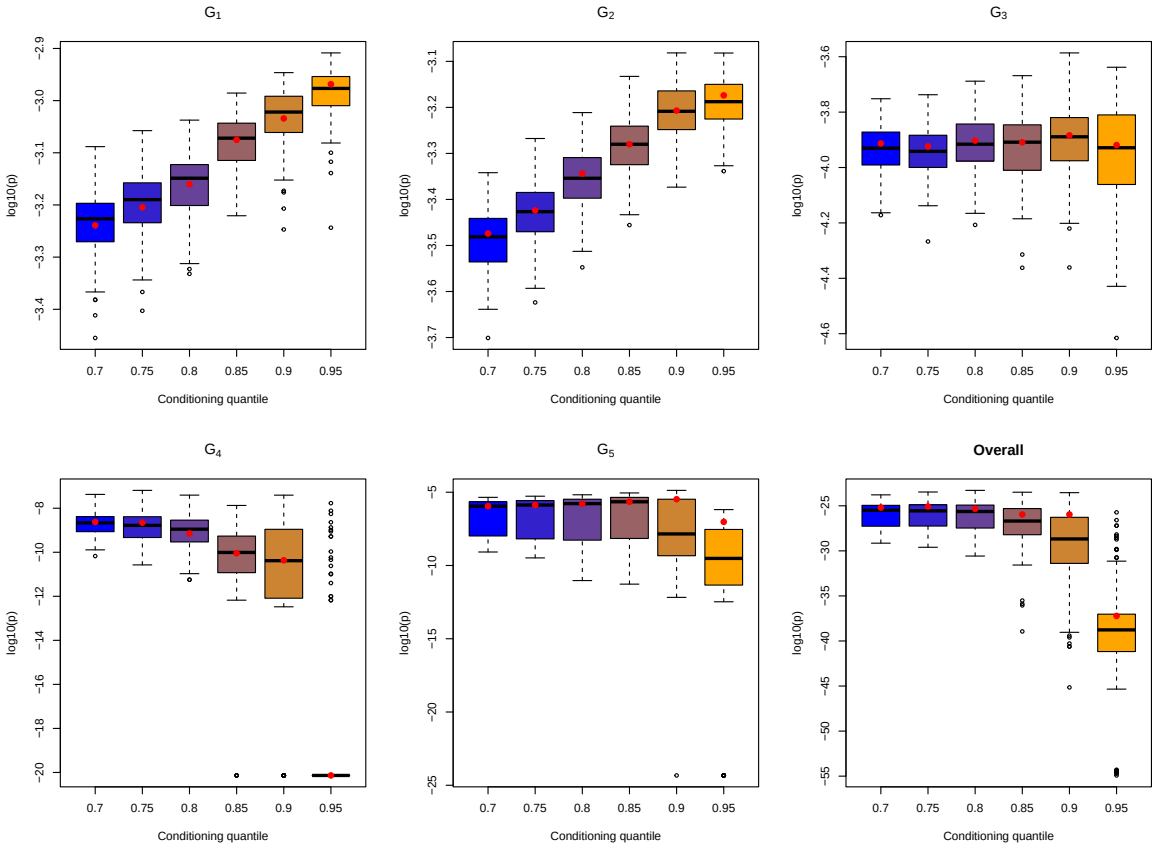


Figure E.3.2: Part 2 subgroup and overall bootstrapped probability estimates on the log scale for C4. The red points indicate the original sample estimates and the colouring of the boxplots indicates the choice of conditioning threshold, with the conditioning quantile indices 1-6 referring to the quantile levels $\{0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$, respectively.

Bibliography

Araújo, I. B. and Pugh, D. T. (2008). Sea levels at Newlyn 1915–2005: analysis of trends for future flooding risks. *Journal of Coastal Research*, 24:203–212.

Asadi, P., Engelke, S., and Davison, A. C. (2018). Optimal regionalization of extreme value distributions for flood estimation. *Journal of Hydrology*, 556:182–193.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

Baranes, H., Woodruff, J., Talke, S., Kopp, R., Ray, R., and DeConto, R. (2020). Tidally driven interannual variation in extreme sea level frequencies in the Gulf of Maine. *Journal of Geophysical Research: Oceans*, 125(10).

