

Quantifying extreme multivariate risks with environmental applications

Nicholas Beck

Doctor of Philosophy

Department of Mathematics and Statistics
McGill University
Montréal, Québec, Canada

December 2020

A thesis submitted to McGill University in partial
fulfillment of the requirements of the degree of
Doctor of Philosophy

©Nicholas Beck, 2020

Abstract

This thesis explores different techniques for quantifying risk. Risk is a concept that permeates several branches of mathematics as well as fields for which statistical analysis is crucial. Examples include economics, actuarial sciences, finance, and hydrology. In particular, risk is examined here through the use of risk measures and statistical modeling.

First, we explore the estimation of multivariate risk measures. Specifically, we develop a semi-parametric estimation procedure for expectiles for extreme levels of risk. Multivariate expectiles and their extremes have been the focus of plentiful research recently, including estimation in extreme scenarios. However, current estimation techniques in an extreme value framework are restricted to the limiting cases of upper tail dependence: independence and comonotonicity. In this thesis, an alternative optimization problem along with a consistent estimation scheme is presented, which can solve for multivariate extreme expectiles without having to assume any underlying dependence structure. Specifically, we show that if the upper tail dependence function, tail index, and tail ratio can be consistently estimated, then one would be able to accurately estimate multivariate extreme expectiles. The finite-sample performance of this methodology is demonstrated using both simulated and real data.

Second, we build a hierarchical Bayesian model for quantifying the prevalence and magnitude of extreme surges on the Atlantic Coast of Canada with limited data. Generalized extreme value distributions are fitted marginally to surge observations at 21 buoys within our domain, and the modeling hierarchy includes latent Gaussian fields whose means and variances are driven by the atmospheric sea-level pressure and the distance between the buoys, respectively.

Incorporating this spatial information allows the model to share information between buoys and, more importantly, allows for spatial interpolation to be conducted at locations with no observations. Additionally, introducing a copula into the hierarchy allows for a continuous representation of extreme surges, which extends the inferential capabilities of the models beyond a site-by-site basis. Using realization of extreme surges simulated from the model and combining it with the physically driven tidal process at each location, we are able to predict potentially catastrophic water levels.

Résumé

Cette thèse porte sur les techniques de quantification du risque. La notion de risque intervient un peu partout en mathématiques et dans des domaines pour lesquels l'analyse statistique est cruciale, dont l'économie, l'actuariat, la finance et l'hydrologie. En particulier, le risque est abordé ici sous l'angle des mesures de risque et de la modélisation statistique.

Nous nous penchons d'abord sur l'estimation de mesures de risque multivariées. Nous proposons plus précisément une procédure d'estimation semi-paramétrique des expectiles aux niveaux de risque extrêmes. Les expectiles multivariés et leurs extrêmes ont été largement étudiés récemment, y compris leur estimation dans des scénarios extrêmes. Dans ce cadre, les techniques d'estimation actuelles ne sont toutefois adaptées qu'aux cas de dépendance caudale supérieure limite : l'indépendance et la comonotonie. Cette thèse met de l'avant un nouveau schéma d'optimisation et une méthode d'estimation convergente permettant le calcul des expectiles extrêmes multivariés sans égard à la structure de dépendance. Nous montrons de fait que si la fonction de dépendance caudale supérieure, l'indice de dépendance caudale et le ratio caudal peuvent être estimés de manière convergente, il en va de même pour les expectiles extrêmes multivariés. Le comportement de ces estimations est exploré à taille finie à l'aide de données réelles et simulées.

Dans un second temps, nous élaborons un modèle bayésien hiérarchique pour quantifier la prévalence et l'ampleur des ondes de tempête extrêmes sur la côte atlantique du Canada à partir de peu de données. Des lois de valeurs extrêmes généralisées univariées sont ajustées aux données relevées par 21 bouées et le modèle hiérarchique s'appuie sur des champs gaussiens latents dont les moyennes et les variances sont respectivement fonction de la pression atmosphérique au niveau de la mer et de la distance entre les bouées. L'ajout de ces informations spatiales au modèle permet le partage d'informations entre les bouées et, mieux encore, d'interpoler en des lieux où l'on ne dispose d'aucune donnée. L'inclusion d'une copule dans la hiérarchie conduit en outre à une représentation continue des ondes extrêmes, ce qui permet d'étendre l'inférence au-delà des sites. En simulant des ondes extrêmes à partir du modèle et en y incorporant le mouvement des marées, on peut ainsi prédire des niveaux d'eau potentiellement catastrophiques en tout lieu.

Acknowledgments

I would like to begin by thanking my supervisors, Professors Christian Genest and Mélina Mailhot, for their invaluable guidance and support throughout the last five years. Their advising has had a profound impact on my growth as a statistician, researcher, and writer and they have my deepest gratitude for the time they have spent helping me on this journey.

I would also like to express my heartfelt appreciation to my collaborators Professors Jonathan Jalbert (Polytechnique Montréal) and Elena Di Bernardino (Université Côte d’Azur). Their contributions to the works presented in this monograph were undoubtedly essential to my success. Additionally, I would like to thank Véronique Tremblay for her supervision and guidance during my research internship at Desjardins General Insurance Group.

To all my colleagues who have been there throughout this process, whether to bounce ideas off of, eat lunch, grab a coffee or discuss nothing of any importance, thank you. The sense of community I felt from the multitude of the graduate students I have crossed paths with has no doubt played a significant role in my experience at McGill. Many of these individuals I am happy to consider as more than just colleagues, but close friends.

To my family, whose endless support throughout all of my academic pursuits have left an indelible impression on my drive and determination, thank you so much and I love you all.

A special thanks goes out to Angela and Jackie, whom I could always count on for some sports talk, candy, or both.

The work presented here has been revised down to the last detail and, for that, I would like to express my gratitude to my examining committee: Professors Alexandra Schmidt, Russell Steele, and Gilles Stupfler (École nationale de la statistique et de l’analyse de l’information, Rennes) for their numerous insights and helpful comments.

Finally I would like to acknowledge the financial support I received throughout my PhD studies, notably from the Fonds de recherche du Québec – Nature et Technologies PhD Research Fellowship, the Schulich Graduate Fellowship, the Society of Actuaries through the James C. Hickman Scholar program as well as the funding I received through Dr. Genest’s research grants from the Canada

Research Chair Program, the Natural Sciences and Engineering Research Council of Canada, the Canadian Statistical Sciences Institute, and the Fonds de recherche du Québec – Nature et Technologies.

Contributions to original knowledge

Chapter 3:

In Chapter 3, the estimation procedure for extreme multivariate L^1 -expectiles, described in 3.4, consists entirely of original scholarship. In Section 3.4.1, an alternative optimization problem for L^1 -expectiles for levels of risk $\alpha \approx 1$ is presented. In Section 3.4.2, the procedure for solving an approximate version of this alternative formulation is described, including the required consistency arguments. These arguments are explicitly detailed in Section 3.4.3, including descriptions of the individual estimators required. In Section 3.4.4, we provide several simulations demonstrating the algorithms satisfactory finite-sample performance as well as a real data application. Potential improvements to the proposed methodology are provided in Section 3.4.5.

Chapter 4:

The application of a HBM to coastal surges on the Atlantic Coast of Canada, including the Gulf and Estuary of the St. Lawrence River, is original work. Moreover, the work helped to advance our understanding of the occurrence of floods in Atlantic Canada, including contributing to the launch of Canada's first comprehensive overland flood insurance product in March 2017.

Contributions of authors

I am solely responsible for the content of the introduction, background information and conclusion (Chapters 1, 2, and 5). The contributions to the original work presented in Chapters 3 and 4 can be described as follows:

Chapter 3

This is a joint work with Professors Méлина Mailhot and Elena Di Bernardino. They introduced me to the notions of elicibility and expectiles, with a focus on multivariate extremes. With their guidance, I formulated the alternative optimization problem from which extreme L^1 -expectiles could be estimated. I then took the lead in establishing the consistency of the estimation procedure and founded my results with simulation studies, which I was solely responsible for programming. Professors Mailhot, Di Bernardino, and Genest also contributed numerous helpful suggestions during the writing process.

Chapter 4

This is joint work done with Professors Christian Genest, Jonathan Jalbert and Méлина Mailhot. They helped me to conceptualize the hierarchical Bayesian model that we would use to model the extreme surges on the Atlantic Coast of Canada. Professor Jalbert, in particular, provided access to several resources through Ouranos from which I was able to build my understanding of hydrology. With their guidance, I built the complete hierarchical model, including the data extraction, data cleaning, preliminary analyses, modeling, diagnostic tools, simulations, applications, and all the required coding. Professors Genest, Mailhot, and Jalbert also provided invaluable feedback during the writing process.

Contents

1	Introduction	1
2	Background	4
2.1	Classic extreme value theory	4
2.1.1	Block maxima	4
2.1.2	Peaks-over-threshold	7
2.2	Multivariate extreme value theory	9
2.2.1	Copulas	11
2.2.2	Tail dependence	14
2.3	Risk measures	16
2.3.1	Univariate risk measures	16
2.3.2	Multivariate extensions	17
2.3.3	Properties of risk measures	21
2.3.4	Estimation of risk measures	24
3	Extreme Multivariate Expectiles	28
3.1	Elicitability	28
3.2	Expectiles	30
3.3	Extensions to higher dimensions	31
3.4	Estimation of extremes for L^1 -expectiles	33
3.4.1	Optimization	36
3.4.2	Approximate optimization problem	39
3.4.3	Consistency	40
3.4.4	Examples	49
3.4.5	Discussion	57
3.5	Conclusion	58
Appendices		61
3.A	Proof of Proposition 3.6	61
4	Environmental Extremes	62
4.1	Geostatistics	62
4.2	Spatial extremes	65
4.2.1	Hierarchical models	65
4.2.2	Residual dependence	68
4.3	Flood risk in Atlantic Canada	71
4.3.1	The data	72
4.3.2	Tidal modeling	73
4.3.3	Data selection and augmenting	78
4.3.4	The statistical model	81
4.3.5	Model estimation	87
4.3.6	Application	92
4.3.7	Discussion	100
4.4	Conclusion	101
Appendices		103
4.A	List of Stations by ID	103
4.B	Justification of 15% rule for including maxima	103
4.C	Effect on shape parameter given different cutoffs for data augmenting	105

5 Final Conclusion and Future Work	106
List of Publications	108
Bibliography	109
Acronyms and Notation	123

List of Figures

2.1	Comparison of lower and upper orthant VaR for Weibull margins joined by a Gumbel with varying Kendall's τ	18
2.2	Performance of $\widehat{\text{VaR}}_{0.75}(X)$ as the sample size increases	25
3.1	Demonstration of BFGS algorithm for solving (3.13) for varying α and dependence structures. Here, $X_i \sim b_i t_2$ where t_2 is a student t -distribution with 2 degrees of freedom and $b_i = 2^{i-1}$ for $i \in \{1, 2, 3\}$. The dotted green line represents the true values as described in Maume-Deschamps et al. (2018).	39
3.2	Comparison between $\hat{\lambda}_n^{23}(0.5, 1.5)$, $\hat{\lambda}_\beta^{23}(0.5, 1.5)$, $\hat{\lambda}_{\mathbf{x}}^{23}(0.5, 1.5)$ and $\lambda^{23}(0.5, 1.5)$. Note that for these estimates $\ell_\theta = \ell_i = n^{0.75}$. The margins are Pareto, specifically, $X_i \sim \text{P}(2, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$	42
3.3	Performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_\beta^{23}(c_2 t^{-\theta}/c_3, 1) dt$ for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_\lambda = n^q$ for $q \in \{0.1, 0.2, \dots, 0.9\}$. They are compared to the true value $\int_{\beta_2/\beta_3}^{\infty} \lambda^{23}(c_2 t^{-\theta}/c_3, 1) dt$ under Pareto margins with a comonotonic dependence structure (green horizontal line). Here $X_i \sim \text{P}(2, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$	44
3.4	Performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_n^{23}(c_2 t^{-\theta}/c_3, 1) dt$ as in Proposition 3.6 for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_\lambda = n^q$ for $q \in \{0.1, 0.2, \dots, 0.9\}$. They are compared to the true value $\int_{\beta_2/\beta_3}^{\infty} \lambda^{23}(c_2 t^{-\theta}/c_3, 1) dt$ under Pareto margins with a comonotonic dependence structure (green horizontal line). Here $X_i \sim \text{P}(2, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$	45
3.5	Performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_\beta^{23}(\hat{c}_2 t^{-\hat{\theta}}/\hat{c}_3, 1) dt$ for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_\lambda = n^q$ for $q \in \{0.1, 0.2, \dots, 0.9\}$. They are compared to the true value $\int_{\beta_2/\beta_3}^{\infty} \lambda^{23}(c_2 t^{-\theta}/c_3, 1) dt$ under Pareto margins with a comonotonic dependence structure (green horizontal line). Here $X_i \sim \text{P}(2, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$	48

3.6	Boxplots for the three-dimensional Pareto simulations. Here $X_i \sim P(3.5, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$. Results for independent margins (left) and comonotonic margins (right). True values Θ^\perp and Θ^+ (dashed green lines) provided for comparison.	52
3.7	Hill Plots for estimating $\hat{\theta}$ using BMI measurements.	56
3.8	Extreme Expectiles for Pima Indians Diabetes Study. Included is the extreme expectile at level $\alpha = 1 - 1/n$ (red triangle).	56
3.9	Boxplots for the three-dimensional Pareto simulations using L-BFGS-B. Margins are comonotonic and remaining settings are equivalent to Figure 3.6.	59
4.1	Map of Atlantic Canada with labels identifying the location of the 21 monitoring stations used in this study and listed in Table A.	72
4.2	Tide comparison at Station 65 (St. John, NB). (a) June 2013. (b) June 2014.	75
4.3	Model fit comparison for different groups of harmonic constituents at each of the 21 monitored locations. Location codes are provided in the Appendix. (a) Adjusted coefficient of determination R^2 . (b) RMSE.	77
4.4	Tidal fitting examples at Stations 365 (Yarmouth, NS; left) and 490 (Halifax, NS; right). (a) Groundhog Day Gale, February 1–3, 1976. (b) Hurricane Juan, September 28–30, 2003.	78
4.5	Asymptotic 95% confidence intervals for the MLE's fitted to each of the selected 21 stations marginally.	80
4.6	Q-Q plot comparison for constrained MLE model at Station 65 (St John, NB).	81
4.7	Boxplots showing the difference between raw correlation estimates and structured correlation estimates based on a Matérn correlogram with parameter $\kappa \in \{0.5, 1.5, 2.5\}$ and $\kappa \rightarrow \infty$	83
4.8	Sea-level pressure provided by CRCM5 for Atlantic Canada. At each grid point, the mean annual minimum pressure is given in kPA. It is the use of this variable as a covariate at the process level of the Bayesian hierarchical model that enables spatial interpolation at unmonitored locations.	84
4.9	Examples of the well mixing of chains for the winning model, $C_{9,\Xi}$ with spatially varying log-scale parameters. Chains for μ_1 (top left), σ_1 (top right), ξ (bottom left) and ρ_Ξ (bottom right) are shown for a variety of starting parameters. The filled in circles represent the starting points of each chain. Here, s_1 corresponds to station 491.	89
4.10	Maps showing the median values of the posterior distribution for the location parameters of the hierarchical Bayesian model at the monitoring stations (upper panel) and in 1361 locations along the Atlantic coast of Canada (lower panel).	94

4.11	Maps showing the median values of the posterior distribution for the scale parameters of the hierarchical Bayesian model at the monitoring stations (upper panel) and in 1361 locations along the Atlantic coast of Canada (lower panel).	95
4.12	Maps showing the values of the standard deviation of the posterior distribution for the location (upper panel) and scale (lower panel) parameters in 1361 locations along the Atlantic Coast of Canada.	96
4.13	Maps showing the 97.5% (upper panel) and 99.9% (lower panel) quantiles of the posterior distribution for the surge heights at 1361 locations along the Atlantic Coast of Canada.	97
4.14	Boxplots representing the distribution of water levels at three locations under scenarios (a) and (b), based on 10,000 realizations of the surge model.	99
4.15	Variability in the fitted constrained shape parameter for different quantiles q_α acting as the cutoff for new data.	105

List of Tables

3.1	Median results of simulation study for Pareto margins in three dimensions. Standard deviations for each estimate are presented in parentheses. Here $X_i \sim P(3.5, 1.25(1+i))$ for $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$	51
3.2	Median results of simulation study for Pareto margins in five dimensions linked with the independence copula. Standard deviations for each estimate are presented in parentheses. Here $X_i \sim P(3.5, 1.25(1+i))$ for $i \in \{1, \dots, 5\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$	53
3.3	Analogous results to Table 3.1 using the full parametric estimators for $\hat{\Theta}$ and \hat{e}_α for independent and comonotonic random vectors. Here $X_i \sim P(3.5, 1.25(1+i))$ for $i \in \{1, \dots, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$	55
3.4	Microbenchmarking comparison for BFGS vs L-BFGS-B. Estimating $e^\perp(\mathbf{X})$ with $d = 3$ and $n = 250$ and 100 replications. Margins Pareto distributed, $X_i \sim P(3.5, 1.25(1+i))$. q_α is the α -level quantile of the distribution of run-times. Entries are presented in seconds. Microbenchmarking was completed using 4 cores from a Intel Core i7-5500U CPU with 2.40 GHz clock speed.	58
4.1	Average SSE of prediction using k -folds cross validation with $k = 8$ for collections of harmonic constituents hc4, hc5, hc7, hc37, hc60 and hc114. For more information on hc4, hc7, hc37, hc60 and hc114, see the TideHarmonics R package. The collection hc5 is simply hc4 with the larger lunar elliptic semidiurnal constituent N2 included.	76
4.2	Model comparison using WAIC for different choices of copulas and log-scale parameters that are either equal or freely varying. Reported in the table are $-2\widehat{\text{lppd}}$, $\widehat{p}_{\text{WAIC},j}$ and $\widehat{\text{WAIC}}_j$ for $j \in \{1, 2\}$, and the value of the MPSRF \hat{R}_M	88

4.3	Comparison of the return level for various return periods (RP, in years) computed with either the individual margins of the Hierarchical Bayesian Model (HBM) or the maximum likelihood estimate based on that single location. Also included are asymptotic 95% confidence intervals for the MLE and 95% credible intervals derived from the HBM, as well as the percent relative change given by $\%RC = 100 \times (HBM - MLE)/MLE$. As stations 3200 and 3260 are unmonitored, the MLEs, and hence the $\%RC$, cannot be computed.	91
4.4	Return level for various return periods (RP, in years) computed at 10 stations using the joint distribution of surges from the Hierarchical Bayesian Model (HBM). The names of the stations are specified by ID in Table A.	92
A	List of monitored stations and their locations, along with the number of annual maxima available at each site.	103

1

Introduction

In all facets of life, individuals and collectives are exposed to risk. Colloquially, risk is viewed in a negative light as the possibility that something bad will happen. Moreover, risk can manifest itself in several ways. Actuaries, hydrologists, and economists are just a few examples of professions in which practitioners must account for risk on a day-to-day basis.

A big component of risk is how it relates to uncertainty. In this sense, statistics presents itself as the perfect tool for quantifying risk. By developing statistical techniques for quantifying the impact and propensity of risky events, one can have a better understanding of how to prepare for, mitigate, and respond to such outcomes. Furthermore, it is often the events that are both rare and impactful which are of greatest interest, e.g., economic recession, financial turmoil, and environmental disasters. These events, which are all extreme in nature, can pose catastrophic consequences for the wellbeing and safety of humanity.

A first step in understanding these events is to model the mechanisms which drive them using statistical distributions. Fitting a particular distribution to a collection of observations can provide insight into their behavior by relating them to well understood random objects. Additionally, for decision making purposes it can be beneficial to supplement the general structure of distributions with quantities that can provide additional information for a given situation. Risk measures are one such tool for accomplishing this task as they can provide clear and digestible summaries of the underlying stochastic behavior.

In this thesis, we explore techniques for measuring risk using both statistical modeling and risk measures. In particular, we are interested in the occurrence of extreme events for several interactive processes. Extreme events are difficult to model and quantify because, by nature, they are rare. This results in having few observations available for modeling purposes which can produce poor fitting models with little predictive accuracy. Moreover, selecting the wrong distribution can have drastic effects on the resulting inference. Risk measures stemming from these ill-fitted distributions pose the same drawbacks. This problem is further exacerbated when one is interested in the behavior of several related quantities. In this scenario, one must also account for the way concurrent events can

impact one another, making the scarcity of data all the more impactful.

The first way we explore multivariate extremes is through the estimation of multivariate expectiles. In particular, we discuss how to consistently estimate L^1 -expectiles for elevated levels of risk using observed data. While estimation of these measures has been discussed in the literature (Maume-Deschamps et al., 2017, 2018), the case of extremes is only addressed under pre-imposed assumptions that the underlying dependence structure is asymptotic independence or comonotonicity. These dependence relations present limiting scenarios of tail dependence which do not represent the full breadth of possibilities when discussing dependence of extreme events. The estimator we have developed can capture any underlying dependence structure without any prior distributional assumptions, thereby mitigating the risks of model misspecification.

In the second problem, we return to statistical modeling, specifically in the context of spatial extremes with environmental applications. Spatial extremes have a rich literature in the environmental sciences (Davison et al., 2012; Tawn et al., 2018; Davison et al., 2019). Using techniques from spatial statistics, one can characterize observations using the physical and atmospheric properties of the locations at which they are recorded, including, for instance, the distances between them. By incorporating these properties into the modeling process, inference on the nature of environmental processes can be elevated, opening the door to several compelling techniques, such as spatial interpolation. Combining these methods with extreme value theory, one can effectively model extremes of spatially driven processes. In particular, we are interested in applying some of these techniques for quantifying the risk of extreme surges in Atlantic Canada.

Chapter 2 provides the necessary background information for the discussions in this thesis. Sections 2.1 and 2.2 introduce extreme value theory in the univariate and multivariate settings, respectively, including block-maxima and peaks-over-threshold techniques. In particular, Section 2.2 discusses different ways to quantify dependence between random variables, including copulas and measures of association. Section 2.3 formally introduces risk measures, including some examples in the univariate and multivariate settings, important properties for risk measures to have in practice, and a discussion of how to estimate risk measures efficiently using data.

In Chapter 3, we begin by adding to the lexicon of desirable properties for risk measures and introduce elicibility, a property which has garnered increased interest from an actuarial standpoint in recent literature (Bellini and Bignozzi, 2015; Ziegel, 2016; Herrmann et al., 2018). Building on elicibility, we turn our attention to expectiles risk measures which are elicitable while possessing several other sought after properties. Beginning in the univariate setting, we then extend the discussion of expectiles to dimensions $d \geq 2$ with several different characterizations. Finally, we develop a consistent semi-parametric estimation scheme for L^1 -expectiles when considering extreme levels of risk.

Introduction

Chapter 4 focuses on spatial environmental extremes. Techniques discussed include classical geostatistics, max-stable processes, and Bayesian hierarchical modeling for problems in the environment. In particular, using a hierarchical Bayesian model, we model the propensity of extreme surges in the Atlantic Coast of Canada. Using this information, we hope to contribute to the rapidly developing flood risk prevention infrastructure in Canada. In fact, the work presented here helped contribute to the launch of a comprehensive overland flood insurance product by a prominent Canadian insurer in March of 2017.

Finally, Chapter 5 summarizes the impacts of the works presented in this thesis and briefly discusses some venues for further investigation.

2

Background

In this chapter, the important background information motivating the work in this thesis is summarized. In particular, Section 2.1 introduces classical extreme value theory, including the block-maxima and peaks-over-threshold methods of modeling extreme observations. Section 2.2 extends these univariate notions to dimensions $d \geq 2$, including a discussion of general dependence structures and measures of tail dependence. Finally, Section 2.3 introduces risk measures, including properties and estimation in the univariate as well as multivariate setting.

2.1 Classic extreme value theory

When one is interested in understanding the potential impacts of a stochastic process, it is natural to be curious about particularly elevated, or “extreme”, levels of risk. In fact, the study of these extreme phenomena is a widely studied topic in statistics with a rich research history. Here we take the opportunity to introduce some of the concepts in extreme value theory. For a more in-depth discussion of extremes, we suggest the works of [Coles \(2001\)](#), [Beirlant et al. \(2004\)](#), [de Haan and Ferreira \(2006\)](#), [Resnick \(2013\)](#), among others.

2.1.1 Block maxima

The first way we consider the study of extremes is through the study of maxima. Specifically, we consider the maximum of a pre-specified number of observations, called a block. Let X_1, X_2, \dots be independent and identically distributed (iid) random variables (rv) with common marginal cumulative distribution function (cdf) F . For a block with n observations, denote the maximum of them $M_n = \max(X_1, \dots, X_n)$. For iid random variables, it is easy to see that the distribution of M_n can be written

2.1 Classic extreme value theory

$$\begin{aligned}
 F_{M_n}(x) &= \Pr(M_n \leq x) = \Pr\{\max(X_1, \dots, X_n) \leq x\} \\
 &= \Pr(X_1 \leq x, \dots, X_n \leq x) \\
 &= \Pr(X_1 \leq x) \cdots \Pr(X_n \leq x) \\
 &= F^n(x).
 \end{aligned}$$

The block maxima method seeks to study the behavior of M_n as the block size $n \rightarrow \infty$. However, as n approaches infinity, F_{M_n} will become degenerate. The following theorem, summarizing results established independently by [Fisher and Tippett \(1928\)](#), [Von Mises \(1936\)](#) and [Gnedenko \(1943\)](#), formalizes a powerful result about the distribution of M_n .

Theorem 2.1. *Let X_1, X_2, \dots be a sequence of iid rv's with common marginal cdf F and denote $M_n = \max(X_1, \dots, X_n)$. If sequences $a_n > 0$, $b_n \in \mathbb{R}$ exist, such that for all $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = G_\xi\{(x - \mu)/\sigma\},$$

for some non-degenerate distribution G_ξ , then G_ξ is of the form

$$G_\xi(x) = \begin{cases} \exp\{-(1 + \xi x)^{-1/\xi}\} & \text{if } \xi \neq 0, \\ \exp\{\exp(-x)\} & \text{if } \xi = 0, \end{cases} \quad (2.1)$$

for location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$. In this case, we say F is in the domain of attraction of G_ξ , denoted $F \in DA(G_\xi)$. Moreover, G_ξ is known as a generalized extreme value distribution (GEVD), comprising three families of distributions: The Fréchet ($\xi < 0$), the Gumbel ($\xi = 0$) and the Weibull ($\xi < 0$).

The significance of the shape parameter ξ should be stressed. First, ξ , also known as the extreme value index, entirely determines the behavior of the tails of G_ξ . It quantifies how “heavy” the tails of G_ξ are. For instance, the shape parameter determines the domain on which the distribution G_ξ is defined, which can be important in applications. The three possibilities are:

- (i) $x \in [0, \infty)$ for $\xi > 0$,
- (ii) $x \in (-\infty, 0]$ for $\xi < 0$, and
- (iii) $x \in (-\infty, \infty)$ for $\xi = 0$.

Second, the shape ξ also relates EVT to the concept of regular variation. A measurable function f is regularly varying at a with tail index θ , denoted $f \in RV_\theta(a)$, if

$$\lim_{t \rightarrow a} f(tx)/f(t) = x^\theta$$

2.1 Classic extreme value theory

for all $x > 0$. If $\theta = 0$, the function f is called *slowly varying*. All regularly varying functions f can be written as

$$f(x) = x^\theta L(x),$$

where L is a slowly varying function. For GEVD's, it can be shown that $F \in DA(G_\xi)$ if and only if $1 - F \in RV_{-1/\xi}(\infty)$.

One of the properties which characterize GEVD's is that of max-stability. A distribution F is said to be max-stable if, for every $n \in \{2, 3, \dots\}$, there are constants α_n and β_n such that

$$F^n(\alpha_n z + \beta_n) = F(z).$$

Max-stability implies that the distribution of the maxima is equivalent to the distribution of the margins, up to a change of scale and location. The following result of [Leadbetter et al. \(1983\)](#) establishes max-stability as a defining property of GEVDs.

Theorem 2.2. *A distribution is max-stable if, and only if, it is a GEVD.*

In data analysis, it is important to realize that the asymptotic requirements of [Theorem 2.1](#) will never be verifiable. As such, parameter estimation can pose a challenge for GEVD's. In particular, the choice of block-size n is critical. If blocks are chosen to be too small, the asymptotic requirements of [Theorem 2.1](#) will not be met and the model fit is likely to be poor. If the blocks are too large, too few observations will be considered, yielding large variances in the estimates. Depending on the type of observations, common choices for block size are daily, monthly or yearly. For example, this could equate to the study of maximal rainfall on the daily, monthly, or yearly time horizon.

To fit the parameters of a GEV, several techniques are possible; see, e.g., [Hosking et al. \(1985\)](#) or [Smith \(1985\)](#). The most commonly used technique, available in several packages of the statistical computing language **R**, is maximum likelihood estimation (MLE). In general, MLE centers around finding the collection of parameters which optimize the underlying parametric distribution relative to the observed data. For a collection of block-maxima $\mathbf{z} = (z_1, \dots, z_m)$ from a rv X having a GEVD, the log-likelihood can be written

$$\ell(\mu, \sigma, \xi; \mathbf{z}) = \begin{cases} -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^m \log(1 + \xi \dot{z}_i) - \sum_{i=1}^m (1 + \xi \dot{z}_i)^{-1/\xi} & \text{if } \xi \neq 0, \\ -m \log \sigma - \sum_{i=1}^m \dot{z}_i - \sum_{i=1}^m \exp(-\dot{z}_i) & \text{if } \xi = 0, \end{cases} \quad (2.2)$$

where $\dot{z}_i = (z_i - \mu)/\sigma$. Additionally, for $\xi \neq 0$ we require $1 + \xi \dot{z}_i > 0$ for all $i \in \{1, \dots, m\}$ else the log-likelihood is $-\infty$. Note that while no analytical solution for the minimum of [\(2.2\)](#) exists, it can be solved for using standard numerical optimization algorithms.

2.1 Classic extreme value theory

Fitted models are often used to provide inference about the underlying random process. While the approximated parameters, in particular ξ , provide some insight into the natures of extremes, interest exists beyond these values. To this end, two quantities of great import in extreme value analysis are the *return level* and the *return period*. These quantities are linked with the following expression:

$$\frac{1}{p} = \Pr(X > z_p),$$

where z_p and p are the return level and return period, respectively. Note that the units of the return period are determined by the block size. For example, if the block size is one year, the return-level is commonly interpreted as the value of X which is expected to be surpassed on average once every p years. More precisely, z_p is the value for which there is a $1/p$ chance of it being surpassed each year. Simply, z_p is the $(1 - 1/p)$ -level quantile of X , i.e.,

$$q_{1-1/p}(X) = \inf \{x \in \mathbb{R} : \Pr(X \leq x) \geq 1 - 1/p\}.$$

For a GEVD, there exists a closed form for the return level, namely,

$$z_p = \begin{cases} \mu - \sigma [1 - \{-\log(1 - 1/p)\}] / \xi & \text{if } \xi \neq 0, \\ \mu - \sigma \log \{-\log(1 - 1/p)\} & \text{if } \xi = 0. \end{cases}$$

These quantities are of particular interest in fields such as finance, insurance and environmental sciences. Consider the construction of a dam on river bordering a small town; if one were to consider daily water levels for the river with a block size of one year, the ideal height at which to build the dam to protect the community for the next 50 years, on average, would be z_{50} return level.

While the block maxima method is often seen in the study of extremes, a drawback is that it can be wasteful of data as there could be several extreme values in a block. Two methods which can mitigate this issue are considering the r largest observations in each block, see, e.g., [Coles \(2001\)](#) Section 3.5, or by considering threshold exceedances. The latter is outlined in the following section.

2.1.2 Peaks-over-threshold

A second way to study extremes, which can curb the issue of data scarcity, is through the peaks-over-thresholds (POT) approach. Instead of considering the singular maximum, or r largest observations, for a pre-specific block size, consider any observation above a pre-described *threshold* as extreme. For a rv X and a threshold u , $X - u$ is called the exceedance of X over u . The POT approach is interested in the asymptotic distribution of $X - u | X > u$, which we will denote F_u . The following theorem provides an asymptotic result for F_u as u approaches infinity.

2.1 Classic extreme value theory

Theorem 2.3 (Coles (2001), Theorem 4.1). *Let X_1, X_2, \dots be sequence of iid rv with common marginal cdf F . If $F \in DA(G_\xi)$ then, for large enough u ,*

$$F_u(x) \approx H(x) = \begin{cases} 1 - (1 + \xi x/\sigma_u)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\sigma_u) & \text{if } \xi = 0, \end{cases}$$

where $\sigma_u = \sigma + \xi(u - \mu)$. The distribution H is called a generalized Pareto distribution (GPD).

If in a given block, there are several observations one would like to consider extreme, it is clear to see the advantage that the POT methodology has over block-maxima. For an appropriately chosen threshold u , one can keep a larger portion of the available observations for modeling purposes. However, in practice, the choice of threshold is not straightforward. The issues are analogous to those for modeling block-maxima. If the threshold is set too low, the asymptotic requirements of Theorem 2.3 will not be satisfied. If the threshold is too high, too few observations will be considered, producing large variances in model fitting. One way to select an appropriate threshold is through a *mean residual life plot*. If $Y_{u_0} = X - u_0 | X > u_0$ has a GPD with parameters σ_{u_0} and ξ then, for $\xi < 1$,

$$\mathbb{E}(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1 - \xi}.$$

Given Theorem 2.3, the distribution of Y_u should also follow a GPD for all $u > u_0$, i.e.,

$$\mathbb{E}(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} - \xi u_0}{1 - \xi} + \frac{\xi}{1 - \xi} u. \quad (2.3)$$

Equation (2.3) thus implies that for $u > u_0$, $\mathbb{E}(Y_u)$ is linear in u . It stands to reason then that the smallest threshold u_0 for which the average of the exceedances is linear in u for $u > u_0$ is a reasonable choice. For more information on threshold selection, the interested reader is referred to, e.g., Dupuis (1999), Coles (2001), and Embrechts et al. (1997). With a threshold selected, and for observed exceedances $\mathbf{y} = (y_1, \dots, y_{n_u})$, GPD parameter fitting can be done via MLE with log-likelihood

$$\ell(\sigma, \xi; \mathbf{y}) = \begin{cases} -k \log \sigma - (1 - 1/\xi) \sum_{i=1}^{n_u} \log(1 + \xi y_i/\sigma) & \text{if } \xi \neq 0, \\ -n_u \log \sigma - \sigma^{-1} \sum_{i=1}^{n_u} y_i & \text{if } \xi = 0. \end{cases}$$

Additionally, in the case $\xi \neq 0$, we require $1 + \xi y_i/\sigma > 0$ for all $i \in \{1, \dots, n_u\}$, else the log-likelihood is $-\infty$.

Similarly to the GEVD's, return-levels and return periods for GPD's are useful metrics when studying threshold exceedances. Deriving these quantities requires one to not only consider the

2.2 Multivariate extreme value theory

distribution H , but also the probability that X is in fact above the threshold u . Thus, the return level $z_p > u$ is defined as the value of X such that

$$\begin{aligned} \frac{1}{p} &= \Pr(X > z_p) = \Pr(X > u) \Pr(X > z_p | X > u) \\ &= \zeta_u \{1 - H(z_p)\} \\ &= \zeta_u \left(1 + \xi \frac{z_p - u}{\sigma}\right)^{-1/\xi}, \end{aligned}$$

where $\zeta_u = \Pr(X > u)$. Thus, z_p can be written

$$z_p = \begin{cases} u + \sigma \{(\zeta_u p)^\xi - 1\} / \xi & \text{if } \xi \neq 0, \\ u + \sigma \log(\zeta_u p) & \text{if } \xi = 0. \end{cases}$$

2.2 Multivariate extreme value theory

The theory presented in Section 2.1 centers around the concept of a single sequence of mutually independent observations. While the results of Theorem 2.1 are indeed powerful, they say nothing about the inter-relationships between several connected extreme values. For example, if one were interested in how extreme gains/losses of individual stocks impacted one another, a more sophisticated theory would be required. Building from the concepts of Section 2.1, let $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$, for $i \in \{1, \dots, n\}$, be an iid collection of independent random vectors from a d -dimensional cdf F with marginal cdf's F_1, \dots, F_d . Denote the j th component-wise maxima $M_{jn} = \max(X_{1j}, \dots, X_{nj})$. Suppose there exists sequences $\mathbf{a}_n = (a_{1n}, \dots, a_{dn}) > \mathbf{0}$ and $\mathbf{b}_n = (b_{1n}, \dots, b_{dn}) \in \mathbb{R}^d$ such that for all $\mathbf{x} \in \mathbb{R}^d$

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{M_{1n} - b_{1n}}{a_{1n}} \leq x_1, \dots, \frac{M_{dn} - b_{dn}}{a_{dn}} \leq x_d \right) = \lim_{n \rightarrow \infty} F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) = G(\mathbf{x}), \quad (2.4)$$

then G is called a *multivariate extreme value distribution* (MEVD) and one says $F \in DA(G)$. In this scenario, the marginal cdf's F_1, \dots, F_d of F are GEVD's, i.e., $F_i \in DA(G_{\xi_i})$, where G_{ξ_i} are of the form (2.1).

The following proposition from Resnick (2013), extends the notion of max-stability and its relation to extreme value distributions to the multivariate setting. First, in the multivariate setting, a d -dimensional function G is max-stable if there exists sequences $\mathbf{a}_k > \mathbf{0}$ and $\mathbf{b}_k \in \mathbb{R}^d$ such that for every $k \in \{2, 3, \dots\}$ and all $\mathbf{x} \in \mathbb{R}^d$

$$G^k(\mathbf{a}_k \mathbf{x} + \mathbf{b}_k) = G(\mathbf{x}).$$

Proposition 2.4 (Resnick (2013), Proposition 5.9). *The class of MEVD's is precisely the class of*

2.2 Multivariate extreme value theory

max-stable functions with non-degenerate marginals.

While for GEVD's there is a well defined parametric family, this is not the case for the class MEVD's. An MEVD is characterized by its *exponent measure*, μ on $[0, \infty]^d$, such that

$$G(\mathbf{x}) = \exp\{-\mu([-\infty, \mathbf{x}]^c)\},$$

where A^c represents the complement of a set A . For simplicity, the marginal distributions of X_1, \dots, X_d are often assumed to be unit Fréchet, i.e., $F_1(x) = \exp(-x^{-1})$. This simplification does not impact the discussion of G (Resnick (2013), Proposition 5.10). For X_1, \dots, X_d having unit Fréchet distributions, the exponent measure can be written $V(\mathbf{x}) = \mu([0, \infty] \setminus [0, \mathbf{x}])$ and the joint distribution G has the form

$$G(\mathbf{x}) = \exp\{-V(\mathbf{x})\},$$

for $\mathbf{x} > \mathbf{0}$, where

$$V(\mathbf{x}) = \int_{S_d} \max(\omega_1/x_1, \dots, \omega_d/x_d) dH(\boldsymbol{\omega})$$

and $S_d = \{\boldsymbol{\omega} \in [0, 1]^d : \omega_1 + \dots + \omega_d = 1\}$ is called the *unit simplex*. The measure H , called the *spectral measure*, on S_d satisfies

$$\int_{S_d} \omega_i dH(\boldsymbol{\omega}) = 1$$

for all $i \in \{1, \dots, d\}$. An advantage of this formulation is that $V(t\mathbf{x}) = t^{-1}V(\mathbf{x})$, i.e., it is homogeneous of order -1 . Similar results have been established for the distribution of multivariate exceedances; see, e.g., Coles (2001), McNeil et al. (2015). For this thesis, we focus primarily on multivariate block maxima.

A concept related to the notion of multivariate extremes, which will be of great importance in this thesis, is that of multivariate regular variation (MRV).

