

# Conditional Models for Spatial Extremes

Robert Shooter, M.Sci.(Hons.), M.Res



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

March 2020

# Abstract

Extreme environmental events endanger human life and cause serious damage to property and infrastructure. For example, Storm Desmond (2015) caused approximately £500m of damage in Lancashire and Cumbria, UK from high winds and flooding, while Storm Britta (2006) damaged shipping vessels and offshore structures in the southern North Sea, and led to coastal flooding. Estimating the probability of the occurrence of such events is key in designing structures and infrastructure that are able to withstand their impacts.

Due to the rarity of these events, extreme value theory techniques are used for inference. This thesis focusses on developing novel spatial extreme value methods motivated by applications to significant wave height in the North Sea and north Atlantic, and extreme precipitation for the Netherlands.

We develop methodology for analysing the dependence structure of significant wave height by utilising spatial conditional extreme value methods. Since the dependence structure of extremes between locations is likely to be complicated, with contributing factors including distance and covariates, we model dependence flexibly; otherwise, the incorrect assumption on the dependence between sites may lead to

inaccurate estimation of the probabilities of spatial extreme events occurring. Existing methods for spatial extremes typically assume a particular form of extremal dependence termed asymptotic dependence, and often have intractable forms for describing the dependence of joint events over large numbers of locations. The model developed here overcomes these deficiencies. Moreover, the estimation of joint probabilities across sites under both asymptotic independence and asymptotic dependence, the two limiting extremal dependence classes, is possible with our model; this is not the case with other methods.

We propose a method for the estimation of marginal extreme precipitation quantiles, utilising a Bayesian spatio-temporal hierarchical model. Our model parameters incorporate an autoregressive prior distribution, and use spatial interpolation to pool information on model parameters across neighbouring sites.

# Acknowledgements

I'd first like to thank my supervisors Jonathan Tawn, Jennifer Wadsworth and Philip Jonathan for guiding me through the three years of my PhD; their help has been invaluable in writing this thesis. In particular, thanks Jon for the encouragement and assistance (sometimes going beyond what I would have expected!) you've provided throughout the project, even when exceptionally busy. Thanks also to Phil for accommodating me in Rhuthun multiple times to collaborate on work - *diolch!* More generally, thank you to Shell for allowing for this project to happen and particular thanks to Emma, David and Matthew, whose suggestions and advice I'm very grateful for.

Thanks also to my parents and wider family for all of the support through my education (and life), I'm very grateful for all of the opportunities I've received with your help; going to university may not have been possible without that. And thank you for putting up with visiting Lancaster for 8 years helping me transport things, etc.

Speaking of Lancaster, I've greatly appreciated the support of the STOR-i CDT (and by extension EPSRC) over the past 4 years, providing advice as well as many

laughs along the way. I'd especially like to thank my year's cohort from STOR-i who have helped in making my time so enjoyable, the Extremes Coffee group for allowing me to vent my grievances at the world every Tuesday and for the (mostly) enjoyable paper-writing experience, and to the STOR-i football attendees who allowed me to embarrass myself regularly and not let me forget about it.

I couldn't write these acknowledgements without mentioning my "live-in chum" for my whole 8 years at Lancaster - Jake. His *unique* brand of humour and conversation have brought me much entertainment (plus occasional boredom) and provided much-needed levity at times, if not occasional bewilderment. I'm also very grateful to all my other long-suffering friends who have often brightened up my days. Anja, you're also alright, I guess.

# Declaration

I declare that the work in this thesis has been done by myself, except where noted below, and has not been submitted elsewhere for the award of any other degree. The word count of this thesis is 64,798 words.

The content in Chapter 4 has been accepted for publication as Tawn J., Shooter, R., Towe, R. and Lamb, R. (2018). Modelling spatial extreme events with environmental applications, *Spatial Statistics* 28:39-58. In this chapter, I contributed the analysis and writing in Sections 4.2.1, 4.2.2 and 4.4.1, and also contributed minor parts of the text in Sections 4.1, 4.2.3, 4.2.4 and 4.3, including Figure 4.3.1. Section 4.4.2 was written principally by R. Towe, with some minor edits made by me.

The content in Chapter 5 has been accepted for publication as Shooter, R., Ross, E., Tawn, J. and Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics* 30:e2562.

The content in Chapter 6 has been submitted for publication as Shooter, R., Ross, E., Tawn, J. and Jonathan, P. (2019). Basin-wide conditional spatial extremes for severe ocean storms.

The content in Chapter 7 has been accepted for publication as Barlow, A. M.,

Rohrbeck, C., Sharkey, P., Shooter R. and Simpson, E. S. (2018). *Extremes* 21:431-439. In this chapter, I wrote Section 7.1 and carried out preliminary analysis of the data used in the chapter, as well as contributing edits to the text in the remaining sections.

Robert Shooter

# Contents

<b>Abstract</b>	<b>I</b>
<b>Acknowledgements</b>	<b>III</b>
<b>Declaration</b>	<b>V</b>
<b>Contents</b>	<b>XII</b>
<b>List of Figures</b>	<b>XXIV</b>
<b>List of Tables</b>	<b>XXV</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Thesis outline . . . . .	7
<b>2 Literature Review</b>	<b>11</b>
2.1 Univariate extreme value theory . . . . .	11
2.1.1 Block maxima approach . . . . .	11
2.1.2 Threshold exceedance approach . . . . .	14

2.1.3	Point process representation . . . . .	17
2.2	Multivariate extreme value theory . . . . .	19
2.2.1	Componentwise maxima . . . . .	19
2.2.2	Multivariate point process approach . . . . .	23
2.2.3	Multivariate generalised Pareto distribution . . . . .	24
2.3	Extremal dependence . . . . .	25
2.3.1	Copulas . . . . .	26
2.3.2	Dependence measures . . . . .	27
2.4	Spatial extremes . . . . .	30
2.4.1	Max-stable processes . . . . .	30
2.4.2	Inference for max-stable processes . . . . .	38
2.4.3	Other recent approaches . . . . .	39
2.5	Conditional extremes methods . . . . .	46
2.5.1	Theoretical background . . . . .	47
2.5.2	Conditional extremes models . . . . .	52
2.5.3	Linking conditional extremes to the Brown-Resnick max-stable process . . . . .	55
<b>3</b>	<b>Comparison of censored likelihood methods under asymptotic inde- pendence</b> . . . . .	<b>65</b>
3.1	Censored likelihood methods . . . . .	66
3.2	Inverted max-stable processes and resulting properties . . . . .	68
3.3	Dependence measures for asymptotically independent processes . . . . .	70

3.4	IMSP censored likelihoods fitted to Gaussian process data . . . . .	72
3.4.1	Results for $\eta(h)$ . . . . .	74
3.4.2	Results for $\chi(u; h)$ . . . . .	82
3.5	Gaussian process censored likelihoods fitted to inverted max-stable process data . . . . .	87
3.5.1	Results for $\eta(h)$ . . . . .	89
3.5.2	Results for $\chi(u; h)$ . . . . .	92
3.6	Summary of results . . . . .	95
<b>4</b>	<b>Modelling spatial extreme events with environmental applications</b>	<b>99</b>
4.1	Introduction . . . . .	99
4.2	Existing methods . . . . .	105
4.2.1	Univariate modelling . . . . .	105
4.2.2	Max-stable processes . . . . .	109
4.2.3	Pareto processes . . . . .	110
4.2.4	Weakness of Pareto processes . . . . .	113
4.3	Conditional extremes . . . . .	115
4.3.1	Asymptotics for conditional multivariate extremes . . . . .	115
4.3.2	Models for conditional spatial extremes . . . . .	119
4.4	Applications . . . . .	125
4.4.1	Offshore risk from waves . . . . .	125
4.4.2	Understanding widespread flood risk . . . . .	134
<b>5</b>	<b>On spatial conditional extremes for ocean storm severity</b>	<b>146</b>

5.1	Introduction . . . . .	146
5.2	Motivating application . . . . .	149
5.3	Spatial conditional extremes . . . . .	152
5.3.1	Characterising extremal dependence . . . . .	152
5.3.2	The conditional extremes model of Heffernan and Tawn (2004)	153
5.3.3	The spatial conditional extremes (SCE) model . . . . .	154
5.3.4	Constraints . . . . .	156
5.4	Inference . . . . .	156
5.4.1	Likelihood for the “free” model . . . . .	157
5.4.2	MCMC for the free model . . . . .	161
5.4.3	Inference for the “parametric- $\alpha$ ” model . . . . .	161
5.4.4	Comparison of free and parametric- $\alpha$ models . . . . .	162
5.5	Application to northern North Sea North-South transect (NNS:N-S) .	163
5.5.1	Free model . . . . .	164
5.5.2	“Parametric- $\alpha$ ” fit . . . . .	166
5.6	Application to other North Sea transects . . . . .	168
5.6.1	Comparison of Model Fits for all Transects . . . . .	169
5.6.2	NNS east-west transect . . . . .	170
5.6.3	CNS transects . . . . .	172
5.7	Discussion and conclusions . . . . .	174
5.8	Supplementary Material . . . . .	177
<b>6</b>	<b>Basin-wide spatial conditional extremes for severe ocean storms</b>	<b>180</b>

6.1	Introduction . . . . .	180
6.2	Conditional extremes . . . . .	183
6.2.1	Extremal dependence . . . . .	183
6.2.2	Multivariate conditional extremes models . . . . .	184
6.2.3	Spatial conditional extremes . . . . .	186
6.2.4	Model parameter variation with distance . . . . .	188
6.3	Inference . . . . .	189
6.3.1	Likelihood and MCMC . . . . .	189
6.3.2	Parametric forms for $\alpha$ , $\beta$ and $\sigma$ . . . . .	190
6.3.3	Parameterising $\mu$ and $\delta$ with distance . . . . .	192
6.3.4	Incorporating anisotropy . . . . .	192
6.4	Applications . . . . .	194
6.4.1	Significant wave height data . . . . .	194
6.4.2	Results . . . . .	196
6.5	Conclusions . . . . .	201

**7 A Bayesian spatio-temporal model for precipitation extremes STOR  
team contribution to the EVA2017 challenge 204**

7.1	Introduction . . . . .	204
7.2	Methodology . . . . .	207
7.2.1	Likelihood . . . . .	207
7.2.2	Prior model . . . . .	209
7.2.3	Threshold selection and estimation . . . . .	210

<i>CONTENTS</i>	XII
7.3 Results and discussion . . . . .	212
<b>8 Conclusions and further work</b>	<b>216</b>
<b>Bibliography</b>	<b>221</b>

# List of Figures

2.1.1 Comparison of univariate extreme value approaches on daily rainfall data at a location in south-west England from 1914-1962. Orange lines separate the data into yearly blocks, the blue line indicates a threshold of 30mm daily rainfall, blue circles indicate data that exceed this threshold. Red points indicate yearly maxima; triangles denote yearly maxima which also exceed 30mm, red circles indicate those maxima which do not exceed this threshold. . . . .	15
2.4.1 Simulations of two-dimensional max-stable processes with standard Gumbel margins on the space $[0, 10] \times [0, 10]$ : (a), (b) and (c) are Smith processes with covariance matrices $\Sigma_1, \Sigma_2$ and $\Sigma_3$ respectively; (d) shows a Schlather process, with correlation function $\rho(\mathbf{h}) = \exp(-\ \mathbf{h}\ )$ .	34
2.4.2 A simulation of a one-dimensional Brown-Resnick max-stable process, with variogram $\gamma(h) = \ h\ $ , on standard Gumbel margins on the line segment $[0, 10]$ . . . . .	37

3.4.1 Comparison of theoretical values of  $\eta(h)$  using estimated parameter values of  $\alpha$  and  $\lambda$ . The red line corresponds to the theoretical result under the estimate obtained from the Wadsworth and Tawn (2012b) censoring scheme, the blue line corresponds to the theoretical result under the estimate obtained from the Ledford and Tawn (1996) censoring scheme and the orange line represents the theoretical value of  $\eta(h)$  for the simulated Gaussian process, using  $n = 1000$  replicates. . . . . 73

3.4.2 Comparison of theoretical values of  $\chi(u; h)$  using estimated parameter values of  $\alpha$  and  $\lambda$  with  $h = 4$  and  $0.9 \leq u \leq 1$ . The red line corresponds to the theoretical result under the estimate obtained from the Wadsworth and Tawn (2012b) censoring scheme, the blue line corresponds to the theoretical values of  $\chi(u; h)$  under the estimate obtained from the Ledford and Tawn (1996) censoring scheme and the orange line represents the theoretical value of  $\chi(u; h)$  for the simulated Gaussian process, using  $n = 1000$  replicates. . . . . 74

3.4.3 Plot of bias of  $\hat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. The dashed black line indicates zero bias. . . . . 76

3.4.4 Plot of variance of  $\hat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 76

3.4.5 Plot of RMSE of  $\widehat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 77

3.4.6 Plot of bias of  $\widehat{\eta}(h)$  with exponential covariance function parameters  $\phi = 1$  (square points) and  $\phi = 5$  (triangular points) in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. The black dashed line indicates zero bias. . . . . 80

3.4.7 Plot of variance of  $\widehat{\eta}(h)$  with exponential covariance function parameters  $\phi = 1$  (square points) and  $\phi = 5$  (triangular points) in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 80

3.4.8 Plot of RMSE of  $\widehat{\eta}(h)$  with exponential covariance function parameters  $\phi = 1$  (square points) and  $\phi = 5$  (triangular points) in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 81

3.4.9 Plot of RMSE of  $\widehat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 200$  replicates. . . . . 81

3.4.10 Plot of bias of  $\widehat{\chi}(u; h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 84

3.4.11 Plot of variance of  $\widehat{\chi}(u; h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 84

3.4.12 Plot of RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 85

3.4.13 Plot of RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.99$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 85

3.4.14 Plot of RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.995$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. . . . . 86

3.4.15 Plot of RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 200$  replicates. . . . . 87

3.5.1 Plot of the bias of  $\hat{\eta}(h)$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable process, from simulations with  $n = 1000$  replicates. The black dashed line indicates zero bias. . . . . 90

3.5.2 Plot of the variance of  $\hat{\eta}(h)$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable process, from simulations with  $n = 1000$  replicates. . . . . 91

3.5.3 Plot of the RMSE of  $\hat{\eta}(h)$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. . . . . 91

3.5.4 Plot of the bias of  $\widehat{\chi}(u; h)$  for  $u = 0.95$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. The black dashed line indicates zero bias. . . . . 93

3.5.5 Plot of the variance of  $\widehat{\chi}(u; h)$  for  $u = 0.95$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. . . . . 93

3.5.6 Plot of the RMSE of  $\widehat{\chi}(u; h)$  for  $u = 0.95$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. . . . . 94

3.5.7 Plot of the RMSE of  $\widehat{\chi}(u; h)$  for  $u = 0.99$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. . . . . 94

3.5.8 Plot of the RMSE of  $\widehat{\chi}(u; h)$  for  $u = 0.995$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. . . . . 95

4.2.1 Illustration of threshold stability property described by relationship (4.2.3).  
 The left panel shows a sample from  $V \sim \text{GPD}(\psi, \xi)$  with the vertical line representing the threshold  $v$  and the red points the exceedances of  $v$ ; the right panel shows these same exceedances (shown as excesses in red) after scaling (here the GPD has parameters  $(\psi, \xi) = (1, 0)$ , and so  $c_v = 1$ ). These scaled excesses are compared against a new sample (in grey) from the original distribution of  $V$ , we note that these two samples have the same distribution. . . . . 107

4.2.2 Illustration of a Pareto process, showing realisations of a process  $X(s)$  (grey lines), where for some chosen threshold  $u$  (blue line), with the realisations where  $\sup_{s \in \mathcal{S}} X(s) > u$  (red lines) being approximately distributed as  $u + T(s)$ . . . . . 112

4.2.3 Illustration of types of extremal spatial behaviour. The top row shows a process which retains the same spatial profile as the event becomes more extreme, corresponding to asymptotic dependence. The bottom row depicts the extreme event becoming more localised as its magnitude increases, commonly seen in practice and corresponding to asymptotic independence. . . . . 115

4.3.1 Illustrations of Pareto type, Gaussian and mixture extremal processes on a space  $\mathcal{S} = (-10, 10)$ . In all cases  $X(0)$  is in an extreme state (equal to the 99.995% marginal quantile), and the latent Gaussian process  $Z(s)$  has mean and standard deviation of  $\mu(h) = \mu_c$  and  $\sigma(h) = \sigma_c$  for  $h > 0$  and correlation function  $\rho(h) = \exp(-h/3)$ . Illustration as follows: (a) Pareto type process with  $\mu_c = -0.4$ ,  $\sigma_c^2 = 1.3$ ; (b) Gaussian process  $\alpha(h) = \exp(-h/3)$ ,  $\mu_c = 0.06$ ,  $\sigma_c^2 = 0.6$ ; (c) mixture process with  $h_{AD} = 3$ ;  $\alpha(h) = \exp(-|h - h_{AD}|/3)$  for  $h > 3$ ,  $\beta(h) = 0$ ,  $\mu_c = -0.05$ ,  $\sigma_c^2 = 1.3$ . . . . . 124

4.4.1 Map of sampling locations in the North Sea from which the data are collected, with the particular transect used for model fitting highlighted in red. . . . . 126

4.4.2 Pointwise estimates of  $\alpha(h)$  from the multivariate conditional extremes fit, conditioned on the west-most location in the transect. Lag  $h = 0$  corresponds to the conditioning site, with  $h = 6$  being the parameter estimate at the most easterly site. Estimates are for integer values of  $h$  and these are shown to be linearly interpolated to show we know that the function is continuous. The dotted lines show 95% confidence intervals for the pointwise estimates. . . . . 129

4.4.3 Pointwise estimates of  $\mu(h)$  and  $\sigma(h)$  with properties shown identical to Figure 4.4.2. . . . . 130

4.4.4 Simulations from the fitted multivariate conditional extremes model; black points are the data on Gumbel margins, whilst red points are simulated data from the fitted model: (a), (b) and (c) show these data when  $h = 1, 3, 6$  respectively. In each case, the  $x$ -axis is the standardised wave height at the conditioning site (the most westerly in the transect), with the  $y$ -axis being the standardised wave height at the other site. . . . . 133

4.4.5 Comparisons of the non-limit conditional probability (4.1.4) for (a) the 99th percentile and (b) the one year return period from the observed data set; (c) and (d) show this conditional probability estimated using our model for the 10 and 100 year return periods respectively. The triangle symbol represents the conditioning gauge for the estimate, this gauge is situated in the river Severn catchment. . . . . 139

4.4.6 Comparison of the three dependence models used to estimate probability of observing at least one event above a  $T$ -year return period for a given year: our model for the dependence (black), under a complete dependence model (blue) and under complete dependence (red). . . . 142

4.4.7 Estimated probability of observing at least  $r$  LRF regions above a  $T$ -year return period in any given year. The black, red, green and blue curves show the cases for when  $r = 1, 2, 3, 4$  respectively. . . . . 144

5.2.1 NNS and CNS locations considered. . . . . 150

5.4.1 Illustration of notation used to describe the disposition of points on the line, enabling pooling of data from pairs of locations by distance. The  $(c, j)$  notation is shown below the line, and distance  $h_{c0j}$  given above each point. Points with the same value of  $h_{c0j}$  are shown in the same colour;  $\Delta$  is the inter-location spacing. . . . . 158

5.5.1 NNS:N-S transect, free model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with end-points shown as solid triangles). . . . . 165

5.5.2 Scatter plots illustrating dependence between values of Laplace-scale storm peak  $H_S$  at different relative distances for NNS:N-S transects, from (a) original sample and (b) simulation under the fitted free model. Black points are the original data on Laplace scale; red points are data simulated under the fitted model. . . . . 166

5.5.3 Observed spatial processes from the NNS:N-S transect with Laplace-scale values at the left-hand location in the interval  $[3.5, 4.5]$ , together with posterior predictive estimates from simulation under the fitted free model, represented using the median (black) and upper and lower limits of a 95% pseudo-likelihood credible interval. Red lines are simulated spatial processes from the fitted model. . . . . 167

5.5.4 NNS:N-S transect, parametric- $\alpha$  model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with endpoints shown as solid triangles). . . . . 168

5.6.1 NNS:E-W transect, parametric  $\alpha(h)$  model: estimates for (a)  $\alpha(h)$ , (b)  $\beta(h)$ , (c)  $\mu(h)$  and (d)  $\sigma(h)$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with endpoints shown as solid triangles). . . . . 171

5.6.2 CNS:N-S transect, parametric- $\alpha$  model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with endpoints shown as solid triangles). . . . . 172

5.6.3 CNS:E-W transect, parametric- $\alpha$  model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with endpoints shown as solid triangles). . . . . 173

5.7.1 Pseudo-likelihood credible intervals for (a) the conditional mean and (b) the conditional standard deviation of the fitted dependence model as a function of distance in kilometres, for conditioning Laplace-scale value of 5, and different transects: NNS:N-W (red), NNS:E-W (magenta), CNS:N-S (blue), CNS:E-W (cyan). . . . . 175

6.4.1 Map of the conditioning site (coloured red) and remote sites (black) for the North Atlantic long transect analysis. . . . . 195

6.4.2 Map of the conditioning site (coloured green) and remote sites (red) for the North Sea ocean basin. Sites not used in this analysis are coloured black. . . . . 196

6.4.3 Parameter estimates, with distance  $d$  for the North Atlantic long transect analysis. The solid black line represents the posterior median, with dashed lines representing the upper and lower limits of the 95% posterior credible interval, for (a)  $\alpha(d)$ , (b)  $\beta(d)$ , (c)  $\mu(d)$ , (d)  $\sigma(d)$  and (e)  $\delta(d)$ . . . . . 198

6.4.4 Parameter estimates, with distance  $d$  (for  $d < 180\text{km}$ ) for the North Atlantic long transect analysis. The solid black line represents the posterior median, with dashed lines representing the upper and lower limits of the 95% posterior credible interval, for (a)  $\alpha(d)$ , (b)  $\beta(d)$ , (c)  $\mu(d)$ , (d)  $\sigma(d)$  and (e)  $\delta(d)$ . . . . . 199

6.4.5 Parameter estimates, with distance  $d$  for the North Sea ocean basin analysis. The solid black line represents the posterior median, with dashed lines representing the upper and lower limits of the empirical 95% posterior credible interval, for (a)  $\alpha(d)$ , (b)  $\beta(d)$ , (c)  $\mu(d)$ , (d)  $\sigma(d)$  and (e)  $\delta(d)$ . . . . . 201

6.5.1 Simulated processes (red) from our fitted spatial conditional extremes model, using median values of the model parameters from the last 1000 iterations of the MCMC chain, and observed processes (black), conditioned on value of $X_0 \approx 4$ for simulated data, and between 0.985 and 0.995 Laplace quantiles for observed data. The grey line shows the mean of the values of $X_0$ conditioned upon. . . . .	203
7.3.1 MCMC chains for the scale and shape parameters for station 10 in June.	213
7.3.2 Location of stations 2, 5, 7 and 10, as well as estimates of the corresponding scale and shape parameters and predicted 0.998 quantiles. . . . .	214

# List of Tables

2.5.1 Table of normalising constants $a_{j i}$ , $b_{j i}$ , as well as the forms of the limiting distribution $G_{ i}$ for a variety of cases of dependence between variables; ‡ refers the reader to Heffernan and Tawn (2004) for details of this limiting form. This table has been adapted from Table 1 of Heffernan and Tawn (2004). . . . .	49
3.6.1 Table of total computational time (in seconds) to obtain estimates of inverted Brown-Resnick max-stable process and Gaussian process (GP) parameters (as specified in Sections 3.4 and 3.5) for all of the simulated data sets described in Sections 3.4 and 3.5. . . . .	96
5.6.1 Table of DIC values for the free fit model and parametric- $\alpha$ model for all of the transect analyses. . . . .	170
7.3.1 Percentage improvement over the benchmark for Challenges 1 and 2 across each month. . . . .	215

# Chapter 1

## Introduction

### 1.1 Motivation

In the recent past, there have been numerous extreme environmental events across the world which have caused significant damage to infrastructure and property, as well as leading to loss of life. Examples of such events include Storm Desmond in December 2015 causing widespread flooding across areas of North West England, record high temperatures in Western Europe during Summer 2019 creating public health issues, and Storm Britta in October/November 2006, which caused damage to numerous offshore platforms and ships in the North Sea (Kettle, 2015). In this thesis, we will concentrate on applying our methods to extreme significant wave height, denoted by  $H_S$ , corresponding to extreme winter storm events in the North Sea and north Atlantic, and extreme precipitation in the Netherlands, but the methodology developed here can be applied to a multitude of other applications, for example in modelling the spatial dependence of extreme temperatures during heatwave events.

In our applications, the inference made on the behaviour of extremes, such as the likelihood of certain events occurring, would help guide design criteria for offshore structures and vessels in the case of significant wave height, and appropriate measures for preventing flooding from extreme precipitation. The former of these is of significant interest to companies which require such information for offshore engineering purposes, such as Shell. Furthermore, we wish to be able to do this in a manner which is theoretically justifiable, and computationally feasible. The latter of these is important in environmental applications as the data are often high-dimensional since interest frequently lies in the joint behaviour of variables at a number of sampling locations.

As the population of the Earth increases and more infrastructure is built, the impacts felt by extreme events become more significant. Thus, it is of key importance that the risk to human life, as well as financial risk, is mitigated by determining the likelihood of these events, and preparing appropriately based on this information. However, the type of events which cause major damage generally tend to occur at a specific location only rarely. Any such event would be in the tail of the marginal distribution for the process at the location, or may be in the tail of the joint distribution of the process, and so accurate inference would be difficult when using classical statistical models fitted to the whole distribution. On top of this, by the very nature of rare events, little data will exist for them. These considerations motivate the use of extreme value theory.

Up until fairly recently, for example by Coles and Tawn (1991), classical block maxima methods were typically used on environmental data, relying on multivariate

extensions of the univariate theory of extremes. Since environmental data are generally of interest over an area of the Earth's surface, this motivates the use of spatial extremes methodology. Theoretical frameworks for considering spatial extreme events have been present since the work of de Haan (1984) who defined the notion of a max-stable process, but there has been a rapid development in the area since the turn of the 21st Century. At first, these developments were largely built upon the max-stable process framework of de Haan (1984), and subsequently Smith (1990) who developed a model motivated by extreme rainfall events, but they have recently become both very numerous and with various modelling properties.

Underlying the inference of spatial extremes are the notions of asymptotic dependence and asymptotic independence, which are the only two limiting forms of extremal dependence. In broad terms, asymptotic dependence means that if an extreme has occurred at a specific location, there is a non-zero probability that an extreme will simultaneously occur somewhere else within the sampling domain however large the extreme events is, whilst under asymptotic independence, the larger the magnitude of the event, the more localised it becomes spatially. Often in applications, this latter type of behaviour is often observed, suggesting that asymptotic independence should be incorporated into a model for spatial extreme values. These two types of extremal dependence will lead to different estimates of probabilities of joint spatial extreme events. Thus, the type of extremal dependence assumed in the model used has a significant impact on the accuracy of any inference. The aim of this thesis is to develop the first spatial extreme value models which are able to incorporate both asymptotic dependence and asymptotic independence.

The max-stable processes of de Haan (1984), Smith (1990) and Schlather (2002), which we will introduce in more detail in Section 2.4.1, assume that only asymptotic dependence is present in the spatial field, so that all extreme values are dependent on one another. This means that any inference drawn from fitting a max-stable process model will be conservative, since the worst-possible case of extreme behaviour will be assumed. Wadsworth and Tawn (2012a) introduced the asymptotically independent counterpart to max-stable processes, known as inverted max-stable processes. As a mix of the two dependence types is more likely to be realistic, they propose to use a mixture model of both max-stable and inverted max-stable processes.

It is this flexibility of dependence structure over space which has been the focus of recent developments in spatial extremes modelling; see Huser and Wadsworth (2018) or Engelke et al. (2019) for examples. A large portion of this thesis will develop a methodology utilising a spatial conditional extremes approach, based on that proposed by Wadsworth and Tawn (2019), which permits both limiting types of dependence. We believe that using such models will allow for more realistic modelling of the environmental extremes that we will consider. Our models will allow the extremal dependence structure to vary with the distance between two points, with asymptotic dependence permitted for some distance between sites, before imposing asymptotic independence at all larger distances. Simulation for extrapolation from the model we propose is straightforward. By simulating from our model, we can estimate the joint probability of extreme events of a certain magnitude occurring.

With the limiting dependence type being modelled as part of this, our estimates should be more realistic than those from spatial extremes models such as max-stable

processes, since we can capture the mixture of types of extremal dependence behaviour that would be suggested by physical considerations. This has major implications in practice, since assessing the joint impact of a storm is of significant operational importance for Shell. An example of this is due to there being a limited amount of helicopters being available for evacuation from offshore platforms in a storm event; not all of these can be evacuated at once. Hence, knowledge of the joint characteristics of the event is important to know which sites should be evacuated, or to be able to make the structures able to withstand the storm so that there is no need for evacuation. Conventional offshore design tends to ignore the effects of the joint spatial occurrences of extreme events, and so our model also has this advantage of being able to consider the probability of joint spatial extreme events.

To carry out inference on the parameters in the models we present, we will largely make use of Bayesian inference. This is of particular importance in the spatial conditional extreme value model, where maximum likelihood techniques would have been very difficult to implement, chiefly because the global maximum is hard to find using these techniques. A reason for this is that some of the model parameters have a level of dependence with one another causing issues with parameter identifiability, and different parameter combinations have similar likelihood values. Combined with the likelihood surface often being relatively flat in our applications, this led to maximum likelihood algorithms being tricky to implement reliably. Thus, Bayesian methods are utilised in Chapters 5 and 6, as we feel this is the best approach to overcome the aforementioned computational issues and provides the best information for design engineers.

This technique will provide a new approach to the modelling of the spatial extreme values of  $H_S$  and may provide insight into the underlying extremal dependence structure which has not been explored before. Previous studies have utilised max-stable processes (e.g., Ross et al., 2017a) to assess the behaviour of extreme significant wave height in the North Sea, finding that the direction from which the wave emanates has an impact on the level of the dependence between extreme events.

Computational aspects will also be explored. A key drawback of many existing spatial extremes methods is the computational time required to fit a model; given that many environmental applications are high-dimensional (comprising of a large number of sites), this creates a significant practical issue. To this end, we will investigate the effects of using a censored likelihood scheme proposed by Wadsworth and Tawn (2012a), which may be appropriate for asymptotically independent data (but could also be used for asymptotically dependent data), compared to using a censored likelihood described by Ledford and Tawn (1996). The former is more computationally efficient but may introduce additional bias making parameter inference significantly worse; we investigate this by conducting a simulation study using misspecification of asymptotically independent models. We also present work for modelling extremes of significant wave height whereby the focus is on improving computational time without losing accuracy of inference, since from a practical perspective, this would be highly beneficial.

This thesis also contains details of a model for marginal inference of extreme precipitation data at a set of sites. These data were provided as part of a data challenge, for which assessing characteristics of the data was difficult. As a consequence, we will

provide a method of constructing a model for precipitation extremes which does not assume any particular behaviour in the dependence structure, and relies on the use of simple techniques for inference, which we believe is advantageous in such a scenario. This method could be applicable in situations where data are of a poor quality in terms of the number of missing data being large.

## 1.2 Thesis outline

We now provide an overview of the content of each chapter in this thesis. Firstly, Chapter 2 provides a detailed background of the requisite extreme value theory techniques in this thesis. First of all, we describe the classical univariate techniques, such as modelling extremes via block maxima and threshold excesses, and discuss the methods of inference using these. We then provide details of multivariate extensions of the univariate methods. These then motivate spatial models, such as max-stable processes, which we discuss. We conclude our overview of extreme value theory by introducing the concept of conditional extreme value theory, that is, modelling the behaviour of variables given that some other variable is extreme. Most of this discussion relates to the model proposed by Heffernan and Tawn (2004), before we link this model to spatial approaches, in order to outline that this is a natural model for spatial extremes; this type of modelling will be prominent within Chapters 4, 5 and 6.

Chapter 3 outlines a simulation study into the effect of misspecification of a spatial extremes model under the assumption of asymptotic independence. The simulation

study is similar to that of Huser et al. (2016), who carry out an investigation of the performance of a wide range of likelihoods under asymptotic dependence. In our work, we compare only the performance of censored likelihood techniques proposed by Ledford and Tawn (1996) and Wadsworth and Tawn (2012a), for which the latter may be more appropriate under asymptotic independence but with both also being applicable under asymptotic dependence. The Wadsworth and Tawn (2012a) censoring scheme is the more computationally efficient of these two approaches. We assess the performance by computing the bias and standard deviation of parameter estimates under both censored likelihoods.

Chapters 4, 5 and 6 focus on the development of a conditional spatial extremes model motivated by the conditional extremes model of Heffernan and Tawn (2004), with a focus on the application of the model to significant wave height data. At short inter-location distances, physical considerations suggest that the extremal dependence between storm severity at two locations exhibits asymptotic dependence, whereas with increasing distance we find asymptotic independence and eventually perfect independence. The conditional spatial extremes model incorporates all these forms of dependence. Compared to alternative descriptions of spatial extreme value processes, the model is advantageous since it admits both asymptotic dependence and asymptotic independence, and is conceptually straightforward. Moreover, our model is able to estimate the probabilities of joint events across locations; it is generally difficult, or impossible, to obtain a closed form expression for these joint probabilities using existing spatial extreme value models.

In Chapter 4, we will explain the limitations of many commonly-used modelling

approaches for spatial extremes and show how spatial models can be developed that overcome these deficiencies by exploiting the flexible conditional multivariate extremes models of Heffernan and Tawn (2004). We then illustrate the benefits of these new spatial models through applications to North Sea wave analysis and to widespread UK river flood risk analysis.

In Chapter 5, we use the spatial conditional extremes model within a Bayesian framework to estimate the extremal dependence of ocean storm severity (quantified by  $H_S$ ) for locations on spatial transects with approximate east-west (E-W) and north-south (N-S) orientations in the northern North Sea (NNS) and central North Sea (CNS), so the model considers space as one-dimensional. For  $H_S$  transformed on to standard Laplace marginal scale, the conditional extremes “linear slope” parameter  $\alpha$  decays approximately exponentially with distance for all transects. Further, the decay of mean extremal dependence with distance is found to be faster in the CNS than in the NNS. The persistence of mean extremal dependence is greatest for the E-W transect in the NNS, which is likely to be due to this transect being approximately aligned with the direction of propagation of the most severe storms in the region.

Chapter 6 presents a two-dimensional extension of the spatial conditional extremes model, described in Chapters 4 and 5, for ocean storm severity in the North Sea and the north Atlantic. In this model we incorporate distance-dependent parameters (with distance between sites considered across 2-dimensional space), with some represented as linear piecewise functions, in order to improve computational efficiency and model flexibility. In the work presented in this chapter, we use a generalised Gaussian distribution to describe the distribution of model residuals, and estimate anisotropy of

extremal dependence using a suitable metric for 2-dimensional distance. We apply the model to characterise the extremal spatial dependence of a two-dimensional spatial neighbourhood spanning hundreds of kilometres, and a one-dimensional transect spanning thousands of kilometres. In doing this, we allow the flexible modelling of the residual distributions, rather than relying on a Gaussian assumption as has been used in previous work.

Chapter 7 concerns our approach to the EVA2017 challenge, the aim of which was to predict extreme precipitation marginal quantiles across several sites in the Netherlands. Our approach uses a Bayesian hierarchical structure, which combines Gamma and generalised Pareto distributions. We impose a spatio-temporal structure in the model parameters via an autoregressive prior. Estimates are obtained using Markov chain Monte Carlo techniques and spatial interpolation. This approach has been successful in the context of the challenge, providing reasonable improvements over the benchmark metric provided by the competition organisers.

Finally, in Chapter 8 we give a summary of our conclusions from the work contained in this thesis, before mentioning scope for further work in the area.

Chapters 4, 5, 6 and 7 are presented as a sequence of papers, and so are constructed to be read as separate from one another; consequently, there may be background methodological information repeated in these chapters.

# Chapter 2

## Literature Review

### 2.1 Univariate extreme value theory

#### 2.1.1 Block maxima approach

First, two approaches to modelling extremes in a univariate setting are given, following the descriptions in Coles (2001). These methods are of key importance when considering marginal aspects of multivariate and spatial extremes.

The first technique that will be considered is to model block maxima, that is, to split a sequence of observations into blocks of equal length and model the maxima of each of these. To do this, first suppose we have an independent and identically-distributed (i.i.d.) sequence of random variables  $X_1, \dots, X_n$  from some common distribution function  $F$  and denote its maximum by  $M_n = \max\{X_1, \dots, X_n\}$ ; minima can be studied in the same framework by noting that

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

A key result for modelling these is the Extremal Types Theorem (Fisher and Tippett, 1928):

**Theorem 1** (Extremal Types Theorem): If there are sequences of constants  $(a_n > 0)_{n=1}^{\infty}$  and  $(b_n)_{n=1}^{\infty}$  such that, for any  $x \in \mathbb{R}$ ,

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x) \quad (n \rightarrow \infty),$$

for some non-degenerate distribution  $G$ , then  $G$  is either a Gumbel, Fréchet or negative Weibull distribution, which are defined as follows (with  $a > 0, b \in \mathbb{R}, \alpha > 0$ ):

- **Gumbel:**  $G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \quad x \in \mathbb{R};$
- **Fréchet:**  $G(x) = \begin{cases} 0, & x \leq b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x > b; \end{cases}$
- **Negative Weibull:**  $G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^\alpha\right]\right\}, & x < b, \\ 1, & x \geq b. \end{cases}$

The Extremal Types Theorem says that if appropriate normalising sequences,  $(a_n)$  and  $(b_n)$ , exist for the sequence of block maxima, then the normalised maxima must converge to one of the three classes of distribution given above. Moreover,  $F$  is said to be in the domain of attraction of  $G$ . Normalisation of the maxima is critical, since using  $(F(x))^n$  ( $x \in \mathbb{R}$ ), which corresponds to  $\mathbb{P}(M_n \leq x)$ , is generally impractical due to  $F$  usually being unknown. Additionally, if  $M_n$  is not normalised appropriately, we obtain a degenerate distribution, since then for all  $x < x^F$ , where  $x^F$  is the end-point of the distribution,  $F^n(x) \rightarrow 0$  as  $n \rightarrow \infty$  and a point mass at  $x^F$  is obtained.

In fact, it can be shown that all three of these distributions can be represented as special cases of the generalised extreme value (GEV) distribution, which has the parametric form

$$G(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} \quad (x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+, \xi \in \mathbb{R} \setminus \{0\}), \quad (2.1.1)$$

where  $\{z\}_+ = \max\{0, z\}$ , with the interpretation that if  $\xi = 0$ , then

$$G(x) = \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\} \quad (x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+).$$

The GEV distribution has the max-stability property, crucial for modelling extremes.

This says that for all  $n \in \mathbb{N}$ , there are constants  $A_n > 0, B_n$  such that

$$G^n(A_n x + B_n) = G(x);$$

furthermore the GEV distribution is the only distribution that satisfies this property.

The max-stability property described above means that, up to type, the maximum of GEV-distributed random variables must also follow a GEV distribution. Another property of the GEV distribution is that the endpoints are defined by the parameters; if  $\xi < 0$ , then the upper endpoint of the distribution is  $\mu - \sigma/\xi$ . If  $\xi > 0$ , then this value is the lower endpoint. We note that an important special case of the GEV distribution is the standard Fréchet distribution, given by  $F(x) = \exp(-1/x)$  for  $x > 0$ , also referred to as a unit Fréchet distribution.

The calculation of quantiles forms a particularly important part of inference from the GEV distribution; they can be used to inform about values which can be expected

to be exceeded over a given time period. The  $1 - p$  quantile can be calculated by

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

obtained by inverting (2.1.1); the interpretation is that the value  $z_p$  is exceeded in the length of one block with probability  $p$ . For example, if we have blocks corresponding to months, then the probability that  $z_p$  is exceeded in any given month is  $p$ . The standard terminology is that  $z_p$  is the return level corresponding to the return period  $1/p$ ; e.g. if  $p = 0.05$ , then  $z_p$  is the value exceeded once every 20 months on average. Coles (2001) provides methods for obtaining the variance of estimates of  $z_p$  if maximum likelihood estimators of  $\mu, \sigma$  and  $\xi$  have been calculated.

### 2.1.2 Threshold exceedance approach

An alternative way of modelling univariate extremes is to assess the characteristics of exceedances of some suitably chosen threshold. This method has the immediate advantage over the block maxima technique by not necessarily wasting as much data, since only one data point in each block is used but there could be more data points in a block that are useful for modelling extremes. We display this difference in Figure 2.1.1 using daily rainfall data from a location in south-west England, as used in Coles and Tawn (1996); the key distinction is to note that for the chosen threshold there are more blue points, corresponding to threshold exceedances (where this has been set to 30mm), than yearly maxima, denoted by red circles and triangles.

Such exceedances are described through the generalised Pareto distribution (GPD).

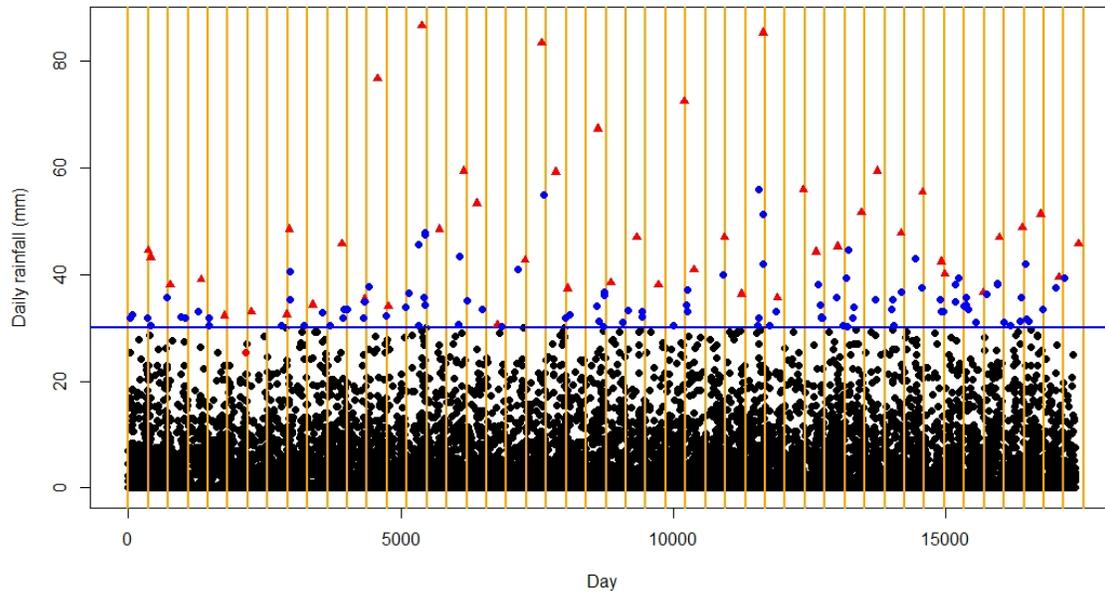


Figure 2.1.1: Comparison of univariate extreme value approaches on daily rainfall data at a location in south-west England from 1914-1962. Orange lines separate the data into yearly blocks, the blue line indicates a threshold of 30mm daily rainfall, blue circles indicate data that exceed this threshold. Red points indicate yearly maxima; triangles denote yearly maxima which also exceed 30mm, red circles indicate those maxima which do not exceed this threshold.

Formally, suppose that for some large  $n$ ,  $\mathbb{P}(M_n \leq x) \approx G(x)$ , where  $G$  is a GEV distribution and  $M_n$  is defined as before. Then for some large enough  $u$ , the threshold, the conditional distribution function of  $(X - u)|\{X > u\}$  is given approximately by

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)_+^{-\frac{1}{\xi}}, \quad y > 0, \quad (2.1.2)$$

where  $\sigma_u = \sigma + \xi(u - \mu) > 0$ , and denote a distribution taking the form (2.1.2) by  $\text{GPD}(\sigma_u, \xi)$ . A justification for this model in describing threshold exceedances is given in Coles (2001), and also Pickands (1975) and Davison and Smith (1990).

The GPD distribution has some important properties. First, we note that the parameter  $\xi$  common to both of the distributions  $G$  and  $H$  is equivalent, with the relevant interpretations of these given in Coles (2001). Secondly, the GPD exhibits threshold stability. Suppose that we have a threshold  $u_0$ , above which a  $\text{GPD}(\sigma_{u_0}, \xi)$  distribution is an appropriate model to model  $(X - u_0)|\{X > u_0\}$ . Then by the definition of the GPD, then for any  $u > u_0$ , the distribution of  $(X - u)|\{X > u\}$  should still be a GPD. However, this will be a  $\text{GPD}(\sigma_u, \xi)$  distribution, where  $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$ , i.e., the scale parameter  $\sigma_u$  is dependent on the threshold  $u$  chosen. This issue can be alleviated via the reparameterisation  $\sigma^* = \sigma_u - \xi u$ , which is constant with respect to  $u$ ; this way of representing the scale parameter aids with inference.

An important consideration when modelling threshold exceedances lies in the choice of threshold  $u$ . If  $u$  is chosen to be too small then there will be a large number of points above the threshold and the asymptotic results will not be suitable, which creates bias. However, if  $u$  is too large then the lack of data points above the threshold leads to parameter estimates having a high variance. Diagnostics such as

mean residual life plots and parameter stability plots, both given in Coles (2001), may help out in this choice. A comprehensive overview of threshold selection methods is given by Scarrott and MacDonald (2012).

As in the case for block maxima, inference for return levels can be made. However, as all threshold exceedances are now used (say there are  $N$  of these), slight alterations need to be made to obtain return levels corresponding to particular periods of time. Otherwise, the return level is simply the value exceeded, on average, once every  $N$  exceedances. Additionally, in order to undo the conditioning for the GPD,  $\mathbb{P}(X > u)$  must also be estimated. Coles (2001) gives details on the procedures for both of these aspects of the inference.

### 2.1.3 Point process representation

It can be shown that the two approaches to univariate extreme value theory described above can be considered as arising from sequences of point processes. The key result in this is as follows, and is as given by Coles (2001).

**Theorem 2:** Suppose that we have a sequence  $X_1, \dots, X_n$  of i.i.d. random variables and sequences  $(a_n > 0)_{n=1}^\infty$  and  $(b_n)_{n=1}^\infty$  such that, for any  $x \in \mathbb{R}$ ,

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x) \quad (n \rightarrow \infty),$$

with  $M_n$  defined as in Section 2.1 and  $G(\cdot)$  takes the form as defined in (2.1.1), with lower endpoint  $x_G$  and upper endpoint  $x^G$ . Then, the sequence of point processes given by

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

converges on regions in the form  $(0, 1) \times [u, \infty)$ , where  $x_G < u < x^G$ , to a Poisson process with intensity measure

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

where  $A = [t_1, t_2] \times [x, x^G]$ , where  $x > x_G$  and  $0 \leq t_1 < t_2 \leq 1$ .

For practical purposes, since the distribution of  $\{X_i\}_{i=1}^n$  is usually unknown, the following formulation, again given by Coles (2001), may be more useful. First, suppose that the sequence of data  $X_1, \dots, X_n$  as in Theorem 2. Then, for some sufficiently large  $u$ , let

$$N_n = \left\{ \left( \frac{i}{n+1}, X_i \right) : i = 1, \dots, n \right\},$$

and consider  $N_n$  on a region of the form  $(0, 1) \times [u, \infty)$ . Then  $N_n$  on this region is approximately a Poisson process with intensity measure

$$\Lambda(A) = (t_2 - t_1) \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \quad (2.1.3)$$

on  $A = [t_1, t_2] \times (x, \infty)$ , for  $x > u$ ,  $0 \leq t_1 < t_2 \leq 1$ .

To utilise this result, firstly the threshold  $u$  above which the Poisson process approximation is appropriate must be decided upon; this can be done using similar techniques to those outlined in Section 2.1.2. Then, set  $A = (0, 1) \times [u, \infty)$  and let the data points that lie within  $A$  be denoted  $\{(t_1, x_1), \dots, (t_{N(A)}, x_{N(A)})\}$ , so that in particular there are  $N(A)$  points in the region  $A$ . For easier interpretation of analysis, if the data arise from  $m$  blocks (e.g.,  $m$  years), then the intensity (2.1.3) can be replaced by

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

The advantage of this is that the parameters  $(\mu, \sigma, \xi)$  which are estimated now correspond to the GEV parameters of the block maxima, rather than the  $m$ -block maxima (e.g., annual maxima, rather than  $m$ -year maxima). Maximum likelihood estimators can then be found using the likelihood

$$L(\mu, \sigma, \xi; \mathbf{x}) \propto \exp \left\{ -m \left[ 1 + \xi \left( \frac{u - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} \prod_{i=1}^{N(A)} \frac{\left\{ 1 + \left( \frac{x_i - \mu}{\sigma} \right) \right\}_+^{-\frac{1}{\xi} - 1}}{\sigma}. \quad (2.1.4)$$

The expression given in (2.1.4) is derived from the likelihood for the Poisson process, namely

$$L(\mu, \sigma, \xi; \mathbf{x}) = \exp \{ -\Lambda(A) \} \prod_{i=1}^{N(A)} \lambda(t_i, x_i),$$

where  $\Lambda(A) = \int_{t_1}^{t_2} \lambda(t, x) dt$ . The approach given here differs slightly from that of fitting a GEV to the block maxima, as in Section 2.1.1, since all of the data larger than the threshold  $u$  are used to estimate the model parameters; this should result in more accurate inference. We also note that in this parameterisation, the parameters are invariant to the choice of  $u$ , unlike the GPD.

## 2.2 Multivariate extreme value theory

### 2.2.1 Componentwise maxima

To motivate multivariate extreme value theory, we describe the concept of componentwise maxima, a multivariate extension of block maxima. Consider a sample of  $d$ -dimensional observations,  $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$ ,  $i = 1, \dots, n$ . Then, as in Beirlant et al. (2004), define  $\mathbf{M}_n = (M_{n1}, \dots, M_{nd})$  to be the vector of componentwise maxima

with elements defined by

$$M_{nj} = \max_{1 \leq i \leq n} X_{ij} \quad (j = 1, \dots, d).$$

The vector  $\mathbf{M}_n$  is not necessarily a data point; each component's maximum may arise from different observations. Then, given an i.i.d. sample  $\mathbf{X}_1, \dots, \mathbf{X}_n$  from a common distribution function  $F$  and any  $\mathbf{x} \in \mathbb{R}^d$ ,

$$\mathbb{P}(\mathbf{M}_n \leq \mathbf{x}) = \mathbb{P}(\mathbf{X}_1 \leq \mathbf{x}, \dots, \mathbf{X}_n \leq \mathbf{x}) = F^n(\mathbf{x}),$$

taking componentwise operations in the above expression, and shall also do this for any vector expressions in the rest of this section.

There is a multivariate analogue to the univariate case of describing how extreme value distributions arise, as stated in Smith et al. (1990). Suppose we have vectors  $\mathbf{a}_n = (a_{n1} > 0, \dots, a_{nd} > 0)$  and  $\mathbf{b}_n = (b_{n1}, \dots, b_{nd})$ , such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x} \right) = G(\mathbf{x}), \quad (2.2.1)$$

for some  $d$ -dimensional distribution  $G$  which is non-degenerate in each margin. Then  $G$  is a multivariate extreme value distribution; we note also that each of the margins follows a univariate GEV distribution as described in Section 2.1. Furthermore, there is a multivariate max-stable analogue; a multivariate distribution function is max-stable if for all  $N \in \mathbb{N}$ , there are vectors  $\mathbf{A}_N > 0$  and  $\mathbf{B}_N$  such that

$$G^N(\mathbf{x}) = G(\mathbf{A}_N \mathbf{x} + \mathbf{B}_N).$$

Just as in the univariate case, a distribution function  $G$  is a multivariate extreme value distribution if and only if it satisfies the max-stability property and the componentwise maxima follows a multivariate extreme value distribution up to type.

Following Tawn (1990) or Beirlant et al. (2004), suppose we have a sequence of  $d$ -dimensional random variables,  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  with unit Fréchet margins and vector of componentwise maxima defined above. The class of limit distributions is given by

$$G(x_1, \dots, x_d) = \exp(-V(x_1, \dots, x_d)) \quad (x_1, \dots, x_d > 0); \quad (2.2.2)$$

these are known as multivariate extreme value (MEV) distributions, and the vector of componentwise maxima must follow this class of distributions. The function  $V$  in (2.2.2) is termed the exponent measure, and is defined as

$$V(x_1, \dots, x_d) = \int_{S_d} \max_{1 \leq j \leq d} \left( \frac{w_j}{x_j} \right) dH(w_1, \dots, w_d) \quad (x_1, \dots, x_d > 0), \quad (2.2.3)$$

where  $H$  is a measure on the  $(d - 1)$ -dimensional unit simplex

$$S_d = \{ \mathbf{w} \in [0, 1]^d : w_1 + \dots + w_d = 1 \quad (j = 1, \dots, d) \},$$

which satisfies

$$\int_{S_d} w_i dH(w_1, \dots, w_d) = 1 \quad (i = 1, \dots, d).$$

Then

The exponent measure  $V$  satisfies two important properties, the first of which is that

$$V(\infty, \dots, \infty, x_k, \infty, \dots, \infty) = \frac{1}{x_k},$$

so that each variable is marginally Fréchet-distributed. The other is homogeneity of order -1, that is, for  $c > 0$ ,

$$V(cx_1, \dots, cx_p) = \frac{1}{c} V(x_1, \dots, x_p); \quad (2.2.4)$$

this property ensures that this class of distributions is max-stable.

As a result of having to differentiate (2.2.2) with respect to each  $x_i$ , likelihood computation is very difficult for MEV distributions for even low-dimensional data, resulting from the need to repeatedly differentiate this distribution function.

An example of an appropriate function  $V$  for the bivariate case is given by

$$V(x, y) = \left( x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}} \right)^\alpha,$$

where  $x > 0, y > 0$  and  $0 < \alpha \leq 1$ . This form is known as the logistic model and was first described by Gumbel (1960). It has the properties that as  $\alpha \rightarrow 0$ , then the associated bivariate distribution function  $G(x, y) \rightarrow \exp\{-\max(x^{-1}, y^{-1})\}$ , corresponding to perfect dependence and when  $\alpha = 1$ ,  $G(x, y) = \exp\{-(x^{-1} + y^{-1})\}$ , which corresponds to independence. Thus, the logistic model is able to capture a wide range of dependence. However, this model assumes that the variables are exchangeable, and so Tawn (1988) develops an asymmetric generalisation of the logistic model, which has exponent measure defined by

$$V(x, y) = \left\{ \left( \frac{\theta}{x} \right)^{\frac{1}{\alpha}} + \left( \frac{\phi}{y} \right)^{\frac{1}{\alpha}} \right\}^\alpha + \frac{1}{x}(1 - \theta) + \frac{1}{y}(1 - \phi),$$

for  $x > 0, y > 0$  with  $0 < \alpha \leq 1$  and  $0 \leq \theta, \phi \leq 1$ . With this model, upon setting  $\theta = \phi = 1$ , the logistic model is obtained; if  $\theta = \phi$ , then this model represents a mixture of the logistic model and independence. Independence is obtained if  $\alpha = 1$ ,  $\theta = 0$  or  $\phi = 0$ , with complete dependence occurring for the case  $\theta = \phi = 1$  as  $\alpha \rightarrow 0$ . A multivariate generalisation of the asymmetric logistic model is given by Tawn (1990).

### 2.2.2 Multivariate point process approach

An alternative method of considering multivariate extremes is to adopt a point process model, as was done earlier for the univariate instance. This multivariate approach is described by Coles and Tawn (1991).

Given an i.i.d. sequence of  $d$ -dimensional random vectors  $\mathbf{X}_1, \mathbf{X}_2, \dots$ , with unit Fréchet marginal distributions and whose joint distribution function lies in the domain of attraction of a multivariate extreme value distribution, define pseudo-radial components

$$R_i = \sum_{j=1}^d \frac{X_{ij}}{n} \quad (i = 1, \dots, n)$$

and angular components

$$W_{ij} = \frac{X_{ij}}{nR_i} \quad (i = 1, \dots, n; j = 1, \dots, d),$$

where  $X_{ij}$  is the  $j$ th component of  $\mathbf{X}_i$ . Then the point process given by

$$P_n = \left\{ \left( \frac{X_{i1}}{n}, \dots, \frac{X_{id}}{n} \right) : i = 1, \dots, n \right\},$$

converges in distribution to a non-homogeneous Poisson process on the space  $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$

with intensity measure

$$\mu(dr \times d\mathbf{w}) = \frac{dr}{r^2} dH(\mathbf{w}),$$

where  $H$  is the measure defined previously.

Coles and Tawn (1991) show how this point process representation leads to the componentwise maxima technique discussed in Section 2.2.1, similarly to as in the univariate point process approach, and note that numerical integration is often required to compute the intensity measure. Moreover, this framework can be used to

model failure probabilities of structures arising from combinations of variables, see Coles and Tawn (1994). However, these methods only work well for variables which satisfy a property known as asymptotic dependence, which is defined in Section 2.3.

### 2.2.3 Multivariate generalised Pareto distribution

Recently, multivariate extensions of the GPD described in Section 2.1.2 have been considered, principally by Rootzen et al. (2018a), Rootzen et al. (2018b) and Kiril-iouk et al. (2019). In this section, we briefly outline the details of this multivariate distribution.