Bashtannyk, D. M. and Hyndman, R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298.

Batstone, C., Lawless, M., Tawn, J. A., Horsburgh, K., Blackman, D., McMillan, A., Worth, D., Laeger, S., and Hunt, T. (2013). A UK best-practice approach for extreme sea-level analysis along complex topographic coastlines. *Ocean Engineering*, 71:28–39.

Bernardara, P., Andreewsky, M., and Benoit, M. (2011). Application of regional frequency analysis to the estimation of extreme storm surges. *Journal of Geophysical Research: Oceans*, 116.

BODC (2020). National Oceanographic Database. https://www.bodc.ac.uk/data/bodc_database/nodb/.

Box, G. E. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Budic, L., Didenko, G., and Dormann, C. F. (2016). Squares of different sizes: effect of geographical projection on model parameter estimates in species distribution modeling. *Ecology and Evolution*, 6(1):202–211.

Calafat, F. M., Wahl, T., Tadesse, M. G., and Sparrow, S. N. (2022). Trends in Europe storm surge extremes match the rate of sea-level rise. *Nature*, 603(7903):841–845.

Carter, D. and Challenor, P. (1981). Estimating return values of environmental parameters. *Quarterly Journal of the Royal Meteorological Society*, 107(451):259–266.

Chatfield, C. (2013). *The Analysis of Time Series: Theory and Practice*. Springer, New York.

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

Cisneros, D., Gong, Y., Yadav, R., Hazra, A., and Huser, R. (2023). A combined statistical and machine learning approach for spatial prediction of extreme wildfire frequencies and sizes. *Extremes*, 26(2):301–330.

Cohen, J. D. and Deeming, J. E. (1985). The national fire-danger rating system: basic equations. https://www.fs.usda.gov/research/treesearch/27298.

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

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

Coles, S. G. and Tawn, J. A. (1990). Statistics of coastal flood prevention. *Philosophical Transactions of the Royal Society of London: Series A*, 332(1627):457–476.

Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society: Series C*, 45(4):463–478.

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

Committee on Climate Change (2018). Managing the coast in a changing climate. https://www.theccc.org.uk/publication/managing-the-coast-in-a-changing-climate/. Accessed 23/02/24.

Crockett, J. L. and Westerling, A. L. (2018). Greater temperature and precipitation extremes intensify Western U.S. droughts, wildfire severity, and Sierra Nevada tree mortality. *Journal of Climate*, 31(1):341–354.

Dale, M., Wicks, J., Mylne, K., Pappenberger, F., Laeger, S., and Taylor, S. (2014). Probabilistic flood forecasting and decision-making: an innovative risk-based approach. *Natural hazards*, 70:159–172.

D'Arcy, E., Murphy-Barltrop, C. J., Shooter, R., and Simpson, E. S. (2023a). A marginal modelling approach for predicting wildfire extremes across the contiguous United States. *Extremes*, 26(2):381–398.

D'Arcy, E. and Tawn, J. A. (2021). Discussion of "towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges" (by Howard, T. and Williams, S. D. P.). *Natural Hazards and Earth System Sciences Discussions*, 21(12):3693–3712.

D'Arcy, E., Tawn, J. A., Joly, A., and Sifnioti, D. E. (2023b). Accounting for seasonality in extreme sea-level estimation. *The Annals of Applied Statistics*, 17(4):3500–3525.

D'Arcy, E., Tawn, J. A., Joly-Laugel, A., and Sifnioti, D. E. (2023c). Supplement to "accounting for seasonality in extreme sea level estimation".

D'Arcy, E., Tawn, J. A., and Sifnioti, D. E. (2022). Accounting for climate change in extreme sea level estimation. *Water*, 14(19):2956.

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

Davis, R. A. and Resnick, S. I. (1989). Basic properties and prediction of max-ARMA processes. *Advances in Applied Probability*, 21(4):781–803.