Definition 2.1 (Multivariate Regular Variation). *The distribution of a random vector \mathbf{X} on $[0, \infty]^d$ is said to be regularly varying if there exists a non-null Radon measure $\mu_{\mathbf{X}}$ on Borel σ -algebra \mathcal{B}_d on $[0, \infty]^d \setminus \{\mathbf{0}\}$, and a normalization function $a : \mathbb{R} \mapsto \mathbb{R}$ which satisfies $\lim_{x \rightarrow +\infty} a(x) = +\infty$, and such that*

$$u\mathbb{P}\left\{\frac{\mathbf{X}}{a(u)} \in \cdot\right\} \xrightarrow[u \rightarrow +\infty]{\nu} \mu_{\mathbf{X}}(\cdot).$$

Definition 2.2 (MRV equivalent definitions). *Let \mathbf{X} be a random vector on \mathbb{R}^d . The following definitions are equivalent:*

- The vector \mathbf{X} has regularly varying tail of index θ .

2.2 Multivariate extreme value theory

- There exists a finite measure μ on the unit sphere, $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$, and a normalizing function $a : (0, \infty) \mapsto (0, \infty)$ such that

$$\lim_{t \rightarrow +\infty} \mathbb{P} \left\{ \|\mathbf{X}\| > xa(t), \frac{\mathbf{X}}{\|\mathbf{X}\|} \in \cdot \right\} = x^{-\theta} \mu(\cdot)$$

for all $x > 0$. The measure μ depends on the chosen norm; it is called the spectral measure of \mathbf{X} .

- There exists a finite measure μ on the unit sphere \mathbb{S}^{d-1} , a slowly varying function L , and a positive real $\theta > 0$ such that

$$\lim_{x \rightarrow +\infty} \frac{x^\theta}{L(x)} \mathbb{P} \left(\|\mathbf{X}\| > x, \frac{\mathbf{X}}{\|\mathbf{X}\|} \in B \right) = \mu(B)$$

for all $B \in \mathcal{B}(\mathbb{S}^{d-1})$ with $\mu(\partial B) = 0$, where \mathcal{B} represents the Borel σ -algebra and ∂B the boundary of the set B .

For more information on the concept of regular variation, see, e.g., [Resnick \(2007\)](#) and [Resnick \(2013\)](#). In particular, the equivalence between the convergence of multivariate extremes and multivariate regular variation has been stated in either [Resnick \(2007\)](#), Proposition 7.1 or [Resnick \(2013\)](#), Proposition 5.15

2.2.1 Copulas

The multivariate distribution of extremes can also be expressed in terms of copulas. In what follows we provide a brief overview of copula theory. For a comprehensive study of copula theory, see, e.g., [Nelsen \(2006\)](#) and [Joe \(2014\)](#). Copulas represent a powerful way in which to model the dependence between random vectors. In essence, a copula is a multivariate distribution on a collection of standard uniform random variables, i.e., for $U_1, \dots, U_d \sim \mathcal{U}(0, 1)$, a copula is a function $C : [0, 1]^d \mapsto [0, 1]$ such that, for all $u_1, \dots, u_d \in [0, 1]$,

$$C(u_1, \dots, u_d) = \Pr(U_1 \leq u_1, \dots, U_d \leq u_d).$$

A copula will also possess the following properties:

- (i) $C(u_1, \dots, u_d) = 0$ whenever $u_j = 0$ for at least one $j \in \{1, \dots, d\}$.
- (ii) $C(u_1, \dots, u_d) = u_j$ if $u_i = 1$ for all $i \in \{1, \dots, d\} \setminus \{j\}$.
- (iii) C is d -nondecreasing on $[0, 1]^d$. That is, for each hyperrectangle $R = \prod_{j=1}^d [a_j, b_j] \subset [0, 1]^d$,

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1 + \dots + i_d} C(u_{1i_1}, \dots, u_{di_d}),$$

2.2 Multivariate extreme value theory

where for $j \in \{1, \dots, d\}$, $u_{j1} = a_j$ and $u_{j2} = b_j$.

In fact, copulas can be used to quantify the dependence between a much wider array of random variables. This result is formalized in the following theorem, attributed to [Sklar \(1959\)](#).

Theorem 2.5 (Sklar's Theorem). *Consider a collection of rv's X_1, \dots, X_d with cdf's F_1, \dots, F_d , respectively, for $d \geq 2$ and joint cdf F . There exists a function $C : [0, 1]^d \mapsto [0, 1]$, called a copula, such that*

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

for all $x_1, \dots, x_d \in \mathbb{R}$. Moreover, if the margins are all continuous, then the function C is unique. Conversely, for a copula C and a collection of marginal cdf's F_1, \dots, F_d , the function F , defined by

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

is a joint cdf for X_1, \dots, X_d .

The proof of this theorem stems from the probability integral transform. For a rv X with continuous cdf F , it can be shown that $F(X) \sim \mathcal{U}(0, 1)$. Thus, for any collection of rv's, the copula can be written

$$C(u_1, \dots, u_d) = \Pr\{X_1 \leq F_1^{-1}(u_1), \dots, X_d \leq F_d^{-1}(u_d)\}.$$

It is important to note that the dependence structure of the rv's X_1, \dots, X_d is entirely determined by the copula C . Moreover, the range of possible dependence strengths is delineated by the Fréchet–Hoeffding bounds. Specifically, any copula C satisfies

$$\max(0, u_1 + \dots + u_d + 1 - d) \leq C(\mathbf{u}) \leq \min(u_1, \dots, u_d). \quad (2.5)$$

Returning to the study of extremes, a copula for a vector extreme events is partly established by the univariate theory. If the margins of G are non-degenerate, then they must be GEVD's. As these distributions are continuous, Sklar's Theorem tells us that G has a unique copula representation. The following theorem introduces extreme value (EV) copulas using its defining characteristic.

Theorem 2.6 ([McNeil et al. \(2015\)](#), Theorem 7.44). *If (2.4) holds for some joint cdf F and some G with GEV margins, then the unique copula C of G satisfies*

$$C^{1/t}(\mathbf{u}^t) = C(\mathbf{u})$$

for all $t > 0$ and $\mathbf{u} \in [0, 1]^d$.

This result can be used to extend the notion of maximum domain of attraction to copulas, denoted the copula domain of attraction (CDA).

2.2 Multivariate extreme value theory

Theorem 2.7 (McNeil et al. (2015) Theorem 7.48). *Let F be a joint cdf of a random vector $\mathbf{X} = (X_1, \dots, X_d)$ with continuous marginals cdfs F_1, \dots, F_d and a copula C . Denote $G = C(G_1, \dots, G_d)$ a MEVD with EV copula C_{EV} . Then $F \in DA(G)$ if and only if $F_i \in DA(G_i)$ for all $i \in \{1, \dots, d\}$ and*

$$\lim_{t \rightarrow \infty} C^t(u_1^{1/t}, \dots, u_d^{1/t}) = C_{EV}(u_1, \dots, u_d) \quad (2.6)$$

for all $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$.

If (2.6) holds for some copula C , we say that $C \in CDA(C_{EV})$. EV copulas can be characterized in several ways. The following is due to Pickands (1981), who states that C is an extreme value copula if and only if it can be written

$$C(\mathbf{u}) = \exp \left\{ A \left(\frac{\log u_1}{\sum_{i=1}^d \log u_i}, \dots, \frac{\log u_d}{\sum_{i=1}^d \log u_i} \right) \sum_{i=1}^d \log u_i \right\},$$

where the map A is called the *Pickands dependence function*.

Other families of copulas do exist, e.g., Elliptical and Archimedean copulas. Elliptical copulas are the implicit dependence structure for multivariate elliptical distributions. A random vector $\mathbf{X} \in \mathbb{R}^d$ is elliptically distributed if $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{S}$, where $\boldsymbol{\mu} \in \mathbb{R}^d$ is a vector of location parameters, $\mathbf{A} \in \mathbb{R}^{n \times k}$ such that $\mathbf{A}\mathbf{A}^\top$ is a covariance matrix, \mathbf{S} is uniformly distributed on \mathbb{S}^{k-1} , and $R \geq 0$ is a radial rv independent of \mathbf{S} . Archimedean copulas are characterized by a generator function, $\psi : [0, \infty) \mapsto [0, 1]$, where

$$C_\psi(u_1, \dots, u_d) = \psi \{ \psi^{-1}(u_1) + \dots + \psi^{-1}(u_d) \}.$$

For C_ψ to be a valid copula, McNeil and Nešlehová (2009) show that the following properties for ψ are required:

- (i) $\psi(0) = 1$ and $\lim_{x \rightarrow \infty} \psi(x) = 0$.
- (ii) ψ is strictly decreasing on the interval $[0, x_\psi]$ where $x_\psi = \inf\{x : \psi(x) = 0\}$.
- (iii) For all $k \in \{1, \dots, d-2\}$, the k th derivative of ψ exists on $(0, \infty)$ and satisfies $(-1)^k \psi^{(k)}(x) \geq 0$. Additionally, $(-1)^{d-2} \psi^{(d-2)}$ is non-decreasing and convex on $(0, \infty)$.
- (iv) The pseudo-inverse $\psi^{-1} : [0, 1] \mapsto [0, \infty]$ is defined as the inverse of ψ on $(0, 1]$ with $\psi^{-1}(0) = x_\psi$.

While the elliptical and Archimedean families of copulas are generally not extreme value, they can exhibit tail dependence; see Section 2.2.2.

In this thesis, we will be interested in estimating the underlying copula structure. Estimation of a copula is a widely discussed topic. There are several well documented methods for estimating C .

2.2 Multivariate extreme value theory

The simplest of them is the empirical copula. For a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ with corresponding ranks $R_{ij} = \sum_{k=1}^n \mathbb{1}(X_{kj} \leq X_{ij})$, the empirical copula can be written

$$\begin{aligned} C_n(\mathbf{u}) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}(R_{i1}/n \leq u_1, \dots, R_{id}/n \leq u_d) \\ &= \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1}\{\hat{F}_{i,n}(X_{ij}) \leq u_j\}. \end{aligned}$$

For more discussion on the empirical copula, see, e.g., [Fermanian et al. \(2004\)](#), [Genest and Rémillard \(2004\)](#), [Rüschendorf \(2009\)](#) or [Segers \(2012\)](#). One issue with the empirical copula is it does not have uniform margins and, therefore, is not a genuine copula. As such, it does not satisfy the Fréchet–Hoeffding bounds (2.5). Two estimators that address this are the checkerboard copula C_n^{\boxtimes} ([Genest et al., 2014, 2017](#); [Kuzmenko et al., 2020](#)) and the empirical beta copula C_n^β ([Segers et al., 2017](#)). In particular, the empirical beta copula can be written

$$C_n^\beta(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{R_{i,j},n}(u_j),$$

where $\mathbf{u} \in [0, 1]^d$ and $F_{r,n}$ is the cumulative distribution function (cdf) of $U_{r:n}$, the r th order statistic of n independent uniform random variables on $[0, 1]$. This also corresponds to $U_{r:n} \sim \text{Beta}(r, n + 1 - r)$. Given that both C_n^{\boxtimes} and C_n^β are genuine copulas, they satisfy the Fréchet–Hoeffding bounds (2.5). Furthermore, [Segers et al. \(2017\)](#) demonstrate that C_n^β outperforms C_n and C_n^{\boxtimes} in terms of both bias and variance.

2.2.2 Tail dependence

In multivariate statistical analysis, practitioners are often interested in summarizing the way rv's impact one another. Classic measures of dependence between two random variables X_1 and X_2 include Pearson correlation

$$\rho(X_1, X_2) = \frac{\mathbb{E}[\{X_1 - \mathbb{E}(X_1)\}\{X_2 - \mathbb{E}(X_2)\}]}{\sqrt{\text{var}(X_1)\text{var}(X_2)}},$$

and Kendall's τ ,

$$\begin{aligned} \tau(X_1, X_2) &= \Pr\{(X_1 - X_1^*)(X_2 - X_2^*) > 0\} - \Pr\{(X_1 - X_1^*)(X_2 - X_2^*) < 0\} \\ &= 2 \Pr\{(X_1 - X_1^*)(X_2 - X_2^*) > 0\} - 1, \end{aligned}$$

where (X_1^*, X_2^*) denotes an independent copy of (X_1, X_2) . These quantities are measures of *overall* association between X_1 and X_2 . However, in EVT, one is more interested in the likelihood of

2.2 Multivariate extreme value theory

extreme events occurring together. For instance, two stock indices recording record lows together, or neighboring cities experiencing droughts. The repercussions of these types events occurring together could be disproportionately worse than if they were to occur separately. In this sense there is less interest in the dependence between probabilistically common events and more so in the dependence between rare events.

One way this can be quantified is through the *upper tail dependence function* (udf)

$$\lambda(\mathbf{x}) = \lim_{t \rightarrow 0^+} \frac{\Pr\{F_1(X_1) \geq 1 - tx_1, \dots, F_d(X_d) \geq 1 - tx_d\}}{t}. \quad (2.7)$$

For a copula C representing the dependence structure of \mathbf{X} the udf can be written

$$\lambda(\mathbf{x}) = \lim_{t \rightarrow 0^+} \frac{\bar{C}(tx_1, \dots, tx_d)}{t},$$

where $\bar{C}(u_1, \dots, u_d) = \Pr\{\bar{F}_1(X_1) \leq u_1, \dots, \bar{F}_d(X_d) \leq u_d\}$ is the survival copula associated to C . The udf quantifies the likelihood that the members of a random vector \mathbf{X} are simultaneously large. In two dimensions, the scenario where $\mathbf{x} = (1, 1)$ is often considered, i.e.,

$$\begin{aligned} \chi := \lambda(1, 1) &= \lim_{t \rightarrow 0^+} \Pr\{F_1(X_1) \geq 1 - t | F(X_2) \geq 1 - t\} \\ &= \lim_{t \rightarrow 0^+} \frac{2t + C(1 - t, 1 - t) - 1}{t}. \end{aligned}$$

The value χ is called the *upper tail dependence coefficient* (udc) and summarizes the asymptotic tail dependence on the pair (X_1, X_2) . When $\chi = 1$ random pairs are said to be comonotonic, for $\chi = 0$ they are said to be asymptotically independent and for $0 < \chi < 1$ the random pairs are said to be asymptotically dependent. See [Joe \(1997\)](#) and [Joe et al. \(2010\)](#) for more information on the udc and udf, respectively. An analogous quantity exists for the lower tail, called the lower tail dependence function.

Related to the udf is the *stable tail dependence function* (stdf),

$$\mathcal{L}(\mathbf{x}) = \lim_{t \rightarrow 0^+} \frac{\Pr\{F(X_1) \geq 1 - tx_1 \text{ or } \dots \text{ or } F(X_d) \geq 1 - tx_d\}}{t},$$

which considers the probability that at least one observation is extreme. In two dimensions, the stdf and udf can be linked by

$$\mathcal{L}(x, y) = x + y - \lambda(x, y).$$

This relationship will be important in [Chapter 3](#). Finally, EV copulas may also be written in terms of

2.3 Risk measures

the stdf, i.e, a copula C is extreme value if and only if there exists a stdf \mathcal{L} such that $C = C_{\mathcal{L}}$, and

$$C_{\mathcal{L}} = \exp\{-\mathcal{L}(-\log u_1, \dots, -\log u_d)\}.$$

2.3 Risk measures

Across a multitude of disciplines, including finance, actuarial science and hydrology, quantifying risky positions is an essential part of day-to-day operations and an important step in making sound decisions. Examples include an insurance company wishing to assess how much capital to reserve in order to pay off future claims or a city deciding where to limit the construction of new structures, be they private or commercial, given the risk of flooding. Completing these tasks is often done through the use of *risk measures*. In fact, risk measures are often *required* in certain industries so that companies operate in good faith of the interests of stakeholders; see for instance, [Solvency II \(2016\)](#), [OSFI \(2019\)](#), or [Basel II \(2004\)](#), among others. A risk measure is defined as a mapping from the set of random variables (measurable functions) to the reals, i.e.,

$$\mathcal{T} : \mathcal{X} \mapsto \mathbb{R}^d.$$

It is often the case that a risk measure is defined in terms of some pre-prescribed risk level $\alpha \in (0, 1)$; in this sense we can denote risk measures as mappings \mathcal{T}_{α} . In [Section 2.3.1](#), we introduce several common univariate risk measures. [Section 2.3.2](#) defines several multivariate extensions of these measures. [Section 2.3.3](#) discusses several definitions which describe properties that could be considered as desirable for risk measures and the heuristics behind them. Finally, [Section 2.3.4](#) provides concepts and examples surrounding the accurate estimation of risk measures using observations.

2.3.1 Univariate risk measures

The two most popular risk measures are the value-at-risk (VaR) and expected shortfall (ES). First, and perhaps currently the most widely-used risk measure, is VaR. In particular, use of VaR is required for regulatory purposes in many cases. For example, [Solvency II \(2016\)](#) requires European insurance companies to retain capital reflecting the 99.5% VaR of their “basic own funds” for a period of one year to mitigate the risk of insolvency and the ability to pay customer claims. Similar regulations for Canada are provided through [OSFI \(2019\)](#). For a rv X with cdf F , the α -level VaR is defined by

$$\text{VaR}_{\alpha}(X) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\} \tag{2.8}$$

$$= \inf\{x \in \mathbb{R} : \bar{F}(x) \leq 1 - \alpha\} \tag{2.9}$$

2.3 Risk measures

since $\bar{F}(x) = \Pr(X > x) = 1 - F(x)$. For continuous rv's, this can be simplified as $\text{VaR}_\alpha(X) = F^{-1}(\alpha)$ where F^{-1} is the generalized inverse of F . The VaR is often referred to simply as a quantile in many other fields of study. In this sense, it can be seen that the VaR is also a commonly used metric for risk in EVT, as the α -level VaR is the same as the $1 - 1/p$ return level.

While the α -level VaR can be interpreted as the value of X for which you are protected at least $100\alpha\%$ of the time, sometimes one wishes to know more. For instance, were a rv X to exceed a specified value, what could we expect to actually happen? A practical example: given that the water level at a dam will surpass the dam's height, by how much will it overflow? The ES, sometimes referred to as the tail value-at-risk (TVaR), measures exactly this. The ES at level α for a random variable X with cdf F is

$$\text{ES}_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_u(X) du.$$

Moreover, from [Acerbi and Tasche \(2002a\)](#), it can be shown that

$$\text{ES}_\alpha(X) = \begin{cases} \mathbb{E}\{X|X > \text{VaR}_\alpha(X)\} & \text{for } X \text{ continuous,} \\ \frac{1}{1-\alpha} (\mathbb{E}\{X; X \geq \text{VaR}_\alpha(X)\} + \text{VaR}_\alpha(X) [\Pr\{X < \text{VaR}_\alpha(X)\} - \alpha]) & \text{for } X \text{ discrete,} \end{cases}$$

where $\mathbb{E}(X; A) = \mathbb{E}(X \cdot \mathbb{1}_A)$. It is easy to see that $\text{ES}_\alpha(X) \geq \text{VaR}_\alpha(X)$ for all $\alpha \in (0, 1)$. For continuous X , the ES is also known as the conditional tail expectation (CTE). A criticism of ES is that it can be heavily impacted by the presence of outliers in a collection of observations. In this sense it is not considered to be robust. A solution to this issue, introduced in [Cont et al. \(2010\)](#), is to cap the range of possible outcomes that are considered past $\text{VaR}_\alpha(X)$. This gives way to the *range value-at-risk* (RVaR) defined for $0 \leq \alpha_1 < \alpha_2 \leq 1$ by

$$\text{RVaR}_{\alpha_1, \alpha_2}(X) = \frac{1}{\alpha_2 - \alpha_1} \int_{\alpha_1}^{\alpha_2} \text{VaR}_u(x) du.$$

By setting $\alpha_2 = 1$, it is easy to see that $\text{RVaR}_{\alpha_1, 1}(X) = \text{ES}_\alpha(X)$.

While VaR is a more prevalent risk measure in industry, ES is increasingly considered a more appropriate choice for reasons which will be discussed in Section 2.3.3. For a further discussion on VaR and ES as well as other related notions to risk measurement, see, e.g., [McNeil et al. \(2015\)](#) or [Albrecher et al. \(2017\)](#).

2.3.2 Multivariate extensions

While univariate risk measures are fairly well understood, the responsibility to accurately assess risk is inherently a multivariate problem. However, once these measures are extended to the multivariate setting, it opens a seemingly boundless list of possibilities. There is often no clear-cut extension for

2.3 Risk measures

VaR by

$$\begin{aligned}\underline{\text{VaR}}_\alpha(\mathbf{X}) &= \partial\{\mathbf{x} \in \mathbb{R}_+^d : F(\mathbf{x}) \geq \alpha\} \\ &= \bigcup_{i=1}^d \left\{ \left(x_1, \dots, x_{i-1}, \underline{\text{VaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}), x_{i+1}, \dots, x_d \right) : x_j \geq \text{VaR}(X_j), \forall j \neq i \right\},\end{aligned}$$

where $\underline{\text{VaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}) = \inf\{x_i \in \mathbb{R}_+ : F_{x_{\setminus i}}(x_i) \geq \alpha\}$ denotes the lower orthant VaR for a rv X_i when $\mathbf{X}_{\setminus i} = \mathbf{x}_{\setminus i}$. Similarly, the upper-orthant VaR is defined

$$\begin{aligned}\overline{\text{VaR}}_\alpha(\mathbf{X}) &= \partial\{\mathbf{x} \in \mathbb{R}_+^d : \bar{F}(\mathbf{x}) \leq 1 - \alpha\} \\ &= \bigcup_{i=1}^d \left\{ \left(x_1, \dots, x_{i-1}, \overline{\text{VaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}), x_{i+1}, \dots, x_d \right) : x_j < \text{VaR}(X_j), \forall j \neq i \right\},\end{aligned}$$

where $\overline{\text{VaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}) = \inf\{x_i \in \mathbb{R}_+ : \bar{F}_{x_{\setminus i}}(x_i) \leq 1 - \alpha\}$ is the upper orthant VaR for a rv X_i when $\mathbf{X}_{\setminus i} = \mathbf{x}_{\setminus i}$. The central idea behind this formulation is to fix all but one of the random variables, and find the value of the free variable which satisfies the required conditions. For more information on the properties of the lower and upper orthant VaR, see [Cossette et al. \(2013\)](#).

It is important to note that using these definitions, the upper and lower orthant VaR defined by (2.10) and (2.11) provide an infinite collection of points, as opposed to a singular value. Such a property could be considered undesirable for applications such as capital allocation, where a singular value, or vector, is required for decision making purposes. While [Cossette et al. \(2013\)](#) discuss methods for allocating capital from this collection of points, [Cousin and Di Bernardino \(2013\)](#) take a different route in developing a multivariate extension of VaR. Here, the authors define an upper and lower orthant VaR as the expected value of the random vector given that it is a member of the level sets (2.10) or (2.11), i.e.,

$$\begin{aligned}\text{eVaR}_\alpha(\mathbf{X}) &= \mathbb{E}\{\mathbf{X} | \mathbf{X} \in \underline{\text{VaR}}_\alpha(\mathbf{X})\} \\ &= \begin{bmatrix} \mathbb{E}\{X_1 | \mathbf{X} \in \underline{\text{VaR}}_\alpha(\mathbf{X})\} \\ \vdots \\ \mathbb{E}\{X_d | \mathbf{X} \in \underline{\text{VaR}}_\alpha(\mathbf{X})\} \end{bmatrix}\end{aligned}\tag{2.12}$$

and

$$\begin{aligned}\overline{\text{eVaR}}_\alpha(\mathbf{X}) &= \mathbb{E}\{\mathbf{X} | \mathbf{X} \in \overline{\text{VaR}}_\alpha(\mathbf{X})\} \\ &= \begin{bmatrix} \mathbb{E}\{X_1 | \mathbf{X} \in \overline{\text{VaR}}_\alpha(\mathbf{X})\} \\ \vdots \\ \mathbb{E}\{X_d | \mathbf{X} \in \overline{\text{VaR}}_\alpha(\mathbf{X})\} \end{bmatrix},\end{aligned}\tag{2.13}$$

2.3 Risk measures

respectively. In these definitions, the resulting output is equal in dimension to the underlying random vector, foregoing the need for further allocation methods.

Just as in the univariate case, the formulations of multivariate VaR provided by [Cossette et al. \(2013\)](#) and [Cousin and Di Bernardino \(2013\)](#) can also be seen as providing insufficient information in catastrophic scenarios. To this end, both definitions of VaR offer a seamless continuation to multivariate ES. For brevity we focus on the lower orthant, though equivalent definitions are available for the upper orthant in each case. First, [Beck and Mailhot \(2018\)](#) extend the work of [Cossette et al. \(2016\)](#), defining the lower orthant TVaR as

$$\underline{\text{TVaR}}_{\alpha}(\mathbf{X}) = \bigcup_{i=1}^d \left\{ (x_1, \dots, x_{i-1}, \underline{\text{TVaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}), x_{i+1}, \dots, x_d) : x_j > \text{VaR}(X_j), \forall j \neq i \right\},$$

where

$$\underline{\text{TVaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}) = \mathbb{E} \left[X_i | X_i > \underline{\text{VaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}), \mathbf{X}_{\setminus i} \leq \mathbf{x}_{\setminus i} \right], \quad x_j \geq \text{VaR}(X_j)$$

for all $j \in \{1, \dots, d\}$ with $j \neq i$. A similar definition exists for the upper orthant TVaR; see [Cossette et al. \(2016\)](#) or [Beck and Mailhot \(2018\)](#) for more information. The connection between $\underline{\text{TVaR}}_{\alpha}(\mathbf{X})$ and its univariate counterpart becomes more apparent with [Proposition 2.8](#).

Proposition 2.8 ([Cossette et al. \(2016\)](#), Proposition 2.2). *For any $i \in \{1, \dots, d\}$ and $x_i \geq \text{VaR}(X_i)$, one has*

$$\underline{\text{TVaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}) = \frac{1}{F_{\setminus i}(\mathbf{x}_{\setminus i}) - \alpha} \int_{\alpha}^{F_{\setminus i}(\mathbf{x}_{\setminus i})} \underline{\text{VaR}}_{u, x_{\setminus i}}(\mathbf{X}) du.$$

In this sense $\underline{\text{TVaR}}_{\alpha}(\mathbf{X})$ exists as a natural continuation of $\underline{\text{VaR}}_{\alpha}(\mathbf{X})$. Conversely, using the levels sets $\underline{L}_{\mathbf{X}}$ and $\bar{L}_{\mathbf{X}}$, multivariate formulations of CTE arise naturally as extensions of [\(2.12\)](#) and [\(2.13\)](#). As defined in [Cousin and Di Bernardino \(2014\)](#), one has

$$\begin{aligned} \underline{\text{CTE}}_{\alpha}(\mathbf{X}) &= \mathbb{E} \{ \mathbf{X} | \mathbf{X} \in \underline{L}_{\mathbf{X}}(\alpha) \} \\ &= \begin{bmatrix} \mathbb{E} \{ X_1 | \mathbf{X} \in \underline{L}_{\mathbf{X}}(\alpha) \} \\ \vdots \\ \mathbb{E} \{ X_d | \mathbf{X} \in \underline{L}_{\mathbf{X}}(\alpha) \} \end{bmatrix}. \end{aligned}$$

The upper orthant CTE can be defined similarly. These measures are interpreted in the same way as their univariate counterparts: given that you have surpassed the stated risk level, how much can you expect to lose?

An alternative formulation of multivariate risk measures that has gained some traction in recent years is *geometric* risk measures. [Chaudhuri \(1996\)](#) formalizes the notion of geometric VaR as

2.3 Risk measures

follows:

$$\overrightarrow{\text{VaR}}_{\alpha}(\mathbf{X}) = \arg \min_{x \in \mathbb{R}^d} \mathbb{E} \{ \Phi_{\alpha}(\mathbf{X} - \mathbf{x}) \}$$

with $\Phi_{\alpha}(\mathbf{x}) = (\|\mathbf{x}\|_2 + \langle \alpha, \mathbf{x} \rangle) / 2$ for a fixed $\alpha \in B$, where $\|\cdot\|_2$ is the Euclidean norm, $\langle \cdot, \cdot \rangle$ is the scalar product, and $B = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 < 1\}$ is the unit ball in \mathbb{R}^d . In contrast to the measures defined above, the risk-level α for geometric risk measures is actually a vector in itself. This feature allows one to assign different levels of risk to each asset in \mathbf{X} while still considering the dependence between them, as opposed to a singular level of risk associated to the collection of random variables. Similarly, a geometric formulation of TVaR and RVaR is possible. See [Herrmann et al. \(2020\)](#) for more information.

This is by no means an exhaustive list of existing risk measures. For more research on multivariate risk measures, see, e.g., [Torres et al. \(2015\)](#), [Di Bernardino et al. \(2015\)](#), [Di Bernardino and Palacios-Rodríguez \(2017\)](#), or [Girard and Stupfler \(2017\)](#).

2.3.3 Properties of risk measures

As a mathematical object, a risk measure is a very vague concept. In Sections 2.3.1 and 2.3.2 we introduced several well known examples of risk measures. However, what is it that makes these measures so widely used and commonly referenced? In practical applications, it is of great interest to discuss properties which one believes define a “good” risk measure. The first such notion we discuss, and one which is widely considered the standard for any useful measure of risk is the property of *coherence*. Introduced in [Artzner et al. \(1999\)](#), a risk measure \mathcal{T} is called *coherent* if all of the following hold:

(A1) Translation invariance: for all $X \in \mathcal{X}$ and $c \in \mathbb{R}$

$$\mathcal{T}(X + c) = \mathcal{T}(X) + c.$$

By adding or subtracting a deterministic quantity to the underlying rv, this alters the associated risk to the random variable by the same amount.

(A2) Subadditivity: for all $X, Y \in \mathcal{X}$

$$\mathcal{T}(X + Y) \leq \mathcal{T}(X) + \mathcal{T}(Y).$$

Subadditivity reflects the notion of risk reduction through diversification. By aggregating assets, the subsequent collective risk will be less than the risks of the individual assets combined.

2.3 Risk measures

(A3) Positive homogeneity: for all $X \in \mathcal{X}$ and $\lambda > 0$,

$$\mathcal{T}(\lambda X) = \lambda \mathcal{T}(X).$$

Instead of diversifying one's portfolio, consider multiples of the same asset. The resulting portfolio risk would be scaled as such.

(A4) Monotonicity: For $X, Y \in \mathcal{X}$ such that $\Pr(X \leq Y) = 1$, then

$$\mathcal{T}(X) \leq \mathcal{T}(Y),$$

almost surely. A variable that always takes larger values, is always riskier.

Given these axioms and their interpretations, it is easy to see these properties provide a defensible definition for a good measure of risk.

With coherence in mind, we elaborate on the adequacy of certain risk measures. In particular, it becomes clear why VaR is so heavily criticized in the literature: VaR is not coherent as it is not always sub-additive. When considering VaR, pooling of risks does not guarantee a reduction of risk. Conversely, ES is sub-additive and therefore coherent. Many academics use this difference, in part, in motivating a switch from VaR to ES in many risk management situations. For more discussion on the differences between VaR and ES, see, e.g., [Tasche \(2002\)](#), [Acerbi and Tasche \(2002b\)](#), [Yamai and Yoshida \(2002\)](#), or [Yamai and Yoshida \(2005\)](#).

While coherence is generally accepted as a well founded definition for axiomatic risk measures, it is not devoid of criticism. These criticisms often center around the sub-additivity property which VaR is criticized for lacking. For example, [Dhaene et al. \(2008\)](#) discuss scenarios under which sub-additivity can be problematic in a risk management scenario. One way to alleviate such pitfalls, is to relax the axioms required for coherence. Introduced by [Föllmer and Schied \(2002\)](#), *convex* risk measures accomplish this by replacing properties (A2) and (A3) with

(A2.5) Convexity: For all $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$,

$$\mathcal{T} \{ \lambda X + (1 - \lambda)Y \} \leq \lambda \mathcal{T}(X) + (1 - \lambda) \mathcal{T}(Y).$$

It should be noted that (A2) and (A3) imply (A2.5), but the converse is not true. In this sense coherent risk measures are a subset of the class of convex risk measures. Convex risk measures are sometimes referred to as *weakly coherent* risk measures because of this. This formulation is also motivated by criticisms of the positive homogeneity property. In certain situations, it is argued that the risk of an asset can increase in a nonlinear fashion with respect to the size of its position. For additional discussion on convex risk measures, see [Deprez and Gerber \(1985\)](#) and [Frittelli and Gianin \(2002\)](#).

2.3 Risk measures

The discussion around risk measures in this thesis generally centers around the concept that the risk associated to a rv X is inextricably linked to its distribution. This is formalized with the axiom of *law invariance*.

(A0) Law Invariance: If $X, Y \in \mathcal{X}$ such that $\Pr(X \leq c) = \Pr(Y \leq c)$ for all c , then $\mathcal{T}(X) = \mathcal{T}(Y)$.

Combined with axioms (A1)-(A4) one has *law-invariant coherent risk measures* (Kusuoka, 2001). Axioms (A0), (A1), (A2.5) and (A4) form the collection of *law-invariant convex risk measures* (Frittelli and Gianin, 2005). However, Kou et al. (2013) take a different approach, choosing to develop a characterization for risk measures which is primarily data driven. Motivated by their interest in the Basel accords (Basel II (2004); Basel III (2010)), they introduce the notion of natural risk statistics. The idea of quantifying risk using statistics instead of risk measures is motivated in part by the appeal of assuaging the risk of model misspecification. Consider a sample $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)})$ from a rv X , where $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_{n_i}^{(i)})$ is the i th sub-sample of X for $i \in \{1, \dots, m\}$ where $n_1 + \dots + n_m = n$ is the total number of observations. The sub-samples of $\mathbf{x}_1, \dots, \mathbf{x}_m$ represent different scenarios of X . A statistic $\hat{\mathcal{T}} : \mathbb{R}^n \mapsto \mathbb{R}$ is called a *natural risk statistic* if it satisfies the following axioms:

(C1) Positive homogeneity and translation scaling: For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $a \geq 0$, $b \in \mathbb{R}$ and $s > 0$

$$\hat{\mathcal{T}}(a\mathbf{x} + b\mathbf{1}) = a\hat{\mathcal{T}}(\mathbf{x}) + sb,$$

where $\mathbf{1} = (1, \dots, 1)$.

(C2) Monotonicity: For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, if $\mathbf{x} \leq \mathbf{y}$ then

$$\hat{\mathcal{T}}(\mathbf{x}) \leq \hat{\mathcal{T}}(\mathbf{y}),$$

where $\mathbf{x} \leq \mathbf{y}$ is interpreted component wise, i.e., $x_i^j \leq y_i^j$ for $i \in \{1, \dots, n_j\}$ and $j \in \{1, \dots, m\}$.

(C3) Scenario-wise comonotonic sub-additivity: For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ which are scenario-wise comonotonic

$$\hat{\mathcal{T}}(\mathbf{x} + \mathbf{y}) \leq \hat{\mathcal{T}}(\mathbf{x}) + \hat{\mathcal{T}}(\mathbf{y}).$$

Here, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are called scenario-wise comonotonic if for all $j \in \{1, \dots, m\}$ and $i, k \in \{1, \dots, n_j\}$, $(x_i^j - x_k^j)(y_i^j - y_k^j) \geq 0$.

(C4) Empirical law invariance: For any collection of permutations ϕ_i of $(1, \dots, n_i)$ for $i \in \{1, \dots, m\}$

$$\hat{\mathcal{T}}(\mathbf{x}) = \hat{\mathcal{T}}\left(x_{\phi_1(1)}^{(1)}, \dots, x_{\phi_1(n_1)}^{(1)}, \dots, x_{\phi_m(1)}^{(m)}, \dots, x_{\phi_m(n_m)}^{(m)}\right).$$

2.3 Risk measures

Coherence is still debatably the most widely discussed characterization of risk measures, though convex risk measures and natural risk statistics have strong heuristic merit. Classifications not explicitly defined in this thesis include: vector-valued coherent risk measures (Jouini et al., 2004), insurance risk measures (Wang et al., 1997), risk measures using Choquet integrals (Song and Yan, 2009) and more. In Chapter 3, we introduce the concept of elicibility (Osband, 1985; Gneiting, 2011) which furthers the discussion of how to define an ideal measure of risk.

2.3.4 Estimation of risk measures

Taking into account law-invariance, the risk measures presented in this thesis are explicitly defined in terms of an underlying rv X or random vector \mathbf{X} . For example, if $X \sim \mathcal{N}(\mu, \sigma^2)$, the α -level VaR can be written

$$\text{VaR}_\alpha(X) = \mu + \sigma \Phi^{-1}(\alpha),$$

where Φ is the standard normal cdf. Similarly, for a bivariate random vector $\mathbf{X} = (X_1, X_2)$ with marginal cdfs F_1 and F_2 and a dependence structure defined by an Archimedean copula with generator ψ , one may write

$$\underline{\text{VaR}}_{\alpha, x_1}(\mathbf{X}) = F_2^{-1} [\psi^{-1} \{ \psi(\alpha) - \psi(u_1) \}],$$

where $u_1 = F_1(x_1)$. Unfortunately, the reality is that practitioners are often presented collections of data without any prior knowledge of the stochastic process from which it was generated. To this end, there are no assurances that any assumed statistical model is in fact correct and choosing the wrong model can lead to drastically incorrect assessments which can be catastrophic for companies, customers, and citizens alike. This helped motivate the formulation of natural risk statistics defined in Section 2.3.3. In this thesis, as a means of mitigating such dangers, risk measures are estimated using nothing more than the data provided and some relatively unrestrictive assumptions. For example, VaR can be estimated using the empirical cdf of a rv X , denoted

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x),$$

i.e.,

$$\widehat{\text{VaR}}_\alpha(X) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\}. \quad (2.14)$$

When defining an estimator for a particular risk measure, one would like to have some guarantees that the estimated value will be reasonably close to the true value were we to know the underlying distribution. Enter the concept of *consistency*. An estimator \mathcal{T}_n for \mathcal{T} is said to be consistent for a

2.3 Risk measures

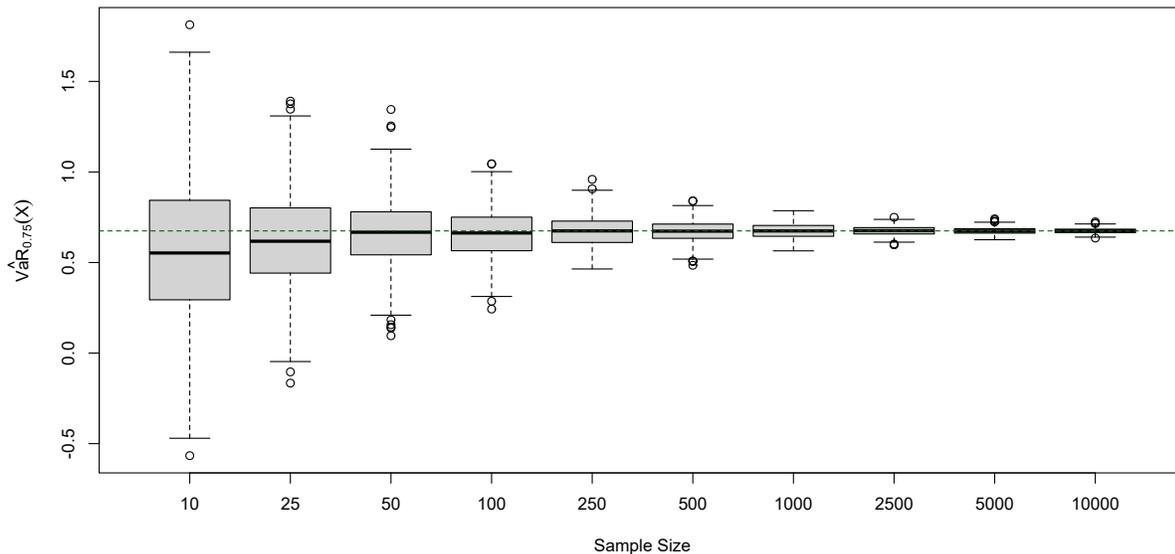


Figure 2.2: Performance of $\widehat{\text{VaR}}_{0.75}(X)$ as the sample size increases

risk measure \mathcal{T} if \mathcal{T}_n converges in probability to \mathcal{T} , i.e.,

$$\lim_{n \rightarrow \infty} \Pr(|\mathcal{T} - \mathcal{T}_n| > \epsilon) = 0,$$

for all $\epsilon > 0$. This is denoted by $\mathcal{T}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathcal{T}$ and is sometimes referred to as weak consistency. Strong consistency implies \mathcal{T}_n converges almost surely (a.s.), i.e.,

$$\Pr\left(\lim_{n \rightarrow \infty} \mathcal{T}_n = \mathcal{T}\right) = 1,$$

and is denoted $\mathcal{T}_n \xrightarrow[n \rightarrow \infty]{a.s.} \mathcal{T}$. It can be shown that convergence almost surely implies convergence in probability, though convergence in probability is usually considered sufficient. Simply put, consistency implies that, with arbitrarily large probability as the sample size tends to infinity, the difference between the estimated value \mathcal{T}_n and the true value \mathcal{T} becomes negligible. For more information on convergence of random variables, see, e.g., [Billingsley \(2008\)](#).