Firstly, using the notation of Section 2.2.1, if we have that the limit (2.2.1) holds, then the  $d$ -dimensional random variable  $\mathbf{X}$  has the property that

$$\left\{ \frac{\mathbf{X} - \mathbf{b}_n}{\mathbf{a}_n} \right\} | \{\mathbf{X} \notin \mathbf{b}_n\}$$

converges, as  $n \rightarrow \infty$ , in distribution to a random variable  $\mathbf{Y}$ , where  $\mathbf{Y}$  follows a multivariate generalised Pareto distribution (MGPD); denote this distribution function by  $H$ .

It is not necessarily the case that the marginal distributions, say  $H_1, \dots, H_d$ , of  $\mathbf{Y} = (Y_1, \dots, Y_d)$  are univariate GPDs. This only arises if the margins are conditioned on being positive, so that

$$H_j^+(y) = \mathbb{P}(Y_j > y | Y_j > 0) = 1 - \left( 1 + \frac{\xi_j y}{\sigma_j} \right)_+^{-\frac{1}{\xi_j}} \quad (j = 1, \dots, d),$$

where  $z_+ = \max(0, z)$  as previously. However, for all  $j$ , if  $\xi_j > 0$ , then  $H_j$  has lower endpoint  $-\sigma_j/\xi_j$ , and has no finite lower endpoint otherwise.

The MGPD  $H$  has various technical properties of use, derived by Rootzen et al. (2018a,b). One of these is the multivariate form of threshold stability, in the sense that if a random variable  $\mathbf{Y} \sim H$  and we have  $\mathbf{u} \geq \mathbf{0}$ ,  $\boldsymbol{\sigma} + \boldsymbol{\xi}\mathbf{u} > \mathbf{0}$  and  $H(\mathbf{u}) < 1$ , then the distribution of  $(\mathbf{Y} - \mathbf{u}) | (\mathbf{Y} \not\leq \mathbf{u})$  also follows an MGPD, with the same shape parameters  $\boldsymbol{\xi}$  as  $H$  but scale parameters  $\boldsymbol{\sigma} + \boldsymbol{\xi}\mathbf{u}$ . Further properties, such as sum-stability and the form of the conditional marginal distributions are discussed in Rootzen et al. (2018a,b) and Kiriliouk et al. (2019).

An important point about the MGPD is that it is possible to link the distribution  $G$  in the limit (2.2.1) to the MGPD  $H$ ; this is through the expression

$$H(\mathbf{x}) = \frac{\log G\{\min(\mathbf{x}, \mathbf{0})\} - \log G(\mathbf{x})}{\log G(\mathbf{0})}.$$

This link implies that the dependence structure of  $H$  is determined entirely by the dependence structure of  $G$  from which  $H$  arises.

## 2.3 Extremal dependence

It is very important to consider the nature of dependence in spatial extremes; in applications, this can be vital as accurate inference of extreme events relies on assessing the characteristics of the dependence between these. The quantities shown in this section will be utilised in subsequent chapters.

Key to understanding this are the notions of asymptotic dependence and asymptotic independence. Coles et al. (1999) give measures of these for bivariate random vectors, and we will detail these, as well as giving an example of a measure for higher-dimensional cases. Before describing these dependence measures, we introduce the

concept of copulas; these are useful tools in understanding the dependence between random variables.

### 2.3.1 Copulas

We outline some key results on copulas, which can be found in Joe (1997). A copula is a multivariate distribution in which each of the margins follows a Uniform(0, 1) distribution. Then for a continuous  $d$ -dimensional distribution function  $F$ , whose  $i$ th marginal distribution is denoted  $F_i$ , then the copula  $C$  associated with  $F$  is a distribution function  $C : [0, 1]^d \rightarrow [0, 1]$  with the property

$$F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d)),$$

where  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ . Additionally, if the margins are continuous and have quantile functions  $F_1^{-1}, \dots, F_d^{-1}$ , then for  $\mathbf{u} = (u_1, \dots, u_d)$ , with each  $u_i \in [0, 1]$ ,

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

is unique. Further, we also note that using the probability integral transform, the Uniform(0, 1) margins may be transformed to any other choice of marginal distribution. The key implication of that is the copula being invariant to transformation. Copulas for multivariate extreme value distributions defined by (2.2.2) must also satisfy the max-stability property; that is,

$$C^m(F_1(x_1), \dots, F_d(x_d)) = C(F_1^m(x_1), \dots, F_d^m(x_d))$$

for all  $\mathbf{x} = (x_1, \dots, x_d)$ .

### 2.3.2 Dependence measures

If a pair of random variables  $(X, Y)$  follow some common marginal distribution  $F$ , then we define

$$\chi = \lim_{z \rightarrow z^F} \mathbb{P}(Y > z | X > z),$$

where  $z^F$  is the upper end-point of  $F$ . Then the measure  $\chi$  provides a natural measure of the dependence between extreme values of  $X$  and  $Y$ ; when  $\chi > 0$  we say that the variables are asymptotically dependent, or exhibit extremal dependence. This may be generalised to the case where  $X \sim F_X$  and  $Y \sim F_Y$  do not follow the same marginal distribution by transforming  $(X, Y)$  to a pair with Uniform(0, 1) margins via the probability integral transform, i.e.,  $(U, V) = (F_X(X), F_Y(Y))$ . Then we have that

$$\chi = \lim_{u \rightarrow 1} \mathbb{P}(V > u | U > u).$$

By considering  $\mathbb{P}(V > u | U > u)$ , an alternative method of calculating  $\chi$  is

$$\begin{aligned} \mathbb{P}(V > u | U > u) &= \frac{\mathbb{P}(U > u, V > u)}{\mathbb{P}(U > u)} \\ &= \frac{1 - 2u + C(u, u)}{1 - u} \\ &= 2 - \frac{1 - C(u, u)}{1 - u} \approx 2 - \frac{\log C(u, u)}{\log u}, \end{aligned}$$

for  $u \approx 1$  and where  $C(\cdot, \cdot)$  is the copula describing the dependence between  $U$  and  $V$  (equivalently  $X$  and  $Y$ ). By defining

$$\chi(u) = 2 - \frac{\log \mathbb{P}(U < u, V < u)}{\log \mathbb{P}(U < u)} \quad (0 \leq u \leq 1),$$

we obtain a sub-asymptotic estimator of  $\chi$ , where  $\chi = \lim_{u \rightarrow 1} \chi(u)$ . The estimator  $\chi(u)$  is useful when assessing the nature of dependence in a dataset, where the

asymptotic properties may not hold.

When  $\chi = 0$ , we say that the variables  $U$  and  $V$  (or  $X$  and  $Y$ ) are asymptotically independent, whilst perfectly dependent variables have  $\chi = 1$ . Furthermore, the sign of  $\chi(u)$  describes whether the variables have positive or negative dependence at the quantile level  $u$ .

However, for asymptotically independent distributions,  $\chi$  cannot determine the strength of dependence, and so an extra measure is needed. First, let  $\bar{F}(x, y) = \mathbb{P}(X > x, Y > y)$  be the joint survivor function of  $X$  and  $Y$ , so

$$\bar{F}(x, y) = 1 - F_X(x) - F_Y(y) + F(x, y) = \bar{C}\{F_X(x), F_Y(y)\},$$

for  $\bar{C}(u, v) = 1 - u - v + C(u, v)$ . Then let

$$\bar{\chi}(u) = \frac{2 \log \mathbb{P}(U > u)}{\log \mathbb{P}(U > u, V > u)} - 1 = \frac{2 \log(1 - u)}{\log \bar{C}(u, u)} - 1 \quad (0 \leq u \leq 1),$$

so that  $-1 < \bar{\chi}(u) \leq 1$ . Similarly to the procedure for  $\chi$ , define  $\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u)$ , with  $-1 < \bar{\chi} \leq 1$ ; in particular asymptotically dependent variables have  $\bar{\chi} = 1$  and variables are asymptotically independent otherwise.

To completely summarise extremal dependence, the pair of measures  $(\chi, \bar{\chi})$  is needed. The case where  $(\chi > 0, \bar{\chi} = 1)$  suggests asymptotic dependence of the variables, and  $\chi$  is considered a measure of the strength of dependence. Conversely,  $(\chi = 0, \bar{\chi} < 1)$  represents the class of asymptotically independent variables; in this case,  $\bar{\chi}$  signifies the strength of dependence between the variables.

Next, the measure  $\bar{\chi}$  is compared to the coefficient of tail dependence,  $\eta$ , discussed by Ledford and Tawn (1996). The quantity  $\eta \in (0, 1]$  arises from the representation,

on Uniform(0, 1) margins,

$$\bar{C}(u, u) \sim L((1-u)^{-1})(1-u)^{1/\eta} \quad (u \rightarrow 1), \quad (2.3.1)$$

where  $L$  is a slowly varying function, i.e.,  $L(tx)/L(x) \rightarrow 1$  as  $x \rightarrow \infty$  for any fixed  $t > 0$ . Then

$$\bar{\chi}(u) \sim \frac{2 \log(1-u)}{\log L((1-u)^{-1}) + \frac{1}{\eta} \log(1-u)} - 1 \rightarrow 2\eta - 1 \quad (u \rightarrow 1);$$

a particular consequence being that  $\bar{\chi} = 2\eta - 1$ . The quantity  $\eta$  provides an alternative measure of the extent of extremal independence, with  $\eta = 1$  corresponding to perfect dependence, and  $\eta = 1/2$  corresponding to perfect independence.

Now suppose that we have i.i.d. Fréchet random variables  $X_1, \dots, X_d$ . Then a measure of extremal dependence is the extremal coefficient, denoted by  $\theta_d$ , and dropping the subscript if  $d = 2$ . The definition, as given in Schlather and Tawn (2003), relies on the homogeneity of order  $-1$  of the exponent measure  $V$  stated in (2.2.4), and is as follows

$$\mathbb{P}(X_1 \leq z, \dots, X_d \leq z) = \exp\left(-\frac{V(1, \dots, 1)}{z}\right) = \exp\left(-\frac{\theta_d}{z}\right) = \left[\exp\left(-\frac{1}{z}\right)\right]^{\theta_d},$$

for  $z > 0$  with  $1 \leq \theta_d \leq d$ . Here it can be seen that the value of  $\theta_d$  gives the effective number of independent variables amongst  $X_1, \dots, X_d$ , so that in particular,  $\theta_d = 1$  for perfectly dependent variables, whilst independent variables have  $\theta_d = d$ , with values between these limits representing different levels of dependence. For bivariate applications, we have that  $\lim_{z \rightarrow \infty} \mathbb{P}(Y > z | X > z) = 2 - \theta$  on Fréchet margins, which gives a natural way of calculating the extent of asymptotic dependence. In Section 2.4.1, we describe values of  $\theta$  permitted for some spatial extremes models,

where dependence between the process at two locations is a function of the distance between the two locations.

A natural multivariate extension of the dependence measure  $\chi$  would be to set

$$\chi_d = \lim_{z \rightarrow \infty} \mathbb{P}(X_2 > z, \dots, X_d > z | X_1 > z).$$

Letting  $C \subseteq \{2, \dots, d\}$ , and noting that  $\mathbb{P}(X_i > z) \sim z^{-1}$  as  $z \rightarrow \infty$ , Eastoe and Tawn (2012) suggest the measure

$$\chi_C = \lim_{z \rightarrow \infty} \left\{ z \mathbb{P} \left( \min_{i \in C} X_i > z \right) \right\},$$

and show that

$$\mathbb{P} \left( \min_{i \in C} X_i > z \right) = L_C \left( \frac{1}{\mathbb{P}(X_1 > z)} \right) \mathbb{P}(X_1 > z)^{\frac{1}{\eta_C}},$$

where  $L_C$  is a slowly varying function, and  $0 < \eta_C \leq 1$ . These provide multivariate extensions of the bivariate measures  $\chi$  and  $\eta$  defined earlier.

## 2.4 Spatial extremes

### 2.4.1 Max-stable processes

Max-stable processes arise as an extension of multivariate extreme value distribution methods, and are commonly used in spatial extreme value applications. Suppose that  $\{W_i(\mathbf{s})\}_{i \in \{1, \dots, n\}}$ , over  $\mathbf{s} \in \mathbb{R}^d$ , is a sequence of  $n$  independent replications of some stationary continuous stochastic process  $W(\cdot)$ . For appropriate sequences  $(a_n(\mathbf{s}) > \mathbf{0})_{n=1}^\infty$  and  $(b_n(\mathbf{s}))_{n=1}^\infty \in \mathbb{R}$ , for all  $\mathbf{s} \in \mathbb{R}^d$ , assume that

$$Z(\mathbf{s}) = \lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} W_i(\mathbf{s}) - b_n(\mathbf{s})}{a_n(\mathbf{s})} \quad (\mathbf{s} \in \mathbb{R}^d).$$

If the limiting process  $Z(\cdot)$  exists and has non-degenerate marginal distributions for all  $\mathbf{s} \in \mathbb{R}^d$ , then  $Z(\cdot)$  is a stationary max-stable process, where the margins follow a GEV distribution. In the particular case where  $W(\cdot)$  has unit Fréchet margins then  $a_n(\mathbf{s}) = n$  and  $b_n(\mathbf{s}) = 0$  for all  $\mathbf{s}$ , the margins of  $Z(\cdot)$  follow a unit Fréchet distribution; we shall work with unit Fréchet margins without loss of generality for the remainder of this section. The joint distribution function of  $Z(\cdot)$  for any subset of sites  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \in \mathbb{R}^d$ , where  $n$  is any element of  $\{1, 2, \dots\}$ , is then given by

$$\mathbb{P}(Z(\mathbf{s}_1) < x_1, \dots, Z(\mathbf{s}_n) < x_n) = G(x_1, \dots, x_n), \quad (2.4.1)$$

where  $G$  is of the form (2.2.2), where the exponent measure  $V$  depends on the distances between the sites. Because  $Z(\cdot)$  is stationary, if we consider the joint distribution (2.4.1) for the set of sites  $\{\mathbf{s}_1 + \boldsymbol{\tau}, \dots, \mathbf{s}_n + \boldsymbol{\tau}\}$ , for any  $\boldsymbol{\tau} \in \mathbb{R}^d$ , the joint distribution obtained is equal to (2.4.1).

We now detail two approaches to the construction of max-stable processes. One construction, by Smith (1990), is as follows. Let  $\{(\mathbf{W}_i, T_i) : i \geq 1\}$  be points of a Poisson process  $\Pi$  on  $\mathbb{R}^d \times \mathbb{R}_+$ . Further suppose that the intensity of  $\Pi$  is given by  $d\Lambda(\mathbf{w}, t) = d\mathbf{w} \times t^{-2} dt$  and  $f$  is a non-negative function on  $\mathbb{R}^d$  such that  $\int_{\mathbb{R}^d} f(\mathbf{s}) d\mathbf{s} = 1$ . Then

$$Z(\mathbf{s}) = \sup_{(\mathbf{w}, t) \in \Pi} t f(\mathbf{s} - \mathbf{w}) \quad (\mathbf{s} \in \mathbb{R}^d) \quad (2.4.2)$$

is a stationary max-stable process with unit Fréchet margins. Smith (1990) interprets this construction as a model for rainfall arising from storms which are centred at the points  $\mathbf{w}$  (uniformly distributed in space), with the function  $f$  defining the shape of these storms, and  $t$  describing the magnitude of a storm. Processes arising from

this construction are often too smooth to provide realistic models for spatial data, however.

Schlather (2002) gives a generalisation of (2.4.2) to permit further models in this framework; we follow the notation as used by Davison et al. (2012). First, let  $\mathbf{o}$  be the origin. If  $W(\cdot)$  is a stationary process on  $\mathbb{R}^d$ ,  $\mathbb{E}[\max\{0, W(\mathbf{o})\}] = \mu \in \mathbb{R}^+$ , and  $\Pi$  is a Poisson process on  $\mathbb{R}^+$  which has intensity measure  $d\Lambda(t) = \mu^{-1}t^{-2}dt$ , then a stationary max-stable process is defined by

$$Z(\mathbf{s}) = \max_{t \in \Pi} t \max\{0, W_i(\mathbf{s})\} \quad (\mathbf{s} \in \mathbb{R}^d), \quad (2.4.3)$$

where the  $W_i(\cdot)$ , for  $i = 1, 2, \dots$ , are i.i.d. replications of  $W(\cdot)$ . The resulting process  $Z(\cdot)$  has unit Fréchet margins. Here  $W(\cdot)$ , when positive, can be interpreted as describing the shape of a storm, if that is the application at hand, with the value of  $t$  being the magnitude of a storm event.

In practice, parameterised models of max-stable processes are used; since bivariate distribution functions are typically the only closed form distribution functions available for max-stable processes, we only provide details of their bivariate exponent measures.

Following Smith (1990), suppose that  $f$  in (2.4.2) is a multivariate normal density function, of dimension  $d$ , with covariance matrix  $\Sigma$ . Then, for locations  $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d$ , the resulting process has bivariate exponent measure

$$V_{\mathbf{h}}(x, y) = \frac{1}{x} \Phi \left( \frac{a(\mathbf{h})}{2} + \frac{1}{a(\mathbf{h})} \log \frac{y}{x} \right) + \frac{1}{y} \Phi \left( \frac{a(\mathbf{h})}{2} + \frac{1}{a(\mathbf{h})} \log \frac{x}{y} \right),$$

where  $\Phi$  is the standard Gaussian distribution function and  $a^2(\mathbf{h}) = \mathbf{h}\Sigma^{-1}\mathbf{h}^T$ , with  $\mathbf{h} = \mathbf{s}_2 - \mathbf{s}_1$ . This model is often referred to as the Smith max-stable process. Equiva-

lently, the Smith max-stable process may be constructed in the form (2.4.3) by taking  $W_i(\mathbf{s}) = f(\mathbf{s} - \mathbf{Y}_i)$ , where  $\mathbf{Y}_i$  arises from a homogeneous Poisson process on  $\mathbb{R}^d$ . In the one-dimensional case,  $\Sigma$  is simply the variance of the Gaussian kernel  $f$  used in the construction of the process.

The Smith model has bivariate extremal coefficient given by  $\theta(\mathbf{h}) = 2\Phi\{a(\mathbf{h})/2\}$  with  $1 \leq \theta(\mathbf{h}) < 2$  for all finite  $h$ ; we see that this has a value of 1 when  $\|\mathbf{h}\| = 0$ , and a limit of 2 as  $\|\mathbf{h}\| \rightarrow \infty$ , where  $\|\cdot\|$  denotes the Euclidean norm. Thus, this max-stable process construction gives asymptotic dependence at all finite distances, with independence only being a limiting case at infinite distance.

Figures 2.4.1a, 2.4.1b and 2.4.1c show simulations of two-dimensional Smith processes with standard Gumbel margins, each on the space  $[0, 10] \times [0, 10]$ , with respective covariance matrices

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \Sigma_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}; \Sigma_3 = \begin{bmatrix} 0.7 & 0.5 \\ 0.5 & 0.7 \end{bmatrix}.$$

We see that the effect of a non-zero covariance term, as in Figures 2.4.1b and 2.4.1c, is to introduce a certain orientation of the events in the process. By changing the diagonal terms, comparing Figures 2.4.1a and 2.4.1c, it is seen that the extreme events can be modelled as being more, or less, localised in the spatial field along the direction of the coordinate axes depending on the magnitude of these diagonal terms.

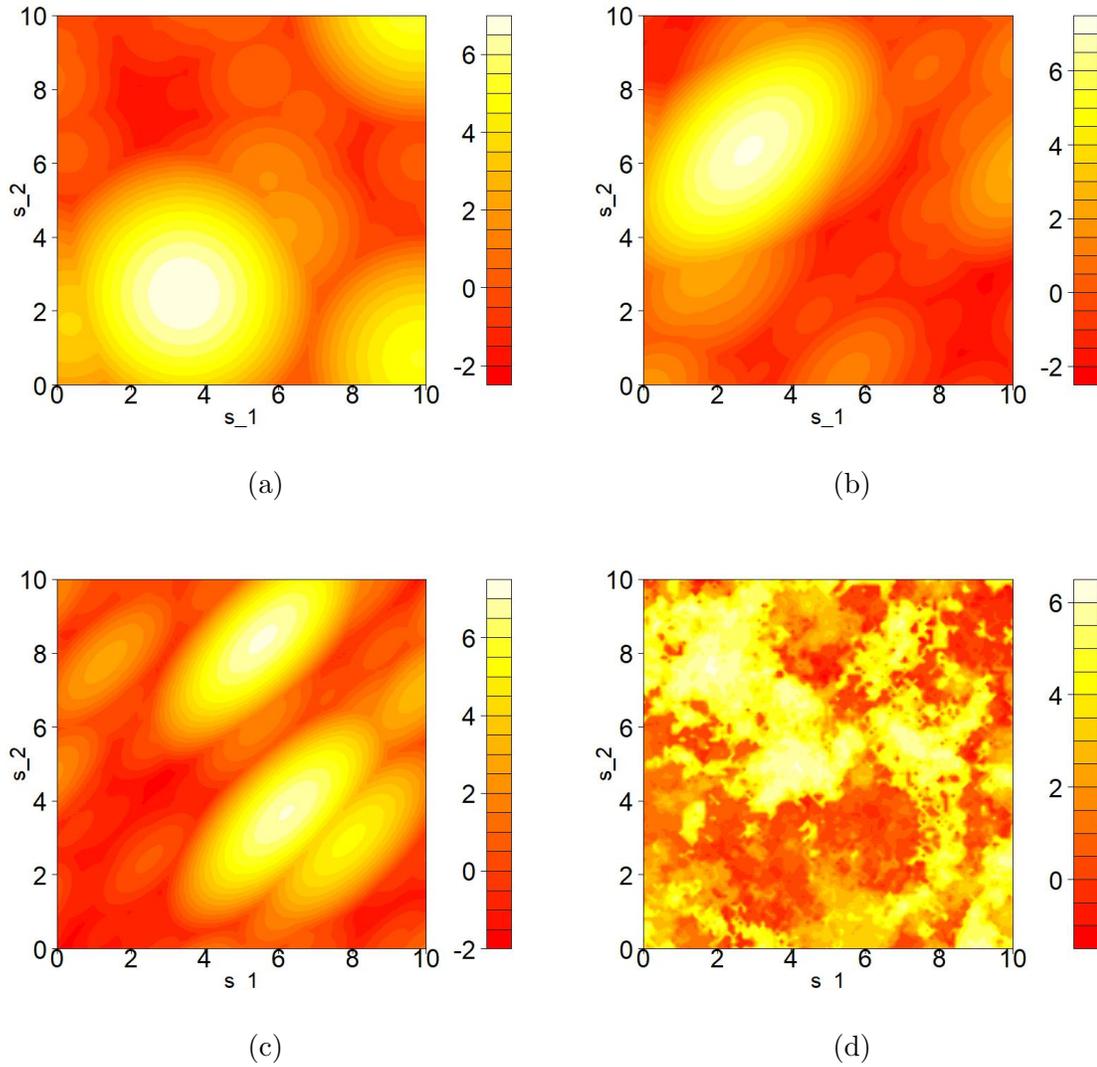


Figure 2.4.1: Simulations of two-dimensional max-stable processes with standard Gumbel margins on the space  $[0, 10] \times [0, 10]$ : (a), (b) and (c) are Smith processes with covariance matrices  $\Sigma_1, \Sigma_2$  and  $\Sigma_3$  respectively; (d) shows a Schlather process, with correlation function  $\rho(\mathbf{h}) = \exp(-\|\mathbf{h}\|)$ .

Suppose now that each  $W_i(\cdot)$  in the representation (2.4.3) is a stationary Gaussian process with correlation function  $\rho(\mathbf{h})$ ; Banerjee et al. (2004) describe a multitude of

choices for  $\rho(\cdot)$ . We then have bivariate exponent measure given by

$$V_{\mathbf{h}}(x, y) = \frac{1}{2} \left( \frac{1}{x} + \frac{1}{y} \right) \left\{ 1 + \sqrt{1 - \frac{2xy(\rho(\mathbf{h}) + 1)}{(x + y)^2}} \right\}.$$

This model was first proposed by Schlather (2002), and is thus commonly termed the Schlather max-stable process. Like the Smith model, we see that the process is stationary. Figure 2.4.1d displays a simulation from the Schlather model with Gumbel margins with correlation function  $\rho(\mathbf{h}) = \exp(-[\|\mathbf{h}\|/\phi]^\alpha)$ , where  $\phi > 0$  and  $0 < \alpha < 2$ ; in this figure,  $\phi = 1$  and  $\alpha = 1$ . Changing  $\phi$  and  $\alpha$  leads to different behaviour of the process, e.g., increasing  $\phi$  leads to longer-range dependence, whilst larger values of  $\alpha$  increase the smoothness of the process.

The process is much less smooth than any realisation obtained from the Smith model, so the model may be more realistic in some scenarios. As  $\theta(\mathbf{h}) \rightarrow 1$  as  $\|\mathbf{h}\| \rightarrow 0$ , the process is near perfectly dependent at small separations. However, if each  $W_i(\cdot)$  is a two-dimensional, stationary and isotropic process, then, as stated by Davison et al. (2012), it is the case that  $\theta(\mathbf{h}) < 1.838$  for all finite  $\mathbf{h}$ , so the Schlather model is unable to capture independence at any distance, and thus cannot capture the whole range of asymptotic dependence. Wadsworth and Tawn (2012b) combine features from the representations (2.4.2) and (2.4.3) to produce a model which has the short-range dependence benefits of the Schlather model whilst also retaining the limiting independence property of the Smith model, in the sense that points an infinite distance apart from one another are independent.

If  $W(\cdot)$  in (2.4.3) is of the form  $W(\mathbf{s}) = \exp[\varepsilon(\mathbf{s}) - \gamma(\mathbf{s})]$ , where  $\varepsilon(\cdot)$  is a stationary

Gaussian process with  $\varepsilon(\mathbf{o}) = 0$  almost surely, which has variogram  $\gamma(\cdot)$ , defined as

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var} \{W(\mathbf{h}) - W(\mathbf{o})\}.$$

The resulting process is termed a Brown-Resnick process, after Brown and Resnick (1977). This has bivariate exponent measure

$$V_{\mathbf{h}}(x, y) = \frac{1}{x} \Phi \left( \frac{a(\mathbf{h})}{2} + \frac{1}{a(\mathbf{h})} \log \left( \frac{y}{x} \right) \right) + \frac{1}{y} \Phi \left( \frac{a(\mathbf{h})}{2} + \frac{1}{a(\mathbf{h})} \log \left( \frac{x}{y} \right) \right), \quad (2.4.4)$$

where  $a^2(\mathbf{h}) = 2\gamma(\mathbf{h})$  and  $\mathbf{h} = \mathbf{s}_2 - \mathbf{s}_1$ . For the Brown-Resnick process, the extremal coefficient is given by  $\theta(\mathbf{h}) = 2\Phi[\sqrt{\gamma(\mathbf{h})/2}]$ ; when  $\gamma(\cdot)$  is unbounded, we have that  $1 \leq \theta(\mathbf{h}) < 2$  for all finite  $\mathbf{h}$ , with  $\theta(\mathbf{h}) \rightarrow 2$  as  $\|\mathbf{h}\| \rightarrow \infty$ . Hence, we obtain the whole range of  $\theta(\mathbf{h})$  between 1 and 2 (Davison et al., 2012), and so the Brown-Resnick max-stable process is able to capture all levels of asymptotic dependence. A particular case of the Brown-Resnick process, assumed in much of the literature, arises if the variogram of  $\varepsilon(\cdot)$  has the form  $\|\mathbf{h}/\lambda\|^\alpha$ , where  $\lambda > 0, 0 < \alpha \leq 2$ , but other forms are possible. Then if  $\alpha = 2$ , we obtain the Smith max-stable process as detailed previously.

In Figure 2.4.2, we display a simulation on the line segment  $[0, 10]$  of a one-dimensional Brown-Resnick process, using the simulation procedure of Dieker and Mikosch (2015), setting  $\lambda = \alpha = 1$  in the above form of the variogram.

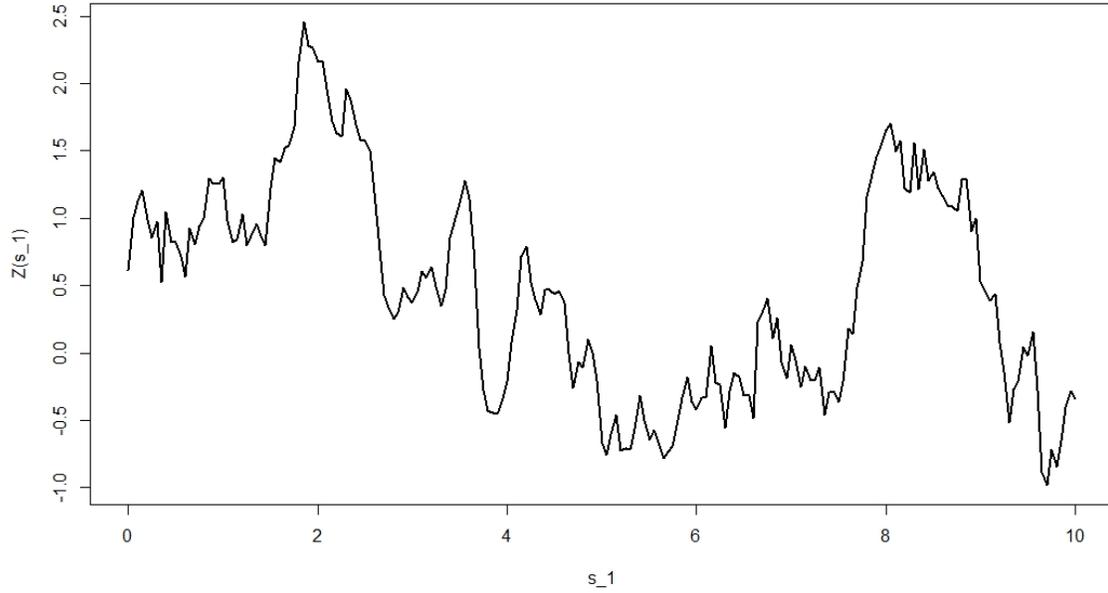


Figure 2.4.2: A simulation of a one-dimensional Brown-Resnick max-stable process, with variogram  $\gamma(h) = \|h\|$ , on standard Gumbel margins on the line segment  $[0, 10]$ .

By considering the extremal coefficients of these three types of max-stable process, we see that these processes exclusively exhibit asymptotically dependence behaviour for any finite distances, with asymptotic independence only achieved in the limit as  $\|\mathbf{h}\| \rightarrow \infty$  for the Smith and Brown-Resnick models, and is not attained for the Schlather max-stable process. Moreover, it is possible to construct max-stable processes with  $\theta(\mathbf{h}) = 2$  for all  $\|\mathbf{h}\| > \tau$  for small  $\tau > 0$ , but these models are fully independent when  $\|\mathbf{h}\| > \tau$ . Therefore, these processes have that  $\eta_{\|\mathbf{h}\|} = 1$  for  $\|\mathbf{h}\| \leq \tau$  and  $\eta_{\|\mathbf{h}\|} = 1/2$  for  $\|\mathbf{h}\| > \tau$ , with  $\eta_{\|\mathbf{h}\|}$  being the pairwise measure  $\eta$  defined in Section 2.3 calculated at sites  $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d$  with  $\mathbf{h} = \mathbf{s}_2 - \mathbf{s}_1$ ; consequently, these processes cannot exhibit dependence in the class of asymptotic independence. Max-stable pro-

cesses of this type arise if  $f(\cdot)$  in (2.4.2) is zero when  $\|(\mathbf{w} - \mathbf{s})\| > \tau$  but  $f(\cdot) > 0$  otherwise.

## 2.4.2 Inference for max-stable processes

The pairwise likelihood is often used for max-stable processes due to the full likelihood requiring the exponent measure  $V$  to be repeatedly differentiated, leading to severe computational issues for even relatively low-dimensional applications. Hence, inference is usually carried out using pairwise likelihood methods, as described by Varin (2008) and Varin et al. (2011). To help explain this approach, first suppose that we have  $n$  sampling locations, at which we observe one realisation of a spatial process  $\mathbf{Z} = [Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)]$ , for  $Z(\cdot)$  defined as in Section 2.4.1, which has density function  $f_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is the vector of (unknown) parameters of interest. Supposing that we can evaluate the bivariate joint density  $f_B$  of  $(Z_r, Z_t) = (Z(\mathbf{s}_r), Z(\mathbf{s}_t))$ , which depends on the separation of the sites  $\mathbf{s}_r$  and  $\mathbf{s}_t$ , then the pairwise likelihood for a realisation  $\mathbf{z}$  of the process  $\mathbf{Z}$  takes the form

$$L(\boldsymbol{\theta}; \mathbf{z}) = \prod_{r=1}^{n-1} \prod_{t=r+1}^n f_B(z_r, z_t; \boldsymbol{\theta}, \mathbf{h}(r, t) = \mathbf{s}_r - \mathbf{s}_t) = \prod_{t>r} f_B(z_r, z_t; \boldsymbol{\theta}, \mathbf{h}(r, t) = \mathbf{s}_r - \mathbf{s}_t),$$

where  $z_j$  is the  $j$ th element of  $\mathbf{z} = (z_1, \dots, z_n)$ . If there are  $m$  i.i.d. observations of  $\mathbf{Z}$ , say  $\mathbf{z}^{(i)} = (z_1^{(i)}, \dots, z_n^{(i)})$  for  $i = 1, \dots, m$ , then the overall pairwise likelihood is

$$L(\boldsymbol{\theta}; \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}) = \prod_{i=1}^m \prod_{t>r} f_B(z_r^{(i)}, z_t^{(i)}; \boldsymbol{\theta}, \mathbf{h}(r, t)). \quad (2.4.5)$$

The particular form of the likelihood for a max-stable process relies on the form given in (2.4.1). If, for sampling locations  $\mathbf{s}_r, \mathbf{s}_t$  in some spatial domain  $\mathcal{S}$ , the bivariate distribution function has the form  $\mathbb{P}(Z(\mathbf{s}_r) < x, Z(\mathbf{s}_t) < y) = F(x, y) =$

$\exp\{-V_{\mathbf{h}(r,t)}(x,y)\}$  for  $x, y > 0$ , where  $V_{\mathbf{h}(r,t)}$  is the exponent measure of the process, we can use the density obtained from this expression in the likelihood (2.4.5) if we have  $m$  i.i.d. observations of  $Z(\cdot)$ , where pairs of sites are considered. In this,  $V_{\mathbf{h}(r,t)}$  may be taken to be one of the forms given in Section 2.4.1.

Denoting the maximum likelihood estimator of  $\boldsymbol{\theta}$  by  $\hat{\boldsymbol{\theta}}$ , obtained by maximising (2.4.5), then the variance matrix of  $\hat{\boldsymbol{\theta}}$  requires use of the so-called sandwich estimator because of the generally incorrect assumption of pairwise independence. Details of the sandwich estimator may be found in Varin et al. (2011).

There have been various methods proposed in order to remove the need of using the pairwise likelihood: Genton et al. (2011) and Huser and Davison (2013) derive triplewise forms of the distribution function for the Smith and Brown-Resnick max-stable processes to improve efficiency, at the cost of computational time; Engelke et al. (2015) use extremal increments of the Brown-Resnick process to form an estimator which utilises Gaussian process, which may be estimated more easily in high dimensions, and Wadsworth and Tawn (2014) make use of having information on the occurrence of maxima in the Stephenson and Tawn (2005) likelihood, which can then be related to the censored likelihood of a Poisson process.

### 2.4.3 Other recent approaches

We now describe some alternative approaches to spatial extremes modelling in the recent literature, and provide a brief summary of each of these.

### Hierarchical modelling

One method of modelling spatial extremes is to make use of a hierarchical model to allow Bayesian inference, such as those developed by Cooley et al. (2007) and Cooley and Sain (2010); both of these articles apply their methods to precipitation data. These comprise of models for the data, the underlying process and the prior distributions, with these three modelling levels being linked together. The benefit of using a hierarchical model is that information can be shared from neighbouring locations, improving inference. Here, we describe the method of Cooley and Sain (2010), and then comment on the differences in the Cooley et al. (2007) hierarchical model.

Firstly, assuming that there are  $N$  sampling locations, the model for the data at each location is based on the point process likelihood (2.1.4), with parameters  $\mu_i, \sigma_i, \xi_i$  corresponding to the  $i$ th sampling location and are common over events. The full likelihood is taken as the product over all sampling locations, assuming independence of the process between all sites, conditional on the marginal parameters  $\{\mu_i, \sigma_i, \xi_i, i = 1, \dots, N\}$ , and combined with a prior suggested by Martins and Stedinger (2000). We note that for the spatial processes considered in Chapters 4, 5 and 6, this conditional independence assumption appears to be unrealistic for applications; however, Sang and Gelfand (2010) allow conditional dependence between sites in their hierarchical model.

Then, the process assumes the following forms for the parameters for  $i \in \{1, \dots, N\}$ :

$$\begin{aligned}\mu_i &\sim N(\mathbf{X}_i^T \boldsymbol{\beta}_\mu + U_{i,\mu}, 1/\tau_\mu^2); \\ \log(\sigma_i) &\sim N(\mathbf{X}_i^T \boldsymbol{\beta}_\sigma + U_{i,\sigma}, 1/\tau_\sigma^2); \\ \xi_i &\sim N(\mathbf{X}_i^T \boldsymbol{\beta}_\xi + U_{i,\xi}, 1/\tau_\xi^2),\end{aligned}$$

with  $N(\eta, \psi^2)$  being a Gaussian distribution with mean  $\eta$  and standard deviation  $\psi$ ;  $\mathbf{X}_i$  represents the covariate information for the  $i$ th sampling site,  $\boldsymbol{\beta}_\theta$  is a vector of the regression coefficients,  $U_{i,\theta}$  is a random effect for parameter  $\theta$  at location  $i$ , and  $\tau_\theta$  is some (fixed) precision, where we take  $\theta$  to represent one of  $\mu, \sigma$  or  $\xi$ . The random effects  $\mathbf{U}_\theta = (U_{1,\theta}, \dots, U_{N,\theta})$  are independent for the different  $\theta$ . Given  $\theta$ , the random effects are spatially dependent, following an autoregressive model across the lattice of sites.

The hierarchical model proposed by Cooley et al. (2007) is broadly similar. However, a GPD is used to model the data, and then latent spatial processes are used for the GPD parameters; a Gaussian process is used to model the process of the log-transformed scale parameter and for the shape parameter. In this framework, the mean vector of the Gaussian process is itself a function of covariates and associated scaling parameters. Under this model, spatial interpolation to new locations can be carried out by using conditional forms of the Gaussian processes. In order to estimate the exceedance rate of the threshold chosen at each site, a further latent spatial process is used, with a binomial distribution used at each location to model the number of declustered threshold exceedances.

### Hierarchical max-stable model of Reich and Shaby (2012)

We now outline an alternative hierarchical max-stable model, introduced by Reich and Shaby (2012). Assume that  $Y(\mathbf{s})$  is some block maximum with marginal distribution  $\text{GEV}(\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}))$ , and that for all  $\mathbf{s} \in \mathbb{R}^d$ ,  $Y(\mathbf{s}) = \mu(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi(\mathbf{s})} \{X(\mathbf{s})^{\xi(\mathbf{s})} - 1\}$ , where  $X(\cdot)$  is the residual max-stable process, having Fréchet margins. Reich and Shaby (2012) model  $X(\mathbf{s})$  as  $U(\mathbf{s})\theta(\mathbf{s})$ , with  $U(\mathbf{s})$  modelled by i.i.d.  $\text{GEV}(1, \alpha, \alpha)$  random variables at each site  $\mathbf{s}$ , where  $0 < \alpha < 1$ , accounting for non-spatial variability, such as measurement error. The process  $\theta(\cdot)$  is taken to be the weighted sum of  $N$  positive i.i.d. random variables  $A_1, \dots, A_N$ , defined below, weighted by  $N$  kernel basis functions  $\{w_j(\mathbf{s}) \geq 0\}$ , for  $j = 1, \dots, N$ , such that  $\sum_{j=1}^N w_j(\mathbf{s}) = 1$  for all  $\mathbf{s} \in \mathbb{R}^d$ . In particular,

$$\theta(\mathbf{s}) = \left[ \sum_{j=1}^N A_j \{w_j(\mathbf{s})\}^{\frac{1}{\alpha}} \right]^{\alpha} \quad (\mathbf{s} \in \mathbb{R}^d),$$

so that  $\theta(\cdot)$  models the spatial variation, and inherits its spatial smoothness from the  $w_j(\mathbf{s})$ . Here, each  $A_k$  has positive stable distribution denoted  $\text{PS}(\alpha)$ , with density  $f(a|\alpha)$  satisfying  $\int_0^\infty \exp(-at)f(a|\alpha)da = \exp(-t^\alpha)$  for  $t \geq 0$ . Then the process  $X(\cdot)$  is max-stable, with Fréchet margins; Reich and Shaby (2012) prove this property. Moreover, for any arbitrary set of  $m$  locations  $\{\mathbf{s}_1, \dots, \mathbf{s}_m\}$ , the joint distribution function of  $X(\cdot)$  is given by

$$\mathbb{P}(X(\mathbf{s}_i) < x_i, i = 1, \dots, m) = \exp \left\{ - \sum_{j=1}^N \left[ \sum_{i=1}^m \left( \frac{w_j(\mathbf{s}_i)}{x_i} \right)^{\frac{1}{\alpha}} \right]^{\alpha} \right\}.$$

The model for  $Y(\cdot)$  is then a hierarchical random effects model, with

$$Y(\mathbf{s}_i) | A_1, \dots, A_N \stackrel{\text{indep.}}{\sim} \text{GEV}[\mu^*(\mathbf{s}_i), \sigma^*(\mathbf{s}_i), \xi^*(\mathbf{s}_i)] \quad (i = 1, \dots, n);$$

$$A_j \stackrel{\text{i.i.d.}}{\sim} \text{PS}(\alpha) \quad (j = 1, \dots, N),$$

where  $\mu^*(\mathbf{s}) = \mu(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi(\mathbf{s})} \{\theta(\mathbf{s})^{\xi(\mathbf{s})} - 1\}$ ,  $\sigma^*(\mathbf{s}) = \alpha\sigma(\mathbf{s})\theta(\mathbf{s})^{\xi(\mathbf{s})}$ ,  $\xi^*(\mathbf{s}) = \alpha\xi(\mathbf{s})$ . The finite-dimensional joint distributions of this model are multivariate GEV distributions, as defined in Chapter 2.2.1. Generally, such joint distributions cannot be expressed in closed form for the max-stable process models in Chapter 2.4.1. Reich and Shaby (2012) utilise MCMC methods for inference under this model, exploiting the conditional independence they assume, with this type of inference being possible representing another advantage of this model over the models in Chapter 2.4.1. Reich and Shaby (2012) show that the Smith max-stable process, an asymptotically dependent process, is a limiting case of the hierarchical max-stable model defined above. In general, however, the hierarchical max-stable models obtained from this method all exhibit asymptotic independence.

### Generalised Pareto processes

Ferreira and de Haan (2014) provide a spatial process analogue of the GPD, termed a generalised Pareto process. Firstly, we note that a simple Pareto process  $X(\cdot)$  can be constructed as

$$X(\mathbf{s}) = RW(\mathbf{s}),$$

where  $R$  is a standard Pareto random variable, and  $W(\cdot)$  is a process which satisfies  $\mathbb{P}[\sup_{\mathbf{s} \in \mathcal{S}} W(\mathbf{s}) > w_0] = 1$  for some constant  $w_0 > 0$ , and  $\mathbb{E}[W(\mathbf{s})] > 0$  for each location

$\mathbf{s}$  in the space of interest  $\mathcal{S}$ . Here,  $R$  and  $W(\cdot)$  are independent. The resulting Pareto process  $X(\cdot)$  has the property that  $\sup_{\mathbf{s} \in \mathcal{S}} W(\mathbf{s})/w_0$  is standard Pareto-distributed. Furthermore, if  $W(\mathbf{s}) = w_0$  for all  $\mathbf{s} \in \mathcal{S}$ , then the process  $X(\cdot)$  is perfectly dependent; however independence between two sites is not possible. Moreover, for a positive spatial process  $Y(\cdot)$ , we have the following link between the distributions of  $X(\cdot)$  and  $Y(\cdot)$ :

$$X(\mathbf{s}) := \lim_{u \rightarrow \infty} \left[ \frac{RY(\mathbf{s})}{u} \middle| \sup_{\mathbf{s} \in \mathcal{S}} RY(\mathbf{s}) > u \right].$$

We then have that a generalised Pareto process  $Z(\cdot)$  can be defined from a simple Pareto process  $X(\cdot)$  by

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \frac{\sigma(\mathbf{s})[X(\mathbf{s})^{\xi(\mathbf{s})} - 1]}{\xi(\mathbf{s})},$$

where  $\mu(\cdot), \xi(\cdot) \in \mathbb{R}$ ,  $\sigma(\cdot) \in \mathbb{R}^+$ .

As well as providing a natural way of describing the behaviour of spatial threshold exceedances, Pareto processes have the advantage over max-stable process that they only require one realisation of  $R$  and  $W(\mathbf{s})$  for simulation, rather than the repeated simulations of a process which are necessary for simulating max-stable processes. Suitable choices of  $W(\cdot)$  can be made in order to make inference simpler than for max-stable processes, though this is still non-trivial in most cases. We note that Pareto processes are in the class of asymptotically dependent processes.

### **Huser and Wadsworth (2018) model for unknown dependence type**

The spatial extremes model developed by Huser and Wadsworth (2018) is able to capture both asymptotic dependence and asymptotic independence behaviour spa-

tially, with the caveat that the process may only model one type of these types of behaviour across all locations. Suppose that  $W(\cdot)$  is a positive stationary spatial process which has Pareto marginal distributions and is asymptotically independent, satisfying (2.3.1); Gaussian processes are an example of such a spatial process when transformed to Pareto margins. Then the spatial dependence model proposed by Huser and Wadsworth (2018) has the form

$$X(\mathbf{s}) = R^\delta W(\mathbf{s})^{1-\delta} \quad (0 \leq \delta \leq 1), \quad (2.4.6)$$

where  $R$  is Pareto random variable, independent of  $W(\cdot)$ . If  $\delta > 1/2$ , then the  $R^\delta$  component is heavier-tailed than  $W(\cdot)^{1-\delta}$ , and we obtain an asymptotically dependent process at all locations. If  $\delta \leq 1/2$ , then the process exhibits asymptotically independent behaviour everywhere. We note that the case  $\delta = 1/2$  is treated specifically by Huser and Wadsworth (2018). Also, we have that as  $\delta \rightarrow 0$ , the copula of the process  $W(\cdot)$  is obtained, whilst as  $\delta \rightarrow 1$ , perfect dependence is seen. The lowest level of dependence possible in the process  $X(\cdot)$  is defined by the corresponding  $W(\cdot)$  process.

An alternative representation can be obtained by taking logarithms of (2.4.6), specifically we have  $\tilde{X}(\mathbf{s}) = \delta \tilde{R} + (1 - \delta) \tilde{W}(\mathbf{s})$ , where  $\tilde{R}$  is a unit exponential random variable, and  $\tilde{W}(\mathbf{s}) = \log[W(\mathbf{s})]$ , which is independent of  $\tilde{R}$ , has unit exponential margins. From this characterisation, and considering  $\tilde{R}$  as a spatial process, it is seen that the process  $\tilde{X}(\cdot)$  can be interpreted as a sum of a perfectly dependent process, arising from  $\tilde{R}$ , and an asymptotically independent process, this being contributed by  $\tilde{W}(\cdot)$ . Here,  $\delta$  is fixed for all distances between sites; in Chapters 4, 5 and 6,

we describe models where the dependence type can be modelled flexibly for different distances, in essence by allowing  $\delta$  to change with the distance between sites.

For inference, Huser and Wadsworth (2018) take  $W(\cdot)$  to be a Gaussian process, since it has a relatively simple representation in high dimensions compared to other spatial processes. We note that since joint distributions are still required to be calculated for inference, the model still is computationally difficult for a large number of sampling locations. The authors also point out that since positive association of extremes still occurs as the distance between sites becomes infinitely large if  $\delta > 1/3$ , the model is better suited to small spatial regions.

## 2.5 Conditional extremes methods

We now introduce an alternative method of approaching multivariate extreme value problems. Heffernan and Tawn (2004) first proposed a conditional extremes model, which has since been generalised by Heffernan and Resnick (2007), with the aim that extrapolation to events of practical interest is made much easier than the methods Chapter 2.2 has outlined. Here, we give an overview of the model, its theoretical justification, and its properties, as conditional extremes modelling forms the basis of the content in Chapters 4, 5 and 6.

## 2.5.1 Theoretical background

### General results

We first provide some theoretical justification for the conditional extremes model, following results from Heffernan and Tawn (2004). Suppose that we have a vector of random variables  $\mathbf{Y} = (Y_1, \dots, Y_d)$  which have Gumbel margins (although any marginal distribution with exponential upper tail may be chosen), and that interest lies in the behaviour of  $\mathbb{P}(\mathbf{Y}_{-i} \leq \mathbf{y}_{-i} | Y_i = y_i)$  for  $i \in \{1, \dots, d\}$ , where  $\mathbf{Y}_{-i}$  denotes the vector  $\mathbf{Y}$  with its  $i$ th component removed. Now, assume that for any  $i \in \{1, \dots, d\}$  we have functions  $\mathbf{a}_{|i} : \mathbb{R} \rightarrow \mathbb{R}^{d-1}$ ,  $\mathbf{b}_{|i} : \mathbb{R} \rightarrow \mathbb{R}^{d-1}$  such that

$$\mathbf{Z}_{|i} = \frac{\mathbf{Y}_{-i} - \mathbf{a}_{|i}(y_i)}{\mathbf{b}_{|i}(y_i)}$$

has the property that

$$\mathbb{P}(\mathbf{Z}_{|i} \leq \mathbf{z}_{|i} | Y_i = y_i) \rightarrow G_{|i}(\mathbf{z}_{|i}) \quad (y_i \rightarrow \infty), \quad (2.5.1)$$

where the joint distribution function  $G_{|i}$  has non-degenerate margins.

It follows under weak assumptions that (2.5.1) implies that for  $y_i > 0$

$$\mathbb{P}(\mathbf{Z}_{|i} \leq \mathbf{z}_{|i}, Y_i - u_i > y_i | Y_i > u_i) \rightarrow G_{|i}(\mathbf{z}_{|i}) \exp(-y_i) \quad (u_i \rightarrow \infty), \quad (2.5.2)$$

so that the random variables  $Y_i - u_i$  and  $\mathbf{Z}_{|i}$  are independent in the limit. The limiting result holds by virtue of the exponential upper tail of the Gumbel distribution and its memoryless property, as well as the limit (2.5.1).

We note at this point that alternatively, Heffernan and Resnick (2007) show that another representation can be attained by normalising  $\mathbf{Y}_{-i}$  by  $u_i$  instead; however,

the Heffernan and Tawn (2004) method is easier from a statistical perspective, and so we concentrate on this approach.

To obtain marginal distributions of  $G_{j|i}$ , define for  $j \in \{1, \dots, d\}$ ,  $j \neq i$ ,

$$G_{j|i}(z_{j|i}) = \lim_{y_i \rightarrow \infty} \mathbb{P}(Z_{j|i} \leq z_{j|i} | Y_i = y_i),$$

where

$$Z_{j|i} = \frac{Y_j - a_{j|i}(y_i)}{b_{j|i}(y_i)},$$

and  $a_{j|i}(\cdot)$ ,  $b_{j|i}(\cdot)$  are the components of  $\mathbf{a}_{j|i}(\cdot)$ , respectively  $\mathbf{b}_{j|i}(\cdot)$ , associated with  $Y_j$ .

Then with this definition,  $G_{j|i}$  is the marginal distribution of the joint distribution  $G_{j|i}$  related to the variable  $Y_j$ .

To choose the normalising functions  $\mathbf{a}_{j|i}(\cdot)$  and  $\mathbf{b}_{j|i}(\cdot)$ , we look at the behaviour of  $F_{j|i}(y_j|y_i) = \mathbb{P}(Y_j < y_j | Y_i = y_i)$ , as  $G_{j|i}$  must be non-degenerate for each  $j \neq i$ . Heffernan and Tawn (2004) then give the following result on properties that  $\mathbf{a}_{j|i}$  and  $\mathbf{b}_{j|i}$  must satisfy.

**Theorem 3:** Suppose that  $\mathbf{Y}$  has a continuous joint density function. Then if for any  $i$ ,  $\mathbf{a}_{j|i}(\cdot)$ , and  $\mathbf{b}_{j|i} > 0$  that satisfy the limit (2.5.1), then the components associated with  $Y_j$ , for  $j \neq i$  satisfy the following (up to type):

- $\lim_{y_i \rightarrow \infty} F_{j|i}(a_{j|i}(y_i)|y_i) = p_{j|i}$ , where  $p_{j|i} \in (0, 1)$  is a constant.
- $b_{j|i}(y_i) = h_{j|i}[a_{j|i}(y_i)|y_i]^{-1}$ , where

$$h_{j|i}(y_j|y_i) = \frac{f_{j|i}(y_j|y_i)}{1 - F_{j|i}(y_j|y_i)} \quad (y_j \in \mathbb{R}),$$

with  $f_{j|i}$  being the density arising from  $F_{j|i}$ .

We note that the normalising functions are not unique, and that these functions are identifiable only up to type. Details of this can be found in Heffernan and Tawn (2004).

### Examples of normalising functions and general form

We now describe some examples of normalising functions described by Theorem 3, and the general form given by Heffernan and Tawn (2004) that these must take. To motivate this, Table 2.5.1, which is adapted from Table 1 of Heffernan and Tawn (2004), lists examples of functions  $a_{j|i}$ ,  $b_{j|i}$  under various conditions.

Dependence type	$a_{j i}(y_i)$	$b_{j i}(y_i)$	Form of $G_{ i}$
Perfect positive dependence	$y_i$	1	Degenerate
Asymptotic dependence (MEV distribution)	$y_i$	1	‡
Asymptotic independence (multivariate Gaussian)	$\rho_{ij}^2 y_i$	$\sqrt{y_i}$	Gaussian
Complete independence	0	1	Gumbel

Table 2.5.1: Table of normalising constants  $a_{j|i}$ ,  $b_{j|i}$ , as well as the forms of the limiting distribution  $G_{|i}$  for a variety of cases of dependence between variables; ‡ refers the reader to Heffernan and Tawn (2004) for details of this limiting form. This table has been adapted from Table 1 of Heffernan and Tawn (2004).

To provide an example of how to find such normalising functions, assume that we have a vector of random variables  $(X_F, Y_F)$ , which has Fréchet marginal distributions,

and joint distribution function

$$G_F(x, y) = \exp \left\{ - \left( x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}} \right)^\alpha \right\} \quad (x, y > 0; 0 < \alpha < 1).$$

Since  $\log(X_F)$  transforms  $X_F$  (equivalently,  $Y_F$ ) to follow a Gumbel marginal distribution, call this transformed variable  $X_G$  (equivalently  $Y_G$ ), we have that

$$G_G(x, y) = \mathbb{P}(X_G < x, Y_G < y) = \mathbb{P}(X_F < e^x, Y_F < e^y) = e^{-V(e^x, e^y)},$$

where  $V(x, y) = (x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}})^\alpha$ . Letting  $y = a(x) + b(x)z$  and  $f_{X_G}$  be the density of  $X_G$ , then

$$\lim_{x \rightarrow \infty} \mathbb{P} \left( \frac{Y_G - a(x)}{b(x)} \leq z | X_G = x \right) = \lim_{x \rightarrow \infty} \frac{\frac{\partial}{\partial x} G_G(x, y)}{f_{X_G}(x)} \Big|_{y=a(x)+b(x)z}$$

This limiting term has the form

$$\frac{\frac{\partial}{\partial x} G_G(x, y)}{f_{X_G}(x)} = \frac{\exp\{-[\exp(-x/\alpha) + \exp(-y/\alpha)]^\alpha - \frac{x}{\alpha}\} \{\exp(-x/\alpha) + \exp(-y/\alpha)\}^{\alpha-1}}{\exp\{-x - \exp(-x)\}}, \quad (2.5.3)$$

where upon substituting  $y = a(x) + b(x)z$ , (2.5.3) is equivalent to

$$\begin{aligned} & \exp \left\{ - \left[ \exp \left( -\frac{x}{\alpha} \right) + \exp \left( -\frac{a(x) + b(x)z}{\alpha} \right) \right]^\alpha - \frac{x}{\alpha} + x + \exp(-x) \right\} \\ & \times \left[ \exp \left( -\frac{x}{\alpha} \right) + \exp \left( -\frac{a(x) + b(x)z}{\alpha} \right) \right]^{\alpha-1} \\ & = \exp \left\{ - \left[ \exp \left( -\frac{x}{\alpha} \right) + \exp \left( -\frac{a(x) + b(x)z}{\alpha} \right) \right]^\alpha + \exp(-x) \right\} \\ & \times \exp \left( -\frac{x}{\alpha} + x \right) \exp \left( \frac{x}{\alpha} - x \right) \left\{ 1 + \exp \left[ -\frac{a(x) - x + b(x)z}{\alpha} \right] \right\}^{\alpha-1}. \end{aligned}$$

Here, the first exponential term in the final equality approaches unity as  $x \rightarrow \infty$ , and so we need to ensure that the final term of the final line above is a non-degenerate distribution function as  $x \rightarrow \infty$ ; clearly we require that  $a(x) = x$ , and where  $b(x) = c$ ,

for some constant  $c > 0$  which can be arbitrarily chosen as  $c = 1$ . These choices of  $a(x)$  and  $b(x)$  are then equivalent to finding functions  $\mathbf{a}_{|1}$ ,  $\mathbf{b}_{|1}$  in the bivariate case of Theorem 3, and a logistic distribution is obtained as the limiting distribution  $G_{|i}$ , that is,  $G_{|i}(z) = \{1 + \exp(-z/\alpha)\}^{\alpha-1}$ .