Davison, A. C., Huser, R., and Thibaud, E. (2019). Spatial extremes. In *Handbook of Environmental and Ecological Statistics*, pages 711–744. Chapman and Hall/CRC.

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*, 52(3):393–425.

Dixon, M. J. and Tawn, J. A. (1994). Extreme sea levels: modelling interaction between tide and surge. *Statistics for the Environment 2: Water Related Issues*.

Dixon, M. J. and Tawn, J. A. (1999). The effect of non-stationarity on extreme sea-level estimation. *Journal of the Royal Statistical Society: Series C*, 48(2):135–151.

Dixon, M. J., Tawn, J. A., and Vassie, J. M. (1998). Spatial modelling of extreme sea-levels. *Environmetrics*, 9(3):283–301.

Duane, A., Castellnou, M., and Brotons, L. (2021). Towards a comprehensive look at global drivers of novel extreme wildfire events. *Climatic Change*, 165:43.

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*, 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.

Egbert, G. D. and Ray, R. D. (2017). Tidal prediction. *Journal of Marine Research*, 75(3):189–237.

Émile, M. and Gumbel, J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Annales de l'ISUP*, 9(2):171–173.

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8(1):241–270.

Enríquez, A. R., Wahl, T., Baranes, H. E., Talke, S. A., Orton, P. M., Booth, J. F., and Haigh, I. D. (2022). Predictable changes in extreme sea levels and coastal flood risk due to long-term tidal cycles. *Journal of Geophysical Research: Oceans*, 127(4).

Environment Agency (2018). Coastal Flood Boundary Conditions for the UK: Update 2018. Technical summary report. https://environment.data.gov.uk/dataset/6e856bda-0ca9-404f-93d7-566a2378a7a8. Accessed 01/10/21.

Environment Agency (2021). Thames Estuary 2100: 10-year review - technical monitoring report. https://www.gov.uk/government/publications/thames-estuary-2100-te2100-monitoring-reviews/thames-estuary-2100-10-year-monitoring-review-2021. Accessed 14/01/24.

Environment Agency (2023a). Daily mean flow (m3/s) time series for Kingston. https://environment.data.gov.uk/hydrology/station/8496ce69-482c-406a-a2f0-ac418ef8f099.

Environment Agency (2023b). The Thames Barrier. https://www.gov.uk/guidance/the-thames-barrier#:~:text=The%20Thames%20Barrier%20has%20been,against%20combined%20tidal%2Ffluvial%20flooding. Accessed 23/02/24.

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

Ferreira, M. (2011). On tail dependence: a characterization for first-order max-autoregressive processes. *Mathematical Notes*, 90:882–893.

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

Gilleland, E. and Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8):1–39.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151.

Graff, J. (1978). Concerning the recurrence of abnormal sea levels. *Coastal Engineering*, 2:177–187.

Guerrero, M. B., Huser, R., and Ombao, H. (2023). Conex–connect: learning patterns in extremal brain connectivity from multichannel eeg data. *The Annals of Applied Statistics*, 17(1):178–198.

Haigh, I. D., Nicholls, R., and Wells, N. (2010). A comparison of the main methods for estimating probabilities of extreme still water levels. *Coastal Engineering*, 57(9):838–849.

Harrison, E., Drake, T., and Ots, R. (2023). *finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling*. R package version 1.0.7.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer, New York.

Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes*, 3:279–290.

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*, 66(3):497–546.

Heffernan, J. E., Tawn, J. A., and Zhang, Z. (2007). Asymptotically (in) dependent multivariate maxima of moving maxima processes. *Extremes*, 10:57–82.

Hendry, A., Haigh, I. D., Nicholls, R. J., Winter, H., Neal, R., Wahl, T., Joly-Laugel, A., and Darby, S. E. (2019). Assessing the characteristics and drivers of compound flooding events around the UK coast. *Hydrology and Earth System Sciences*, 23(7):3117–3139.

Hijmans, R. J. (2019). *geosphere: Spherical Trigonometry*. R package version 1.5-10.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.