Returning to the example of VaR, it can be shown that the estimator (2.14) is consistent for $\text{VaR}_\alpha(X)$; see, e.g., [van der Vaart \(1998\)](#). In Figure 2.2, the finite-sample performance of (2.14) is demonstrated for $X \sim \mathcal{N}(0, 1)$ and varying sample sizes. As can be seen, for increasingly large sample sizes the accuracy of (2.14) improves until there is minimal variability around the true value $\text{VaR}_{0.75}(X) = 0.6745$. In dimensions $d \geq 2$, [Cuevas and Rodríguez-Casal \(2004\)](#), and [Cuevas et al. \(2006\)](#) develop techniques for the estimation of the level sets $\underline{L}_\alpha(X)$ and $\bar{L}_\alpha(\mathbf{X})$. Focusing on the lower level sets $\underline{L}_\alpha(\mathbf{X})$, the authors show that for an appropriate estimator, F_n , of the joint cdf F of

2.3 Risk measures

\mathbf{X} , the boundaries of the empirical level set

$$\underline{L}_\alpha^n(\mathbf{X}) = \{\mathbf{x} \in \mathbb{R}^d : F_n(\mathbf{x}) \geq \alpha\}$$

is a consistent estimator for $\partial \underline{L}_\alpha(\mathbf{X}) = \underline{\text{VaR}}_\alpha(\mathbf{X})$. [Di Bernardino et al. \(2013\)](#) extended these works to the non-compact setting with specific interest in multivariate risk theory. To deal with the non-compactness assumption, the authors introduce truncated versions of the lower-orthant VaR

$$L_\alpha^T(\mathbf{X}) = \{\mathbf{x} \in [0, T]^d : F(\mathbf{x}) \geq \alpha\} \text{ and } L_\alpha^{n, T_n}(\mathbf{X}) = \{\mathbf{x} \in [0, T_n]^d : F_n(\mathbf{x}) \geq \alpha\},$$

where $T_n \rightarrow \infty$ as $n \rightarrow \infty$. Next, define the Hausdorff distance between two sets A_1 and A_2 ,

$$d_H(A_1, A_2) = \inf \{\rho > 0 : A_1 \subset B(A_2, \rho), A_2 \subset B(A_1, \rho)\}$$

for $B(A, \rho) = \cup_{\mathbf{x} \in A} B(\mathbf{x}, \rho)$ where $B(\mathbf{x}, \rho) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq \rho\}$ is the closed ball centered at \mathbf{x} of radius ρ . The authors establish under certain regularity conditions, including that F_n is continuous for almost all samples of size n and

$$\|F - F_n\|_\infty \rightarrow 0 \text{ a.s.},$$

that

$$d_H(\underline{\text{VaR}}_{\alpha_1}(\mathbf{X})^{T_n}, \underline{\text{VaR}}_{\alpha_1}^n(\mathbf{X})^{T_n}) = \mathcal{O}(\|F - F_n\|_\infty) \text{ a.s.} \quad (2.15)$$

In turn, they use this result to show that the nonparametric estimator

$$\underline{CTE}_\alpha^n(\mathbf{x}) = \begin{bmatrix} \frac{\sum_{i=1}^n X_{i1} \mathbb{1}\{\mathbf{X}_i \in L_\alpha^n(\mathbf{X})\}}{\sum_{i=1}^n \mathbb{1}\{\mathbf{X}_i \in L_\alpha^n(\mathbf{X})\}} \\ \vdots \\ \frac{\sum_{i=1}^n X_{id} \mathbb{1}\{\mathbf{X}_i \in L_\alpha^n(\mathbf{X})\}}{\sum_{i=1}^n \mathbb{1}\{\mathbf{X}_i \in L_\alpha^n(\mathbf{X})\}} \end{bmatrix}$$

is consistent for $\underline{CTE}_\alpha(\mathbf{X})$. Equivalently, [Beck and Mailhot \(2018\)](#) use (2.15) to establish a consistent estimator for the orthant based TVaR. First they show

$$\underline{\text{VaR}}_{\alpha, x_{\setminus i}}^n(\mathbf{X}) \xrightarrow[n \rightarrow \infty]{a.s.} \underline{\text{VaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}), \quad (2.16)$$

where $\underline{\text{VaR}}_{\alpha, x_{\setminus i}}^n(\mathbf{X}) = \{x \in \mathbb{R} : F_{n, x_{\setminus i}}(x) \geq \alpha\}$ and $F_{n, x_{\setminus i}} : x_i \mapsto F_n(\mathbf{x})$. Using (2.16), they show

$$\underline{\text{TVaR}}_{\alpha, x_{\setminus i}}^n(\mathbf{X}) \xrightarrow[n \rightarrow \infty]{a.s.} \underline{\text{TVaR}}_{\alpha, x_{\setminus i}}(\mathbf{X}),$$

2.3 Risk measures

where

$$\underline{\text{TVaR}}_{\alpha, x_{\setminus i}}^n(\mathbf{X}) = \frac{1}{m} \sum_{j=1}^m \underline{\text{VaR}}_{u_j, x_{\setminus i}}(\mathbf{X}),$$

by the dominated convergence theorem. For further research on the estimation of risk measures, see, e.g., [Jones and Zitikis \(2003\)](#), [Ahn and Shyamalkumar \(2011\)](#), [Cai et al. \(2015\)](#), [Lauer and Zähle \(2016\)](#), [Torres et al. \(2017\)](#), [Di Bernardino and Palacios-Rodríguez \(2017\)](#), [Daouia et al. \(2018\)](#).

3

Extreme Multivariate Expectiles

In this chapter, we discuss the estimation of multivariate expectiles. Expectiles have received increased interest in the literature recently due to the fact that they are the only law-invariant risk measure which is both coherent *and* elicitable. Specifically, we develop a consistent estimator for multivariate expectiles when considering extreme levels of risk $\alpha \approx 1$. In Section 3.1, we introduce the notion of elicibility and demonstrate how VaR is an elicitable risk measure. In Section 3.2, univariate expectiles are introduced. Section 3.3 provides several extensions of expectiles for dimensions $d \geq 2$: Σ -expectiles, L_p -expectiles, and geometric expectiles. In Section 3.4, we describe an estimation procedure which provides consistent estimates of L_1 -expectiles for levels of $\alpha \approx 1$. Finally, we end the chapter with a brief discussion and concluding remarks in Section 3.4.5 and 3.5, respectively.

3.1 Elicibility

In Section 2.3.3 several properties which one would consider necessary to define an ideal risk measure are discussed. However, in recent years, increasing attention has been shifted to the study of *elicibility*. Elicibility stems from the notion of backtesting, which refers to the idea of validating and comparing the effectiveness of an estimation procedure using past observations. This in turn helps in deciding the most effective method for making forecasts about future events. However, as discussed in Gneiting (2011), it is not always sensible to compare forecasting methods. This brings us to the property of elicibility. First discussed in Osband (1985) and later coined by Lambert et al. (2008), a collection of risk measures for which meaningful comparisons of point forecasts is possible is called *elicitable*.

3.1 Elicitability

Explicitly defining elicibility requires the concept of *consistency*:

Definition 3.1. A scoring function $S : \mathbb{R}^d \times \mathbb{R}^d \mapsto [0, \infty)$ is consistent for a risk measure \mathcal{T} relative to \mathcal{X} if

$$\mathbb{E}_{\mathbf{X}}\{S(\mathbf{t}, \mathbf{X})\} \leq \mathbb{E}_{\mathbf{X}}\{S(\mathbf{x}, \mathbf{X})\} \quad (3.1)$$

for all $\mathbf{X} \in \mathcal{X}$, all $\mathbf{t} \in \mathcal{T}(\mathbf{X})$ and all $\mathbf{x} \in \mathbb{R}^d$. The scoring function is strictly consistent if it is consistent and equality in (3.1) implies that $\mathbf{x} \in \mathcal{T}(\mathbf{X})$.

For example, if one were to consider the $\mathcal{T}(X) = \mathbb{E}(X)$ then the mean square error $S(x, y) = (x - y)^2$ is consistent relative to $\mathbb{E}(X)$. In fact, for random variables with second moment $\mathbb{E}(X^2) < \infty$, it is strictly consistent. Using consistency, Definition 3.2 formalizes the concept of elicibility.

Definition 3.2. A risk measure \mathcal{T} is elicitable relative to a class of rv \mathcal{X} if there exists a scoring function S that is strictly consistent for \mathcal{T} relative to \mathcal{X} .

One way to interpret Definition 3.2 is that a risk measure $\mathcal{T} : \mathcal{X} \mapsto \mathbb{R}^d$ is elicitable if it may be written as the solution of the optimization problem

$$\mathcal{T}(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E}\{S(\mathbf{x}, \mathbf{X})\}.$$

For a discussion on the properties of elicibility, see [Bellini and Bignozzi \(2015\)](#). For a discussion of elicibility and its relationship to coherence, see [Ziegel \(2016\)](#). Under this framework, consider VaR. Indeed, the α -level VaR is an elicitable risk measure. For a rv X , VaR can be written in its elicitable form as

$$\text{VaR}_\alpha(X) = \arg \min_{x \in \mathbb{R}} \mathbb{E}\{\alpha(X - x)_+ + (1 - \alpha)(X - x)_-\},$$

where $x_+ = \max(0, x)$ and $x_- = \max(0, -x) = -\min(0, x)$. Unfortunately, VaR is not coherent. Conversely, while ES is coherent, it was shown by [Gneiting \(2011\)](#) that it is not elicitable. As there exists no scoring function which is consistent for ES, evaluating and comparing the performance of ES based on point forecasts would not be sensible, making it unsuitable for forecasting-based model selection. However, recent results from [Fissler and Ziegel \(2016\)](#), show that for a real-valued random variable X with finite first moment and unique quantiles, the pair $(\text{VaR}_\alpha(X), \text{ES}_\alpha(X))$ is 2-elicitable. Such a result would allow for the joint use of VaR and ES in forecasting-based model selection. For more information on k -elicibility and higher order elicibility, see [Lambert et al. \(2008\)](#) and [Fissler and Ziegel \(2016\)](#), respectively.

3.2 Expectiles

The introduction of elicibility has led many to wonder about the viability of both VaR and ES as tools for effectively quantifying risk. Much research has been devoted to the development of risk measures while taking into account the concept of elicibility. For instance, [Bellini and Bignozzi \(2015\)](#) discuss a family of risk measures known as *generalized quantiles*. A generalized quantile can be written

$$q_\alpha^*(X) = \arg \min_{x \in \mathbb{R}} \mathbb{E}[\alpha \mathcal{C}_1\{(X - x)_+\} + (1 - \alpha) \mathcal{C}_2\{(X - x)_-\}],$$

where $\mathcal{C}_1, \mathcal{C}_2$ are convex functions. It is immediate that any measure defined in such a way is elicitable with corresponding strictly consistent scoring function $S(x, y) = \alpha \mathcal{C}_1\{(x - y)_+\} + (1 - \alpha) \mathcal{C}_2\{(x - y)_-\}$. Also note that VaR is a generalized quantile where $\mathcal{C}_1(x) = \mathcal{C}_2(x) = x$.

Another member of this family, and the focus of this chapter, are expectiles. In one dimension, the α -level expectile for a rv X corresponds to a generalized quantile with $\mathcal{C}_1(x) = \mathcal{C}_2(x) = x^2$, i.e.,

$$e_\alpha(X) = \arg \min_{x \in \mathbb{R}} \mathbb{E} \{ \alpha (X - x)_+^2 + (1 - \alpha) (X - x)_-^2 \} \quad (3.2)$$

$$= \arg \min_{x \in \mathbb{R}} \mathbb{E} [|\alpha - \mathbf{1}\{(X - x) \leq 0\}| (X - x)^2]$$

$$= \arg \min_{x \in \mathbb{R}} \frac{1}{2} \mathbb{E} [|X - x| \{ |X - x| + (2\alpha - 1) \}]. \quad (3.3)$$

First introduced by [Newey and Powell \(1987\)](#), expectiles are a measure which can be both elicitable *and* coherent. In fact, for $\alpha > 0.5$, it is the only risk measure that is coherent, elicitable and law-invariant ([Ziegel, 2016](#)). As risk managers are often interest in elevated levels of risk, this lines up perfectly with the needs of industry.

For interpretability, it is interesting to consider expectiles through their optimality condition of the first order. Taking the derivative of (3.2) with respect to x , one can describe $e_\alpha(X)$ as the solution of

$$\frac{1 - \alpha}{\alpha} = \frac{\mathbb{E}\{(X - x)_+\}}{\mathbb{E}\{(X - x)_-\}}. \quad (3.4)$$

Thus, $e_\alpha(X)$ can be interpreted as the value of X which equates the ratio of positive scenarios relative negative scenarios to $(1 - \alpha)/\alpha$. The remainder of this chapter will focus on a multivariate extension of expectiles, and in particular the estimation of them for elevated levels of α . For a further discussion of univariate expectiles in risk management and asymptotic properties, see [Bellini and Di Bernardino \(2017\)](#).

3.3 Extensions to higher dimensions

As was discussed in Section 2.3.2, extending risk measures to dimensions $d \geq 2$ allows for a variety of constructions. For expectiles, we provide two examples of multivariate extensions defined in Maume-Deschamps et al. (2017). First, L^p -expectiles:

Definition 3.3. Let $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$ be random vector and let $1 \leq p \leq \infty$. The α -level L^p -expectile of \mathbf{X} is then

$$\begin{aligned} \mathbf{e}_\alpha^p(\mathbf{X}) &= \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left\{ \alpha \|\mathbf{X} - \mathbf{x}\|_p^2 + (1 - \alpha) \|\mathbf{X} - \mathbf{x}\|_p^2 \right\} \\ &= \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left[\alpha \left\{ \sum_{i=1}^d (X_i - x_i)_+^p \right\}^{2/p} + (1 - \alpha) \left\{ \sum_{i=1}^d (X_i - x_i)_-^p \right\}^{2/p} \right]. \end{aligned}$$

The choice of p has an important effect on the resulting measure. For example, for $p = 2$ one recovers a vector of univariate expectiles,

$$\begin{aligned} \mathbf{e}_\alpha^2(\mathbf{X}) &= \arg \min_{\mathbf{x} \in \mathbb{R}^d} \sum_{i=1}^d \mathbb{E} \{ \alpha (X_i - x_i)_+^2 + (1 - \alpha) (X_i - x_i)_-^2 \} \\ &= (e_\alpha(X_1), \dots, e_\alpha(X_d))^\top. \end{aligned}$$

For $p = 2$, the underlying dependence structure of the random vector is not taken into account.

The second construction presented here are matrix expectiles. Also referred to as Σ -expectiles, they are defined in Definition 3.4.

Definition 3.4. Let $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$ be random vector such that $\mathbb{E}|X_i X_j| < \infty$ for all pairs $(i, j) \in \{1, \dots, d\}^2$, and $\Sigma = (\pi_{ij})_{i,j \in \{1, \dots, d\}}$ a real square matrix of full rank, symmetric and positive semi-definite that satisfies

- (i) for all $i \in \{1, \dots, d\}$, $\pi_{ii} = \pi_i > 0$, and
- (ii) for all $i, j \in \{1, \dots, d\}$, $\pi_i > \pi_{ij}$.

The α -level Σ -expectile of a random vector \mathbf{X} is any vector

$$\mathbf{e}_\alpha^\Sigma(\mathbf{X}) \in \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \{ \alpha (\mathbf{X} - \mathbf{x})_+^\top \Sigma (\mathbf{X} - \mathbf{x})_+ + (1 - \alpha) (\mathbf{X} - \mathbf{x})_-^\top \Sigma (\mathbf{X} - \mathbf{x})_- \},$$

where A^\top denotes the transpose of a matrix A . If the solution is unique, the Σ -expectile of \mathbf{X} is

$$\mathbf{e}_\alpha^\Sigma(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \{ \alpha (\mathbf{X} - \mathbf{x})_+^\top \Sigma (\mathbf{X} - \mathbf{x})_+ + (1 - \alpha) (\mathbf{X} - \mathbf{x})_-^\top \Sigma (\mathbf{X} - \mathbf{x})_- \}. \quad (3.5)$$

3.3 Extensions to higher dimensions

For equality to hold in (3.5), it is sufficient for $\pi_{ij} > 0$ for all $(i, j) \in \{1, \dots, d\}^2$. Matrix expectiles are interesting because the flexibility in the choice of Σ provides multiple venues for application. If one were interested in some collection of spatially related random variables, Σ could be a function of a distance matrix. If one were interested in emphasizing the tail dependence between the random vectors, Σ could be a matrix of tail dependence coefficients.

A third extension of expectiles to the multivariate framework uses a geometric interpretation. Building off the works of Chaudhuri (1996) and his geometric interpretation of multivariate VaR, Herrmann et al. (2018) extend the score function representation in (3.3) to $S_\alpha : \mathbb{R}^d \mapsto [0, \infty)$, where

$$S_\alpha(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2 (\|\mathbf{x} - \mathbf{y}\|_2 + \langle \mathbf{x}, \mathbf{y} \rangle),$$

for $\alpha \in B$, giving the *multivariate geometric expectile* defined by

$$\vec{e}_\alpha(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \{ S_\alpha(\mathbf{X}, \mathbf{x}) \}.$$

By allowing α to take values in the d -dimensional unit ball, the authors are capable of assigning different levels of risk to each asset in \mathbf{X} while still considering the dependence between them. See Herrmann et al. (2018) for more information.

Moving forward we consider a special case of multivariate expectiles which satisfy both Definitions 3.3 and 3.4. Setting all the entries of Σ to be 1, denoted $\Sigma = \mathbf{1}_d$, gives

$$\begin{aligned} e_\alpha(\mathbf{X}) &:= e^{\mathbf{1}_d}(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \{ \alpha (\mathbf{X} - \mathbf{x})_+^\top \mathbf{1}_d (\mathbf{X} - \mathbf{x})_+ + (1 - \alpha) (\mathbf{X} - \mathbf{x})_-^\top \mathbf{1}_d (\mathbf{X} - \mathbf{x})_- \} \\ &= \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left[\alpha \left\{ \sum_{i=1}^d (X_i - x_i)_+ \right\}^2 + (1 - \alpha) \left\{ \sum_{i=1}^d (X_i - x_i)_- \right\}^2 \right]. \end{aligned} \quad (3.6)$$

This is also equivalent to the L^p -expectile for $p = 1$. For simplicity, these will henceforth be referred to as L^1 -expectiles. Note that this construction only takes into account pairwise dependencies.

Similar to univariate expectiles, the interpretation of (3.6) can be divined by considering its first order condition. Taking derivatives with respect to \mathbf{x} , it is seen that $e_\alpha(\mathbf{X})$ satisfies the following system of equations:

$$\alpha \sum_{i=1}^d \mathbb{E} [(X_i - x_i)_+ \mathbf{1}\{X_k > x_k\}] = (1 - \alpha) \sum_{i=1}^d \mathbb{E} [(X_i - x_i)_- \mathbf{1}\{X_k < x_k\}]$$

or, equivalently,

$$\frac{1 - \alpha}{\alpha} = \frac{\mathbb{E} [\|\mathbf{X} - \mathbf{x}\|_+ \mathbf{1}\{X_k > x_k\}]}{\mathbb{E} [\|\mathbf{X} - \mathbf{x}\|_- \mathbf{1}\{X_k < x_k\}]}$$

for all $k \in \{1, \dots, d\}$. The similarities to (3.4) are clear; $e_\alpha(\mathbf{X})$ can thus be interpreted as the

3.4 Estimation of extremes for L^1 -expectiles

value of \mathbf{X} that sets the ratio of expected positive scenario outcomes relative to negative ones to $(1 - \alpha)/\alpha$ for all members of the random vector \mathbf{X} . The goal in many risk management applications is to minimize the ratio $(1 - \alpha)/\alpha$, which requires one to consider elevated levels of α . Calculating expectiles for elevated levels of α will be discussed in the remainder of this chapter. We end the section by listing a few properties of L^1 expectiles.

Proposition 3.1. *For any random vector $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$ with finite second moment, $e_\alpha(\mathbf{X})$ has the following properties:*

(i) *Positive homogeneity: for any $a > 0$,*

$$e_\alpha(a\mathbf{X}) = ae_\alpha(\mathbf{X}).$$

(ii) *Translation invariance: for any vector $\mathbf{b} = (b_1, \dots, b_d) \in \mathbb{R}^d$,*

$$e_\alpha(\mathbf{X} + \mathbf{b}) = e_\alpha(\mathbf{X}) + \mathbf{b}.$$

(iii) *Law invariance: for any random vector $\mathbf{Y} = (Y_1, \dots, Y_d) \in \mathbb{R}^d$ with finite second moment such that pairs (X_i, X_j) and (Y_i, Y_j) have the same distribution for all $i, j \in \{1, \dots, d\}^2$, then*

$$e_\alpha(\mathbf{X}) = e_\alpha(\mathbf{Y})$$

for all $\alpha \in [0, 1]$.

(iv) *Pseudo-invariance by linear transformations: for any vectors $\mathbf{a} = (a_1, \dots, a_d) \in [0, \infty)^d$ and $\mathbf{b} = (b_1, \dots, b_d) \in \mathbb{R}^d$,*

$$e_\alpha(A\mathbf{X} + \mathbf{b}) = Ae_\alpha^{A\mathbf{1}_d A}(\mathbf{X}) + \mathbf{b},$$

where $A = \text{diag}(\mathbf{a}^\top)$, a diagonal square matrix with \mathbf{a} across the diagonal.

For a proof of these properties and a discussion of additional properties of $e_\alpha(\mathbf{X})$, see [Maume-Deschamps et al. \(2017\)](#). The listed properties may also be extended for any matrix Σ satisfying Definition 3.4.

3.4 Estimation of extremes for L^1 -expectiles

As noted in [Maume-Deschamps et al. \(2017\)](#), multivariate expectiles can rarely be solved for explicitly. However, the estimation of $e_\alpha(\mathbf{X})$ is possible using noisy observations. In particular, the authors show that one can estimate L^1 -expectiles from observational data using Robbins–Monro stochastic optimization ([Robbins and Monro, 1951](#)). They show that in general scenarios, this algorithm performs quite well when solving for $e_\alpha(\mathbf{X})$. However, one area which the Robbins–Monro

3.4 Estimation of extremes for L^1 -expectiles

algorithm fails is estimating $e_\alpha(\mathbf{X})$ for elevated risk levels $\alpha \approx 1$. As discussed in Section 3.3, there is a great deal of interest in evaluating risk for extreme levels of α . To this end, extra care must be given when estimating $e_\alpha(\mathbf{X})$ for $\alpha \approx 1$. For a discussion of extreme expectiles in the univariate case, see, e.g., [Bellini and Bignozzi \(2015\)](#), [Mao et al. \(2015\)](#), [Bellini and Di Bernardino \(2017\)](#), [Daouia et al. \(2018, 2019\)](#).

In the case of multivariate extreme L^1 -expectiles (MEEs), [Maume-Deschamps et al. \(2018\)](#) lay the foundation we wish to build upon. Their discussion centers around random variables with equivalent tails, which is formalized in Assumption 1.

Assumption 1. *Assume that \mathbf{X} has equivalent regularly varying marginal tails, i.e.,*

(i) $\bar{F}_1 \in RV_{-\theta}(\infty)$, with $\theta > 0$.

(ii) *The tails of X_1, \dots, X_d are equivalent. That is, for all $i \in \{1, \dots, d\}$*

$$\lim_{x \rightarrow \infty} \bar{F}_i(x) / \bar{F}_1(x) = c_i.$$

The conditions of Assumption 1 imply that $\bar{F}_i \in RV_{-\theta}(\infty)$ for all $i \in \{1, \dots, d\}$. The following proposition provides the main result for L^1 -expectiles when considering levels of risk $\alpha \rightarrow 1$.

Proposition 3.2 ([Maume-Deschamps et al. \(2018\)](#), Proposition 2.3). *Assume that \mathbf{X} has a regular varying multivariate distribution in the sense of Definition 2.2 and that Assumption 1 holds. Consider the L_1 -expectile $e_\alpha(\mathbf{X}) = (e_\alpha^i(\mathbf{X}))_{i \in \{1, \dots, d\}}$. Then any limit vector $(\eta, \beta_2, \dots, \beta_d)$ of $((1 - \alpha) / \bar{F}_1 \{e_\alpha^1(\mathbf{X})\}, e_\alpha^2(\mathbf{X}) / e_\alpha^1(\mathbf{X}), \dots, e_\alpha^d(\mathbf{X}) / e_\alpha^1(\mathbf{X}))$ satisfies the following system*

$$\frac{1}{\theta - 1} - \eta \frac{\beta_k^\theta}{c_k} = - \sum_{i=1, i \neq k}^d \left\{ \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt - \eta \frac{\beta_k^{\theta-1}}{c_k} \beta_i \right\}, \quad \forall k \in \{1, \dots, d\}. \quad (3.7)$$

With Proposition 3.2, one can provide an alternative system of equations when solving for $\lim_{\alpha \rightarrow 1} e_\alpha(\mathbf{X})$. Moreover, it has been shown that for the limiting cases of tail dependence, comonotonicity and asymptotic independence, closed form solutions to the system in (3.7) exist. These results are summarized in the following two lemmas:

Lemma 3.1 ([Maume-Deschamps et al. \(2018\)](#), Lemma 2.5). *Under the assumptions of Proposition 3.2, consider the L_1 -expectile $e_\alpha(\mathbf{X}) = (e_\alpha^i(\mathbf{X}))_{i \in \{1, \dots, d\}}$. If $\mathbf{X} = (X_1, \dots, X_d)$ is such that the pairs (X_i, X_j) are asymptotically independent for all $i \neq j \in \{1, \dots, d\}$, then the limit vector $(\eta, \beta_2, \dots, \beta_d)$ of $((1 - \alpha) / \bar{F}_1 \{e_\alpha^1(\mathbf{X})\}, e_\alpha^2(\mathbf{X}) / e_\alpha^1(\mathbf{X}), \dots, e_\alpha^d(\mathbf{X}) / e_\alpha^1(\mathbf{X}))$ satisfies*

$$\eta = \frac{1}{(\theta - 1) \left\{ 1 + \sum_{j=2}^d c_j^{1/(\theta-1)} \right\}} \text{ and } \beta_k = c_k^{1/(\theta-1)} \quad (3.8)$$

for all $k \in \{1, \dots, d\}$.

3.4 Estimation of extremes for L^1 -expectiles

Lemma 3.2 (Maume-Deschamps et al. (2018), Lemma 2.4). *Under the assumptions of Proposition 3.2, consider the L_1 -expectile $e_\alpha(\mathbf{X}) = (e_\alpha^i(\mathbf{X}))_{i \in \{1, \dots, d\}}$. If $\mathbf{X} = (X_1, \dots, X_d)$ is a comonotonic random vector then the limit vector $(\eta, \beta_2, \dots, \beta_d)$ of $((1 - \alpha)/\bar{F}_1\{e_\alpha^1(\mathbf{X})\}, e_\alpha^2(\mathbf{X})/e_\alpha^1(\mathbf{X}), \dots, e_\alpha^d(\mathbf{X})/e_\alpha^1(\mathbf{X}))$ satisfies*

$$\lim_{\alpha \rightarrow 1} \frac{1 - \alpha}{\bar{F}_k\{e_\alpha^k(\mathbf{X})\}} = \frac{1}{(\theta - 1)} \text{ and } \beta_k = c_k^{1/\theta} \quad (3.9)$$

for all $k \in \{1, \dots, d\}$. Note that for $k = 1$ this gives $\eta = 1/(\theta - 1)$.

For simplicity, we denote $\Theta^\perp = (\eta^\perp, \beta_2^\perp, \dots, \beta_d^\perp)$ and $\Theta^+ = (\eta^+, \beta_2^+, \dots, \beta_d^+)$, where the entries of Θ^\perp and Θ^+ are given by (3.8) and (3.9), respectively. By rearranging the results provided in Lemmas 3.1 and 3.2, asymptotic representations of MEEs under asymptotically independent and comonotonic random vectors can be defined by

$$e_\alpha^\perp(\mathbf{X}) \underset{\alpha \rightarrow 1}{\sim} \text{VaR}_\alpha(X_1)(\theta - 1)^{-1/\theta} \left(1 + \sum_{j=2}^d c_j^{1/(\theta-1)}\right)^{-1/\theta} \left(1, c_2^{1/(\theta-1)}, \dots, c_d^{1/(\theta-1)}\right) \quad (3.10)$$

and

$$e_\alpha^+(\mathbf{X}) \underset{\alpha \rightarrow 1}{\sim} \text{VaR}_\alpha(X_1)(\theta - 1)^{-1/\theta} \left(1, c_2^{1/\theta}, \dots, c_d^{1/\theta}\right), \quad (3.11)$$

respectively. Moreover, estimators $\hat{e}_\alpha^\perp(\mathbf{X})$ and $\hat{e}_\alpha^+(\mathbf{X})$ for (3.10) and (3.11), respectively, can be written in the same way, using plug-in estimates for $\text{VaR}_\alpha(X_1)$, θ and c_2, \dots, c_d . The consistency of said estimators is established in Theorem 3.3

Theorem 3.3 (Maume-Deschamps et al. (2018), Theorem 4.6). *Under the assumptions of Proposition 3.2, assume further that the function $U : y \mapsto \text{VaR}_{1-1/y}(X_1)$ satisfies*

$$\frac{U(ux)}{U(x)} \underset{x \rightarrow \infty}{\sim} u^\gamma [1 + h_{-\beta}(u)b(x) + o\{b(x)\}],$$

where $\beta > 0$, $b \in RV_{-\beta}(\infty)$ and $h_{-\beta}(u) = (1 - u^{-\beta})/\beta$. Next, choose $k = k(n)$ and $\alpha = \alpha(n)$ such that

- (i) $k(n) \rightarrow \infty$, $k(n)/n \rightarrow 0$ and $\alpha(n) \rightarrow 1$, $n\{1 - \alpha(n)\} \rightarrow c > 0$ as $n \rightarrow \infty$;
- (ii) $\sqrt{k(n)} \left[1 + \log^2 \frac{k(n)}{n\{1 - \alpha(n)\}}\right] \rightarrow \infty$ as $n \rightarrow \infty$.

Then, if each pair in \mathbf{X} is asymptotically independent,

$$\hat{e}_\alpha^\perp(\mathbf{X})/e_\alpha^\perp(\mathbf{X}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1.$$

3.4 Estimation of extremes for L^1 -expectiles

If \mathbf{X} is comonotonic, then

$$\hat{e}_\alpha^+(\mathbf{X})/e_\alpha^+(\mathbf{X}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1.$$

3.4.1 Optimization

The results of Theorem 3.3 represent a great step forward in the estimation of extreme multivariate expectiles. However, being restricted to the limiting cases of tail dependence is suboptimal. Moreover, it requires one to assume in advance that the given data come from a distribution with asymptotically independent or comonotonic margins. We wish to expand on their methodology so that one may estimate $e_\alpha(\mathbf{X})$ for $\alpha \approx 1$ under any tail dependence structure without any distributional assumptions. Using Proposition 3.2, we reformulate the optimization problem in (3.6).

Definition 3.5. Consider a d -dimensional random vector \mathbf{X} with regularly varying multivariate distribution in the sense of Definition 2.2 with margins possessing equivalent tails as given in Assumption 1. Let $\boldsymbol{\beta}^* = (\beta_1, \dots, \beta_d)$ with $\beta_1 = 1$ and $\boldsymbol{\Theta} = (\eta, \boldsymbol{\beta}_{2:d}) = (\eta, \beta_2, \dots, \beta_d)$. Let $c_1 = 1$ and $\Lambda = (\theta, c_2, \dots, c_d, \lambda(\cdot))$, where c_i is the tail ratio for $i \in \{2, \dots, d\}$, θ is the tail index and λ is the udf associated to the random vector \mathbf{X} in (2.7).

Define the loss function

$$L_\Lambda(\boldsymbol{\Theta}) := \frac{1}{2} \|F_\Lambda(\boldsymbol{\Theta})\|_2^2, \quad (3.12)$$

where

$$F_\Lambda(\boldsymbol{\Theta}) = \left(F_\Lambda^{(1)}(\boldsymbol{\Theta}), \dots, F_\Lambda^{(d)}(\boldsymbol{\Theta}) \right) = \left(g_\Lambda^{(1)}(\boldsymbol{\Theta}) + f_\Lambda^{(1)}(\boldsymbol{\Theta}), \dots, g_\Lambda^{(d)}(\boldsymbol{\Theta}) + f_\Lambda^{(d)}(\boldsymbol{\Theta}) \right),$$

and

$$g_\Lambda^{(k)}(\boldsymbol{\Theta}) = \frac{1}{\theta - 1} - \eta \frac{\beta_k^\theta}{c_k},$$

$$f_\Lambda^{(k)}(\boldsymbol{\Theta}) = \sum_{i \neq k}^d \left\{ \int_{\frac{\beta_i}{\beta_k}}^\infty \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt - \eta \frac{\beta_k^{\theta-1}}{c_k} \beta_i \right\}$$

for all $k \in \{1, \dots, d\}$. Define an optimal vector $\boldsymbol{\Theta}^*$, obtained by optimizing the loss function L_Λ in (3.12), i.e.,

$$\boldsymbol{\Theta}^* = \arg \min_{\boldsymbol{\Theta}} L_\Lambda(\boldsymbol{\Theta}). \quad (3.13)$$

With Definition 3.5, we have formulated an explicit optimization problem for L^1 -expectiles when $\alpha \rightarrow 1$. Corollary 3.1 establishes that the optimization problem (3.13) can be used to calculate MEEs.

3.4 Estimation of extremes for L^1 -expectiles

Corollary 3.1. Consider a d -dimensional random vector \mathbf{X} with regularly varying multivariate distribution in the sense of Definition 2.2. Consider the L_1 -expectile $\mathbf{e}_\alpha(\mathbf{X}) = (e_\alpha^i(\mathbf{X}))_{i \in \{1, \dots, d\}}$ as in (3.6). Under Assumption 1 and the assumption that the vector $((1-\alpha)/\bar{F}_1\{e_\alpha^1(\mathbf{X})\}, e_\alpha^2(\mathbf{X})/e_\alpha^1(\mathbf{X}), \dots, e_\alpha^d(\mathbf{X})/e_\alpha^1(\mathbf{X}))$ has a unique limit point $\Theta = (\eta, \beta_2, \dots, \beta_d)$ for $\alpha \rightarrow 1$, then Θ satisfies the optimization problem (3.13) in Definition 3.5. Moreover, the optimization problems (3.6) and (3.13) are equivalent for $\alpha \rightarrow 1$, i.e.,

$$\lim_{\alpha \rightarrow 1} \frac{e_\alpha^i(\mathbf{X})}{\tilde{e}_\alpha^i(\mathbf{X})} = 1, \text{ for } i \in \{1, \dots, d\}, \text{ where } \tilde{\mathbf{e}}_\alpha(\mathbf{X}) := \text{VaR}_\alpha(X_1)\eta^{1/\theta}(1, \beta_2, \dots, \beta_d)^\top.$$

Proof. From Proposition 3.2 and the uniqueness of the limit $\Theta = (\eta, \beta_2, \dots, \beta_d)$ we get that Θ satisfies the optimization problem in (3.13). Furthermore, from Proposition 5.1 in Maume-Deschamps et al. (2018), we know that $\mathbf{e}_\alpha(\mathbf{X}) \sim \text{VaR}_\alpha(X_1)\eta^{1/\theta}(1, \beta_2, \dots, \beta_d)^\top$, for $\alpha \rightarrow 1$. Hence, the result is proved. \square

To solve this problem, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) descent algorithm will be used. The algorithm is a member of the family of quasi-Newton optimization methods and allows one to avoid calculating second derivatives, a feat which can drastically improve computation time.

Algorithm 1 BFGS for Estimation of Extreme Expectiles

(Step 0) Put counter $k := 0$ and choose initial values

$$\Theta^0 \in \mathbb{R}^d, H_0 \in \mathbb{R}^{d \times d} \text{ initial inverse Hessian matrix, } \sigma \in (0, 1/2), \rho \in (\sigma, 1), \text{ and } \epsilon \geq 0.$$

(Step 1) Let L_Λ as in Definition 3.5. If $\|\nabla L_\Lambda(\Theta^k)\| \leq \epsilon$: STOP.

(Step 2) Calculate the direction $\mathbf{d}^k = -H_k \nabla L_\Lambda(\Theta^k)$.

(Step 3) Determine the step size $t_k > 0$ such that the Wolfe conditions are satisfied, i.e.,

$$\begin{aligned} L_\Lambda(\Theta^k + t_k \mathbf{d}^k) &\leq L_\Lambda(\Theta^k) + \sigma t_k \nabla L_\Lambda(\Theta^k) \\ \nabla L_\Lambda(\Theta^k + t_k \mathbf{d}^k)^\top \mathbf{d}^k &\geq \rho \nabla L_\Lambda(\Theta^k)^\top \mathbf{d}^k. \end{aligned}$$

where \top denotes the transpose.

(Step 4) Update the following:

$$\begin{aligned} \bullet \Theta^{k+1} &:= \Theta^k + t_k \mathbf{d}^k & \bullet \mathbf{y}^k &:= \nabla L_\Lambda(\Theta^{k+1}) - \nabla L_\Lambda(\Theta^k) \\ \bullet \mathbf{s}^k &:= \Theta^{k+1} - \Theta^k & \bullet H_{k+1} &:= (\mathbb{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^\top) H_k (\mathbb{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^\top) + \rho_k \mathbf{s}_k \mathbf{s}_k^\top \end{aligned}$$

where $\rho_k = (\mathbf{y}_k^\top \mathbf{s}_k)^{-1}$.

(Step 5) Set $k \leftarrow k + 1$ and go to **(Step 1)**.

3.4 Estimation of extremes for L^1 -expectiles

The BFGS descent algorithm is given in [Algorithm 1](#). The gradient of L_Λ can then be written

$$\begin{aligned} \nabla L_\Lambda(\cdot) &= \nabla \left[\frac{1}{2} \sum_{k=1}^d \left\{ g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right\}^2 \right] \\ &= \begin{bmatrix} \sum_{k=1}^d \left\{ \left(g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right) \times \frac{\partial}{\partial \eta} \left(g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right) \right\} \\ \sum_{k=1}^d \left\{ \left(g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right) \times \frac{\partial}{\partial \beta_2} \left(g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right) \right\} \\ \vdots \\ \sum_{k=1}^d \left\{ \left(g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right) \times \frac{\partial}{\partial \beta_d} \left(g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot) \right) \right\} \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} \frac{\partial}{\partial \Theta_j} g_\Lambda^{(k)}(\Theta) &= \begin{cases} -\beta_k^\theta / c_k, & j = 1, \\ -\eta \theta \beta_k^{\theta-1} / c_k, & j = k \neq 1, \\ 0, & \text{otherwise,} \end{cases} \\ \frac{\partial}{\partial \Theta_j} f_\Lambda^{(k)}(\Theta) &= \begin{cases} -\frac{\beta_k^{\theta-1}}{c_k} \sum_{i \neq k} \beta_i, & j = 1, \\ \sum_{i \neq k} \left[\lambda^{ik} \left(\frac{c_i}{c_k} \left(\frac{\beta_i}{\beta_k} \right)^{-\theta}, 1 \right) \frac{\beta_i}{\beta_k^2} - \eta \frac{(\theta-1)\beta_k^{\theta-2}}{c_k} \beta_i \right], & j = k \neq 1, \\ -\lambda^{jk} \left(\frac{c_j}{c_k} \left(\frac{\beta_j}{\beta_k} \right)^{-\theta}, 1 \right) \frac{1}{\beta_k} - \eta \frac{\beta_k^{\theta-1}}{c_k}, & \text{otherwise,} \end{cases} \end{aligned}$$

for $j, k \in \{1, \dots, d\}$.