The normalising functions suggested in this example and by Table 2.5.1 take a particular form. Heffernan and Tawn (2004) give these forms as

$$\mathbf{a}_{|i}(y_i) = \tilde{\mathbf{a}}_{|i} y_i + \mathbb{I}_{\{\tilde{\mathbf{a}}_{|i}=0, \tilde{\mathbf{b}}_{|i}<0\}} [\tilde{\mathbf{c}}_{|i} - \tilde{\mathbf{d}}_{|i} \log(y_i)]; \quad (2.5.4)$$

$$\mathbf{b}_{|i}(y_i) = y_i^{\tilde{\mathbf{b}}_{|i}}, \quad (2.5.5)$$

where  $\tilde{\mathbf{a}}_{|i}$ , and similar terms, are vector constants and  $\mathbb{I}$  denoting an indicator function on the subscripted set. The vectors  $\tilde{\mathbf{a}}_{|i}$ ,  $\tilde{\mathbf{b}}_{|i}$ ,  $\tilde{\mathbf{c}}_{|i}$ ,  $\tilde{\mathbf{d}}_{|i}$  have components such that  $\tilde{a}_{j|i}, \tilde{d}_{j|i} \in [0, 1]$ ,  $\tilde{b}_{j|i} \in (-\infty, 1)$ ,  $\tilde{c}_{j|i} \in \mathbb{R}$ , where  $j \neq i$ . We note that the purpose of the indicator function in (2.5.4) is to account for the possible presence of negative association between variables.

Heffernan and Resnick (2007) generalise this formulation; they show that under weak assumptions for the joint distribution of  $\mathbf{Y}_{-i}$ , the normalising functions  $\mathbf{a}_{|i}(\cdot)$  and  $\mathbf{b}_{|i}(\cdot)$  must be regularly varying with specific constraints; in particular, if Laplace marginal distributions are used, then each component of  $\mathbf{a}_{|i}$  must be regularly varying of index 1, and each component of  $\mathbf{b}_{|i}(\cdot)$  regularly varying of index less than 1.

Keef et al. (2013b) provide an alternative formulation if using Laplace, rather than Gumbel, marginal distributions, that the normalising functions take the form  $\mathbf{a}_{|i}(y_i) = \boldsymbol{\alpha}_{|i} y$ ,  $\mathbf{b}_{|i}(y_i) = y_i^{\boldsymbol{\beta}_{|i}}$ . In this instance, for each  $j \neq i$ ,  $\alpha_{j|i} \in [-1, 1]$ ,  $\beta_{j|i} \in (-\infty, 1)$ ; if  $-1 \leq \alpha_{j|i} < 0$  corresponds to negative association between the  $j$  and

$i$ th variables (conditional on the  $i$ th variable being sufficiently large), with positive association of these variables occurring if  $0 < \alpha_{j|i} \leq 1$ . This representation will be used in subsequent chapters.

## 2.5.2 Conditional extremes models

Having introduced the theoretical background of the conditional extremes approach of Heffernan and Tawn (2004), we now describe how this is used for modelling the behaviour of extremes, and interpretation of the model.

### Heffernan and Tawn (2004) model and properties

The limiting distribution in (2.5.1), along with its independence property, are key to application of the model. Assume that for  $i = 1, \dots, d$ , there is some threshold  $u_i$  for which the limit (2.5.2) holds exactly; i.e., for  $y_i > 0$ ,

$$\mathbb{P}(\mathbf{Y}_{-i} < \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i}, Y_i - u_i > y_i | Y_i > u_i) = G_{|i}(\mathbf{z}_{|i}) \exp(-y_i).$$

The conditional extremes dependence model, with Laplace margins and setting  $\mathbf{a}_{|i}(y_i) = \boldsymbol{\alpha}_{|i}y_i$ ,  $\mathbf{b}_{|i} = y_i^{\boldsymbol{\beta}_{|i}}$  as proposed by Keef et al. (2013b), is thus given by

$$\mathbf{Y}_{-i} | \{Y_i = y_i\} = \boldsymbol{\alpha}_{|i}y_i + y_i^{\boldsymbol{\beta}_{|i}}\mathbf{Z}_{|i}, \quad (2.5.6)$$

for all  $y_i > u_i$ , where  $\mathbf{Z}_{|i}$  is independent of  $Y_i$ , and hence amounts to a non-linear regression model once  $Y_i$  is sufficiently large.

There is no specific form implied for the distribution of  $\mathbf{Z}_{|i}$  in order to calculate  $\boldsymbol{\alpha}_{|i}$  and  $\boldsymbol{\beta}_{|i}$ ; indeed Heffernan and Tawn (2004) outline how this may be modelled empirically. It is, however, convenient to use a  $(d - 1)$ -dimensional multivariate Gaussian

distribution for ease of computation in the estimation of  $\alpha_{|i}$  and  $\beta_{|i}$ . Other suggestions for modelling  $Z_{|i}$ , whilst retaining computational ease of estimating  $\alpha_{|i}$  and  $\beta_{|i}$ , include using a mixture of multivariate Gaussian distributions via a Dirichlet process (Lugrin et al., 2016b), and using a Gaussian copula on kernel-smoothed marginal distributions (Towe et al., 2019). In general, Heffernan and Tawn (2004) give that the mean and standard deviation vectors of  $\mathbf{Y}_{-i}|Y_i = y_i > u_i$  are given by  $\alpha_{|i}y_i + y_i^{\beta_{|i}}\boldsymbol{\mu}_{|i}$  and  $y_i^{\beta_{|i}}\boldsymbol{\sigma}_{|i}$  respectively, where  $\boldsymbol{\mu}_{|i}$  and  $\boldsymbol{\sigma}_{|i}$  are vectors of the marginal means and standard deviations. Thus, if for a pair of variables  $(Y_i, Y_j)$ ,  $Z_{j|i}$  is taken to follow a Gaussian distribution with mean  $\mu_{j|i}$  and standard deviation  $\sigma_{j|i}$  (i.e.,  $\mathbf{Z}_{|i}$  follows a multivariate Gaussian distribution with some appropriate covariance structure), then

$$Y_{j|i} \sim \text{N}(\alpha_{j|i}y_i + y_i^{\beta_{j|i}}\mu_{j|i}, y_i^{2\beta_{j|i}}\sigma_{j|i}^2).$$

A drawback of the model is the issue of self-consistency of parameters when conditioning upon different variables, i.e., that each model of  $\mathbf{Y}_{-i}|Y_i$  for  $i = 1, \dots, d$  is consistent with the others. Heffernan and Tawn (2004) describe a variety of properties necessary for this to hold under asymptotic dependence, however they note that conditions for the asymptotic independence case are difficult to characterise. Hence, it is suggested that no additional structure is imposed to ensure self-consistency, since the data are from a valid joint distribution, thus there should not be any great departure from self-consistency upon conditioning. Moreover, Heffernan and Tawn (2004) find that the performance of the model suffers when imposing self-consistency conditions. Liu and Tawn (2014) discuss self-consistency further, describing how definitions of self-consistency may be different for different subsets of the sample space and con-

straints required for these.

### Simulation from Heffernan and Tawn (2004) model

Heffernan and Tawn (2004) provide details on how to simulate easily from the conditional extremes model (2.5.6), conditioning on a particular variable. Their simulation algorithm allows estimation of probabilities of events via Monte Carlo approximation from the samples generated. However, we focus on the rejection sampling method of Keef et al. (2013b) which generates events that may arise from any variable being large.

Using the same notation as above, denote the sample space by  $\mathcal{Y} = \{\mathbf{y} \in \mathbb{R}^d : y_i > u_i \text{ for some } i = 1, \dots, d\}$ , and partition this space into subsets  $\mathcal{Y}_i = \{\mathbf{y} \in \mathbb{R}^d : (y_i > u_i) \cap (y_i = \max[\mathbf{y}])\}$  for  $i = 1, \dots, d$ , where  $\max(\mathbf{y})$  denotes the maximum component of  $\mathbf{y}$ . Suppose that we wish to simulate  $M$  events; then the number of samples which lie in  $\mathcal{Y}_i$  follows a multinomial distribution which has  $M$  samples and event probabilities  $\mathbb{P}(\mathbf{Y} \in \mathcal{Y}_i)/\mathbb{P}(\mathbf{Y} \in \mathcal{Y})$  for each  $i = 1, \dots, d$ . Probabilities of the type  $\mathbb{P}(\mathbf{Y} \in \mathcal{Y}_i)$  may be calculated by using the fitted conditional extremes model, as described by Heffernan and Tawn (2004). The simulation algorithm of an event of the form  $\mathbf{Y} | (\mathbf{Y} \in \mathcal{Y}_i)$  is then prescribed by Keef et al. (2013b) as follows.

1. Generate  $E \sim \text{Exp}(1)$ , and set  $Y_i^* = u_i + E$ .
2. Independently of  $Y_i^*$ , choose a realisation  $\mathbf{Z}_{|i}^*$  of the residuals  $\mathbf{Z}_{|i}$ ; this is appropriate only if estimating  $G_{|i}$  by the empirical distribution of  $\mathbf{Z}_{|i}$ .
3. Set  $Y_j^* = a_{j|i}(Y_i^*) + b_{j|i}(Y_i^*)Z_{j|i}^*$ , for  $j = 1, \dots, d, j \neq i$ , using the fitted param-

ters for  $\mathbf{a}_i(\cdot)$  and  $\mathbf{b}_i(\cdot)$ .

4. If  $Y_i^* < \max_{j \in \{1, \dots, d\} \setminus \{i\}} Y_j^*$ , return to step 1; stop otherwise.

A benefit of simulation being carried out in this manner is that the probability of a given variable having the largest non-exceedance probability is able to be different across variables; this means that different levels of dependence between variables can be accounted for. However, when using this procedure for the simulation of spatial fields, Wadsworth and Tawn (2019) note a number of deficiencies in this method, such as only being able to simulate conditional upon extreme values at sampling locations rather than arbitrary sets of locations, and possibly requiring a large number of rejections to obtain suitable samples. Wadsworth and Tawn (2019) propose a simulation algorithm for the spatial extension of the conditional extremes model to overcome these simulation issues, as well as other issues, by utilising importance sampling rather than the rejection sampler of Keef et al. (2013b).

### 2.5.3 Linking conditional extremes to the Brown-Resnick max-stable process

Finally, we look at how conditioning on extremes events can be used to better understand the properties of existing max-stable process models; this motivates the use of the conditional spatial extremes models introduced later in this thesis. We calculate limiting conditional distributions of the Brown-Resnick max-stable process, as defined in Section 2.4.1, and find that we obtain natural closed form expressions upon doing so.

**Bivariate case**

For data  $(X, Y)$  arising from a Brown-Resnick process with Laplace margins as defined in Section 2.4.1, suppose we are interested in the behaviour of  $\mathbb{P}(Y - u < z | X = u)$ , with limiting distributional form  $G(z)$  as  $u \rightarrow \infty$ , noting that

$$\lim_{u \rightarrow \infty} \frac{\frac{\partial}{\partial x} \mathbb{P}(Y < x + z, X < x)|_{x=u}}{\frac{1}{2}e^{-u}} = G(z). \quad (2.5.7)$$

To transform between a Fréchet-distributed variable  $X_F$  to a variable  $X_L$  following a Laplace distribution, we consider, for  $e^{-\frac{1}{X_F}} > 0.5$ ,

$$1 - \frac{1}{2}e^{-X_L} = e^{-\frac{1}{X_F}} \quad \Rightarrow \quad X_L = -\log \left\{ 2 \left( 1 - e^{-\frac{1}{X_F}} \right) \right\}, \quad X_F = -\frac{1}{\log(1 - e^{-X_L}/2)}. \quad (2.5.8)$$

By (2.5.8) we have that  $X_F \sim 2e^{X_L}$  as  $X_L \rightarrow \infty$ ; this arises by considering  $\log(1 - w)$  for  $w \approx 0$ .

Thus,

$$G_L(x, y) = \mathbb{P}(X_L < x, Y_L < y) \sim \mathbb{P}(X_F < 2e^x, Y_F < 2e^y) = e^{-V(h(x), h(y))},$$

where  $V$  is as defined by (2.4.4), dropping the subscript  $\mathbf{h}$ , and  $h(w) = 2e^w$ , noting the distinction of this function from the vector  $\mathbf{h}$ . Hence,

$$\frac{\partial G_L(x, y)}{\partial x} = -h'(x)V_1(h(x), h(y))e^{-V(h(x), h(y))}, \quad (2.5.9)$$

where  $h'(w) \sim 2e^w$  (taking the derivative with respect to  $w$ ) as  $w \rightarrow \infty$ , and  $V_1$  is the derivative of  $V$  with respect to the first argument. Dropping the argument from the function  $a^2(\mathbf{h}) = \gamma(\mathbf{h})/2$ , with  $\gamma(\cdot)$  as defined in Section 2.4.1, for notational convenience, we have

$$V(s, t) = \frac{1}{s}\Phi \left\{ a + \frac{1}{2a} \log \left( \frac{t}{s} \right) \right\} + \frac{1}{t}\Phi \left\{ a + \frac{1}{2a} \log \left( \frac{s}{t} \right) \right\},$$

so  $V_1$  takes the form

$$\begin{aligned} V_1(s, t) = & -\frac{1}{s^2} \Phi \left\{ a + \frac{1}{2a} \log \left( \frac{t}{s} \right) \right\} - \frac{1}{2as^2} \phi \left\{ a + \frac{1}{2a} \log \left( \frac{t}{s} \right) \right\} \\ & + \frac{1}{2sta} \phi \left\{ a + \frac{1}{2a} \log \left( \frac{s}{t} \right) \right\}, \end{aligned}$$

where  $\Phi$  is the standard Gaussian distribution function and  $\phi$  is the standard Gaussian density function. Recalling that  $h(w) = 2e^w$ , we have

$$\begin{aligned} V_1(h(x), h(y)) \approx & -\frac{1}{(2e^x)^2} \Phi \left\{ a + \frac{1}{2a} (y - x) \right\} - \frac{1}{2a(2e^x)^2} \phi \left\{ a + \frac{1}{2a} \log (y - x) \right\} \\ & + \frac{1}{2a(2e^x)(2e^y)} \Phi \left\{ a + \frac{1}{2a} \log (x - y) \right\} \end{aligned}$$

Then, letting  $x = u, y = u + z$ ,

$$V_1(h(u), h(u + z)) \sim e^{-2u} \left[ -\frac{1}{4} \Phi \left( a + \frac{z}{2a} \right) - \frac{1}{8a} \phi \left( a + \frac{z}{2a} \right) + \frac{e^{-z}}{8a} \phi \left( a - \frac{z}{2a} \right) \right].$$

Now, note that

$$\begin{aligned} e^{-z} \phi \left( a - \frac{z}{2a} \right) &= \frac{e^{-z}}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( a - \frac{z}{2a} \right)^2 \right\} \\ &= \frac{e^{-z}}{\sqrt{2\pi}} \times \exp \left\{ -\frac{1}{2} a^2 + \frac{z}{2} - \frac{1}{2} \left( \frac{z}{2a} \right)^2 \right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{a^2}{2} - \frac{z}{2} - \frac{1}{2} \left( \frac{z}{2a} \right)^2 \right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[ a^2 + z + \left( \frac{z}{2a} \right)^2 \right] \right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( a + \frac{z}{2a} \right)^2 \right\} \\ &= \phi \left( a + \frac{z}{2a} \right). \end{aligned}$$

Hence, for large  $u$ ,

$$\begin{aligned} \frac{\frac{\partial}{\partial x} \mathbb{P}(Y < u + z, X < u)}{\frac{1}{2}e^{-u}} &\sim \frac{-h'(u)V_1(h(u), h(u+z))e^{-V(h(u), h(u+z))}}{\frac{1}{2}e^{-u}} \\ &\sim \frac{-(2e^u) \times \left\{ -\frac{1}{4}e^{-2u}\Phi\left(a + \frac{z}{2a}\right) \right\}}{\frac{1}{2}e^{-u}} \\ &= \Phi\left(a + \frac{z}{2a}\right), \end{aligned}$$

where the second line is due to the fact that  $\lim_{u \rightarrow \infty} V(h(u), h(u+z)) = 0$ . Thus, we have that  $G(z) = \Phi\left(a + \frac{z}{2a}\right)$ , i.e., the limiting marginal distribution is a Gaussian distribution. Moreover, note that the conditional normalising functions are given by  $a_i(y_i) = y_i$  and  $b_i(y_i) = 1$ , so the process is asymptotically dependent for all values of  $\mathbf{h}$ .

### Trivariate case

For data  $\mathbf{W} = (W_1, W_2, W_3)$  on Fréchet margins arising from a three-dimensional Brown-Resnick process, Huser and Davison (2013) give that

$$\mathbb{P}(W_1 \leq w_1, W_2 \leq w_2, W_3 \leq w_3) = G_F(w_1, w_2, w_3) = \exp\{-V(w_1, w_2, w_3)\}, \quad (2.5.10)$$

where

$$\begin{aligned} V(w_1, w_2, w_3) &= \frac{1}{w_1} \Phi_2\{\eta(w_1, w_2), \eta(w_1, w_3); R_1\} \\ &\quad + \frac{1}{w_2} \Phi_2\{\eta(w_2, w_1), \eta(w_2, w_3); R_2\} \\ &\quad + \frac{1}{w_3} \Phi_2\{\eta(w_3, w_1), \eta(w_3, w_2); R_3\}. \end{aligned}$$

Here,  $\Phi_2(\cdot, \cdot; R)$  is a bivariate normal distribution function with mean  $\mathbf{0}$ , unit variances and correlation matrix  $R$ , and

$$\eta(w_i, w_j) = \sqrt{\frac{\gamma_{ij}}{2}} - \frac{\log\left(\frac{w_i}{w_j}\right)}{\sqrt{2\gamma_{ij}}},$$

with  $\gamma_{ij} = \gamma(\mathbf{s}_i - \mathbf{s}_j)$  for  $\mathbf{s}_i \in \mathbb{R}^d$ , and where  $\gamma(\cdot)$  is the variogram of the process.

Furthermore, we have

$$R_1 = \frac{\gamma_{12} + \gamma_{13} - \gamma_{23}}{2\sqrt{\gamma_{12}\gamma_{13}}}, \quad R_2 = \frac{\gamma_{12} + \gamma_{23} - \gamma_{13}}{2\sqrt{\gamma_{12}\gamma_{23}}}, \quad R_3 = \frac{\gamma_{13} + \gamma_{23} - \gamma_{12}}{2\sqrt{\gamma_{12}\gamma_{23}}}.$$

In this case, we are interested in the behaviour as  $u \rightarrow \infty$  of

$$\mathbb{P}(\mathbf{Y} - \mathbf{1}u < \mathbf{z} | X = u) \rightarrow G(\mathbf{z}) \tag{2.5.11}$$

for variables  $(X, \mathbf{Y})$  on Laplace margins, where  $X = X_1$  and  $\mathbf{Y} = (X_2, X_3)$ . We calculate expression (2.5.11) via

$$\lim_{u \rightarrow \infty} \frac{\frac{\partial}{\partial x} \mathbb{P}(\mathbf{Y} < \mathbf{1}u + \mathbf{z}, X < x) |_{x=u}}{\frac{1}{2}e^{-u}} = G(\mathbf{z}).$$

We again utilise the result of (2.5.8) and set  $h(w) = 2e^w$ , giving

$$G_L(x_1, x_2, x_3) = \exp\{-V(h(x_1), h(x_2), h(x_3))\},$$

so that

$$\frac{\partial G_L(x_1, x_2, x_3)}{\partial x_1} = -h'(x_1)V_1(h(x_1), h(x_2), h(x_3))e^{-V(h(x_1), h(x_2), h(x_3))}.$$

Now, define  $a'_i(\cdot)$  [or, where appropriate,  $a^{(i)}(\cdot)$ ] to be the partial derivative of  $a(\cdot)$

with respect to the  $i^{\text{th}}$  argument of  $a$ , we have

$$\begin{aligned}
V_1(x_1, x_2, x_3) &= -\frac{1}{x_1^2} \Phi_2 \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\
&\quad + \frac{\eta_1(x_1, x_2)}{x_1} \Phi_2^{(1)} \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\
&\quad + \frac{\eta_1(x_1, x_3)}{x_1} \Phi_2^{(2)} \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\
&\quad + \frac{\eta_2(x_2, x_1)}{x_2} \Phi_2^{(1)} \{ \eta(x_2, x_1), \eta(x_2, x_3); R_2 \} \\
&\quad + \frac{\eta_2(x_3, x_1)}{x_3} \Phi_2^{(1)} \{ \eta(x_3, x_1), \eta(x_3, x_2); R_3 \}.
\end{aligned}$$

Since

$$\eta_1(x_i, x_j) = \frac{1}{\sqrt{2\gamma_{ij}}} \times \left\{ -\frac{1/x_j}{x_i/x_j} \right\} = -\frac{1}{x_i\sqrt{2\gamma_{ij}}} \text{ and } \eta_2(x_i, x_j) = \frac{1}{x_j\sqrt{2\gamma_{ij}}},$$

then  $V_1$  can be reformulated as

$$\begin{aligned}
V_1(x_1, x_2, x_3) &= -\frac{1}{x_1^2} \Phi_2 \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\
&\quad - \frac{1}{x_1^2\sqrt{2\gamma_{12}}} \Phi_2^{(1)} \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\
&\quad - \frac{1}{x_1^2\sqrt{2\gamma_{13}}} \Phi_2^{(2)} \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\
&\quad + \frac{1}{x_1x_2\sqrt{2\gamma_{12}}} \Phi_2^{(1)} \{ \eta(x_2, x_1), \eta(x_2, x_3); R_2 \} \\
&\quad + \frac{1}{x_1x_3\sqrt{2\gamma_{13}}} \Phi_2^{(1)} \{ \eta(x_3, x_1), \eta(x_3, x_2); R_3 \}. \tag{2.5.12}
\end{aligned}$$

Now, note that for continuous random variables  $V$  and  $W$ ,

$$\frac{\partial}{\partial v} \mathbb{P}(V < v, W < w) = \mathbb{P}(W \leq w | V = v) f_V(v),$$

where  $f_V$  is the density function of  $V$ , so if  $(V, W)$  follow a bivariate normal distribu-

tion with standard Gaussian margins and correlation matrix  $R_1$ ,

$$\begin{aligned} \Phi_2^{(1)} \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} &= \mathbb{P}(W \leq \eta(x_1, x_3) | V = \eta(x_1, x_2)) \times \phi \{ \eta(x_1, x_2) \} \\ &= \Phi \left\{ \frac{\eta(x_1, x_3) - R_1 \eta(x_1, x_2)}{\sqrt{1 - R_1^2}} \right\} \phi \{ \eta(x_1, x_2) \}; \end{aligned} \quad (2.5.13)$$

this follows since if  $(T_1, T_2) \sim N(\mathbf{0}, \Sigma)$ , where  $\Sigma$  is the  $2 \times 2$  correlation matrix with off-diagonal elements  $\rho$ , then  $\mathbb{E}(T_1 | T_2 = t_2) = \rho t_2$  and  $\text{Var}(T_1 | T_2 = t_2) = 1 - \rho^2$ , so that  $T_1 | T_2 = t_2 \sim N(\rho t_2, 1 - \rho^2)$ . Then perform the obvious standardisation. Moreover, similar terms to (2.5.13) can be found for other terms requiring the partial derivative of the bivariate normal density. Hence, (2.5.12) becomes

$$\begin{aligned} V_1(x_1, x_2, x_3) &= -\frac{1}{x_1^2} \Phi_2 \{ \eta(x_1, x_2), \eta(x_1, x_3); R_1 \} \\ &\quad - \frac{1}{x_1^2 \sqrt{2\gamma_{12}}} \phi \{ \eta(x_1, x_2) \} \Phi \left[ \frac{\eta(x_1, x_3) - R_1 \eta(x_1, x_2)}{\sqrt{1 - R_1^2}} \right] \\ &\quad - \frac{1}{x_1^2 \sqrt{2\gamma_{13}}} \phi \{ \eta(x_1, x_3) \} \Phi \left[ \frac{\eta(x_1, x_2) - R_1 \eta(x_1, x_3)}{\sqrt{1 - R_1^2}} \right] \\ &\quad + \frac{1}{x_1 x_2 \sqrt{2\gamma_{12}}} \phi \{ \eta(x_2, x_1) \} \Phi \left[ \frac{\eta(x_2, x_3) - R_2 \eta(x_2, x_1)}{\sqrt{1 - R_2^2}} \right] \\ &\quad + \frac{1}{x_1 x_3 \sqrt{2\gamma_{13}}} \phi \{ \eta(x_3, x_1) \} \Phi \left[ \frac{\eta(x_3, x_2) - R_3 \eta(x_3, x_1)}{\sqrt{1 - R_3^2}} \right]. \end{aligned}$$

Substituting  $h(w) = 2e^w$ , then we have

$$\begin{aligned}
V_1(h(x_1), h(x_2), h(x_3)) &= -\frac{1}{(2e^{x_1})^2} \Phi_2 \{ \eta(2e^{x_1}, 2e^{x_2}), \eta(2e^{x_1}, 2e^{x_3}); R_1 \} \\
&\quad - \frac{1}{(2e^{x_1})^2 \sqrt{2\gamma_{12}}} \phi_1 \{ \eta(2e^{x_1}, 2e^{x_2}) \} \Phi_1 \left[ \frac{\eta(2e^{x_1}, 2e^{x_3}) - R_1 \eta(2e^{x_1}, 2e^{x_2})}{\sqrt{1 - R_1^2}} \right] \\
&\quad - \frac{1}{(2e^{x_1})^2 \sqrt{2\gamma_{13}}} \phi_1 \{ \eta(2e^{x_1}, 2e^{x_3}) \} \Phi_1 \left[ \frac{\eta(2e^{x_1}, 2e^{x_2}) - R_1 \eta(2e^{x_1}, 2e^{x_3})}{\sqrt{1 - R_1^2}} \right] \\
&\quad + \frac{1}{2e^{x_1} 2e^{x_2} \sqrt{2\gamma_{12}}} \phi_1 \{ \eta(2e^{x_2}, 2e^{x_1}) \} \Phi_1 \left[ \frac{\eta(2e^{x_2}, 2e^{x_3}) - R_2 \eta(2e^{x_2}, 2e^{x_1})}{\sqrt{1 - R_2^2}} \right] \\
&\quad + \frac{1}{2e^{x_1} 2e^{x_3} \sqrt{2\gamma_{13}}} \phi_1 \{ \eta(2e^{x_3}, 2e^{x_1}) \} \Phi_1 \left[ \frac{\eta(2e^{x_3}, 2e^{x_2}) - R_3 \eta(2e^{x_3}, 2e^{x_1})}{\sqrt{1 - R_3^2}} \right].
\end{aligned}$$

Let  $a_{ij} = \sqrt{\gamma_{ij}/2} = a_{ji}$ ; then we have that

$$\eta(2e^{x_i}, 2e^{x_j}) = \sqrt{\frac{\gamma_{ij}}{2}} - \frac{\log\left(\frac{2e^{x_i}}{2e^{x_j}}\right)}{\sqrt{2\gamma_{ij}}} = a_{ij} - \frac{(x_i - x_j)}{2a_{ij}}.$$

Also let  $x_1 = u$ ,  $x_2 = u + z_2$ ,  $x_3 = u + z_3$ , so that, by a similar calculation found in the bivariate case,

$$\begin{aligned}
e^{-z_2} \phi_1 \{ \eta(2e^{x_2}, 2e^{x_1}) \} &= e^{-z_2} \phi \{ \eta(2e^{u+z_2}, 2e^u) \} \\
&= e^{-z_2} \phi \left\{ a_{21} - \frac{z_2}{2a_{21}} \right\} \\
&= \phi \left\{ a_{12} + \frac{z_2}{2a_{12}} \right\}.
\end{aligned}$$

This leads to

$$\begin{aligned}
V_1(h(u), h(u+z_2), h(u+z_3)) &= -\frac{1}{4e^{2u}} \Phi_2 \left\{ a_{12} + \frac{z_2}{2a_{12}}, a_{13} + \frac{z_3}{2a_{13}}; R_1 \right\} \\
&\quad - \frac{1}{8a_{12}e^{2u}} \phi \left\{ a_{12} + \frac{z_2}{2a_{12}} \right\} \Phi \left[ \frac{a_{13} + \frac{z_3}{2a_{13}} - R_1(a_{12} + \frac{z_2}{2a_{12}})}{\sqrt{1-R_1^2}} \right] \\
&\quad - \frac{1}{8a_{13}e^{2u}} \phi \left\{ a_{13} + \frac{z_3}{2a_{13}} \right\} \Phi \left[ \frac{a_{12} + \frac{z_2}{2a_{12}} - R_1(a_{13} + \frac{z_3}{2a_{13}})}{\sqrt{1-R_1^2}} \right] \\
&\quad + \frac{1}{8a_{12}e^{2u}} \phi \left\{ a_{12} + \frac{z_2}{2a_{12}} \right\} \Phi \left[ \frac{a_{23} + \frac{(z_3-z_2)}{2a_{23}} - R_2(a_{12} - \frac{z_2}{2a_{12}})}{\sqrt{1-R_2^2}} \right] \\
&\quad + \frac{1}{8a_{13}e^{2u}} \phi \left\{ a_{13} + \frac{z_3}{2a_{13}} \right\} \Phi \left[ \frac{a_{32} + \frac{(z_2-z_3)}{2a_{23}} - R_3(a_{13} - \frac{z_3}{2a_{13}})}{\sqrt{1-R_3^2}} \right].
\end{aligned}$$

Then the limiting form is given by

$$\begin{aligned}
G(\mathbf{z}) &= \lim_{u \rightarrow \infty} \frac{-h'(u)V_1(h(u), h(u+z_2), h(u+z_3))e^{-V(h(u), h(u+z_2), h(u+z_3))}}{\frac{1}{2}e^{-u}} \\
&= \lim_{u \rightarrow \infty} -4e^{2u} \times V_1(h(u), h(u+z_2), h(u+z_3)) \\
&= \Phi_2 \left\{ a_{12} + \frac{z_2}{2a_{12}}, a_{13} + \frac{z_3}{2a_{13}}; R_1 \right\} \\
&\quad - \frac{1}{2a_{12}} \phi \left\{ a_{12} + \frac{z_2}{2a_{12}} \right\} \Phi \left[ \frac{a_{13} + \frac{z_3}{2a_{13}} - R_1(a_{12} + \frac{z_2}{2a_{12}})}{\sqrt{1-R_1^2}} \right] \\
&\quad - \frac{1}{2a_{13}} \phi \left\{ a_{13} + \frac{z_3}{2a_{13}} \right\} \Phi \left[ \frac{a_{12} + \frac{z_2}{2a_{12}} - R_1(a_{13} + \frac{z_3}{2a_{13}})}{\sqrt{1-R_1^2}} \right] \\
&\quad + \frac{1}{2a_{12}} \phi \left\{ a_{12} + \frac{z_2}{2a_{12}} \right\} \Phi \left[ \frac{a_{23} + \frac{(z_3-z_2)}{2a_{23}} - R_2(a_{12} - \frac{z_2}{2a_{12}})}{\sqrt{1-R_2^2}} \right] \\
&\quad + \frac{1}{2a_{13}} \phi \left\{ a_{13} + \frac{z_3}{2a_{13}} \right\} \Phi \left[ \frac{a_{32} + \frac{(z_2-z_3)}{2a_{23}} - R_3(a_{13} - \frac{z_3}{2a_{13}})}{\sqrt{1-R_3^2}} \right]. \tag{2.5.14}
\end{aligned}$$

Whilst it is not obvious that the final four terms in the limit (2.5.14) analytically sum to zero, numerical tests we have carried out suggest these terms do cancel each other out for each of a wide range of parameters that we have tested. Thus, the limiting distribution appears to be given by  $\Phi_2 \left\{ a_{12} + \frac{z_2}{2a_{12}}, a_{13} + \frac{z_3}{2a_{13}}; R_1 \right\}$ . Hence, we see

that Gaussian closed-form expressions arise for the limiting conditional distributions of established multivariate, equivalently spatial, methods for extremes. In particular, we see that the form of  $G$  for higher dimensions is a multivariate Gaussian distribution, thus the assumptions of the Heffernan and Tawn (2004) model would be correct in this instance, despite the original process being a Brown-Resnick max-stable process. Moreover, upon extending this to the spatial case, the residual process  $\mathbf{Z}$  would follow a Gaussian process. The overall spatial process would be asymptotically dependent, since we require normalising functions  $\mathbf{a}_i(y_i) = y_i$ , and  $\mathbf{b}_i(y_i) = 1$ . This motivates our use of the Heffernan and Tawn (2004) conditional extreme value model for use in spatial applications, since we can see that with a Gaussian process assumed for  $\mathbf{Z}$ , our model is a natural extension of the Brown-Resnick max-stable process model when considered conditionally.

## Chapter 3

# Comparison of censored likelihood methods under asymptotic independence

This chapter will present an investigation into the effects of different censored pairwise likelihoods for data simulated from asymptotically independent random fields with known parameters. The purpose of the investigation will be to compare the performance of the censored likelihood method proposed by Ledford and Tawn (1996) against a suggested method justified for asymptotically independent data by Wadsworth and Tawn (2012b); the latter of these is computationally easier but may introduce additional bias. As a measure of performance, we shall compute dependence measures  $\eta(h)$  and  $\chi(u; h)$ , as described in Section 2.3, based on the estimated parameters in each case, and ascertain their root mean square error (RMSE), bias and standard deviation from the true values of these. The investigation will comprise of both fit-

ting an inverted Brown-Resnick max-stable process, defined in Section 3.2, pairwise likelihood to data simulated from a Gaussian process (both of which are asymptotically independent processes), and vice-versa. A similar study under the assumption of asymptotic dependence has been performed by Huser et al. (2016), using a wider range of likelihood approaches than will be presented here.

### 3.1 Censored likelihood methods

Recall that, from Section 2.4.1, pairwise likelihood methods are commonly utilised for inference on max-stable processes (MSPs). Since inverted max-stable processes (IMSPs) are constructed from the max-stable process models introduced in Section 2.4.1, they suffer from the same computational issues that MSPs have. Thus, pairwise likelihood approaches are also used for inference on IMSPs.

However, fitting a pairwise likelihood for all available data will induce bias in parameter estimation. The class of IMSPs are spatial copulas which permit a wide range of forms of  $\eta(h)$ , informing about the rate of convergence to  $\chi(u; h) = 0$ . As such, data below an appropriate threshold are not informative about this behaviour. In practice, a censored likelihood approach is taken to overcome this issue, using only the full bivariate density in a region of the (pairwise) sample space assumed to provide a good approximation to the limiting model and using alternative likelihood contributions elsewhere. We will consider two constructions of such censoring methods for extreme values.

Suppose the random field of interest is  $X(\cdot)$ , over some spatial domain  $\mathcal{S}$ , and

consider a realisation  $\mathbf{x}$  of the random field, using similar notation to Section 2.4.2. Let  $F_{ij}(\cdot, \cdot; \boldsymbol{\theta})$  be the pairwise distribution function given by  $F(x_i, x_j; \boldsymbol{\theta})$ , and denote by  $f_{ij}(\cdot, \cdot; \boldsymbol{\theta})$  its associated pairwise density function. Choosing some suitable censoring threshold  $v$ , one possible pairwise censoring approach proposed by Ledford and Tawn (1996) has pairwise likelihood contributions given by

$$L_{ij}^{(LT)}(\boldsymbol{\theta}) = \begin{cases} F_{ij}(v, v; \boldsymbol{\theta}) & \text{if } \max(x_i, x_j) \leq v; \\ \frac{\partial}{\partial z_j} F_{ij}(v, x_j; \boldsymbol{\theta}) & \text{if } x_i < v, x_j > v; \\ \frac{\partial}{\partial z_i} F_{ij}(x_i, v; \boldsymbol{\theta}) & \text{if } x_j < v, x_i > v; \\ f_{ij}(x_i, x_j; \boldsymbol{\theta}) & \text{if } \min(x_i, x_j) > v. \end{cases} \quad (3.1.1)$$

These contributions may then be used in a likelihood of the form (2.4.2) to provide the censored likelihood.

An alternative censoring scheme that we shall consider is given by Wadsworth and Tawn (2012b), where the pairwise censored likelihood contributions are

$$L_{ij}^{(WT)}(\boldsymbol{\theta}) = \begin{cases} f_{ij}(x_i, x_j; \boldsymbol{\theta}) & \text{if } \max(x_i, x_j) > v; \\ F_{ij}(v, v; \boldsymbol{\theta}) & \text{if } \max(x_i, x_j) \leq v, \end{cases} \quad (3.1.2)$$

with the same notation used as for the Ledford and Tawn (1996) censoring scheme. This form is motivated for data arising from processes which exhibit asymptotic independence, since then approximating the likelihood by the full likelihood is appropriate when just one of the variables is large, as an asymptotically independent process is likely to be extreme in just one component. Thus, despite the fact that using the full likelihood in such instances may introduce more bias to parameter estimates than

using partial contributions, such as those suggested by Ledford and Tawn (1996), it may be the case that the bias introduced is sufficiently small that the computational benefits of only calculating two likelihood contributions may outweigh any additional error in parameter estimation.

In order to investigate this, we will simulate data from two asymptotically independent processes, a Gaussian process, and an inverted Brown-Resnick max-stable process. We then specify the censored likelihoods described above arising from an inverted Brown-Resnick process or Gaussian process, respectively, so that each data set has a misspecified censored likelihood; this is done as data usually arise from an unknown process, so misspecification provides a more natural assessment of performance. By computing the bias, variance and root mean square errors of comparing estimated values of dependence measures with the true values of these measures, then a direct comparison of the two censoring schemes under misspecification can be made and conclusions drawn on which method is preferred.

## **3.2 Inverted max-stable processes and resulting properties**

Recalling that max-stable processes provide a framework for modelling spatial extremes under the assumption of asymptotic dependence, Wadsworth and Tawn (2012b) provide a method of obtaining a random field which exhibits asymptotically independent behaviour. By inverting the copula associated with a max-stable process, a corresponding inverted max-stable process (IMSP) which has asymptotically inde-

pendent behaviour can be found. The copula is inverted by transforming the original copula by a monotonically decreasing function; the particular result that we use here is that if  $Z_F(\cdot)$  is a max-stable process with unit Fréchet margins, then for  $\mathbf{s} \in \mathcal{S}$ , where  $\mathcal{S}$  is the spatial domain of interest,

$$Z_E^*(\mathbf{s}) = \frac{1}{Z_F(\mathbf{s})} \quad (3.2.1)$$

defines an IMSP with standard exponential margins. Then, for an MSP with bivariate distribution function given by  $\mathbb{P}(Z_F(\mathbf{s}_1) < x, Z_F(\mathbf{s}_2) < y) = \exp(-V(x, y))$  for  $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$  and exponent measure defined as in Section 2.4.1, the associated IMSP obtained through (3.2.1) has the property that

$$\mathbb{P}(Z_E^*(\mathbf{s}_1) > x, Z_E^*(\mathbf{s}_2) > y) = \exp\{-K(x, y)\} = \exp\left\{-V\left(\frac{1}{x}, \frac{1}{y}\right)\right\}. \quad (3.2.2)$$

In particular, for a Brown-Resnick max-stable process, we have that

$$V\left(\frac{1}{x}, \frac{1}{y}\right) =: K(x, y) = x\Phi\left\{\frac{a}{2} + \frac{1}{a}\log\left(\frac{x}{y}\right)\right\} + y\Phi\left\{\frac{a}{2} + \frac{1}{a}\log\left(\frac{y}{x}\right)\right\}, \quad (3.2.3)$$

using the notation of Section 2.4.1. As a result of (3.2.2), we note that the distribution function of an IMSP on exponential margins is

$$F_E(x, y) = 1 - \exp(-x) - \exp(-y) + \exp\{-K(x, y)\}. \quad (3.2.4)$$

We also have that

$$K_1(x, y) := \frac{\partial K(x, y)}{\partial x} = \Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{x}{y}\right) + \frac{1}{a}\phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{x}{y}\right) - \frac{y}{ax}\phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{y}{x}\right)$$

and, similarly,

$$K_2(x, y) := \frac{\partial K(x, y)}{\partial y} = \Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{y}{x}\right) + \frac{1}{a}\phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{y}{x}\right) - \frac{x}{ay}\phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{x}{y}\right).$$

Hence, we obtain

$$\begin{aligned} \frac{\partial K(x, y)}{\partial x \partial y} = K_{12}(x, y) &= \frac{1}{a^2 y} \left( \frac{a}{2} + \frac{1}{a} \log \frac{x}{y} \right) \phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{x}{y} \right) \\ &+ \frac{1}{a^2 x} \left( \frac{a}{2} + \frac{1}{a} \log \frac{y}{x} \right) \phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{y}{x} \right) \\ &- \frac{1}{ay} \phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{x}{y} \right) \\ &- \frac{1}{ax} \phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{y}{x} \right). \end{aligned}$$

Note that the above derivation utilises the fact that  $\phi'(z) = -z\phi(z)$ . We can then use these in the bivariate density function, derived from (3.2.4), given by

$$f_E(x, y) = K_1(x, y)K_2(x, y)e^{-K(x, y)} - K_{12}(x, y)e^{-K(x, y)}. \quad (3.2.5)$$

### 3.3 Dependence measures for asymptotically independent processes

We now describe the forms of the dependence measures for IMSPs and Gaussian processes, which we shall use as the basis of comparison of the misspecified models. Recall the dependence measures  $\bar{\chi}$ ,  $\eta$  and  $\theta$  defined in Section 2.3.2, for which we shall utilise spatial counterparts in the subsequent analyses of this chapter.

Firstly, to calculate  $\eta(h)$  theoretically for an IMSP  $Z_E^*(\cdot)$ , we use the property that  $\eta(h) = 1/\theta(h)$  (Wadsworth and Tawn, 2012b), where  $\theta(h)$  is the extremal coefficient of the associated MSP; for a Brown-Resnick max-stable process and  $h > 0$ , this is given by

$$\theta(h) = 2\Phi \left( \sqrt{\frac{\gamma(h)}{2}} \right), \quad (3.3.1)$$

where  $\gamma(\cdot)$  is the semi-variogram of the process. In particular for this study, this will be assumed to have the form

$$\gamma(h) = \left(\frac{h}{\lambda}\right)^\alpha \quad (\lambda > 0, \alpha \in (0, 2]).$$

We can also calculate  $\chi(u; h)$  by considering

$$\begin{aligned} \chi(u; h) &= \frac{\mathbb{P}(Z_E^*(\mathbf{s}_1) > F_E^{-1}(u), Z_E^*(\mathbf{s}_2) > F_E^{-1}(u))}{1 - u} = \frac{\exp[-K(F_E^{-1}(u), F_E^{-1}(u))]}{1 - u} \\ &= (1 - u)^{K(1,1)-1}, \end{aligned}$$

where  $h = \|\mathbf{s}_2 - \mathbf{s}_1\|$ ,  $F_E$  is a standard exponential distribution and  $K$  is defined as in (3.2.3). The final equation follows by considering  $F_E^{-1}(u) = -\log(1 - u)$  and the homogeneity of order  $-1$  of  $K$ . In this expression,  $h$  may be considered as fixed with respect to the choice of  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , and calculated for a range of thresholds  $u \in [0, 1]$ .

For a Gaussian process  $X_G(\cdot)$ , we have that

$$\eta(h) = \frac{1 + \rho(h)}{2},$$

as given by Ledford and Tawn (1996) where  $\rho(\cdot)$  is the underlying covariance function and calculate  $\chi(u; h)$  for  $h = \|\mathbf{s}_2 - \mathbf{s}_1\|$  (where  $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$ ) by

$$\chi(u; h) = \frac{\mathbb{P}(X_G(\mathbf{s}_1) > F_G^{-1}(u), X_G(\mathbf{s}_2) > F_G^{-1}(u))}{1 - u},$$

where  $F_G$  is a standard normal distribution function. If  $X_G(\cdot)$  is a centred Gaussian process, this quantity may be calculated numerically using the joint distribution function with mean vector  $\mathbf{0}$ , and covariance matrix determined by  $\rho(\cdot)$ .

### 3.4 IMSP censored likelihoods fitted to Gaussian process data

To check the performance of the two censoring methods described above, we calculate the dependence measures  $\eta(h)$  and  $\chi(u; h)$ , based on the theoretical results from Section 3.3 for the model which has been simulated from, and compare these to the corresponding measures calculated from parameter estimates arising from the misspecified model fitted via the two censored likelihood schemes. When fitting these, we use (3.2.4) and (3.2.5), where we have

$$\frac{\partial F(z_i, z_j)}{\partial z_i} = e^{-z_i} - K_1(z_i, z_j)e^{-K(z_i, z_j)},$$

and  $K$  and  $K_1$  defined as in Section 3.2. A similar result holds for  $\frac{\partial F}{\partial z_j}$ , with  $K_2$  also defined as before. When using the censoring methods outlined in Section 3.1, we set  $v$  to be the 0.95 quantile.

Data are simulated from a one-dimensional Gaussian process at 31 uniformly-spaced locations on the line segment  $\mathcal{S} = [0, 12]$  and we fit the incorrect model using the censored likelihoods (3.1.1) and (3.1.2). When simulating the Gaussian process, we use the exponential covariance function  $\rho(h) = \exp(-h/\phi)$ , where  $h = \|\mathbf{s}_j - \mathbf{s}_k\|$  for each sampling location  $\mathbf{s}_i \in \mathcal{S}$ .

Before presenting the results of a more thorough simulation study, we look at examples of the comparison between the theoretical values of  $\eta(h)$  and  $\chi(u; h)$  for data simulated from the Gaussian process detailed above. For these examples, a Gaussian process has been simulated with 1000 replications and with  $\phi = 2$  in the exponential

covariance function  $\rho(\cdot)$ . Figure 3.4.1 shows an example comparison of theoretical values of  $\eta(h)$  for the inverted Brown-Resnick max-stable processes with parameters estimated by both censoring schemes, setting  $v = 0.95$ , along with the corresponding true value of  $\eta(h)$  for the true Gaussian process model, given in Section 3.3. Figure 3.4.2 then shows an example of a comparison of theoretical  $\chi(u; h)$  for the estimated inverted Brown-Resnick max-stable processes, along with the theoretical values for the true model; this has been performed with  $h = 4$ . Note that the estimates in Figures 3.4.1 and 3.4.2 have been calculated from only one simulation of the process; the simulation studies below utilise 100 of these simulations.

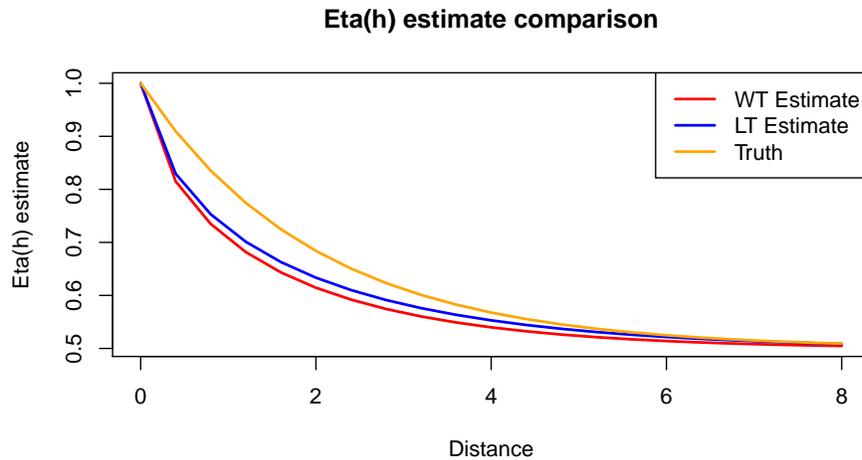


Figure 3.4.1: Comparison of theoretical values of  $\eta(h)$  using estimated parameter values of  $\alpha$  and  $\lambda$ . The red line corresponds to the theoretical result under the estimate obtained from the Wadsworth and Tawn (2012b) censoring scheme, the blue line corresponds to the theoretical result under the estimate obtained from the Ledford and Tawn (1996) censoring scheme and the orange line represents the theoretical value of  $\eta(h)$  for the simulated Gaussian process, using  $n = 1000$  replicates.

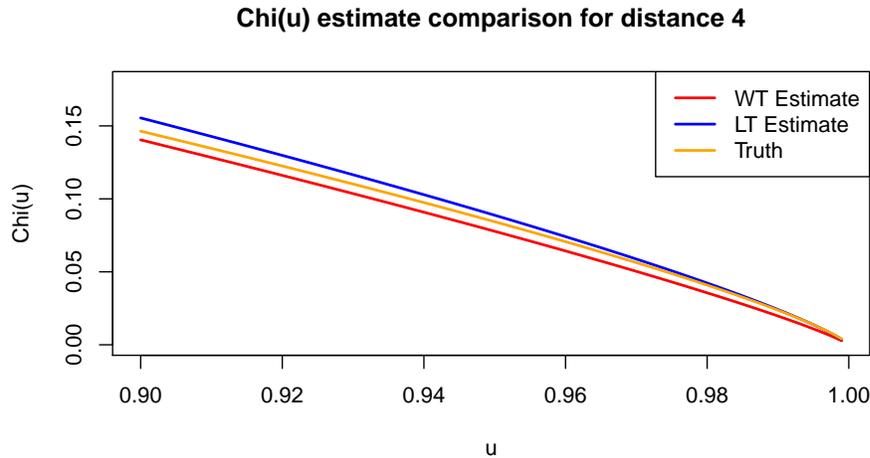


Figure 3.4.2: Comparison of theoretical values of  $\chi(u; h)$  using estimated parameter values of  $\alpha$  and  $\lambda$  with  $h = 4$  and  $0.9 \leq u \leq 1$ . The red line corresponds to the theoretical result under the estimate obtained from the Wadsworth and Tawn (2012b) censoring scheme, the blue line corresponds to the theoretical values of  $\chi(u; h)$  under the estimate obtained from the Ledford and Tawn (1996) censoring scheme and the orange line represents the theoretical value of  $\chi(u; h)$  for the simulated Gaussian process, using  $n = 1000$  replicates.

### 3.4.1 Results for $\eta(h)$

To compare the performance of the censoring schemes in relation to computing  $\eta(h)$ ,  $n = 100$  Gaussian processes, each with 1000 replications, are simulated with range parameters  $\phi = 2$  and  $\phi = 5$  using the exponential covariance function. Then the bias,

$$B = \mathbb{E}[\hat{\eta}(h)] - \eta_0(h),$$

and variance, given by

$$V = \mathbb{E}[(\hat{\eta}(h) - \mathbb{E}[\hat{\eta}(h)])^2] = \text{Var}(\hat{\eta}(h)),$$

are calculated. These can then be used to calculate the root mean square error, defined by

$$\text{RMSE}[\hat{\eta}(h)] = \sqrt{B^2 + V} = \sqrt{\frac{\sum_{i=1}^n (\hat{\eta}_i(h) - \eta_0(h))^2}{n}}.$$

In the above,  $\hat{\eta}_k(h)$  denotes the estimate of  $\eta(h)$  at a given distance  $h$  from simulation  $k \in \{1, \dots, n\}$  under one of the censoring schemes, whilst  $\eta_0(h)$  represents the true value (under the Gaussian process) of  $\eta(h)$ . Figures 3.4.3 - 3.4.5 show the bias, variance and RMSE of the estimates under the two censoring schemes for the case where  $\phi = 2$ . In each figure, results from applying the Ledford and Tawn (1996) censoring method are shown in blue, and results from the Wadsworth and Tawn (2012b) method displayed in red.

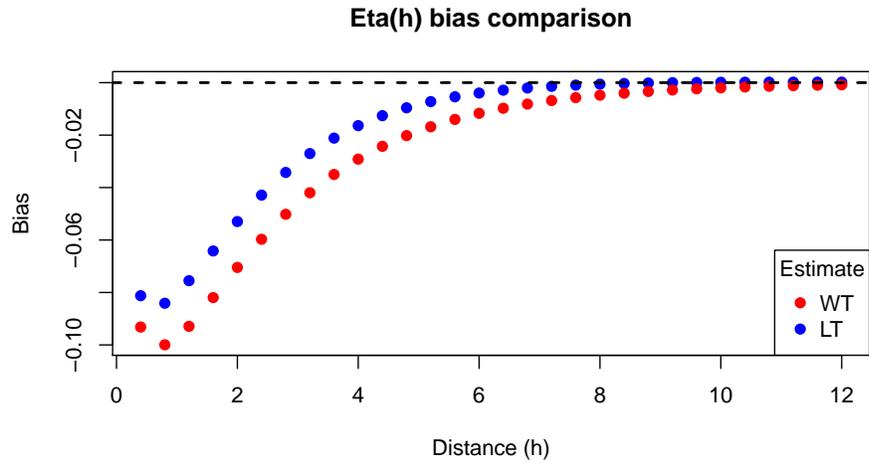


Figure 3.4.3: Plot of bias of  $\hat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. The dashed black line indicates zero bias.

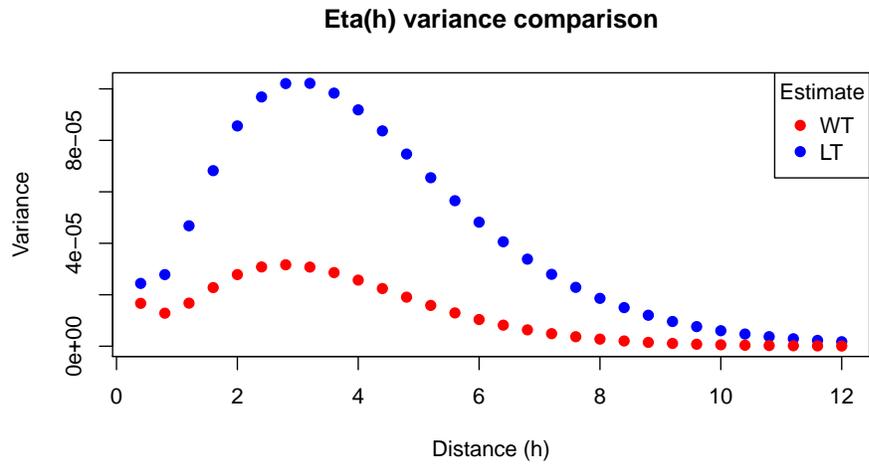


Figure 3.4.4: Plot of variance of  $\hat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

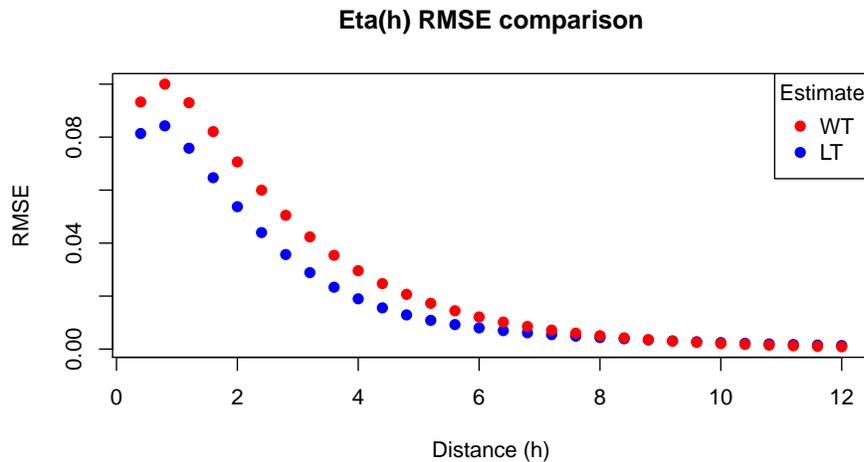


Figure 3.4.5: Plot of RMSE of  $\hat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

Figure 3.4.3 shows that while both censoring methods have led to estimates of  $\eta(h)$  that have negative bias, it is the Ledford and Tawn approach which has less bias, with both tending towards zero bias as the distance  $h$  increases. This is expected, since both the misspecified model estimates and the true model have  $\eta(h) \rightarrow 1/2$  as  $h \rightarrow \infty$ . However, it is seen in Figure 3.4.4 that the Wadsworth and Tawn censoring scheme has lower variance for each distance, though the variance of each is very small at all distances. Thus, the best comparison is perhaps made through comparing the RMSEs of estimation of  $\eta(h)$  under the two schemes, though we note that the bias is the dominating factor in its value. Using Figure 3.4.5, we see that the Ledford and Tawn approach has a smaller RMSE up to a distance of  $h \approx 10$ , beyond which the estimates arising from implementing the Wadsworth and Tawn censoring scheme become preferred. However,  $\eta_0(h = 10) \approx 0.504$ , so that the process is approximately

independent at these distances and so differences in estimation of  $\hat{\eta}(h)$  are likely to be negligible. In general, the difference between the RMSEs of  $\hat{\eta}(h)$  arising from the two censoring schemes appears to be relatively small for each  $h$ , so using the Wadsworth and Tawn approach would not necessarily be wholly inappropriate in this instance.

We also considered the case where  $\phi = 5$  in the exponential covariance function of the simulated Gaussian process. We see from the triangular points in Figures 3.4.6 and 3.4.7 that the bias and variance follow the same pattern of behaviour as with  $\phi = 2$ , leading to the behaviour of the RMSE shown by the triangular points in Figure 3.4.8 being similar to that shown in Figure 3.4.5. As before, estimates from the Ledford and Tawn (1996) censoring method are shown in blue, and Wadsworth and Tawn (2012b) estimates are coloured red. The main distinction appears to be that the difference in RMSE is perhaps smaller for  $\phi = 5$  than for  $\phi = 2$  at small distances, but is then larger for  $h$  greater than approximately 2, for which  $\eta_0(h = 2) \approx 0.835$ , so that this difference could be important in assessing the nature of dependence. We note that the values of RMSE approach zero for estimates from both censoring schemes more slowly than for  $\phi = 2$ . Hence, here the Ledford and Tawn (1996) scheme would be preferred in this case. Similar simulation runs were also carried out for  $\phi = 1$ , shown by squares in Figures 3.4.6, 3.4.7 and 3.4.8, suggesting that the RMSE of estimates decays to zero more quickly than either the cases where  $\phi = 2$  or  $\phi = 5$ ; this is likely tied to the fact that  $\eta_0(h) \rightarrow 1/2$  more quickly when  $\phi = 1$  than with the other parameter values.