Hitz, A. S., Davis, R. A., and Samorodnitsky, G. (2017). Discrete extremes. *arXiv pre-print: 1707.05033*.

Holden, Z. A., Swanson, A., Luce, C. H., Jolly, W. M., Maneta, M., Oyler, J. W., Warren, D. A., Parsons, R., and Affleck, D. (2018). Decreasing fire season precipitation increased recent western US forest wildfire activity. *Proceedings of the National Academy of Sciences*, 115(36).

Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, Cambridge; New York.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Howard, T. and Williams, S. D. P. (2021). Towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges. *Natural Hazards and Earth System Sciences*, 21(12):3693–3712.

Hsing, T., Hüsler, J., and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78(1):97–112.

Huser, R. and Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of agricultural, biological, and environmental statistics*, 21(3):470–491.

Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(1):e1537.

International Atomic Energy Agency (2015). The Fukushima Daiichi Accident: Report by the Director General. https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1710-ReportByTheDG-Web.pdf. Accessed 23/01/24.

Ivek, T. and Vlah, D. (2023). Reconstruction of incomplete wildfire data using deep generative models. *Extremes*, 26:251–271.

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

Joe, H. (2014). *Dependence Modeling with Copulas*. CRC press.

Jonathan, P., Randell, D., Wu, Y., and Ewans, K. (2014). Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88:520–532.

Jones, M. W., Smith, A., Betts, R., Canadell, J. G., Prentice, I. C., and Le Quéré, C. (2020). Climate change increases risk of wildfires. *ScienceBrief Review*, 116:117–120.

Joseph, M. B., Rossi, M. W., Mietkiewicz, N. P., Mahood, A. L., Cattau, M. E., St. Denis, L. A., Nagy, R. C., Iglesias, V., Abatzoglou, J. T., and Balch, J. K. (2019). Spatiotemporal prediction of wildfire size extremes with Bayesian finite sample maxima. *Ecological Applications*, 29(6):e01898.

Kaczmarska, J. M., Isham, V. S., and Northrop, P. (2015). Local generalised method of moments: an application to point process-based rainfall models. *Environmetrics*, 26(4):312–325.

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.

Kelly, K. and Šavrič, B. (2021). Area and volume computation of longitude–latitude grids and three-dimensional meshes. *Transactions in GIS*, 25(1):6–24.

Killick, R. and Eckley, I. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19.

Koh, J. (2023). Gradient boosting with extreme-value theory for wildfire prediction. *Extremes*, 26(2):273–299.

Koh, J., Pimont, F., Dupuy, J.-L., and Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics*, 17(1):560–582.

Krawchuk, M. A., Moritz, M. A., Parisien, M.-A., Van Dorn, J., and Hayhoe, K. (2009). Global pyrogeography: the current and future distribution of wildfire. *PloS one*, 4(4):e5102.

Kyselý, J., Picek, J., and Beranová, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72(1-2):55–68.

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.

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

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

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1):29–52.

Masselink, G. and Russell, P. (2007). Coastal erosion and coastal geomorphology. *MCCIP (Marine Climate Group Impacts Partnership) Annual Report Card*, 2008.

Masselink, G., Russell, P., Rennie, A., Brooks, S., and Spencer, T. (2020). Impacts of climate change on coastal geomorphology and coastal erosion relevant to the coastal and marine environment around the UK. *MCCIP Science Review*, 2020:158–189.

Menéndez, M. and Woodworth, P. L. (2010). Changes in extreme high water levels based on a quasi-global tide-gauge data set. *Journal of Geophysical Research: Oceans*, 115.

Mhalla, L., Opitz, T., and Chavez-Demoulin, V. (2019). Exceedance-based nonlinear regression of tail dependence. *Extremes*, 22:523–552.

Mooyaart, L. and Jonkman, S. N. (2017). Overview and design considerations of storm surge barriers. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 143(4):06017001.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R. (2021). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, 126(3).

Murphy, C., Tawn, J. A., and Varty, Z. (2023). Automated threshold selection and associated inference uncertainty for univariate extremes. *arXiv*, 2310.17999.