For more information, including convergence results, on the BFGS algorithm see [Nocedal and Wright \(1999\)](#). It is important to note that, in the general case, the true optimization problem (3.13) does not satisfy the convexity assumption required for global convergence of the BFGS algorithm. Nevertheless, the choice to use BFGS is motivated by its simplicity of implementation and further validated by the satisfactory numerical results provided in Section 3.4.4. More sophisticated algorithms which extend BFGS would indeed be possible, including L-BFGS-B ([Byrd et al., 1995](#)) and modified BFGS ([Li and Fukushima, 2001a,b](#)), the latter of which allows one to relax the convexity assumption and still achieve global convergence. These methods are briefly discussed in Section 3.4.5.

In [Figure 3.1](#), the accuracy of the BFGS algorithm applied to (3.12), is demonstrated. It can be seen that the BFGS algorithm performs quite well for the presented optimization problem, recapturing the true MEEs for a range of risk levels.

3.4 Estimation of extremes for L^1 -expectiles

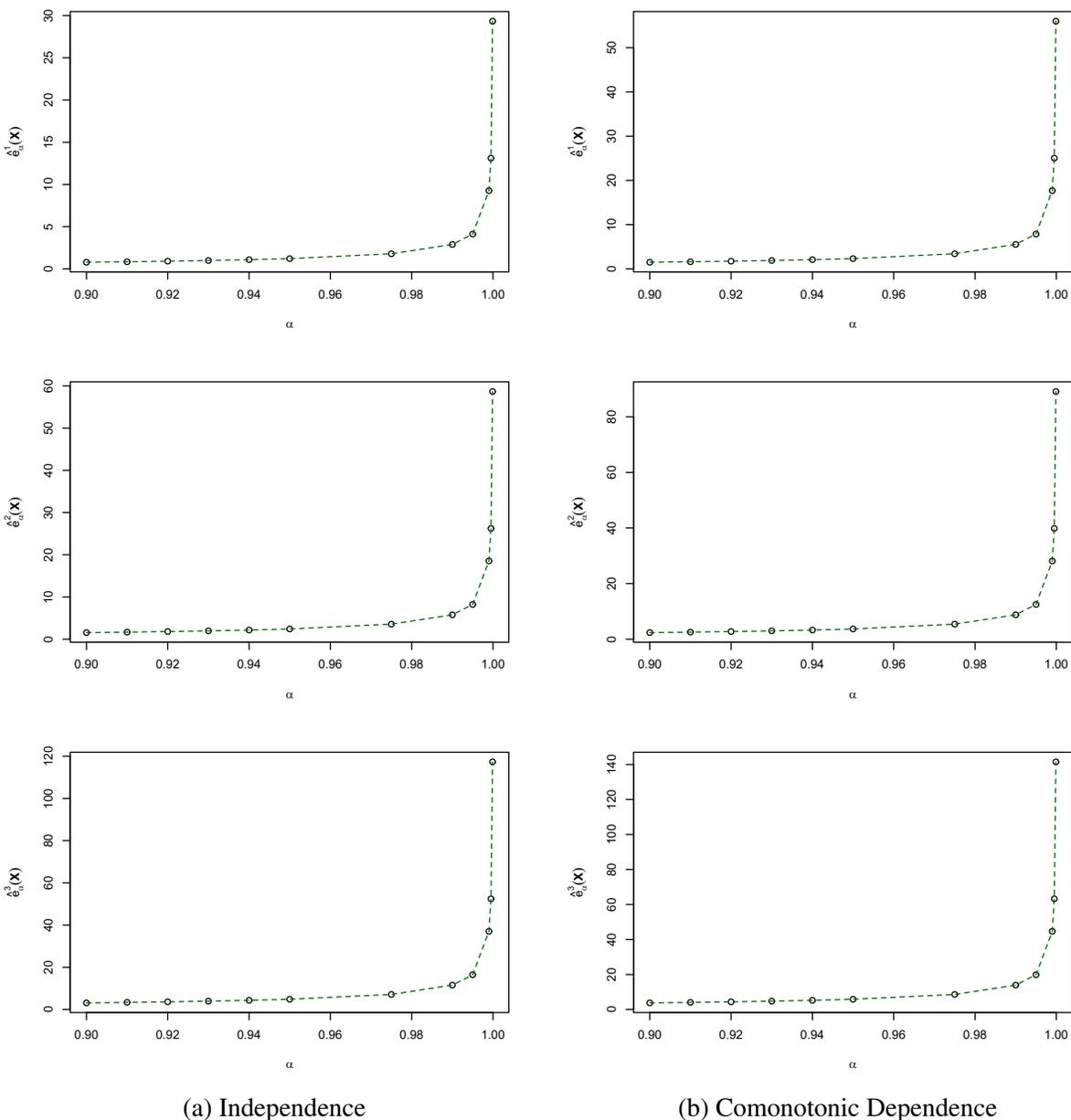


Figure 3.1: Demonstration of BFGS algorithm for solving (3.13) for varying α and dependence structures. Here, $X_i \sim b_i t_2$ where t_2 is a student t -distribution with 2 degrees of freedom and $b_i = 2^{i-1}$ for $i \in \{1, 2, 3\}$. The dotted green line represents the true values as described in [Maume-Deschamps et al. \(2018\)](#).

3.4.2 Approximate optimization problem

While preliminary results of the reformulated optimization problem in Definition 3.5 are encouraging, they are predicated on knowing the underlying distribution of the data and the values of λ , θ and c_2, \dots, c_d . As such information is rarely available in practice, directly solving (3.12) is not

3.4 Estimation of extremes for L^1 -expectiles

possible. Here we consider an approximated version of (3.12) as a proxy for (3.13). In other words, we focus on solving

$$\hat{\Theta}^* = \arg \min_{\Theta \in \mathbb{R}^d} L_{\hat{\Lambda}}(\Theta), \quad (3.14)$$

for some vector of estimators $\hat{\Lambda} = (\hat{\theta}, \hat{c}_2, \dots, \hat{c}_d, \hat{\lambda})$. Specifically, convergence of the estimated optimum can be shown in the following way. The first step will be to show that $\hat{\Lambda} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Lambda$, and subsequently that $L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} L_{\Lambda}(\Theta)$ and $\nabla L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \nabla L_{\Lambda}(\Theta)$. Then, we will prove the consistency of every iteration of the BFGS algorithm (see Theorem 3.7)

$$\hat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Theta^k, \quad \forall k \in \{1, 2, \dots\}.$$

Finally by using a two-step procedure (see Algorithm 2), we can study the behaviour of the proposed method (see Corollary 3.3).

3.4.3 Consistency

Before we can establish the consistency of the estimator $\hat{\Theta}^*$, we first require consistent estimators for the components of Λ . Also note that given the form (3.12), a consistent estimator for

$$\int_{\beta_i/\beta_k}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt \quad (3.15)$$

is also required. For n observations from a random vector \mathbf{X} , denote the random sample $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$ for $i \in \{1, \dots, d\}$. Next, denote $X_{j:1,n} \leq \dots \leq X_{j:n,n}$ the order statistics associated to the realizations of the j th random variable of \mathbf{X} for $j \in \{1, \dots, d\}$. Also define $\ell = \ell(n)$ an intermediate integer sequence such that $\ell = o(n)$ and $\ell \xrightarrow[n \rightarrow \infty]{} \infty$. First, we estimate the tail index θ via Hill's estimator (Hill, 1975):

$$\hat{\gamma} = \frac{1}{\ell_\theta} \sum_{i=1}^{\ell_\theta} \log(X_{1:n-i+1,n}) - \log(X_{1:n-\ell_\theta,n}) \quad \text{and} \quad \hat{\theta} = \frac{1}{\hat{\gamma}} \quad (3.16)$$

for an intermediate integer sequence $\ell_\theta = \ell_\theta(n)$. The tail ratio, c_i , can be estimated for $i \in \{2, \dots, d\}$ by

$$\hat{c}_i = \left(\frac{X_{i:n-\ell_i+1,n}}{X_{1:n-\ell_i+1,n}} \right)^{-\hat{\theta}}, \quad (3.17)$$

where $\ell_i = \ell_i(n)$ is an intermediate sequence and $\hat{\theta}$ as in (3.16). The consistency of $\hat{\theta}$ and \hat{c}_i is established, for instance, in Deheuvels et al. (1988) or de Haan and Ferreira (2006), and Maume-Deschamps et al. (2018), respectively. This leaves the udf as the sole remaining piece. Given the

3.4 Estimation of extremes for L^1 -expectiles

definition of the udf given in Section 2.2, we consider estimators of λ which take the form

$$\hat{\lambda}(\mathbf{x}) = \frac{n}{\ell_\lambda} \hat{C} \left(\frac{\ell_\lambda}{n} \mathbf{x} \right),$$

where \hat{C}_n is some estimator of the survival copula of the random vector \mathbf{X} . In this scenario, to accurately estimate the upper dependence requires one to first define a consistent estimator for \hat{C} . In Figure 3.2, the performance of $\hat{\lambda}$ under the empirical, empirical beta and checkerboard copulas, introduced in Chapter 2, are compared using independent Pareto type I distributions. Specifically, we consider $X_i \sim P(2, 1.25(1 + i))$ with cdf

$$F_i(x) = 1 - \left(\frac{x}{1.25(1 + i)} \right)^{-2}$$

for $i \in \{1, 2, 3\}$. As can be seen, both the empirical beta and checkerboard copulas provide a far more accurate estimates of λ and do not exhibit any of the oscillating behavior that is seen when considering $\hat{\lambda}_n$. The performance is also better at smaller sample sizes when larger limiting subsequence is considered. Choosing to build on the results of Segers et al. (2017) and Kiriliouk et al. (2018), we focus on developing our estimation procedure using the empirical beta copula, though it seems that the checkerboard copula would be a strong choice as well. In two dimensions we can write the empirical beta copula as

$$C_n^\beta(u, v) = \frac{1}{n} \sum_{i=1}^n F_{R_{i,1},n}(u) F_{R_{i,2},n}(v).$$

Moreover, as C_n^β is a genuine copula, using the bivariate relationship $\bar{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ gives pairwise udf

$$\hat{\lambda}_\beta^{ik}(x_i, x_k) = \frac{n}{\ell_\lambda} \left\{ \frac{\ell_\lambda}{n} x_i + \frac{\ell_\lambda}{n} x_k - 1 + \frac{1}{n} \sum_{j=1}^n F_{R_{j,i},n} \left(1 - \frac{\ell_\lambda}{n} x_i \right) F_{R_{j,k},n} \left(1 - \frac{\ell_\lambda}{n} x_k \right) \right\}. \quad (3.18)$$

Proposition 3.4 establishes the consistency of the udf based on the empirical beta copula.

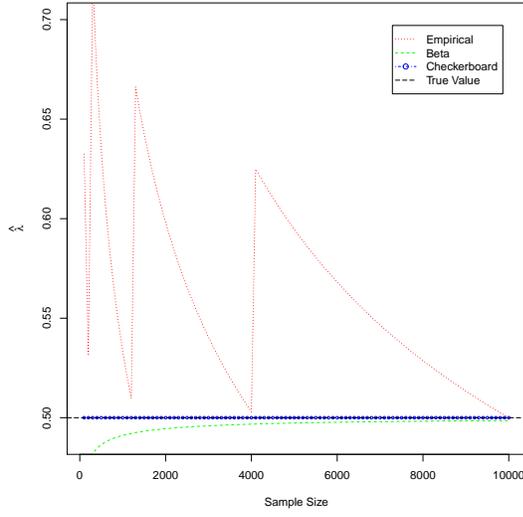
Proposition 3.4. *For the estimator of the udf based on the empirical beta copula, as defined in (3.18), one has that*

$$\hat{\lambda}_\beta^{ik}(x_i, x_k) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \lambda^{ik}(x_i, x_k).$$

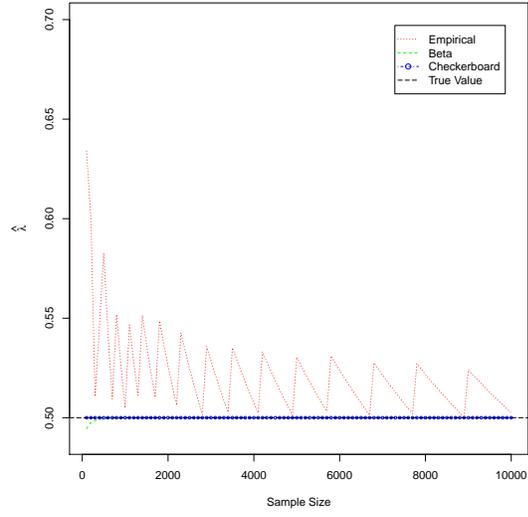
That is, $\hat{\lambda}_\beta^{ik}$ is a consistent estimator for λ^{ik} .

Proof. The proof is straightforward from the work of Kiriliouk et al. (2018). They define the

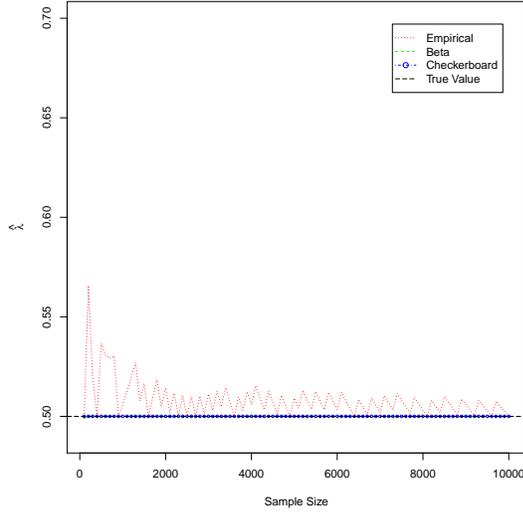
3.4 Estimation of extremes for L^1 -expectiles



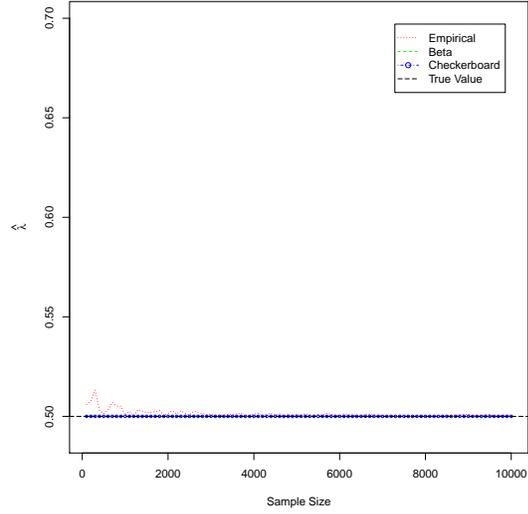
(a) Limiting subsequence $\ell_\lambda = n^{0.25}$



(b) Limiting subsequence $\ell_\lambda = n^{0.40}$



(c) Limiting subsequence $\ell_\lambda = n^{0.50}$



(d) Limiting subsequence $\ell_\lambda = n^{0.75}$

Figure 3.2: Comparison between $\hat{\lambda}_n^{23}(0.5, 1.5)$, $\hat{\lambda}_\beta^{23}(0.5, 1.5)$, $\hat{\lambda}_{\mathbf{x}}^{23}(0.5, 1.5)$ and $\lambda^{23}(0.5, 1.5)$. Note that for these estimates $\ell_\theta = \ell_i = n^{0.75}$. The margins are Pareto, specifically, $X_i \sim P(2, 1.25(1+i))$ for $i \in \{1, 2, 3\}$.

empirical process

$$B_{n,\ell_\lambda}^\beta = \sqrt{\ell_\lambda}(\mathcal{L}_{n,\ell_\lambda}^\beta - \mathcal{L}),$$

where

$$\mathcal{L}_{n,\ell_\lambda}^\beta(\mathbf{x}) = \frac{n}{\ell_\lambda} \left\{ 1 - C_n^\beta \left(1 - \frac{\ell_\lambda}{n} \mathbf{x} \right) \right\}$$

is the stdf defined under the empirical beta copula. In particular, they established the weak convergence (as defined in [van der Vaart and Wellner \(1996\)](#)) $B_{n,\ell_\lambda}^\beta \rightsquigarrow B$ on $\ell^\infty([0, 1])$, where B is some

3.4 Estimation of extremes for L^1 -expectiles

stochastic process with continuous trajectories. Using this result, and the bivariate relationship that for any stdf $\lambda(x, y) = x + y - \mathcal{L}(x, y)$, the empirical process for the udf based on the empirical beta copula will have the same (up to a sign) asymptotic distribution. This implies that

$$\hat{\lambda}_\beta^{ik}(x_i, x_k) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \lambda^{ik}(x_i, x_k),$$

establishing the result. □

In particular, Proposition 3.4 gives

$$\hat{\lambda}_\beta \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \lambda \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right).$$

Finally, we require a consistent estimator for (3.15). The following proposition establishes that the integral of (3.18) is such an estimator.

Proposition 3.5. *Consider the estimator of λ as proposed in (3.18), then*

$$\int_{\beta_i/\beta_k}^{\infty} \hat{\lambda}_\beta^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \int_{\beta_i/\beta_k}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt.$$

Proof. The proof boils down to an application of the Dominated Convergence Theorem (Royden, 1988; Billingsley, 2008). First, the convergence of the integrand follows from Proposition 3.4. Second, as the empirical beta copula is a proper copula, it is easy to see that its udf satisfies the bounds

$$\lambda^{ik,+} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \leq \hat{\lambda}_\beta^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \leq \lambda^{ik,+} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right)$$

and hence

$$0 \leq \hat{\lambda}_\beta^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \leq \min \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right).$$

See for instance Proposition 2.1 in Kiriliouk et al. (2018). Finally, it is shown in Maume-Deschamps et al. (2018) (see the proof of Lemma 3.4) that

$$\begin{aligned} \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik,+} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt &= \frac{c_i}{c_k} \left(\frac{\beta_i}{\beta_k} \right)^{-\theta+1} \left(\left\{ \frac{\beta_k}{\beta_i} \left(\frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} - 1 \right\}_+ \frac{c_k}{c_i} \left(\frac{\beta_k}{\beta_i} \right)^{-\theta} \right. \\ &\quad \left. + \frac{1}{\theta-1} \left[1 + \left\{ \frac{\beta_i}{\beta_k} \left(\frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} - 1 \right\}_+ \right]^{-\theta+1} \right) < \infty. \end{aligned}$$

Thus, by dominated convergence,

$$\int_{\beta_i/\beta_k}^{\infty} \lambda_\beta \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \int_{\beta_i/\beta_k}^{\infty} \lambda \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt,$$

3.4 Estimation of extremes for L^1 -expectiles

as required. □

To our knowledge, a consistent estimator for the integral of the udf has not been established in prior literature. In Figure 3.3, an illustration of the finite-sample performance of the estimator is provided. In particular, for each sample size and subsequence combination, the calculation of $\int_{\beta_2/\beta_3}^{\infty} \lambda_{\beta}^{23} (c_2 t^{-\theta}/c_3, 1) dt$ was replicated 500 times. As can be seen for samples of size $n > 400$ and subsequences n^q for $q > 0.4$, the estimator performs well. For a final comparison, we consider an estimator of (3.15) using the standard empirical copula. In fact, said estimator has a closed form which is derived in Proposition 3.6. The proof is left for Appendix 3.A.

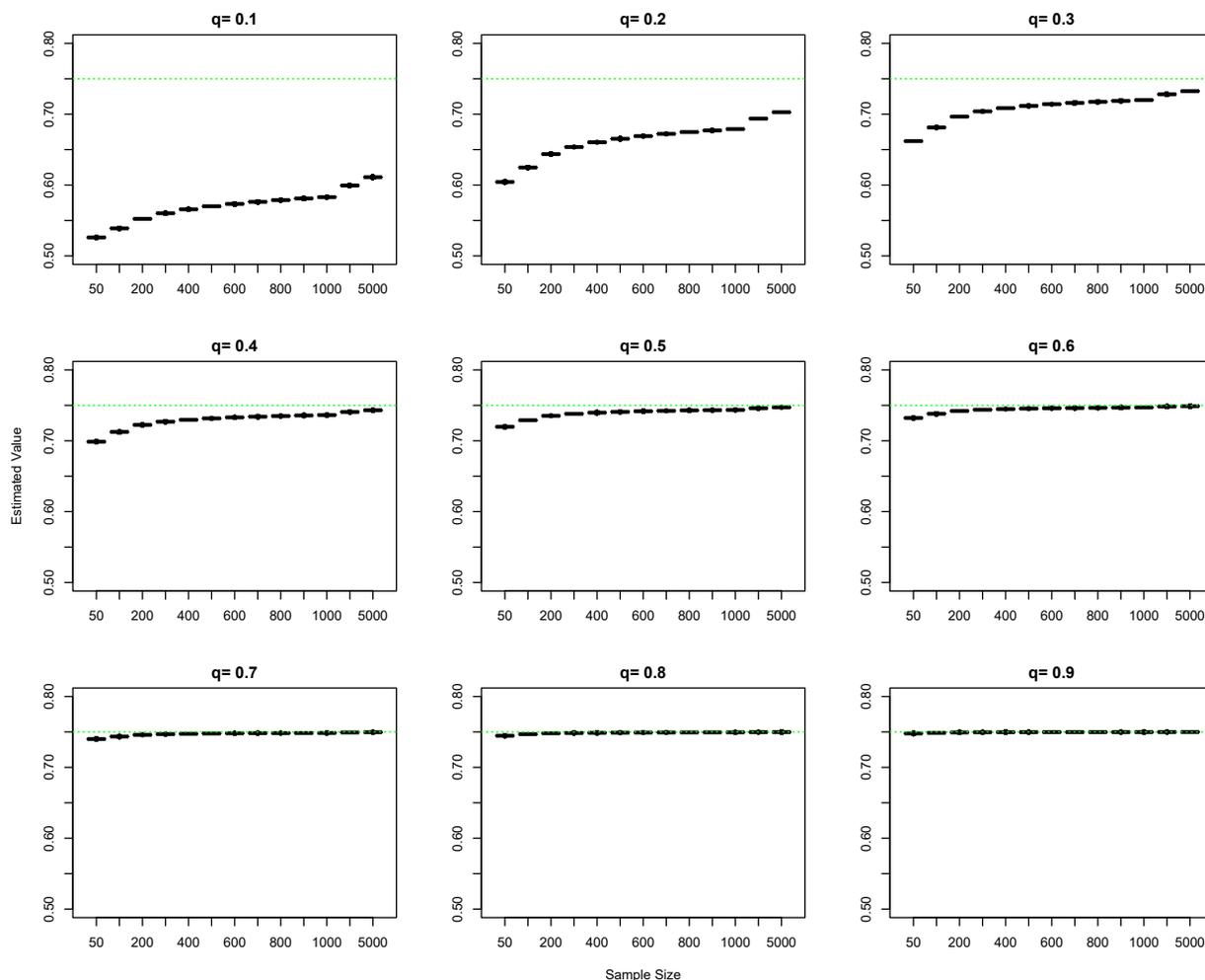


Figure 3.3: Performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_{\beta}^{23} (c_2 t^{-\theta}/c_3, 1) dt$ for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_{\lambda} = n^q$ for $q \in \{0.1, 0.2, \dots, 0.9\}$. They are compared to the true value $\int_{\beta_2/\beta_3}^{\infty} \lambda_{\beta}^{23} (c_2 t^{-\theta}/c_3, 1) dt$ under Pareto margins with a comonotonic dependence structure (green horizontal line). Here $X_i \sim P(2, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$.

3.4 Estimation of extremes for L^1 -expectiles

Proposition 3.6. *Under the empirical copula*

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1} \left\{ R_{ij}^{(n)} / n \leq u_j \right\},$$

it holds that

$$\int_{\beta_i/\beta_k}^{\infty} \hat{\lambda}_n^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt = \frac{1}{\ell_\lambda} \sum_{j=1}^n \left[\left\{ \frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right\}^{-1/\theta} - \frac{\beta_i}{\beta_k} \right]_+ \times \mathbb{1} (X_{kj} \geq X_{k:n-\ell_\lambda, n}),$$

where ℓ_λ is an intermediate integer subsequence.

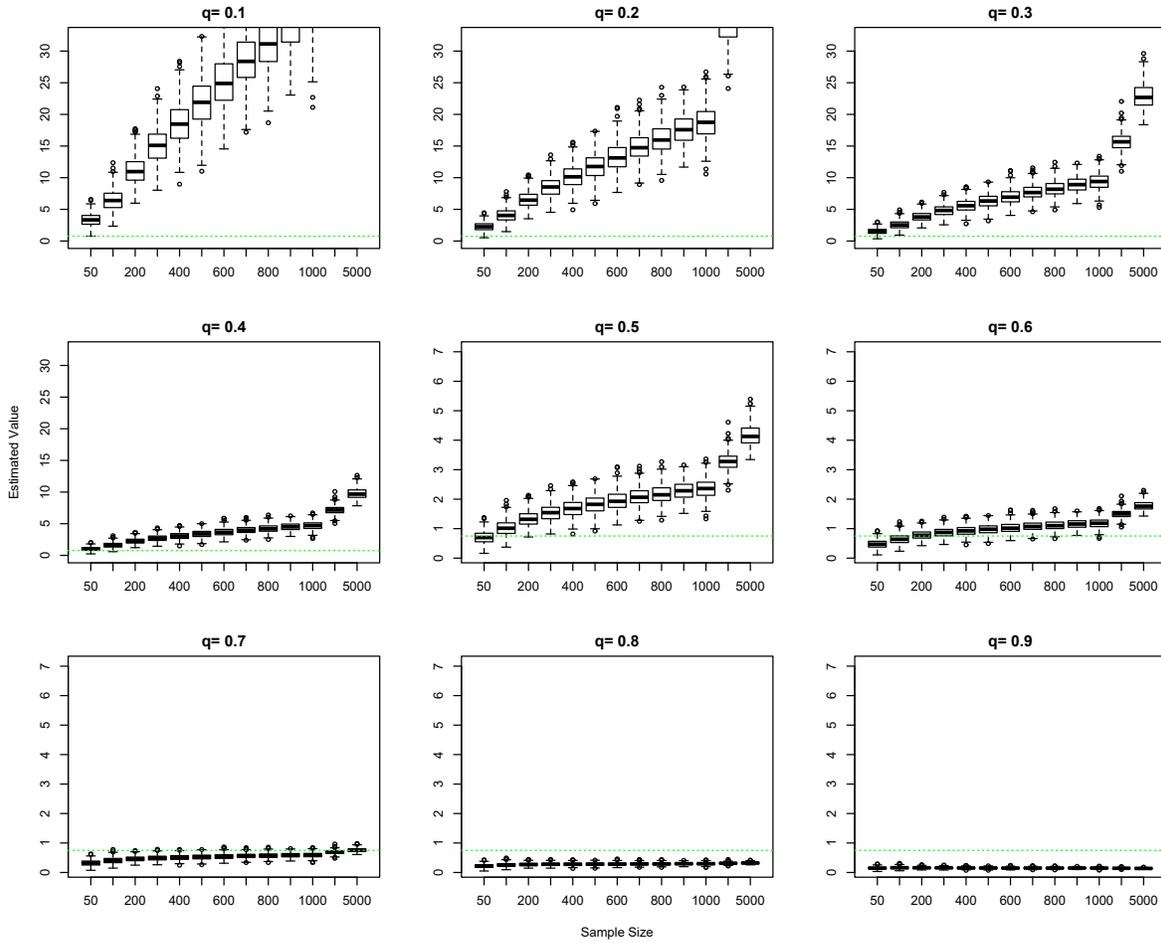


Figure 3.4: Performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_n^{23} (c_2 t^{-\theta} / c_3, 1) dt$ as in Proposition 3.6 for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_\lambda = n^q$ for $q \in \{0.1, 0.2, \dots, 0.9\}$. They are compared to the true value $\int_{\beta_2/\beta_3}^{\infty} \lambda^{23} (c_2 t^{-\theta} / c_3, 1) dt$ under Pareto margins with a comonotonic dependence structure (green horizontal line). Here $X_i \sim P(2, 1.25(1+i))$ for $i \in \{1, 2, 3\}$.

3.4 Estimation of extremes for L^1 -expectiles

Figure 3.4 illustrates the performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_n (c_2 t^{-\theta}/c_3, 1) dt$ as in Proposition 3.6. Comparing to Figure 3.3, one sees that the empirical beta copula far surpasses the standard empirical copula in terms of accuracy.

Before establishing the consistency of our estimation procedure, the next step is to combine the results of Deheuvels et al. (1988), Maume-Deschamps et al. (2018), Propositions 3.4 and 3.5 for $\hat{\theta}$, \hat{c}_2, \dots, c_d , $\hat{\lambda}_\beta$ and $\int \hat{\lambda}_\beta$, respectively.

Corollary 3.2. *Taking $\hat{\Lambda} = (\hat{\theta}, \hat{c}_2, \dots, \hat{c}_d, \hat{\lambda}_\beta^{ik})$ as defined in (3.16), (3.17) and (3.18), one has*

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}_\beta^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) dt \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt.$$

Proof. For simplicity, we establish a similar result for $\hat{\lambda}_\beta^{ik}$, i.e.,

$$\hat{\lambda}_\beta^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right).$$

The proof makes use of the continuous mapping theorem (CMT); see, e.g., van der Vaart (1998), Theorem 2.3. Several iterations of the triangle inequality yield

$$\begin{aligned} & \left| \hat{\lambda}_\beta^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| \\ &= \left| \hat{\lambda}_\beta^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| \\ &= \left| \hat{\lambda}_\beta^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) \right| \\ &\quad + \left| \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) + \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| \\ &\leq \underbrace{\left| \hat{\lambda}_\beta^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) \right|}_{(i)} + \underbrace{\left| \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) \right|}_{(ii)} \\ &\quad + \underbrace{\left| \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right|}_{(ii)} + \underbrace{\left| \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right|}_{(iii)} \\ &\xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0, \end{aligned}$$

where

- (i) by convergence of $\hat{\lambda}_\beta^{ik}$ (Proposition 3.4),
- (ii) by the CMT on λ^{ik} and the consistency of \hat{c}_i , $i \in \{1, \dots, d\}$ (Maume-Deschamps et al. (2018), Proposition 4.2), and

3.4 Estimation of extremes for L^1 -expectiles

(iii) by the CMT on λ^{ik} and the consistency of $\hat{\theta}$ (de Haan and Ferreira (2006), Theorem 3.2.2).

The proof of the main statement follows in the same way, or as a result of the dominated convergence theorem, similar to Proposition 3.5. \square

For illustration, simulations are presented in Figure 3.5. It is clear that adding extra uncertainty through calculation of $\hat{\theta}$ and \hat{c}_i for $i \in \{2, \dots, d\}$ decreases the accuracy of the full estimator. The variability of these quantities is well documented and this would naturally require an increased number of observations to estimate with confidence. Moreover, the selection of the subsequences ℓ_θ and ℓ_i for $i \in \{2, \dots, d\}$ is crucial. While in a simulation study the subsequences were chosen to be fixed for ease of replication, in a real data study the subsequences could be calibrated using Hill plots.

Finally, in Theorem 3.7, we proceed with the main result.

Theorem 3.7. *Let $\hat{\Lambda} = (\hat{\theta}, \hat{c}_2, \dots, \hat{c}_d, \hat{\lambda}_{\text{Beta}}^{ik})$ as in (3.16), (3.17) and (3.18) respectively. Then*

$$L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} L_{\Lambda}(\Theta) \quad \text{and} \quad \nabla L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \nabla L_{\Lambda}(\Theta). \quad (3.19)$$

Moreover, given identical starting values Θ^0 , H_0 , $\sigma \in (0, 1/2)$, $\rho \in (\sigma, 1)$ and $\epsilon \geq 0$, for any step k , it holds that

$$\hat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Theta^k. \quad (3.20)$$

Proof. Noting the structure of L_{Λ} and $\nabla L_{\Lambda}(\Theta)$, the consistency results in (3.19) are a direct result of the consistency of $\hat{\theta}$, \hat{c}_i for $i \in \{2, \dots, d\}$, Propositions 3.4, 3.5 and Slutsky's Theorem. For (3.20), we use induction. First, fix Θ^0 , H_0 , $\sigma \in (0, 1/2)$, $\rho \in (\sigma, 1)$ and $\epsilon \geq 0$. Recalling the iteration step of BFGS algorithm, we have

$$\hat{\mathbf{d}}^0 = -H_0 \nabla L_{\hat{\Lambda}}(\Theta^0) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} -H_0 \nabla L_{\Lambda}(\Theta^0) = \mathbf{d}^0.$$

Similarly, it is clear that $\hat{t}_0 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} t_0$. It follows that

$$\hat{\Theta}^1 = \Theta^0 + \hat{t}_0 \hat{\mathbf{d}}^0 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Theta^0 + t_0 \mathbf{d}^0 = \Theta^1,$$

as well as $\hat{\mathbf{y}}^0 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbf{y}^0$, $\hat{\mathbf{s}}^0 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbf{s}^0$ and $\hat{H}_1 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} H_1$. Next, assume $\hat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Theta^k$. Then,

$$\hat{\Theta}^{k+1} = \hat{\Theta}^k + \hat{t}_k \hat{\mathbf{d}}_k = \hat{\Theta}^k + \hat{t}_k \hat{H}_k^{-1} \nabla L_{\hat{\Lambda}}(\hat{\Theta}^k) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Theta^k + t_k H_k^{-1} \nabla L_{\Lambda}(\Theta^k) = \Theta^{k+1}.$$

Thus, for any step $k \in \{1, 2, \dots\}$, $\hat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Theta^k$ by induction. \square

3.4 Estimation of extremes for L^1 -expectiles

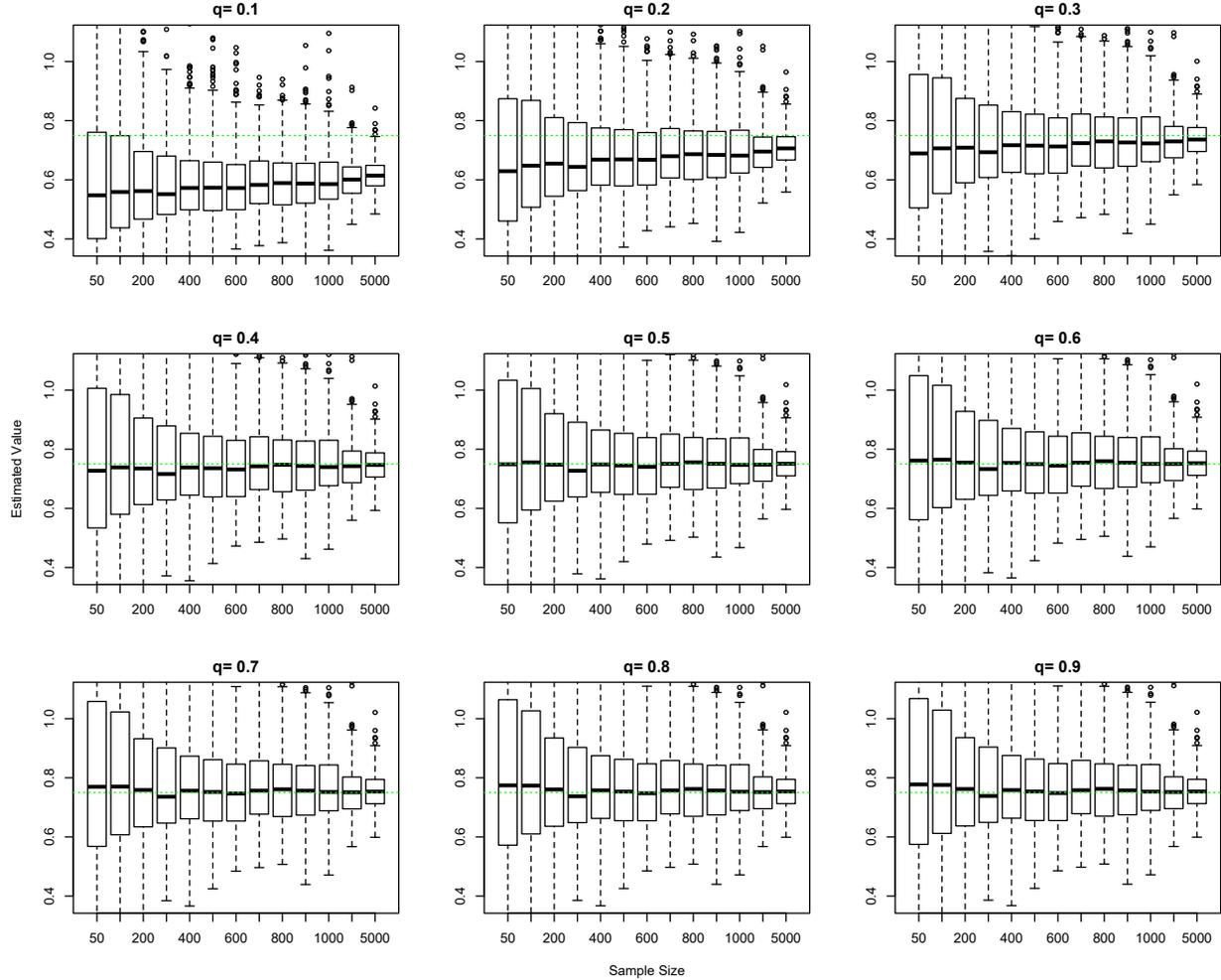


Figure 3.5: Performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_{\beta}^{23}(\hat{c}_2 t^{-\hat{\theta}}/\hat{c}_3, 1) dt$ for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_{\lambda} = n^q$ for $q \in \{0.1, 0.2, \dots, 0.9\}$. They are compared to the true value $\int_{\beta_2/\beta_3}^{\infty} \lambda^{23}(c_2 t^{-\theta}/c_3, 1) dt$ under Pareto margins with a comonotonic dependence structure (green horizontal line). Here $X_i \sim P(2, 1.25(1+i))$ for $i \in \{1, 2, 3\}$.

We now proceed by using an iterated two-step procedure. We first provide an adequate estimate of the true loss function (and its gradient) and then proceed with the optimization procedure. This is formalized in [Algorithm 2](#).

Algorithm 2 Two-steps estimation procedure for MEEs

(Step 1) Taking the limit $n \rightarrow \infty$

Establish the consistency of $\hat{\Lambda}$ and subsequently $L_{\hat{\Lambda}}$ and $\nabla L_{\hat{\Lambda}}$ (see [Theorem 3.7](#)).

(Step 2) Taking the limit $k \rightarrow \infty$

Optimize the consistently approximated problem from Step 1 using the BFGS algorithm.

The resulting asymptotic behavior of our estimation scheme using [Algorithm 2](#) is established in

3.4 Estimation of extremes for L^1 -expectiles

Corollary 3.3. It is important to note the limit order in the non-exchangeable iterated limit (3.21).