Figure 3.4.9 displays the estimates of  $\eta(h)$  and  $\chi(u; h)$  when using a smaller sample size, in this case using  $n = 200$  replicates in each simulated Gaussian process. It was

found that the bias of the estimates continues to dominate the variance; the variance does increase as sample size decreases but only by a small amount. The principal effect of reducing the sample size appears to be that the RMSE decreases to zero with increasing  $h$  more slowly. This effect can be seen by comparing Figures 3.4.5 and 3.4.9, noting that the maximum values of the RMSEs in the estimates are similar as well, but appear to be slightly larger. Moreover, the maximum differences between the two RMSEs of the estimates is seen to increase by a small amount, and occurs for larger  $h$  than in the  $n = 1000$  case. Sample sizes of  $n = 100$  and  $n = 500$  were also used; this led to very similar behaviour of the RMSE as described when comparing  $n = 200$ , with bias dominating the variance in each case. The effects upon setting  $n = 100$  were slightly more pronounced than those described for the  $n = 200$  case, with the effects less pronounced when  $n = 500$ . An even smaller sample size may lead to a case where variance would not be dominated by the bias, but this has not been tested.

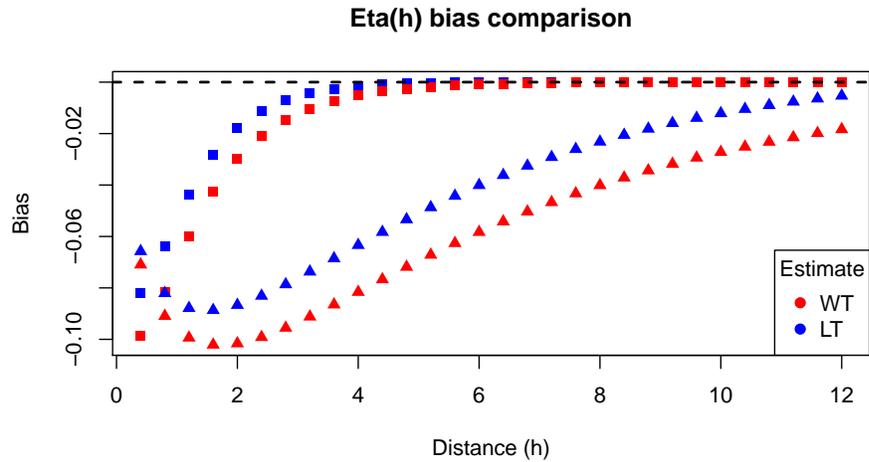


Figure 3.4.6: Plot of bias of  $\hat{\eta}(h)$  with exponential covariance function parameters  $\phi = 1$  (square points) and  $\phi = 5$  (triangular points) in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates. The black dashed line indicates zero bias.

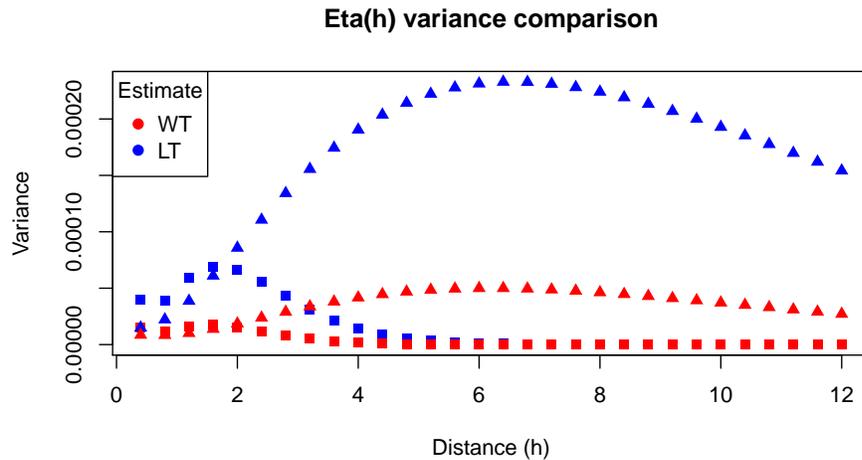


Figure 3.4.7: Plot of variance of  $\hat{\eta}(h)$  with exponential covariance function parameters  $\phi = 1$  (square points) and  $\phi = 5$  (triangular points) in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

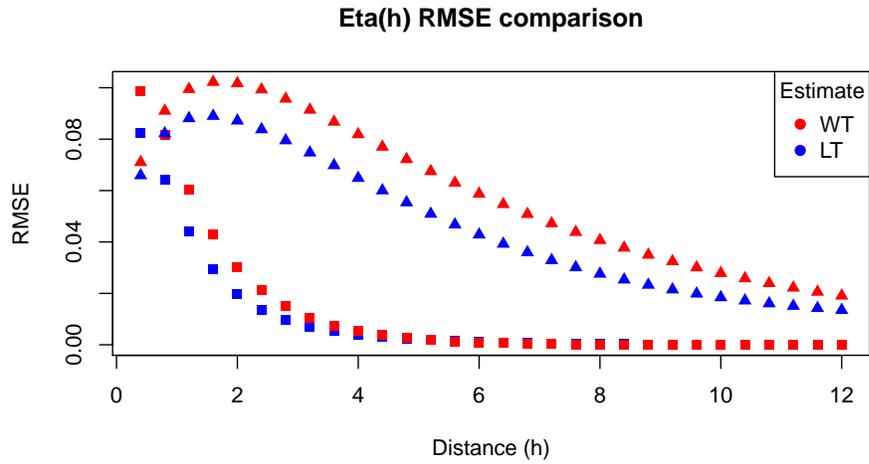


Figure 3.4.8: Plot of RMSE of  $\hat{\eta}(h)$  with exponential covariance function parameters  $\phi = 1$  (square points) and  $\phi = 5$  (triangular points) in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

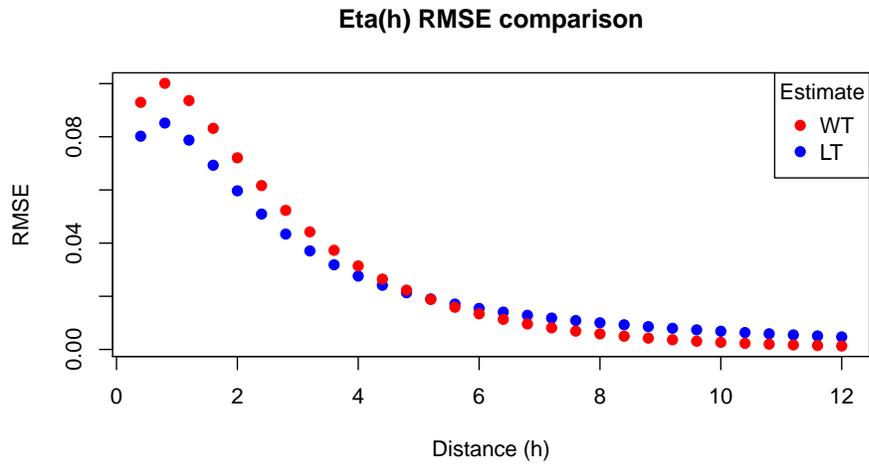


Figure 3.4.9: Plot of RMSE of  $\hat{\eta}(h)$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 200$  replicates.

### 3.4.2 Results for $\chi(u; h)$

We now consider the comparison of  $\chi(u; h)$ , under the two censoring schemes, for thresholds  $u = 0.95, 0.99, 0.995$ , taking  $\phi = 2$  in this comparison. We display the resulting bias, variance and RMSE, which are computed similarly to those for  $\eta(h)$  in 3.4.1, for the case  $u = 0.95$  with  $h = 0.4, 0.8, \dots, 12$  (taking  $u$  as fixed each time) in Figures 3.4.10, 3.4.11 and 3.4.12. As in Section 3.4.2, the results from applying the Ledford and Tawn (1996) censoring method are shown in blue, with results from the Wadsworth and Tawn (2012b) method displayed in red. Since the results for bias and variance for thresholds  $u = 0.99, 0.995$  were broadly similar for those for  $u = 0.95$ , we simply show the resulting RMSEs of estimation of  $\chi(u; h)$  in Figures 3.4.13 and 3.4.14, denoting the true value in each case by  $\chi_0(u; h)$ .

We see in Figure 3.4.10 that the Ledford and Tawn (1996) censoring scheme generally has a lower magnitude of bias in estimation of  $\chi(0.95; h)$  for most values of  $h$ ; it is only at the largest values of  $h$  considered here that the two values of bias become approximately equal and are close to zero. Again, this is expected since  $\chi(u; h) \rightarrow 0$  as  $h \rightarrow \infty$  for both the true model and the misspecified models. Moreover, it is seen that the Wadsworth and Tawn (2012b) censoring scheme has negative bias for its estimates of  $\chi(0.95; h)$  at all points, whilst the Ledford and Tawn censoring scheme exhibits positive bias. However, the variance of  $\widehat{\chi}(0.95; h)$  seen in Figure 3.4.11, whilst very small for both censoring schemes, is lower for all  $h$  under the Wadsworth and Tawn (2012b) method. Consequently, the RMSEs shown in Figure 3.4.12 show that, apart from  $h \leq 0.4$ , for which  $\chi_0(0.95; h) > 0.515$ , the RMSE of  $\widehat{\chi}(0.95; h)$  is lower

for the Ledford and Tawn method when  $h$  is lower than a value of approximately  $h = 4$ , above which the RMSE of the Wadsworth and Tawn approach is lower. For  $h > 4$ ,  $\chi_0(0.95; h) < 0.09$ , suggesting that the process is approaching independence, for which  $\chi(u; h) = 1 - u$ , as well as having  $\widehat{\chi}(0.95; h)$  take values in a similar range for these distances, so that these results are perhaps to be expected with few conclusions to be drawn from this behaviour.

Looking at the results of the RMSEs for  $u = 0.99$  and  $u = 0.995$  from Figures 3.4.13 and 3.4.14, we see that there is a similar pattern for the behaviour of the RMSE of  $\widehat{\chi}(u; h)$  for the two censored likelihoods. Again, the estimates arising from the Ledford and Tawn approach are lower up to a certain value of  $h$ , these being  $h \approx 5$  when  $u = 0.99$  (for which  $\chi_0(0.99; h) \approx 0.018$ ) and  $h \approx 6$  when  $u = 0.995$  (here,  $\chi_0(0.995; h) \approx 0.0075$ ), before the estimates from the Wadsworth and Tawn (2012b) method are lower. However, when this arises, both values of the RMSE are very close to zero, and so the differences are negligible, again arising from the fact that  $\chi_0(u; h)$  takes very small values for these values of  $h$  and  $u$ .

Results from simulating from a Gaussian process with  $\phi = 5$  in the exponential covariance function suggest that the behaviour of the RMSE with distance  $h$  is largely similar, with the main difference being that the distance  $h$  at which the Wadsworth and Tawn censored likelihood begins to perform better with respect to RMSE is larger in this case. Again, this is due to the fact that both the estimates and true values of  $\chi(u; h)$  will be close to zero for the value of  $h$  where this occurs. Figures for this case have been omitted.

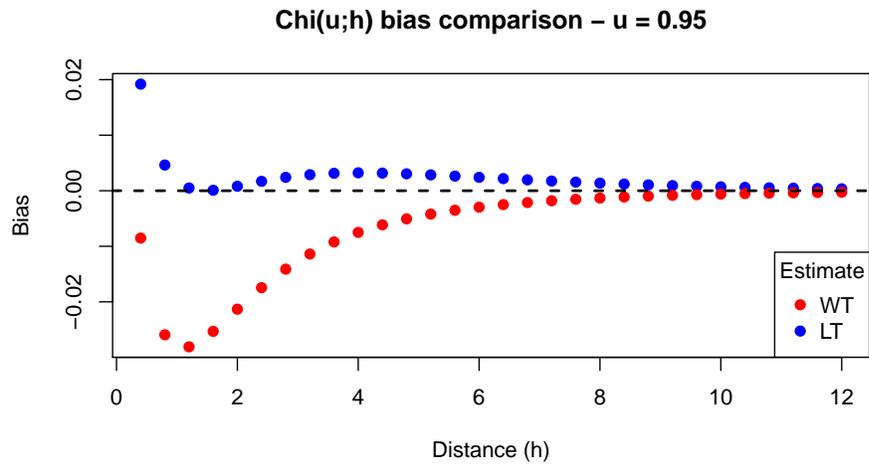


Figure 3.4.10: Plot of bias of  $\hat{\chi}(u; h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

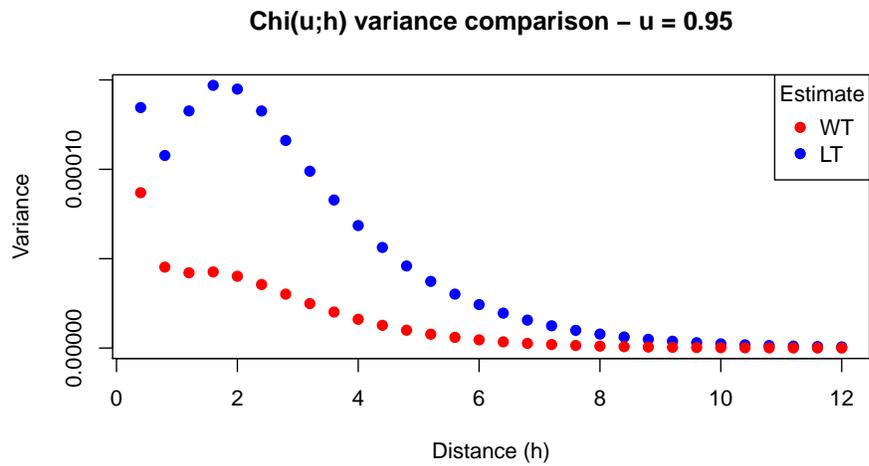


Figure 3.4.11: Plot of variance of  $\hat{\chi}(u; h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

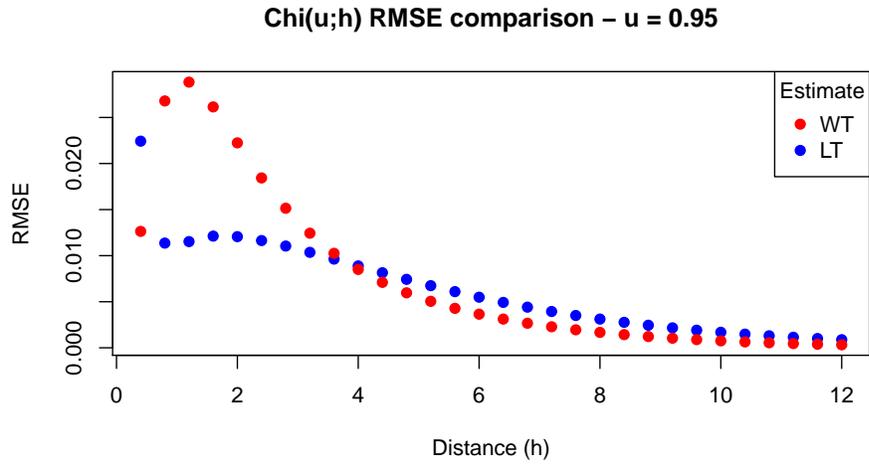


Figure 3.4.12: Plot of RMSE of  $\hat{\chi}(u;h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

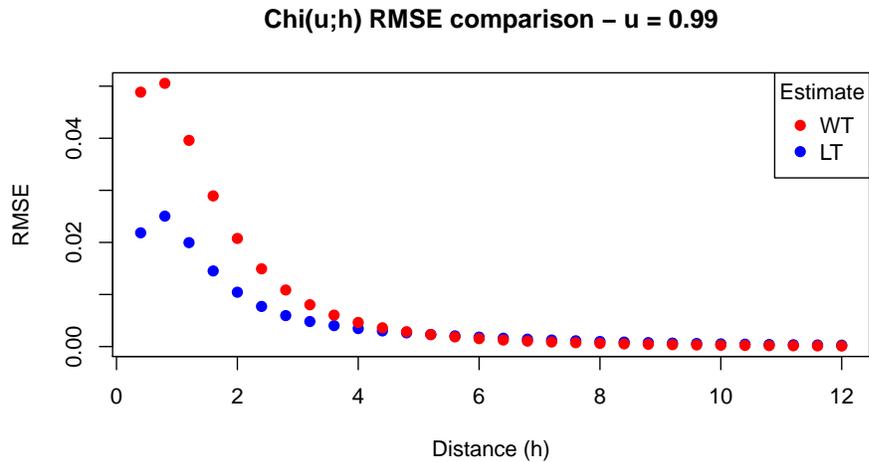


Figure 3.4.13: Plot of RMSE of  $\hat{\chi}(u;h)$  for  $u = 0.99$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

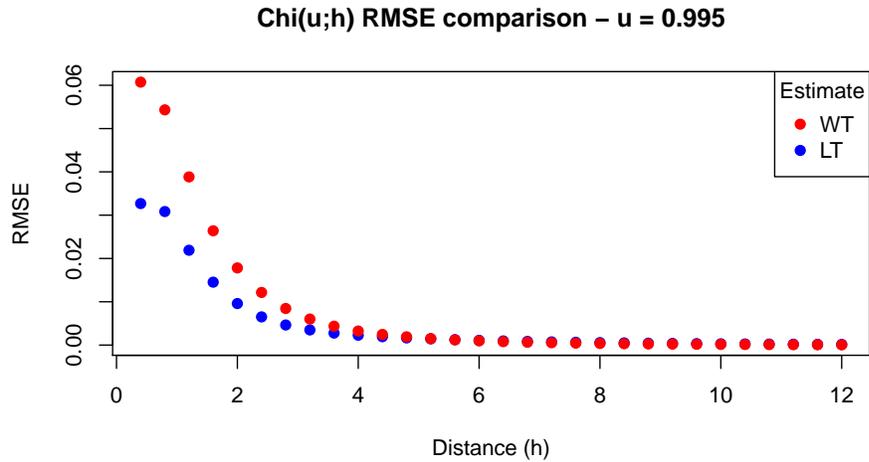


Figure 3.4.14: Plot of RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.995$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 1000$  replicates.

Figure 3.4.15 shows the behaviour of  $\chi(0.95; h)$  when estimating from 100 simulations of Gaussian processes with  $n = 200$  replications instead. In this case, the variance becomes a more important component in calculating the RMSE of the estimates. Indeed, as well as the RMSE being higher for both sets of estimates for all  $h$ , it appears that the Wadsworth and Tawn (2012b) estimates have a lower RMSE for most values of  $h$ . This effect is also seen for  $u = 0.99$  and  $u = 0.995$ . When setting  $n = 100$ , this effect becomes greater, with the Wadsworth and Tawn estimates performing better for a larger range of  $h$ . The converse occurs for  $n = 500$ .

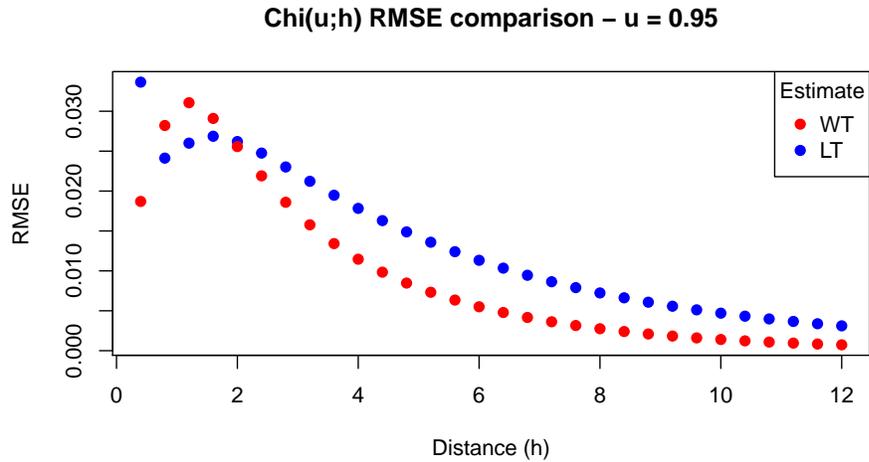


Figure 3.4.15: Plot of RMSE of  $\hat{\chi}(u;h)$  for  $u = 0.95$  with exponential covariance function parameter  $\phi = 2$  in the simulated Gaussian processes, from simulations with  $n = 200$  replicates.

### 3.5 Gaussian process censored likelihoods fitted to inverted max-stable process data

We now consider the effects of carrying out a similar analysis to that considered in Section 3.4, but with the misspecification considered in the opposite manner. That is, we simulate data from an inverted Brown-Resnick MSP with semi-variogram defined by (3.3.1) and then use these data in the censored pairwise likelihoods corresponding to a Gaussian process, which has exponential covariance function as in Section 3.4; as such we will make performance comparisons using the estimates of  $\phi$  obtained. Useful expressions for these censored likelihoods are as follows. The pairwise distribution functions and pairwise densities for a Gaussian process are straightforward; these are

just the bivariate normal distribution function and density with appropriate covariance matrix arising from the chosen covariance function. These functions may then be evaluated as appropriate in the censoring methods. The partial derivative terms require more work; we use the following result, assuming that  $(X, Y)$  are random variables from a centred Gaussian process with unit variances and correlation  $\rho$  and joint distribution function  $F_{XY}$ :

$$\begin{aligned} \frac{\partial}{\partial y} F_{XY}(u, y)|_{y=z} &= \frac{\partial}{\partial y} \int_{-\infty}^u \int_{-\infty}^y f_Y(v) f_{X|Y}(x|v) dv dx |_{y=z} \\ &= f_Y(z) \int_{-\infty}^u f_{X|Y}(x|z) dx \\ &= \phi(z) \Phi\left(\frac{u - z\rho}{\sqrt{1 - \rho^2}}\right) \end{aligned}$$

where  $\phi$  is the standard normal density function, and  $\Phi$  is the standard normal distribution function. In the above calculation,  $f_{X|Y}$  and  $f_Y$  represent the conditional density  $X|Y$  and marginal density of  $Y$  respectively, the latter of which is a standard normal density function. We can use this result on some spatial domain  $\mathcal{S}$  by considering  $X$  as the Gaussian process marginally at a site  $\mathbf{s}_i \in \mathcal{S}$ , and similarly  $Y$  as the Gaussian process at  $\mathbf{s}_j$ ; this is then used in the relevant censored pairwise likelihood contributions.

Again, we simulate data at 31 uniformly-spaced points over the line segment  $\mathcal{S} = [0, 12]$ . This was performed by simulating 100 Brown-Resnick max-stable processes, each with  $n = 1000$  replications, using the simulation method of Dieker and Mikosch (2015), and then computing the corresponding inverted max-stable process. We will display results, setting  $v$  to be the 0.95 quantile in the censored likelihoods in 3.1, from processes simulated using combinations of Brown-Resnick max-stable pro-

cess parameters ( $\alpha = 1, \lambda = 1$ ), but we note that results calculated from parameter combinations ( $\alpha = 0.5, \lambda = 0.5$ ), ( $\alpha = 0.5, \lambda = 1$ ) and ( $\alpha = 1, \lambda = 2$ ) suggest that increasing (respectively, decreasing)  $\lambda$  leads to the value of  $h$  at which the Wadsworth and Tawn approach is to be preferred becomes larger (respectively, smaller). Changing the value of  $\alpha$  appeared to have no significant effects on the behaviour of the estimates.

The same results of changing the sample size were found as in Section 3.4; for estimates of  $\eta(h)$  there was little change in the RMSE of the estimates due to bias continuing to dominate the variance, whilst the Wadsworth and Tawn (2012b) approach tends to perform better than the Ledford and Tawn (1996) method when estimating  $\chi(u; h)$  for many values of  $h$ , due to the variance of the estimates becoming a more important factor. Estimates of dependence measures from applying the Ledford and Tawn censoring method are again shown in blue, and results from the Wadsworth and Tawn method displayed in red, throughout this section.

### 3.5.1 Results for $\eta(h)$

As in Section 3.4.1, we first display results comparing the bias, variance and root mean square of estimates of the dependence measure  $\eta(h)$  from our fitted models, compared to the true value at each  $h$ ; plots of these may be found in Figures 3.5.1, 3.5.2 and 3.5.3.

Figure 3.5.1 shows that the magnitude of the bias of  $\hat{\eta}(h)$  when estimated using the Ledford and Tawn (1996) censored likelihood is lower than that for the Wadsworth and Tawn (2012b) method for  $h < 4$ . However, for  $h \approx 5$  and greater, the magnitude

of the bias is lower for the Wadsworth and Tawn method. However,  $\eta(h) < 0.6$  for  $h > 6$ , indicating that independence is being approached for such distances suggesting a possible reason for this behaviour. It can be seen that the variance of  $\hat{\eta}(h)$  is larger when the Ledford and Tawn method is applied for all distances  $h$ , albeit the values at each  $h$  are very small so that there is little difference between these estimates. Finally, we see from Figure 3.5.3 that for all  $h$  less than approximately 4.4, the RMSE of  $\eta(h)$  is lower when estimated using the Ledford and Tawn censored likelihood, but the Wadsworth and Tawn (2012b) estimates have a lower value of RMSE for all other distances.

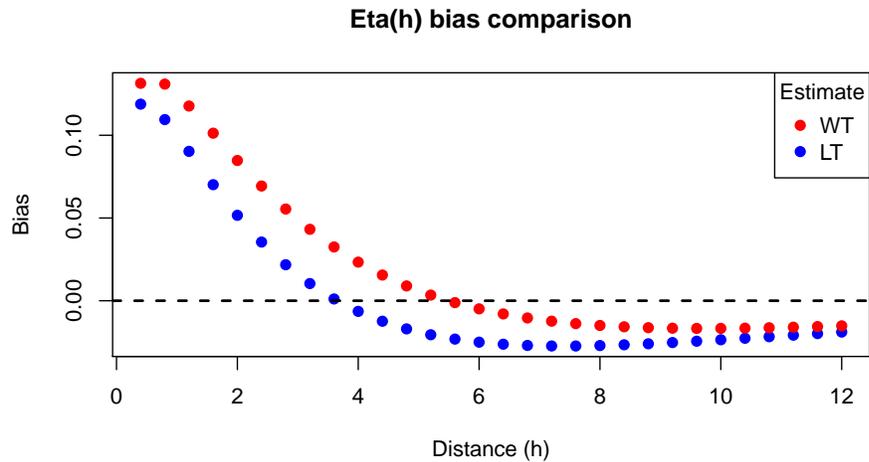


Figure 3.5.1: Plot of the bias of  $\hat{\eta}(h)$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable process, from simulations with  $n = 1000$  replicates. The black dashed line indicates zero bias.

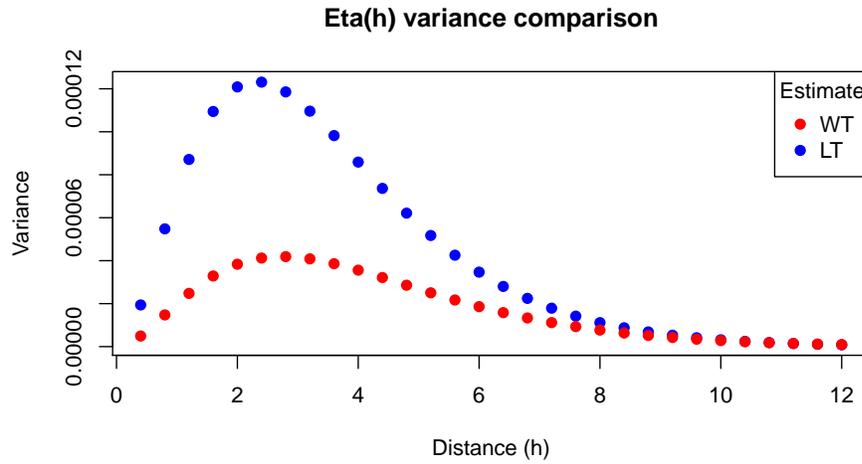


Figure 3.5.2: Plot of the variance of  $\hat{\eta}(h)$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable process, from simulations with  $n = 1000$  replicates.

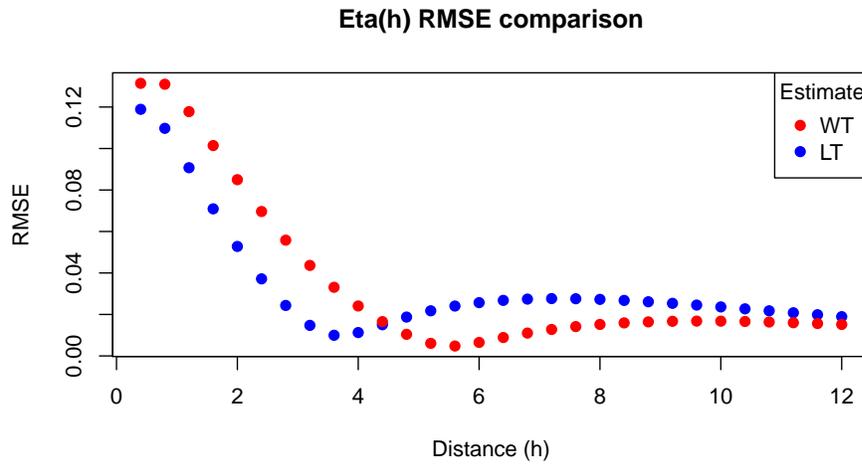


Figure 3.5.3: Plot of the RMSE of  $\hat{\eta}(h)$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates.

### 3.5.2 Results for $\chi(u; h)$

We now look at the results obtained for  $\hat{\chi}(u; h)$  when using the Gaussian process censored likelihoods on inverted Brown-Resnick MSP data; Figures 3.5.4, 3.5.5 and 3.5.6 show the bias, variance and RMSE, respectively, of  $\hat{\chi}(u; h)$  obtained from estimates of  $\phi$  under the two censoring approaches when  $u = 0.95$ . The magnitude of the bias is lower for the estimates which are calculated from the results of the Wadsworth and Tawn censoring scheme when, approximately,  $h > 2.4$ . The variance of the estimates of  $\chi(u; h)$  is seen to be lower for all  $h$ , which combined with the results from the bias means that the RMSE of  $\hat{\chi}(u; h)$  is lower when using the Ledford and Tawn (1996) approach for  $h < 2$ , with the resulting RMSE lower for the Wadsworth and Tawn (2012b) approach for all  $h > 2.4$ , with the RMSE tending towards zero as  $h \rightarrow \infty$  in both instances, and the difference between them becoming smaller for increasing  $h$ . This is expected, however, as  $\chi(u; h) \approx 0$  for large  $h$  so definitive conclusions may be hard to draw from this.

Figures 3.5.7 and 3.5.8 show the RMSE of  $\hat{\chi}(u; h)$  when  $u = 0.99$  and  $u = 0.995$  respectively follows similar behaviour to that described above, with the only differences being that the value of  $h$  at which the RMSE of the Wadsworth and Tawn estimates of  $\chi(u; h)$  become lower is slightly different for these choices of  $u$ . When  $u = 0.99$ , this distance appears to be slightly lower than 3, and this is approximately the same in the case where  $u = 0.995$ .

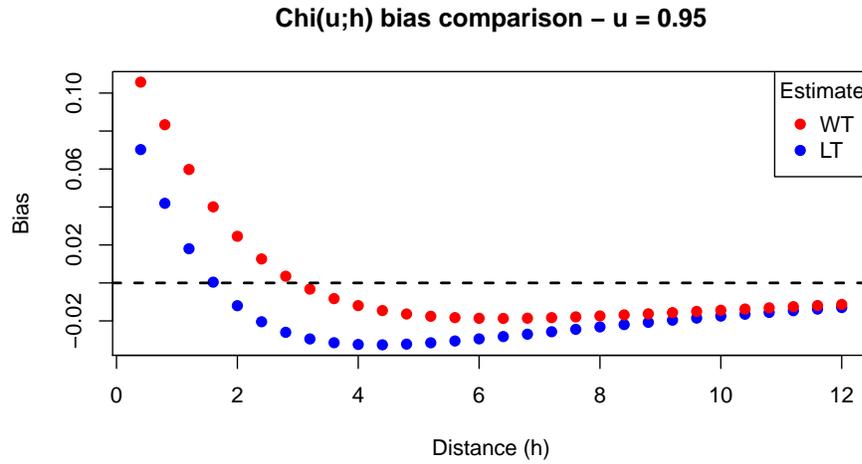


Figure 3.5.4: Plot of the bias of  $\hat{\chi}(u;h)$  for  $u = 0.95$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates. The black dashed line indicates zero bias.

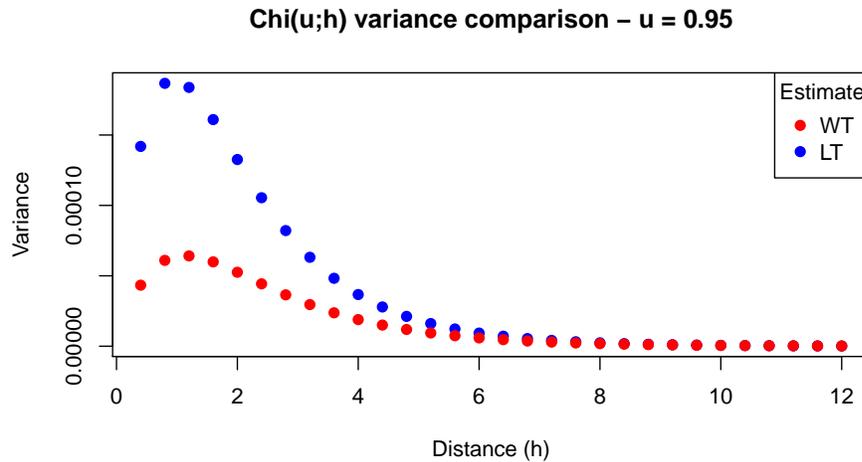


Figure 3.5.5: Plot of the variance of  $\hat{\chi}(u;h)$  for  $u = 0.95$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates.

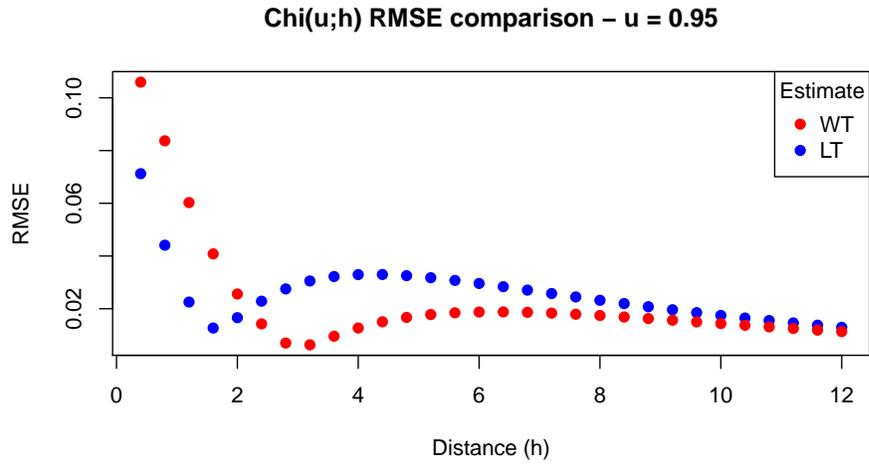


Figure 3.5.6: Plot of the RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.95$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates.

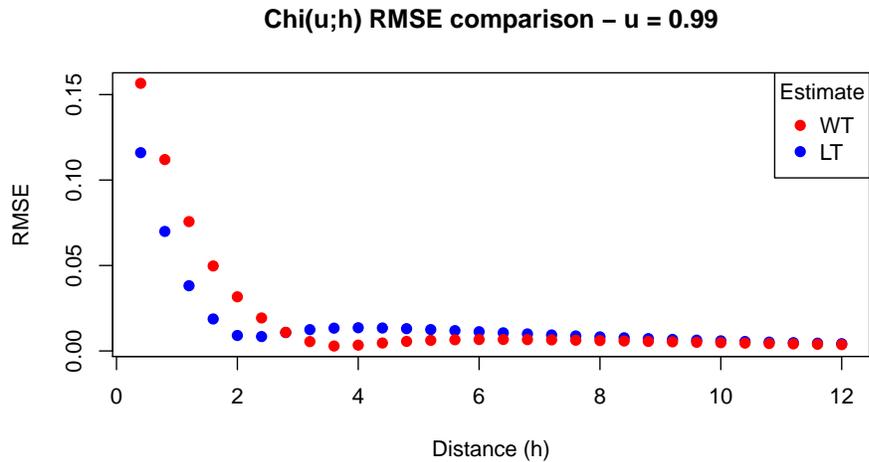


Figure 3.5.7: Plot of the RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.99$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates.

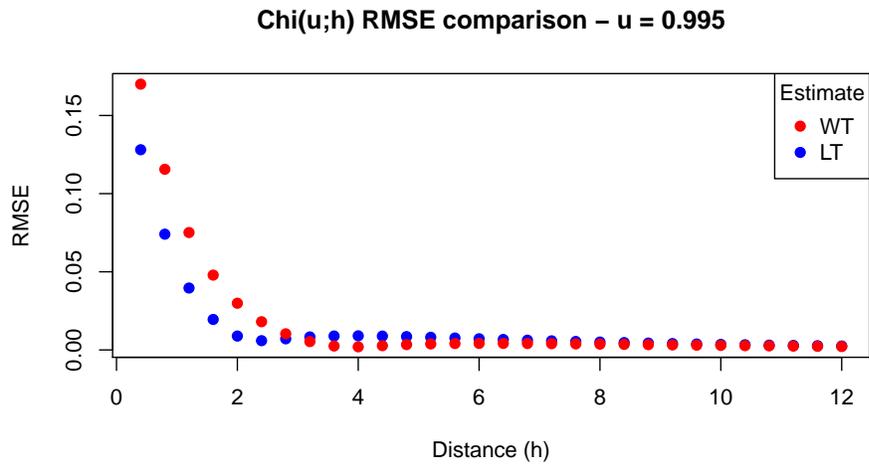


Figure 3.5.8: Plot of the RMSE of  $\hat{\chi}(u; h)$  for  $u = 0.995$  with  $\alpha = 1, \lambda = 1$  set in the simulation of the true inverted Brown-Resnick max-stable processes, from simulations with  $n = 1000$  replicates.

### 3.6 Summary of results

We now summarise the results presented in Sections 3.3 and 3.4, and provide details of the computational time of each of the experiments for the two censored likelihoods in finding the maximum likelihood estimates of all 100 simulated sets of data under these likelihoods; these details may be found in Table 3.6.1.

We see in Table 3.6.1 that the Wadsworth and Tawn (2012b) censored likelihood reduces the computational time, as would be expected, by approximately 60 seconds for both the Gaussian process likelihood and Brown-Resnick IMSP likelihood; this corresponds to an approximate average of 0.6 seconds of reduction in the computational time of likelihood maximisation for each data set. Since this difference is

relatively small, we will not give significant weight to this argument here, but note that this may become a more important factor if utilising more observations or more sampling locations as the difference in computational time becomes more pronounced.

Censoring method	Computational time (seconds)	
	GP likelihood	IMSP likelihood
Ledford and Tawn (1996)	660.39	1135.37
Wadsworth and Tawn (2012b)	591.99	1078.98

Table 3.6.1: Table of total computational time (in seconds) to obtain estimates of inverted Brown-Resnick max-stable process and Gaussian process (GP) parameters (as specified in Sections 3.4 and 3.5) for all of the simulated data sets described in Sections 3.4 and 3.5.

To compare results, we focus on the RMSEs obtained from carrying out the experiments described in Sections 3.4 and 3.5, but note that there are some subtleties in the behaviour over  $h$  of the bias and variance of the dependence measure estimates, pointed out above. These results suggest that the principal differences in performance of the censored likelihoods are in the estimation of short-range and long-range dependence. At short range, i.e. for small  $h$ , it appears that in general, the Ledford and Tawn (1996) approach provides estimates of the dependence measures which have a lower RMSE than those for the Wadsworth and Tawn (2012b) method, with the converse holding generally for large  $h$ . The value of  $h$  at which the preference changes depends on the true process. For more strongly dependent simulated processes, i.e.

larger values of  $\phi$  when simulating the Gaussian processes in Section 3.4 or larger  $\lambda$  in Section 3.5, the higher this value of  $h$  generally becomes. When considering  $\widehat{\chi}(u; h)$ , the effect of the value of  $u$  on the value of  $h$  appears to be that larger values of  $u$  lead to larger values of  $h$  at which the change in preference occurs; a particular reason for this is unclear.

We further note that the differences between the two censoring schemes at each distance  $h$  are also different depending on the strength of dependence of the simulated processes. When these simulated processes are more strongly dependent, it appears that the Wadsworth and Tawn censoring scheme performs worse in comparison to the Ledford and Tawn approach when estimating the measures of dependence than when the processes exhibit weaker dependence. This may be due to the fact that the Wadsworth and Tawn (2012b) censored likelihood is motivated for asymptotically independent variables, but so that this method may perform worse for variables which exhibit higher levels of dependence.

Thus, overall, whilst we note that there are situations in which the Wadsworth and Tawn censored likelihood approach performs better at determining measures of dependence than the Ledford and Tawn approach, this in general occurring when the distance between sites is large and the dependence measures are close to their limiting values, and the improvement in performance is somewhat negligible when this is the case. However, there is some evidence that upon using smaller sample sizes, the Wadsworth and Tawn approach may provide better estimates of the dependence measure  $\chi(u; h)$ . Since the Wadsworth and Tawn estimates of dependence measures generally tend to perform significantly worse for small  $h$ , it would therefore be sensible

to conclude for large sample sizes that the Ledford and Tawn censoring scheme would be more suitable to use if there is no prior knowledge of the true underlying process, as the computational advantage gained from the Wadsworth and Tawn (2012b) censored likelihood is small but the dependence behaviour modelled may be somewhat worse. On the other hand, it appears for small sample sizes that it is difficult to assess which method provides preferred estimates of the dependence measures, since both outperform the other in various scenarios. An area of further study would be to investigate the impacts, on both computational time and the estimation of dependence measures, of using more sampling locations and smaller sample sizes than  $n = 100$ , simulating from a wider range of processes and using different censoring thresholds  $v$ .

# Chapter 4

## Modelling spatial extreme events with environmental applications

### 4.1 Introduction

In many environmental applications data are collected from a number of spatial locations, for example numerous locations across an ocean basin or locations across a river network. Historically interest has been in the extremal behaviour at individual sites. However, our interest lies in developing a framework in which it is possible to estimate probabilities of joint events over space. For example, for wave heights we may want to know the probability of no offshore structure being damaged in a storm, and for river levels the probability that the total damages from a flood exceed £1 billion. Probabilities of the occurrence of extreme spatial events are of particular interest to the reinsurance industry for deriving aggregate financial loss distributions, and also to governments in terms of risk assessment and emergency planning.

To answer such questions we take an asymptotically justified model for the joint occurrence of extreme values of an event over space. Our reason for this is that we aim to extrapolate to spatial events that are larger than any previously observed, so we cannot rely on empirical evidence alone. Asymptotic theory therefore provides a principled approach to develop our models and understanding. Such a spatial model requires both marginal distributions and the dependence structure of the spatial process to be explicitly characterised. It is the challenge of modelling the extremal dependence structure that will be the primary focus of this paper. As closed form probabilities cannot be derived for the spatial events of interest to us, we aim to develop methods that enable straightforward simulation of extreme spatial events from which probabilities can be derived using Monte Carlo methods.

Let  $\{Y(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$  denote a stationary spatial process indexed by  $s$  over a set  $\mathcal{S}$  with marginal distribution function  $F$  which has upper endpoint  $y_F$ . In practice we observe replicates of  $\{Y(s) : s \in \mathcal{S}\}$  at a finite set of points  $\{Y(s_j) : j = 1, \dots, n\}$ , and at times  $t = 1, \dots, n$ . Hence  $Y_t(s)$  denotes the process observed at time  $t$  at location  $s$ . We are interested in the extreme values of  $Y$  over the entire set of  $\mathcal{S}$ . For this paper, we assume that the entire spatial process is independent and identically distributed in time, i.e.,  $\{Y_i(s); s \in \mathcal{S}\}$  is independent of  $\{Y_j(s); s \in \mathcal{S}\}$  for all  $i, j = 1, \dots, n$  with  $i \neq j$ . Thus our focus is on the spatial dependence behaviour of the process only. However, unlike in many applications of spatial statistics, we have a large number of independent and identically distributed replicates of the spatial process from which to make our inference.

In many spatial extreme value problems the aim is to characterise the extremal

behaviour of the spatial process  $Y(s)$ . A complication is that without a natural ordering scheme in more than the one dimension the definition of an extreme event is not well-defined. A range of approaches can be taken, as follows.

**Max-Stable Processes** Consider componentwise maxima over  $n$  independent and identically distributed copies of  $\{Y(s), s \in \mathcal{S}\}$ , i.e.,

$$\{M_n(s); s \in \mathcal{S}\} = \{\max_{1 \leq t \leq n} Y_t(s); s \in \mathcal{S}\}. \quad (4.1.1)$$

Here, and throughout this paper, operations are carried out componentwise, i.e., site specifically.

**Pareto Processes** Consider the process obtained by characterising the limiting behaviour of

$$\{Y(s); s \in \mathcal{S} \mid \max_{s \in \mathcal{S}} Y(s) > u\} \quad (4.1.2)$$

as  $u \rightarrow y_F$ .

**Conditional Extremes Processes** We propose to characterise the behaviour of

$$\{Y(s); s \in \mathcal{S} \mid Y(s_0) > u\} \quad (4.1.3)$$

for any  $s_0 \in \mathcal{S}$  as  $u \rightarrow y_F$ .

When suitably linearly normalised,  $\{M_n(s); s \in \mathcal{S}\}$  converges (as  $n \rightarrow \infty$ ) to a max-stable process; see Smith (1990), Schlather (2002), Padoan et al. (2010), Davison et al. (2012). This is the most widely used approach to spatial extremes due to its historical link to the families of univariate and multivariate extreme value distributions (all finite dimensional distributions of a max-stable process are multivariate extreme

distributions) and also for its elegant mathematical properties. However, this approach cannot be used to answer questions about original events for  $Y(s)$  since  $M_n(s)$  is a composition of a number of different events, and hence this formulation cannot be used to answer our motivating questions. Furthermore, the spatial dependence structure for  $M_n(s)$  is restrictive and so fails to accommodate a wide class of events including Gaussian processes; see the discussion of  $\chi(\tau)$  below.

Using the underlying mathematical formulation of max-stable processes, Ferreira and de Haan (2014) obtain a limiting form of the process (4.1.2), which we outline in Section 4.2.3. Note that Dombry and Ribatet (2015) alternatively condition on other functionals of the process being extreme, and obtain a class of limiting processes known as  $\ell$ -Pareto processes.

Our proposal differs in two ways from that used for Pareto or  $\ell$ -Pareto processes. We condition on the extreme event in conditional representation (4.1.3) being large at a specific site. We also exploit the normalisation structure from Heffernan and Tawn (2004) in the conditional approach (4.1.3) that uses a different normalisation of  $Y(s)$  to achieve a more general (and more flexible) limiting representation. We will take the conditional extremes process approach (4.1.3) which we outline in Section 4.3.2. However, we also give further details of max-stable and Pareto processes to help explain their weaknesses for our needs and to show how our approach differs from them.

To help to first identify the differences between the approaches, let us introduce two pairwise spatial extremal dependence measures,  $\{\chi(\tau), \bar{\chi}(\tau)\}$ , which are natural extensions of multivariate measures defined by Coles et al. (1999) to stationary spatial

processes. Consider a pair of sites  $(s, s + \tau)$ , each in  $\mathcal{S}$ . Then  $\chi(\tau)$  is defined by the following limit probability

$$\chi(\tau) = \lim_{y \rightarrow y_F} \mathbb{P}(Y(s + \tau) > y \mid Y(s) > y), \quad (4.1.4)$$

if it exists. Additionally,  $\bar{\chi}(\tau)$  is determined by the following asymptotic equivalence, as  $y \rightarrow y_F$

$$\mathbb{P}(Y(s + \tau) > y \mid Y(s) > y) \sim \mathcal{L}(1/\bar{F}(y)) \{\bar{F}(y)\}^{[1-\bar{\chi}(\tau)]/[1+\bar{\chi}(\tau)]},$$

where  $\mathcal{L}$  is a slowly varying function at infinity and  $\bar{F}(y) = 1 - F(y)$ . Here  $0 \leq \chi(\tau) \leq 1$  and  $-1 < \bar{\chi}(\tau) \leq 1$ . For each of  $\chi(\tau)$  and  $\bar{\chi}(\tau)$ , larger values correspond to stronger levels of extremal dependence.

If  $\chi(\tau) > 0$ , then  $\bar{\chi}(\tau) = 1$  and the largest values of the process can occur simultaneously at two sites  $\tau$  apart, a property known as asymptotic dependence at lag  $\tau$ . However, if  $\chi(\tau) = 0$  then in the limit the largest values at sites  $\tau$  distance apart must occur in different spatial events, and the process is said to have asymptotic independence at  $\tau$ . For processes with  $\chi(\tau) = 0$ , the quantity  $\bar{\chi}(\tau)$  is a helpful measure for determining the level of asymptotic independence since it controls the rate at which  $\mathbb{P}(Y(s + \tau) > y \mid Y(s) > y)$  converges to zero. In particular,  $0 < \bar{\chi}(\tau) \leq 1$  corresponds to positive extremal dependence,  $\bar{\chi}(\tau) = 0$  to near extremal independence, and  $-1 < \bar{\chi}(\tau) < 0$  to negative extremal dependence.

Determining the pair  $\{\chi(\tau), \bar{\chi}(\tau)\}$ , for all  $\tau$ , provides a good summary of the extremal properties of the process. Some spatial extreme value modelling approaches preclude certain types of extremal dependence. For example, for all non-degenerate max-stable processes or Pareto processes that are dependent at lag  $\tau$  then  $\{\chi(\tau), \bar{\chi}(\tau)\} =$

$(c_\tau, 1)$ , for some  $0 < c_\tau < 1$ . However, for all non-degenerate Gaussian processes  $\{\chi(\tau), \bar{\chi}(\tau)\} = (0, \rho(\tau))$ , where  $\rho(\tau)$  is the correlation of the Gaussian process at lag  $\tau$ . Thus max-stable and Pareto processes are asymptotically dependent, whereas Gaussian processes are asymptotically independent. These measures show that max-stable and Pareto processes fail to capture the spatial extremal dependence features of Gaussian processes. Consequently, if the data were from a Gaussian process but a max-stable process model was fitted then there will be an over-estimation of the risk of jointly large events. Therefore a broader class of spatial extreme value models is required if we are to capture the dependence structures of both these important classes of spatial process. The models we will introduce here have this capability, as well as having sufficient structure in order to model our applications well.

The conditional multivariate extreme value model of Heffernan and Tawn (2004) estimates the form of extremal dependence structure (asymptotic dependence or asymptotic independence) as part of the fitting procedure. The model can handle high dimensional problems (Winter et al., 2016), extremal temporal dependence (Winter and Tawn, 2017), missing values (Keef et al., 2009) and negative dependence (Keef et al., 2013a). Examples of the environmental applications include heatwaves, hydrology and oceanography (Jonathan et al., 2013; Keef et al., 2009; Towe et al., 2017; Winter and Tawn, 2016). Here we outline how these multivariate methods can be extended to a spatial framework and clarify what they offer over existing spatial extreme value models.

Section 4.2 details existing statistical models for spatial extreme values. Section 4.3 presents the conditional multivariate extreme value model of Heffernan and Tawn

(2004) and outlines how this model can be extended to handle spatial extreme problems. Finally, Section 4.4 details two applications of the methodology to oceanography and hydrology; the first of these relates to understanding the extremal dependence of significant wave heights over the North Sea and the second addresses questions on widespread risk of flooding raised by the UK Government's 2016 National Flood Resilience Review.

## 4.2 Existing methods

### 4.2.1 Univariate modelling

Underpinning the two main distributions of univariate extreme value theory are representational characterisations of max-stability and threshold-stability which uniquely define these distributions. Here we recap these features in the univariate case, as they provide the core structure for the existing spatial extremal theory.

Much classical extreme value theory is based on the property of max-stability that leads to the extremal types theorem of Fisher and Tippett (1928). For independent and identically distributed univariate random variables  $\{Y_i; i = 1, \dots, n\}$ , with continuous but otherwise arbitrary distribution function  $F$  with upper endpoint  $y_F$ , let  $M_n = \max\{Y_1, \dots, Y_n\}$ . If there are normalising sequences  $a_n > 0$  and  $b_n$  such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x) \quad (n \rightarrow \infty), \quad (4.2.1)$$

where  $G$  is a non-degenerate distribution function, then  $G$  is of the form

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

with parameters  $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$  corresponding to location, scale and shape parameters and  $\{z\}_+ = \max\{0, z\}$ . This is known as the generalised extreme value (GEV) distribution, and is denoted  $\text{GEV}(\mu, \sigma, \xi)$ . This class of distributions uniquely satisfies the max-stability property which says that for all  $m \in \mathbb{N}$  and  $x \in \mathbb{R}$ , there are constants  $A_m > 0, B_m$  such that

$$\{G(A_mx + B_m)\}^m = G(x).$$

Thus the GEV is the only non-degenerate distribution that is closed to the operation of maximisation.

An alternative approach to modelling univariate extremes is to focus on the exceedances of a threshold  $u$ . Pickands (1975) showed that if there is a non-degenerate limit (4.2.1), then there exists a normalising function  $c(u) > 0$  such that as  $u \rightarrow y_F$ ,

$$\frac{Y - u}{c(u)} \mid Y > u \xrightarrow{d} V,$$

where convergence is in distribution and  $V$  is non-degenerate. Then  $V$  follows a generalised Pareto distribution, which we denote  $\text{GPD}(\psi, \xi)$ , with distribution function

$$H(x) = 1 - \left(1 + \frac{\xi x}{\psi}\right)_+^{-\frac{1}{\xi}}, \quad (x > 0), \quad (4.2.2)$$

with scale parameter  $\psi > 0$  and shape parameter  $\xi \in \mathbb{R}$ .

The characterising property of the GPD is that of threshold stability (Davison and Smith, 1990), that is, for any  $v > 0$ , there exists a function  $c(v) > 0$  such that

$$\left\{ \frac{V - v}{c(v)} \right\} \mid V > v \stackrel{d}{=} V. \quad (4.2.3)$$

Thus scaled excesses of a higher threshold  $v$  by  $V$  have the same distribution as

$V$ . This is illustrated in Figure 4.2.1. The GPD is the only distribution with this threshold-stability property.

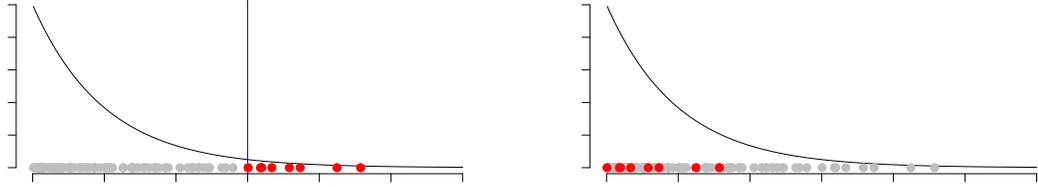


Figure 4.2.1: Illustration of threshold stability property described by relationship (4.2.3). The left panel shows a sample from  $V \sim \text{GPD}(\psi, \xi)$  with the vertical line representing the threshold  $v$  and the red points the exceedances of  $v$ ; the right panel shows these same exceedances (shown as excesses in red) after scaling (here the GPD has parameters  $(\psi, \xi) = (1, 0)$ , and so  $c_v = 1$ ). These scaled excesses are compared against a new sample (in grey) from the original distribution of  $V$ , we note that these two samples have the same distribution.

Based on this asymptotic justification, we make the modelling assumption that the distribution of excesses of  $Y(s)$  over a high threshold  $u$  follows the limiting distribution for excesses exactly, i.e.,

$$Y(s) - u \mid Y(s) > u \stackrel{d}{=} V(s) \mid V(s) > 0 \quad (s \in \mathcal{S}).$$

Consequently, the margins of  $Y(s)$  are  $\text{GPD}(\psi, \xi)$  distributed above the threshold  $u$ , where  $\psi$  and  $\xi$  do not depend on  $s \in \mathcal{S}$  as the  $Y(s)$  process is stationary. Since the above assumption provides no information on the marginal behaviour below  $u$ , the empirical distribution is used below this threshold (Coles and Tawn, 1991). The

resulting model for the marginal distribution function is

$$F(x) = \begin{cases} \tilde{F}(x) & \text{if } x \leq u \\ 1 - [1 - \tilde{F}(u)] \left[1 + \frac{\xi(x-u)}{\psi}\right]_+^{-\frac{1}{\xi}} & \text{if } x > u, \end{cases}$$

where  $\tilde{F}(x)$  is the empirical distribution function of all of the data at all sites. Due to stationarity of the process, data at all locations can be used to estimate  $F$ .

The study of dependence structure is typically undertaken via copulas (Nelsen, 2006), which requires the marginal distributions to be identical and uniformly distributed. Although we have identical margins, we prefer to transform them to non-uniform margins, via the pointwise transformation

$$X_t(s) = K^{-1}\{F(Y_t(s))\} \quad (s \in \mathcal{S}, t = 1, \dots, n),$$

so that  $X_t(s)$  is a spatial process, independent over time, and with marginal distribution function  $K$ . We perform this transformation as the extremal dependence properties of  $X_t(s)$  are more simply expressed for some non-uniform marginal choices.