Murphy-Barltrop, C. J. R. and Wadsworth, J. L. (2022). Modelling non-stationarity in asymptotically independent extremes. *arXiv*, 2203.05860.

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2023). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *arXiv*, 2303.13237.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York.

NOC (2021). National Tidal and Sea Level Facility. https://www.ntslf.org/.

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*, 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.

Northrop, P. J., Jonathan, P., and Randell, D. (2016). Threshold modeling of non-stationary extremes. In *Extreme Value Modeling and Risk Analysis*, pages 107–128. Chapman and Hall/CRC.

O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *The Annals of Probability*, 15(1):281–291.

Office for Nuclear Regulation (2014). Safety Assessment Principles - 2014 edition (Revision 1, January 2020). https://www.onr.org.uk/saps/saps2014.pdf. Accessed 23/01/24.

Office for Nuclear Regulation (2021). NS-TAST-GD-013 – Annex 2 - Meteorological Hazards. https://www.onr.org.uk/consultations/2021/external-hazards/ns-tast-gd-013-annex-2.pdf. Accessed 23/01/24.

Opitz, T. (2023). Editorial: EVA 2021 data challenge on spatiotemporal prediction of wildfire extremes in the USA. *Extremes*, 26:241–250.

Parey, S., Hoang, T. T. H., and Dacunha-Castelle, D. (2013). The importance of mean and variance in predicting changes in temperature extremes. *Journal of Geophysical Research: Atmospheres*, 118(15):8285–8296.

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

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Pollard, J., Spencer, T., and Brooks, S. (2019). The interactive relationship between coastal erosion and flood risk. *Progress in Physical Geography: Earth and Environment*, 43(4):574–585.

Prandle, D. and Wolf, J. (1978). Surge-tide interaction in the southern North Sea. *Elsevier Oceanography Series*, 23:161–185.

Preisler, H. K., Brillinger, D. R., Burgan, R. E., and Benoit, J. (2004). Probability based models for estimation of wildfire risk. *International Journal of wildland fire*, 13(2):133–142.

Pugh, D. and Vassie, J. (1978). Extreme sea levels from tide and surge probability. *Coastal Engineering*, 16:911–930.

Pugh, D. and Woodworth, P. (2014). *Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes*. Cambridge University Press.

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research*, 55:1890–1911.

R Core Team (2016). *stats: R statistical functions*. R package version 3.3.1.

Rahimi, S., Sharifi, Z., and Mastrolonardo, G. (2020). Comparative study of the effects of wildfire and cultivation on topsoil properties in the Zagros forest, Iran. *Eurasian Soil Science*, 53:1655–1668.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336.

Richards, J. and Huser, R. (2022). A unifying partially-interpretable framework for neural network-based extreme quantile regression. *arXiv preprint: 2208.07581*.

Richards, J., Tawn, J. A., and Brown, S. (2022). Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, 16(4):2693–2713.

Robinson, M. E. and Tawn, J. A. (1997). Statistics for extreme sea currents. *Journal of the Royal Statistical Society: Series C*, 46(2):183–205.

Robinson, M. E. and Tawn, J. A. (2000). Extremal analysis of processes sampled at different frequencies. *Journal of the Royal Statistical Society, Series B*, 62(1):117–135.

Rodrigues, M., Costafreda-Aumedes, S., Comas, C., and Vega-García, C. (2019a). Spatial stratification of wildfire drivers towards enhanced definition of large-fire regime zoning and fire seasons. *Science of The Total Environment*, 689:634–644.

Rodrigues, M., González-Hidalgo, J. C., Peña-Angulo, D., and Jiménez-Ruano, A. (2019b). Identifying wildfire-prone atmospheric circulation weather types on mainland Spain. *Agricultural and Forest Meteorology*, 264:92–103.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1988). A point process model for rainfall: further developments. *Proceedings of the Royal Society of London. A*, 417(1853):283–298.