Corollary 3.3. *Under the assumptions of Corollaries 3.1, 3.2 and Theorem 3.7, and further assuming that BFGS algorithm solves for the global minimum of (3.12), then, it holds that*

$$\lim_{k \rightarrow \infty} (\text{plim}_{n \rightarrow \infty} \hat{\Theta}^k) = \Theta^*. \quad (3.21)$$

Proof. Under the assumption that BFGS solves for the global minimum of (3.12), i.e., $\Theta^k \xrightarrow[k \rightarrow \infty]{} \Theta^*$, we have

$$\begin{aligned} \lim_{k \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \mathbb{P} \left(\|\hat{\Theta}^k - \Theta^*\| \geq \epsilon \right) \right\} &= \lim_{k \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \mathbb{P} \left(\|\hat{\Theta}^k - \Theta^k + \Theta^k - \Theta^*\| \geq \epsilon \right) \right\} \\ &\leq \lim_{k \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \mathbb{P} \left(\|\hat{\Theta}^k - \Theta^k\| + \|\Theta^k - \Theta^*\| \geq \epsilon \right) \right\} \\ &\leq \lim_{k \rightarrow \infty} \left[\lim_{n \rightarrow \infty} \left\{ \underbrace{\mathbb{P} \left(\|\hat{\Theta}^k - \Theta^k\| \geq \epsilon/2 \right)}_{(i)} \right. \right. \\ &\quad \left. \left. + \underbrace{\mathbb{P} \left(\|\Theta^k - \Theta^*\| \geq \epsilon/2 \right)}_{(ii)} \right\} \right] \\ &= 0, \end{aligned}$$

where firstly (i) converges to zero as $n \rightarrow \infty$ by Theorem 3.7 and subsequently (ii), which does not depend on n , converges to 0 as $k \rightarrow \infty$ by assumption. \square

3.4.4 Examples

We substantiate the results of Section 3.4.3 with a series of numerical examples. First, we present a simulation study. For comparison, we consider the limiting cases of tail dependence: asymptotic independence and comonotonicity. As closed form solutions exist for Θ^\perp and Θ^+ , provided in Lemmas 3.1 and 3.2, respectively, they will provide a good measuring stick for our estimation procedure. Equivalently, we can compare the explicit expectiles estimates using the values of $e_\alpha^\perp(\mathbf{X})$ and $e_\alpha^+(\mathbf{X})$ given in (3.10) and (3.11), respectively. In this case, the estimator for MEEs is written

$$\hat{e}_\alpha(\mathbf{X}) = \widehat{\text{VaR}}_\alpha^*(X_1) \hat{\eta}^{1/\hat{\theta}} (1, \hat{\beta}_2, \dots, \hat{\beta}_d)^\top$$

where

$$\widehat{\text{VaR}}_\alpha^*(X_1) = X_{1:n-k(n)+1,n} \left\{ \frac{k(n)}{(1-\alpha)n} \right\}^{1/\hat{\theta}}, \quad (3.22)$$

3.4 Estimation of extremes for L^1 -expectiles

called Weissman's estimator (Weissman, 1978), for an intermediate integer sequence $k(n)$ is better suited than $\widehat{\text{VaR}}_\alpha(X)$ for $\alpha \approx 1$. For more information regarding $\widehat{\text{VaR}}_\alpha^*$, see, e.g., Embrechts et al. (1997), Beirlant et al. (2004). However, comparisons using $\hat{\Theta}^\perp$ and $\hat{\Theta}^+$ will more effectively demonstrate the capabilities of our algorithm as they do not rely on the performance of (3.22).

For the simulations, we considered a random vector \mathbf{X} made up of Pareto type I marginals. Specifically, $X_i \sim \text{Pa}(b_i, a)$, where $a = 3.5$ and $b_i = 1.25(i + 1)$ for $i \in 1, \dots, d$ and $d \in \{3, 5\}$. In this scenario, we have that $\bar{F}_1, \dots, \bar{F}_d \in \text{RV}_{3.5}(\infty)$ and

$$c_i = \lim_{x \rightarrow \infty} \frac{\bar{F}_i(x)}{\bar{F}_1(x)} = \left(\frac{b_i}{b_1} \right)^a = \begin{cases} 1.000, & \text{when } i = 1, \\ 4.134, & \text{when } i = 2, \\ 11.314, & \text{when } i = 3, \\ 24.705, & \text{when } i = 4, \\ 46.765, & \text{when } i = 5, \end{cases}$$

where $\bar{F}_1, \dots, \bar{F}_d$ are the margins sf's for X_1, \dots, X_d , respectively. Simulations were conducted for samples of size $n = 50, 100, 250, 500, 1000, 2500, 5000$ and repeated 500 times each. The limiting integer subsequences for θ , c_i for $i \in \{2, \dots, d\}$ and λ were $\ell_\theta = \ell_2 = \dots = \ell_d = n^{0.75}$ and $\ell_\lambda = n^{0.5}$, respectively. The risk level was set to $\alpha = 1 - 1/5000 = 0.9998$ to reflect the largest sample size considered. First, we consider the three-dimensional case. Under asymptotic independence and comonotonicity the true system solutions are

$$\Theta^\perp = (0.0740, 1.764, 2.639) \quad \text{and} \quad \Theta^+ = (0.400, 1.500, 2.000),$$

respectively, with corresponding expectiles

$$e_{0.9998}^\perp(\mathbf{X}) = (13.545, 23.894, 35.744) \quad \text{and} \quad e_{0.9998}^+(\mathbf{X}) = (21.932, 32.899, 43.865).$$

The results of the simulations are summarized in Table 3.1 and Figure 3.6. In the three-dimensional scenario, our proposed estimation procedure performs quite well. At lower samples, the results are only slightly biased though display a fair amount of standard error. As the sample size increases however, the algorithm is able to successfully estimate the true values of both Θ and $e_\alpha(\mathbf{X})$ for $\alpha = 0.9998$ with minimal standard errors. It should be noted that for the comonotonic simulations, the slight bias and small standard errors would mean that the true value would likely not be contained in any reasonable confidence interval, though the bias is generally quite small and continues to shrink as the sample size increases. A possible explanation for this occurrence could be that the structure of a comonotonic random vector makes for less informative samples given the way it is generated; every realization of the random vector comes from the same uniform observation. See Section 3.4.4 for more information on simulation for a comonotonic dependence structure.

3.4 Estimation of extremes for L^1 -expectiles

Table 3.1: Median results of simulation study for Pareto margins in three dimensions. Standard deviations for each estimate are presented in parentheses. Here $X_i \sim P(3.5, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$.

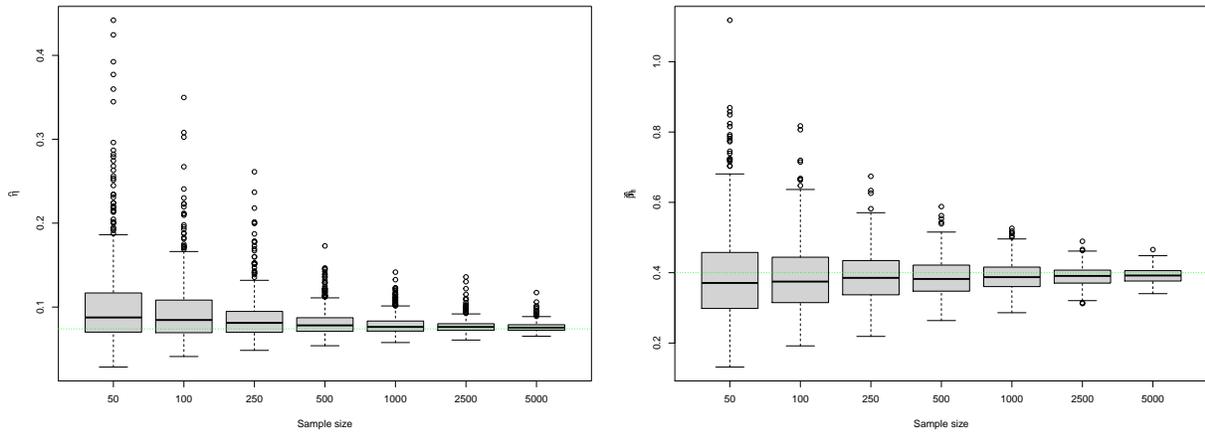
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.088 (0.056)	1.811 (0.382)	2.726 (0.588)	14.814 (9.370)	27.332 (20.79)	40.634 (36.25)
100	0.085 (0.039)	1.769 (0.264)	2.629 (0.426)	14.438 (5.674)	25.757 (11.88)	38.158 (19.61)
250	0.081 (0.028)	1.771 (0.189)	2.660 (0.304)	14.210 (3.700)	25.359 (7.075)	37.745 (11.78)
500	0.078 (0.017)	1.758 (0.132)	2.660 (0.211)	13.725 (2.488)	24.333 (4.833)	36.347 (7.912)
1000	0.076 (0.012)	1.770 (0.110)	2.644 (0.160)	13.786 (1.749)	24.466 (3.529)	36.468 (5.638)
2500	0.076 (0.008)	1.766 (0.073)	2.639 (0.115)	13.682 (1.105)	24.308 (2.169)	36.117 (3.536)
5000	0.075 (0.006)	1.765 (0.052)	2.639 (0.091)	13.636 (0.834)	24.060 (1.652)	36.006 (2.701)
Θ	0.074	1.764	2.639	13.545	23.894	35.744

(a) Independent Margins

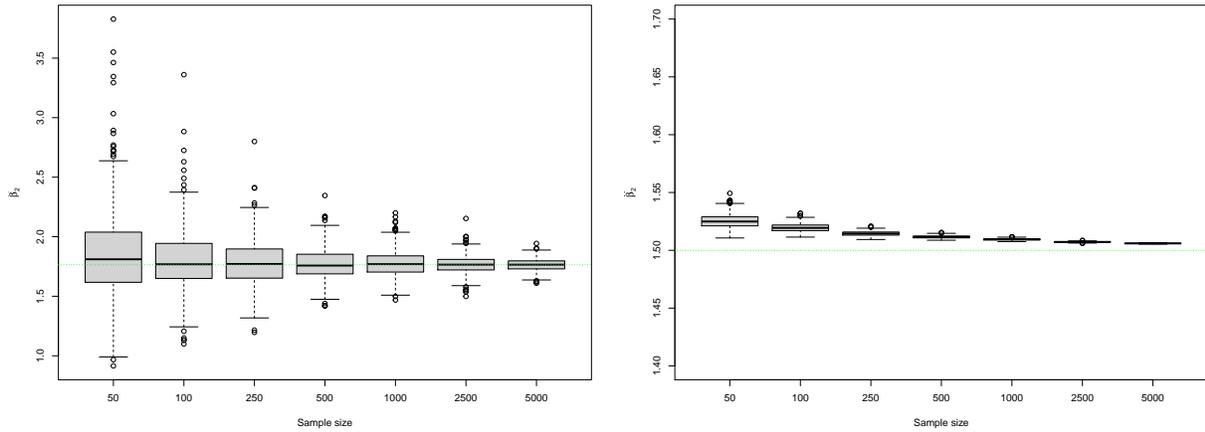
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.371 (0.137)	1.525 (0.006)	2.062 (0.016)	22.523 (20.60)	34.338 (31.81)	46.446 (43.49)
100	0.375 (0.099)	1.519 (0.004)	2.049 (0.009)	21.737 (11.68)	33.028 (17.89)	44.534 (24.29)
250	0.385 (0.072)	1.514 (0.002)	2.036 (0.005)	22.109 (7.243)	33.485 (11.03)	45.029 (14.89)
500	0.382 (0.054)	1.512 (0.001)	2.029 (0.003)	21.580 (4.952)	32.618 (7.514)	43.789 (10.12)
1000	0.387 (0.041)	1.509 (0.001)	2.024 (0.002)	21.797 (3.555)	32.900 (5.383)	44.114 (7.238)
2500	0.391 (0.028)	1.507 (0.000)	2.019 (0.001)	21.834 (2.277)	32.908 (3.441)	44.069 (4.619)
5000	0.392 (0.022)	1.506 (0.000)	2.016 (0.001)	21.810 (1.761)	32.845 (2.658)	43.976 (3.563)
Θ	0.400	1.500	2.000	21.932	32.899	43.865

(b) Comonotonic Margins

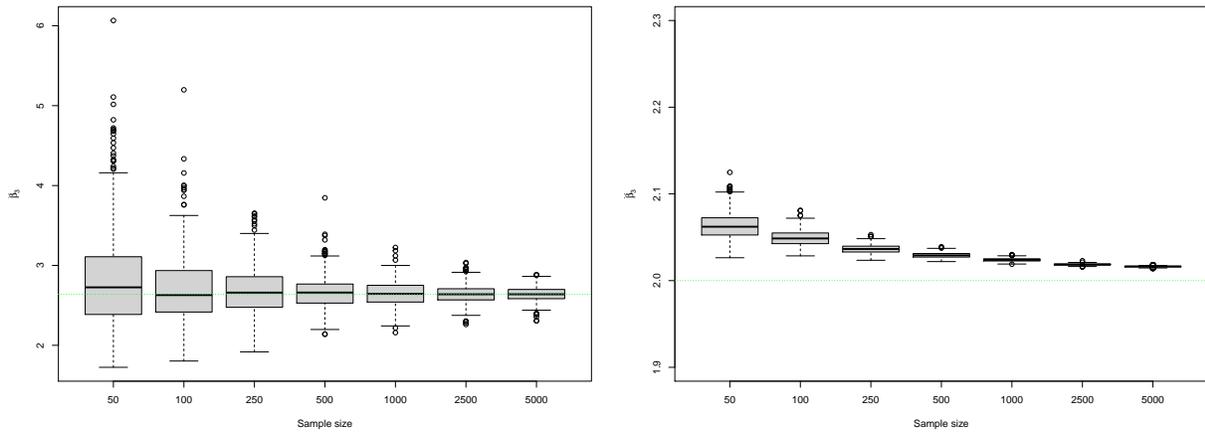
3.4 Estimation of extremes for L^1 -expectiles



(a) Results for $\hat{\eta}$



(b) Results for $\hat{\beta}_2$



(c) Results for $\hat{\beta}_3$

Figure 3.6: Boxplots for the three-dimensional Pareto simulations. Here $X_i \sim P(3.5, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$. Results for independent margins (left) and comonotonic margins (right). True values Θ^\perp and Θ^+ (dashed green lines) provided for comparison.

3.4 Estimation of extremes for L^1 -expectiles

Table 3.2: Median results of simulation study for Pareto margins in five dimensions linked with the independence copula. Standard deviations for each estimate are presented in parentheses. Here $X_i \sim P(3.5, 1.25(1+i))$ for $i \in \{1, \dots, 5\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$.

n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
50	0.040 (0.058)	1.779 (0.773)	2.672 (1.159)	3.641 (1.405)	4.662 (1.887)
100	0.037 (0.033)	1.760 (0.404)	2.622 (0.602)	3.553 (0.851)	4.565 (1.136)
250	0.032 (0.017)	1.779 (0.288)	2.674 (0.410)	3.667 (0.547)	4.783 (0.757)
500	0.032 (0.012)	1.783 (0.205)	2.668 (0.300)	3.661 (0.410)	4.690 (0.545)
1000	0.031 (0.008)	1.782 (0.142)	2.653 (0.206)	3.661 (0.286)	4.723 (0.386)
2500	0.031 (0.005)	1.769 (0.109)	2.653 (0.148)	3.638 (0.210)	4.695 (0.282)
5000	0.030 (0.004)	1.763 (0.073)	2.645 (0.108)	3.618 (0.154)	4.667 (0.205)
Θ	0.029	1.764	2.639	3.607	4.655

(a) Parameter estimates.

n	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$	$\hat{e}_\alpha^4(\mathbf{X})$	$\hat{e}_\alpha^5(\mathbf{X})$
50	12.301 (7.676)	21.755 (22.48)	33.493 (31.78)	46.092 (43.99)	58.649 (53.44)
100	11.546 (4.261)	19.608 (8.535)	29.990 (12.93)	41.265 (19.67)	53.211 (26.50)
250	10.968 (2.510)	19.893 (5.057)	29.810 (7.971)	41.149 (11.05)	52.947 (15.67)
500	10.761 (1.870)	19.108 (3.689)	28.967 (5.726)	38.747 (8.207)	50.317 (11.57)
1000	10.653 (1.261)	19.185 (2.324)	28.377 (3.698)	38.982 (5.352)	50.274 (7.274)
2500	10.660 (0.784)	18.776 (1.591)	28.107 (2.491)	38.583 (3.622)	49.900 (4.956)
5000	10.581 (0.628)	18.653 (1.108)	27.999 (1.823)	38.329 (2.632)	49.508 (3.671)
$e_{0.9998}(\mathbf{X})$	10.390	18.330	27.420	37.475	48.372

(b) Expectiles estimates.

The results of the simulations for $d = 5$ are presented in Table 3.2. It can be seen that the algorithm scales well to the increased dimension. This could be explained in part by the fact we are only ever dealing with pairwise dependencies through the pairwise udf λ^{ik} .

To further substantiate the performance of our estimation procedure, we compare the results in Table 3.1 to simulations using the fully parametric estimators

$$\hat{\eta}^\perp = \frac{1}{(\hat{\theta} - 1) \left(1 + \sum_{j=2}^d \hat{c}_j^{\frac{1}{\hat{\theta}-1}}\right)} \quad \text{and} \quad \hat{\beta}_k^\perp = \hat{c}_k^{\frac{1}{\hat{\theta}-1}}$$

and

$$\hat{\eta}^+ = \frac{1}{\hat{\theta} - 1} \quad \text{and} \quad \hat{\beta}_k^+ = \hat{c}_k^{\frac{1}{\hat{\theta}}},$$

based on (3.8) and (3.9), respectively, from Maume-Deschamps et al. (2018). Using these fully

3.4 Estimation of extremes for L^1 -expectiles

parametric estimators for the system solutions, extreme expectiles estimators can be written

$$\hat{e}_\alpha^\perp(\mathbf{X}) = \widehat{\text{VaR}}_\alpha(X_1) \left(\frac{1}{\hat{\theta} - 1} \right)^{\frac{1}{\theta}} \left(\frac{1}{1 + \sum_{i=1}^d \hat{c}_i^{\frac{1}{\theta-1}}} \right) \left(1, \hat{c}_2^{\frac{1}{\theta-1}}, \dots, \hat{c}_d^{\frac{1}{\theta-1}} \right)^\top$$

and

$$\hat{e}_\alpha^+(\mathbf{X}) = \widehat{\text{VaR}}_\alpha(X_1) \left(\frac{1}{\hat{\theta} - 1} \right)^{\frac{1}{\theta}} \left(1, \hat{c}_2^{\frac{1}{\theta}}, \dots, \hat{c}_d^{\frac{1}{\theta}} \right)^\top$$

for independent and comonotonic random vectors, respectively, where $\widehat{\text{VaR}}_\alpha(X_1)$ is Weissman's estimator (Weissman, 1978) in Equation (3.22). Analogously to Table 3.1, in Table 3.3 we consider median results of simulation study for independent and comonotonic Pareto margins in three dimensions. Standard deviations for each estimate are presented in parentheses. Here $X_i \sim P(3.5, 1.25(1 + i))$ for $i \in \{1, 2, 3\}$.

Unsurprisingly, the fully parametric estimators perform better in particular for smaller sample sizes and have smaller standard errors. However, our semi-parametric method (see Table 3.1) does not require any pre-imposed dependence assumption and still provides satisfactory results. Note that the standard errors being zero for $\hat{\beta}_2$ and $\hat{\beta}_3$ for the comonotonic case are a result of the estimators structures and the distributional assumptions, *i.e.*,

$$\hat{\beta}_i^+ = \hat{c}_i^{\frac{1}{\theta}} = \frac{X_{i:n-\ell_i+1,n}}{X_{i:n-\ell_i+1,n}} = \frac{F_i^{-1}(U_{n-\ell_i+1,n})}{F_1^{-1}(U_{n-\ell_i+1,n})} = \frac{b_i}{b_1}$$

for $i \in \{2, 3\}$, where $U_{n-\ell_i+1,n}$ is the $n - \ell_i + 1, n$ th order statistic from the standard uniform sample used to generate the data. For comonotonic random vectors, each realization of the copula is generated using the same uniform realization for each component, *i.e.*, $X_{ji} = F_j^{-1}(U_i)$ for $j \in \{1, \dots, d\}$ and $i \in \{1, \dots, n\}$.

In the full nonparametric framework, Maume-Deschamps et al. (2017) use Robbins-Monro's stochastic optimization to solve the system of equations

$$\alpha \sum_{i=1}^d \mathbb{E} [(X_i - x_i)_+ \mathbb{1}\{X_k > x_k\}] = (1 - \alpha) \sum_{i=1}^d \mathbb{E} [(X_i - x_i)_- \mathbb{1}\{X_k < x_k\}], \quad \forall k \in \{1, \dots, d\}.$$

For a comparison with our results under the same independent Pareto model, the interested reader is referred to Figure 7 in Maume-Deschamps et al. (2017) (page 40). It is easy to see that for elevated levels of α , the nonparametric approach solved through Robbins-Monro's stochastic optimization performs quite poorly.

3.4 Estimation of extremes for L^1 -expectiles

Table 3.3: Analogous results to Table 3.1 using the full parametric estimators for $\hat{\Theta}$ and \hat{e}_α for independent and comonotonic random vectors. Here $X_i \sim P(3.5, 1.25(1+i))$ for $i \in \{1, \dots, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$.

n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.075 (0.022)	1.794 (0.231)	2.672 (0.407)	13.917 (8.231)	25.373 (18.82)	38.092 (31.18)
100	0.074 (0.016)	1.770 (0.167)	2.645 (0.304)	13.569 (5.119)	24.443 (10.67)	36.305 (17.80)
250	0.075 (0.012)	1.770 (0.124)	2.656 (0.214)	13.736 (3.293)	24.610 (6.598)	36.895 (11.03)
500	0.074 (0.009)	1.765 (0.094)	2.639 (0.157)	13.540 (2.281)	23.933 (4.580)	35.459 (7.501)
1000	0.074 (0.007)	1.767 (0.075)	2.636 (0.126)	13.588 (1.626)	23.915 (3.340)	35.931 (5.452)
2500	0.074 (0.005)	1.767 (0.053)	2.637 (0.089)	13.560 (1.051)	23.957 (2.108)	35.685 (3.441)
5000	0.074 (0.004)	1.762 (0.042)	2.637 (0.074)	13.530 (0.804)	23.903 (1.602)	35.736 (2.667)
True	0.074	1.764	2.639	13.544	23.894	35.744

(a) Independent Margins

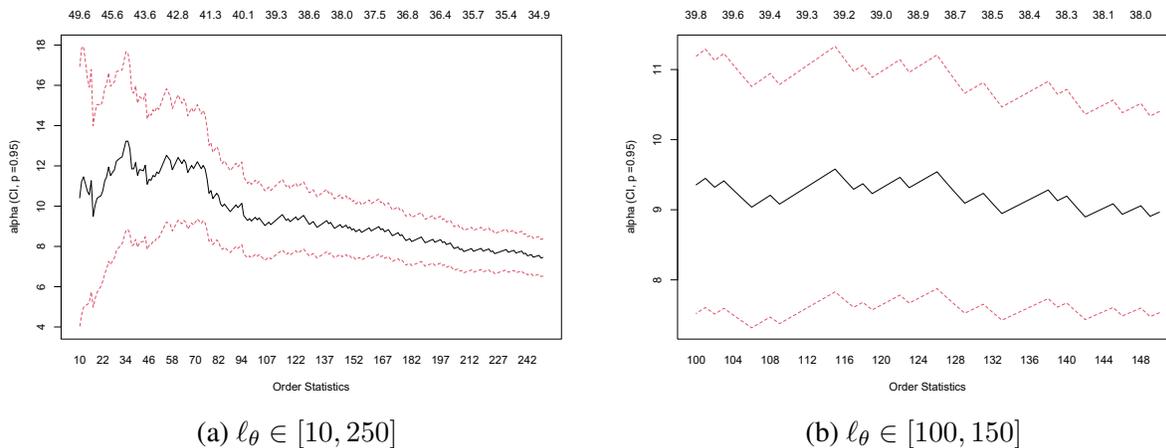
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.411 (0.151)	1.500 (0)	2.000 (0)	23.228 (21.68)	34.843 (32.53)	46.457 (43.37)
100	0.404 (0.106)	1.500 (0)	2.000 (0)	22.222 (12.08)	33.333 (18.12)	44.444 (24.16)
250	0.407 (0.076)	1.500 (0)	2.000 (0)	22.460 (7.411)	33.690 (11.12)	44.921 (14.82)
500	0.399 (0.057)	1.500 (0)	2.000 (0)	21.842 (5.040)	32.763 (7.560)	43.684 (10.08)
1000	0.401 (0.043)	1.500 (0)	2.000 (0)	22.006 (3.605)	33.009 (5.407)	44.011 (7.210)
2500	0.401 (0.029)	1.500 (0)	2.000 (0)	21.980 (2.303)	32.970 (3.454)	43.960 (4.606)
5000	0.400 (0.023)	1.500 (0)	2.000 (0)	21.939 (1.776)	32.908 (2.664)	43.878 (3.552)
Θ	0.400	1.500	2.000	21.933	32.899	43.865

(b) Comonotonic Margins

Finally, we also present an application of MEEs to a real data set. In particular, we consider the Pima Indians Diabetes Database (<https://www.kaggle.com/uciml/pima-indians-diabetes-database>). This data set has been used in previous studies of multivariate risk measures; see, e.g., Girard and Stupfler (2015), Chaouch and Goga (2010). The data set consists of nine variables (Pregnancies, Glucose, Diastolic Blood Pressure (DBP), Skin Thickness, Insulin Levels, body mass index (BMI), Diabetes Pedigree Function, Age and Outcome) for $n = 768$ individuals from the Pima Indian population. The study was conducted in an attempt to understand the prevalence of diabetes in this population. For comparison, we consider the two-dimensional case for pairs (X_{i1}, X_{i2}) where X_{i1} is the BMI and X_{i2} is the DBP for individual $i \in \{1, \dots, 768\}$.

The first step is to estimate $\hat{\theta}$ and \hat{c}_2 . Using the selection procedure described, for instance, by Cai et al. (2015), we consider a range of possibilities for intermediate sequences ℓ_θ and ℓ_2 for which the estimators showed some stability. Then, to gain robustness, we average the estimates corresponding to the selected ranges. The interested reader is also referred to Di Bernardino and Prieur (2018). For $\hat{\theta}$ this turned out to be values of $\ell_\theta \in [100, 150]$. The corresponding Hill plots are presented in Figure 3.7. Considering these values the tail index was estimated to be $\hat{\theta} = 9.126$.

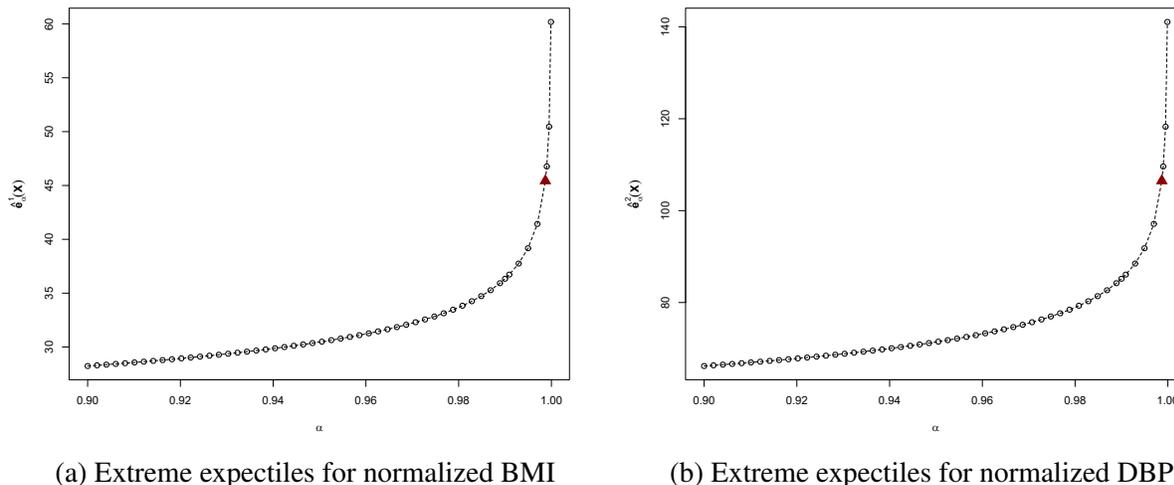
3.4 Estimation of extremes for L^1 -expectiles



(a) $\ell_\theta \in [10, 250]$

(b) $\ell_\theta \in [100, 150]$

Figure 3.7: Hill Plots for estimating $\hat{\theta}$ using BMI measurements.



(a) Extreme expectiles for normalized BMI

(b) Extreme expectiles for normalized DBP

Figure 3.8: Extreme Expectiles for Pima Indians Diabetes Study. Included is the extreme expectile at level $\alpha = 1 - 1/n$ (red triangle).

The same procedure is followed for \hat{c}_2 and the resulting estimated value is $\hat{c}_2 = 1103.046$. In Figure 3.8, the resulting expectiles are plotted.

Recalling the first order conditions of L^1 expectiles, $e_\alpha(\mathbf{X})$ is the value of \mathbf{x} such that

$$\frac{\sum_{i=1}^d \mathbb{E}[(X_i - x_i)_+ \mathbb{1}\{X_k > x_k\}]}{\sum_{i=1}^d \mathbb{E}[(X_i - x_i)_- \mathbb{1}\{X_k < x_k\}]} = \frac{1 - \alpha}{\alpha} \quad \forall k \in \{1, 2\}.$$

This could be interpreted as the value of \mathbf{x} which fixes the ratio of participation in positive scenarios to negative scenarios at $(1 - \alpha)/\alpha$ for all variables. Consider the MEE at level $\alpha = 1 - 1/768 = 0.9987$; we have $\hat{e}_{0.9987}(\mathbf{X}) = (45.433, 106.493)$. As was noted in similar studies of this data set, e.g., [Chaouch and Goga \(2010\)](#), an individual with a BMI above 31.7 kg/m^2 would qualify as

3.4 Estimation of extremes for L^1 -expectiles

severely overweight and a DBP value above 95 mm Hg would be at risk of hypertension. Both these factors are considered to impact the prevalence of diabetes.

3.4.5 Discussion

As mentioned in Section 3.4.1, in the general case the loss function in (3.12) could be non-convex, and thus, might not satisfy the assumptions required for global convergence of the BFGS algorithm. Nevertheless, we chose this method for its relative simplicity and ease of implementation, as well as the satisfactory numerical results it has provided for our problem. To build a more rigorous estimation procedure, one could consider implementing methods which build upon the foundation of the original BFGS algorithm. For instance, [Li and Fukushima \(2001a,b\)](#) discuss modifications to the BFGS algorithm which allow one to relax the assumption of convexity and still ensure global convergence. Specifically, they discuss the importance of Lipschitz continuity and using a damping method on the iterates of the approximate inverse Hessian matrix. Equivalently, [Yuan et al. \(2017\)](#) discuss modifications to the Wolfe conditions in [Algorithm 1](#) which ensures global convergence of BFGS for general functions. Such methods could further improve our results. However, while BFGS is already included in many R packages, the above methods are not readily available. Moreover, manually programming them and doing so efficiently would require a specific work which is beyond the scope of this thesis.

Additionally, while the BFGS algorithm works well for our problem, as exemplified in the simulation studies, it is important to note that its intended use is for unconstrained optimization problems. Clearly for MEEs of positively-valued random vectors it is required that $\Theta > \mathbf{0}$ componentwise. A further improvement to this problem would be incorporating a box-constrained methodology for solving (3.13) and, more importantly, (3.14). A clear choice in this case would be box-constrained BFGS (or BFGS-B). This algorithm extends the basic BFGS algorithm by incorporating simple box constraints into the optimization procedure, allowing us to integrate additional properties of MEEs. More specifically, as comonotonicity and independence represent the limiting cases of dependence in the upper tail, all expectiles will be bounded above and below by comonotonic and independent expectiles. Moreover, from [Lemmas 3.1 and 3.2](#) we know that for asymptotically independent and comonotonic random vectors the solutions to (3.13) are given by (3.8) and (3.9), respectively. This allows one to bound any solution Θ^* of (3.13) to the d -dimensional box $[\Theta_1^L, \Theta_1^U] \times \dots \times [\Theta_d^L, \Theta_d^U]$ where $\Theta_i^L = \min(\Theta_i^-, \Theta_i^+)$ and $\Theta_i^U = \max(\Theta_i^-, \Theta_i^+)$ for $i \in \{1, \dots, d\}$. However, because the estimation is done without prior knowledge of the underlying distribution, the box-constraints in the algorithm must also be estimated. To this end, the bounds of each box will be slightly relaxed. In other words, the algorithm would look for a solution in

$$[\hat{\Theta}_1^L, \hat{\Theta}_1^U] \times \dots \times [\hat{\Theta}_d^L, \hat{\Theta}_d^U],$$

3.5 Conclusion

where $\hat{\Theta}_i^L = (1 - \gamma_L) \min(\hat{\Theta}_i^-, \hat{\Theta}_i^+)$ and $\hat{\Theta}_i^U = (1 + \gamma_U) \max(\hat{\Theta}_i^-, \hat{\Theta}_i^+)$ for some $\gamma_L, \gamma_U > 0$.

Another change that could be made to the general BFGS algorithm is to incorporate limited memory storage of the inverse hessian H_k . This would prove most beneficial when the dimension of the problem is large and we would want to limit the number of operations on large inverse Hessian matrices. Combining these two modifications gives the limited-memory box-constrained BFGS algorithm (L-BFGS-B). Preliminary simulation results using L-BFGS-B applied to (3.14) were conducted with the results summarized in Figure 3.9. It can be seen that the algorithm performs quite well, though not drastically differently than BFGS. However, initial results demonstrated a marked improvement in computation speed; see Table 3.4. L-BFGS-B is seen to be almost twice as fast as BFGS for $d = 3$. For the complete simulations presented in Figures 3.6 and 3.9, the computation time was 4 hours and 1.7 hours, respectively, using 50 cores from an AMD Opteron Processor 6380 with 2.5 GHz clockspeed. However, these simulations were conducted on a shared computation server and the results are likely to be noisier. This time difference is likely to increase with d . While the numerical results are encouraging, theoretical backing would be required to further support this methodology. Nonetheless, we believe further research on incorporating L-BFGS-B is merited as its properties are better suited for the optimization of (3.13) and, more importantly, (3.14). For more information on L-BFGS or L-BFGS-B, see, e.g., Nocedal and Wright (1999), Liu and Nocedal (1989) and Byrd et al. (1995).

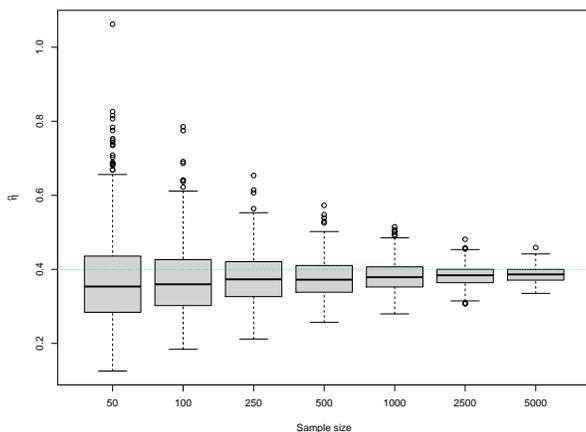
Table 3.4: Microbenchmarking comparison for BFGS vs L-BFGS-B. Estimating $e^\perp(\mathbf{X})$ with $d = 3$ and $n = 250$ and 100 replications. Margins Pareto distributed, $X_i \sim P(3.5, 1.25(1 + i))$. q_α is the α -level quantile of the distribution of run-times. Entries are presented in seconds. Microbenchmarking was completed using 4 cores from a Intel Core i7-5500U CPU with 2.40 GHz clock speed.

	Min	$q_{0.25}$	Mean	$q_{0.50}$	$q_{0.75}$	Max
BFGS	6.621	13.764	19.401	17.237	20.606	80.647
L-BFGS-B	4.711	8.456	11.728	10.589	13.383	43.874

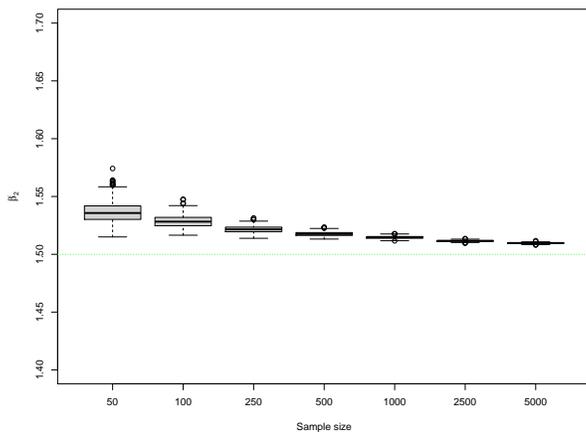
3.5 Conclusion

In this chapter, we provided a review of expectiles in one or more dimensions, including discussion on how to consistently estimate L^1 -expectiles. In particular, we introduce an effective semi-parametric method for estimating multivariate expectiles for extreme risk levels $\alpha \approx 1$. Using the empirical beta copula, taking advantage of the fact that it is a genuine copula, we are able to establish the consistency of the approximate optimization problem in (3.14) under the BFGS algorithm when using empirical estimators for the tail index, tail ratio and udf. We have also proposed a new consistent estimator for the integral of the udf. Numerical simulations demonstrate the satisfactory finite-sample performance of the algorithm when compared to the true multivariate extreme expectiles stemming from comonotonic and independent dependence structures, where closed form

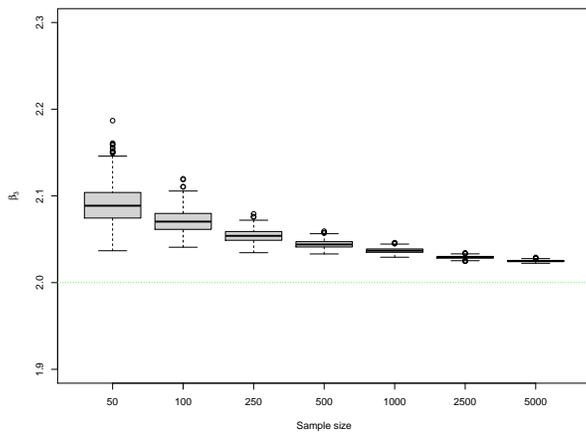
3.5 Conclusion



(a) Results for $\hat{\eta}$



(b) Results for $\hat{\beta}_2$



(c) Results for $\hat{\beta}_3$

Figure 3.9: Boxplots for the three-dimensional Pareto simulations using L-BFGS-B. Margins are comonotonic and remaining settings are equivalent to Figure 3.6.

3.5 Conclusion

solutions exist. Furthermore, the set-up provided by the system of equations in Proposition 3.2 allows us to mitigate the curse of dimensionality as we are only ever required to consider pairwise dependencies. As such, the number of operations required for each iteration of the optimization algorithm will be $\binom{d}{2}$.

In the future, it would be interesting to compare the performance of the algorithm for a variety of estimators; for instance, if one were to consider the checkerboard copula instead of the empirical beta copula. Finally, while finite-sample results are encouraging, the estimation procedure could be further streamlined by considering an algorithm which could take advantage of additional properties of our optimization problem. For example, L-BFGS-B could improve estimation by bounding solutions to (3.14) by the MEEs under the limiting cases of upper tail dependence.

Furthermore the present work could open interesting perspectives in functional statistics field. Indeed, the univariate well known conditional expectile/quantile model (see, *e.g.*, Daouia et al. (2013), Usseglio-Carleve (2018), Girard et al. (2019)) can be adapted to our semi-parametric estimation scheme for new conditional multidimensional L^1 -expectile from heavy-tailed distributions when functional covariate information is available. The interested reader is also referred for instance to Gardes et al. (2010), Gardes and Stupfler (2019) where functional nonparametric estimators for conditional extreme quantiles are proposed.

More precisely, let $(X_1, Z_1), \dots, (X_n, Z_n)$ be n independent copies of a random pair $(\mathbf{X}, \mathbf{Z}) \in \mathbb{R}^d \times \mathcal{E}$ where (\mathcal{E}, d) is a not necessarily finite-dimensional Polish space endowed with a semi-metric d . For instance, \mathcal{E} can be the standard p -dimensional space \mathbb{R}^p , a space of continuous functions over a compact metric space, or a Lebesgue space $L_p(\mathbb{R})$, to name a few. From Equation (3.6), a future development of this work would be to consider the functional conditional multidimensional L^1 -expectile extension given by

$$e_{\alpha}^1(\mathbf{X}, z) := \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left\{ \alpha \left(\sum_{i=1}^d |X_i - x_{i+}| \right)^2 + (1 - \alpha) \left(\sum_{i=1}^d |X_i - x_{i-}| \right)^2 \middle| \mathbf{Z} = z \right\}.$$

With this, the presented work could be adapted to estimate extreme $e_{\alpha}^1(\mathbf{X}, z)$ by using the extrapolation technique when $\alpha \rightarrow 1$ together with a nonparametric estimation of the conditional tail copula (see, *e.g.*, Gardes and Girard (2015)). The main objective of this semi-parametric approach for extreme functional $e_{\alpha}^1(\mathbf{X}, z)$ would be to balance the trade-off between the high sensitivity to dimension of the nonparametric models and the relative lack of flexibility of the completely parametric models (see, *e.g.*, Goia and Vieu (2016)).