The most convenient choice of  $K$  depends on the context: the Fréchet or Pareto distributions are typically assumed for max-stable distributions (Resnick, 1987); for conditional extremes, Heffernan and Tawn (2004) use Gumbel margins; for joint tail modelling, Wadsworth and Tawn (2014) used exponential margins while Keef et al. (2013a) showed that Laplace margins allow negative dependence to be incorporated the most parsimoniously. Critically, Gumbel, exponential and Laplace distributions all have exponential upper tails, so if negative dependence is avoided (which is reasonable in most spatial extremes applications) they are essentially identical approaches for our purposes. Here we take  $X_t(s)$  to have Gumbel marginals, so that

$K(x) = \exp\{-\exp(-x)\}$ , as this gives the clearest link to the max-stable results; since  $\exp\{X_t(s)\}$  has Fréchet margins. Thus, results in Fréchet margins translate to results in Gumbel margins via a log transformation.

We now have that  $\{X_t(s); s \in \mathcal{S}\}$  is a stationary spatial process with Gumbel margins. Although the copula/dependence structure of this process is restricted by the stationarity of the process, the range of choice of models is nonetheless vast. We saw, in the univariate case, that looking at the extremes of the variable reduced the class of possible continuous distributions to either the GEV or GPD depending on the extremal feature that is studied. For the dependence structure similar simplifications arise by imposing max-stability and threshold stability in spatial contexts. We explore these two strategies in Sections 4.2.2 and 4.2.3 respectively.

## 4.2.2 Max-stable processes

Given that  $\{X_t(s); s \in \mathcal{S}\}$  has Gumbel margins, it follows from (4.1.1) and (4.2.1) that we can take  $a_n = 1$  and  $b_n = \log n$  which gives  $Z(s)$ , defined by

$$Z(s) \stackrel{d}{=} \lim_{n \rightarrow \infty} \left\{ \max_{t=1, \dots, n} X_t(s) - \log n \right\} \quad (s \in \mathcal{S}),$$

to be a max-stable process with Gumbel margins. As a consequence of the  $Z(s)$  process being max-stable, for any  $d$  sites  $\{s_1, \dots, s_d\}$  in  $\mathcal{S}$  then  $\{Z(s_1), \dots, Z(s_d)\}$  with distribution function  $G$  is max-stable, i.e., for all  $m \in \mathbb{N}$  and  $\mathbf{x} \in \mathbb{R}^d$ ,

$$\{G(\mathbf{x} + \log m)\}^m = G(\mathbf{x}),$$

so the joint distribution is stable with respect to taking componentwise maxima. From the characterisation of de Haan (1984) and Schlather (2002), the max-stable process

$Z(\cdot)$  takes the form

$$Z(s) = \max_{i \geq 1} \{R_i + W_i(s)\} \quad (s \in \mathcal{S}), \quad (4.2.4)$$

where  $\{R_i, i \in \mathbb{N}\}$  are the points of a Poisson process on  $\mathbb{R}$  with intensity  $\exp(-x)dx$  and the  $W_i(s)$  over  $i$  are independent and identically distributed stochastic processes with continuous sample paths such that

$$\mathbb{E}[\exp\{W_i(s)\}] = 1 \quad (i \in \mathbb{N}, s \in \mathcal{S}).$$

Note that the additive structure is identical to the usual product structure, with the difference arising due to the change in choice of marginal distributions. When  $W(\cdot)$  is a Gaussian process with a particular moment structure, this gives the Brown-Resnick process for  $Z(\cdot)$  (Brown and Resnick (1977); Davison et al. (2012)). A weakness with this model is that  $G$  can only be specified via a series of evaluations of the multivariate normal distribution function (Genton et al., 2011), though reductions in the numerical difficulties can be achieved using methods of Wadsworth and Tawn (2014) that require additional information about which segments of  $Z(s)$  arise from the same  $Y_t(s)$  process.

### 4.2.3 Pareto processes

An alternative asymptotic characterisation for spatial extremes is to use the threshold exceedance analogue of max-stable processes, namely generalised Pareto processes (Ferreira and de Haan, 2014). The strategy behind this development is a spatial extension of the argument that led to the GPD in the univariate case, i.e., we condition

on an extreme event occurring and then study the properties of this extreme event as the threshold that determines the extreme event tends to a limiting value. Specifically, define the process  $T(s)$  by

$$\{T(s); s \in \mathcal{S}\} \stackrel{d}{=} \lim_{u \rightarrow \infty} \left[ \{X(s) - u; s \in \mathcal{S}\} \mid \sup_{s \in \mathcal{S}} X(s) > u \right].$$

Then  $T(s)$  is a Pareto process, with the property that  $\sup_{s \in \mathcal{S}} T(s)$  is distributed as a standard exponential random variable but that  $T(s)$  can be negative for some values of  $s \in \mathcal{S}$ . Critically, for all  $v > 0$ ,  $T(s)$  then satisfies

$$\{T(s) - v \mid \sup_{s \in \mathcal{S}} T(s) > v\} \stackrel{d}{=} T(s),$$

so that  $T(\cdot)$  satisfies the threshold-stability property. Pareto processes are the only such processes that possess this property. This property is illustrated in Figure 4.2.2, which shows a set of realisations of the process  $X(s)$  in black with a subset (indicated in red) corresponding to realisations with  $\sup_{s \in \mathcal{S}} X(s) > u$ . Thus each of the red realisations is approximately a Pareto process, i.e.,  $u + T(s)$ .

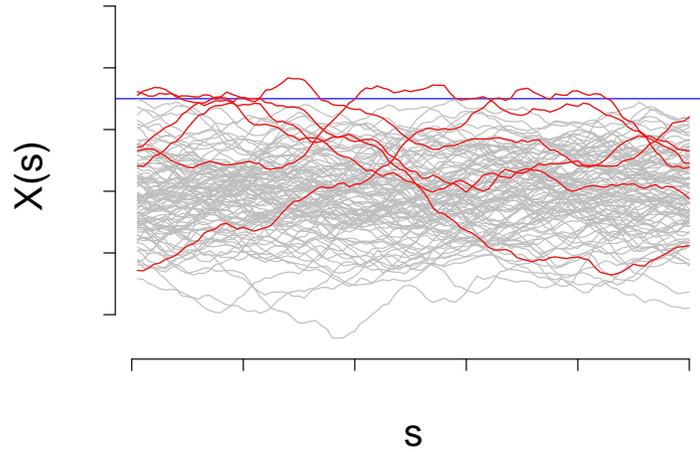


Figure 4.2.2: Illustration of a Pareto process, showing realisations of a process  $X(s)$  (grey lines), where for some chosen threshold  $u$  (blue line), with the realisations where  $\sup_{s \in \mathcal{S}} X(s) > u$  (red lines) being approximately distributed as  $u + T(s)$ .

To help study Pareto processes it is helpful to draw on the max-stable characterisation (4.2.4) of Ferreira and de Haan (2014). A Pareto process is simply one of the latent processes that underpin the  $Z(s)$  process. It follows that we can represent the Pareto process  $T(s)$  by

$$T(s) = R + W(s), \quad (4.2.5)$$

where  $R$  is a standard exponential random variable which is independent of a stochastic process  $W(\cdot)$ , satisfying  $\sup_{s \in \mathcal{S}} W(s) = 0$ . A common choice for this is to set  $W(\cdot)$  to be a Gaussian process, such as the Gaussian process family used for Brown-Resnick processes (Brown and Resnick, 1977). In this case,  $W(\cdot)$  is a conditional Gaussian process, conditional on  $\sup_{s \in \mathcal{S}} W(s) = 0$ . A benefit of working with Pareto processes over max-stable processes is that the process is derived from a single realisation of

$W(\cdot)$  and  $R$ . Therefore, conditionally on  $R$ , the  $T(s)$  process is a conditional Gaussian process which is a massive simplification of inference relative to max-stable processes. However, the conditioning for  $W(\cdot)$  is complex as it applies over all  $s \in \mathcal{S}$ , which makes computation non-trivial.

#### 4.2.4 Weakness of Pareto processes

Assuming that the process  $X(s)$ , when it exceeds a threshold  $u$ , is exactly a Pareto process means that for large  $u$ ,  $X(s) = u + T(s)$ . Hence, for some  $s_0, s \in \mathcal{S}$ , we have

$$X(s_0) = u + R + W(s_0) \text{ and } X(s) = u + R + W(s),$$

where  $R$  is a standard exponential random variable and  $W(s)$  is independent of  $R$ , so that when  $X(s_0)$  is large,

$$X(s) = X(s_0) + \{W(s) - W(s_0)\}.$$

Then  $X(s_0)$  is interpretable as the size of the event and  $\{W(s) - W(s_0)\}$  as the spatial profile of the event. Critically, the shape and size of these extreme events are independent for Pareto processes. Thus events are equally likely to retain the same type of spatial profile whatever their size at a point  $s_0$ . An illustration of this is shown in the top row of panels in Figure 4.2.3, with the profile of the events unchanged as the size of events increases (left to right panels). As a consequence, Pareto processes are asymptotically dependent at all lags, as

$$\lim_{x \rightarrow \infty} \mathbb{P}(X(s) > x | X(s_0) > x) > 0 \quad (s_0, s \in \mathcal{S}).$$

However, in practice we almost never observe such processes. Instead, we often see events becoming more localised, as seen in the bottom row of panels in Figure 4.2.3.

Here we see events of the small initial magnitude and profile as in the top row become more spatially localised around the maximum value as the maximum value of the field increases. For this type of process, which include Gaussian processes,

$$\lim_{x \rightarrow \infty} \mathbb{P}(X(s) > x | X(s_0) > x) = 0 \quad (s_0, s \in \mathcal{S}, s \neq s_0),$$

so the process is asymptotically independent at all lags.

It may be that both of these formulations are too simplistic and the process is asymptotically dependent up to a certain lag  $h_{AD}$ , then asymptotically independent when the lag exceeds  $h_{AD}$ , such as in the models of Bacro et al. (2016). Consequently, we want an inference method which does not pre-determine that the process is asymptotically dependent at all lags, so that  $h_{AD} = \infty$  (like max-stable and Pareto processes), or asymptotically independent at all lags with  $h_{AD} = 0$  (like Gaussian processes). In particular, we would like to have the flexibility to determine the lag  $h_{AD}$  at which this transition occurs. The models introduced in Section 4.3 do precisely that.

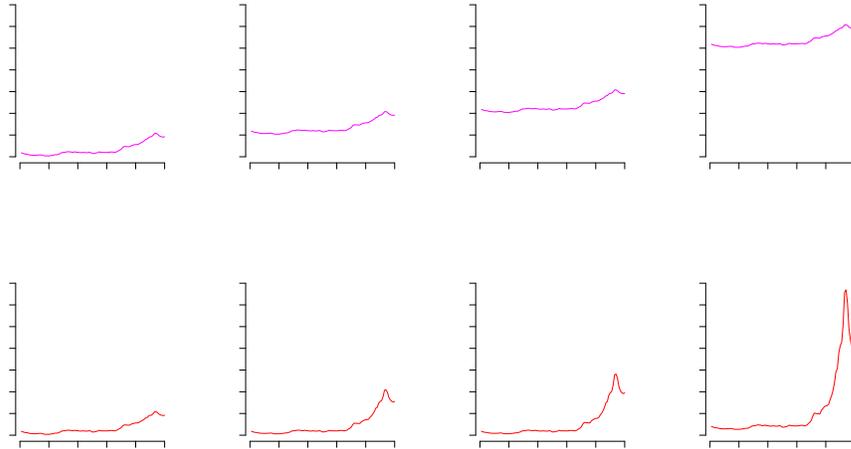


Figure 4.2.3: Illustration of types of extremal spatial behaviour. The top row shows a process which retains the same spatial profile as the event becomes more extreme, corresponding to asymptotic dependence. The bottom row depicts the extreme event becoming more localised as its magnitude increases, commonly seen in practice and corresponding to asymptotic independence.

## 4.3 Conditional extremes

### 4.3.1 Asymptotics for conditional multivariate extremes

Consider a vector random variable  $\mathbf{X} = (X_1, \dots, X_d)$  with Gumbel marginals; for  $i < j$ , we shall use the notation  $\mathbf{X}_{i:j} = (X_i, \dots, X_j)$ . For simplicity, we will assume that all the variables are non-negatively dependent and that  $\mathbf{X}$  has a joint density.

Heffernan and Tawn (2004) propose an asymptotically justified conditional multivariate extremes approach for modelling the extremes of a vector  $\mathbf{X}$  given  $X_1$  is large.

To explore the conditional distribution  $\mathbb{P}\{\mathbf{X} \leq \mathbf{x} \mid X_1 > u\}$  for large  $u$ , we use an asymptotically justified form for this distribution as  $u \rightarrow \infty$ . If  $\mathbf{x}$  is fixed, in general the limit distribution will be a degenerate distribution. Hence  $\mathbf{X}$  needs to be normalised appropriately so that the limiting conditional distribution is non-degenerate as  $u \rightarrow \infty$ . Heffernan and Resnick (2007) propose that  $\mathbf{X}_{2:d}$  is linearly normalised as a function of either  $X_1$  or  $u$ . Normalising by  $X_1$  leads to simpler limit models, thus we use the approach of Heffernan and Tawn (2004) and carry out this normalisation.

Heffernan and Tawn (2004) assume that there exist functions  $a: \mathbb{R} \rightarrow \mathbb{R}^{d-1}$  and  $b: \mathbb{R} \rightarrow \mathbb{R}_+^{d-1}$ , such that for  $x > 0$ ,

$$\mathbb{P}\left(\frac{\mathbf{X}_{2:d} - a(X_1)}{b(X_1)} \leq \mathbf{z}_{2:d}, X_1 - u > x \mid X_1 > u\right) \rightarrow G_{2:d}(\mathbf{z}_{2:d}) \exp(-x), \quad (4.3.1)$$

as  $u \rightarrow \infty$  with  $\mathbf{z}_{2:d} \in \mathbb{R}^{d-1}$  and where  $G_{2:d}$  is a joint distribution function that is non-degenerate in each margin. A key property of the limit (4.3.1) is that the limiting distribution factorises, corresponding to large values of  $X_1$  being independent of the associated normalised  $\mathbf{X}_{2:d}$ .

Under weak assumptions on the joint distribution of  $\mathbf{X}$ , Heffernan and Resnick (2007) show that, componentwise,  $a$  and  $b$  must be regularly varying functions satisfying certain constraints, which for Gumbel margins corresponds to each of the components of  $a$  (respectively  $b$ ) being regularly varying functions of index 1 (respectively less than 1). Within this structure Heffernan and Tawn (2004) found that a simple form for  $a$  and  $b$  holds for a very broad range of copulas. In particular, they assume that

$$a(x) = \boldsymbol{\alpha}_{2:d}x \text{ and } b(x) = x^{\boldsymbol{\beta}_{2:d}}$$

where  $\boldsymbol{\alpha}_{2:d} = (\alpha_2, \dots, \alpha_d) \in [0, 1]^{d-1}$  and  $\boldsymbol{\beta}_{2:d} = (\beta_2, \dots, \beta_d) \in [0, 1]^{d-1}$ . This canonical parametric subfamily of  $a$  and  $b$  provides a parsimonious, yet flexible, family for statistical modelling.

Different types of extremal dependence lead to different values of the extremal dependence parameters  $\boldsymbol{\alpha}_{2:d}$  and  $\boldsymbol{\beta}_{2:d}$ . For  $2 \leq j \leq d$ , when  $\alpha_j = 1$  and  $\beta_j = 0$  the variables  $(X_1, X_j)$  are asymptotically dependent; when  $\alpha_j < 1$ , these variables are asymptotically independent. Within the asymptotic independence case a further resolution of the dependence structure is possible, with  $0 < \alpha_j < 1$  or  $\alpha_j = 0$  and  $\beta_j > 0$  corresponding to positive dependence, and near independence when  $\alpha_j = \beta_j = 0$ . When there is a multivariate normal copula (with  $\rho_{ij} > 0$  corresponding to the correlation parameter between variables  $i$  and  $j$ ), then  $\alpha_j = (\rho_{1j})^2$ ,  $\beta_j = 1/2$  and  $G_{2:d}$  is the joint distribution function of a multivariate normal distribution which has mean vector  $\mathbf{0}$ , variance (for the  $j$ th variable) of  $2\rho_{1j}^2(1 - \rho_{1j}^2)$  and a correlation between variables  $i$  and  $j$  of  $(\rho_{ij} - \rho_{1i}\rho_{1j})/[(1 - \rho_{1i}^2)(1 - \rho_{1j}^2)]^{1/2}$ ; see Heffernan and Tawn (2004).

Unfortunately there is no finite parametric form for  $G_{2:d}$  or its marginal distributions, so a range of approaches have been taken. Heffernan and Tawn (2004) use empirical estimates for  $G_{2:d}$ ; Lugrin et al. (2016a) utilise a mixture of Gaussian distributions, while Towe et al. (2016) use a Gaussian copula with kernel smoothed marginal distributions. Here, we make the assumption that  $G_{2:d}$  is multivariate normal with margins  $N(\mu_j, \sigma_j^2)$  for  $j = 2, \dots, d$ . Under this assumption,

$$X_j \mid \{X_1 = x\} \sim N(\alpha_j x + \mu_j x^{\beta_j}, \sigma_j^2 x^{2\beta_j}) \quad (x > u, j = 2, \dots, d), \quad (4.3.2)$$

with parameters  $\boldsymbol{\alpha}_{2:d}$ ,  $\boldsymbol{\beta}_{2:d}$ ,  $\boldsymbol{\mu}_{2:d} = (\mu_2, \dots, \mu_d)$  and  $\boldsymbol{\sigma}_{2:d} = (\sigma_2, \dots, \sigma_d)$ .

### Inference

In order to estimate the dependence parameters  $\boldsymbol{\alpha}_{2:d}$  and  $\boldsymbol{\beta}_{2:d}$ , a pseudo-likelihood is constructed with  $\mathbf{X}_{2:d} \mid X_1 = x$  (for  $x > u$ ) treated as independent with marginals of the joint conditional distribution stated in equation (4.3.2). The estimation of these dependence parameters is performed through maximum pseudo-likelihood for the  $n_u$  pairs for which  $X_1 > u$ . The likelihood is then

$$L(\boldsymbol{\alpha}_{2:d}, \boldsymbol{\beta}_{2:d}, \boldsymbol{\mu}_{2:d}, \boldsymbol{\sigma}_{2:d}) \propto \prod_{i=2}^d \prod_{j=1}^{n_u} \frac{1}{x_{ij}^{\beta_i} \sigma_i} \exp \left\{ - \frac{\left( x_{ij} - [\alpha_i x_{1j} + \mu_i x_{1j}^{\beta_i}] \right)^2}{2x_{ij}^{2\beta_i} \sigma_i^2} \right\},$$

for  $-\infty < \mu_i < \infty$ ,  $\sigma_i > 0$ ,  $-1 \leq \alpha_i \leq 1$ , and  $-\infty < \beta_i < 1$  for  $i = 2, \dots, d$ , and where  $x_{ij}$  denotes component  $i$  for the  $j$ th exceedance of  $u$  by  $X_1$ . The maximum pseudo-likelihood estimates are denote by  $\hat{\boldsymbol{\alpha}}$ ,  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\sigma}}$ . Then realisations of  $\mathbf{Z}_{2:d} \sim G_{2:d}$  are given by

$$\mathbf{z}_{2:d}^{(j)} = \left( \frac{x_{ij} - \hat{\alpha}_i x_{1j}}{(x_{1j})^{\hat{\beta}_i}}, i = 2, \dots, d \right) \text{ for } j = 1, \dots, n_u \quad (4.3.3)$$

where  $x_{1j} > u$  for each  $j$ . This sample of  $\mathbf{Z}_{2:d}$  is used to obtain an empirical estimate of the joint distribution function  $G_{2:d}$ . Consequently, we have a model for the joint tail behaviour of  $\mathbf{X}$ , when  $X_1$  is large. This enables us to make inferences beyond the range of the observed data with large  $X_1$ ; for more details of fitting these models over different conditioning variables and methods for simulating jointly rare events see Heffernan and Tawn (2004) and Keef et al. (2013a).

A limitation of the inference for models in the conditional multivariate extremes approach is that self-consistency of the different conditional distributions is not ensured. This may lead to inconsistencies when calculating joint exceedance probabilities such as

$$\begin{aligned}\mathbb{P}(X_1 > u, X_2 > u) &= \mathbb{P}(X_1 > u | X_2 > u) \cdot \mathbb{P}(X_2 > u) \\ &= \mathbb{P}(X_2 > u | X_1 > u) \cdot \mathbb{P}(X_1 > u),\end{aligned}$$

since the models for  $X_1 | X_2 > u$  and  $X_2 | X_1 > u$  are not necessarily equal. Liu and Tawn (2014) discussed this problem, making a range of proposals to reduce this problem. One proposal which removes the issue is to assume that  $(X_1, X_2)$  are exchangeable, which implies for that the associated parameters and distributions are equal for each conditional distribution. For non-exchangeable pairs though, whilst removing the self-consistency problems, this induces biased inference.

### 4.3.2 Models for conditional spatial extremes

This section gives an indication only of how some aspects of the multivariate conditional extremes methods could be extended to the spatial setting. For simplicity, it is assumed that  $X(s)$  is isotropic as well as stationary and with Gumbel marginals, and let  $h = |s - s_0|$  be the distance between two sites  $s_0, s \in \mathcal{S}$ . A consequence of these standard spatial statistics assumptions is that the joint distribution of pairs  $\{X(s_1), X(s_2)\}$  are exchangeable variables, for all pairs  $s_1, s_2 \in \mathcal{S}$ , and hence there are none of the issues of self-consistency that are present in multivariate cases.

The natural spatial extension of the Heffernan and Tawn (2004) conditional multi-

variate extremes representation to the spatial context assumes that there exist normalisation functions  $\alpha(h) \in [0, 1]$  and  $\beta(h) \in [0, 1]$  for all  $h > 0$ , with  $\alpha(0) = 1$ ,  $\beta(0) = 0$ , such that as  $u \rightarrow \infty$ ,

$$\left\{ \frac{X(s) - \alpha(h)X(s_0)}{X(s_0)^{\beta(h)}} : s \in \mathcal{S}, X(s_0) - u > x \right\} \mid X(s_0) > u \\ \xrightarrow{d} \{\mu(h) + \sigma(h)Z(s) : s \in \mathcal{S}, E\},$$

where,  $\mu(\cdot)$  and  $\sigma(\cdot)$  are deterministic functions with  $\sigma(h) > 0$  for  $h \neq 0$  and  $\mu(0) = \sigma(0) = 0$ ;  $Z(\cdot)$  is a random process with  $\mathbb{E}[Z(s)] = 0$  and  $\text{Var}[Z(s)] = 1$  for all  $s \in \mathcal{S}$  and  $E$  is a standard exponential random variable that is independent of the process  $Z(\cdot)$ .

Assuming that this limit result holds exactly for a large choice of threshold  $u$  gives a model structure

$$X(s) \mid \{X(s_0) > u\} = \alpha(h)X(s_0) + X(s_0)^{\beta(h)}W(s - s_0) \quad (s \in \mathcal{S}), \quad (4.3.4)$$

where  $\{X(s_0) - u\} \mid X(s_0) > u$  follows a standard exponential distribution and is independent of  $W(\cdot)$ , where  $W(s) := \mu(h) + \sigma(h)Z(s)$  is a spatial isotropic process with  $W(0) = 0$ , marginal mean  $\mu(h)$ , marginal variance  $\sigma^2(h)$  and correlation function  $\rho(\cdot)$ . As in the multivariate conditional extremes case, we will make a modelling assumption that  $W(\cdot)$  is a Gaussian process with a correlation structure to be estimated. This Gaussian assumption may appear to be a very strong assumption but it is the assumed process for all Brown-Resnick max-stable processes (Davison et al., 2012), for the type of processes given in Engelke et al. (2015) and in a conditional form for Pareto processes (Ferreira and de Haan, 2014).

The key is then to make inference on  $\alpha(h), \beta(h), \mu(h), \sigma(h)$  and the correlation structure of  $W(\cdot)$  so that inference can be drawn on the process (4.1.3) (after back transformation from  $X(s)$  to  $Y(s)$ ). There are some interesting special cases of this model:

**Pareto type process** If  $\alpha(h) = 1$  and  $\beta(h) = 0$  for all  $h \geq 0$ , then model (4.3.4) is exactly that given by the process of Engelke et al. (2015) and is strongly related to the Pareto process, given by expression (4.2.5), as it is essentially the same process but subject to different conditioning constraints. It is asymptotically dependent at all lags.

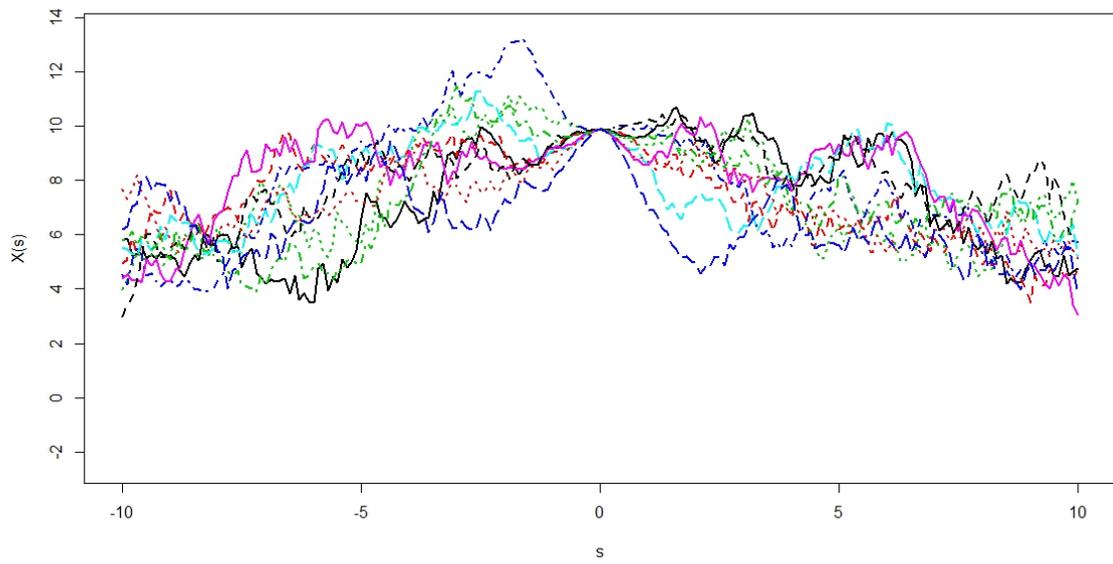
**Gaussian process** From results in Section 4.3.1 on multivariate normal copulas,  $\{\alpha(h)\}^{1/2}$  satisfies the properties of a valid spatial correlation function and  $\beta(h) = 1/2$  for  $h > 0$ , then model (4.3.4) is exactly the limiting conditional extremal process of a Gaussian process; it is asymptotically independent for all positive lags.

**Mixture process** If  $(\alpha(h), \beta(h)) = (1, 0)$  for all  $h \leq h_{AD}$  but  $\alpha(h) < 1$  for  $h > h_{AD}$  then the process is asymptotically dependent up to lag  $h_{AD}$  and asymptotically independent otherwise.

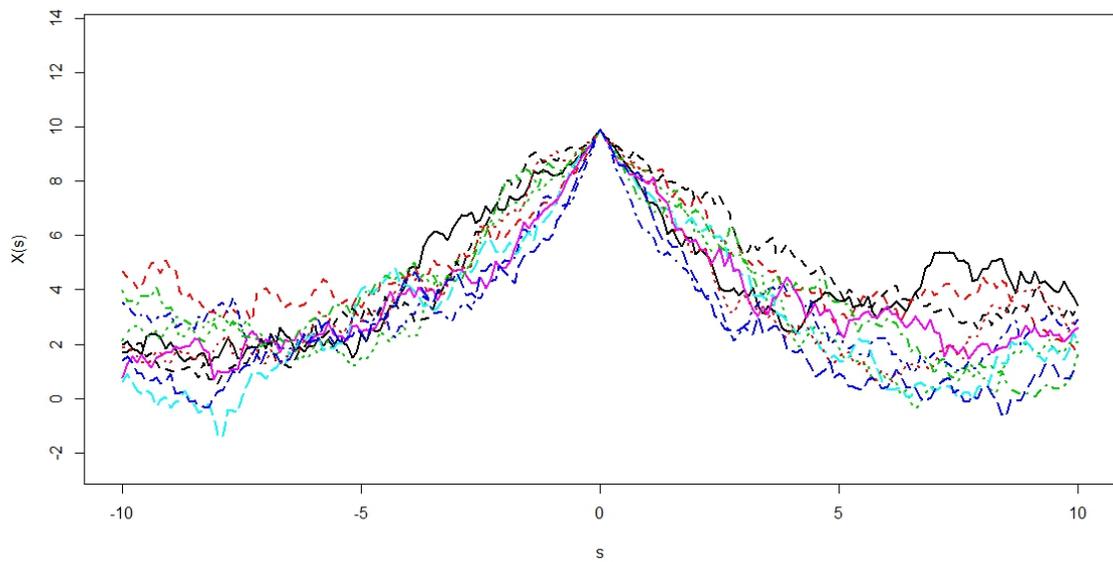
The aim therefore is to identify if any of these structures is present in an application. To help give insight into these three different sub-classes of model (4.3.4), in Figure 4.3.1 we show repeated simulations of a 1-dimensional process with  $X(0)$  equal to the marginal 99.995% quantile, thus all simulations are equal for  $s = 0$ . Firstly, we can see that the three types of process behave differently from one another in the

location of a large event, with all replications for a given process type having broadly similar behaviour. Secondly, note that if  $X(0)$  was more or less extreme the only effect would be a vertical shift of the process when the process is in on-extreme states.

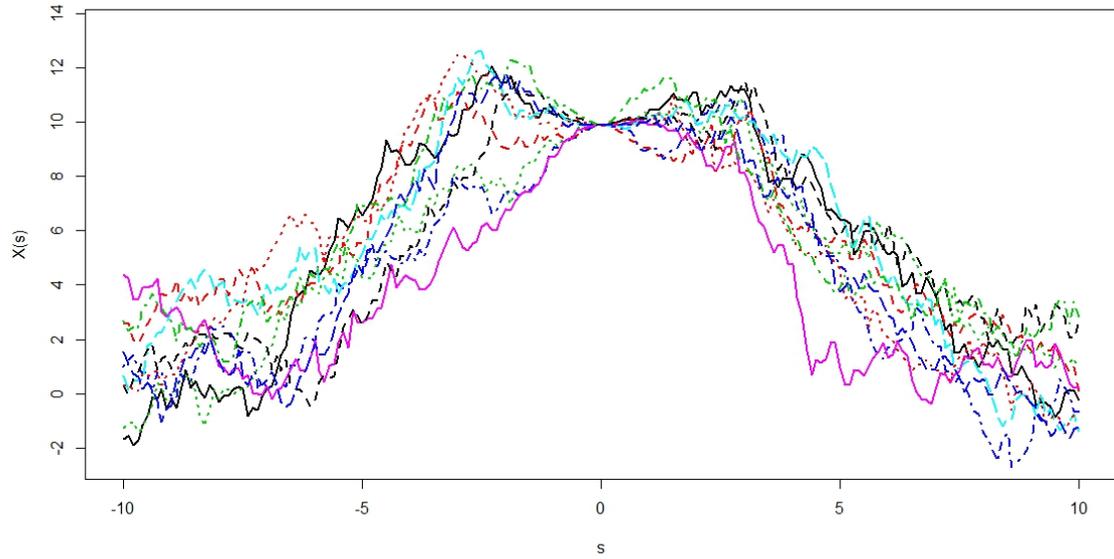
Pareto type processes remain of the same order of magnitude over the space  $\mathcal{S}$ . Specifically, it has a mean negative drift away from an extreme level, with here, due to the choice of correlation function and the Gaussian process for  $Z(s)$ , in the neighbourhood of  $s = 0$  the extremal process is a Brownian motion with negative drift in distance  $|s|$  from the extreme event. Consequently there is a positive probability of  $X(\tau)$  being large given  $X(0)$  is large for all  $s \in \mathcal{S}$ , hence the process is asymptotically dependent for all lags  $\tau$  as defined by definition (4.1.4). In contrast, for the extremal Gaussian process events decay much more rapidly, essentially geometrically, until the process returns to a non-extremal state. Thus, it can be seen that the process is asymptotically independent for all lags  $\tau$ , but with the rate of convergence of the non-limit probability in definition (4.1.4) to 0 is dependent on  $\tau$ . The mixture type processes behave like Pareto type processes up to lag  $h_{AD}$  from the extreme event at  $s = 0$ , but then decay more rapidly to until the process returns to a non-extremal state. Hence the mixture process is seen to be asymptotically dependent up to lag  $h_{AD}$  and asymptotically independent for larger lags.



(a)



(b)



(c)

Figure 4.3.1: Illustrations of Pareto type, Gaussian and mixture extremal processes on a space  $\mathcal{S} = (-10, 10)$ . In all cases  $X(0)$  is in an extreme state (equal to the 99.995% marginal quantile), and the latent Gaussian process  $Z(s)$  has mean and standard deviation of  $\mu(h) = \mu_c$  and  $\sigma(h) = \sigma_c$  for  $h > 0$  and correlation function  $\rho(h) = \exp(-h/3)$ . Illustration as follows: (a) Pareto type process with  $\mu_c = -0.4$ ,  $\sigma_c^2 = 1.3$ ; (b) Gaussian process  $\alpha(h) = \exp(-h/3)$ ,  $\mu_c = 0.06$ ,  $\sigma_c^2 = 0.6$ ; (c) mixture process with  $h_{AD} = 3$ ;  $\alpha(h) = \exp(-|h - h_{AD}|/3)$  for  $h > 3$ ,  $\beta(h) = 0$ ,  $\mu_c = -0.05$ ,  $\sigma_c^2 = 1.3$ .

## 4.4 Applications

### 4.4.1 Offshore risk from waves

#### Background

The accurate modelling of extreme wave heights is of key importance in the design of offshore structures. Such structures must be constructed adhering to strict guidelines, which themselves rely on the assessment of how often extreme events occur. Methods for spatial extremes are useful for enabling the likelihood over sites to be constructed for improved marginal parameter inference and for spatial risk assessment over a network of offshore structures. For the former, we need a reliable spatial dependence model to ensure valid inferences are made for the smoothly varying marginal parameter models (Randell et al., 2015). For the latter, companies with offshore interests often have more than one asset to insure and so having a joint risk assessment that gives the probability that none of the assets will be affected in their lifetime is required.

The aim of our analysis is to test the viability of the conditional spatial extremes methods set out in Section 4.3.2 for application to significant wave data (defined as four times the standard deviation of the sea-surface) in the North Sea region shown in Figure 4.4.1. The data come from a numerical model driven by observational wind data but have been filtered and transformed to give one observation per storm event and to have the marginal wave directional effects removed. This leaves 1680 storm events where the event is extreme for at least one of the 150 locations on the grid. A

description of the data and pre-processing is given in Randell et al. (2016) with these data representing for Shell Research their test-bed for spatial analysis methods.

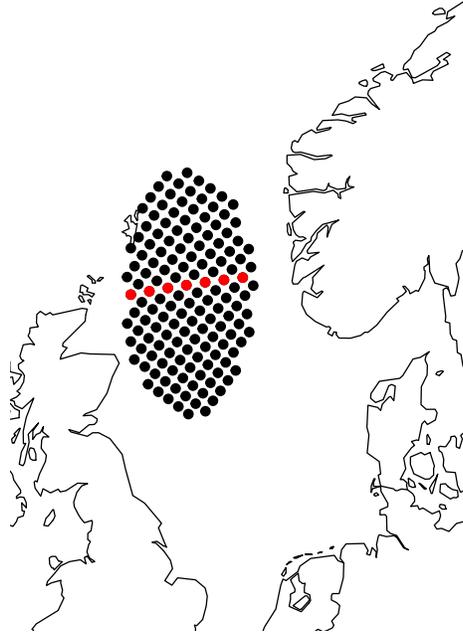


Figure 4.4.1: Map of sampling locations in the North Sea from which the data are collected, with the particular transect used for model fitting highlighted in red.

Directionality of the waves is found to be present in the spatial dependence structure, so for simplicity we perform our spatial inference on a directional transect through the grid, reducing the field to approximately 1 dimension. The transect used is orientated east-west in the centre of the grid and consists of 7 sites; this is highlighted in Figure 4.4.1. The use of transects for this ocean basin is similar to that as used in Ross et al. (2017a), though max-stable processes are fitted in that case.

## Methods

We apply the multivariate conditional extremes model of Section 4.3.1 to identify the potential structure for the spatial functions  $\alpha(h)$ ,  $\beta(h)$ ,  $\mu(h)$  and  $\sigma(h)$ . For illustrative purposes, we only condition on the west-most site in this transect and then fit the model to the other locations in the transect. This is not necessary, however, and more information can be extracted by suitably combining the different conditional distributions. Similar studies using other transects are expected to give weaker levels of extremal dependence as our selected transect direction aligns with most major storm tracks.

To obtain estimates for the model, some assumptions are made for the form of  $G_{2:7}$  in limit (4.3.1). Specifically, to correspond to the Gaussian process formulation in Section 4.3.2, we take  $G_{2:7}$  to be the distribution function of a multivariate normal with mean and standard deviation vectors  $(\mu_2, \dots, \mu_7)$  and  $(\sigma_2, \dots, \sigma_7)$  and with correlation function at lag  $h$  taken to be  $\rho^h$ . This model is fitted jointly over sites, with a multivariate normal likelihood, unlike in all previous applications of Heffernan and Tawn (2004) which use the pseudo-likelihood in Section 4.3.1. For each fitted parameter  $\theta$ , we set  $\theta(i) = \theta_{i+1}$  for  $i = 1, \dots, 6$  so that, for example,  $\alpha(1) = \alpha_2$ .

In fitting the conditional extremes model, the 0.8 quantile of  $X(s)$  has been selected as the conditioning threshold  $u$ . This value was chosen for  $u$  as this seemed to satisfy the required approximate independence property of limit (4.3.1) both for that level and that it holds for all higher threshold choices. In practice, the threshold choice is a compromise between being sufficiently low to utilise enough data whilst being suitably

high so that the asymptotic argument in (4.3.1) provides a good approximation.

## Results

Exploratory analysis using the model described in Section 4.4.1 showed that there was no evidence for  $\beta(h)$  to vary with  $h > 0$ , and so we take  $\beta(h) = \beta_c$ , where  $0 \leq \beta_c < 1$  is some constant, for  $h > 0$ ; our estimated model gives  $\hat{\beta}_c = 0.17$ . Also, we found  $\hat{\rho} = 0.9$ . The corresponding  $\alpha(i), i = 1, \dots, 6$ , estimates are shown in Figure 4.4.2, with the values presented here as pointwise estimates of the function  $\alpha(h)$ . The estimates are consistent with the physical characteristics that may be expected from extreme waves. For  $0 \leq h < h_{AD}$  such that  $\alpha(h) = 1$  the process is asymptotically dependent, then it would be anticipated that a nearby location is likely to experience an extreme wave of the same order of magnitude if the conditioning site has observed such an event. We see that if this holds then  $0 \leq h_{AD} < 1$  based on the 95% confidence intervals for the pointwise estimates. We also see that the degree of dependence is estimated to decrease as the distance between sites increases, which is physically realistic. The decay of the pointwise estimates for  $\alpha(h)$ , for  $h > h_{AD}$ , seems smooth and the analysis suggests a simple parametric form for  $\alpha(h)$  of the form

$$\alpha(h) = \begin{cases} 1 & \text{if } h < h_{AD} \\ \exp\{-\gamma(h - h_{AD})\} & \text{if } h \geq h_{AD}. \end{cases}$$

Previous spatial modelling of significant wave heights has utilised models of max-stable processes, see Section 4.2.2. However, these are asymptotically dependent, i.e.,  $\alpha(h) = 1$  for all  $h$ . We can see from Figure 4.4.2 that this is not a good model for  $h \geq 1$  for these wave data.

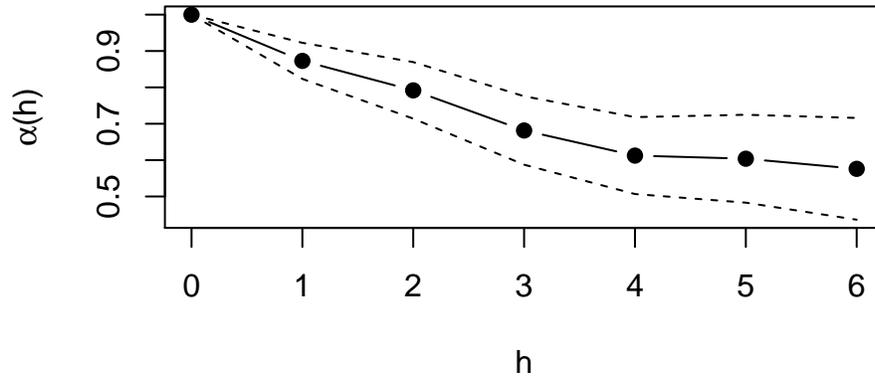


Figure 4.4.2: Pointwise estimates of  $\alpha(h)$  from the multivariate conditional extremes fit, conditioned on the west-most location in the transect. Lag  $h = 0$  corresponds to the conditioning site, with  $h = 6$  being the parameter estimate at the most easterly site. Estimates are for integer values of  $h$  and these are shown to be linearly interpolated to show we know that the function is continuous. The dotted lines show 95% confidence intervals for the pointwise estimates.

Next, consider the estimated mean and standard deviation functions of the limit process  $W(\cdot)$ . Pointwise estimates for  $\mu(h)$  and  $\sigma(h)$  are given in Figure 4.4.3. Both functions behave very similarly; as the distance between the two sites increases, the limit process increases in mean and standard deviation but with decreasing rate for larger distances. This form of  $\sigma(h)$  is as expected since the unpredicted variability is likely to increase as the extremal dependence weakens, but the former is a feature that justifies investigation in future research to understand why this property arises. On this initial analysis, however, it appears that  $\mu(h) \propto \sigma(h)$  would form a good spatial

model.

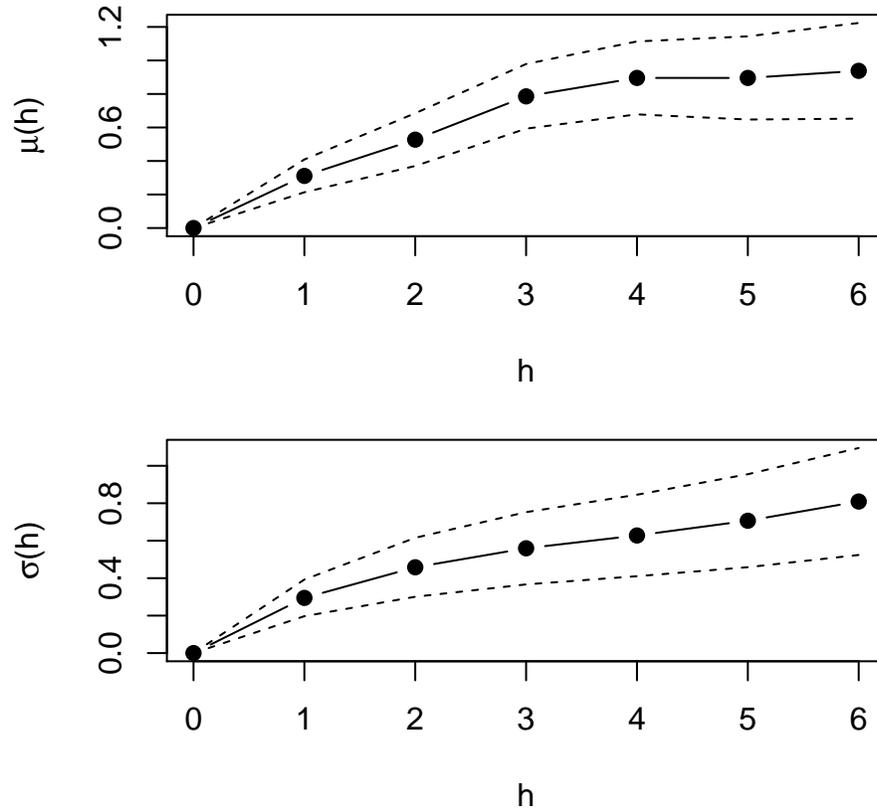
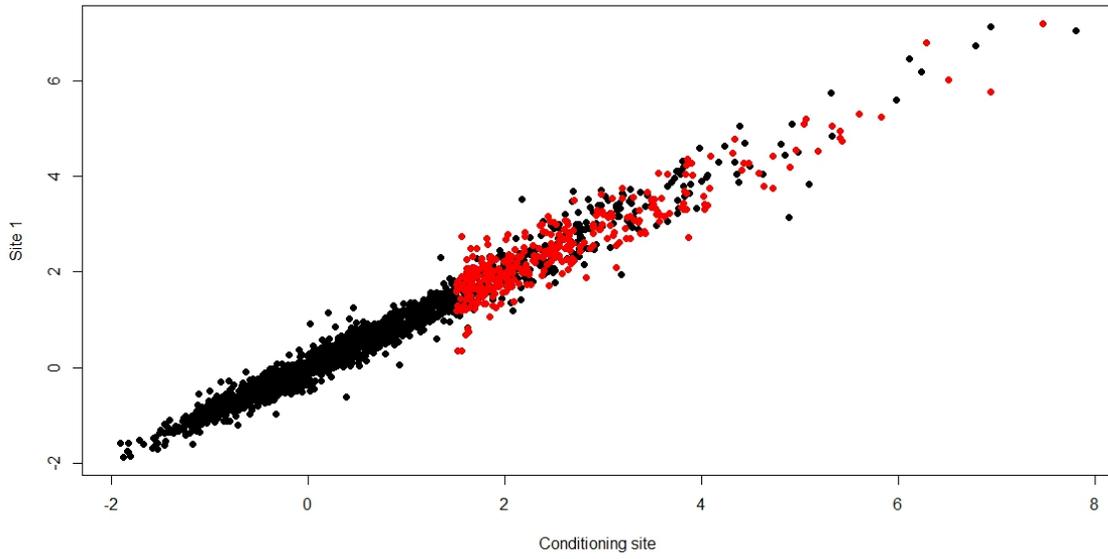


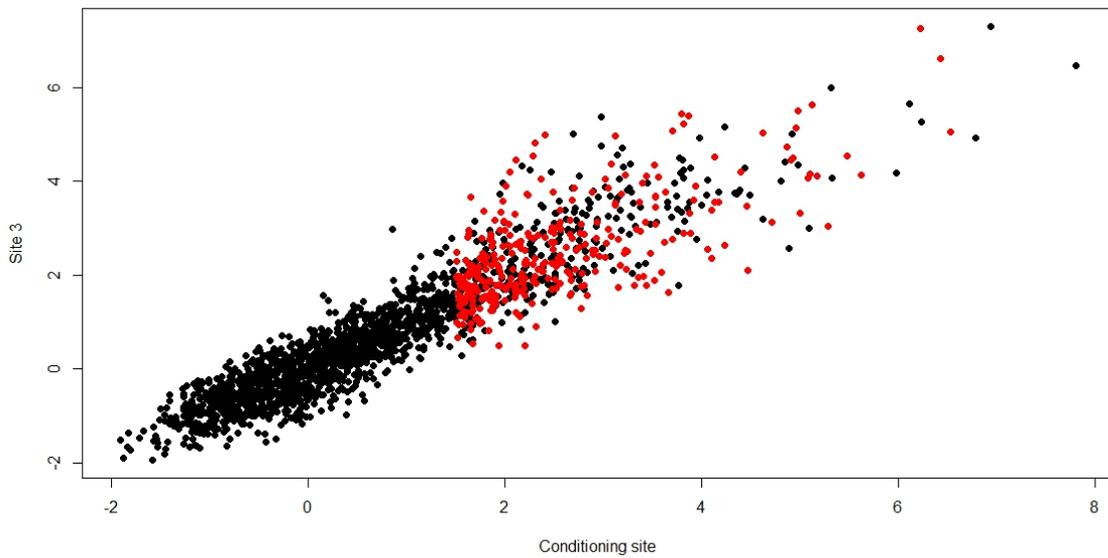
Figure 4.4.3: Pointwise estimates of  $\mu(h)$  and  $\sigma(h)$  with properties shown identical to Figure 4.4.2.

To assess whether the estimates of  $\alpha(h)$  and  $\beta(h)$  are reasonable, we simulate using our fitted model realisation of  $\{X(s), X(s+h)\}$ , for  $h = 1, 3, 6$ , where  $X(s)$  is the standardised (to Gumbel margins) wave height at the most westerly site of the transect and is above the modelling threshold  $u$ . The observed data (black), 1680 points from these joint distributions with Gumbel margins, together with 336

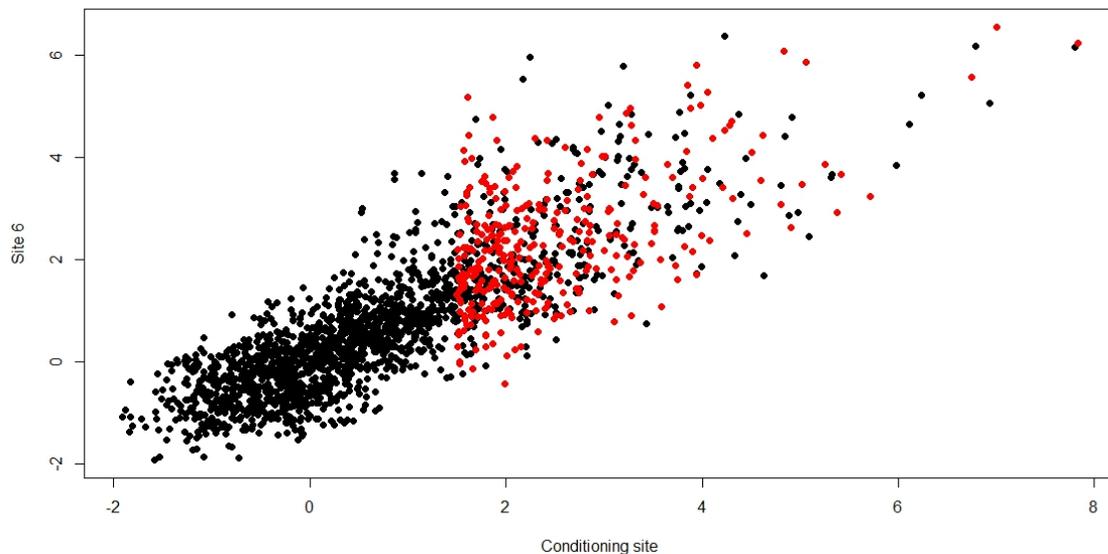
simulated points with  $X(s) > u$  (red) are shown in Figure 4.4.4. It appears from these simulations that the fitted model provides a reasonable fit to the data; for each pair of sites, the distribution of extreme wave data appears to have been captured well. Hence, the model appears to be appropriate for modelling significant wave height in the North Sea on this particular transect. More work will be undertaken to establish if this is the case for further transects in this ocean basin and also to determine how to pool information across transects to estimate the functions  $\alpha(\cdot), \beta(\cdot), \mu(\cdot), \sigma(\cdot)$  that change smoothly over distance or separation depending on whether isotropy is found to hold for extreme wave events.



(a)



(b)



(c)

Figure 4.4.4: Simulations from the fitted multivariate conditional extremes model; black points are the data on Gumbel margins, whilst red points are simulated data from the fitted model: (a), (b) and (c) show these data when  $h = 1, 3, 6$  respectively. In each case, the  $x$ -axis is the standardised wave height at the conditioning site (the most westerly in the transect), with the  $y$ -axis being the standardised wave height at the other site.

## 4.4.2 Understanding widespread flood risk

### Background

Understanding flood risk is an important issue for insurance companies, the government, as well as local communities. Previous events have shown that flood events can affect large spatial areas and have devastating impacts on transport and infrastructure (Shaw et al., 2010). Therefore, it is of paramount interest to understand the features of these events and plan future defences to be able to withstand physically plausible events that we have not yet observed.

Flooding is a continuous spatial process but restricted to the river network; however as is common with environmental problems we only have access to observations at a finite number of locations. Therefore, we want to be able to make predictions from these pointwise locations that are consistent with the underlying spatial process (Davison et al., 2012). Furthermore, the dependence structure of measurements of river flow is highly complex; this is because river flow gauges considered spatially distant through standard metrics such as Euclidean distance can in fact be similar because they lie within the same catchment (Asadi et al., 2015; Shaw et al., 2010). Previous studies such as Asadi et al. (2015) have used the max-stable processes (see Section 4.2.2), however this approach does not suit large scale studies. Other approaches such as Keef et al. (2009); Lamb et al. (2010); Towe et al. (2016) have adopted the conditional multivariate extremes model stated in Section 4.3.1 to understand widespread flood risk.

### **National Flood Resilience Review**

During winter 2015, consecutive storms Desmond, Eva and Frank hit the UK causing widespread flooding across large regions of northern England. These storms required significant responses from the emergency services and in some cases the army to help with the protection of property as well as infrastructure (Lamb et al., 2015). Due to the unprecedented effect of these storms and often the rapid response required, the UK government set up the National Flood Resilience Review (NFRR). The aim of the NFRR was to gain a better understanding of the drivers of flooding in the UK as well as the current methods to deal with the associated risks and damages caused by flooding (Government, 08 September 2016).

In particular, the scientific advisory group of the NFRR wanted to understand more about the likelihood of flooding in the UK and move towards thinking about risks at a national scale rather than location by location. To better understand the risk of widespread flooding, a comprehensive analysis of UK river flow gauges was required. As we are interested in understanding the characteristics of widespread flooding in the UK, a flexible spatial extreme value model that is able to accommodate the known features within the data is required. For example, this needs to model that flood events can be both localised as well as national and not all sites are likely to be extreme concurrently. The Heffernan and Tawn (2004) conditional multivariate extreme value model, stated in Section 3, satisfies both of these modelling requirements.

## Methods

Observations of river flow gauges were obtained from the National River Flow Archive maintained by the Centre of Ecology and Hydrology, as well as from Environment Agency records. Before any statistical modelling was undertaken, a quality assurance of the data was performed. This quality assurance required the data to have at least 20 years of observations with a relatively small percentage of missing values, this requirement enabled robust estimation of the parameters of the associated statistical models (see Sections 2.1 and 3). Furthermore, gauges were removed from the analysis if unnatural changes in the time series were observed, for example if a dam was installed further upstream. This results in unnatural changes of the time series at downstream gauge being present in the time series (Shaw et al., 2010). This quality assurance process resulted in 916 suitable gauging records. To maintain consistency with previous studies of UK flooding, an event was defined to last for a period of time of up to 7 days (Keef et al., 2009). The statistical analysis includes extensions to the Heffernan and Tawn methodology as stated in Section 4.3.1 such as the handling of missing values as well as efficient simulation techniques for high dimensional data sets and methods to model the rate of the number of extreme events per year (Keef et al., 2013a). These aspects are key when modelling spatial river flow data sets with more details of these methods found in Keef et al. (2013a). In order to assess the validity of the statistical models, comparisons such as those shown in Figure 4.4.4 were made. From the statistical analysis, 10000 years worth of events were simulated in Gumbel

margins, we denote these by

$$\{\tilde{X}_t(s_i); i = 1, \dots, 916, t = 1, \dots, 10000n_y\}, \quad (4.4.1)$$

where  $n_y$  is the average number of events in the region per year. This simulated event set includes events that are larger than those observed in the data for at least one site but with the dependence structure of these events being consistent with the features from the observed extreme events (Keef et al., 2009). This simulated event set then allows us to estimate a number of summary statistics for a range of severities of events to help us characterise the behaviour of flooding across the UK.

### Conditional probability calculation

In order to test the validity of simulations from the conditional extreme value model, we compare the calculation of conditional probabilities from both the observed and simulated data sets. For all return levels, the non-limit conditional probability in equation (4.1.4) is calculated relative to a conditioning gauge, which in this case is situated on the river Severn. For the empirical data, the conditional probability was calculated relative to the 99th percentile (approximately a 5 month level) as well as to a level equivalent to the one year return level, the estimates of this can be seen in Figures 4.4.5 (a) and (b) respectively. For the Heffernan and Tawn (2004) model estimates of the conditional probability the empirical conditional probability from the simulated data set was evaluated for both a 10 and 100 year return level, see Figures 4.4.5 (c) and (d) respectively.

In both cases, the strongest dependence is seen with nearby gauges as well as those

that lie within the river Severn catchment. However, the spatial dependence is not stationary, as distant gauges can still have strong extremal dependence, which is larger than those gauges nearby. This feature is due to the similarity of their catchments with the catchment of the conditioning gauge. Focusing on Figures 4.4.5 (a) and (b), when we consider higher levels the conditional probability decreases, this suggests that as events become more severe, they are also becoming more localised. Higher conditional probabilities from the observed data sets cannot be considered as there is insufficient data to produce stable estimates. This decaying conditional probability characteristic though is also observed for the higher levels considered in Figures 4.4.5 (c) and (d), which show our model-based estimates. There is also a smooth transition in Figure 4.4.5 between the estimates of the conditional probabilities from the observed and simulated data sets.