Rodríguez-Pérez, J. R., Ordóñez, C., Roca-Pardiñas, J., Vecín-Arias, D., and Castedo-Dorado, F. (2020). Evaluating lightning-caused fire occurrence using spatial generalized additive models: a case study in central Spain. *Risk Analysis*, 40(7):1418–1437.

Rohrbeck, C., Simpson, E. S., and Tawn, J. A. (2023). Editorial: EVA 2023 data challenge. *Extremes*, (to appear).

Rohrbeck, C. and Tawn, J. A. (2021). Bayesian spatial clustering of extremal behavior for hydrological variables. *Journal of Computational and Graphical Statistics*, 30(1):91–105.

Rokach, L. and Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, New York.

Rootzén, H. and Katz, R. W. (2013). Design life level: quantifying risk in a changing climate. *Water Resources Research*, 49(9):5964–5972.

Ross, E., Sam, S., Randell, D., Feld, G., and Jonathan, P. (2018). Estimating surge in extreme North Sea storms. *Ocean Engineering*, 154:430–444.

Sá, A. C. L., Turkman, M. A. A., and Pereira, J. M. C. (2018). Exploring fire incidence in Portugal using generalized additive models for location, scale and shape (gamlss). *Modeling Earth Systems and Environment*, 4(1):199–220.

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

Schlather, M. (2001). Limit distributions of norms of vectors of positive iid random variables. *The Annals of Probability*, 29(2):862–881.

Schneider, L. F., Krajina, A., and Krivobokova, T. (2021). Threshold selection in univariate extreme value analysis. *Extremes*, 24:881–913.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X. (2012). *Changes in climate extremes and their impacts on the natural physical environment*, book section 3, pages 109–230. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B. (2021). *Weather and Climate Extreme Events in a Changing Climate.*, book section 11, page 1513–1766. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sharkey, P. and Winter, H. C. (2019). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *Environmetrics*, 30(1):e2529.

Sharples, J. J., McRae, R. H. D., Weber, R. O., and Gill, A. M. (2009). A simple index for assessing fire danger rating. *Environmental Modelling & Software*, 24(6):764–774.

Siegert, S., Stephenson, D. B., Sansom, P. G., Scaife, A. A., Eade, R., and Arribas, A. (2016). A Bayesian framework for verification and recalibration of ensemble forecasts: How uncertain is NAO predictability? *Journal of Climate*, 29(3):995–1012.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107:513–532.

Sklar, A. (1959). Functions of distribution to N dimensions and their margins. *Publications of the Statistics Institute of University of Paris*, 8(1):229–231.

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

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

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

Son, R., Kim, H., Wang, S.-Y., Jeong, J.-H., Woo, S.-H., Jeong, J.-Y., Lee, B.-D., Kim, S. H., LaPlante, M., Kwon, C.-G., and Yoon, J.-H. (2021). Changes in fire weather climatology under 1.5 °C and 2.0 °C warming. *Environmental Research Letters*, 16(3):034058.

Southworth, H., Heffernan, J. E., and Metcalfe, P. D. (2020). *texmex: Statistical modelling of extreme values*. R package version 2.4.8.

Stephenson, A. G., Saunders, K., and Tafakori, L. (2018). The MELBS team winning entry for the EVA2017 competition for spatiotemporal prediction of extreme rainfall using generalized extreme value quantiles. *Extremes*, 21:477–484.

Stewart, E. J., Kjeldsen, T. R., Jones, D. A., and Morris, D. G. (2008). The flood

estimation handbook and UK practice: past, present and future. In *Flood Risk Management: Research and Practice*. CRC Press.

Svensson, C. and Jones, D. A. (2004). Dependence between sea surge, river flow and precipitation in south and west Britain. *Hydrology and Earth System Sciences*, 8(5):973–992.

Syakur, M., Khotimah, B. K., Rochman, E., and Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, 336:012017.

Tawn, J. A. (1988a). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Tawn, J. A. (1988b). An extreme-value theory model for dependent observations. *Journal of Hydrology*, 101:227–250.