Appendix

3.A Proof of Proposition 3.6

Proof.

$$\begin{aligned}
& \int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}_n^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt \\
&= \int_{\frac{\beta_i}{\beta_k}}^{\infty} \frac{1}{\ell_\lambda} \sum_{j=1}^n \mathbb{1} \left\{ 1 - F_{n,i}(X_{ij}) < \frac{\ell_\lambda c_i t^{-\theta}}{n}, 1 - F_{n,k}(X_{kj}) < \frac{\ell_\lambda}{n} \right\} dt \\
&= \int_{\frac{\beta_i}{\beta_k}}^{\infty} \frac{1}{\ell_\lambda} \sum_{j=1}^n \mathbb{1} \left\{ \frac{n - R_{ij}^{(n)} + 1}{n} < \frac{\ell_\lambda c_i t^{-\theta}}{n} \right\} \mathbb{1}(X_{kj} \geq X_{k:n-\ell_\lambda, n}) dt \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \int_{\frac{\beta_i}{\beta_k}}^{\infty} \mathbb{1} \left\{ \left(\frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-1/\theta} > t \right\} \mathbb{1}(X_{kj} \geq X_{k:n-\ell_\lambda, n}) dt \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \int_{\frac{\beta_i}{\beta_k}}^{\left(\frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-1/\theta}} \mathbb{1} \left\{ \left(\frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-1/\theta} \geq \frac{\beta_i}{\beta_k} \right\} \mathbb{1}(X_{kj} \geq X_{k:n-\ell_\lambda, n}) dt \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \left[\left\{ \frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right\}^{-1/\theta} - \frac{\beta_i}{\beta_k} \right] \mathbb{1} \left\{ \left(\frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-1/\theta} \geq \frac{\beta_i}{\beta_k} \right\} \\
&\quad \times \mathbb{1}(X_{kj} \geq X_{k:n-\ell_\lambda, n}) \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \left[\left\{ \frac{n - R_{ij}^{(n)} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right\}^{-1/\theta} - \frac{\beta_i}{\beta_k} \right]_+ \mathbb{1}(X_{kj} \geq X_{k:n-\ell_\lambda, n}).
\end{aligned}$$

□

4

Environmental Extremes

Depending on the local climate and geography, populations can be exposed to a litany of natural disasters. For example: tornados, droughts and floods all pose varying risks to different communities. The impacts of these events can be catastrophic to the local economy and the well being of individuals. In this chapter, we seek to build upon our tools for risk analysis by quantifying the risks associated to extreme events in the environment. The modeling of environmental events is inherently spatial in nature as data provided for this purpose are often drawn from locations where their geographic properties are key. To this end, the discussions that follow are predicated on the concept of *spatial processes*. Consider $s \in \mathbb{R}^d$ a location in d -dimensional Euclidean space and denote $Z(s)$ a rv at location s . For a subdomain $\mathcal{D} \subset \mathbb{R}^d$, denote

$$\mathcal{Z} = \{Z(s) : s \in \mathcal{D}\}$$

the random process on \mathcal{D} . More specifically, \mathcal{Z} is called a spatial process or field. For all $s \in \mathcal{D}$, $Z(s)$ provides some information about random occurrences at location $s \in \mathcal{D}$.

In Section 4.1, we introduce how spatial processes are used to study spatial phenomena in the environment through geostatistics. Section 4.2 focuses on the intersection between geostatistics and extreme value theory, in particular focusing on the difference between large- and small-scale variation. In Section 4.3, we provide an application of these techniques for quantifying the risk of extreme surges in Atlantic Canada. Finally, concluding remarks are provided in Section 4.4.

4.1 Geostatistics

The study of spatial processes defined by their geographical locations is often referred to as *geostatistics*. For an in-depth discussion of geostatistics, see [Cressie \(1993\)](#), [Diggle and Ribeiro \(2007\)](#), or [Banerjee et al. \(2014\)](#). Geostatistics can also include the study of phenomena in time, as well as space-time, though the following discussion will be predominantly limited to processes in space.

4.1 Geostatistics

As one is often interested in using realizations of spatial processes for inference, certain operating assumptions make this task more manageable. In particular, we introduce the notion of *stationarity*. There are two main types of stationary: Second-order (weak) stationary and strict (strong) stationarity.

Definition 4.1. A random process Z is said to be second-order or weakly stationary if the following hold

(i) For all $\mathbf{s} \in \mathcal{D}$,

$$\mathbb{E}\{Z(\mathbf{s})\} = \mu.$$

(ii) For all $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{D}$,

$$\text{cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = C(\mathbf{s}_1 - \mathbf{s}_2),$$

where C is called a stationary covariance function or a covariogram.

Moreover, if $C(\mathbf{s}_1 - \mathbf{s}_2)$ is only a function of $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$, then we say that C is isotropic.

Definition 4.2. A random process Z is said to be strict-sense or strongly stationary if for any finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, the joint distribution of $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ satisfies

$$F_{\mathbf{s}_1+h, \dots, \mathbf{s}_n+h}(z_1, \dots, z_n) = F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(z_1, \dots, z_n)$$

for all $m \geq 1$ and $h \in \mathbb{R}^d$.

When $Z(\mathbf{s})$ has finite second moment, strong stationarity implies second-order stationarity. The property of isotropy implies that the variability between spatial locations is solely determined by the distance between them and not the direction. The spatial processes discussed in this chapter often require stationarity and isotropy. In particular, one often considers processes of the form

$$Y(\mathbf{s}) = \alpha(\mathbf{s}) + Z(\mathbf{s}),$$

where $\alpha(\mathbf{s})$ is a deterministic function determining the mean behavior and $Z(\mathbf{s})$ is a second-order stationary, isotropic and zero-mean stochastic process with some spatially driven covariance structure. In particular, the mean behavior is often formulated as a regression relation, i.e., one may write $\alpha(\mathbf{s}) = \mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta}$, where $\mathbf{X}(\mathbf{s})$ is a column vector of spatially defined covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients.

One of the principal goals in many geostatistics applications is to provide inference beyond the original scope of the data provided. For example, providing inference on precipitation outcomes at an unmonitored area using observations available at a neighboring one. This is referred to as spatial interpolation or *kriging*. The simplest type of kriging is called *ordinary kriging* and consists

4.1 Geostatistics

of making predictions at unmonitored locations \mathbf{s}_0 using weighted averages of observed values, i.e.,

$$\hat{z}(\mathbf{s}_0) = \sum_{i=1}^n w(\mathbf{s}_i) z(\mathbf{s}_i),$$

where $z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)$ are the observed values of the spatial process Z at locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ and $w(\mathbf{s}_1), \dots, w(\mathbf{s}_n)$ are the kriging weights which incorporate some behavior about the spatial process as a whole. Thus, kriging requires one to have an understanding of how to quantify the dependence/variability of rv's in space.

As discussed in Section 2.2, the dependence between rv's can be quantified in several ways. This information, which describes the way rv's behave relative to one another, can greatly impact our perception of the risks associated to them. In geostatistics, there are three quantities often used to model spatial variability/dependence: the *variogram*, the *covariogram* and the *correlogram*. For two locations, $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{D}$, the variogram and covariogram are defined by

$$2\gamma(\mathbf{s}_1, \mathbf{s}_2) = \text{var}\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\} \quad (4.1)$$

and

$$\begin{aligned} C(\mathbf{s}_1, \mathbf{s}_2) &= \text{cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} \\ &= \mathbb{E}\{Z(\mathbf{s}_1)Z(\mathbf{s}_2)\} - \mathbb{E}\{Z(\mathbf{s}_1)\}\mathbb{E}\{Z(\mathbf{s}_2)\}, \end{aligned} \quad (4.2)$$

respectively. The quantity $\gamma(\mathbf{s}_1, \mathbf{s}_2)$ is referred to as the *semivariogram*. If the spatial process Z is stationary and isotropic, then (4.1) and (4.2) can be simplified to

$$2\gamma(\mathbf{h}) = \mathbb{E} [\{Z(\mathbf{s}_1) - Z(\mathbf{s}_1 + \mathbf{h})\}^2] \quad \text{and} \quad C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{h}),$$

respectively, where $\mathbf{s}_1 - \mathbf{s}_2 = \mathbf{h}$. In this case, the correlogram can be defined as

$$\rho(\mathbf{s}_1, \mathbf{s}_2) = \text{corr}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = C(\mathbf{h})/C(\mathbf{0}).$$

Second order stationarity is often assumed for many spatial processes as most kriging requires that the semivariogram, γ , be a function of only the difference \mathbf{h} . In the case of Gaussian processes, often used in geostatistics, second-order and strict stationarity coincide. Isotropy is often assumed as it simplifies the estimation of the semivariogram. Under both weak stationarity and isotropy, the traditional empirical variogram is defined as

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|\mathcal{N}(\mathbf{h})|} \sum_{(\mathbf{s}, \mathbf{s}') \in \mathcal{N}(\mathbf{h})} \{Z(\mathbf{s}) - Z(\mathbf{s}')\}^2,$$

4.2 Spatial extremes

where $\mathcal{N}(\mathbf{h}) = \{(\mathbf{s}, \mathbf{s}') : \mathbf{s} - \mathbf{s}' = \mathbf{h}\}$ and $|A|$ is the number of distinct elements in a set A . Often, the empirical variogram is used to inform variogram model selection. Several families of variogram models exist. One of the most well known families of variograms is defined by the Matérn variogram model (Matérn, 1960; Stein, 1999), viz.

$$\gamma(\mathbf{h}) = c_0 + (c_1 - c_0) \left\{ 1 - \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{\mathbf{h}}{\alpha}\right)^{\kappa} K_{\kappa} \left(\frac{\mathbf{h}}{\alpha}\right) \right\},$$

where K_{κ} is the modified Bessel function of the second kind (Bowman, 1958) and

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx$$

is the gamma function. The parameter ν controls the smoothness of γ . Special cases of γ include the exponential and Gaussian variograms for $\nu = 0.5$ and $\nu \rightarrow \infty$, respectively.

In general, while second order-stationarity is often assumed for kriging purposes, spatial interpolation is possible under a weaker set of assumptions, namely *intrinsic* stationarity. Intrinsic stationarity requires only that the increments $Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})$ be second order-stationary. In this case, the semivariogram is still defined in terms of \mathbf{h} . Note however that intrinsically stationary processes do not have covariance functions.

4.2 Spatial extremes

Use of geostatistics in the extreme value framework poses some challenges. The first owes itself to the differing interests between classical geostatistics and EVT. Geostatistics often centers around the study of distributional means whereas EVT focuses on the tail of a distribution. The Gaussian framework often employed in geostatistics is thus unsuitable for extreme value analysis and the interest in max-stable random variables, as defined in Section 2.1.

In this section we discuss ways in which geostatistical methods and EVT have been combined for the purpose of studying spatially driven extremes and, in particular, different ways to model spatial behavior.

4.2.1 Hierarchical models

Hierarchical models are often used to bridge the gap between geostatistics and EVT by introducing a modeling hierarchy which incorporates geostatistical concepts into extreme value models. Often, these models consider the spatial variability in the data with some type of latent process, often assumed Gaussian. For example, Diggle et al. (1998) use latent Gaussian fields to model the intensity of Poisson distributed random data. They apply the same latent Gaussian structure to binomial data.

4.2 Spatial extremes

These hierarchies are constructed using a series of conditional statements. For instance, given a vector of parameters $\Theta = (\theta_1, \theta_2, \theta_3)$, a simple 3-layer hierarchy can be written

$$\mathbf{Z}|\theta_1 \sim p(\cdot; \theta_1), \quad \theta_1|\theta_2 \sim p(\cdot; \theta_2), \quad \text{and} \quad \theta_2 \sim p(\cdot; \theta_3),$$

where the parameters θ_3 are fixed. Here $\mathbf{Z}|\theta_1$ is called the data layer and $p(\cdot; \theta_1)$ defines the likelihood function. The process layer $\theta_1|\theta_2$ incorporates spatiality using concepts from geostatistics. Specifically, the unobserved parameters are modelled using the relationship

$$\theta_1 = \mathbf{X}^\top \boldsymbol{\beta} + \epsilon,$$

where \mathbf{X} is a collection of spatially driven covariates, $\boldsymbol{\beta} \in \theta_2$ are regression coefficients and $\epsilon \sim \mathcal{N}(0, \Sigma)$ is a stationary, often isotropic, Gaussian field with covariance structure defined by the distance between the locations of the observed data. The inclusion of this latent field fosters information sharing between stations, strengthening model fit, and allows for spatial interpolation to be conducted. Finally, θ_2 is called the prior layer of the model.

This methodology lends itself perfectly to the Bayesian framework of modeling. The Bayesian paradigm is particularly powerful in the context of environmental extremes. For one, it allows prior information one might have about the area of study to be incorporated into the fitting. Given that environmental phenomena are physical processes, expert knowledge could prove invaluable. Secondly, as the parameters themselves are treated as random variables, model fitting using Bayesian methods can incorporate the model error in a more tenable manner, allowing it to propagate naturally throughout the estimation process. As such, post-fitting analyses will have a more coherent interpretation.

Conditional on the observed realizations \mathbf{z} of \mathbf{Z} , the distribution of Θ can be written, using Bayes' rule,

$$p(\Theta|\mathbf{z}) = \frac{p(\mathbf{z}|\theta_1)p(\theta_1|\theta_2)p(\theta_3)}{\int_{\vartheta} p(\mathbf{z}|\Theta)dP(\Theta)}, \quad (4.3)$$

where ϑ is the parameter space of Θ . Examples of hierarchical modeling for environmental extremes include wind speeds (Casson and Coles, 1999; Fawcett and Walshaw, 2006), precipitation (Cooley et al., 2007; Sang and Gelfand, 2009; Jalbert et al., 2017; García et al., 2018), wildfires (Mendes et al., 2010) and streamflows (Yan and Moradkhani, 2015; Lima et al., 2016).

Markov chain Monte Carlo

HBM's in the form of (4.3) often have complicated structures which make closed form representations of parameter distributions difficult to obtain, and thus difficult to sample from. To handle this, Markov Chain Monte Carlo (MCMC) techniques are used to produce samples from these unknown

4.2 Spatial extremes

distributions. MCMC techniques function by producing a Markov chain which spans the support of the parameter space ϑ . Requiring the full posterior distribution of the parameters is avoided by considering the relationship

$$p(\Theta|\mathbf{z}) \propto p(\mathbf{z}|\boldsymbol{\theta}_1, \xi)p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_3), \quad (4.4)$$

given that the denominator in (4.3) does not depend on Θ .

In many of the models referenced in this thesis, including our own application in Section 4.3, Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) and the Metropolis–Hastings algorithm (Hastings, 1970; Chib and Greenberg, 1995) are chosen for sampling purposes. The combination of these techniques is often referred to as a *Metropolis–Hastings-within-Gibbs* algorithm. For a collection of parameters $\Theta = (\theta_1, \dots, \theta_p)$, Algorithm 3 describes how to generate N realizations from the posterior distribution using the Metropolis–Hastings-within-Gibbs algorithm. The idea here is to conduct a simple random walk where at each step one generates a potential realization from the posterior distribution. Proposals are generated using a *candidate* distribution, often assumed to be normal, i.e., $\theta_j^{cand} \sim \mathcal{N}(\theta_j^{(i)}, \sigma_j^2)$ for some $\sigma_j^2 > 0$. When the candidate distribution

3 Metropolis-within-Gibbs Sampling Algorithm

Initialize $\Theta = \Theta^{(0)}$.

for $i \in \{1, \dots, N\}$

for $j \in \{1, \dots, p\}$

 (1) Denote $\boldsymbol{\theta}_j^{*,(i)} = (\theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_j^{(i-1)}, \theta_{j+1}^{(i-1)}, \dots, \theta_p^{(i-1)})$

 (2) Sample a possible realization from the posterior distribution, θ_j^{cand} , from a candidate distribution $q(\theta_j^{(i)}|\boldsymbol{\theta}_j^{*,(i)})$.

 (3) Calculate acceptance probability

$$\alpha\left(\theta_j^{cand}|\theta_j^{(i-1)}\right) = \min\left\{1, \frac{q(\theta_j^{(i-1)}|\theta_j^{cand})\pi(\theta_j^{cand}|\mathbf{z}, \boldsymbol{\theta}_j^{*,(i)})}{q(\theta_j^{cand}|\theta_j^{(i-1)})\pi(\theta_j^{(i-1)}|\mathbf{z}, \boldsymbol{\theta}_j^{*,(i)})}\right\}.$$

 Generate $U_{i,j} \sim \mathcal{U}(0, 1)$

if $U_{i,j} < \alpha\left(\theta_j^{cand}|\theta_j^{(i-1)}\right)$ **then**

 Accept the candidate: $\theta_j^{(i)} \leftarrow \theta_j^{cand}$

else

 Reject the candidate: $\theta_j^{(i)} \leftarrow \theta_j^{(i-1)}$

end if

end for

end for

4.2 Spatial extremes

is symmetric, as is the case with the normal distribution, the acceptance probability simplifies to

$$\alpha\left(\theta_j^{cand}|\theta_j^{(i-1)}\right) = \min\left\{1, \frac{\pi(\theta_j^{cand}|\mathbf{z}, \boldsymbol{\theta}_j^{*,(i)})}{\pi(\theta_j^{(i-1)}|\mathbf{z}, \boldsymbol{\theta}_j^{*,(i)})}\right\}.$$

For HBM's, Gibbs sampling allows for sequential sampling of the posterior parameter distribution using simplifications provided by the conditional relationships described in (4.4) while the Metropolis–Hastings algorithm allows one to accept or reject a candidate value as a feasible realization from the true posterior based on the realized improvement of the likelihood of the predictive posterior density, $\pi(\cdot|\mathbf{z}, \boldsymbol{\theta}_j^{*,(i)})$. Diagnostics for MCMC techniques will be postponed until Section 4.3.

4.2.2 Residual dependence

In environmental applications, there are two types of spatial behavior which are of interest: regional and local. The HBM's referenced in Section 4.2.1, focus on modeling the regional behavior of a spatial process. Moreover, these models forego the modeling of local variation. After accounting for spatial dependence through some latent process, the distributions of the observed data are assumed conditionally independent given the parameter or vector of parameters $\boldsymbol{\theta}_1$. If the goal of the analysis is to produce return levels, this assumption is not so impactful (Davison et al., 2012; Ribatet et al., 2016). However, as discussed in Tawn et al. (2018) and Towe et al. (2018), the consideration of events at a site-by-site basis, independent of the surrounding area, is overly simplistic when impact across the entire domain is of interest. This is especially true for environmental events for which large clusters of locations can be impacted by a single event, e.g., a tornado ravaging the coast of a country or a river overflowing and flooding multiple communities. The dependence which helps describes the impact of these localized events on a collection of locations is referred to as *residual dependence*. Below we present a few techniques for modeling residual dependence, after marginal effects have been accounted for.

Max-stable processes

The first way in which the residual dependence between spatially defined rv's can be dealt with is through *max-stable processes*. Building on the concepts of Sections 2.1 and 2.2, max-stable processes extend the notion of max-stability to the infinite-dimensional setting. A construction of max-stable processes was first provided by de Haan (1984) using a spectral representation. See also de Haan and Ferreira (2006). This process is described by two components: a stochastic process $\mathcal{X} = \{X(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ and a Poisson process Π with intensity $d\lambda/\lambda^2$ on $(0, \infty)$. Let $\{X_i(\mathbf{s})\}_{i \in \mathbb{N}}$ be independent realizations of a process $\mathbf{X}(\mathbf{s})$ with $\mathbb{E}\{X(\mathbf{s})\} = 1$ and let $\zeta_i \in \Pi$ be the realizations of

4.2 Spatial extremes

a Poisson process. Then $\mathcal{Z} = \{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, where

$$Z(\mathbf{s}) = \max_{i \geq 1} \zeta_i X_i(\mathbf{s})$$

is a max-stable process with unit Fréchet margins and a distribution function

$$\Pr\{Z(\mathbf{s}) \leq z(\mathbf{s}), \mathbf{s} \in \mathcal{D}\} = \exp\left(-\mathbb{E}\left[\sup_{\mathbf{s} \in \mathcal{D}} \left\{\frac{X(\mathbf{s})}{z(\mathbf{s})}\right\}\right]\right),$$

where $-\mathbb{E}[\sup_{\mathbf{s} \in \mathcal{D}} \{X(\mathbf{s})/z(\mathbf{s})\}]$ is the infinite-dimensional analogue to μ , introduced in Section 2.2. The choice of the process \mathcal{X} will determine the resulting max-stable process \mathcal{Z} .

For instance, [Smith \(1990\)](#) demonstrates how Gaussian processes can be incorporated into the construction of [de Haan \(1984\)](#). If one were to set $X_i(\mathbf{s}) = f(\mathbf{s} - \mathbf{p}_i)$, where \mathbf{p}_i is a Poisson Process with unit rate on \mathcal{D} and f is a multivariate normal density with covariance matrix Σ , the resulting max-stable process is called the Gaussian extreme value process. The bivariate distribution of this process can be written in a closed form ([Hüsler and Reiss, 1989](#)) for $z_1, z_2 > 0$ as

$$\Pr\{Z(\mathbf{s}_1) \leq z_1, Z(\mathbf{s}_2) \leq z_2\} = \exp\left[-\frac{1}{z_1} \Phi\left\{\frac{a}{2} + \frac{\log(z_2/z_1)}{a}\right\} - \frac{1}{z_2} \Phi\left\{\frac{a}{2} + \frac{\log(z_1/z_2)}{a}\right\}\right], \quad (4.5)$$

where $a = \sqrt{(\mathbf{s}_1 - \mathbf{s}_2)^\top \Sigma^{-1} (\mathbf{s}_1 - \mathbf{s}_2)}$ is the dependence parameter and Φ is the standard normal cdf. [de Haan and Pereira \(2006\)](#) define analogous processes when f is a t or Laplace density.

Conversely, [Kablichko et al. \(2009\)](#) suggest setting $X_i(\mathbf{s}) = \exp\{e_i(\mathbf{s}) - \sigma^2(\mathbf{s})/2\}$, where $e_i(\mathbf{s})$ is an intrinsically stationary Gaussian process and $\sigma^2(\mathbf{s}) = \text{var}\{e(\mathbf{s})\}$. This formulation gives the Brown–Resnick process ([Brown and Resnick, 1977](#)). Brown–Resnick processes are a particularly useful class of max-stable processes in geostatistics as their dependence structure is determined by the variogram. The bivariate cdf of Brown–Resnick processes is similar to (4.5), replacing the dependence parameter a with the square root of the semivariogram, i.e.,

$$\Pr\{Z(\mathbf{s}_1) \leq z_1, Z(\mathbf{s}_2) \leq z_2\} = \exp\left[-\frac{1}{z_1} \Phi\left\{\frac{\sqrt{\gamma(h)}}{2} + \frac{\log(z_2/z_1)}{\sqrt{\gamma(h)}}\right\} - \frac{1}{z_2} \Phi\left\{\frac{\sqrt{\gamma(h)}}{2} + \frac{\log(z_1/z_2)}{\sqrt{\gamma(h)}}\right\}\right],$$

where h is the Euclidean distance between \mathbf{s}_1 and \mathbf{s}_2 .

Applications of max-stable processes for environmental extremes include [Reich and Shaby \(2012\)](#), [Thibaud et al. \(2016\)](#), [Oesting et al. \(2017\)](#), [Shin et al. \(2019\)](#), and [Albrecher et al. \(2020\)](#).

4.2 Spatial extremes

Estimation of max-stable processes has been discussed using a variety of methodologies, including nonparametric estimators (Vettori et al., 2018), M-estimators (Einmahl et al., 2016), and generalized least squares estimators (Buhl and Klüppelberg, 2019). While these methods perform well in lower dimensional settings, likelihood-based inference is often preferred in higher dimensions for its large-sample properties. Unfortunately, full likelihood methods suffer from increased computational burden as the dimension of the problem increases (Castruccio et al., 2016; Huser et al., 2019). Conversely, while composite likelihood methods have been suggested to circumvent this issue (Padoan et al., 2010; Huser and Davison, 2013; Sang and Genton, 2014), they do so with a loss of efficiency (Huser et al., 2016). Simulation from max-stable processes also proves difficult depending on the properties of the spatial domain (Oesting et al., 2012, 2018).

Copulas

As discussed in Section 2.2, copulas represent a convenient way of modeling the dependence between marginal random variables. Given that we are interested in rare events, an EV copula would make a natural choice for quantifying these large scale dependencies (Hsing, 1989). Unfortunately, due to their construction, such copulas can be unwieldy to work with in increased dimension (Davison et al., 2012; Davison and Huser, 2015). While methods have been described to mitigate such issues (Vettori et al., 2020), given that problems in the context of spatial extremes often consider a multitude of locations, this makes any practical use of extreme value copulas infeasible at present. Even if the beginning number of observation locations is small, the interest in interpolating to unmonitored locations provides a further set of obstacles.

In the papers of Sang and Gelfand (2010) and Bracken et al. (2016), a Gaussian copula is introduced to the HBM framework so as to model the residual dependence in precipitation data. A Gaussian copula is seen as a viable choice because of its simplicity; its parametrization using a pairwise correlation matrix makes incorporating spatial information straightforward. Moreover, being elliptical, Gaussian copulas are closed under the margins (Genest et al., 2007). This provides an efficient way of dealing with missing data. When evaluating the likelihood of the Gaussian copula for a given year, one can simply integrate out the locations for which there is no annual maximum and still be left with a Gaussian copula of lower dimension with the corresponding dependence structure. Specifically, consider a stationary spatial process $\mathbf{Z} = (Z_1, \dots, Z_d)$ for locations $\mathbf{s}_1, \dots, \mathbf{s}_d$, whose underlying copula C_{Ξ} is elliptical with correlation matrix Ξ . For a given year t , consider the realization $\mathbf{z}(t) = (z_1(t), \dots, z_d(t))$ and define the set $\mathcal{I} = \{i_1, \dots, i_{d'}\}$ of indices of size $d' \leq d$ for which there is an annual maximum recorded at locations $\mathbf{s}_{i_1}, \dots, \mathbf{s}_{i_{d'}}$. The underlying copula $C_{\Xi'}$ of the vector $\mathbf{Z}' = (Z_{i_1}, \dots, Z_{i_{d'}})$ is then of lower dimension than, but from the same elliptical family as, C_{Ξ} ; in fact, Ξ' is deduced from Ξ by extracting the appropriate $d' \times d'$ submatrix.

4.3 Flood risk in Atlantic Canada

However, a key feature of extreme value copulas is tail dependence, i.e., the propensity of extreme events to occur simultaneously. Gaussian copulas are asymptotically independent, and thus assume all extreme events occur independently of each other. Student's t copulas represent an interesting compromise in this arena. Similar to Gaussian copulas, t copulas are elliptical. They can easily incorporate spatial information into their covariance matrices and are closed under the margins. Additionally, while not extreme value, they still capture tail dependence. Specifically, it was shown by [Demarta and McNeil \(2005\)](#) that the upper tail dependence index of a random pair (Z_i, Z_j) whose underlying copula is Student's t with ν degrees of freedom is given by $2t_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{1-\rho_{ij}}/\sqrt{1+\rho_{ij}}\right)$, where ρ_{ij} is the Spearman's correlation between Z_i and Z_j while $t_{\nu+1}$ refers to the cumulative distribution function of the standard Student's t distribution with $\nu+1$ degrees of freedom. The level of tail dependence decreases as ν increases, and the limiting case $\nu \rightarrow \infty$ corresponds to the Gaussian copula, which exhibits no tail dependence.

For more information on the extreme value properties of the Student's t copula, see [Nikoloulopoulos et al. \(2009\)](#). For an example of the t copula being used with a HBM, see [Ghosh and Mallick \(2011\)](#).

4.3 Flood risk in Atlantic Canada

In this section, we use a combination of extreme value theory, spatial statistics, hierarchical Bayesian methods and dependence modeling to quantify the risk of flooding in Atlantic Canada.

Flooding represents one of the biggest and most impactful risks globally. The impact of floods is devastating, both in terms of the number of victims and damages. According to [Winsemius et al. \(2016\)](#), between 1980 and 2013, economic losses due to floods exceeded \$1 trillion (2013 USD) with more than 220,000 lives lost. In 2018 alone, based on a study conducted by Munich Re, 46% of the world's disasters were categorized as flooding events ([L w, 2019](#)).

Clearly, accurately predicting and mitigating the effects of overland floods is important, particularly in light of climate change. With the observed increase in the risk of flooding, one would expect an equal increase in protective measures. To an extent, this has been the case in the United States and generally in Europe, where flooding has been met with developments at both the governmental level and in the private (e.g., insurance) sector over the past several decades. From a statistical modeling standpoint, however, relatively little has been done to quantify the risk of overland flooding on a large and sparsely monitored domain.

Motivated by the need for insurers to disclose their risk exposure, the analysis described here contributes to the development of tools for assessing the risk and economic impact of overland flooding in Canada. It was not until the Alberta floods of 2013 that Canadian insurers began to develop products for protection against overland flooding. In the wake of this disaster, only a third of the estimated \$6 billion in damages were covered by insurance, none of which was paid out for

4.3 Flood risk in Atlantic Canada

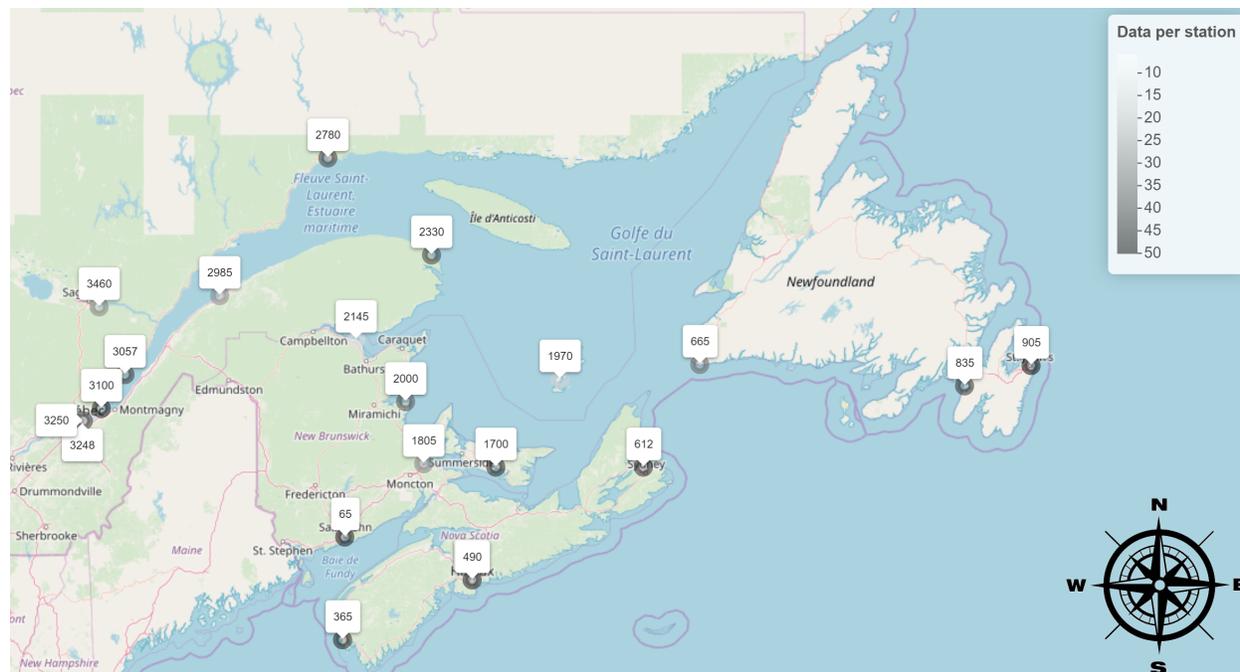


Figure 4.1: Map of Atlantic Canada with labels identifying the location of the 21 monitoring stations used in this study and listed in Table A.

overland flooding (Teufel et al., 2017). The first overland flood product in Canada was launched in 2015 by Aviva but was initially available only in Alberta and Ontario. In 2017, Desjardins became the first Canadian insurer to pay out for overland flooding. These floods, caused by a confluence of melting snow and bad weather in April and May, cost insurers over \$220 million for events occurring in Québec and Ontario (Teufel et al., 2019).

4.3.1 The data

Our application focuses on the Atlantic Coast of Canada, including the Gulf and Estuary of the St. Lawrence River. For brevity, this domain will be referred to as “Atlantic Canada”, including Québec, even though this expression is generally reserved for the region of Canada comprising the four provinces located on the Atlantic coast.

To quantify the risk of coastal flooding in this vast region, shown in Figure 4.1, we rely on water level observations. After some preliminary processing, described in Sections 4.3.2 and 4.3.3, these data will be used to fit a spatial HBM with GEV margins to the area. The water level data we used were freely available from Fisheries and Oceans Canada (<http://www.tides.gc.ca>). Their measurements are made using buoys placed at over a thousand locations, some of which date back 1848. However, data sparsity poses a challenge as the number of observations spans on average six years per station.

4.3 Flood risk in Atlantic Canada

Given that the initial motivation for this study was in insurance, the consideration of the buoys was driven by coastal population density. In the end, 21 buoys were selected. Information about their locations, including station name, station ID and available annual maxima are provided in Appendix 4.A. The buoys can be identified by their ID numbers in Figure 4.1, where darkness of the circles represents the number of annual maxima available. The selected buoys provide data ranging from 1966 to 2015. However, many of the buoys were not operational for the entirety of this period resulting in a fair amount of missing data throughout the series. On average, 37 years of observations on average were available over these 21 locations.

4.3.2 Tidal modeling

The water levels measured by the buoys can be broadly decomposed into a deterministic part, namely the tide, and a stochastic part called the surge. Large surges occurring in tandem with high tides can cause flooding events. In September 2003, for instance, Hurricane Juan hit Atlantic Canada and caused an estimated \$300 million in damages in the area. In Halifax, there were waves of up to 20 m in height at their maximum, and surges between 1.5 and 2 m occurring near high-tide; see, e.g., Fogarty (2004) for details. The interaction between tide and surge is thus crucial to adequately quantify flood risk on the Atlantic Coast of Canada. Given that the surges represent the stochastic portion of our observations, this is what we model using the HBM detailed in Section 4.3.4.

Before modeling surges, we must first extract them from the water level observations by removing the tidal component. Several models have been developed for tidal prediction. Each of these methods operates by representing the tidal process through a collection of *harmonic constituents*. Harmonic constituents are defined as “one of the elements in a mathematical expression for the tide-producing force and in corresponding formulas for the tide or tidal current. Each constituent represents a periodic change or variation in the relative positions of the Earth, Moon, and Sun.”¹ More information on harmonic constituents can be found on the website of the National Oceanic and Atmospheric Administration (NOAA) (https://tidesandcurrents.noaa.gov/about_harmonic_constituents.html). For our study of Atlantic surges, we consider three tidal models: WebTide, the Tidal Model Driver (TMD) and TideHarmonics.

The first option, WebTide, was developed by Dupont et al. (2002) for Eastern Canada. They incorporated into it an inverse/forward finite-element method model which fits a mesh to the region of interest taking into account a prescribed level of error. Their model considers five major harmonic constituents, M2, N2, S2, K1, and O1, incorporates data from several sources and assimilates them into two groups, one for fitting and one for validation. While the authors claim a 10 cm root mean square error (RMSE), given the rather coarse resolution of the model, it cannot make tidal predictions at our collection of locations.

¹<https://tidesandcurrents.noaa.gov/glossary.html#constituent>

4.3 Flood risk in Atlantic Canada

A second option is the TMD Matlab toolkit. Developed by Oregon State University, TMD incorporates its own harmonic analysis model. Their regional model assimilates multiple sources of data and fits the tide $h(t, x)$, at time t for a specific harmonic constituent x , by

$$h(t, x) = pu(t, x)\text{Re}[h(x)e^{i\{\omega_x(t-t_0)+V_0(t_0)+ph(t,x)\}}],$$

where $V_0(t_0)$ is the astronomical argument for the constituent at time t_0 , $pu(t, x)$ and $ph(t, x)$ are nodal corrections, and Re refers to the real part of the complex argument. See [Egbert et al. \(1994\)](#) or [Egbert and Erofeeva \(2002\)](#) for more information. Using this formulation, the authors fit the harmonic constituents M2, N2, S2, K1, O1, K2, P1 and Q1. Unfortunately, the resolution of the model was again too coarse to provide predictions, except at a few of our buoys.

In the end, we opted for the R package TideHarmonics of [Stephenson \(2016\)](#), which provides estimates of the tidal process based on user-provided input data. To model the tide $T(\mathbf{s}, t)$ at time t for station \mathbf{s} , [Stephenson \(2016\)](#) considers a harmonic analysis on N constituents, viz.

$$T(\mathbf{s}, t) = M(\mathbf{s}, t) + \sum_{n=1}^N A_n \cos[\pi\{\omega_n t - \psi_n(\mathbf{s})\}/180],$$

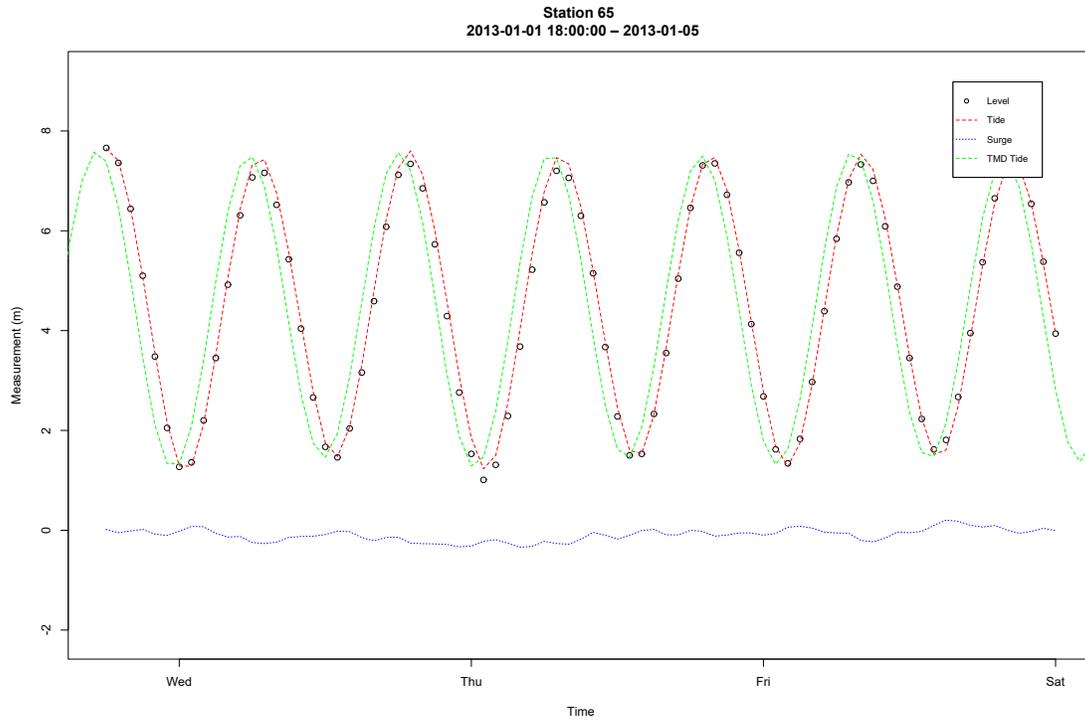
where $M(\mathbf{s}, t)$ is the observed time-varying mean sea-level, ω_n is the known angular frequency of harmonic constituent n , and A_n and $\psi_n(\mathbf{s})$ are the amplitude and phase shift of that constituent respectively. First, $M(\mathbf{s}, t)$ is estimated using the data. Second, the values of A_n and ψ_n are fitted using standard least squares regression. For more details on the model, see the TideHarmonics vignette at <https://cran.r-project.org/web/packages/TideHarmonics/vignettes/austides.pdf>.

To validate our choice, we compared the tidal patterns generated by TideHarmonics at the monitoring location in St. John, NB, where TMD could also be used. As transpires from [Figure 4.2a](#), the TMD model has difficulty representing the observed water levels, even though it incorporates satellite altimetry in addition to gage data. By contrast, the TideHarmonics package does capture the tidal behavior quite effectively without resorting to anything but the raw data from Fisheries and Oceans Canada. In our experience, TideHarmonics is preferable to TMD even when it performs poorly, as illustrated in [Figure 4.2b](#).