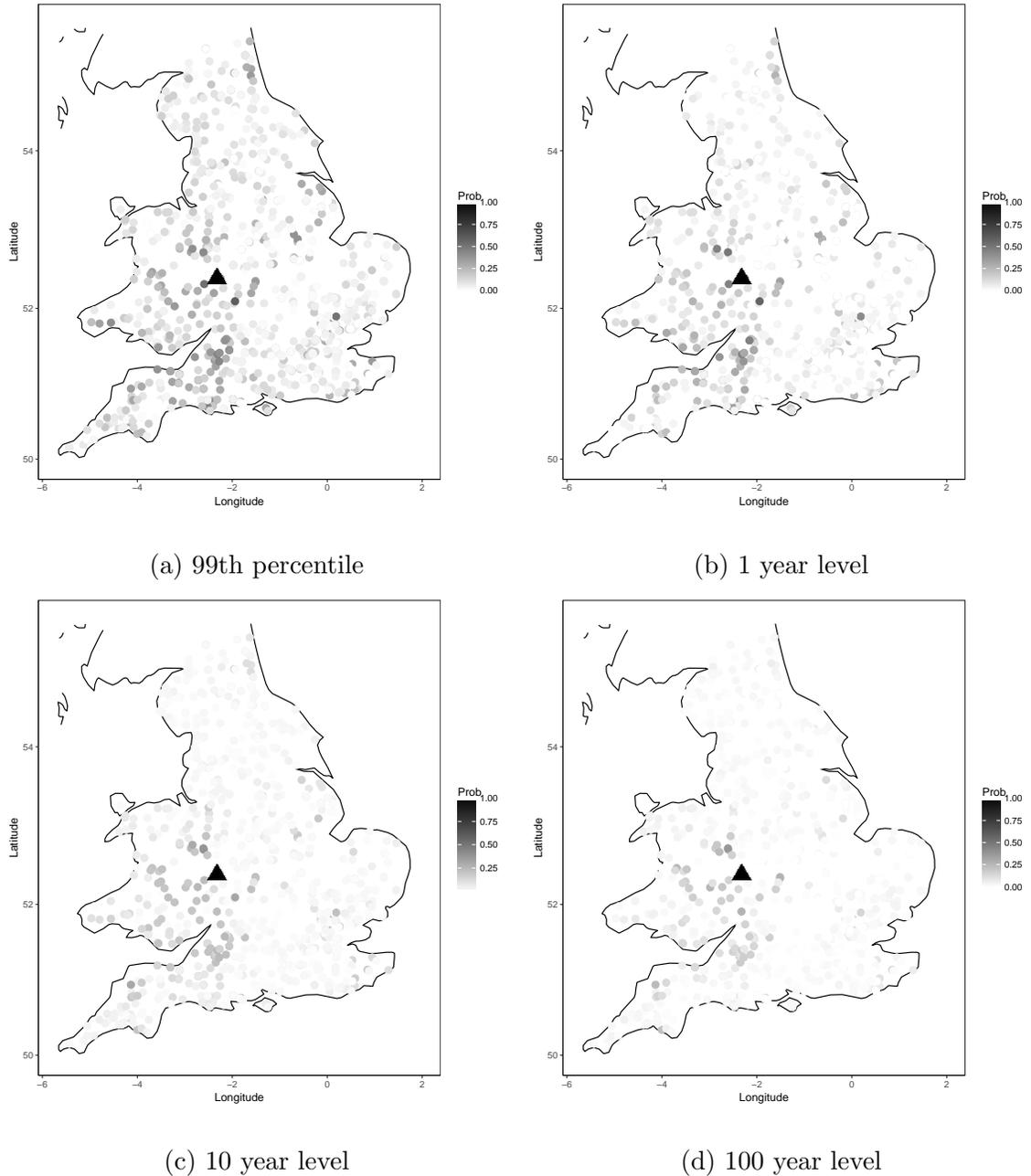


Figure 4.4.5: Comparisons of the non-limit conditional probability (4.1.4) for (a) the 99th percentile and (b) the one year return period from the observed data set; (c) and (d) show this conditional probability estimated using our model for the 10 and 100 year return periods respectively. The triangle symbol represents the conditioning gauge for the estimate, this gauge is situated in the river Severn catchment.

If the statistical model had assumed asymptotic dependence between river flow gauges, the conditional probabilities shown in Figure 4.4.5 would be estimated as invariant to conditioning return level. Therefore, if the 99% quantile was used to fit the model, comparing Figures 4.4.5 (a) and (d) shows that this leads to an error in spatial extremal dependence estimation. In this particular case, there would be massive over-estimation of the spatial extremal dependence between river flow gauges. These comparisons confirm that the conditional extreme value model of Heffernan and Tawn (2004) is accurately capturing the extremal dependence observed in spatially extreme river flows.

### Scenario evaluation for the National Flood Resilience Review

The analysis of the observed and simulated data sets in Section 4.4.2 confirmed that the features of the observed data set are being captured in the models represented by the simulated event set. As a result, we are able to use the simulated event set as a proxy for a long observational record to answer fundamental questions for flood risk management posed by the NFRR such as:

*What is the chance of an extreme river flow occurring at one or more gauges across England and Wales, somewhere within the national river gauge network in any one year?*

To frame this question in terms of our notation, we need, for an arbitrary year  $t$ , to estimate  $1 - \mathbb{P}(M_{Y(s_i),t} < y_{s_i,T}; i = 1, \dots, 916)$ , where  $M_{Y(s),t}$  is the annual maximum in year  $t$  for the river flow in site  $s$  and  $y_{s,T}$  is the  $T$  year return level at site  $s$ . This

probability is identical to  $1 - \mathbb{P}(M_{X(s_i),t} < x_T; i = 1, \dots, 916)$ , where  $x_T$  is the  $T$  year return level on Gumbel margins. We estimate the second term in this probability using the simulated sample (4.4.1) as

$$\hat{\mathbb{P}}(M_{Y(s_i),t} < y_{s_i,T}; i = 1, \dots, 916) = \frac{1}{k} \sum_{j=1}^k \mathbb{1} \left( \max_{i=1, \dots, 916} M_{\tilde{X}(s_i);j} < x_T \right), \quad (4.4.2)$$

where  $k = 10000n_y$  and  $\mathbb{1}(A)$  is the indicator function of event  $A$ .

The estimates of  $1 - \mathbb{P}(M_{Y(s_i),t} < y_{s_i,T}; i = 1, \dots, 916)$  are shown as  $T$  varies in Figure 4.4.6 using the modelled dependence with estimator (4.4.2) and under the two limiting cases that assume all of the 916 gauges are either completely independent or completely dependent. Here the complete independence case assumes that there is no association between when flooding occurs at each of the 916 gauges, whereas the complete dependence assumes that each of the 916 gauges behave identically. The benefit of the conditional extremes approach is that we are able to estimate the probability whatever  $T$ , i.e., even for events with return periods that are greater than the severity of the events captured in the observed data set. For the NFRR, the key feature of this analysis was that the probability of observing a 1 in 100 year event at any of the 916 gauging stations in any given year is 0.78, so its very likely a 100 year event occurs somewhere in this region.

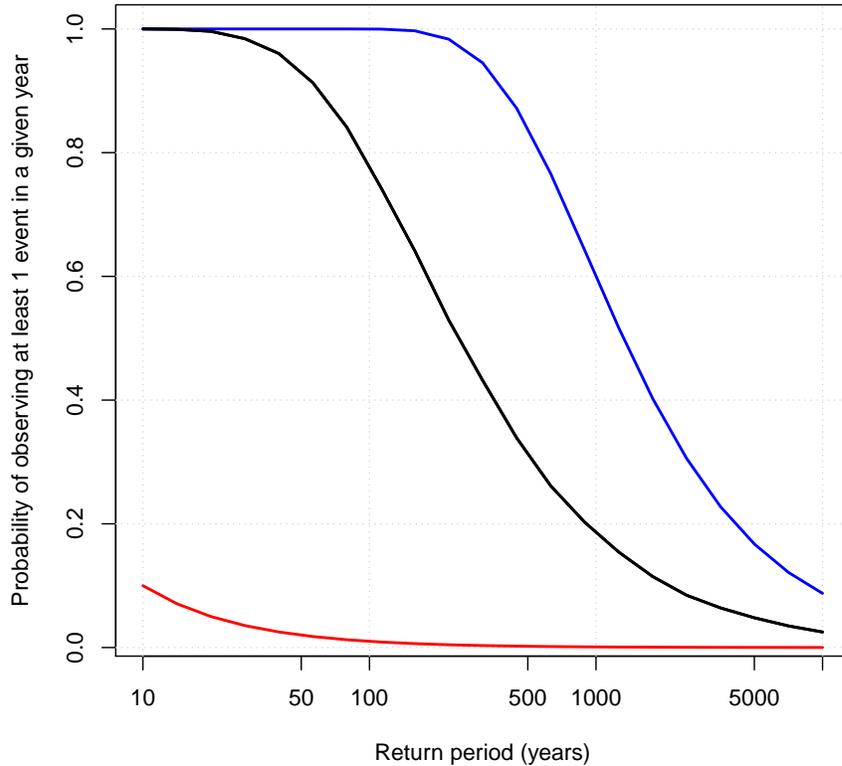


Figure 4.4.6: Comparison of the three dependence models used to estimate probability of observing at least one event above a  $T$ -year return period for a given year: our model for the dependence (black), under a complete dependence model (blue) and under complete dependence (red).

This analysis considered only those locations where there are gauges with river flow measurements; current research is addressing how this question can be answered for every place along the river network, i.e., to estimate  $1 - \mathbb{P}(M_{Y(s),t} < y_{s,T}; \text{ for } s \in \mathcal{S})$ . It should be also noted that our study focusses on England and Wales, reflecting the scope of the NFRR (flood risk management is a devolved matter in the United

Kingdom, with separate arrangements in place in Scotland).

*What is the chance of an extreme river flow occurring in one or more Local Resilience Forums, somewhere within the national river gauge network in any one year?*

The analysis shown in Figure 4.4.6 considered the probability of observing a flood event at any gauge across the river network. However, for emergency planning purposes, interest lies in determining the spatial extent of potential events. Within the England and Wales, responses to natural hazards are managed through 42 Local Resilience Forums (LRFs), which we denote by  $\{L_p; p = 1, \dots, R = 42\}$ . Therefore, it seems natural to define events in accordance to the number of LRFs that receive a  $T$  year event at some gauge. Let  $M_{X,t}(L_p) = \max_{i \in L_p} M_{X(s_i),t}$ , i.e., it is the maximum level, on Gumbel scale, over the  $p$ th LRF and let  $M_{X,t}(L_{(r)})$  be the  $r$  largest value of  $M_{X,t}(L_p), p = 1, \dots, R$  in year  $t$ , so  $M_{X,t}(L_{(1)}) > \dots > M_{X,t}(L_{(R)})$ . To understand the regional extent of spatial flood events, we are interested in whether in an arbitrary year  $t$ , at least  $r$  LRFs have exceedances of the marginal  $T$  return level, i.e., the  $\{M_{X,t}(L_{(r)}) > x_T\}$ . We estimate this probability using the simulated sample (4.4.1) by

$$\hat{\mathbb{P}}(M_{X,t}(L_{(r)}) > x_T) = \frac{1}{k} \sum_{j=1}^k \mathbb{1} \{M_{\tilde{X},j}(L_{(r)}) > x_T\},$$

where  $k = 10000n_y$ .

Estimates of the probability for  $r = 1, \dots, 4$  are shown in Figure 4.4.7. As expected the estimates for at least  $r = 1$  region being above a  $T$ -year return period in any given year is consistent with the analysis shown in Figure 4.4.6. Most interesting is that in

any given year there is 0.35 probability of at least a 1 in 100 year event occurring in at least four LRFs.

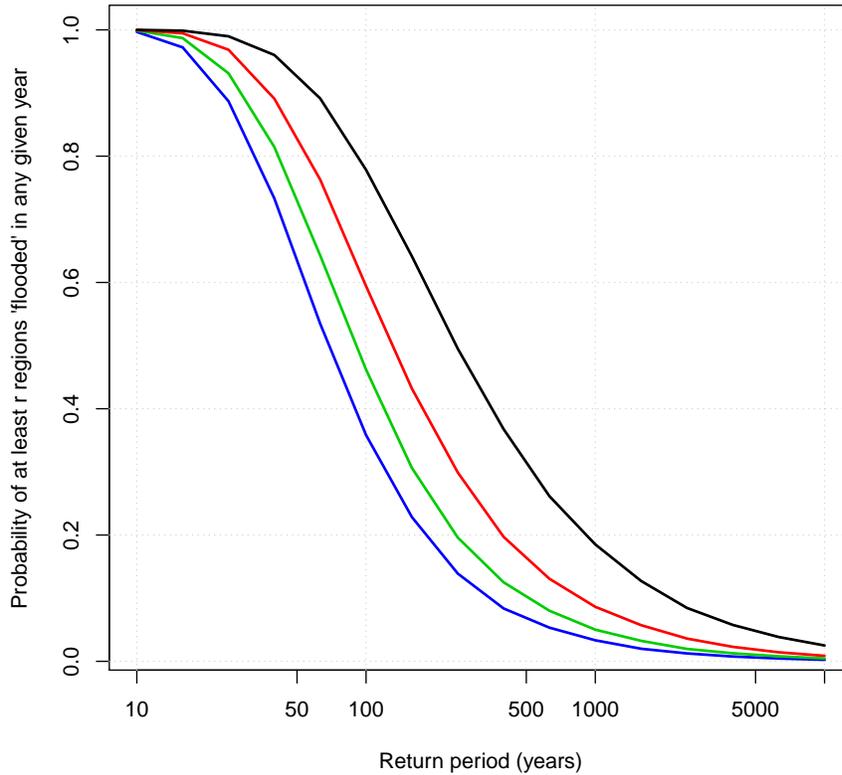


Figure 4.4.7: Estimated probability of observing at least  $r$  LRF regions above a  $T$ -year return period in any given year. The black, red, green and blue curves show the cases for when  $r = 1, 2, 3, 4$  respectively.

Both of the questions proposed by the NFRR highlighted that flooding is more common than one might expect. The typical communication of return period is a single site measure. The conditional spatial extreme value model of Heffernan and Tawn (2004) allows us to provide robust answers to these national scale questions

through carefully capturing the complex dependence structure of a high dimensional set of river flow gauges. The uncertainty around the estimates of the conditional probability as well as the point estimates shown in Figures 4.4.6 and 4.4.7 from the NFRR can easily be assessed by bootstrap methods.

The questions proposed by the NFRR were answered by modelling the spatial dependence of gauges on the river network. However, ultimate interest lies in estimating the chance of observing a flood in a given year at any location along the river network. Answering this question is an ongoing research question, which involves exploiting information about the river network as well as modelling the joint dependence of river flow with that of the process of rainfall.

# Chapter 5

## On spatial conditional extremes for ocean storm severity

### 5.1 Introduction

Quantifying extreme ocean environments is important for safe and reliable construction and operation of offshore and coastal infrastructure. Extreme value analysis provides a framework within which the marginal and dependence characteristics of extreme ocean environments can be estimated, and joint inferences corresponding to very long periods of observation made in the presence of non-stationarity with respect to covariates.

The spatial structure of ocean surface roughness within a storm is of particular concern when inferences are based on observations from multiple locations in a neighbourhood. For a given ocean basin, when the distance between two locations is small relative to the spatial extent of a storm low pressure field, it is reasonable to expect

that large values of ocean surface roughness (for a period of time of the order of an hour, quantified in terms of significant wave height  $H_S$ ) at the two locations will be dependent. Moreover, the extent of this spatial dependence will potentially itself be non-stationary with respect to covariates, such as storm direction and season. A reasonable statistical description of  $H_S$  on a neighbourhood of locations should therefore admit appropriately flexible descriptions of extremal spatial dependence. Incorrect specification or estimation of the dependence structure can lead to misleading joint predictions of  $H_S$  on the neighbourhood. We note a number recent articles on spatial extremes with at least some synoptic content, including Davison et al. (2012), Reich and Shaby (2012), Ribatet (2013), Huser and Wadsworth (2018) and Tawn et al. (2018).

A number of recent studies explore the extremal spatial dependence of  $H_S$ . For example, Kereszturi et al. (2016) assesses the extremal dependence of North Sea storm severity using the summary statistics  $\chi$  and  $\bar{\chi}$  (or equivalently  $\eta$ , Coles et al., 1999), outlined in Section 5.3. Estimates for these summary statistics were used to categorise observed extremal dependence as either asymptotic dependence (AD, suggesting that extreme events tend to occur simultaneously) or asymptotic independence (AI, suggesting that extreme events are unlikely to occur together); further discussion of these concepts is given in Section 5.3.1. In Kereszturi et al. (2016), it was found that, in most cases considered, asymptotic independence seemed to be the more appropriate assumption, compared to the assumption of asymptotic dependence. Kereszturi (2016) and Ross et al. (2017a) extend this assessment to include the estimation of a number of max-stable process (MSP) and inverted MSP models (Wadsworth and Tawn,

2012b), including the so-called Smith (Smith, 1990), Schlather (Schlather, 2002) and Brown-Resnick (Brown and Resnick, 1977) models, and corresponding models for the inverted processes. For all models considered, there is evidence that the extremal dependence of  $H_S$  at two locations varies with the distance between the locations, and their relative orientation.

By construction, MSP models considered in Kereszturi (2016) and Ross et al. (2017a) exhibit AD exclusively, whereas inverted MSP models only exhibit AI. In general, we do not know a priori which form of extremal dependence is more appropriate: a decision concerning the form of extremal dependence present in the sample must therefore be made before parameter estimation; this is less than ideal, although estimation of  $\chi$  and  $\bar{\chi}$  can aid this choice. We note alternative AD models including those of Reich and Shaby (2012), Ferreira and de Haan (2014), Rootzen et al. (2018b), Kiriliouk et al. (2019). A number of more sophisticated hybrid models have been proposed (e.g. Wadsworth and Tawn 2012b, Wadsworth et al. 2017, Huser and Wadsworth 2018) spanning dependence classes, but these tend to be rather computationally challenging to estimate in practice.

The conditional extremes model of Heffernan and Tawn (2004) provides an alternative approach to characterising extremal spatial dependence admitting both AI and AD. The conditional extremes model also allows the incorporation of covariate effects (e.g. Jonathan et al., 2014). In the current work, we propose an extension of the conditional extremes method to a spatial setting, known as the spatial conditional extremes (SCE) model. SCE provides a framework to quantify the extreme marginal and dependence structure of  $H_S$  for locations in a neighbourhood, including

the behaviour of extremal dependence of  $H_S$  at different locations as a function of the relative displacements of locations. Model estimation can be achieved using a relatively straightforward Markov chain Monte Carlo (MCMC) scheme, and unlike for MSP models, does not require composite likelihood techniques for parameter estimation and hence does not incur parameter bias, as detailed in Tawn et al. (2018) and Wadsworth and Tawn (2019).

The layout of the article is as follows. In Section 5.2, we present motivating applications involving samples of  $H_S$  on spatial neighbourhoods in the northern and central North Sea. Section 5.3 outlines the spatial conditional extremes model. Parameter estimation is performed using Bayesian inference as described in Section 5.4; details of parameter constraints from Keef et al. (2013a), and the Metropolis-within-Gibbs sampling scheme, are given in the Appendix. Results of the application of the SCE model to the north-south transect of the northern North Sea sample are given in Section 5.5, with corresponding results for the east-west transect (for the northern North Sea), and north-south and east-west transects for the central North Sea reported in Section 5.6. Section 5.7 provides discussion and conclusions.

## 5.2 Motivating application

We consider hindcast data for storm peak significant wave height (henceforth  $H_S$  for brevity) from two neighbourhoods, one in the northern North Sea (NNS) and one in the central North Sea (CNS), as illustrated in Figure 5.2.1. In each neighbourhood, values for  $H_S$  are available on north-south (N-S) and east-west (E-W) transects intersecting at a central location.

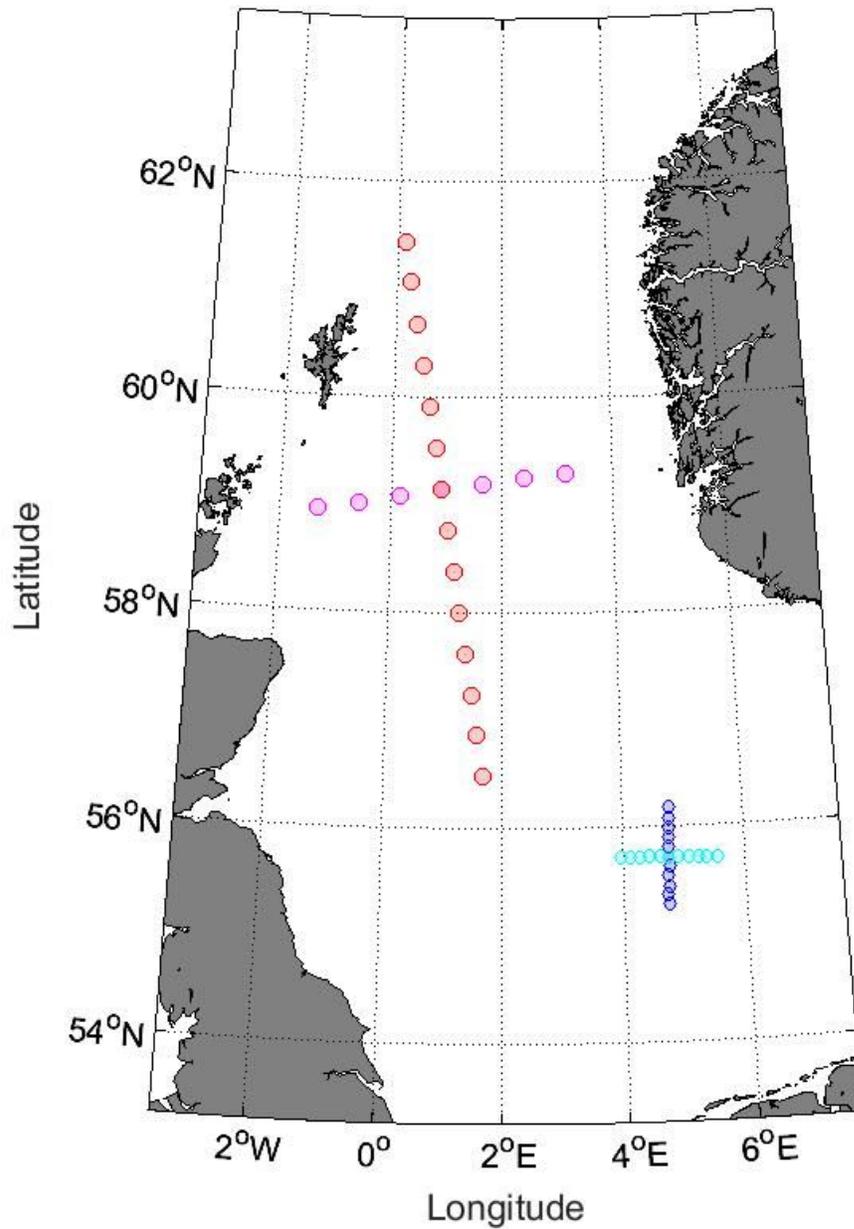


Figure 5.2.1: NNS and CNS locations considered.

The NNS sample corresponds to winter storms (occurring in winter months October-March) from the NEXTRA hindcast (Oceanweather 2002) for 20 locations on the two transects. Storm intervals for a total of 1680 storms during the period 1 Oct 1964 to 31 Mar 1995 were isolated from up- and down-crossings of a sea state significant wave height threshold for the central location, using the procedure outlined in Ewans

and Jonathan (2008). Storm peak significant wave height for each storm interval at each location provided a sample of  $1680 \times 20$  observations for further analysis. For each storm-location combination, the direction (from which waves emanate, measured clockwise from North) at the time of the storm peak, referred to as the storm direction, was also retained. The spatial extremal characteristics of this sample have been examined previously in Ross et al. (2017a); further discussion and illustrations of the data are available there.

The CNS sample corresponds to hindcast storm peak events (occurring at any time of year) for a period of 37 years from 10 January 1979 to 30 December 2015 for 21 locations on the two transects. The hindcast uses CFSR wind fields (Saha et al. 2014) and a MIKE21 spectral wave simulator model (Sorensen et al. 2005) to generate storm time-series at each location. Storm periods were again identified as exceedances of a threshold, non-stationary with respect to season and direction, using the procedure of Ewans and Jonathan (2008) for the central location. In this way, a total of 3104 storm events were isolated per location for further analysis.

As will be explained further in Section 5.3, the SCE model is most conveniently considered for data with marginal standard Laplace distributions. For simplicity, we therefore choose to transform the NNS and CNS samples to standard Laplace scale prior to spatial conditional extremes analysis, as suggested by Keef et al. (2013b), for example. This is achieved by estimating non-stationary marginal models (directional for NNS and directional-seasonal for CNS), following the approach of Ross et al. (2017a) and Ross et al. (2017b), independently per location. Transformed data then follow a standard Laplace distribution for each location.

Figure 5.2.1 illustrates that the inter-location spacing for the NNS hindcast is considerably larger than for the CNS hindcast. For this reason, it is important we compare the variation of extremal spatial dependence between locations explicitly as a function of physical distance (here in kilometres, km). Scatter plots of Laplace-scale storm peak  $H_S$  for pairs of locations separated by distances of 43.0, 171.8 and 300.7 km along the NNS north-south (NNS:N-S) transect, coloured red in Figure 5.2.1, are shown by the black points in Figure 5.5.2 (see Section 5.5.1).

## 5.3 Spatial conditional extremes

### 5.3.1 Characterising extremal dependence

Key concepts in assessing extremal dependence are the notions of asymptotic dependence (AD) and asymptotic independence (AI). Typically, these are assessed through calculating two quantities,  $\chi$  and  $\bar{\chi}$ , introduced by Coles et al. (1999). For bivariate data  $(X, Y)$  with common margins, the quantity  $\chi$  is calculated as

$$\chi = \lim_{u \rightarrow u_F} \mathbb{P}(Y > u | X > u),$$

where  $u_F$  is the upper endpoint of the common marginal distribution  $F$  of the random variables. Then  $\bar{\chi}$  is defined by Coles et al. (1999) as  $\bar{\chi} = 2\eta - 1$ . Here,  $\eta$ , known as the coefficient of tail dependence, is defined by Ledford and Tawn (1996) from the asymptotic approximation, as  $z \rightarrow u_F$ ,

$$\mathbb{P}(X > z, Y > z) \sim \mathcal{L} \left( \frac{1}{\mathbb{P}(X > z)} \right) [\mathbb{P}(X > z)^{1/\eta}],$$

where  $\mathcal{L}(w)$  is a slowly varying function, so that  $\mathcal{L}(tw)/\mathcal{L}(w) \rightarrow 1$  as  $w \rightarrow \infty$  for  $t > 0$ . Coles et al. (1999) provide details on how to calculate estimates for  $\chi$  and  $\bar{\chi}$ . Then  $\chi > 0$  defines the extent of AD present, whereas  $\chi = 0$  suggests the variables exhibit AI. In the latter case,  $\bar{\chi}$  measures the extent of AI present. Tawn et al. (2018) present a spatial equivalent for these measures. Crucially, the spatial characteristics under these two limiting extremal behaviour types are very different; under AD, two (or more) extreme events may occur at separate sites simultaneously, whilst under AI this is not the case. Realistically, a spatial field is likely to exhibit a mixture of these behaviours: at short inter-location distance, asymptotic dependence may prevail; for sites a large distance apart, asymptotic independence is more likely, leading to independence at very large distances. The SCE model accommodates both these possibilities.

### 5.3.2 The conditional extremes model of Heffernan and Tawn (2004)

In its simplest form, for a sample from a pair  $(X, Y)$  of random variables with Laplace marginal distributions, for  $x$  larger than some suitable threshold  $u$ , the model proposed by Heffernan and Tawn (2004) is

$$Y|\{X = x\} = a(x) + b(x)Z, \quad (5.3.1)$$

where  $Z$  is a residual process with typically unknown distribution function  $G$ . A particular form that may be utilised when working with Laplace margins is to set  $a(x) = \alpha x$  and  $\beta(x) = x^\beta$ , for  $-1 \leq \alpha \leq 1$  and  $0 \leq \beta \leq 1$ . This form of the conditional extremes model is used as the basis for the rest of this paper. We also assume that the unknown residual distribution  $G$  is Gaussian.

This model may be extended to a general multivariate case. Let  $Z$  be a multivariate Gaussian distribution with marginal distributions  $N(\mu_j, \sigma_j^2)$  ( $j = 0, \dots, n$ ) for a set of spatial random variables  $(X_0, \dots, X_n)$  with standard Laplace margins. Then we have a multivariate model given by

$$(X_1, \dots, X_n) | \{X_0 = x\} \sim \text{MVN}(\boldsymbol{\alpha}x + \boldsymbol{\mu}x^\beta, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T), \quad (5.3.2)$$

where  $x > u$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^T$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ , and  $\mathbf{B} = \text{diag}(x^{\beta_1}, x^{\beta_2}, \dots, x^{\beta_n})$ , and  $\boldsymbol{\Sigma}$  is the variance-covariance matrix of the residuals  $Z$ . In expression (5.3.2), vector operations are carried out component-wise.

We then have marginal models for  $j = 1, \dots, n$  given by

$$X_j | \{X_0 = x\} \sim N(\alpha_j x + \mu_j x^{\beta_j}, \sigma_j^2 x^{2\beta_j}).$$

Equation (5.3.2) corresponds to the multivariate extension of Equation (5.3.1), in which information about parameters  $\boldsymbol{\theta} = \{\alpha_i, \beta_i, \mu_i, \sigma_i\}_{i=1}^n$  can be shared between random variables. The increased number of parameters, as compared to Equation (5.3.1), means that this model is more computationally-challenging to estimate.

### 5.3.3 The spatial conditional extremes (SCE) model

The SCE model is a spatial extension of the conditional extremes model, following the work of Tawn et al. (2018) and Wadsworth and Tawn (2019). First suppose that  $X(\cdot)$ , the process of interest, is stationary and isotropic and has Laplace marginal distributions. Also suppose that we have sampling locations  $s, s_0 \in \mathcal{S}$ , where  $\mathcal{S}$  is some spatial domain. Then for  $h = |s - s_0|$ , the distance or lag between two sites, we

have

$$X(s) \mid \{X(s_0) > u\} = \alpha(h)X(s_0) + X(s_0)^{\beta(h)}Z(s - s_0). \quad (5.3.3)$$

For a set of fixed spatial locations, Equations (5.3.2) and (5.3.3) are equivalent if we assume that  $Z$  is a residual Gaussian process with mean function  $\mu(h)$  and covariance incorporating  $\sigma(h)$ , as described in Equations (5.4.1) and (5.4.3). Of key importance is that different combinations of parameter values correspond to different types of spatial dependence. We have AD at all distances  $h$  when  $\alpha(h) = 1$  and  $\beta(h) = 0$  for all  $h \geq 0$ , while a mixture of limiting dependence classes is observed if  $(\alpha(h), \beta(h)) = (1, 0)$  for  $h \leq h_{AD}$  but also  $\alpha(h) < 1$  for  $h > h_{AD}$ , for some distance  $h_{AD}$ . The process exhibits AD up to distance  $h_{AD}$  and AI thereafter. Hence, the proposed framework is able to estimate extremal dependence flexibly.

The model set out in expression (5.3.3) gives the behaviour of the process conditional on the process being extreme at  $s_0$ . We need this model to hold for all  $s_0 \in \mathcal{S}$ , and for all of these conditional distributions to be self-consistent with one another. Although the original multivariate conditional extremes models of Heffernan and Tawn (2004) do not impose additional assumptions about pairwise exchangeability, our choice of a stationary isotropic model imposes the required structure on the different conditional models to yield the required self-consistency.

Although not key to developments in this paper, a natural question is whether the conditional models stem from a valid stochastic process. This is clarified by Wadsworth and Tawn (2019). They show that extreme events arising from a valid stochastic process can be generated over space, in such a way that events can be

extreme at any spatial location. Therefore, although the SCE model is not explicitly specified as a stochastic process over space, it is specified implicitly for a process that has an extreme event somewhere in  $\mathcal{S}$ . In this paper, we focus only on questions relating to the behaviour of the process given that there is an extreme event somewhere in  $\mathcal{S}$ . Wadsworth and Tawn (2019) discuss an extension for which this condition is removed.

### 5.3.4 Constraints

For a given  $h$ , we constrain the possible values of pairs of parameters  $(\alpha(h), \beta(h))$  as suggested by Keef et al. (2013a), and outlined in the Appendix. The motivation for this constraint is to impose an ordering of conditional distributions associated with asymptotic independence ( $\alpha(h) < 1$ ) and asymptotic positive dependence ( $\alpha(h) = 1, \beta(h) = 0$ ). In practice, this means that certain combinations of  $(\alpha(h), \beta(h))$  are inadmissible. We also impose gradient-based constraints on  $(\alpha(h), \beta(h))$  following Lugrin (2018), in order to improve the identifiability of the parameter combinations. The motivation for these constraints is ensuring that the derivative, with respect to  $x$ , of  $\mathbb{E}(X(h)|X(0) = x) = \alpha(h)x + \mu(h)x^{\beta(h)}$  is positive, for  $x \geq u$ , with  $u$  some suitable threshold; we then have the constraints  $\alpha(h) + \mu(h)\beta(h)x^{\beta(h)-1} \geq 0$  and  $\alpha(h) \geq 0$  for all  $h$ .

## 5.4 Inference

We consider two variants of the SCE model, differing by the manner in which “linear slope” parameters  $\{\alpha_k\}$  are estimated. In the more general form, outlined in Sec-

tion 5.4.1, these parameters are estimated freely given the sample data, likelihood function and constraints from Section 5.3.4. In the restricted parametric form, outlined in Section 5.4.2, the decay of  $\alpha$  with distance  $h$  follows a prescribed physically-plausible exponential form described by only two parameters. We first consider the more general “free” model.

### 5.4.1 Likelihood for the “free” model

Consider  $p+1$  equally-spaced points on a transect. Suppose we condition on the value of  $H_S$  at a point on the line, marked in black in the two examples of Figure 5.4.1. Our goal is to fit a joint distribution for the values of  $H_S$  at all remaining points, conditioned on an extreme value observed at the conditioning point.

As the set of remaining random variables depends on the conditioning point chosen, we require two indices to define locations: an index  $c \in \{0, 1, 2, \dots, p\}$  to indicate the “conditioning” point, and an index  $j \in \{1, 2, \dots, p\}$  for the remaining points on the line, which we henceforth call “remote” points. The conditioning point will therefore always have an index of the form  $(c, 0)$ , as illustrated in Figure 5.4.1, where  $c = 0$  in the upper image, and  $c = 2$  in the lower.

We indicate the location of the conditioning point as  $s_{c0}$ , and the location of remote points using  $\{s_{cj}\}$ . The distances of remote points to the conditioning point are then denoted by  $\{h_{c0j}\}$ , with  $h_{c0j} = |s_{cj} - s_{c0}|$ . Similarly, distances between remote points  $(c, j)$  and  $(c, j')$  are denoted  $\{h_{cjj'}\}$  with  $h_{cjj'} = |s_{cj'} - s_{cj}|$ ; example values of  $(c, j)$  and  $h_{c0j}$  are indicated in Figure 5.4.1. In the case of the lower image in Figure 5.4.1, note that there are locations that sit a common distance from the conditioning

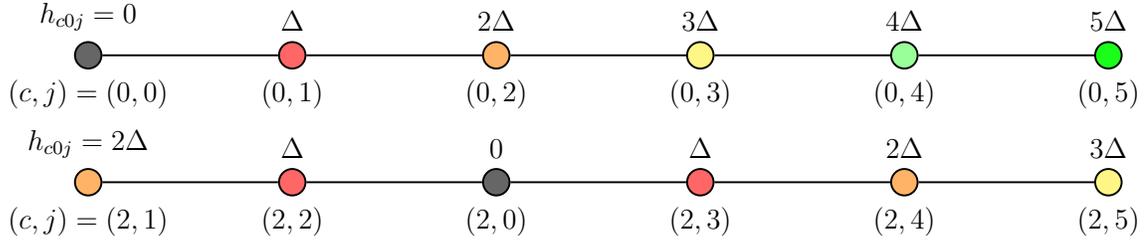


Figure 5.4.1: Illustration of notation used to describe the disposition of points on the line, enabling pooling of data from pairs of locations by distance. The  $(c, j)$  notation is shown below the line, and distance  $h_{c0j}$  given above each point. Points with the same value of  $h_{c0j}$  are shown in the same colour;  $\Delta$  is the inter-location spacing.

point (with the same value of  $h_{c0j}$ , shown as discs of the same colour). We assume that conditional dependence is isotropic on a transect, so that the parameters of the SCE model are at most a function of inter-location distances only. Specifically, the parameters  $\alpha$ ,  $\beta$ ,  $\mu$  and  $\sigma$  are functions of distance from conditioning location, and the residual dependence between remote locations will in addition be a function of distances between remote locations. We seek a model for the joint dependence structure for any number of locations conditional on an extreme value at the conditioning location. For definiteness, consider first the case of two remote locations  $(c, j)$  and  $(c, j')$  (with  $j' \neq j$ ) and conditioning location  $(c, 0)$ , and corresponding random variables  $(X_{cj}, X_{cj'}, X_{c0})$ . We can then write the SCE model as

$$(X_{cj}, X_{cj'}) | \{X_{c0} = x_{c0}\} \sim \text{MVN}_2(\mathcal{M}_{cjj'}, \mathcal{C}_{cjj'}), \quad x_{c0} > q_\tau \quad (5.4.1)$$

where  $q_\tau$  is the quantile of a standard Laplace distribution with non-exceedance prob-

ability  $\tau$ ,

$$\mathcal{M}_{cjj'} = [\alpha(h_{c0j}), \alpha(h_{c0j'})]x_{c0} + [\mu(h_{c0j}), \mu(h_{c0j'})]x_{c0}^{[\beta(h_{c0j}), \beta(h_{c0j'})]}$$

and

$$\begin{aligned} \mathcal{C}_{cjj'} &= \begin{bmatrix} x_{c0}^{\beta(h_{c0j})} & 0 \\ 0 & x_{c0}^{\beta(h_{c0j'})} \end{bmatrix} \begin{bmatrix} \sigma(h_{c0j}) & 0 \\ 0 & \sigma(h_{c0j'}) \end{bmatrix} \begin{bmatrix} 1 & \rho^{h_{cjj'}} \\ \rho^{h_{cjj'}} & 1 \end{bmatrix} \\ &\times \begin{bmatrix} \sigma(h_{c0j}) & 0 \\ 0 & \sigma(h_{c0j'}) \end{bmatrix}^T \begin{bmatrix} x_{c0}^{\beta(h_{c0j})} & 0 \\ 0 & x_{c0}^{\beta(h_{c0j'})} \end{bmatrix}^T \end{aligned} \quad (5.4.2)$$

and  $\rho$  is the between-neighbour residual correlation parameter. We can extend the model to three or more remote locations, or reduce it for one remote location in the obvious way. Hence we can construct a sample Gaussian likelihood  $L$  under the model for all observations, with conditioning variate exceeding  $q_\tau$ , of all possible combinations of two or more locations on the line. We note that in Equation (5.4.2), any correlation function  $K(\cdot)$  could be used in the third matrix; for this work, we specifically use an exponential correlation function, so that  $K(h_{cjj'}) = \rho^{h_{cjj'}}$ .

The likelihood  $L$  is a function of  $\{\alpha(h_{c0j}), \beta(h_{c0j}), \mu(h_{c0j}), \sigma(h_{c0j})\}$ , and  $\rho$  (for different distances  $\{h_{cjj'}\}$  between remote locations). Since the locations are equally-spaced, the values of  $\alpha$ ,  $\beta$ ,  $\mu$  and  $\sigma$  can only be estimated for given distances  $h = k\Delta$ , for lag index  $k = 1, 2, \dots, p$ , where  $\Delta$  is the location spacing for the application (expressed in kilometres). For ease of discussion below, we can therefore write  $L \triangleq L(\theta)$  for the full parameter set as

$$\theta = \{\{\alpha_k, \beta_k, \mu_k, \sigma_k\}_{k=1}^p, \rho\}, \quad (5.4.3)$$

where parameters are indexed by lag  $k$  not distance  $h$ , so that  $\alpha_k = \alpha(k\Delta)$ , etc.

In practice, we also pool all available observations corresponding to unique combinations of distances (i.e., from different choices of conditioning location  $(c, 0)$ ) in the SCE likelihood; we thereby exploit the sample well, in a computationally-favourable manner. Hence, we no longer have the true likelihood under our model but instead a pseudo-likelihood, since the same observation (of each location on a transect) may enter more than one conditioning likelihood contribution (corresponding to conditioning on extreme values at a particular location). Using a pseudo-likelihood as if it is a likelihood is widely known to give point estimates that are asymptotically consistent, but that measures of uncertainty are underestimated. In our Bayesian inference, we expect to underestimate posterior credibility intervals using these pooled data.

Various approaches are available to adjust estimated uncertainty, either by inflating variances or modifying the pseudo-likelihood. In Bayesian inference, the methods of Ribatet et al. (2012) provide an appropriate approach to valid inference for any selected model. In this paper, however, we use the raw pseudo-likelihood for presentation of results, which we justify as follows. The paper focuses on model selection between the free model introduced in this section and a nested parametric model, introduced in Section 5.4.3, with the actual uncertainties of the parameters being of secondary importance relative to the selection of the better model. When using the pseudo-likelihood in place of the full likelihood, inference for the free model will give parameter estimates with credible intervals which are too narrow. Thus, if our subsequent parametric model estimates fall inside these intervals, it suggests that the parametric model provides a better fit than the free model. We emphasise that credible intervals referred to in this work correspond to pseudo-likelihood credible

intervals.

### 5.4.2 MCMC for the free model

We use Bayesian inference to estimate the joint posterior distribution of parameters  $\theta$  from Equation (5.4.3). In our experience, Bayesian inference with reasonable prior specification and MCMC scheme, provides a more reliable approach to parameter estimation, than maximum likelihood techniques. An outline of the procedure, discussion of the priors used and an algorithm, are given in Section 5.8. In brief, we proceed as follows.

First, we use random search to find a reasonable starting value for  $\theta$ . Then, to improve on the starting solution, we use a Metropolis-within-Gibbs algorithm iteratively to sample each of the elements of  $\theta$  in turn. Then we use a *grouped adaptive* random walk Metropolis-within-Gibbs algorithm iteratively to convergence, judged to have occurred when trace plots for parameters and their dependence stabilise. Within the grouped adaptive algorithm, we jointly update the parameters  $(\alpha_k, \beta_k, \mu_k, \sigma_k)$  for each  $k$ , following the adaptive approach of Roberts and Rosenthal (2009) to make correlated proposals. We also adjust proposal standard deviation such that the acceptance rate is optimised for all parameters.

### 5.4.3 Inference for the “parametric- $\alpha$ ” model

Though the constraints of Section 5.3.4 go some way to improving identifiability of suitable parameter combinations, it is still difficult to obtain plausible results in some cases for the free SCE model. Therefore, we shall consider a parametric form for

$\alpha$  based on physical considerations, whereby  $\alpha(h)$  should in general decrease with increasing  $h$ , but also to reduce the dimension of the parameter space, helping parameter identifiability. Specifically, we explore the performance of a SCE model where  $\alpha$  is parameterised as a function of distance, writing

$$\alpha_k = \exp \left\{ - \left( \frac{k}{\kappa_1} \right)^{\kappa_2} \right\}, \quad k = 1, 2, \dots, p \quad (5.4.4)$$

with parameters  $\kappa_1, \kappa_2 > 0$ . The resulting likelihood is  $L \triangleq L(\theta^*)$  with adjusted parameter set

$$\theta^* = \{ \kappa_1, \kappa_2, \{ \beta_k, \mu_k, \sigma_k \}_{k=1}^p, \rho \}. \quad (5.4.5)$$

The MCMC procedure for the parametric- $\alpha$  model is similar to that for the free model, except that  $\kappa_1, \kappa_2$  are separated from the grouped parameters  $(\beta_k, \mu_k, \sigma_k)$  for each  $k$ .

#### 5.4.4 Comparison of free and parametric- $\alpha$ models

To compare results from free and parametric- $\alpha$  models, we use the Deviance Information Criterion (DIC), as proposed by Spiegelhalter et al. (2002), a Bayesian analogue of the Akaike Information Criterion (Akaike 1974). Defining  $D(\theta) = -2 \log L(\theta)$ , where  $L$  is our pseudo-likelihood, we measure model complexity using

$$p_D = \overline{D(\theta)} - D(\bar{\theta}),$$

where  $\overline{D(\theta)}$  is the average of the deviances (calculated after burn-in) and quantifies lack-of-fit. Further,  $\bar{\theta}$  is the average of posterior estimates of  $\theta$ , and note that this is an estimate for the posterior mean. Explicitly, from the final  $m$  iterations of the

MCMC chain, we calculate

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \theta^{(i)} \quad \text{and} \quad \overline{D(\theta)} = \frac{1}{m} \sum_{i=1}^m D(\theta^{(i)}),$$

where component-wise averages are taken in the first equation. The DIC is then calculated as

$$DIC = p_D + \overline{D(\theta)} = 2\overline{D(\theta)} - D(\bar{\theta}),$$

with lower values preferred.

## 5.5 Application to northern North Sea North-South transect (NNS:N-S)

We now apply the free model and parametric- $\alpha$  model to data for the NNS:N-S transect. We start by considering the free model in some detail (in Section 5.5.1), demonstrating that the fitted model explains the data well. Next, in Section 5.5.2, we consider the corresponding parametric- $\alpha$  model, and show that this also fits well, as well as using the DIC, as defined in Section 5.4.4, to show that the fit of free and parametric- $\alpha$  models is similar. The analysis is extended to other transects and locations in Section 5.6. Throughout this section, we adopt a conditioning threshold with non-exceedance probability  $\tau = 0.9$  for the SCE model, after testing the stability of inferences to other choices of threshold. Threshold choice of course involves a bias-variance trade-off: increasing sample size for tail modelling versus inclusion of points from outside the tail region. We note that parameter estimates were relatively stable for choices of extreme value threshold above  $\tau = 0.8$  and below either  $\tau = 0.9$  (for

NNS data) or  $\tau = 0.95$  (for CNS data).

### 5.5.1 Free model

The inference scheme introduced in Section 5.4 is used to estimate parameters  $\theta$  (see Equation (5.4.3)) for the NNS:N-S transect. Posterior mean and pseudo-likelihood credible intervals for estimates of each of  $\alpha(h)$ ,  $\beta(h)$ ,  $\mu(h)$  and  $\sigma(h)$  from the final 1000 iterations (out of a total of 20000 iterations) of the MCMC algorithm described in Section 5.4.2 are shown in Figure 5.5.1. Trace plots showing convergence of MCMC chains are given in Section 5.8. We note that the parameter  $\rho$  has a posterior mode of approximately 0.73 and a 95% pseudo-likelihood credible interval with width of approximately 0.09. We see from Figure 5.5.1 that  $\alpha$  decays exponentially with  $h$ ; this motivates the adoption of the parametric- $\alpha$  model in Section 5.5.2. In particular, we see that  $\alpha(h) \neq 1$  for any  $h$ , so this suggests asymptotic independence is present for all distances  $h$ . We see that  $\mu(h)$  mirrors the behaviour of  $\alpha(h)$  to some extent, in that for  $h < 200$  km,  $\mu$  increases fairly quickly, before stabilising and possibly decreasing again; this illustrates the anticipated dependence between estimates for  $\alpha$  and  $\mu$  in the conditional extremes model. The parameter  $\beta$  is relatively constant with  $h$ , taking values between 0.3 and 0.4, whereas  $\sigma$  increases in general with  $h$ . The behaviour of  $\alpha(h)$  and  $\sigma(h)$  appears reasonable given physical intuition and evidence from the data (the black points) in Figure 5.5.2: extremal dependence reduces as distance between conditioning and remote sites increases, yet the overall variability at each location is constant given that  $H_G$  at each location has been transformed to standard Laplace scale.

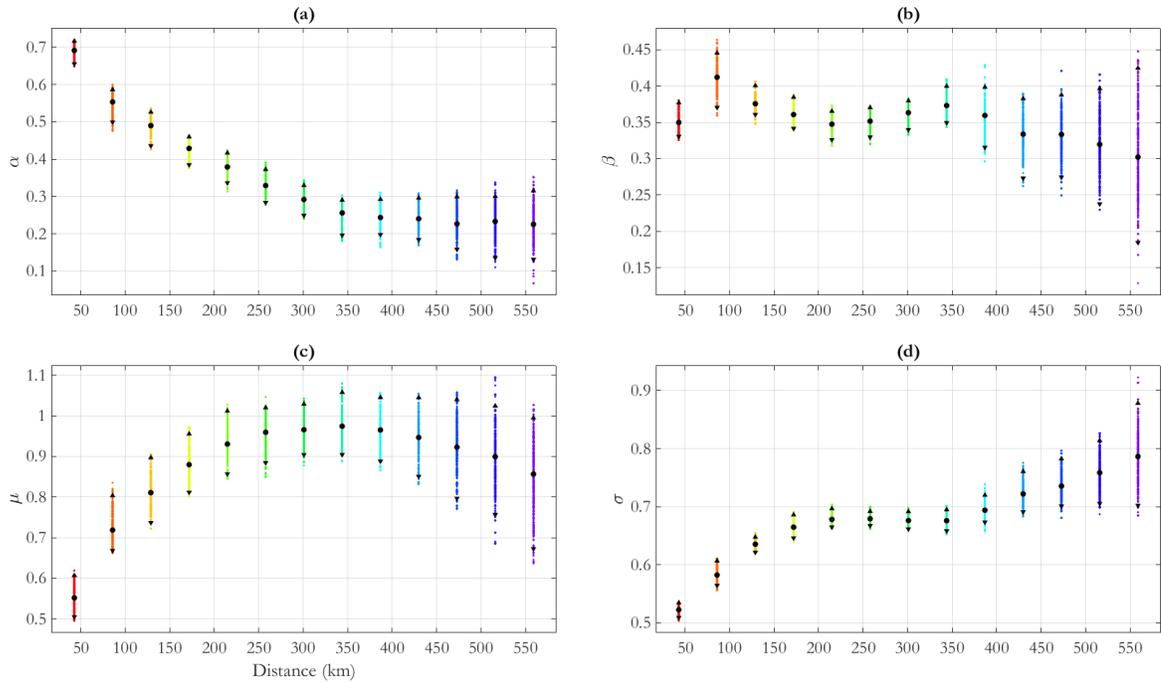


Figure 5.5.1: NNS:N-S transect, free model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with end-points shown as solid triangles).

Figures 5.5.2 and 5.5.3 display diagnostics for the fitted model. Figure 5.5.2 shows the original data on Laplace scale (in black), at three different separations  $h$  of remote and conditioning points. Data simulated under the fitted model are overlaid in red; there is good general agreement. Figure 5.5.3 shows observed sequences of  $H_S$  values along transects with conditioning value (of  $H_S$  at either end-point of the transect) between 3.5 and 4.5 on Laplace scale in blue, as well as two simulated spatial processes from the fitted model, shown in red. The figure also shows the corresponding 95% pseudo-likelihood credible interval under the fitted SCE model with conditioning values between 3.5 and 4.5; again there is general agreement between observation and simulation under fitted model; in particular the simulated processes appear to have

similar smoothness to the observed processes.

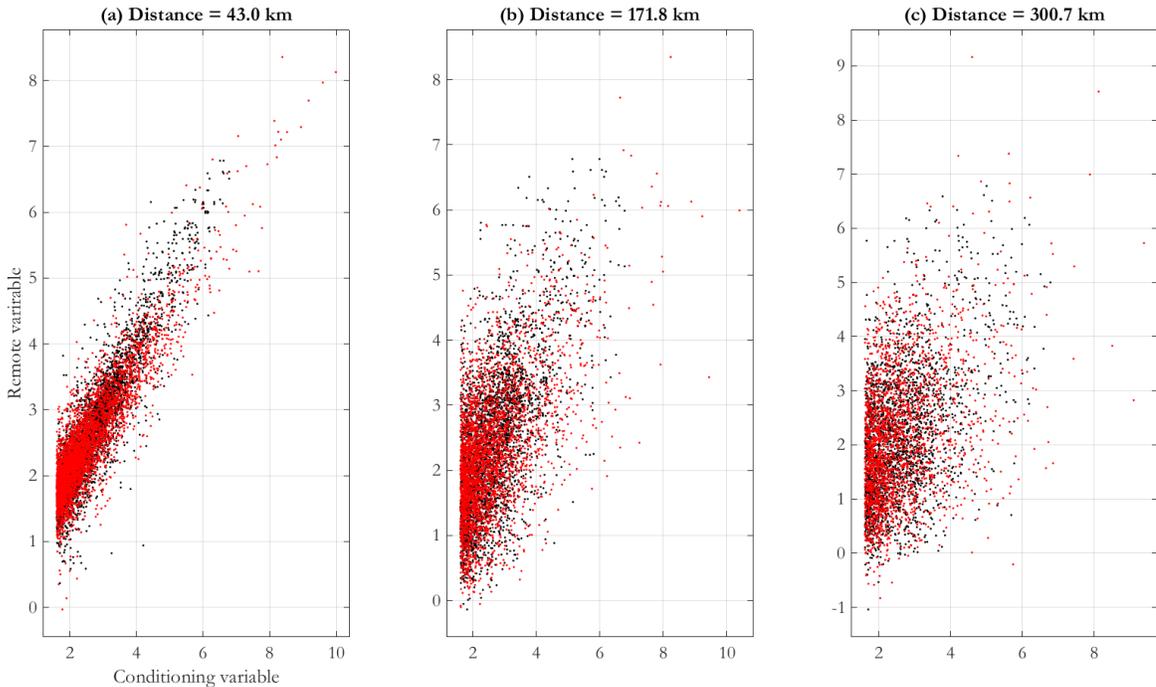


Figure 5.5.2: Scatter plots illustrating dependence between values of Laplace-scale storm peak  $H_S$  at different relative distances for NNS:N-S transects, from (a) original sample and (b) simulation under the fitted free model. Black points are the original data on Laplace scale; red points are data simulated under the fitted model.

### 5.5.2 “Parametric- $\alpha$ ” fit

Figure 5.5.1 suggests an exponential decay of parameter  $\alpha$  with distance  $h$  in the free model. Here, we examine the performance of the SCE model with the functional form for  $\alpha(h)$  given in Equation (5.4.4) and with parameters  $\theta^*$  to estimate (as defined in (5.4.5)). Comparing Figures 5.5.1 and 5.5.4 shows that pseudo-likelihood credible intervals for  $\alpha(h)$  are considerably narrower in the parametric- $\alpha$  model. This is not surprising, since the parametric- $\alpha$  model has a smaller number of parameters. More-

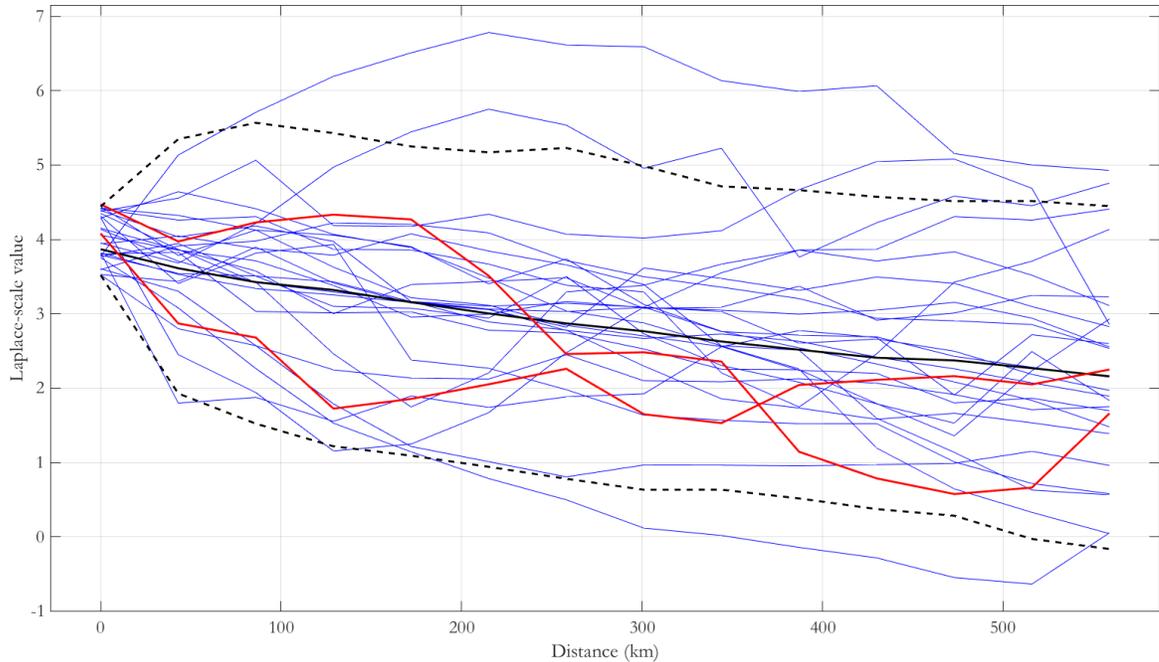


Figure 5.5.3: Observed spatial processes from the NNS:N-S transect with Laplace-scale values at the left-hand location in the interval  $[3.5, 4.5]$ , together with posterior predictive estimates from simulation under the fitted free model, represented using the median (black) and upper and lower limits of a 95% pseudo-likelihood credible interval. Red lines are simulated spatial processes from the fitted model.

over, the parametric decay of  $\alpha$  in the parametric- $\alpha$  model restricts its possible values for any  $h$ . Further, they show that posterior mean estimates for  $\alpha(h)$ ,  $\beta(h)$ ,  $\mu(h)$  and  $\sigma(h)$  are similar under the two models.