Tawn, J. A. (1992). Estimating probabilities of extreme sea-levels. *Journal of the Royal Statistical Society: Series C*, 41(1):77–93.

Tawn, J. A. and Vassie, J. M. (1989). Extreme sea levels: the joint probabilities method revisited and revised. *Proceedings of the Institution of Civil Engineers*, 87(3):429–442.

Trace-Kleeberg, S., Haigh, I. D., Walraven, M., and Gourvenec, S. (2023). How should storm surge barrier maintenance strategies be changed in light of sea-level rise? a case study. *Coastal Engineering*, 184:104336.

Wadey, M. P., Haigh, I., Nicholls, R. J., Brown, J. M., Horsburgh, K., Carroll, B., Gallop, S. L., Mason, T., Bradshaw, E., et al. (2015). A comparison of the 31 January–1 February 1953 and 5–6 December 2013 coastal flood events around the UK. *Frontiers in Marine Science*, 2(84).

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. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B*, 79(1):149–175.

Wahl, T., Haigh, I. D., Woodworth, P. L., Albrecht, F., Dillingh, D., Jensen, J., Nicholls, R. J., Weisse, R., and Wöppelmann, G. (2013). Observed mean sea level changes around the North Sea coastline from 1800 to present. *Earth-Science Reviews*, 124:51–67.

Weiss, J. and Bernardara, P. (2013). Comparison of local indices for regional frequency analysis with an application to extreme skew surges. *Water Resources Research*, 49(5):2940–2951.

Williams, J., Horsburgh, K. J., Williams, J. A., and Proctor, R. N. (2016). Tide and skew surge independence: new insights for flood risk. *Geophysical Research Letters*, 43(12):6410–6417.

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

Wong, S. D., Broader, J. C., and Shaheen, S. A. (2020). Review of California wildfire evacuations from 2017 to 2019. Technical report, UC Office of the President: University of California Institute of Transportation Studies.

Wong, T. E., Sheets, H., Torline, T., and Zhang, M. (2022). Evidence for increasing frequency of extreme coastal sea levels. *Frontiers in Climate*, 4(796479).

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. CRC press.

Woodworth, P. L., Menéndez, M., and Roland Gehrels, W. (2011). Evidence for century-timescale acceleration in mean sea levels and for recent changes in extreme sea levels. *Surveys in Geophysics*, 32(4):603–618.

Woodworth, P. L. and Player, R. (2003). The permanent service for mean sea level: an update to the 21st century. *Journal of Coastal Research*, 19(2):287–295.

Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Arnold, J. R., DeAngelo, B., Doherty, S., Easterling, D. R., Edmonds, J., Edmonds, T., Hall, T., et al. (2017). Climate science special report: fourth national climate assessment, volume I. Technical report, U.S. Global Change Research Program, Washington, DC, USA.

Youngman, B. D. (2019). Generalized additive models for exceedances of high thresholds with an application to return level estimation for US wind gusts. *Journal of the American Statistical Association*, 114(528):1865–1879.

Youngman, B. D. (2022). evgam: an R package for generalized additive extreme value models. *Journal of Statistical Software*, 103(3):1–26.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447.

Zhang, Y., Lim, S., and Sharples, J. J. (2017). Wildfire occurrence patterns in ecoregions of New South Wales and Australian Capital Territory, Australia. *Natural Hazards*, 87:415–435.

Zhuang, Y., Fu, R., Santer, B. D., Dickinson, R. E., and Hall, A. (2021). Quantifying contributions of natural variability and anthropogenic forcings on increased fire weather risk over the western United States. *Proceedings of the National Academy of Sciences*, 118(45):e2111875118.

Ziel, R. H., Bieniek, P. A., Bhatt, U. S., Strader, H., Rupp, T. S., and York, A. (2020). A comparison of fire weather indices with MODIS fire days for the natural regions of Alaska. *Forests*, 11(5):516.

Zsamboky, M., Fernández-Bilbao, A., Smith, D., Knight, J., and Allan, J. (2011). Impacts of climate change on disadvantaged UK coastal communities. Technical report, Joseph Rowntree Foundation.