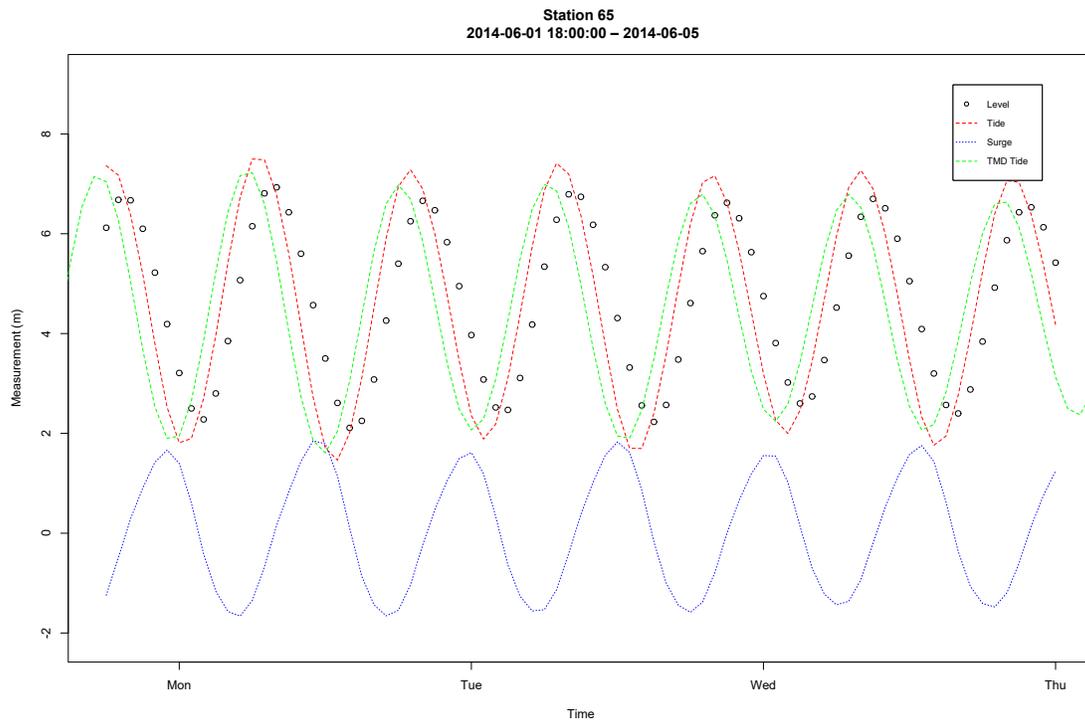
A critical element in using TideHarmonics is the selection of the number N of harmonic constituents. At least 409 such constituents have been identified in the literature and bundled in scientifically meaningful groups. Different organizations throughout the world ran studies to build and validate their own regional prediction models. In particular, the National Oceanic and Atmospheric Administration (NOAA), based in the United States, uses 37 constituents.

For this study, we chose to align with the NOAA variable selection, in part to keep our results comparable with this organization, but also because this choice was supported by the results of

4.3 Flood risk in Atlantic Canada



(a) June 2013



(b) June 2014

Figure 4.2: Tide comparison at Station 65 (St. John, NB). (a) June 2013. (b) June 2014.

4.3 Flood risk in Atlantic Canada

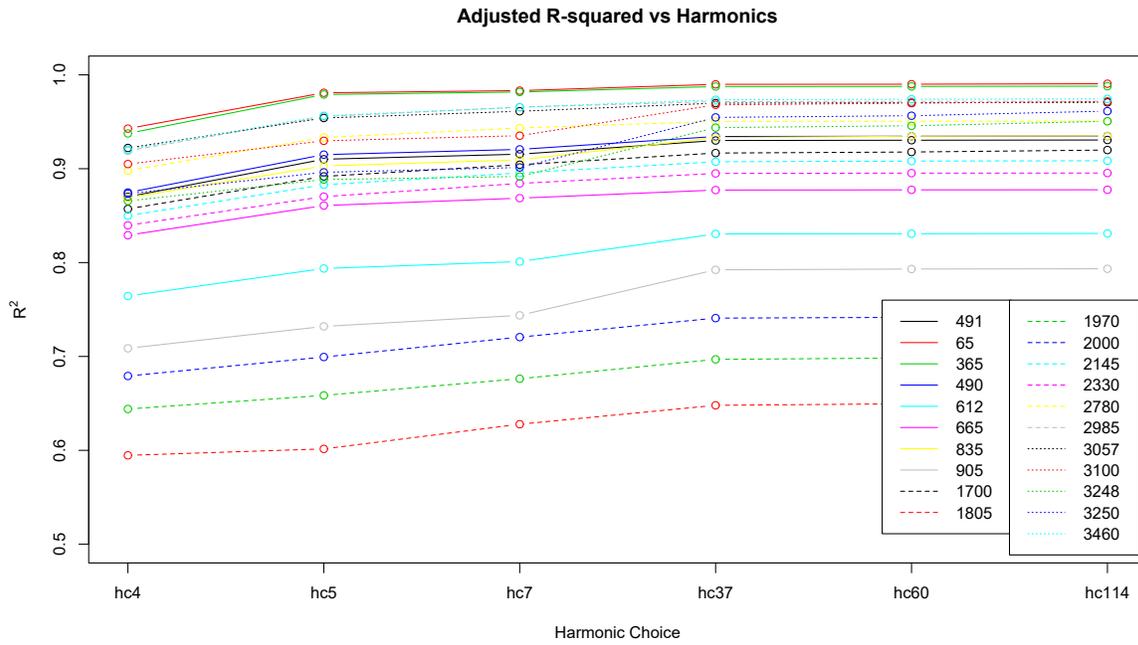
our own diagnostic analysis. Specifically, we compared the evolution of the adjusted coefficient of determination R^2 and the RMSE as the number of constituents increases. Results are shown in Figure 4.3 for all 21 monitored locations. Additionally, we used a k -folds cross validation procedure to test the predictive capabilities of each collection of harmonic constituents. For $k = 8$, the results can be found in Table 4.1. As can be seen, the inclusion of a fifth harmonic, the larger lunar elliptic semidiurnal constituent N2, drastically improves the fit of many of the tidal sequences, in accordance with Dupont et al. (2002). One can also see that while the inclusion of more harmonics monotonically improves both metrics, fitting more than the 37 harmonic constituents used by the NOAA provides minimal improvement. For a detailed description of these constituents, see <https://tidesandcurrents.noaa.gov/glossary.html>.

Examples of the fitted tidal series are presented in Figure 4.4. As a means of validating the results, we see that the tidal fitting process accurately captures historical extreme weather events in two locations. The left panel depicts The Groundhog Day gale of 1976 as experienced in Yarmouth, NS.

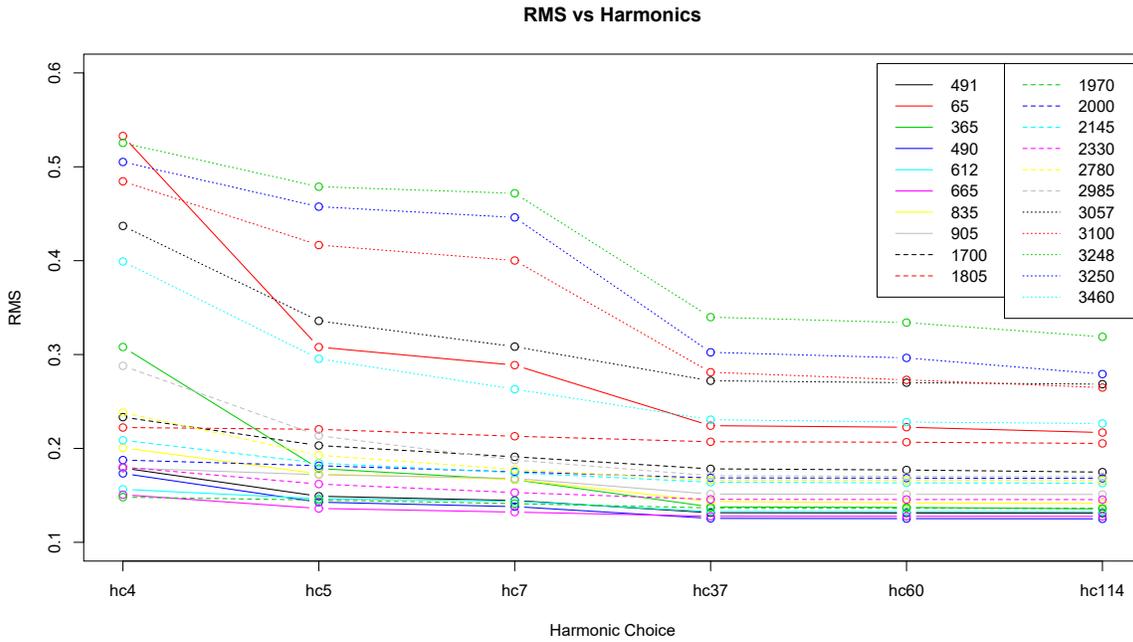
Table 4.1: Average SSE of prediction using k -folds cross validation with $k = 8$ for collections of harmonic constituents hc4, hc5, hc7, hc37, hc60 and hc114. For more information on hc4, hc7, hc37, hc60 and hc114, see the TideHarmonics R package. The collection hc5 is simply hc4 with the larger lunar elliptic semidiurnal constituent N2 included.

Station	491	65	365	490	612	665	835
hc4	303.81	12483.69	4647.07	1509.70	1200.77	925.66	1815.24
hc5	208.52	3434.55	1547.75	1022.33	1053.74	754.65	1360.13
hc7	191.42	2906.01	1351.08	958.14	1015.90	709.14	1267.20
hc37	161.01	1284.39	909.77	798.69	864.15	662.69	942.40
hc60	160.52	1250.32	903.22	795.88	863.05	661.74	922.26
hc114	159.96	1148.87	880.23	792.87	861.71	661.06	916.76
Station	905	1700	1805	1970	2000	2145	2330
hc4	1736.48	2775.94	1315.86	188.75	1470.49	733.31	1473.09
hc5	1602.11	2099.79	1294.78	180.47	1377.01	575.59	1187.50
hc7	1535.58	1857.69	1208.82	170.74	1276.56	511.68	1055.88
hc37	1262.18	1632.04	1152.54	160.68	1180.30	458.75	954.53
hc60	1257.65	1611.61	1147.91	160.12	1176.72	455.54	951.73
hc114	1256.26	1569.05	1133.22	159.81	1172.40	453.74	950.56
Station	2780	2985	3057	3100	3248	3250	3460
hc4	2454.60	2665.86	9162.00	12513.59	2050.50	12426.00	5197.93
hc5	1622.99	1476.28	5473.36	9249.21	1723.14	10210.56	2891.97
hc7	1371.19	1145.99	4642.16	8549.17	1675.00	9745.88	2262.74
hc37	1210.60	960.60	3672.61	4204.64	943.38	4517.30	1750.86
hc60	1206.68	953.70	3621.98	3968.38	943.28	4356.86	1714.12
hc114	1204.93	949.23	3584.36	3727.12	902.78	3877.08	1693.10

4.3 Flood risk in Atlantic Canada



(a) R^2



(b) RMSE

Figure 4.3: Model fit comparison for different groups of harmonic constituents at each of the 21 monitored locations. Location codes are provided in the Appendix. (a) Adjusted coefficient of determination R^2 . (b) RMSE.

4.3 Flood risk in Atlantic Canada

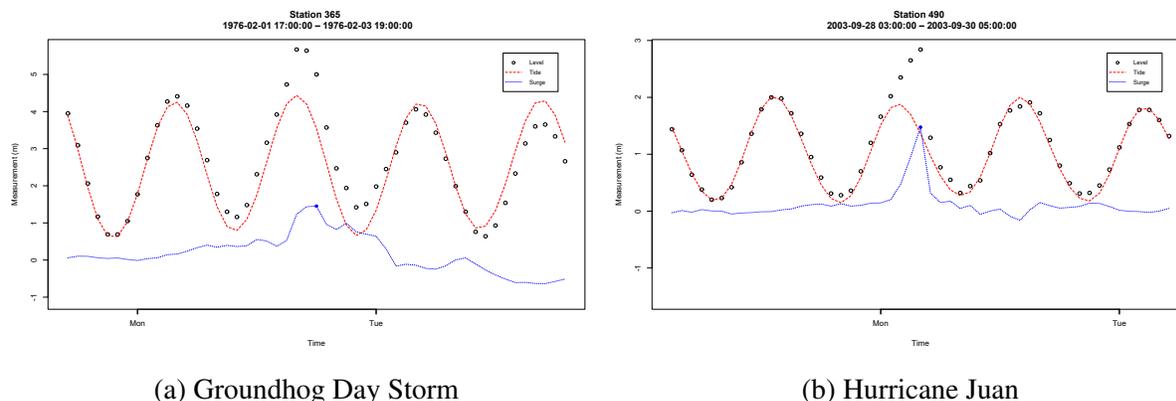


Figure 4.4: Tidal fitting examples at Stations 365 (Yarmouth, NS; left) and 490 (Halifax, NS; right). (a) Groundhog Day Gale, February 1–3, 1976. (b) Hurricane Juan, September 28–30, 2003.

At the time, a powerful winter storm with winds of almost 200 km/h in tandem with a rare event within the tidal cycle known as Saros decimated the northeastern United States and southeastern Canada. In Canada, damages were estimated to be around \$10 million. The right panel accurately represents Hurricane Juan, mentioned earlier, with a maximum surge of 1.5 m occurring at around 1 am local time.

4.3.3 Data selection and augmenting

To account for the vast amount of missing data and curb the risk of mislabeling an observed value as extreme, maxima were only recorded for years where there were no more than 15% of the hourly measurements missing. This choice seems appropriate because when too few data are available, there is a significant risk that the actual maximum is lost in the missing data. Using the 15% cutoff, as argued in Appendix 4.B, provided reasonable assurance that the recorded annual maxima was the true maximum and that it stemmed from a GEV distribution. This process provided us with 647 annual maxima. Had the data been complete over all 50 years and 21 locations, there would have been $1050 = 50 \times 21$ values, meaning we only have 61.6% of the total possible annual maxima. Note that in the case of Station 3248 (Vieux Québec, QC), this initially left only two annual maxima available for use. This station was included nonetheless as it will later serve to demonstrate the data-sharing capabilities of the employed spatial model.

Unfortunately, a strict application of this rule could cause important data points to be excluded from the model. For example, at station 490, there would then be no annual maximum recorded for the year 2003, when Hurricane Juan occurred. Regardless of the amount of missing data, such an event is crucial to the accurate modeling of flood risk in Atlantic Canada.

To address this issue, we revisited the years with insufficient data and checked if a maximum from that year could feasibly fall in line with the initially recorded maxima. In such cases, the

4.3 Flood risk in Atlantic Canada

4 Data augmentation algorithm.

1. Extract an initial data matrix using only maxima that appear in years with less than 15% of the data missing.
2. Assume that for each buoy s_i with $i \in \{1, \dots, 21\}$, the maximum annual surge Z_i follows a generalized extreme value (GEV) distribution with parameters $\theta_i = (\mu_i, \sigma_i) \in \mathbb{R} \times (0, \infty)$ and $\xi \in \mathbb{R}$, i.e.,

$$Z_i \sim \mathcal{G}\mathcal{E}\mathcal{V}(\mu_i, \sigma_i, \xi). \quad (4.6)$$

3. For the observed maxima, \mathbf{z} , fit the marginals jointly through MLE by assuming that they are mutually independent and share a common shape parameter, ξ . The corresponding log-likelihood is

$$\begin{aligned} \ell(\boldsymbol{\mu}, \boldsymbol{\sigma}, \xi | \mathbf{z}) = \sum_{i=1}^{21} \left[-n_i \log \sigma_i - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{n_i} \log \left\{ 1 + \xi \left(\frac{z_{ji} - \mu_i}{\sigma_i} \right) \right\} \right. \\ \left. - \sum_{j=1}^{n_i} \left\{ 1 + \xi \left(\frac{z_{ji} - \mu_i}{\sigma_i} \right) \right\}^{-1/\xi} \right]. \end{aligned}$$

Call the resulting estimates $\hat{\xi}$ and $\theta_i = (\hat{\mu}_i, \hat{\sigma}_i)$ for each $i \in \{1, \dots, 21\}$.

4. For each station s_i with $i \in \{1, \dots, 21\}$, look through the raw hourly data in all years where no maximum was recorded because there was more than 15% of data missing. For each such year j , find the maximum z_{ij}^* and include it in the dataset if its value is above the α quantile of the fitted marginal distribution $\mathcal{G}\mathcal{E}\mathcal{V}(\theta_i, \xi)$.
-

maximum was added to the data set even if more than 15% of the data were missing. Details of the procedure are given in Algorithm 4.

In Step 2 of Algorithm 4, the assumption that the shape parameter is the same for all buoys was made after an initial fit of the marginal models through maximum likelihood. As the sample sizes are small, our analysis revealed large standard errors for the marginal shape parameter estimates. As all but one pair of (asymptotic) 95% confidence intervals intersected (see Figure 4.5), there was no statistical basis to support the notion that the shape parameter varies spatially. The assumption of a common shape parameter is frequent in the hydrological literature; see, e.g., [Sang and Gelfand \(2009, 2010\)](#), [Ghosh and Mallick \(2011\)](#) or [Reza Najafi and Moradkhani \(2013\)](#).

An application of this procedure with $\alpha = 0.05$ in Step 4 led to the inclusion of $n_\alpha = 138$ values, bringing the number of annual maxima from 647 to 785. Other values of α are clearly possible and, in our case, n_α decreases almost linearly as $\alpha \rightarrow 1$. For example, $n_{0.1} = 128$ and $n_{0.25} = 104$. In preliminary marginal analysis (Appendix 4.C), the maximum likelihood estimate of the common shape parameter as calculated in Step 3 did not vary substantially, ranging from 0.03

4.3 Flood risk in Atlantic Canada

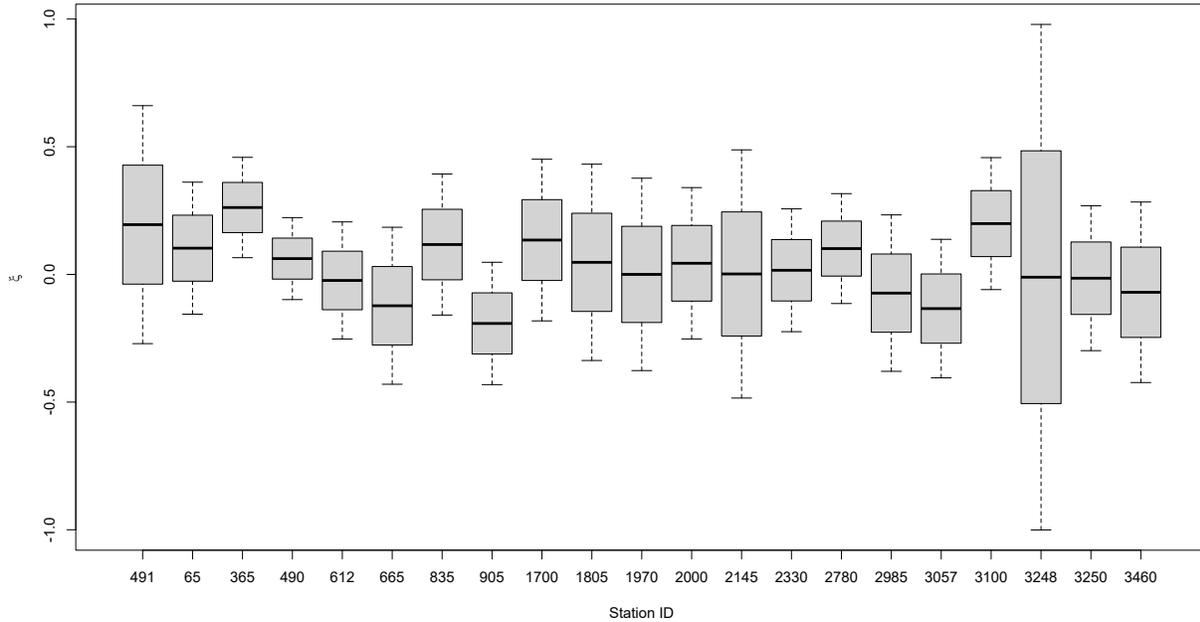


Figure 4.5: Asymptotic 95% confidence intervals for the MLE's fitted to each of the selected 21 stations marginally.

to 0.07. Overall, $\alpha = 0.05$ seemed like a reasonable choice given the intent to found the inference on a sufficiently broad base.

It is important to note that we assumed that there was no trend in time for the annual maximum surges. At location s_i for some $i \in \{1, \dots, 21\}$, we had at our disposal a series of at most 8760 hourly surges for each (non-leap) year t , say $X_{i,1}(t), \dots, X_{i,8760}(t)$, and the corresponding annual maximum $Z_i(t) = \max\{X_{i,1}(t), \dots, X_{i,8760}(t)\}$ from which to estimate θ_i and ξ . However, as discussed, the series is typically shorter than that, due to missing values. Looking for a temporal trend in the annual maximum surges, we used the Mann–Kendall test for trend. At the 5% level, and accounting for multiple comparisons by controlling the false discovery rate, there was no evidence to reject the null hypothesis of no temporal trend in the annual maximum surges at any of the 21 locations.

Spurious data*

As a final check, we examined the augmented data to make sure that none of the included maxima is a result of some type of measurement error. This was accomplished by cross-referencing the data augmented Q-Q plots based on marginal GEV's, fitted with MLE, with historical records; when a proposed maximum seemed particularly large, evidence for its occurrence was sought in, e.g., the press, as was the case for Hurricane Juan. While this procedure was by no means exhaustive, we

4.3 Flood risk in Atlantic Canada

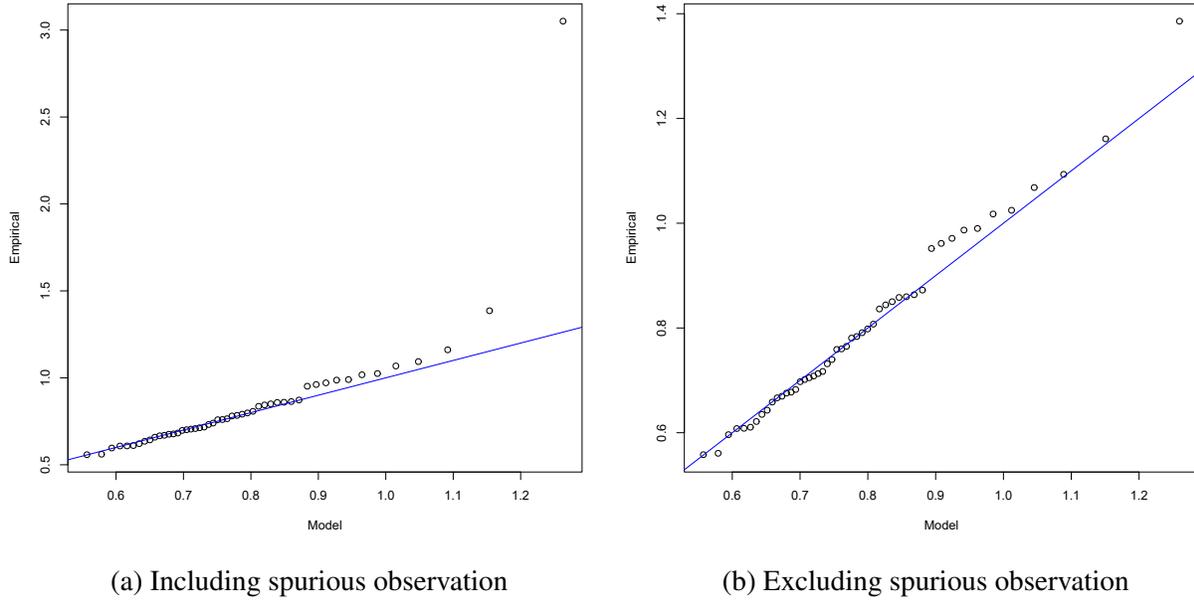


Figure 4.6: Q-Q plot comparison for constrained MLE model at Station 65 (St John, NB).

believe that the inclusion of this ancillary step improved the quality of our data, and subsequently the credibility of the model.

An example of its application at station 65 (St. John, NB) can be found in Figure 4.6. After the augmenting procedure, the maximum annual surge for 2014 was recorded at 3 m. Given the disparity between this observation and all other annual maxima for that station, clearly this would mark a significant event. However, no record of such an event could be found in our inquiries. The point was thus removed from the collection of observations and the resulting Q-Q plot in Figure 4.6b demonstrates a much better fit.

4.3.4 The statistical model

For the extracted and augmented annual maxima surges, we consider a HBM as described in Section 4.2.1. Specifically, let $\mathbf{Z} = (Z_1, \dots, Z_{21})$ be the random vector of maximum annual surges at each of the 21 buoys considered. Denote $\Delta = (\Theta, \eta, \xi) = (\theta_1, \dots, \theta_{21}, \eta, \xi)$, where for each $i \in \{1, \dots, 21\}$, $\theta_i = (\mu_i, \sigma_i)$. The extremal spatial Bayesian hierarchical model for Δ was constructed by assuming a joint distribution for $\mathbf{Z} | (\Theta, \xi)$ and a hierarchical prior for Θ involving a vector, η , of hyperparameters. Conditional on the observed values of \mathbf{Z} , the posterior for can be written

$$p(\Delta | \mathbf{Z}) \propto p(\mathbf{Z} | \Theta, \xi) p(\Theta | \eta, \xi) p(\eta, \xi). \quad (4.7)$$

4.3 Flood risk in Atlantic Canada

The data, process and prior layers are detailed as follows.

Data level

As discussed in Section 4.3.3, the marginal distributions of Z_1, \dots, Z_{21} will follow a GEV distribution with the same structure as (4.6). As we are interested in inference beyond individual sites, we include a dependence structure for the residual dependence.

Given the dimension of the problem and the prevalence of missing data, we opted for a Student's t copula, denoted $C_{\nu, \Xi}$. Owing to the difficulty in estimating the degrees of freedom, this parameter was taken as fixed and different values of ν were tested for best fit. As for the correlation structure induced by Ξ , it was assumed to be in the form of an exponential correlogram, i.e.,

$$\Xi_{ij} = \exp(-d_{ij}/\rho_{\Xi}),$$

where d_{ij} is the straight line distance between s_i and s_j when considering the Earth's curvature. This is also known as the great circle distance between s_i and s_j , which is very common in geostatistics. Other choices of correlograms were briefly considered within the Matérn class,

$$\Xi_{ij}(\kappa) = \frac{2^{1-\kappa}}{\Gamma(\kappa)} \left\{ \sqrt{2\kappa}(d_{ij}/\rho_{\Xi}) \right\}^{\kappa} K_{\kappa} \left\{ \sqrt{2\kappa}(d_{ij}/\rho_{\Xi}) \right\}.$$

In this parametrization, $\kappa = 0.5$ corresponds to the exponential correlogram, and the limiting case $\kappa \rightarrow \infty$ coincides with the Gaussian correlogram. To justify the choice $\kappa = 0.5$, we compared, for all distinct $i, j \in \{1, \dots, 21\}$, the two following estimates of Ξ_{ij} :

- (i) the unstructured correlation estimate $\tilde{\Xi}_{ij}$ obtained by inversion of Kendall's tau;
- (ii) the structured estimate $\hat{\Xi}_{ij}(\kappa)$ derived from the median of the posterior for ρ_{Ξ} for $\kappa \in \{0.5, 1.5, 2.5\}$ and $\kappa \rightarrow \infty$.

The resulting boxplots displayed in Figure 4.7 show the minimal influence of κ . The two most extreme outliers in the plots of Figure 3.7 correspond to pairs of stations that are geographically near: Halifax (490) and Bedford (491) on one hand, Vieux Québec (3248) and Lauzon (3250) on the other hand. In each case, the distance between the stations is smaller than 4 km and in the unstructured case, the correlation estimates happen to be near zero, whereas they should be close to 1. This oddity does not reflect reality but is due to the very small number of pairs of simultaneous observations available at these two locations (15 and 2, respectively).

4.3 Flood risk in Atlantic Canada

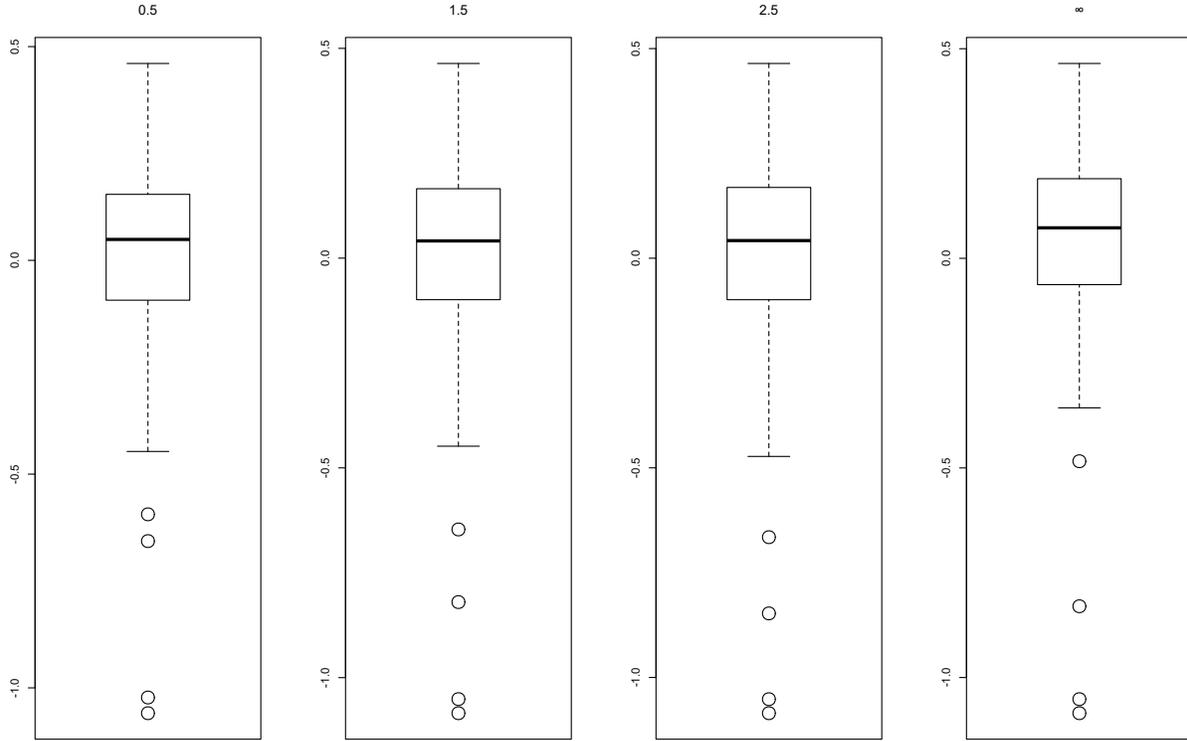


Figure 4.7: Boxplots showing the difference between raw correlation estimates and structured correlation estimates based on a Matérn correlogram with parameter $\kappa \in \{0.5, 1.5, 2.5\}$ and $\kappa \rightarrow \infty$

Process level

In the process layer we model the location parameters $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{21})$ and log scale parameters $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_{21}) = (\log(\sigma_1), \dots, \log(\sigma_{21}))$ using latent spatial Gaussian fields, i.e., we set

$$\boldsymbol{\mu} \sim \mathcal{N}_{21}(\mathbf{X}_{\boldsymbol{\mu}}\boldsymbol{\beta}_{\boldsymbol{\mu}}, \tau_{\boldsymbol{\mu}}^2 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}) \quad \text{and} \quad \boldsymbol{\Phi} \sim \mathcal{N}_{21}(\mathbf{X}_{\boldsymbol{\Phi}}\boldsymbol{\beta}_{\boldsymbol{\Phi}}, \tau_{\boldsymbol{\Phi}}^2 \boldsymbol{\Sigma}_{\boldsymbol{\Phi}}). \quad (4.8)$$

As discussed in Section 4.2.1, these hierarchical structures are useful because the matrices $\mathbf{X}_{\boldsymbol{\mu}}$ and $\mathbf{X}_{\boldsymbol{\Phi}}$ can include covariate information about the region to improve inference capabilities. The two strongest drivers of extreme surges are known to be wind speed and atmospheric sea-level pressure (Muis et al., 2016). To our knowledge, wind data are unavailable on a fine-resolution grid in the domain of interest. In contrast, we were able to incorporate atmospheric sea-level pressure into the model. Due to the lack of observational data, however, sea-level pressure data had to be obtained from a reconstruction of past meteorology by the Canadian Regional Climate Model 5 (CRCM5).

The CRCM5 is the fifth-generation Canadian RCM (Martynov et al., 2013) developed by the *Centre pour l'étude et la simulation du climat à l'échelle régionale* (ESCIER) at the Université du

4.3 Flood risk in Atlantic Canada



Figure 4.8: Sea-level pressure provided by CRCM5 for Atlantic Canada. At each grid point, the mean annual minimum pressure is given in kPA. It is the use of this variable as a covariate at the process level of the Bayesian hierarchical model that enables spatial interpolation at unmonitored locations.

Québec à Montréal. The simulated data were generated for the Northeastern part of North America and archived at every three hours. More information on this climate reconstruction can be found in [Bresson et al. \(2017\)](#).

In Figure 4.8, the mean annual minimum pressure provided by CRCM5 is given in kilopascals (kPA) at each grid point in the region of study. The 21 stations are displayed there, along with a sub-grid of the pressure covariate. As can be seen, sea level pressure lowers from South-West to North-East. The inclusion of this feature at the process level of the Bayesian hierarchical model enables realistic spatial interpolation at unmonitored locations.

As is well known, lower pressure induces large surges. We extracted from the reconstructed sea-level pressure data various summary statistics at each location s_i with $i \in \{1, \dots, 21\}$, e.g., the mean, median, minimum, maximum, standard deviation, and range. We then regressed maximum likelihood estimates of the marginal parameters μ and Φ using these summary statistics and simple transforms thereof. Using standard diagnostic tools, we concluded that the mean annual minimum sea-level pressure x_i at location s_i was the best predictor and that on balance, there was little benefit to the inclusion of additional variables. Therefore, we assumed that

$$\mathbb{E}(\mu_i) = \beta_{0,\mu} + \beta_{1,\mu}x_i \quad \text{and} \quad \mathbb{E}(\phi_i) = \beta_{0,\Phi} + \beta_{1,\Phi}x_i.$$

4.3 Flood risk in Atlantic Canada

We further assumed that for each $i, j \in \{1, \dots, d\}$, the (i, j) th entries of the correlation matrices Σ_{μ} and Σ_{Φ} are respectively given by $(\Sigma_{\mu})_{ij} = \exp(-d_{ij}/\rho_{\mu})$ and $(\Sigma_{\Phi})_{ij} = \exp(-d_{ij}/\rho_{\Phi})$, where d_{ij} is again the greater circle distance between s_i and s_j . Note that these will correspond to exponential correlograms with range parameters ρ_{μ} and ρ_{Φ} , respectively (Cressie, 1993). Finally, τ_{μ}^2 and τ_{Φ}^2 stand for the variance of μ_i and ϕ_i , respectively; these are assumed to be the same for all $i \in \{1, \dots, 21\}$, which amounts to saying that the variance of the Gaussian fields which generate the location parameters μ and the log-scale parameters Φ is constant over space.

This set up allows for the sharing of information between locations based on their proximity and is useful when interpolating our results to unmonitored locations. With these latent fields, we can fit GEVs to locations where no observations exists, allowing us to complete a more comprehensive analysis of flood risk on the Atlantic coast.

Prior level

To complete the model specification, prior distributions are required on $\xi, \tau_{\mu}^2, \tau_{\Phi}^2, \beta_{0,\mu}, \beta_{1,\mu}, \beta_{0,\Phi}, \beta_{1,\Phi}, \rho_{\mu}, \rho_{\phi},$ and ρ_{Ξ} . They were all taken to be mutually independent and non-informative, i.e., $\pi(\xi) \propto 1, \pi(\tau_{\mu}^2) \propto 1/\tau_{\mu}^2, \pi(\tau_{\Phi}^2) \propto 1/\tau_{\Phi}^2, \pi(\beta_{0,\mu}) \propto 1, \pi(\beta_{1,\mu}) \propto 1, \pi(\beta_{0,\Phi}) \propto 1, \pi(\beta_{1,\Phi}) \propto 1,$ as well as $\rho_{\mu}, \rho_{\phi}, \rho_{\Xi} \sim \mathcal{U}[0, \max_{ij} d_{ij}/3]$ with the restriction that $\rho_{\Xi} \leq \min(\rho_{\mu}, \rho_{\Phi})$ to avoid non-identifiability issues.

Additionally, the choice of priors for $\beta_{0,\mu}, \beta_{1,\mu}, \beta_{0,\Phi}, \beta_{1,\Phi}, \tau_{\mu}^2,$ and τ_{Φ}^2 facilitate a more efficient fitting procedure as the resulting posterior distributions will have closed forms for sample generation. Consider for μ where $\pi(\tau_{\mu}^2) \propto 1/\tau_{\mu}^2$ corresponds to an improper inverse gamma distribution, $\tau_{\mu}^2 \sim \lim_{a,b \rightarrow 0} \text{IG}(a, b)$. Recall, $Y \sim \text{IG}(a, b)$ has a inverse gamma distribution with density

$$f(y) = \frac{b^a}{\Gamma(a)x^{a+1}} \exp\left(-\frac{b}{x}\right).$$

Given the hierarchical structure of (4.7), the posterior of $\beta_{\mu}|\tau_{\mu}^2$ and τ_{μ}^2 can be written, using Bayes' Rule,

$$\begin{aligned} p(\beta_{\mu}, \tau_{\mu}^2 | \mu) &= \frac{p(\beta_{\mu}, \tau_{\mu}^2, \mu)}{p(\mu)} \\ &= \frac{p(\mu | \beta_{\mu}, \tau_{\mu}^2) p(\beta_{\mu}, \tau_{\mu}^2)}{p(\mu)} \\ &\propto p(\mu | \beta_{\mu}, \tau_{\mu}^2) p(\beta_{\mu} | \tau_{\mu}^2) p(\tau_{\mu}^2). \end{aligned}$$

4.3 Flood risk in Atlantic Canada

Using the distributional assumption in (4.8), we have

$$\begin{aligned}
p(\boldsymbol{\beta}_\mu, \tau_\mu^2 | \boldsymbol{\mu}) &\propto \frac{1}{\tau_\mu^2} \cdot 1 \cdot \left(\frac{1}{\tau_\mu^2}\right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\tau_\mu^2} (\boldsymbol{\mu} - \mathbf{X}_\mu \boldsymbol{\beta}_\mu)^\top \Sigma_\mu^{-1} (\boldsymbol{\mu} - \mathbf{X}_\mu \boldsymbol{\beta}_\mu) \right\} \\
&= \left(\frac{1}{\tau_\mu^2}\right)^{\frac{n-p}{2}+1} \exp \left(-\frac{1}{2\tau_\mu^2} \boldsymbol{\mu}^\top \Sigma_\mu^{-1} \boldsymbol{\mu} \right) \left(\frac{1}{\tau_\mu^2}\right)^{\frac{p}{2}} \\
&\quad \times \exp \left\{ -\frac{1}{2\tau_\mu^2} \underbrace{(-\boldsymbol{\mu}^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu - \boldsymbol{\beta}_\mu^\top \mathbf{X}_\mu^\top \Sigma_\mu^{-1} \boldsymbol{\mu} + \boldsymbol{\beta}_\mu^\top \mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu)}_{(a)} \right\}.
\end{aligned}$$

To simplify the expression further, we complete the square in (a)

$$\begin{aligned}
&-\boldsymbol{\mu}^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu - \boldsymbol{\beta}_\mu^\top \mathbf{X}_\mu^\top \Sigma_\mu^{-1} \boldsymbol{\mu} + \boldsymbol{\beta}_\mu^\top \mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu \\
&= \boldsymbol{\beta}_\mu^\top \mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu - 2\boldsymbol{\mu}^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu \\
&= \boldsymbol{\beta}_\mu^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \boldsymbol{\beta}_\mu - 2 \underbrace{\boldsymbol{\mu}^\top \Sigma_\mu^{-1} \mathbf{X}_\mu (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu)^{-1} (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu)}_{=\hat{\boldsymbol{\beta}}^\top} \boldsymbol{\beta}_\mu \\
&= \boldsymbol{\beta}_\mu^\top \mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu \boldsymbol{\beta}_\mu - 2\hat{\boldsymbol{\beta}}^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \boldsymbol{\beta}_\mu + \hat{\boldsymbol{\beta}}^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \hat{\boldsymbol{\beta}}_\mu - \hat{\boldsymbol{\beta}}_\mu^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \hat{\boldsymbol{\beta}}_\mu \\
&= (\boldsymbol{\beta}_\mu - \hat{\boldsymbol{\beta}}_\mu)^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) (\boldsymbol{\beta}_\mu - \hat{\boldsymbol{\beta}}_\mu) - \hat{\boldsymbol{\beta}}_\mu^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \hat{\boldsymbol{\beta}}_\mu.
\end{aligned}$$

Returning to the joint posterior, one finds that

$$\begin{aligned}
p(\boldsymbol{\beta}_\mu, \tau_\mu^2 | \boldsymbol{\mu}) &\propto \underbrace{\left(\frac{1}{\tau_\mu^2}\right)^{\frac{n-p}{2}+1} \exp \left[-\frac{1}{2\tau_\mu^2} \left\{ \boldsymbol{\mu}^\top \Sigma_\mu^{-1} \boldsymbol{\mu} - \hat{\boldsymbol{\beta}}_\mu^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \hat{\boldsymbol{\beta}}_\mu \right\} \right]}_{\propto p(\tau_\mu^2 | \boldsymbol{\mu})} \\
&\quad \times \underbrace{\left(\frac{1}{\tau_\mu^2}\right)^{\frac{p}{2}} \exp \left\{ -\frac{1}{2\tau_\mu^2} (\boldsymbol{\beta}_\mu - \hat{\boldsymbol{\beta}}_\mu)^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) (\boldsymbol{\beta}_\mu - \hat{\boldsymbol{\beta}}_\mu) \right\}}_{\propto p(\boldsymbol{\beta}_\mu | \tau_\mu^2, \boldsymbol{\mu})},
\end{aligned}$$

implying that

$$\tau_\mu^2 | \boldsymbol{\mu} \sim \text{IG} \left\{ \frac{n-p}{2}, \boldsymbol{\mu}^\top \Sigma_\mu^{-1} \boldsymbol{\mu} - \hat{\boldsymbol{\beta}}_\mu^\top (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu) \hat{\boldsymbol{\beta}}_\mu \right\}$$

and

$$\boldsymbol{\beta}_\mu | \tau_\mu^2, \boldsymbol{\mu} \sim \mathcal{N} \left\{ \hat{\boldsymbol{\beta}}_\mu, \tau_\mu^2 (\mathbf{X}_\mu^\top \Sigma_\mu^{-1} \mathbf{X}_\mu)^{-1} \right\}.$$

4.3.5 Model estimation

In total, eight models were fitted, corresponding to the independence copula Π and Student's t copula $C_{\nu, \Xi}$ with varying degrees of freedom ν . In addition, a special case of the models in which $\phi_1 = \dots = \phi_{21} = \varphi$ was run for comparison purposes. For each choice of copula, the model was fitted using the Metropolis–Hastings–within–Gibbs algorithm described in Section 4.2.1. Specifically, we produced four independent chains of 30,000 observations for each model with the first 5,000 iterations of each chain being discarded, considered as warm-up. To deal with serial dependence between iterations, we kept every 10th realization. In the end, we are left with 10,000 realizations from the posterior distribution.