The informal discussion above suggests that the quality of fit of free and parametric- $\alpha$  models is similar. To compare these models more formally, we use the DIC introduced in Section 5.4.4. Values for parameter estimates and likelihood from the last  $m=1000$  MCMC iterations are used to estimate the DIC for the two models; the DIC for the free model was calculated to be 27514.22, and for the parametric- $\alpha$  model

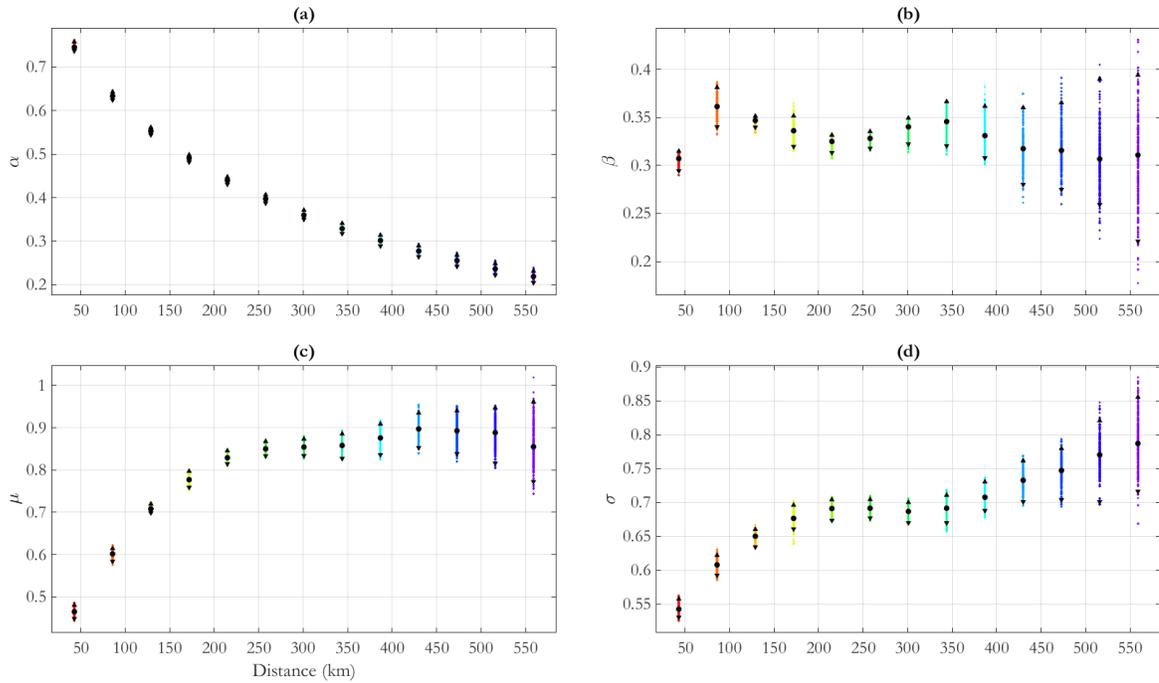


Figure 5.5.4: NNS:N-S transect, parametric- $\alpha$  model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with end-points shown as solid triangles).

27501.68. Since the DIC for the parametric- $\alpha$  model is smaller than for the free model, we infer in this case that the parametric- $\alpha$  model is to be preferred, and that the difference between free and parametric- $\alpha$  fits is small. However, the parametric- $\alpha$  model has the additional advantage that the computational time is decreased due to the smaller number of parameters to estimate in this version of the SCE model.

## 5.6 Application to other North Sea transects

The wave environment in the NNS and CNS is known not to be isotropic (e.g. Feld et al. 2015); we might therefore suspect that the extremal spatial dependence in these neighbourhoods might also be sensitive to transect orientation. Inspection of

Figure 5.2.1 shows that fetches in the NNS are in general longer than in the CNS; further, water depths in the NNS are greater than those in the CNS. It is not unreasonable therefore to anticipate that extremal spatial dependence may be different in different regions of the North Sea. Moreover, for the data considered here, the CNS data are available on a finer grid than for the NNS data, so we may be able to pick out finer-scale features of the dependence structure. Furthermore, the lengths of transects and their spatial resolutions vary, offering the possibility of detecting finer-scale effects (in the CNS) and longer-range effects (for transects with largest distances  $h$ ). This motivates estimating SCE models for the NNS:E-W transect, and the CNS:N-S and CNS:E-W transects.

Below, we start by comparing DIC values for free and parametric- $\alpha$  models. Since it was found that the performance and characteristics of the models were similar for all transects, subsequent discussion of parameter behaviour with  $h$  is restricted to the parametric- $\alpha$  model. As in Section 5.5, all MCMC chains are of length 20000, and we utilise the final 1000 iterations for inference.

### 5.6.1 Comparison of Model Fits for all Transects

We compare DIC values for free and parametric- $\alpha$  model parameterisations to assess in particular whether the parametric- $\alpha$  model is a reasonable general representation for all transects, relative to the free model. Table 5.6.1 gives values for the DIC for each of the transects considered in this work.

From the table, we see that the DIC is lower for the parametric- $\alpha$  model for NNS transects; for the CNS transects, the free model produces lower values for the

Model	DIC	
	Free	Parametric- $\alpha$
NNS:N-S	27514.22	27501.68
NNS:E-W	7360.93	7356.75
CNS:N-S	23471.67	23476.13
CNS:E-W	23809.94	23827.10

Table 5.6.1: Table of DIC values for the free fit model and parametric- $\alpha$  model for all of the transect analyses.

DIC. However, comparing the differences between DIC values per transect with the variability of the corresponding negative log-likelihoods from the MCMC, we see that differences in the DIC are small in each case. We conclude that there is little material difference between free and parametric- $\alpha$  fits for any of the transects.

### 5.6.2 NNS east-west transect

We first apply the parametric- $\alpha$  model to NNS:E-W, coloured magenta in Figure 5.2.1, using a non-exceedance probability of  $\tau = 0.9$  when applying the SCE model, as in Section 5.5. Posterior estimates for model parameters are shown in Figure 5.6.1. This transect has fewer sites available for analysis than NNS:N-S in Section 5.5, and hence fewer data may be pooled together for estimation. Therefore, we would naturally expect model parameter uncertainties to be larger. From the figure it is clear that the pseudo-likelihood credible intervals are wider than for NNS:N-S, at similar  $h$ . The

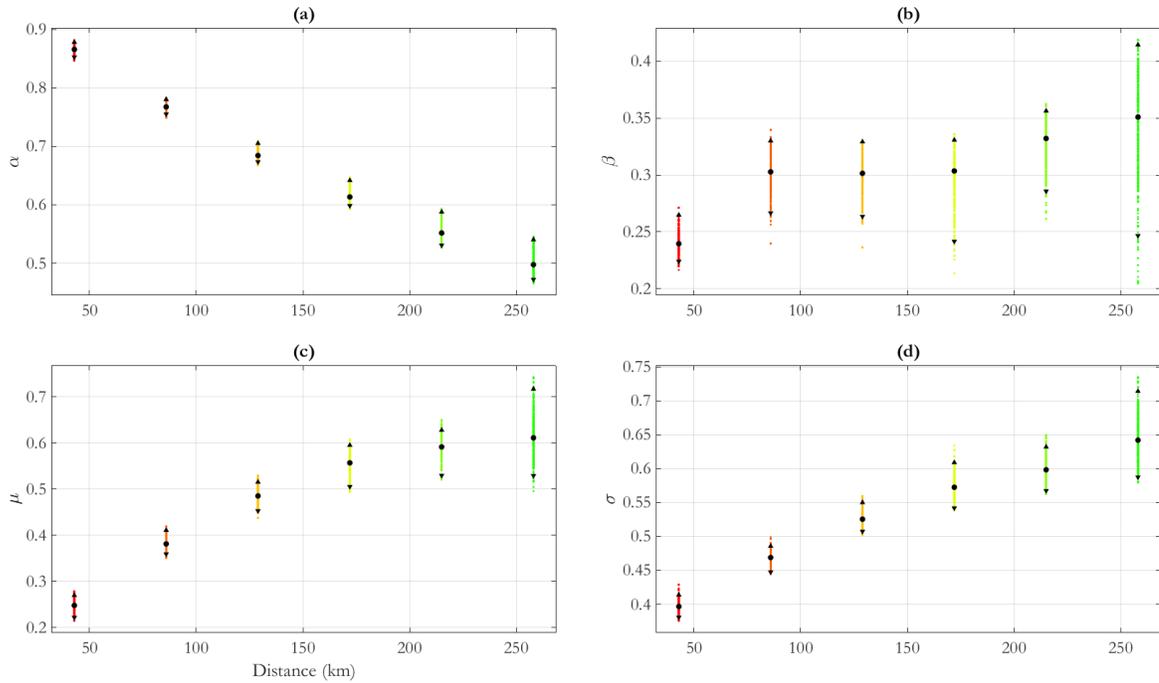


Figure 5.6.1: NNS:E-W transect, parametric  $\alpha(h)$  model: estimates for (a)  $\alpha(h)$ , (b)  $\beta(h)$ , (c)  $\mu(h)$  and (d)  $\sigma(h)$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with end-points shown as solid triangles).

behaviour of parameter estimates for  $\mu$  and  $\sigma$  with  $h$  are similar to those observed for NNS:N-S. However, in NNS:E-W,  $\beta$  increases with distance. The figure also illustrates that estimates for  $\alpha(h)$  on NNS:E-W are larger; in particular,  $\alpha(h \approx 50 \text{ km}) \approx 0.9$ , suggesting that dependence is much higher at short range for NNS:E-W than for NNS:N-S, for which  $\alpha(h \approx 50 \text{ km}) \approx 0.6$ . Further, the rate of decay of  $\alpha$  with  $h$  is smaller for NNS:E-W than for NNS:N-S. These findings are plausible given physical intuition: the largest events in the NNS are Atlantic storms travelling approximately E-W. It is reasonable then to expect that spatial dependence along E-W transects may be higher than for transects with other orientations.

### 5.6.3 CNS transects

For the central North Sea north-south transects (CNS:N-S, coloured dark blue in Figure 5.2.1; and CNS:E-W coloured cyan), the separation  $\Delta$  of locations is smaller than for NNS transects. Furthermore, as more data are available at each site for this ocean basin, we set  $\tau = 0.95$  for the SCE model. Parameter estimates from the parametric- $\alpha$  model are shown in Figure 5.6.2 for CNS:N-S. Compared to NNS

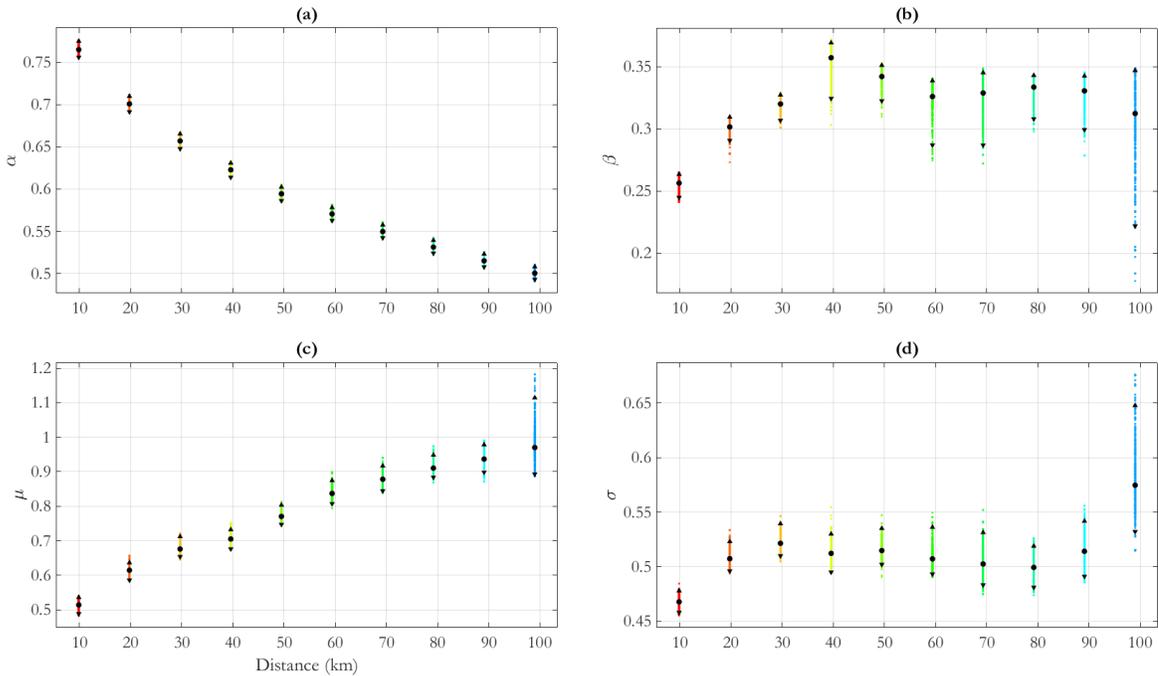


Figure 5.6.2: CNS:N-S transect, parametric- $\alpha$  model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with end-points shown as solid triangles).

transects,  $\alpha$  decreases quickly with  $h$ . At  $h \approx 100$  km the value of  $\alpha$  is approximately 0.5, close to that estimated for the NNS:N-S transect at  $h \approx 150$  km, but at  $h \approx 250$  km for NNS:E-W. The behaviour of  $\mu$  and  $\sigma$  with  $h$  is similar to earlier cases, and  $\beta$  is approximately constant at approximately 0.3, and  $\sigma$  at 0.52. Pseudo-likelihood

credible intervals for estimates increase with  $h$ .

For the CNS:E-W transect, posterior estimates for the SCE parameters are shown in Figure 5.6.3; this transect is slightly longer than the CNS:N-S transect. The pa-

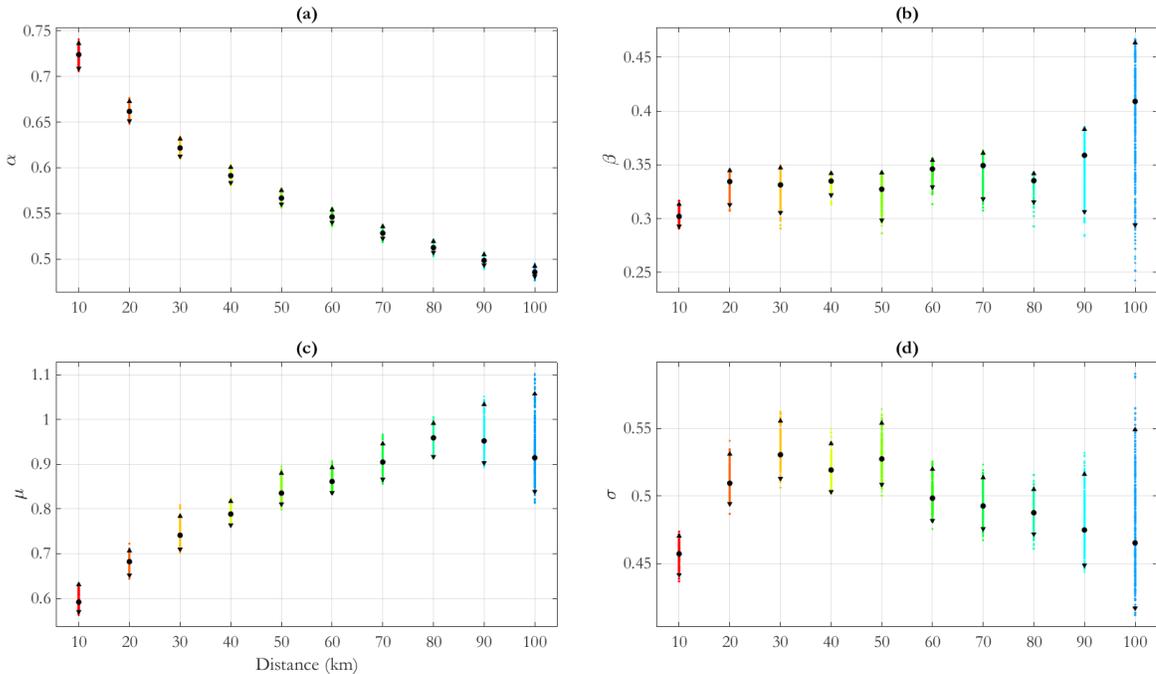


Figure 5.6.3: CNS:E-W transect, parametric- $\alpha$  model: parameter estimates for (a)  $\alpha$ , (b)  $\beta$ , (c)  $\mu$  and (d)  $\sigma$  with distance  $h$ , summarised using posterior means (disk) and 95% pseudo-likelihood credible intervals (with end-points shown as solid triangles).

parameter  $\mu$  increases with  $h$ , and  $\beta$  is approximately constant at approximately 0.33. There is some evidence that  $\sigma(h)$  decreases for  $h > 50$  km. The general behaviour of  $\alpha$  with  $h$  is similar to that for the CNS:N-S transect, with a somewhat slower decay.

We note that the behaviour of  $\mu(h)$  in these analyses, and in the results of Chapter 6, is fairly hard to determine. In general, however, it appears that  $\mu$  increases for a small range of  $h$ , before displaying evidence that the function either reaches a plateau or begins to decrease. The initial rate of increase of  $\mu(h)$  and the distance  $h$  for which

this happens appears to change for each transect; since this behaviour is difficult to determine, this motivates the use of unconstrained linear piecewise forms for this parameter in Chapter 6, from which similar effects are seen.

## 5.7 Discussion and conclusions

In this work, we use a spatial conditional extremes model to investigate the extremal dependence of significant wave height  $H_S$  along straight line transects of different lengths with different spatial orientations and resolutions in the northern and central North Sea. The analyses described in Sections 5.5 and 5.6 suggest that the general nature of extremal dependence is similar for all transects. It appears that the linear dependence parameter  $\alpha$  in the SCE model decays with separation  $h$  of locations, and that this decay is approximately exponential (recalling that  $H_S$  is expressed on standard Laplace scale). The parameter  $\mu$  increases with  $h$ , potentially to a finite asymptote, while the parameter  $\beta$  appears to remain approximately constant as a function of  $h$ . There is some evidence that  $\sigma$  increases initially with  $h$ , but no consistent subsequent behaviour is observed.

Features of the extremal dependence vary by region and transect orientation. For instance, we note that the estimate of  $\rho$ , the residual dependence parameter, for the NNS:N-S transect (with a posterior mode of approximately 0.73 and a 95% pseudo-likelihood credible interval width of approximately 0.06) is different from its value for the other three transects (for which  $\rho$  is estimated to have a mode of approximately 0.5 in each case, and 95% pseudo-likelihood credible intervals of width of ap-

proximately 0.06). Figure 5.7.1(a) illustrates the behaviour of the conditional mean  $\alpha(h)x + \mu(h)x^{\beta(h)}$  from the SCE model for a (Laplace-scale) conditioning value  $x = 5$ , approximately corresponding to the 0.997 quantile. Figure 5.7.1(b) shows the corresponding evolution of the conditional standard deviation  $\sigma(h)x^{\beta(h)}$ . From Figure 5.7.1

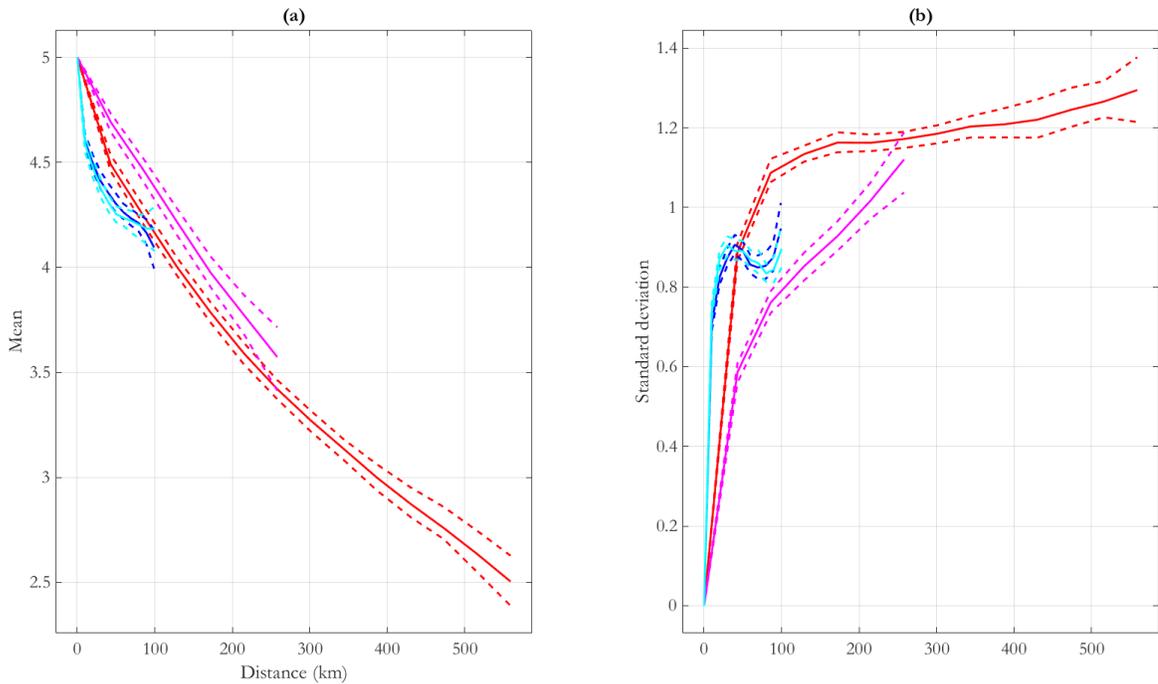


Figure 5.7.1: Pseudo-likelihood credible intervals for (a) the conditional mean and (b) the conditional standard deviation of the fitted dependence model as a function of distance in kilometres, for conditioning Laplace-scale value of 5, and different transects: NNS:N-W (red), NNS:E-W (magenta), CNS:N-S (blue), CNS:E-W (cyan).

it is clear that extremal dependence of  $H_S$  in the NNS is more persistent than in the CNS, and that extremal dependence on the NNS:E-W transect is more persistent than on the NNS:N-S transect (see also Section 5.6.2). That is, longer-range extremal dependence is observed for the E-W transect in the NNS; the same conclusion was drawn by Ross et al. (2017a) in their analysis of related data for the same region, using

one- and two-dimensional max-stable process models. It will be interesting to extend the current SCE model to two-dimensional neighbourhoods of locations, particularly to investigate whether directional differences, related to differences due to transect orientation reported here, are observed.

From an intuitive perspective, we expect the value of SCE parameter  $\alpha$  to decay to zero for large  $h$ , since for large  $h$  the value at the conditioning location should not affect the value at the remote location. For the same reason, we expect  $\beta(h)$  and  $\mu(h)$  to decay to zero, and  $\sigma(h)$  to asymptote to a finite value; see Wadsworth and Tawn (2019) for discussion of the modelling of spatial independence at long range. We plan to examine this by exploring the characteristics of storm peak  $H_S$  on long transects extending over at least 1000 km.

Inspection of Equations (5.3.2) or (5.4.2) readily shows that identification of SCE model parameters is problematic in general, although considerations such as those of Keef et al. (2013a) help restrict the admissible set of parameter values. Imposing an exponential form on the decay of  $\alpha(h)$  with  $h$  was found at least not to be detrimental in the current work. Inspection of the resulting behaviour of parameter estimates in the figures above suggests that further parameterisation of  $\mu(h)$  in particular may be useful.

Understanding the extremal spatial dependence of ocean storms is important for the reliable characterisation of extreme storms and their impact on marine and coastal facilities and habitats. From a statistical perspective, the ocean environment provides a useful test bed for models for spatial extremal dependence over a range of distances. From an offshore engineering perspective, the findings of studies such as the present

work can lead to more informed procedures to accommodate the effects of spatial dependence in engineering design guidelines, since these often require the estimation of events occurring with a given annual probability. The spatial conditional extremes model would seem to offer a relatively straightforward method to help achieve this.

## 5.8 Supplementary Material

This section summarises the constraints of Keef et al. (2013a) used in the conditional extremes model described in this chapter and the MCMC procedure used for parameter estimation.

### The constraints of Keef et al. (2013a)

We also constrain the possible parameter values of  $\alpha(h)$  and  $\beta(h)$ , for  $h > 0$ , as suggested by Keef et al. (2013a). The constraint of interest, for  $\alpha(h)$ ,  $\beta(h)$  and some given  $h$ , in this work is Case 1 of Theorem 1 as given by Keef et al. (2013a); namely that we require either

$$\alpha(h) \leq \min\{1, 1 - \beta(h)z_h(q)v^{\beta(h)-1}, 1 - v^{\beta(h)-1}z_h(q) + v^{-1}z_h^+(q)\}$$

or

$$1 - \beta(h)z_h(q)v^{\beta(h)-1} < \alpha(h) \leq 1, \text{ and}$$

$$(1 - \beta(h)^{-1})\{\beta(h)z_h(q)\}^{1/(1-\beta(h))}(1 - \alpha(h))^{-\beta(h)/(1-\beta(h))} + z_h^+(q) > 0.$$

Here,  $z_h(q)$  is the  $q$ th quantile of the distribution of standardised residuals from the conditional extremes model at distance  $h$  with non-exceedance probability  $q$ . Sim-

ilarly,  $z_h^+(q)$  is the  $q$ th quantile of the distribution of standardised residuals from the conditional extremes model assuming asymptotic positive dependence (i.e., forcing  $\alpha(h) = 1$ ,  $\beta(h) = 0$ ) at distance  $h$  with non-exceedance probability  $q$ . In practice, as suggested by Keef et al. (2013a), it is sufficient to satisfy the constraints above for  $q = 1$  and  $\nu$  equal to the maximum observed value of the conditioning variate.

## MCMC procedure

The MCMC method implemented in Section 5.4.2 is adapted from the method of Roberts and Rosenthal (2009). Suppose the parameters of interest are  $\boldsymbol{\theta} = \{\alpha_k, \beta_k, \mu_k, \sigma_k\}_{k=1}^p \cup \{\rho\}$ , where  $p$  is the number of sampling locations. The total number of parameters is therefore  $n_P = 4p + 1$ . We impose uniform prior distributions for each of these parameters; explicitly,  $\pi(\alpha_k) \sim \text{Unif}(0, 1)$ ,  $\pi(\beta_k) \sim \text{Unif}(0, 1)$ ,  $\pi(\mu_k) \sim \text{Unif}(-2, 2)$  and  $\pi(\sigma_k) \sim \text{Unif}(0, 3)$  for all  $k = 1, \dots, p$ , and  $\pi(\rho) \sim \text{Unif}(0, 1)$ . A total of  $n$  updates of  $\boldsymbol{\theta}$  will be performed.

First we obtain a random starting solution  $\boldsymbol{\theta}^{(0)}$  by sampling the elements of  $\boldsymbol{\theta}$  from their prior distributions, verifying that the starting solution has a valid likelihood (defined in Section 5.4.1).

Writing  $\theta_k^{(i)}$  as the value of the  $k$ th parameter of  $\boldsymbol{\theta}$  at the  $i$ th iteration, we then use an adaptive random walk Metropolis-within-Gibbs scheme for  $n_S$  iterations. That is, for  $i = 2, \dots, n_S$ , where  $n_S < n$ , we update each  $\theta_k^{(i)}$  in turn. If  $i \leq 2n_P$ , we propose a candidate value  $\theta_k^{(i)c}$  from distribution

$$Q_1 = N(\theta_k^{(i-1)}, 0.1^2).$$

For  $i > 2n_P$  (and  $i \leq n_S$ ) we propose  $\theta_k^{(i)c}$  from distribution  $Q_2$  defined by

$$Q_2 = (1 - \beta)N(\theta_k^{(i-1)}, 2.38^2\Sigma_i) + \beta N(\theta_k^{(i-1)}, 0.1^2),$$

where  $\beta = 0.05$ , as proposed by Roberts and Rosenthal (2009), and  $\Sigma_i$  is the empirical covariance of the parameter  $\theta_k$  from the previous  $i$  iterations.

For  $i > n_S$ , we use a grouped adaptive random walk Metropolis-within-Gibbs scheme, updating quartets  $\theta_{G_k}^{(i)} = (\alpha_k^{(i)}, \beta_k^{(i)}, \mu_k^{(i)}, \sigma_k^{(i)})$  jointly, before updating  $\rho$  independently as before. If  $i \leq n_S + 2n_P$ , and the quartet state is  $\theta_{G_k}^{(i)}$ , we propose candidates  $\theta_{G_k}^{(i)c}$  from distribution

$$Q_3 = MVN(\theta_{G_k}^{(i-1)}, (0.1^2)/4).$$

If  $i > n_S + 2n_P$ , we propose  $\theta_{G_k}^{(i)c}$  from distribution

$$Q_4 = (1 - \beta)MVN(\theta_{G_k}^{(i-1)}, 2.38^2\Sigma_i) + \beta MVN(\theta_{G_k}^{(i-1)}, 0.1^2/4),$$

where again  $\beta = 0.05$  and  $\Sigma_i$  is the empirical variance-covariance matrix of the parameters  $\theta_{G_k}^{(i)}$  from the previous  $i$  iterations. Finally we update  $\rho$ .

Throughout, a candidate state is accepted using the standard Metropolis-Hastings acceptance criterion. Since prior distributions for parameters are uniform, and proposals symmetric, this is effectively just a likelihood ratio. That is, we accept the candidate state with probability  $\min(1, L^c/L)$ , where  $L$  and  $L^c$  are the likelihoods evaluated at the current and candidate states respectively, with candidates lying outside their prior domains rejected.

# Chapter 6

## Basin-wide spatial conditional extremes for severe ocean storms

### 6.1 Introduction

Many models for spatial extremes require that the type of extremal dependence exhibited (e.g., asymptotic dependence or asymptotic independence, described for example in Coles et al., 1999) must be decided beforehand. Here, we describe a conditional spatial extremes model able to capture both types of asymptotic behaviour with no prior information required. The conditional spatial extremes framework is a useful tool in assessing the risk involved in the construction of coastal and offshore structures, enabling correct assessment of extremal dependence and providing better estimation of the joint risk of potentially damaging extreme events occurring from ocean storms than currently-used methods.

The current study involves the characterisation of extremal spatial dependence of

extreme ocean storm severity, quantified for a storm event using storm peak significant wave height ( $H_S$ ). For two sampling locations a short distance apart (relative to the size of a storm), we may expect that an extreme value of  $H_S$  may arise at each location from the same storm event, characteristic of asymptotic dependence (AD). If two sites are far apart, it is unlikely that extremes occurring at the two locations would be simultaneously large; corresponding to asymptotic independence (AI). Previous studies (Kereszturi et al., 2016 and Ross et al., 2017a) have shown that the nature and extent of extremal dependence in an ocean basin changes with distance between locations.

The traditional approach to spatial extremes has been to consider max-stable processes (MSPs); see Brown and Resnick (1977), Smith (1990) and Schlather (2002) for details on how to apply these models. Crucially, MSP models typically make the assumption that the spatial process is asymptotically dependent, and hence such models may be inappropriate for modelling  $H_S$  over an ocean basin. We note that other AD spatial extremes models have been proposed, such as Reich and Shaby (2012) and Ferreira and de Haan (2014); the processes described in the latter of these are termed Pareto processes. There have been multiple models proposed in the recent spatial extremes literature which are able to model either class of extremal dependence, e.g. Wadsworth and Tawn (2012b), Wadsworth et al. (2017) and Huser and Wadsworth (2018). However, these models suffer from drawbacks; either the fitted model must assume a certain type of extremal dependence across the entire spatial domain in which it is fitted, as is the case for the Huser and Wadsworth (2018) model for example, or the model is rather computationally challenging to

fit, such as the Wadsworth and Tawn (2012b) approach. The conditional spatial extremes model we present here overcomes these issues. Careful parameterisations of distance effects enables the spatial extremes problem to be well-described in terms of a relatively small number of parameters, even when the number of measurement locations is high. This greatly reduces the computational burden when fitting across hundreds of sampling locations compared to broadly equivalent MSPs. Moreover, we incorporate both types of extremal dependence into our model parsimoniously without prior specification, with the dependence class changing with distance in the spatial domain. More detailed overviews of spatial extremes modelling may be found in Davison et al. (2012), Ribatet (2013) and Tawn et al. (2018), for example.

Our model builds upon the work of Wadsworth and Tawn (2019) and Shooter et al. (2019), adopting the concept of having known functional forms for some of the spatial conditional extremes parameters. It was found in Shooter et al. (2019) that assuming a parametric form, requiring only two parameters, for the slope parameter,  $\alpha$  in the spatial conditional extremes model was adequate to capture the behaviour of the parameter as a function of distance between sites, whilst having the significant benefit of reducing computational time, since the  $\alpha$  parameter otherwise needs to be fitted separately at each modelling location. We build upon this idea by imposing either parametric or piecewise-linear forms for more model parameters, motivated by theoretical considerations and evidence from preliminary analysis. The resulting reduction in model complexity for given sample size allows us to consider analyses incorporating many sampling locations. Such analysis would have been computationally demanding in the MSP framework, as well as biased due to the assumption of

asymptotic dependence. The current work includes other novel features, including the adoption of a generalised Gaussian, or delta-Laplace distribution, to describe the marginal distribution of model residuals coupled with a spatial Gaussian copula. In our application to a two-dimensional grid of locations in the northern North Sea, we also account for possible anisotropy in the spatial domain given previous evidence of this from other studies (Kereszturi et al., 2016; Ross et al., 2017a; Shooter et al., 2019).

The article is presented as follows. In Section 6.2, we discuss the conditional extremes model of Heffernan and Tawn (2004) and its spatial extension as proposed by Wadsworth and Tawn (2019). Section 6.3 then summarises the inferential scheme used for parameter estimation. Section 6.4 outlines the performance of our model in application to a long-distance west-east zonal transect in the north Atlantic, and to a two-dimensional spatial neighbourhood of locations in the northern North Sea. Discussion and conclusions are given in Section 6.5.

## 6.2 Conditional extremes

### 6.2.1 Extremal dependence

A key issue in modelling spatial extremes is assessing the nature of dependence between extreme events; that is, if we observe an extreme event, we are interested in the information provided by this event about the probability of observing further simultaneous extreme events. We naturally expect that over short distances, it is quite likely that an extreme event being observed at one location may be related to an extreme

observation at another. On the other hand, extremes observed at distant locations are likely to be independent of one another. To quantify these effects, measures of extremal dependence are utilised.

To describe extremal dependence, Coles et al. (1999) introduce the measures  $\chi$  and  $\bar{\chi}$ , most easily calculated through their sub-asymptotic forms  $\chi(u)$  and  $\bar{\chi}(u)$ , where  $u \in [0, 1]$ . For bivariate Uniform random variables  $(U, V)$ , which may be obtained by applying the probability integral transform, these are defined as

$$\chi(u) = 2 - \frac{\log \mathbb{P}(U < u, V < v)}{\log \mathbb{P}(U < u)} \quad \text{and} \quad \bar{\chi}(u) = \frac{2 \log \mathbb{P}(U > u)}{\log \mathbb{P}(U > u, V > v)} - 1.$$

Then  $\chi$  and  $\bar{\chi}$  may be obtained by taking the respective limits of these functions, as  $u \rightarrow 1$ . The nature of extremal dependence between  $U$  and  $V$  may then be described by considering  $\chi$  and  $\bar{\chi}$  together. If  $\chi = 0$ , then if  $-1 \leq \bar{\chi} < 1$ , the random variables are asymptotically independent, and the value of  $\bar{\chi}$  signifies the level of dependence. On the other hand, if  $\bar{\chi} = 1$  and  $0 < \chi \leq 1$ , then the pair  $(U, V)$  exhibit asymptotic dependence, with  $\chi$  providing a measure of this. For a full description of extremal dependence types, we refer the reader to Ledford and Tawn (1996) and Coles et al. (1999). Spatial extensions of these measures are discussed in Tawn et al. (2018).

### 6.2.2 Multivariate conditional extremes models

Suppose that we have a vector of random variables  $(X_0, \mathbf{X})$ , where  $X_0$  and  $\mathbf{X} = (X_1, \dots, X_p)$  have Gumbel marginal distributions, again obtainable through the probability integral transform. Then Heffernan and Tawn (2004) assume that there exist

functions  $\mathbf{a} : \mathbb{R} \rightarrow \mathbb{R}^p$ ,  $\mathbf{b} : \mathbb{R} \rightarrow \mathbb{R}^p$  such that, defining

$$\mathbf{Z} = \frac{\mathbf{X} - \mathbf{a}(X_0)}{\mathbf{b}(X_0)},$$

where all operations are taken to be componentwise, then we have that, for  $x > 0$ ,

$$\lim_{u \rightarrow \infty} \mathbb{P}(\mathbf{Z} \leq \mathbf{z}, X_0 - u > x | X_0 > u) = G(\mathbf{z}) \exp(-x), \quad (6.2.1)$$

where  $G$  is a joint distribution which has non-degenerate margins. This form for the conditional extremes model is asymptotically justified; see Heffernan and Tawn (2004); Heffernan and Resnick (2007) for details.

Keef et al. (2013a) show that if the margins of  $\mathbf{X}$  are instead assumed to be Laplace-distributed (obtainable through using the probability integral transform), then canonical functional forms for  $\mathbf{a}(\cdot)$  and  $\mathbf{b}(\cdot)$  are  $\mathbf{a}(x) = \boldsymbol{\alpha}x$  and  $\mathbf{b}(x) = x^\boldsymbol{\beta}$  (for  $x > 0$ ), where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ . In this representation, each  $\alpha_i \in [-1, 1]$  and  $\beta_i \in (-\infty, 1]$ . Different values for these parameters correspond to different classes of extremal dependence; this is discussed in the spatial case below, but we note that we shall assume positive dependence in this work, and thus restrict  $\alpha_i \in [0, 1], \beta_i \in [0, 1]$  for all  $i \in \{1, \dots, p\}$ . Then, choosing some suitably high threshold  $u$ , we have that for all  $x_0 > u$ , the conditional extremes model may be represented as

$$\mathbf{X} | \{X_0 = x_0\} = \boldsymbol{\alpha}x_0 + x_0^\boldsymbol{\beta}\mathbf{Z}, \quad (6.2.2)$$

where  $\mathbf{Z}$  is independent of  $X_0$ , following from (6.2.1).

### 6.2.3 Spatial conditional extremes

We may extend the model described in (6.2.2) to a spatial context, as described by Tawn et al. (2018) and Wadsworth and Tawn (2019). Suppose that we have a stationary and isotropic spatial process  $X(\cdot)$  over some spatial domain  $\mathcal{S}$ , which has Laplace marginal distributions, and that we have sampling locations  $s, s_0 \in \mathcal{S}$ . Then letting  $d = \|s - s_0\|$  and assuming positive dependence between variables, we have in general that, for all  $x_0 > u$ ,

$$X(s) \mid \{X(s_0) = x_0\} = \alpha(d)x_0 + x_0^{\beta(d)}Z(s - s_0), \quad (6.2.3)$$

where  $\alpha : \mathbb{R}^+ \rightarrow [0, 1]$ ,  $\beta : \mathbb{R}^+ \rightarrow [0, 1]$  and  $Z(\cdot)$  is a residual process independent of  $X(\cdot)$ . For inference purposes, if we have spatial data  $(X_0, \mathbf{X})$  with  $\mathbf{X} = (X_1, X_2, \dots, X_p)$ , observed at sampling locations  $s_0, s_1, \dots, s_p$ , we let  $d_j = \|s_j - s_0\|$  for  $j = 1, 2, \dots, p$  and then set  $\alpha_j = \alpha(d_j)$  and  $\beta_j = \beta(d_j)$ .

We note that, in particular, we require particular conditions on the residual process  $Z(\cdot)$ . To this end, we follow Wadsworth and Tawn (2019) and suppose that the process  $Z(\cdot)$  has delta-Laplace margins with parameters  $\delta, \sigma, \mu$  also dependent on  $d$ . That is, adopting similar notation as for  $\alpha_j = \alpha(d_j)$ , we have

$$f_{Z_j}(z_j) = \frac{\delta}{2\kappa_j\sigma_j\Gamma\left(\frac{1}{\delta_j}\right)} \exp\left\{-\left|\frac{z - \mu_j}{\kappa_j\sigma_j}\right|^{\delta_j}\right\},$$

for  $j = 1, 2, \dots, p$ ,  $\delta_j, \sigma_j, \kappa_j \in \mathbb{R}^+$ ,  $\mu_j \in \mathbb{R}$ ,  $z_j \in \mathbb{R}$  and where  $\kappa_j^2 = \frac{\Gamma\left(\frac{1}{\delta_j}\right)}{\Gamma\left(\frac{3}{\delta_j}\right)}$ . Here,  $\Gamma(\cdot)$  represents the gamma function. The purpose of the parameter  $\kappa_j$  is to assist with parameter identifiability, and note that the mean and variance of this distribution are respectively  $\mu_j$  and  $\sigma_j^2$ , regardless of the choice of  $\delta_j$ . We denote this distribution as

$\text{DL}(\mu_j, \sigma_j^2, \delta_j)$ . With this notation, upon setting  $\delta_j = 2$ , we have a Gaussian density function, whereas setting  $\delta_j = 1$  leads to the density of a Laplace distribution. The standard Laplace distribution, with variance 2, corresponds to the case  $\sigma_j^2 = 2$  in our notation.

Of particular importance is the requirement that as  $d \rightarrow \infty$ , we approach perfect independence between  $X(s)$  and  $X(s_0)$ . Inspection of (6.2.3) suggests that we should simply be left with random Laplace random variables; i.e.,  $\lim_{d \rightarrow \infty} \delta(d) = 1$ ,  $\lim_{d \rightarrow \infty} \alpha(d) = \lim_{d \rightarrow \infty} \beta(d) = \lim_{d \rightarrow \infty} \mu(d) = 0$ ,  $\lim_{d \rightarrow \infty} \sigma(d) \rightarrow \sqrt{2}$ .

Consider a vector of random variables  $\mathbf{X} = (X_0, \dots, X_p)$ , corresponding to  $p + 1$  spatial locations, with standard Laplace marginal distributions, i.e.,  $X_j \sim \text{DL}(0, 2, 1)$  for  $j = 0, \dots, p$ . We then assume, conditional on  $X_0 = x_0$ , for  $x_0 > u$ , that  $\mathbf{X}$  follows a multivariate extension of the delta-Laplace distribution,

$$(X_1, \dots, X_p) | \{X_0 = x_0\} = \boldsymbol{\alpha}x_0 + x_0^{\beta} \mathbf{Z},$$

where  $\mathbf{Z} \sim \text{DL}_p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\delta}; \boldsymbol{\Sigma})$ , with  $\boldsymbol{\Sigma}$  representing a Gaussian copula dependence structure via a correlation matrix between residual components, such that

$$F_{\mathbf{Z}}(\mathbf{z}) = \Phi_p \left( \Phi^{-1}(F_{Z_1}(z_1)), \Phi^{-1}(F_{Z_2}(z_2)), \dots, \Phi^{-1}(F_{Z_p}(z_p)); \mathbf{0}, \boldsymbol{\Sigma} \right), \quad (6.2.4)$$

where  $F_Y$  represents the cumulative distribution function of  $Y$ , and  $\Phi$  is the cumulative distribution function of a standard Gaussian distribution. Note that the  $(j, k)$ th element of  $\boldsymbol{\Sigma}$  is denoted by  $[\boldsymbol{\Sigma}]_{j,k}$ . Hence, marginally,  $Z_j \sim \text{DL}(\mu_j, \sigma_j^2, \delta_j)$ , so that  $Z_j = Z_j^{(DL)} \sigma_j + \mu_j$ , where  $Z_j^{(DL)} \sim \text{DL}(0, 1, \delta_j)$ . Writing  $X_j^c$  to represent  $X_j | \{X_0 = x_0\}$ , we have  $X_j^c = \alpha_j x_0 + x_0^{\beta_j} Z_j \sim \text{DL}(m_j, s_j^2, \delta_j)$ , where  $m_j = \alpha_j x_0 + x_0^{\beta_j} \mu_j$  and

$s_j = x_0^{\beta_j} \sigma_j$ . Hence,

$$\mathbf{X}^c = (\mathbf{X}|\{X_0 = x_0\}) \sim \text{DL}_p(\mathbf{m}, \mathbf{s}^2, \boldsymbol{\delta}; \boldsymbol{\Sigma}), \quad (6.2.5)$$

where  $\mathbf{m} = (m_1, m_2, \dots, m_p)$ ,  $\mathbf{s} = (s_1, s_2, \dots, s_p)$  and  $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_p)$ .

Model (6.2.5) is able to describe different types of extremal dependence, inferred from the values of parameters  $(\alpha_j, \beta_j)$  for  $j \in \{1, \dots, p\}$ . If  $(\alpha_j, \beta_j) = (1, 0)$ , then the random variables  $X_0$  and  $X_j$  are asymptotically dependent, whereas if  $\alpha_j < 1$ , the random variables exhibit asymptotic independence. Further discussion can be found in Tawn et al. (2018).

#### 6.2.4 Model parameter variation with distance

The  $p+1$  measurement locations are assumed to have coordinates  $\mathbf{r}_j$  for  $j = 0, 1, \dots, p$ . Parameters  $\alpha, \beta, \mu, \sigma, \delta$  are assumed to be continuous functions of the distance between a remote location ( $j = 1, 2, \dots, p$ ) and the conditioning location ( $j = 0$ ). Thus, for example,  $\alpha_j = \alpha(d(\mathbf{r}_j, \mathbf{r}_0))$ , where  $d(\mathbf{r}, \mathbf{r}')$  is a measure of the distance between locations  $\mathbf{r}$  and  $\mathbf{r}'$ . In addition, we assume that  $[\boldsymbol{\Sigma}]_{jk} = \rho^{d(\mathbf{r}_j, \mathbf{r}_k)}$ , for some  $\rho \in [0, 1]$ .

Inspection of Equation (6.2.5) shows that, at zero distance, we must have  $\alpha(0) = 1$ ,  $\beta(0) = 0$ ,  $\mu(0) = 0$ , and  $\sigma(0) = 0$ ; since the distribution at zero distance will be a point mass, we cannot quantify  $\delta(0)$ . Furthermore, the discussion in Section 6.2.3 suggests that we require specific behaviour for large distances. We therefore adopt parametric forms for the variation of  $\alpha, \beta$  and  $\sigma$  with distance, and piecewise linear forms for the variation of  $\mu$  and  $\delta$ . Details are provided in Section 6.3.2.

We also anticipate a degree of anisotropy in the variation of model parameters

between locations. For this reason, we also parameterise the distance function  $d$  such that  $d(\mathbf{r}, \mathbf{r}') \equiv d(\mathbf{r}, \mathbf{r}'; \boldsymbol{\theta})$ , for parameters  $\boldsymbol{\theta}$  to be estimated; details are provided in Section 6.3.2.

## 6.3 Inference

### 6.3.1 Likelihood and MCMC

By (6.2.4) and accounting for the Jacobian, the joint density  $f_{\mathbf{X}^c}(\mathbf{x}) = \mathbb{P}(\mathbf{X} = \mathbf{x} | X_0 = x_0)$ , with marginal distributions  $f_{X_j^c}$  can be written

$$f_{\mathbf{X}^c}(\mathbf{x}) = \phi_p(w_1, w_2, \dots, w_p; \mathbf{0}, \boldsymbol{\Sigma}) \prod_{j=1}^p \frac{f_{X_j^c}(x_j)}{\phi(w_j)}, \quad (6.3.1)$$

where  $\phi$  denotes the standard Gaussian density,  $\phi_p$  represents the  $p$ -dimensional Gaussian density with given mean vector and correlation matrix, and  $w_j = \Phi^{-1}\{F_{X_j^c}(x_j)\}$  with  $x_j \in \mathbb{R}$  for  $j = 1, 2, \dots, p$ . From (6.3.1) above, the negative log-density is given by

$$\begin{aligned} \ell(\mathbf{x}; \boldsymbol{\theta}) &= -\log\{f_{\mathbf{X}^c}(\mathbf{x})\} \\ &= -\log \phi_p(\mathbf{w}; \mathbf{0}, \boldsymbol{\Sigma}) - \sum_{j=1}^p \log f_{X_j^c}(x_j) + \sum_{j=1}^p \log \phi(w_j), \end{aligned}$$

where  $\mathbf{w}' = (w_1, w_2, \dots, w_p)$ .

Hence, the sample negative log-likelihood given a sample  $\{x_{ij}\}_{i=1, j=0}^{n,p}$  is

$$\begin{aligned} \ell &= \frac{np}{2} \log(2\pi) + \frac{n}{2} \log |\boldsymbol{\Sigma}| + \sum_{j=1}^p \log \left( 2s_j \kappa \Gamma \left[ \frac{1}{\delta_j} \right] \right) - n \sum_{j=1}^p \log \delta_j \\ &+ \sum_{i=1}^n \left\{ \mathbf{w}'_i \boldsymbol{\Sigma}^{-1} \mathbf{w}_i + \sum_{j=1}^p \left| \frac{x_{ij} - m_j}{\kappa_j s_j} \right|^{\delta_j} + \sum_{j=1}^p \log \phi(w_{ij}) \right\}, \end{aligned} \quad (6.3.2)$$

where  $\kappa_j^2 = \Gamma(1/\delta_j)/\Gamma(3/\delta_j)$  for each  $j$  and  $\mathbf{w}'_i = (w_{i1}, w_{i2}, \dots, w_{ip})$ .

For parameter estimation, we use the negative log-likelihood (6.3.2) in an adaptive MCMC algorithm similar to that proposed by Roberts and Rosenthal (2009), a variant of which was used in Shooter et al. (2019). Furthermore, as we assume positive dependence, to ensure we obtain consistent parameter estimates, we use the conditional extremes model constraints as detailed by Keef et al. (2013a); details of how the constraints are applied in a spatial setting are provided in Section 5.8. Uniform prior distributions shall be adopted for model parameters; these are chosen so as to permit only sensible ranges for the parameters, whilst being uninformative within these ranges.

### 6.3.2 Parametric forms for $\alpha$ , $\beta$ and $\sigma$

Previous studies (see Shooter et al., 2019) and further investigatory work have shown that it appears reasonable to assume certain parametric forms for some of the parameters in model (6.2.5), namely  $\alpha, \beta$  and  $\sigma$ . Furthermore, as detailed in Section 6.2.4, we expect certain behaviours of these parameters which assists in our choice of parametric forms. Letting  $d$  be an arbitrary distance calculated using the anisotropic distance measure  $d(\cdot, \cdot)$  from Section 6.2.4, whilst noting an isotropic distance measure could also be used, we thus propose the following parametric forms for  $\alpha(d), \beta(d)$  and

$\sigma(d)$  for  $d > 0$ :

$$\begin{aligned}\alpha(d) &= \exp \left\{ - \left( \frac{d}{K_{A1}} \right)^{K_{A2}} \right\} && (K_{A1}, K_{A2} > 0); \\ \beta(d) &= \frac{K_{B1} d^{K_{B2}} \exp \left( - \frac{d}{K_{B3}} \right)}{\max_{d>0} \left\{ d^{K_{B2}} \exp \left( - \frac{d}{K_{B3}} \right) \right\}} && (0 < K_{B1} < 1; K_{B2}, K_{B3} > 0); \\ \sigma(d) &= \sqrt{2} \left\{ 1 - \exp \left[ - \left( \frac{d}{K_{S1}} \right)^{K_{S2}} \right] \right\} && (K_{S1}, K_{S2} > 0).\end{aligned}$$

Using these parametric forms, the full behaviour of  $\alpha, \beta$  and  $\sigma$  with distance can be inferred by estimating a small number of  $K$  parameters, even when the number of locations,  $p$ , involved is large. This improves the computational tractability of the inference considerably. Taking  $\alpha(d)$  as an example, rather than estimating  $\alpha_j$  separately for each remote location (i.e.,  $p$  times), we simply need to estimate  $K_{A1}$  and  $K_{A2}$  using data across all locations. Hence, we reduce the number of parameters to estimate for  $\alpha, \beta$  and  $\sigma$  over all locations from  $3p$  to just 7, whilst trying to ensure that inference in both cases is similar. We evaluate these functions at each location  $s_1, s_2, \dots, s_p$ , as outlined in Sections 6.2.3 and 6.2.4.

Our parameterisation of  $\alpha(d)$  does not admit asymptotic dependence, since  $\alpha(d) \neq 1$  for  $d > 0$ . A possible parametric form for  $\alpha(d)$  able to capture such dependence is described by Wadsworth and Tawn (2019), although for suitable large  $K_{A1}$  and  $K_{A2} \geq 2$ , then  $\alpha(d) \approx 1$  for  $d \approx 0$ . However, values of  $\alpha$  near unity for non-zero values of  $d$  were never observed during our analysis; the smallest distance we shall consider is  $d \approx 40\text{km}$ . Finer-scale grids of locations were not used in these analyses but could be used in further studies. We also note that our form for  $\sigma(d)$  does not permit observation of a nugget effect, but we do not see any evidence that this impacts

the accuracy of our inference.

### 6.3.3 Parameterising $\mu$ and $\delta$ with distance

The constraints discussed in Section 6.2.4 and previous studies (see Shooter et al. (2019)) suggest that the parametric forms outlined in Section 6.3.2 are appropriate to describe parameter behaviour with distance  $d$  for  $\alpha$ ,  $\beta$  and  $\sigma$ . We are less sure about the behaviour of  $\mu$  and  $\delta$  with  $d$ . For this reason, we choose to specify this behaviour non-parametrically, in terms of piecewise linear representations  $\mu(d)$  and  $\delta(d)$ . The specification of  $\delta(d)$  is analogous to that of  $\mu(d)$ , described next. We specify a set of  $n_d$  equally-spaced distances  $d_k$ ,  $k = 1, 2, \dots, n_d$ , covering the domain, with corresponding values  $\mu_k$  for  $\mu$ , such that  $\mu_k = \mu(d_k)$ . Then, for an arbitrary distance  $d \in (d_1, d_{n_d}]$ , we define

$$\mu(d) = \frac{(d - d_{k^*})\mu_{k^*} + (d_{k^*+1} - d)\mu_{k^*+1}}{(d_{k^*+1} - d_{k^*})}, \quad (6.3.3)$$

where  $k^* = \underset{k}{\operatorname{argmax}} \{d_k < d\}$ . Parameter estimates for  $\mu_k$ ,  $\delta_k$ ,  $k = 1, 2, \dots, n_d$ , are sought during inference. For  $d > d_{n_d}$ , we note that we are unable to model the function using these piecewise forms, and analysis may not show the functions  $\mu(d)$  and  $\delta(d)$  attaining their expected limits. We note that this form is similar in structure to a linear spline (see de Boor, 1978 for details of this).

### 6.3.4 Incorporating anisotropy

Previous work (for example, Shooter et al., 2019) has shown that the spatial extremal dependence of storm severity exhibits some anisotropy. To investigate this further

in the current work, we choose to represent the distance between two locations with coordinates  $\mathbf{r}$  and  $\mathbf{r}'$  as

$$d(\mathbf{r}, \mathbf{r}'; \nu_1, \nu_2) = \left( (\mathbf{r}' - \mathbf{r}) \mathbf{S}^{-1} (\mathbf{r}' - \mathbf{r})' \right)^{1/2} \quad (6.3.4)$$

where  $\mathbf{S} = \begin{bmatrix} 1 & \nu_2 \\ \nu_2 & \nu_1 \end{bmatrix}$  with  $\nu_1 > 0$  and  $\nu_1 - \nu_2^2 > 0$ ; the parameters  $\nu_1$  and  $\nu_2$  are to be estimated. Isotropy corresponds to the case  $\nu_1 = 1, \nu_2 = 0$ .

Locations of points on the surface of the Earth are typically specified in terms of longitude-latitude coordinates. Temporarily adopting oceanographic notation, the shortest distance (e.g. in metres) on the surface of a spherical Earth between locations with longitude-latitude coordinates  $(\lambda, \varphi)$  and  $(\lambda', \varphi')$  can be calculated using the spherical law of cosines. In the current work (see Equation (6.3.4)) we characterise spatial anisotropy in terms of a quadratic form in the Cartesian  $x$  and  $y$  components of displacement between locations. It is therefore convenient to adopt a local Cartesian description of displacement on the surface of a sphere, following e.g. Vallis (2017). For locations with longitude-latitude coordinates  $(\lambda, \varphi)$  and  $(\lambda', \varphi')$ , we locate the local origin of coordinates at  $(\bar{\lambda}, \bar{\varphi}) = ((\lambda + \lambda')/2, (\varphi + \varphi')/2)$ , with  $x$  axis running West-East (in the Northern Hemisphere) and  $y$  axis poleward. Then, to a good approximation when  $|\lambda' - \lambda|$  and  $|\varphi' - \varphi|$  are small, the local Cartesian displacement between the points can be written

$$\mathbf{r}' - \mathbf{r} = (a \cos(\bar{\varphi})(\lambda' - \lambda), a(\varphi' - \varphi)), \quad (6.3.5)$$

where  $a$  is the radius of the spherical Earth. We adopt this model to estimate the distance between all pairs of locations in this work.

## 6.4 Applications

### 6.4.1 Significant wave height data

We apply our spatial conditional extremes model to two data sets: one corresponding to a long west-east zonal transect, and the other from a spatial grid in the northern North Sea. The first sample is comprised of hindcast values of storm peak significant wave height at a total of 274 locations on a west-east transect of near-constant latitude of approximately  $63^\circ\text{N}$  passing to the south of Iceland and the north of the Faroe Islands, for longitudes from approximately  $-25^\circ$  to  $+5^\circ$ , extending from west of Iceland to the Norwegian coast, a map of this is depicted in Figure 6.4.1. The data were taken from the NORA10 hindcast (Breivik et al., 2013). At each location, marginal directional-seasonal extreme value analysis of storm peak values was performed as described in Shooter et al. (2019), and the storm peak data subsequently transformed to standard Laplace marginal scale. A subset of 40 approximately equally-spaced locations was selected for the spatial conditional extremes analysis reported here. Since the locations lie on a line of constant latitude, complications of spherical trigonometry do not arise, and we are free to use differences between longitudes of locations as a measure of distance. We choose the conditioning point to be the most westerly point on the transect, marked red in Figure 6.4.1.

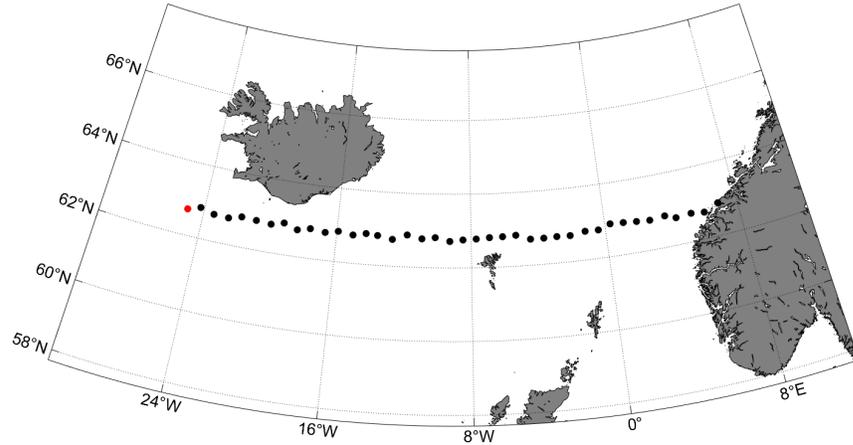


Figure 6.4.1: Map of the conditioning site (coloured red) and remote sites (black) for the North Atlantic long transect analysis.

The second sample corresponds to hindcast values of storm peak significant wave height at a total of 150 locations in a spatial neighbourhood of the northern North Sea, between the UK and Norway, previously reported by Ross et al. (2017a) and Shooter et al. (2019). The data were taken from the NEXTRA hindcast (Oceanweather, 2002). At each location, marginal directional-seasonal extreme value analysis of storm peak values was performed as described in Shooter et al. (2019), and the storm peak data transformed to standard Laplace marginal scale. A subset of 40 approximately equally-spaced locations was selected for the spatial conditional extremes analysis reported here, as illustrated in Figure 6.4.2.

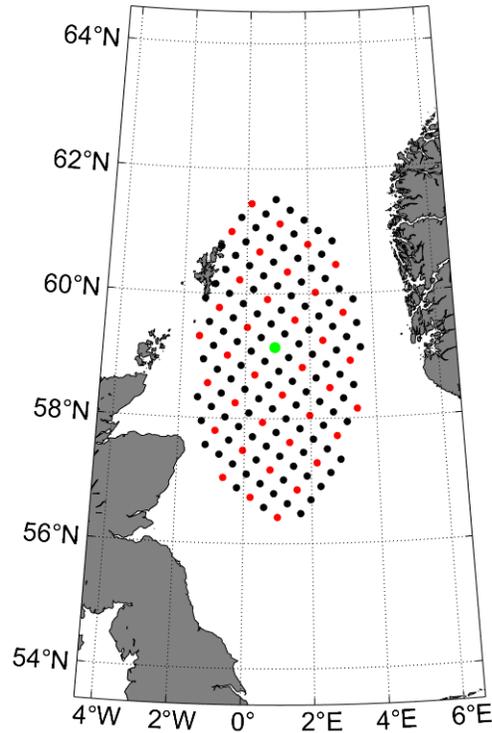


Figure 6.4.2: Map of the conditioning site (coloured green) and remote sites (red) for the North Sea ocean basin. Sites not used in this analysis are coloured black.

## 6.4.2 Results

### North Atlantic zonal transect

The MCMC algorithm, using the anisotropic version of our model, outlined in Section 6.3.1 was executed for 50000 iterations, with convergence occurring at approximately 30000 iterations. Results displayed here correspond to the final 1000 iterations, although we note that the MCMC chain was judged to have converged before this point.

Marginal summaries of posterior distributions of parameters with distance are shown in Figure 6.4.3. The solid black line indicates the posterior median, with the 95% credible interval shown by dashed lines, corresponding to an analysis using a threshold with non-exceedance probability 0.9 at the conditioning location. Figure 6.4.4 shows an expanded view of Figure 6.4.3 for distances  $d < 180\text{km}$ ;  $\alpha(d) > 0.4$  for this interval of  $d$ , suggesting that extremal dependence is fairly strong for distances of this order. Referring to Figure 6.4.3, by the end of the transect, for  $d > 1100\text{km}$ , we observe values of  $(\alpha(d), \beta(d))$  near zero, which suggests almost complete independence at these distances. For  $d > 200\text{km}$ , the piecewise linear form for  $\mu(d)$  is decreasing with  $d$ , and at a distance of  $d = 1200\text{km}$ , appears to be near zero. Our functional form for  $\sigma(d)$  increases to its limiting value of  $\sqrt{2}$  quickly, suggesting independent Laplace residuals for  $d > 70\text{km}$ , as seen in Figure 6.4.4. This characteristic coincides with relatively low values of  $\delta(d)$ , particularly for  $d > 300\text{km}$ . For some  $d$ ,  $\delta(d)$  has credible intervals with lower bounds truncated at  $\delta(d) = 1$ , stemming from our choice of prior distribution. Hence, it is possible that allowing more flexibility in the domain of  $\delta(d)$  could be considered. In this analysis, the posterior value of  $\rho$  was found to be approximately 0.98; this corresponds to a linear correlation value of 0.44 at a distance of 1200km.

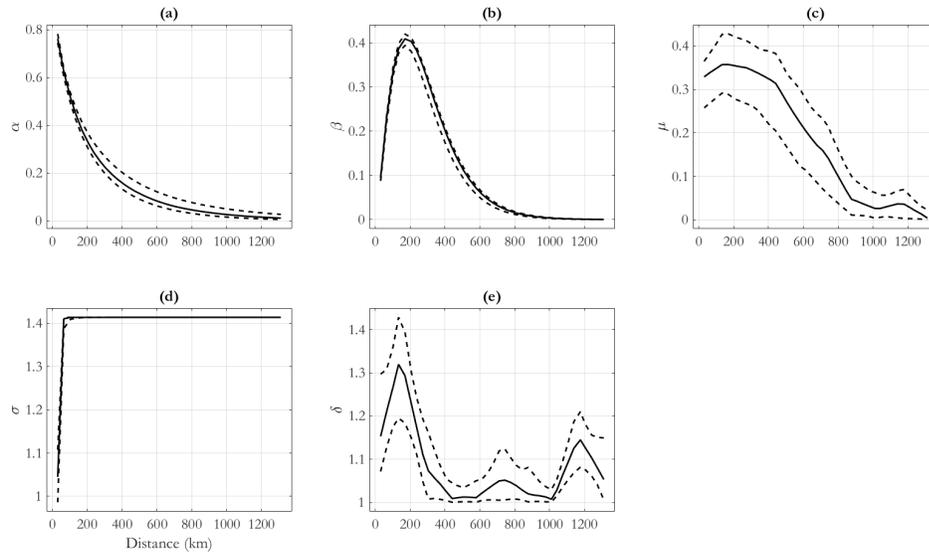


Figure 6.4.3: Parameter estimates, with distance  $d$  for the North Atlantic long transect analysis. The solid black line represents the posterior median, with dashed lines representing the upper and lower limits of the 95% posterior credible interval, for (a)  $\alpha(d)$ , (b)  $\beta(d)$ , (c)  $\mu(d)$ , (d)  $\sigma(d)$  and (e)  $\delta(d)$ .

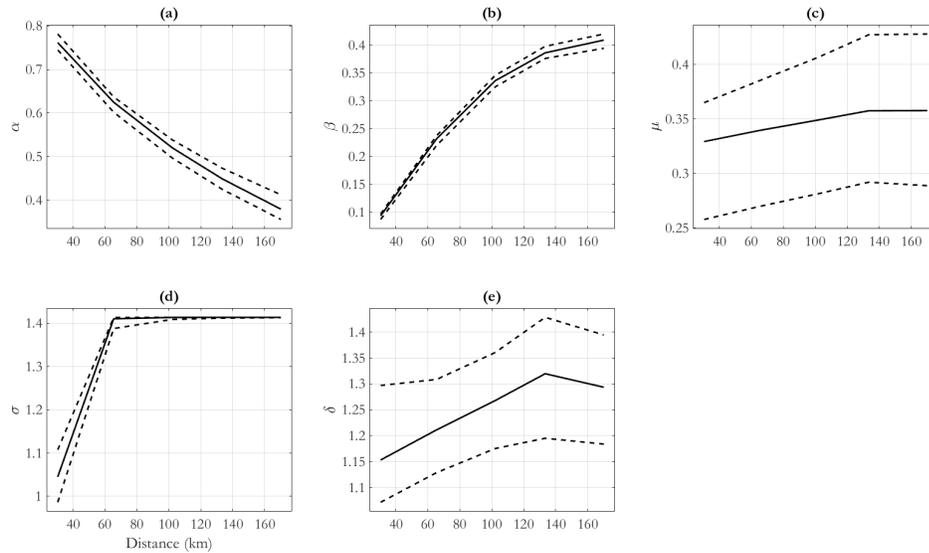


Figure 6.4.4: Parameter estimates, with distance  $d$  (for  $d < 180\text{km}$ ) for the North Atlantic long transect analysis. The solid black line represents the posterior median, with dashed lines representing the upper and lower limits of the 95% posterior credible interval, for (a)  $\alpha(d)$ , (b)  $\beta(d)$ , (c)  $\mu(d)$ , (d)  $\sigma(d)$  and (e)  $\delta(d)$ .

### North Sea ocean basin

The MCMC scheme was executed for 50000 iterations, with convergence occurring at approximately 25000 iterations. Posterior distributions of parameters are estimated here based on the final 1000 iterations. A threshold value at the conditioning site corresponding to the 0.85 Laplace quantile was used. Marginal summaries of the posterior distributions of parameters with distance are shown in Figure 6.4.5; solid lines show the posterior median and dashed lines indicate the upper and lower limits of the 95% posterior credible interval. In this figure, we display distance  $d$  using the local Cartesian coordinate frame using (6.3.5), assuming isotropy. In Figure 6.4.5, we see

fairly rapid decay of  $\alpha(d)$ , with  $\alpha(50\text{km}) \approx 0.6$  and  $\alpha(d = 300\text{km}) \approx 0.1$ . The parameter  $\beta(d)$  increases with distance initially, reaching a maximum of approximately 0.36 at  $d \approx 130\text{km}$ , and then decays. The value of  $\mu(d)$  increases with  $d$ , whereas the value of  $\delta$  decreases towards unity (the lower limit for its Uniform prior distribution). The parameter  $\sigma(d)$  follows the specified functional form, but does not reach its limit value of  $\sqrt{2}$  within the spatial range of this analysis. The posterior medians of anisotropy parameters  $\nu_1$  and  $\nu_2$  are approximately 1.04 and 0.0 respectively, indicating a small degree of anisotropy, consistent with the behaviour seen in Shooter et al. (2019). The 95% credible intervals for the parameters  $\nu_1$  and  $\nu_2$  were approximately (1.01, 1.10) and (-0.04, 0.04) respectively. Indeed, the behaviour of  $\alpha(d)$  and  $\beta(d)$  is similar to that seen for north-south transect analysis of this ocean basin in Shooter et al. (2019), but rather different to the east-west transect results from that paper.

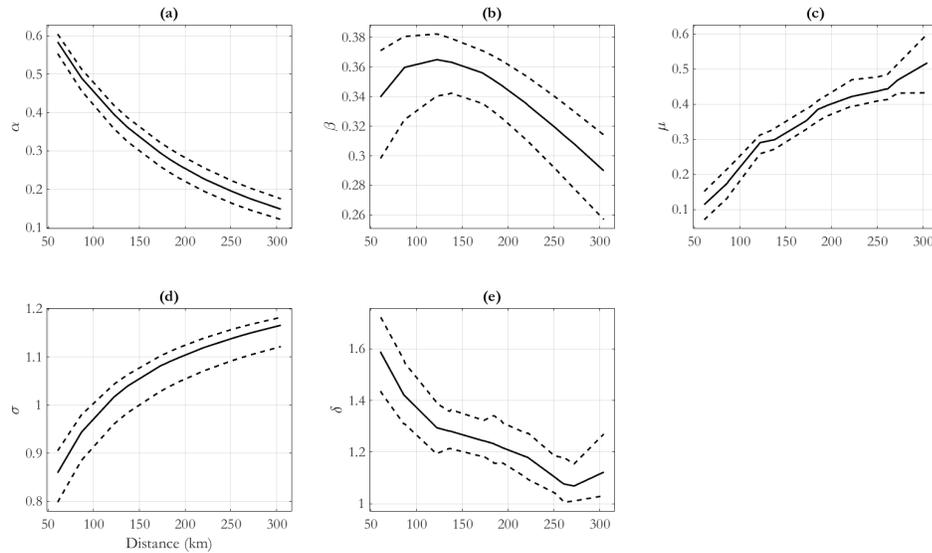


Figure 6.4.5: Parameter estimates, with distance  $d$  for the North Sea ocean basin analysis. The solid black line represents the posterior median, with dashed lines representing the upper and lower limits of the empirical 95% posterior credible interval, for (a)  $\alpha(d)$ , (b)  $\beta(d)$ , (c)  $\mu(d)$ , (d)  $\sigma(d)$  and (e)  $\delta(d)$ .

## 6.5 Conclusions

In this work, we use a spatial conditional extremes model to study ocean storm severity in the North Sea and the North Atlantic. The model describes marginal non-stationarity with respect to storm direction and season, and captures spatial anisotropy of the extremal dependence structure. The spatial conditional extremes model incorporates inter-location distance-dependent parameters, some represented as linear piecewise functions, others with pre-specified parametric forms. These allow asymptotic dependence at short inter-location distances, leading to asymptotic

independence, and eventually perfect independence, as distance increases. We allow flexible modelling of the residual distribution via a generalised Gaussian distribution, in preference to the Gaussian assumption used in previous work.

The importance of allowing for different forms of extremal dependence with distance is illustrated in Figure 6.5.1, which shows observations (in black) from the long zonal transect in the North Atlantic, with values of  $\approx 4$  at the conditioning, on Laplace scale. The mean value of the response at remote locations decreases with increasing distance, indicative of asymptotic independence. Under the assumption of asymptotic dependence, often made in spatial extremes modelling (e.g. using max-stable processes), the mean would be expected to return to the value of  $\approx 4$  (the grey line), at some finite distance. Realisations drawn under the estimated spatial conditional extremes model, admitting asymptotic independence are given in red; these appear to be consistent with the observations. We note that the greater smoothness of the red simulated processes compared to the observed processes is due to including all sampling locations across the transect for the observations, whilst only simulating data at the subset of sites which have been chosen for fitting the spatial conditional extremes model.

The current work suggests a number of potential avenues for further method development and application. From a methodological perspective, we are keen to make the spatial conditional extremes formulation consistent with our expectations regarding the spatial variation of extreme ocean storms. This probably requires more sophisticated representations of covariate effects within the spatial conditional extremes model, consistent with the quality of observational and calibrated simulator data

available. From an applications perspective, the approach would appear to be ideally suited for characterisation of spatial ocean surface roughness, as measured by satellite altimetry (Young and Ribal, 2019).

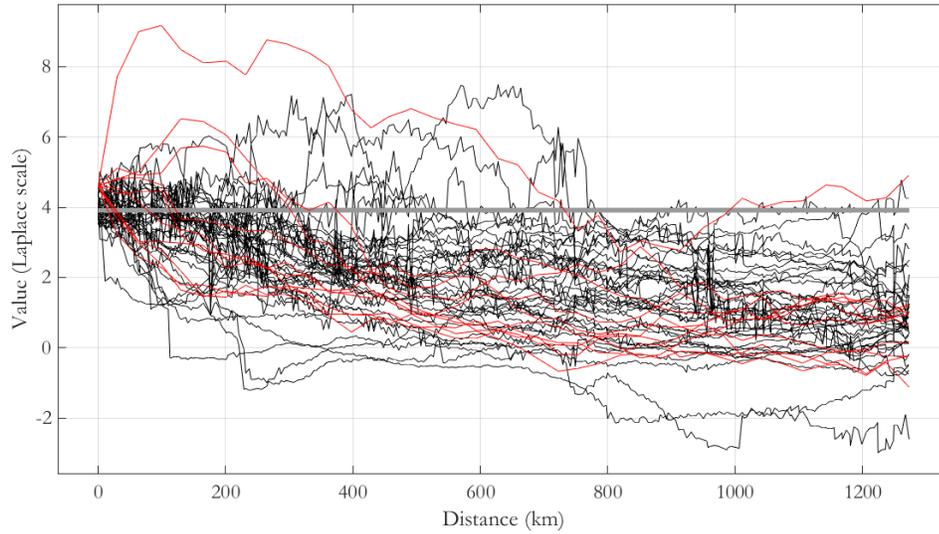


Figure 6.5.1: Simulated processes (red) from our fitted spatial conditional extremes model, using median values of the model parameters from the last 1000 iterations of the MCMC chain, and observed processes (black), conditioned on value of  $X_0 \approx 4$  for simulated data, and between 0.985 and 0.995 Laplace quantiles for observed data. The grey line shows the mean of the values of  $X_0$  conditioned upon.

# Chapter 7

## A Bayesian spatio-temporal model for precipitation extremes STOR team contribution to the EVA2017 challenge

### 7.1 Introduction

Recently, there have been numerous examples of devastating rainfall events - these include Storm Desmond, which hit northern England and Scotland, and Hurricane Harvey which affected the southern United States. In both cases, a large amount of damage and disruption was caused by severe flooding. By better understanding the probability of extreme rainfall events occurring, we can prepare more suitably for these potential flood events by adapting infrastructure appropriately.

The challenge data is comprised of precipitation readings for multiple weather stations in the Netherlands; the training set consists of data collected between 1972 and 1995 whilst the validation set was collected from 1996 to 2016, with different numbers of observations for each site. A detailed description of the data is provided in Wintenberger (2018). The aim of the competition is to predict extreme quantiles for the years 1996 to 2016 and predictions are assessed via a predefined error metric; see Wintenberger (2018).

There exists a rich literature within the extreme value theory framework for modelling precipitation extremes. A classical approach is to utilise block maxima. Suppose that we have independent and identically distributed (i.i.d.) random variables  $X_1, \dots, X_n$ , with  $M_n = \max\{X_1, \dots, X_n\}$ . When normalised appropriately, and as  $n \rightarrow \infty$ ,  $M_n$  follows a generalised extreme value (GEV) distribution (Fisher and Tippett, 1928), which has distribution function

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

where  $\{z\}_+ = \max\{0, z\}$ , and has parameters  $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ , corresponding to location, scale and shape parameters respectively.

An alternative technique is to follow Pickands (1975) and use exceedances of a threshold  $u$ . For some suitably large  $u$ , the conditional distribution function of  $(X_i - u) \mid (X_i > u)$  is approximately given by the generalised Pareto distribution (GPD), which has the form

$$H(x) = 1 - \left( 1 + \frac{\xi x}{\psi} \right)_+^{-\frac{1}{\xi}}, \quad x > 0, \quad (7.1.1)$$

where  $(\psi, \xi) \in \mathbb{R}_+ \times \mathbb{R}$  are the scale and shape parameters respectively. In the context

of the challenge at hand, both the GEV and GPD may be fitted separately at each site to give a model fit whereby any dependence is ignored.

By considering the physical process of rainfall, one can expect that nearby locations will exhibit similar behaviour, which invites improved inference by sharing information across sites. One popular method for the modelling of spatial extremes is to use max-stable processes (Brown and Resnick, 1977; Smith, 1990; Schlather, 2002). These arise as the limiting process from replications of spatial processes which have been suitably normalised (de Haan, 1984) and have been used to analyse rainfall data previously; see, for example, Davison et al. (2012) and Reich and Shaby (2012). However, such processes assume dependence of the extremes across sites; an investigation of pairwise dependence using scatter plots showed no clear evidence for this behaviour across the spatial grid. Moreover, max-stable models are difficult to fit and this would have been further impeded by the lack of data available at some sites.

Another approach is to impose spatial structure on the model parameters via a Bayesian hierarchical model; this is closer in nature to the method we propose. Spatial hierarchical models have been used previously to model spatial count data (Diggle et al., 1998) and, more recently, have been utilised in extreme value analysis. Cooley et al. (2007) describe a model, applied to rainfall data, whereby a GPD is fitted at the sampling locations, and allow the model parameters to vary according to a spatial process structure - in particular the authors use a Gaussian process for this. A spatio-temporal hierarchical modelling method for extreme events is given by Sang and Gelfand (2009), who apply their methods to precipitation data.

In this paper, we define a Bayesian hierarchical model which accounts for the

spatial and seasonal variation in the data. Our approach captures the frequency of non-zero events of precipitation and introduces an extremal mixture model, combining gamma and generalised Pareto distributions, for positive amounts of rainfall. Spatio-temporal structure in the parameters for the extremal mixture model is imposed via a separate autoregressive prior for each of them, which takes the form of a Gaussian Markov random field. Model estimates are then obtained using spatial interpolation and Markov chain Monte Carlo (MCMC) techniques. Cooley et al. (2007) defines a similar approach for continuous space, whereas we consider a finite number of sites and additionally incorporate seasonality.

The remainder of this article is structured as follows. Section 7.2 details our Bayesian framework and its estimation: in Sections 7.2.1 and 7.2.2 respectively, we specify our likelihood and prior models; in Section 7.2.3, we discuss parameter estimation. In Section 7.3, we discuss the results obtained using our method for modelling rainfall extremes, and highlight areas for potential improvements.

## 7.2 Methodology

### 7.2.1 Likelihood

Interest lies in modelling the daily rainfall amounts for each site and month. Due to seasonality in the rainfall data, the weak extremal dependence of the daily amount of rainfall across sites and the nature of the challenge, we model each month and site individually. Specifically, daily rainfall events within a month at a site are assumed to be i.i.d. Our model is motivated by an analysis of the sites for which data have

been recorded for at least five years.

Let  $R_{j,m}$  denote the random variable corresponding to the daily rainfall amount at site  $j$  for a day in month  $m = 1, \dots, 12$ . We consider the transformed random variable

$$\tilde{R}_{j,m} = \log(1 + R_{j,m}). \quad (7.2.1)$$

Wadsworth et al. (2010) show that such a transformation may increase the rate of convergence of the distribution tails to an extreme value form, in particular for distributions which appear as heavy-tailed as our rainfall data. Predictions on the extreme quantiles of  $R_{j,m}$  are later obtained in Section 7.3 by reversing this transformation. We note that the transformed observations are non-negative and an observation of  $R_{j,m} = 0$  remains unchanged.

We infer on the distribution of  $\tilde{R}_{j,m}$  by defining a hierarchical model. The first model component considers occurrences of non-zero amounts of rainfall on a day,  $\tilde{R}_{j,m} > 0$ , and we denote their probability by  $p_{j,m}$ . A temporal trend in  $p_{j,m}$  was investigated, but we did not find evidence of this for any site. Next, we consider the distribution  $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} > 0)$ . There exists a rich literature on modelling positive rainfall amounts, such as Wilks (2006), So et al. (2015) and Yunus et al. (2017). By investigating QQ plots, we find that an estimated gamma distribution works quite well for non-extreme amounts of precipitation. However, most of the observed monthly extremes are not captured well.

To improve the model fit, we define an extremal mixture model (Frigessi et al., 2002; Behrens et al., 2004; MacDonald et al., 2011) which combines the gamma distri-

bution with a GPD as defined in (7.1.1). Given a threshold  $u_{j,m}$ ,  $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} \leq u_{j,m})$  follows a truncated gamma distribution, while  $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} > u_{j,m})$  is generalised Pareto distributed. Formally, let  $G_{j,m} \sim \text{Gamma}(\alpha_{j,m}, \beta_{j,m})$  with shape  $\alpha_{j,m}$  and rate  $\beta_{j,m}$ , and  $H_{j,m} \sim \text{GPD}(\psi_{j,m}, \xi_{j,m})$  with scale  $\psi_{j,m} = \tilde{\psi}_{j,m} - \xi u_{j,m}$  and shape  $\xi_{j,m}$ . The reparametrisation of the scale parameter in  $H_{j,m}$  removes the effect of the threshold on inference and has been used in previous studies (Fawcett and Walshaw, 2006). Then, the cumulative distribution function of  $\tilde{R}_{j,m} \mid (\tilde{R}_{j,m} > 0)$  is given by

$$\mathbb{P}(\tilde{R}_{j,m} > r \mid \tilde{R}_{j,m} > 0) = \begin{cases} \mathbb{P}(G_{j,m} > r) & r \leq u_{j,m}, \\ \mathbb{P}(G_{j,m} > u_{j,m}) \mathbb{P}(H_{j,m} > r - u_{j,m}) & r > u_{j,m}. \end{cases} \quad (7.2.2)$$

Combining the model components defined above, the event  $\tilde{R}_{j,m} > r$ , for  $r > u_{j,m}$ , occurs with probability

$$\mathbb{P}(\tilde{R}_{j,m} > r) = p_{j,m} \mathbb{P}(G_{j,m} > u_{j,m}) \mathbb{P}(H_{j,m} > r - u_{j,m}).$$

Due to the empirical mean of  $R_{j,m} \mid (R_{j,m} > 0)$  being similar for all  $j$ , we fix  $\alpha_{j,m}$ ,  $m = 1, \dots, 12$  in the gamma distribution to be constant across sites and, thus, refer to this parameter as  $\alpha_m$  in the rest of this paper.

## 7.2.2 Prior model

Prior selection is critical in this analysis due to the varying degrees of data availability at each site; inference at sites where data are lacking or unavailable will be dominated by the prior distribution. We considered uninformative, improper uniform priors on  $\log \alpha_m$ ,  $\log \beta_{j,m}$ ,  $\log \tilde{\psi}_{j,m}$  and  $\xi_{j,m}$ . However, these produced unrealistic estimates of

$\xi_{j,m}$ , mostly due to the difficulty in estimating  $\xi_{j,m}$  given short data records. Studies on extreme rainfall often feature the prior used in Martins and Stedinger (2000) which constrains the shape parameter to be in a sensible interval.

We instead introduce a prior aimed at exploiting the spatial and seasonal structure of the model parameters. We assume that parameters for neighbouring sites and adjacent months are likely to be similar. Explicitly, we propose for  $\phi_{j,m}$ , an arbitrary parameter at site  $j$  and month  $m$ , that

$$\phi_{j,m} \sim \mathcal{N} \left( \frac{\phi_{j,m-1} + \phi_{j,m+1} + \sum_{j' \neq j} \phi_{j',m} d_{j,j'}}{2 + \sum_{j' \neq j} d_{j,j'}}, \frac{1}{(2 + \sum_{j' \neq j} d_{j,j'}) \tau_\phi} \right), \quad (7.2.3)$$

where  $\tau_\phi > 0$  denotes the precision for parameter  $\phi$ , common to all sites and months, where  $\phi$  is one of our model parameters. The constant  $d_{j,j'} \geq 0$  describes our prior belief concerning the degree of similarity of  $\phi_{j,m}$  and  $\phi_{j',m}$ . This prior is a variant of the Intrinsic Autoregressive (IAR) prior as described in Banerjee et al. (2004) and allows us to pool information across neighbouring sites and months, which helps to produce more stable parameter estimates and to reduce uncertainty in these estimates. The cyclical nature of the sequence of months means that values 0 and 13 for  $m - 1$  and  $m + 1$  should be replaced by the values 12 and 1 respectively in order to ensure that December and January are correctly identified as being adjacent months. We define a flat, conjugate Gamma(1, 0.001) prior for  $\tau_\phi$ .

### 7.2.3 Threshold selection and estimation

We detail our approach to estimate the model defined in Sections 7.2.1 and 7.2.2 in the following. First, we consider  $p_{j,m}$ , which can be estimated independently from the

remaining parameters due to the hierarchical model structure. Next, the selection of the thresholds  $u_{j,m}$  is described. Finally, we infer on the remaining model parameters via an MCMC algorithm which is outlined at the end of this subsection.

For sites with more than five years of data, we estimate  $p_{j,m}$  empirically due to the high number of observations available. We infer on the remaining sites via spatial interpolation. Let  $\mathcal{J}$  denote the indices of the sites with at least five years of data. We further define a pairwise weighting between arbitrary sites  $j$  and  $j'$  by introducing the weight

$$d_{j,j'} = \exp(-\|\mathbf{x}_j - \mathbf{x}_{j'}\|), \quad (7.2.4)$$

where  $\mathbf{x}_j$  denotes the longitude and latitude coordinates of site  $j$  and  $\|\cdot\|$  corresponds to the Euclidean distance. As the study region is small, the curvature of the earth is negligible and the Euclidean distance in the two-dimensional space is close to the true distance between the sites. Then for a site  $j \notin \mathcal{J}$ , the estimate  $\hat{p}_{j,m}$  for  $p_{j,m}$  is derived as

$$\hat{p}_{j,m} = \sum_{j' \in \mathcal{J}} d_{j,j'} \hat{p}_{j',m}. \quad (7.2.5)$$

The weights  $d_{j,j'}$  defined in (7.2.4) are identical to the ones which we set in the prior density (7.2.3). As the weighting function (7.2.4) produces larger values for locations close together, a higher weight is given to neighbouring sites.

We now consider how to select the thresholds,  $u_{j,m}$ , of our model (7.2.2). These thresholds must be large enough for the asymptotic argument of Pickands (1975) to approximately hold whilst also low enough so that we have a sufficient number of observations for reliable model fitting. We use the classical fixed threshold approach

as described in Coles (2001) for the sites in  $\mathcal{J}$ . Specifically, by inspection of threshold stability plots, we find the smallest threshold above which the GPD is an appropriate model for the exceedances. For the other sites, we estimate these thresholds in an equivalent manner to (7.2.5). Other threshold selection methods are outlined by Scarrott and MacDonald (2012).

The parameters of our gamma-GPD mixture model are estimated using MCMC methods. We sample from the posterior distribution using a Metropolis-within-Gibbs scheme. In particular, proposal values of each parameter are generated sequentially from a Gaussian distribution and accepted with a probability defined as the posterior ratio of the proposed state relative to the current state of the Markov chain. The hyperparameter  $\tau_\phi$  in (7.2.3) is updated by sampling from the full conditional Gamma posterior as described by Knorr-Held (2003). We tune the parameters of the MCMC algorithm to ensure an acceptance rate of 20-25% in accordance with the optimality criterion of Roberts et al. (1997).

### 7.3 Results and discussion

We begin this section by considering the results of the MCMC implementation. We run our MCMC chains for 20000 iterations, and discard the first 5000 iterations as burn-in to aid convergence. Examples of the chains produced are provided in Figure 7.3.1 for scale and shape parameters  $\psi_{10,6}$  and  $\xi_{10,6}$ . Estimates of these parameters were obtained using the posterior means of their respective MCMC chains. These plots demonstrate that good mixing has been achieved for this case; similar results were

obtained across other stations and months.

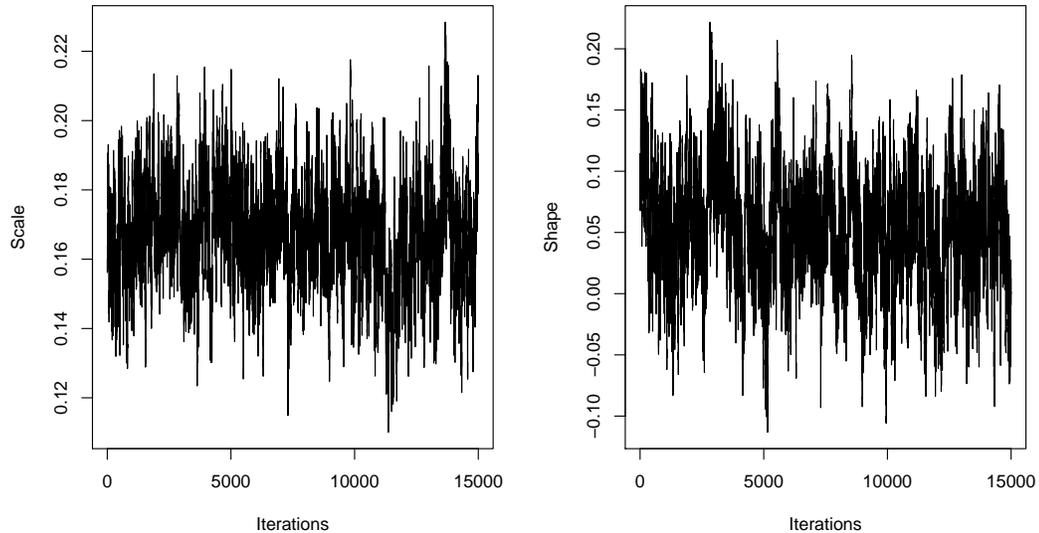


Figure 7.3.1: MCMC chains for the scale and shape parameters for station 10 in June.

We now explore the monthly variation in the estimated model parameters by focussing on results at four nearby stations. The locations of these stations are shown in the top left panel of Figure 7.3.2. The data set contains over 8000 observations for stations 2 and 5, and no observations for stations 7 and 10. The top right and bottom left panels of Figure 7.3.2 show our estimates of the scale and shape parameters, respectively, at these four locations. These plots demonstrate the seasonality in the parameter estimates, with higher values of both the scale and shape generally corresponding to summer and autumn months. This effect is maintained in the predicted 0.998 quantiles, shown in the bottom right panel of Figure 7.3.2, which are typically highest between June and October. A similar trend was observed at other sites, particularly those with limited data where estimates are more heavily influ-

enced by information from other locations, due to the spatial smoothing imposed by the model.

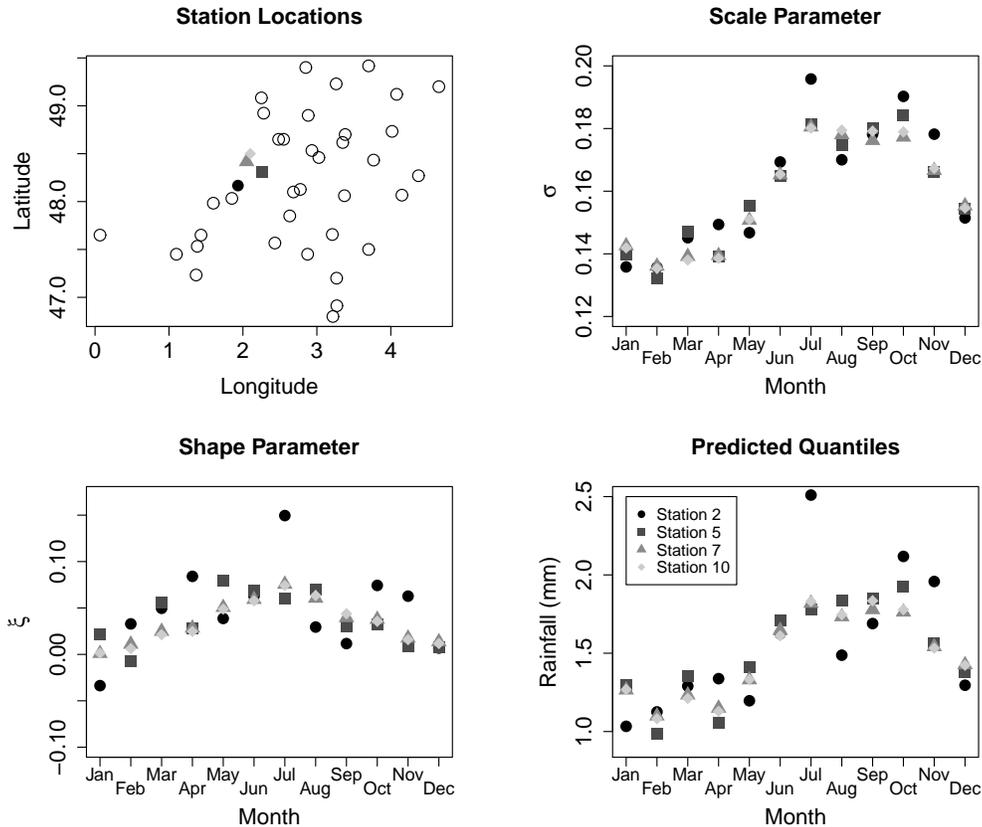


Figure 7.3.2: Location of stations 2, 5, 7 and 10, as well as estimates of the corresponding scale and shape parameters and predicted 0.998 quantiles.

We now consider our estimates in the context of the competition, which used the quantile loss function by Koenker (2005). In particular, as in the challenge, we consider the percentage improvement provided by our method over benchmark predictions. The competition was split into two challenges: Challenge 1 involved only sites where observations were available, with the benchmark quantile estimates being given by the monthly maxima at each station; Challenge 2 included predictions

for all sites, with the benchmark for those sites with no data being taken as the average of the quantiles predicted in Challenge 1 for each month. Our method gave a 59.9% improvement over the benchmark for Challenge 1, and a 57.7% improvement for Challenge 2. Table 7.3.1 shows the performance of our approach using this same metric, but with the results separated by month.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Challenge 1	57.7	71.1	60.0	65.0	43.7	62.8	65.9	77.0	38.7	38.4	52.2	33.4
Challenge 2	54.4	69.3	57.4	61.9	43.1	60.7	64.2	75.4	37.9	36.4	49.3	31.3

Table 7.3.1: Percentage improvement over the benchmark for Challenges 1 and 2 across each month.

As is to be expected, our method performed better in Challenge 1, where only predictions for sites with observations were considered, across all months. Looking at these results separately for each month allows us to identify possible areas for improvement. In particular, the scores for September, October and December are lower than for other months, suggesting that the method could be improved by focussing on the modelling of autumn and winter months.

# Chapter 8

## Conclusions and further work

The results presented in this thesis show that the accurate modelling of the dependence structure and behaviour of environmental extremes is quite complex in nature. In applications, this behaviour can have a large impact on assessing the probability of extreme events of a specific magnitude occurring; this could, for example, provide information for the design criteria for offshore structures.

In Chapters 4, 5 and 6, we provided details of a conditional spatial extreme value model which is able to capture the two limiting extremal dependence types as the distance between points change. Previous spatial models have not had this capability. Inference from the model is relatively straightforward, based around a non-linear regression framework. This has been incorporated into a Bayesian methodology, providing more computational flexibility than maximum likelihood techniques, the latter of which proved to be difficult to fit due to the large number of parameters contained in this spatial model.

The models we have described also have a theoretical underpinning with as few

strong modelling assumptions used as possible, particularly in Chapter 7, where a simple model for prediction of extreme quantiles of rainfall was constructed in a relatively straightforward manner. It is encouraging that the results of Chapter 6 for the long transect in the north Atlantic appear to largely exhibit the expected theoretical results, such as standard Laplace random variables, asymptotically independent of extreme events at the conditioning location, being obtained at large distances, and suggests that the spatial conditional extremes models are suitable from this perspective.

One major issue highlighted here is that the typical assumption of asymptotic dependence across all sites in a spatial field appears to be inappropriate for the data we have considered here. In particular, it appears that even at short distances between sites we do not observe asymptotic dependence for significant wave height data in the North Sea and north Atlantic Ocean. It may be anticipated that other environmental data sets, such as extremes of temperature may indeed see such behaviour; if this were the case then the lag-asymptotically dependent modelling proposals of Wadsworth and Tawn (2019) may prove useful in modelling such data.

We have also shown that when dealing with spatial data, the assumptions made when modelling have a significant impact on inference. For instance, our model in Chapter 5 suggests that there is indeed a significant difference in dependence structure when comparing east-west transects in the North Sea with north-south transects. However, in Chapter 6, we find that after transformation of the coordinates using a local Cartesian framework, the difference in dependence is possibly lower with direction, perhaps due to the difficulty of capturing anisotropic behaviour in this way;

investigation of this apparent effect would be scope for further work in utilising this distance metric.

In various chapters, we have also focussed on the computational aspects of the modelling procedure. In Chapter 3, we showed that the use of a censoring scheme motivated by the assumption of asymptotic independence did indeed mean that the computational time was significantly reduced, but at the cost of a relatively large increase in the bias of the estimated parameters; suggesting that care is necessary when implementing censored likelihoods for extremes. Furthermore, in Chapter 6, we suggest the use of linear piecewise functions within the conditional spatial extreme value model presented in Chapter 4 in order to reduce the number of parameters to estimate; if the model described in Chapter 5 were used on the full two-dimensional grid of locations in the North Sea used in these chapters, the computational expense would be huge. Computational efficiency is important in the applications that have been considered, since decisions on infrastructure may need to be made across a large number of locations.

Other future development work on the spatial conditional extremes models which we present here may include the incorporation of spatial pooling, such as that highlighted in Chapter 5, into the model presented in Chapter 6. However, the obvious practical implication of this would be that the computational time required to fit the spatial conditional extremes model would be multiplied by the number of sites to be pooled over; indeed, this has been found to be the case when initially attempting to implement this on the model. As such, an efficient method of pooling, or one where enough information can be gained to make the additional computing time worthwhile,

would be highly desirable should spatial pooling methods be implemented.

Another avenue of research could be the use of splines to model the spatial conditional extremes model parameters, similar to methods used by Jones et al. (2016) and Randell et al. (2016) for example, which can be viewed as an extension of the use of linear piecewise functions in Chapter 6. A benefit of using splines is that such methods may allow a more flexible structure in the spatial conditional extreme value model parameters as they change with distance. Moreover, the concept of using linear piecewise functions in Chapter 6 could be expanded by modelling all of the spatial conditional extremes model parameters using either splines or linear piecewise functions. By doing this, we would relax our assumptions that the parameters should take particular forms and allow a more flexible model. However, it is possible that in allowing more flexibility parameter identifiability may suffer; in the work for Chapter 5, it was seen that it was difficult to obtain similar values for some of the model parameters from different model fits without parametric forms being imposed. Thus, extra constraints on the parameters may be required if using such models does not permit physically sensible solutions.

The incorporation of covariates into the two-dimensional conditional spatial extremes model presented in Chapter 6 could also be investigated. An obvious starting point for this would be to allow the spatial dependence structure to change with direction. For instance, Chapter 5 suggests that there may a difference of extremal dependence behaviour of significant wave height in the northern North Sea in a west-east direction compared to a north-south direction; this behaviour is also suggested by Kereszturi et al. (2016) and Ross et al. (2017a), for example. Whilst the model

presented in Chapter 6 attempts to describe such effects by utilising a matrix to incorporate anisotropy, the model fitted to the northern North Sea data appears to suggest only a small difference in the dependence structure with direction. Since the parameters of this matrix may be difficult to estimate accurately, the use of covariate effects in the spatial conditional extremes model may provide a better assessment of the dependence structure in such applications; Jonathan et al. (2013) and Winter et al. (2017) propose methods of incorporating covariates into the Heffernan and Tawn (2004) conditional extremes model for instance. Moreover, by utilising covariates, the change of extremal behaviour with direction may be more explicitly defined with simpler interpretation of results from the model.

# Bibliography

- H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19:716–723, 1974.
- P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. *Ann. Appl. Stat.*, 9:2023–2050, 2015.
- J. Bacro, C. Gaetan, and G. Toulemonde. A flexible dependence model for spatial extremes. *J. Stat. Plan. Infer.*, 172:36–52, 2016.
- S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton, 2004.
- C. N. Behrens, H. F. Lopes, and D. Gamerman. Bayesian analysis of extreme events with threshold estimation. *Stat. Modelling*, 4:227–244, 2004.
- J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of extremes: theory and applications*. Wiley, Chichester, UK, 2004.
- O. Breivik, O. J. Aarnes, J.-R. Bidlot, A. Carrasco, and O. Saetra. Wave extremes in the North East Atlantic from ensemble forecasts. *J. Climate*, 26:7525–7540, 2013.

- B. M. Brown and S. I. Resnick. Extreme values of independent stochastic processes. *J. Appl. Probab.*, 14:732–739, 1977.
- S. G. Coles. *An Introduction To Statistical Modelling of Extreme Values*. Springer, 2001.
- S. G. Coles and J. A. Tawn. Modelling multivariate extreme events. *J. Roy. Stat. Soc. B*, 53:377–92, 1991.
- S. G. Coles and J. A. Tawn. Statistical methods for multivariate extremes: An application to structural design (with discussion). *J. Roy. Stat. Soc. C Appl. Stat.*, 43(1):1–31, 1994.
- S. G. Coles and J. A. Tawn. A Bayesian analysis of extreme rainfall data. *J. Roy. Stat. Soc. C Appl. Stat.*, 45:463–478, 1996.
- S. G. Coles, J. E. Heffernan, and J. A. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2:339–365, 1999.
- D. Cooley and S. R. Sain. Spatial hierarchical modeling of precipitation extremes from a regional climate model. *J. Agr. Biol. Environ. Stat.*, 15(3):381–402, 2010.
- D. Cooley, D. Nychka, and P. Naveau. Bayesian spatial modelling of extreme precipitation return levels. *J. Am. Statist. Assoc.*, 102:824–840, 2007.
- A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *J. R. Statist. Soc. B*, 52:393, 1990.

- A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Statist. Sci.*, 27:161–186, 2012.
- C. de Boor. *A Practical Guide To Splines*. Springer Applied Mathematical Sciences, 1978.
- L. de Haan. A spectral representation for max-stable processes. *Ann. Prob.*, 12:1194–1204, 1984.
- A. B. Dieker and T. Mikosch. Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes*, 18:301–314, 2015.
- P. J. Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics. *J. Roy. Stat. Soc. C Appl. Stat.*, 47(3):299–350, 1998.
- C. Dombry and M. Ribatet. Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface*, 8(1):9–17, 2015.
- E. F. Eastoe and J. A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *J. Roy. Stat. Soc. C Appl. Stat.*, 58:25–45, 2012.
- S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. Roy. Stat. Soc. B*, 77(1):239–265, 2015.
- S. Engelke, T. Opitz, and J.L. Wadsworth. Extremal dependence of random scale constructions. *Extremes*, Jul 2019.

- K. C. Ewans and P. Jonathan. The effect of directionality on northern North Sea extreme wave design criteria. *J. Offshore Arct. Eng.*, 130:041604:1–041604:8, 2008.
- L. Fawcett and D. Walshaw. A hierarchical model for extreme wind speeds. *J. Roy. Stat. Soc. C Appl. Stat.*, 55:631–646, 2006.
- G. Feld, D. Randell, Y. Wu, K. Ewans, and P. Jonathan. Estimation of storm peak and intra-storm directional-seasonal design conditions in the North Sea. *J. Offshore Arct. Eng.*, 137:021102:1–021102:15, 2015.
- A. Ferreira and L. de Haan. The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737, 11 2014.
- R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, 24:180–190, 1928.
- A. Frigessi, O. Haug, and H. Rue. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5:219–235, 2002.
- M. G. Genton, Y. Ma, and H. Sang. On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98:481–488, 2011.
- Her Majesty’s Government. National Flood Resilience Review. Technical report, 08 September 2016.
- E. J. Gumbel. Bivariate exponential distributions. *J. Am. Statist. Assoc.*, 55(292):698–707, 1960.

- J. E. Heffernan and S. I. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17:537–571, 2007.
- J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values (with discussion). *J. Roy. Stat. Soc. B*, 66:497–546, 2004.
- R. Huser and A. C. Davison. Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518, 2013.
- R. Huser, A. C. Davison, and M. G. Genton. Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103, 2016.
- R. G. Huser and J. L. Wadsworth. Modeling spatial processes with unknown extremal dependence class. *J. Am. Statist. Assoc.*, 114:434–444, 2018.
- P. Jonathan, K. C. Ewans, and D. Randell. Joint modelling of environmental parameters for extreme sea states incorporating covariate effects. *Coastal Eng.*, 79:22–31, 2013.
- P. Jonathan, K. C. Ewans, and D. Randell. Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25:172–188, 2014.
- M. Jones, D. Randell, K. Ewans, and P. Jonathan. Statistics of extreme ocean environments: non-stationary inference for directionality and other covariate effects. *Ocean Eng.*, 119:30–46, 2016.
- C. Keef, J. Tawn, and C. Svensson. Spatial risk assessment for extreme river flows. *J. Roy. Stat. Soc. C Appl. Stat.*, 58:601–618, 2009.

- C. Keef, I. Papastathopoulos, and J. A. Tawn. Estimation of the conditional distribution of a vector variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Multivariate Anal.*, 115:396–404, 2013a.
- C. Keef, J. A. Tawn, and R. Lamb. Estimating the probability of widespread flood events. *Environmetrics*, 24:13–21, 2013b.
- M. Kereszturi. *Assessing and modelling extremal dependence in spatial extremes*. PhD thesis, Lancaster University, U.K., 2016.
- M. Kereszturi, J. Tawn, and P. Jonathan. Assessing extremal dependence of North Sea storm severity. *Ocean Eng.*, 118:242–259, 2016.
- A. J. Kettle. Review Article: Storm Britta in 2006: offshore damage and large waves in the North Sea. *Nat. Hazards Earth Syst. Sci. Discuss*, 3:5493–5510, 2015.
- A. Kiriliouk, H. Rootzen, J. Segers, and J. L. Wadsworth. Peaks over thresholds modelling with multivariate generalized Pareto distributions. *Technometrics*, 61: 123–135, 2019.
- L. Knorr-Held. Some remarks on Gaussian Markov random field models for disease mapping. In *Highly Structured Stochastic Systems*, pages 203–207. Oxford University Press, Oxford, 2003.
- R. Koenker. *Quantile regression*. Cambridge University Press, New York, 2005.
- R. Lamb, C. Keef, J. A. Tawn, S. Laeger, I. Meadowcroft, S. Surendran, P. Dunning,

- and C. Batstone. A new method to assess the risk of local and widespread flooding on rivers and coasts. *J. Flood Risk Manag.*, 3:323–336, 2010.
- R. Lamb, M. Szönyi, and P. May. Flooding after Storm Desmond. *JBA Trust*, 2015.
- A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.
- Y. Liu and J. A. Tawn. Self-consistent estimation of conditional multivariate extreme value distributions. *J. Multivariate Anal.*, 127:19–35, 2014.
- T. Lugrin. *Semi-parametric Bayesian risk estimation for complex extremes*. PhD thesis, EPFL, Lausanne, Switzerland, 2018.
- T. Lugrin, A. C. Davison, and J. A. Tawn. Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, 19:491–515, 2016a.
- T. Lugrin, A. C. Davison, and J. A. Tawn. Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, 19(3):491–515, 2016b.
- A. MacDonald, C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell. A flexible extreme value mixture model. *Comput. Statist. Data Anal.*, 55:2137–2157, 2011.
- E. S. Martins and J. R. Stedinger. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, 36(3):737–744, 2000.
- R. B. Nelsen. *An introduction to copulas*. Springer, 2006.

- Oceanweather. *North European Storm Study User Group Extension and Reanalysis Archive*. Oceanweather Inc., 2002.
- S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood - based inference for max - stable processes. *J. Am. Statist. Assoc.*, 105:263–277, 2010.
- J. Pickands. Statistical inference using extreme order statistics. *Ann. Stat.*, 3:119–131, 1975.
- D. Randell, G. Feld, K. Ewans, and P. Jonathan. Distributions of return values for ocean wave characteristics in the South China Sea using directional-seasonal extreme value analysis. *Environmetrics*, 26:442–450, 2015.
- D. Randell, K. Turnbull, K. Ewans, and P. Jonathan. Bayesian inference for non-stationary marginal extremes. *Environmetrics*, 27:439–450, 2016.
- B. J. Reich and B. A. Shaby. A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.*, 6:1430–1451, 2012.
- S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Applied Probability vol. 4. Springer-Verlag, New York, 1987.
- M. Ribatet. Spatial extremes: max-stable processes at work. *J. Soc. Fr. Statistique*, 154:156–177, 2013.
- M. Ribatet, D. Cooley, and A.C. Davison. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Stat. Sinica*, 22:813–845, 2012.

- G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *J. Comp. Graph. Stat.*, 18:349–367, 2009.
- G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.*, 7(1):110–120, 1997.
- H. Rootzen, J. Segers, and J. L. Wadsworth. Multivariate generalized Pareto distributions: parametrizations, representations, and properties. *J. Multivariate Anal.*, 165:117–131, 2018a.
- H. Rootzen, J. Segers, and J. L. Wadsworth. Multivariate peaks over thresholds models. *Extremes*, 21:115–145, 2018b.
- E. Ross, M. Kereszturi, M. van Nee, D. Randell, and P. Jonathan. On the spatial dependence of extreme ocean storm seas. *Ocean Eng.*, 145:1–14, 2017a.
- E. Ross, D. Randell, K. Ewans, G. Feld, and P. Jonathan. Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. *Ocean Eng.*, 142:315–328, 2017b.
- S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker. The NCEP Climate Forecast System Version 2. *J. Climate*, 27(6):2185–2208, 2014.
- H. Sang and A. E. Gelfand. Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.*, 16(3):407–426, 2009.

- H. Sang and A. E. Gelfand. Continuous spatial process models for spatial extreme values. *J Agr. Biol. Envir. St.*, 15:49–65, 2010.
- C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *Revstat*, 10:33–60, 2012.
- M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5:33–44, 2002.
- M. Schlather and J. A. Tawn. A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156, 03 2003.
- E. M. Shaw, K. J. Beven, N. A. Chappell, and R. Lamb. *Hydrology in Practice*. CRC Press, 2010.
- R. Shooter, E. Ross, J. Tawn, and P. Jonathan. On spatial conditional extremes for ocean storm severity. *Environmetrics*, 30(6):e2562, 2019.
- R. L. Smith. Max-stable processes and spatial extremes. *Unpublished article, available electronically from [www.stat.unc.edu/postscript/rs/spatex.pdf](http://www.stat.unc.edu/postscript/rs/spatex.pdf)*, 1990.
- R. L. Smith, J. A. Tawn, and H. K. Yuen. Statistics of multivariate extremes. *Int. Stat. Rev.*, pages 47–58, 1990.
- B.-J. So, H.-H. Kwon, D. Kim, and S. O. Lee. Modeling of daily rainfall sequence and extremes based on a semiparametric Pareto tail approach at multiple locations. *J. Hydrology*, 529:1442–1450, 2015.

- O. R. Sorensen, H. Kofoed-Hansen, M. Rugbjerg, and L. S. Sorensen. A third-generation spectral wave model using an unstructured finite volume technique. *Proceedings of the 29th International Conference on Coastal Engineering*, 2005.
- D. J. Spiegelhalter, N. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit (with discussion). *J. Roy. Stat. Soc. B*, 64:1–34, 2002.
- A. Stephenson and J. A. Tawn. Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227, 2005.
- J. A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75:397–415, 1988.
- J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77:245–253, 1990.
- J. A. Tawn, R. Shooter, R. Towe, and R. Lamb. Modelling spatial extreme events with environmental applications. *Spat. Stat.*, 28:39–58, 2018.
- R. Towe, E. Eastoe, J. Tawn, and P. Jonathan. Statistical downscaling for future extreme wave heights in the North Sea. *Ann. Appl. Stat.*, 11:2375–2403, 2017.
- R. P. Towe, J. A. Tawn, R. Lamb, C. G. Sherlock, and Y. Liu. Improving statistical models for flood risk assessment. In *E3S Web of Conferences*, volume 7, page 01011. EDP Sciences, 2016.
- R. P. Towe, J. A. Tawn, R. Lamb, and C. G. Sherlock. Model-based inference of condi-

- tional extreme value distributions with hydrological applications. *Environmetrics*, 0(0):In press, 2019.
- G. Vallis. *Atmospheric and oceanic fluid dynamics*. Cambridge University Press, Cambridge, UK, 2017.
- C. Varin. On composite marginal likelihoods. *Adv. Stat. Anal.*, 92:1–28, 2008.
- C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Stat. Sin.*, 21:5–42, 2011.
- J. L. Wadsworth and J. A. Tawn. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *J. Roy. Stat. Soc. B*, 2012a.
- J. L. Wadsworth and J. A. Tawn. Dependence modelling for spatial extremes. *Biometrika*, 99:253–272, 2012b.
- J. L. Wadsworth and J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15, 2014.
- J. L. Wadsworth and J. A. Tawn. Spatial conditional extremes. *In submission*, 2019.
- J. L. Wadsworth, J. A. Tawn, and P. Jonathan. Accounting for choice of measurement scale in extreme value modelling. *Ann. Appl. Stat.*, 4:1558–1578, 2010.
- J. L. Wadsworth, J. A. Tawn, A. C. Davison, and D. M. Elton. Modelling across extremal dependence classes. *J. Roy. Stat. Soc. C Appl. Stat.*, 79:149–175, 2017.
- D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series ; 59. Elsevier, Amsterdam, 2nd ed. edition, 2006. ISBN 0127519661.

- O. Wintenberger. Editorial: special issue on the extreme value analysis conference challenge “prediction of extremal precipitation”. *Extremes*, 21(3):425–429, 2018.
- H. C. Winter and J. A. Tawn. Modelling heatwaves in central France: a case-study in extremal dependence. *J. Roy. Stat. Soc. C Appl. Stat.*, 65:345–365, 2016.
- H. C. Winter and J. A. Tawn. kth-order Markov extremal models for assessing heat-wave risks. *Extremes*, 20:393–415, 2017.
- H. C. Winter, J. A. Tawn, and S. J. Brown. Modelling the effect of the El Niño-southern oscillation on extreme spatial temperature events over Australia. *Ann. Appl. Stat.*, 10:2075–2101, 2016.
- H. C. Winter, S. J. Brown, and J. A. Tawn. Characterising the changing behaviour of heatwaves with climate change. *Dyn. Stats. Clim. Sys.*, 1:dzw006, 2017.
- I. R. Young and A. Ribal. Multiplatform evaluation of global trends in wind speed and wave height. *Science*, 364(6440):548–552, 2019.
- R. M. Yunus, M. M. Hasan, N. A. Razak, Y. Z. Zubairi, and P. K. Dunn. Modelling daily rainfall with climatological predictors: Poisson-Gamma generalized linear modelling approach. *Int. J. Climatol.*, 37(3):1391–1399, 2017.