To assess the performance of the algorithm for each individual model, the samples from the four chains were used to calculate the multivariate potential scale reduction factor (MPSRF), \hat{R}_M . The MPSRF is a scalar value, introduced in Brooks and Gelman (1998) as a multivariate extension to the potential scale reduction factor of Gelman and Rubin (1992), which quantifies performance by considering how well the individual chains mix. This is accomplished by comparing the within- and between-chain covariance matrices, where each chain is produced using different starting values. In particular, starting values should be chosen to be overdispersed with respect to the target distribution. In each of our chains, Θ was first chosen by jittering the maximum likelihood estimates from Section 4. The remaining hyperparameters and ξ were initialized so as to cover a realistic range of possibilities. Values of \hat{R}_M close to 1 imply that the chains mix well and that each of the chains produced samples closely representing the target distribution whereas large values of \hat{R}_M would imply that additional simulation could improve fit. Being close to 1, the values of \hat{R}_M reported in Table 4.2 suggest that the results of the MCMC algorithm are reliable for all of our models. Examples of the well-mixing of the individual chains for the winning model can be found in Figure 4.9.

To compare the fit of each model, we used the Watanabe–Akaike Information Criterion (WAIC) (Watanabe, 2009; Gelman et al., 2014). In practice, WAIC compares models by considering numerical estimates of the log predictive pointwise density (lppd). The lppd is written

$$\text{lppd} = \sum_{i=1}^{50} \log \int p(\mathbf{Z}_i | \Delta) p_{\text{post}}(\Delta) d\Delta,$$

where $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{i21})$ is the vector of observed maxima in year i . For a collection of realizations $\Delta_1, \dots, \Delta_{10000}$ from the posterior distribution, the lppd is estimated via

$$\widehat{\text{lppd}} = \sum_{i=1}^{50} \log \left\{ \frac{1}{N} \sum_{j=1}^{10000} p(\mathbf{Z}_i | \Delta_j) \right\}.$$

4.3 Flood risk in Atlantic Canada

Table 4.2: Model comparison using WAIC for different choices of copulas and log-scale parameters that are either equal or freely varying. Reported in the table are $-2\widehat{\text{lppd}}$, $\widehat{p}_{\text{WAIC},j}$ and $\widehat{\text{WAIC}}_j$ for $j \in \{1, 2\}$, and the value of the MPSRF \widehat{R}_M .

Model	$-2\widehat{\text{lppd}}$	$\widehat{p}_{\text{WAIC},1}/\widehat{p}_{\text{WAIC},2}$	$\widehat{\text{WAIC}}_1/\widehat{\text{WAIC}}_2$	\widehat{R}_M
Φ as per Eq. (4.8)				
Π	-468.246	35.588/38.504	-397.070/ - 391.239	1.01
$C_{4,\Xi}$	-489.849	70.317/75.599	-349.2147/ - 338.6504	1.01
$C_{5,\Xi}$	-497.107	65.244/69.539	-366.619/ - 358.030	1.02
$C_{7,\Xi}$	-501.759	60.784/64.088	-380.192/ - 373.583	1.01
$C_{8,\Xi}$	-501.821	59.980/63.153	-381.861/ - 375.516	1.02
$C_{9,\Xi}$	-501.540	59.789/62.708	-381.962/ - 376.123	1.01
$C_{10,\Xi}$	-500.702	59.245/62.618	-382.212/ - 375.465	1.02
$C_{15,\Xi}$	-495.762	60.243/63.679	-375.277/ - 368.403	1.01
$\phi_1 = \dots = \phi_{21}$				
Π	-319.997	26.265/29.096	-267.467/ - 261.805	1.01
$C_{1,\Xi}$	-268.491	61.080/78.938	-146.331/ - 110.614	1.01
$C_{5,\Xi}$	-372.844	42.625/48.522	-287.594/ - 275.800	1.01
$C_{6,\Xi}$	-374.812	42.634/48.055	-289.543/ - 278.702	1.01
$C_{7,\Xi}$	-375.615	43.107/47.919	-289.401/ - 279.777	1.01
$C_{9,\Xi}$	-374.801	43.478/48.371	-287.846/ - 278.060	1.01
$C_{10,\Xi}$	-374.001	43.465/48.674	-287.071/ - 276.654	1.01

The models are also penalized based on complexity. In particular, there are two adjustments which have been proposed in the literature, namely

$$p_{\text{WAIC},1} = 2 \sum_{i=1}^{50} \left[\log \left[\mathbb{E} \{p(\mathbf{Z}_i|\Delta)\} - \mathbb{E} \{\log p(\mathbf{Z}_i|\Delta)\} \right] \right] \quad \text{and}$$

$$p_{\text{WAIC},2} = \sum_{i=1}^{50} \text{var}\{\log p(\mathbf{Z}_i|\Delta)\},$$

which are estimated by

$$\widehat{p}_{\text{WAIC},1} = 2 \sum_{i=1}^{50} \left[\log \left\{ \frac{1}{10000} \sum_{j=1}^{10000} p(\mathbf{Z}_i|\Delta_j) \right\} - \frac{1}{10000} \sum_{j=1}^{10000} \log p(\mathbf{Z}_i|\Delta_j) \right] \quad \text{and}$$

$$\widehat{p}_{\text{WAIC},2} = \sum_{i=1}^{50} V_{j=1}^{10000} \{\log p(\mathbf{Z}_i|\Delta_j)\},$$

respectively, where $V_{j=1}^N a_j = \sum_{j=1}^N (a_j - \bar{a})^2 / (N - 1)$ is the sample variance. Table 4.2 reports the estimated values of WAIC using the samples from our MCMC simulations, i.e., for $j \in \{1, 2\}$,

$$\widehat{\text{WAIC}}_j = -2\widehat{\text{lppd}} + 2\widehat{p}_{\text{WAIC},j}.$$

4.3 Flood risk in Atlantic Canada

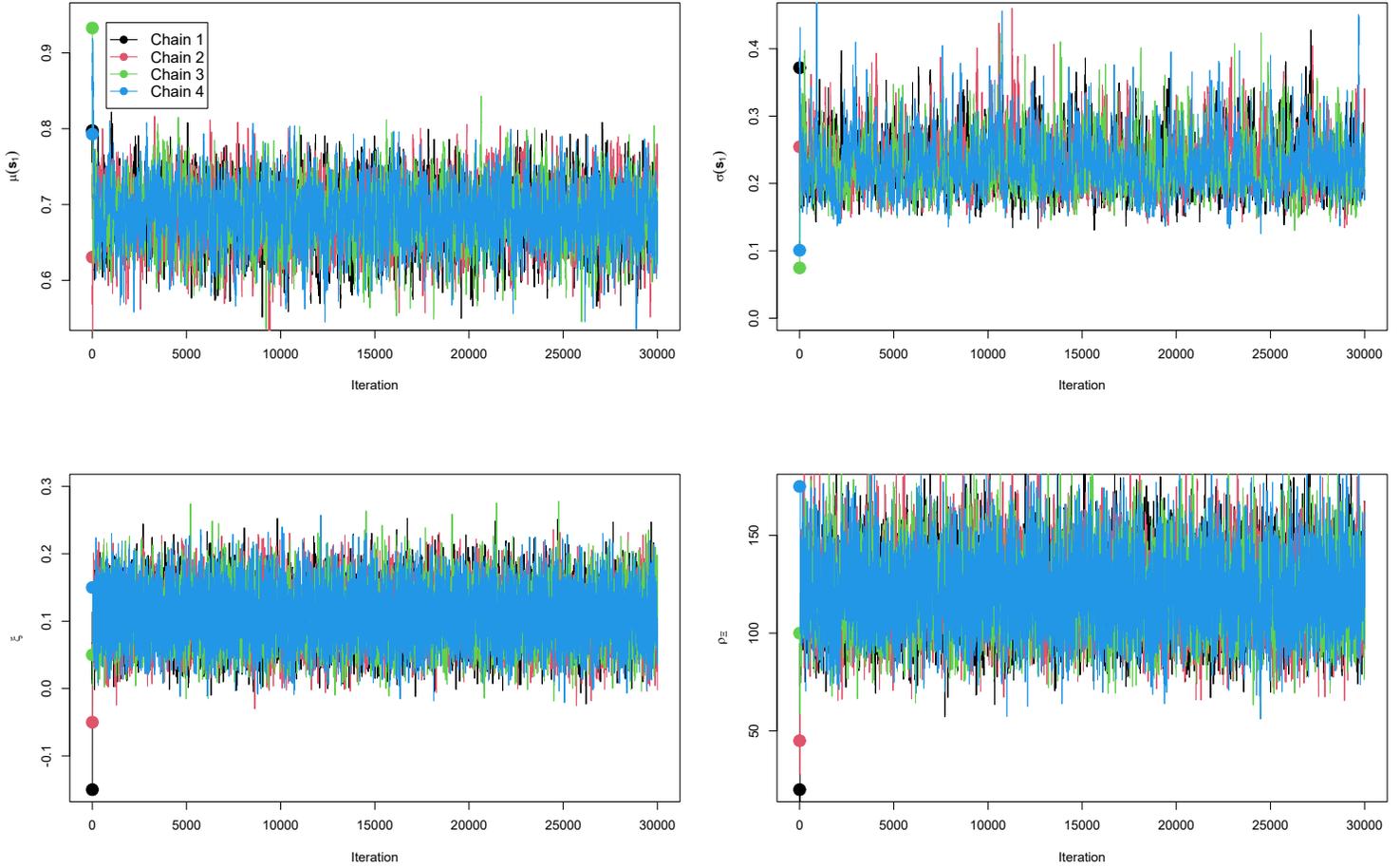


Figure 4.9: Examples of the well mixing of chains for the winning model, $C_{9,\Xi}$ with spatially varying log-scale parameters. Chains for μ_1 (top left), σ_1 (top right), ξ (bottom left) and ρ_Ξ (bottom right) are shown for a variety of starting parameters. The filled in circles represent the starting points of each chain. Here, s_1 corresponds to station 491.

On the basis of these results, it appears that *stricto sensu*, the independence copula Π provides the best overall fit. However, as illustrated by [Genest et al. \(2009\)](#) and others, it is very difficult to detect tail dependence based on a sample of 50 observations. Given that in the present case, the sample size is frequently smaller than 20, it seems that incorporating the t copula into the model is not unreasonable, even if the assumption of independence would be more parsimonious. Moreover, while incorporating this feature into the model may lead to an overestimation of risk, it constitutes a conservative assumption that is preferred by risk regulators. For these reasons, it seems prudent to allow for asymptotic dependence between locations.

Therefore, we ultimately opted for a Student's t copula with 9 degrees of freedom. This choice was further substantiated by Kendall's test, which rejected the null hypothesis of independence at

4.3 Flood risk in Atlantic Canada

the 5% level for 41 of the 210 possible pairs of locations at which data were collected. Only one location (Vieux Québec, QC) was independent of all the others, but this could be attributed to the lack of power of the test, given that there were only six maxima available for that station.

Estimates for the return level of annual maxima surges are provided in Table 4.3 using two different methods. Maximum likelihood estimates were obtained by fitting the marginals independently; they are available only at the four monitored stations. The estimates based on the HBM are the medians of the posterior distributions of the specified return levels. To generate these posteriors, return levels were computed using realizations from the posterior distributions of the GEV parameters at the specified locations. These parameters are affected by the inclusion of dependence structures at both the data and process level.

To compare these two approaches, Table 4.3 also includes the percent relative change observed by going from the MLE to the HBM approach, viz. $\%RC = 100 \times (HBM - MLE)/MLE$, at the original stations. At the selected stations, the HBM consistently predicts heavier return levels than the MLE. This results from the choice of the median as the representative of the posterior distributions of the return levels. Because these posteriors are asymmetric, the median may not coincide with the mean or the mode. A benefit of the Bayesian modeling approach is that one has a choice between these different summaries of the posterior distribution; depending on this choice, which would be guided by the context, the HBM may produce higher or smaller return levels than those provided by the MLE.

An added benefit of Bayesian modeling is that it produces posterior distribution estimates, which makes the calculation of uncertainty much easier. To illustrate this point, 95% credible intervals based on the HBM are provided in Table 4.3. For comparison purposes, asymptotic 95% confidence intervals for the MLEs are also reported. Observe how wide some of the latter become for longer return periods, to the point of including negative values in the cases of Bedford (491) and Vieux Québec (3248), where the number of observations was very small.

It is also interesting to see how the results are impacted by the inclusion of spatiality into the Bayesian model. As an illustration, consider the 50-year return level for Vieux Québec and Lauzon, located just 4 km apart on the St. Lawrence River. When estimation is done by maximum likelihood, there is a difference of approximately 70 cm in the calculated returns for these two locations. This seems like a rather large discrepancy for buoys in the same body of water that are so close. In contrast, the HBM generates returns that are more homogeneous. Note also how the sharing of information between locations in the HBM leads to 95% credible intervals whose widths, for any given return period, do not vary much as a function of the record length.

Table 4.3 further illustrates the ability of the HBM approach to generate return levels at unmonitored locations. This key feature is not shared by the maximum likelihood approach, as the latter does not allow for information to be shared across neighboring locations where the initial model

4.3 Flood risk in Atlantic Canada

Table 4.3: Comparison of the return level for various return periods (RP, in years) computed with either the individual margins of the Hierarchical Bayesian Model (HBM) or the maximum likelihood estimate based on that single location. Also included are asymptotic 95% confidence intervals for the MLE and 95% credible intervals derived from the HBM, as well as the percent relative change given by $\%RC = 100 \times (\text{HBM} - \text{MLE})/\text{MLE}$. As stations 3200 and 3260 are unmonitored, the MLEs, and hence the $\%RC$, cannot be computed.

RP	MLE	Conf Int	HBM	Cred Int	%RC	RP	MLE	Conf Int	HBM	Cred Int	%RC
10	0.94	(0.85, 1.02)	1.18	(1.03, 1.39)	25	10	1.08	(0.89, 1.26)	1.25	(1.09, 1.50)	16
25	1.06	(0.92, 1.20)	1.44	(1.23, 1.76)	36	25	1.23	(0.86, 1.60)	1.53	(1.30, 1.90)	24
50	1.15	(0.97, 1.34)	1.65	(1.38, 2.07)	43	50	1.37	(0.79, 1.96)	1.75	(1.46, 2.23)	28
100	1.25	(1.00, 1.50)	1.87	(1.54, 2.40)	50	100	1.53	(0.65, 2.41)	1.99	(1.63, 2.59)	30
250	1.39	(1.03, 1.75)	2.20	(1.77, 2.89)	58	250	1.77	(0.34, 3.21)	2.34	(1.87, 3.13)	32
500	1.50	(1.03, 1.96)	2.46	(1.94, 3.32)	65	500	1.99	(-0.01, 3.99)	2.62	(2.05, 3.59)	32
1000	1.61	(1.03, 2.19)	2.75	(2.12, 3.83)	71	1000	2.23	(-0.50, 4.96)	2.93	(2.24, 4.13)	31

(a) Station 490: Halifax, NS						(b) Station 491: Bedford Institute, NS					
RP	MLE	Conf Int	HBM	Cred Int	%RC	RP	MLE	Conf Int	HBM	Cred Int	%RC
10	1.81	(1.47, 2.14)	2.26	(2.02, 2.59)	25	10	2.30	(2.07, 2.52)	2.30	(2.08, 2.61)	0
25	1.95	(1.34, 2.57)	2.63	(2.30, 3.13)	35	25	2.57	(2.20, 2.95)	2.68	(2.38, 3.14)	4
50	2.06	(1.13, 2.99)	2.93	(2.52, 3.57)	42	50	2.78	(2.23, 3.32)	2.99	(2.61, 3.59)	8
100	2.17	(0.85, 3.49)	3.25	(2.75, 4.07)	50	100	2.98	(2.22, 3.73)	3.32	(2.85, 4.09)	12
250	2.31	(0.36, 4.26)	3.71	(3.06, 4.80)	61	250	3.24	(2.14, 4.33)	3.80	(3.18, 4.83)	17
500	2.41	(-0.09, 4.92)	4.09	(3.30, 5.44)	69	500	3.43	(2.04, 4.82)	4.19	(3.44, 5.48)	22
1000	2.52	(-0.61, 5.65)	4.49	(3.55, 6.17)	78	1000	3.62	(1.90, 5.34)	4.61	(3.71, 6.21)	27

(c) Station 3248: Vieux Québec, QC						(d) Station 3250: Lauzon, QC					
RP	MLE	HBM	Cred Int	%RC		RP	MLE	HBM	Cred Int	%RC	
10	—	2.13	(1.82, 2.58)	—		10	—	2.26	(1.95, 2.69)	—	
25	—	2.50	(2.09, 3.11)	—		25	—	2.63	(2.22, 3.24)	—	
50	—	2.80	(2.31, 3.56)	—		50	—	2.94	(2.43, 3.70)	—	
100	—	3.11	(2.54, 4.05)	—		100	—	3.26	(2.65, 4.20)	—	
250	—	3.57	(2.84, 4.79)	—		250	—	3.72	(2.96, 4.98)	—	
500	—	3.94	(3.08, 5.41)	—		500	—	4.10	(3.20, 5.65)	—	
1000	—	4.34	(3.33, 6.13)	—		1000	—	4.51	(3.46, 6.39)	—	

(e) Station 3200: Berthier-sur-Mer, QC						(f) Station 3260: Saint-Romuald, QC					
RP	MLE	HBM	Cred Int	%RC		RP	MLE	HBM	Cred Int	%RC	
10	—	—	—	—		10	—	—	—	—	
25	—	—	—	—		25	—	—	—	—	
50	—	—	—	—		50	—	—	—	—	
100	—	—	—	—		100	—	—	—	—	
250	—	—	—	—		250	—	—	—	—	
500	—	—	—	—		500	—	—	—	—	
1000	—	—	—	—		1000	—	—	—	—	

was fit. See, e.g., stations 3200 and 3260 at Berthier-sur-Mer and Saint-Romuald, QC, respectively. The methods used to accomplish this are described in Section 4.3.6.

Finally, Table 4.4 reports the return level for return periods $T \in \{10, 25, 50, 100, 250, 500, 1000\}$ (in years) computed at 10 stations using the joint distribution of surges from the HBM. For each parameter realization from the posterior distribution, Algorithm 6 was used to generate a sample of 10,000 surges at each of the 21 monitored stations. The empirical quantiles corresponding to each return period were then computed for each station individually. While consistent estimation of the multivariate quantiles defined in Beck and Mailhot (2018) would also be possible, it would be computationally prohibitive in such high dimension.

4.3 Flood risk in Atlantic Canada

Table 4.4: Return level for various return periods (RP, in years) computed at 10 stations using the joint distribution of surges from the Hierarchical Bayesian Model (HBM). The names of the stations are specified by ID in Table A.

RP	490	491	665	905	2985	3100	3248	3250	3200	3260
10	1.18	1.29	0.85	0.87	1.25	2.17	2.26	2.33	2.20	2.28
25	1.47	1.61	0.96	1.01	1.46	2.58	2.66	2.77	2.63	2.68
50	1.70	1.85	1.06	1.11	1.66	2.92	2.99	3.12	2.98	3.01
100	1.92	2.14	1.17	1.23	1.87	3.27	3.33	3.50	3.44	3.44
250	2.28	2.59	1.37	1.40	2.15	3.91	3.90	4.09	4.04	3.93
500	2.66	3.10	1.52	1.61	2.46	4.35	4.34	4.64	4.52	4.32
1000	3.27	3.87	1.63	1.89	2.62	4.82	4.78	4.88	5.29	4.89

Comparing the results in Tables 4.3 and 4.4, we can see once again the effect of sharing information across stations spatially. Closer stations tend to have much closer return periods than in the case when independent MLEs are fitted. Moreover, comparing the empirical results to those produced by the fitted margins, it can be seen that the returns are generally slightly higher. This could be a result of the samples used for Table 4.4 first being generated using the fitted copula, whereas the effect of the copula is only felt through its effect on the model fit in the HBM case.

4.3.6 Application

Here we present various applications of fitted model. In particular, we demonstrate

- (i) how to use the latent Gaussian fields to provide GEV parameter estimates at unmonitored locations,
- (ii) how the model can easily generate realizations of extreme surges from the collection of GEVs both at monitored and unmonitored locations, and
- (iii) how to use the generated surges to recover (potentially) extreme water level realizations.

Using the model for interpolation purposes

As discussed in Section 4.1, one of the main interests in geostatistics application is providing inference at locations where no data are recorded. Using the latent Gaussian fields in (4.8) and a simple algorithm, GEVs were fitted to 1340 unmonitored locations along the Atlantic Coast. The procedure is described in Algorithm 5 for interpolating the location parameters across the domain. A similar procedure was used for interpolating the log-scale parameters. To describe this procedure, let $\mathcal{S} = \{s_1, \dots, s_{21}\}$ be the set of locations at which data were recorded.

The results of Algorithm 5 are illustrated in Figures 4.10 and 4.11 for the medians of the posterior distribution for location and scale parameters of the original 21 stations. Moreover, the

4.3 Flood risk in Atlantic Canada

5 Spatial interpolation algorithm.

(1) Denote $\mathcal{S}^* = \{s_{22}, \dots, s_d\}$ the set of new locations of interest. The joint distribution for the expanded vector of means $\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*}$ including the new locations then takes the form $\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*} \sim \mathcal{N}_d(\mathbf{X}_{\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*}} \boldsymbol{\beta}_{\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*}}, \tau_{\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*}}^2 \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*}})$, where $\mathbf{X}_{\boldsymbol{\mu}_{\mathcal{S} \cup \mathcal{S}^*}} = (\mathbf{X}_{\boldsymbol{\mu}_{\mathcal{S}}}, \mathbf{X}_{\boldsymbol{\mu}_{\mathcal{S}^*}})^\top$ and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}^*} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}^*}} \end{bmatrix},$$

where for all $i, j \in \{1, \dots, d\}$, the (i, j) th entry of $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ is of the form $\exp(-d_{ij}/\rho_{\boldsymbol{\mu}})$.

(2) Conditioning on a realization of the location parameters generated from the model, denoted $\boldsymbol{\mu}_{\mathcal{S}} = \mathbf{m}$, the remaining location parameters can be generated from the conditional Gaussian field $\boldsymbol{\mu}_{\mathcal{S}^*} | \boldsymbol{\mu}_{\mathcal{S}} = \mathbf{m} \sim \mathcal{N}_{d-21}(\bar{\boldsymbol{\mu}}_{\mathcal{S}^*}, \tau_{\boldsymbol{\mu}_{\mathcal{S}^*}}^2 \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}_{\mathcal{S}^*}})$, where

$$\bar{\boldsymbol{\mu}}_{\mathcal{S}^*} = \mathbf{X}_{\boldsymbol{\mu}_{\mathcal{S}^*}} \boldsymbol{\beta}_{\boldsymbol{\mu}_{\mathcal{S}^*}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}^*} \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}}^{-1} (\mathbf{m} - \mathbf{X}_{\boldsymbol{\mu}_{\mathcal{S}}} \boldsymbol{\beta}_{\boldsymbol{\mu}_{\mathcal{S}}}) \quad \text{and} \quad \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}_{\mathcal{S}^*}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}^*}, \mathcal{S}^*} - \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}^*} \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}}, \mathcal{S}^*}.$$

maps in Figure 4.12 show the values of the standard deviation of the posterior distribution for the location (upper panel) and scale (lower panel) parameters in these 1361 locations.

Finally, Figure 4.13 shows maps of the 97.5% (upper panel) and 99.9% (lower panel) quantiles of the posterior distribution for the surge heights at the 1361 locations of interest along the Atlantic Coast of Canada. It transpires from these plots that the St. Lawrence River is most susceptible to large surges whereas the Eastern Coast of Newfoundland seems to be the least at risk. Observe also how the standard deviations displayed in Figure 4.12 are smaller in areas surrounding the 21 original monitoring stations, reflecting the smaller uncertainty associated with interpolation near locations where data are available.

Simulation of the fitted surge process

Once fitting and the subsequent interpolation are complete, we have at our disposal a mapping of 1361 interconnected GEV distributions corresponding to locations scattered across the Atlantic Coast of Canada. For the most part, these distributions were obtained without the benefit of data being recorded at their respective locations. With the entire domain of interest modeled, the surge process can be simulated. In particular, given the structure of the proposed model, sample generation is very simple and straightforward. The procedure is outlined in Algorithm 6.

With the generated samples, we have 10,000 realizations of maximal annual surges in all locations of interest. As there is no temporal trend in the surges, they can be interpreted as 10,000 realizations of a specific year, or as realizations for the next 10,000 years. Both will have their uses in applications. However, another step is required first.

4.3 Flood risk in Atlantic Canada

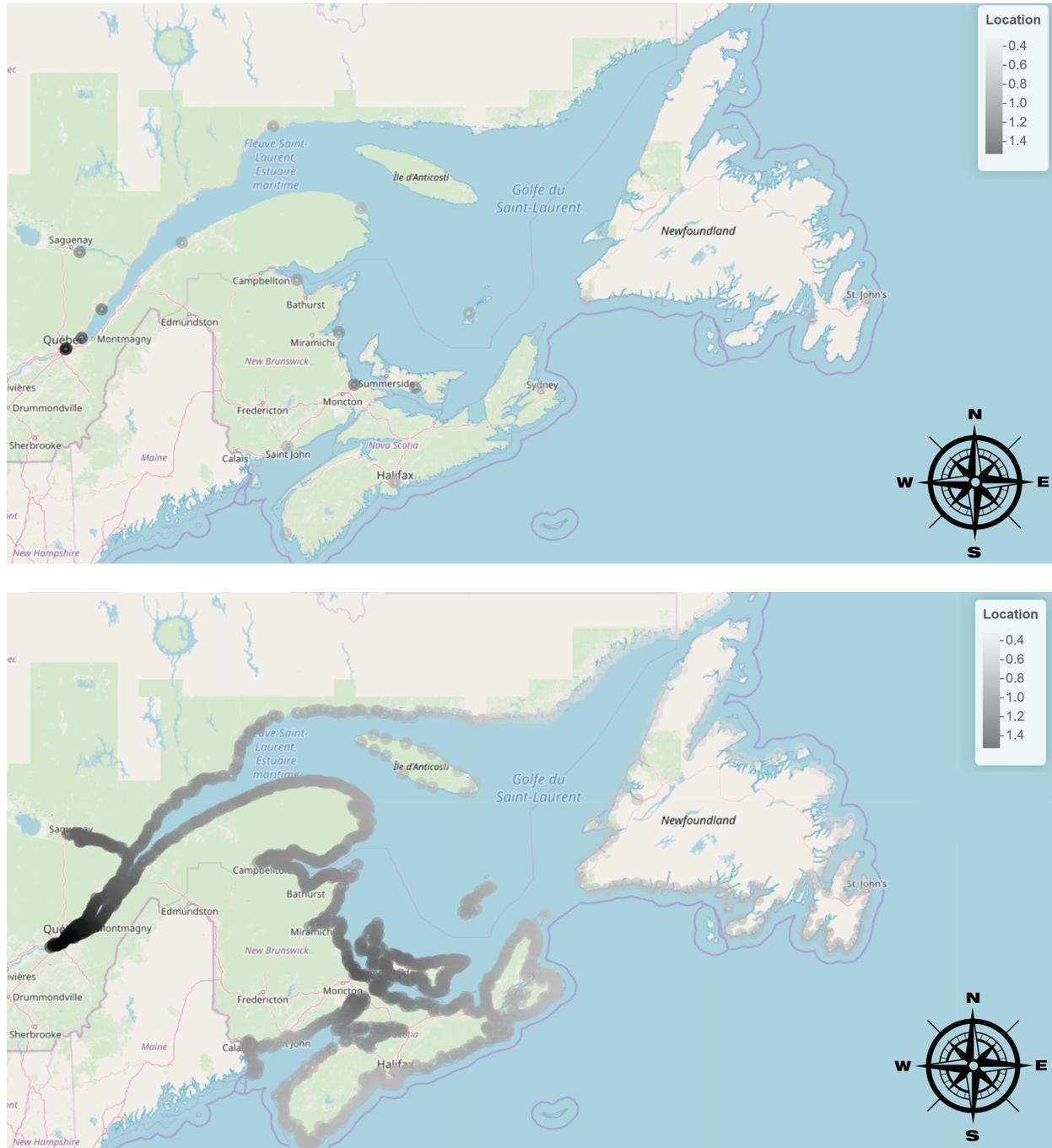


Figure 4.10: Maps showing the median values of the posterior distribution for the location parameters of the hierarchical Bayesian model at the monitoring stations (upper panel) and in 1361 locations along the Atlantic coast of Canada (lower panel).

4.3 Flood risk in Atlantic Canada

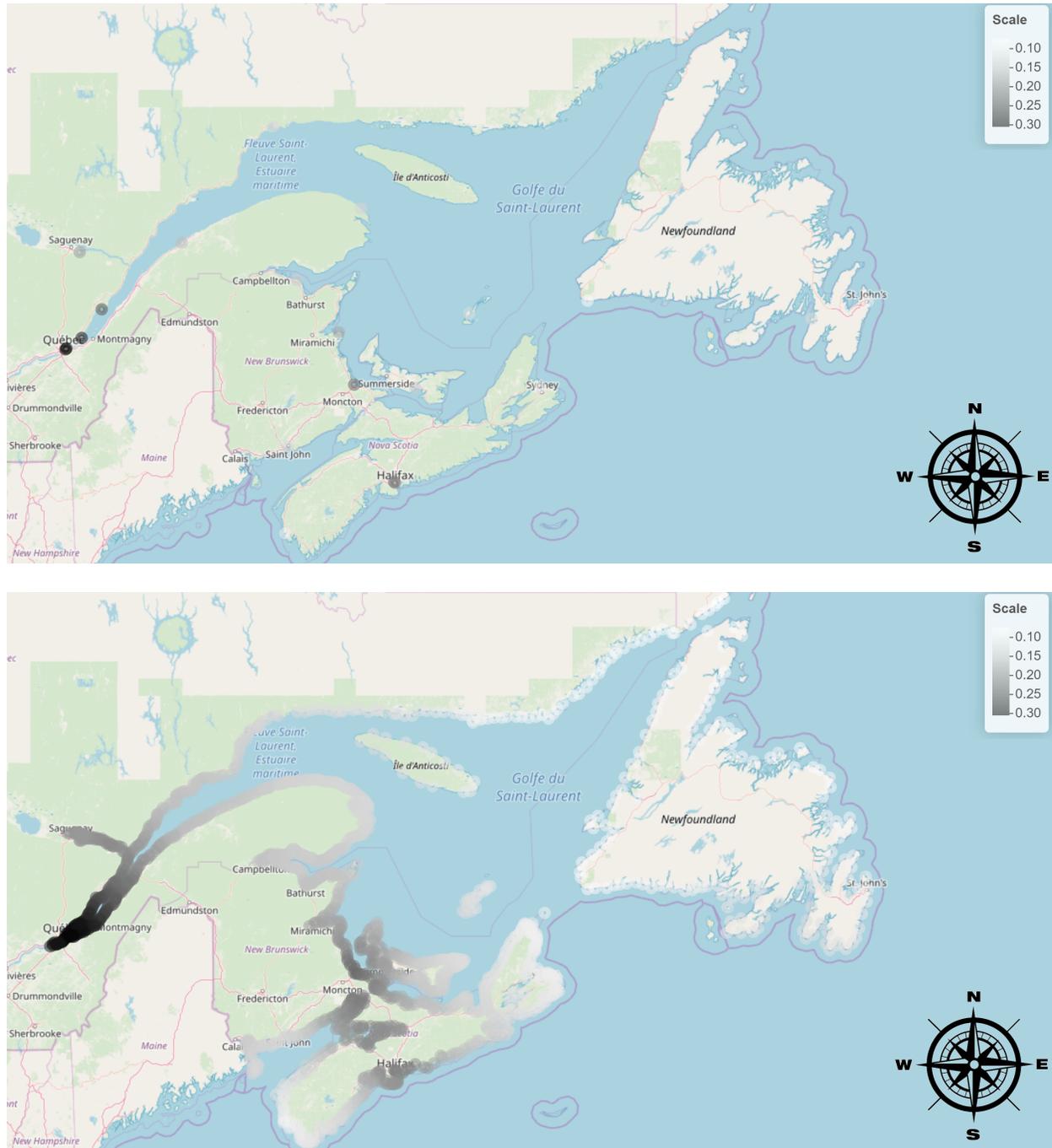


Figure 4.11: Maps showing the median values of the posterior distribution for the scale parameters of the hierarchical Bayesian model at the monitoring stations (upper panel) and in 1361 locations along the Atlantic coast of Canada (lower panel).

Predicting water levels from extreme surges along the coast line

While the fitted model provides information about the frequency and severity of extremal surges, the time at which they occur is critical to determining the corresponding water level, i.e., the combination

4.3 Flood risk in Atlantic Canada

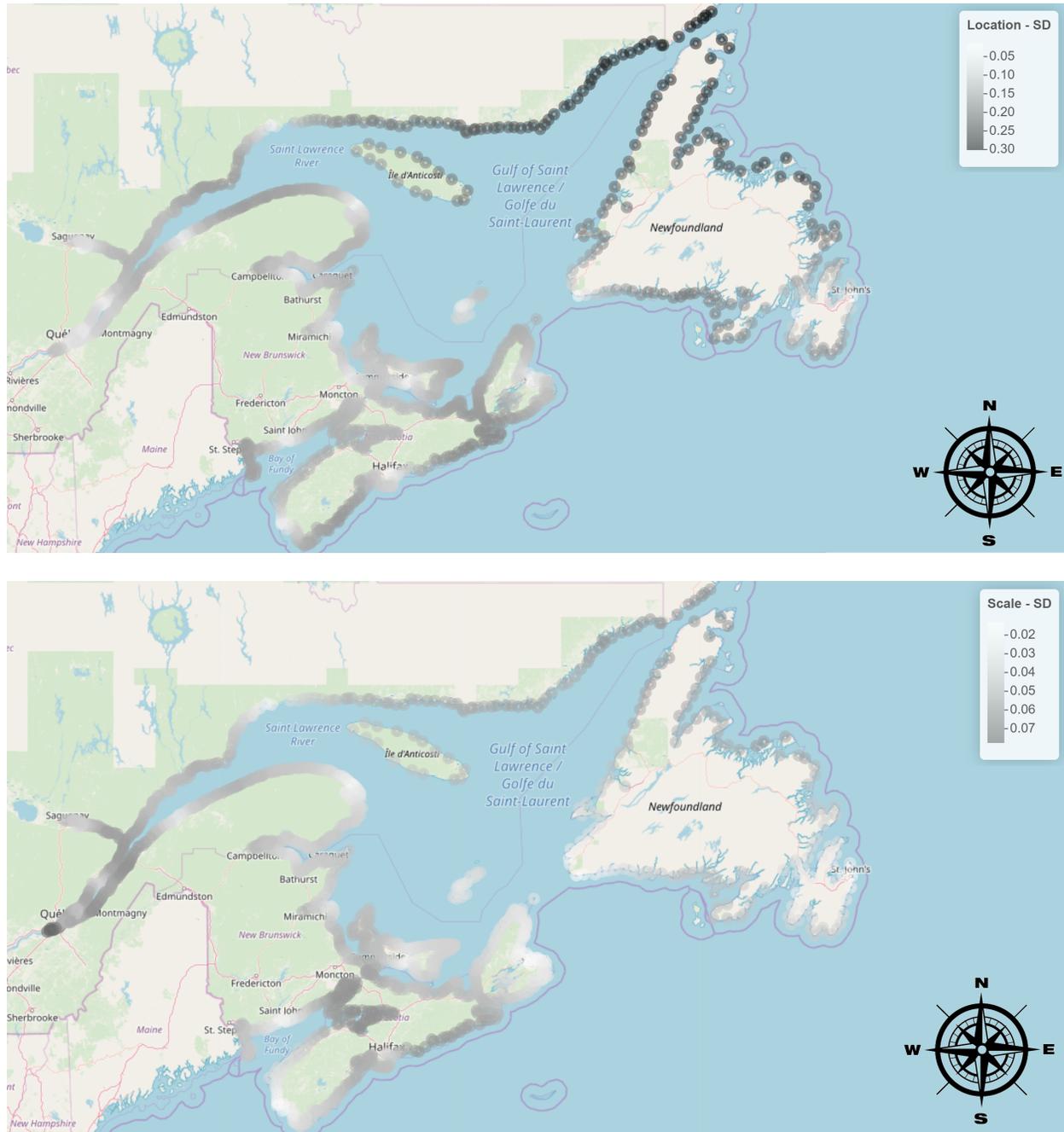


Figure 4.12: Maps showing the values of the standard deviation of the posterior distribution for the location (upper panel) and scale (lower panel) parameters in 1361 locations along the Atlantic Coast of Canada.

of tide and surge, and the associated risk of flooding. Consider for instance Station 3250 at Lauzon, across the river from Québec City. The range of tide at this location is 6.2 m with a minimum and maximum tide over the 1966–2015 period being -0.12 m and 6.10 m, respectively. Were the

4.3 Flood risk in Atlantic Canada

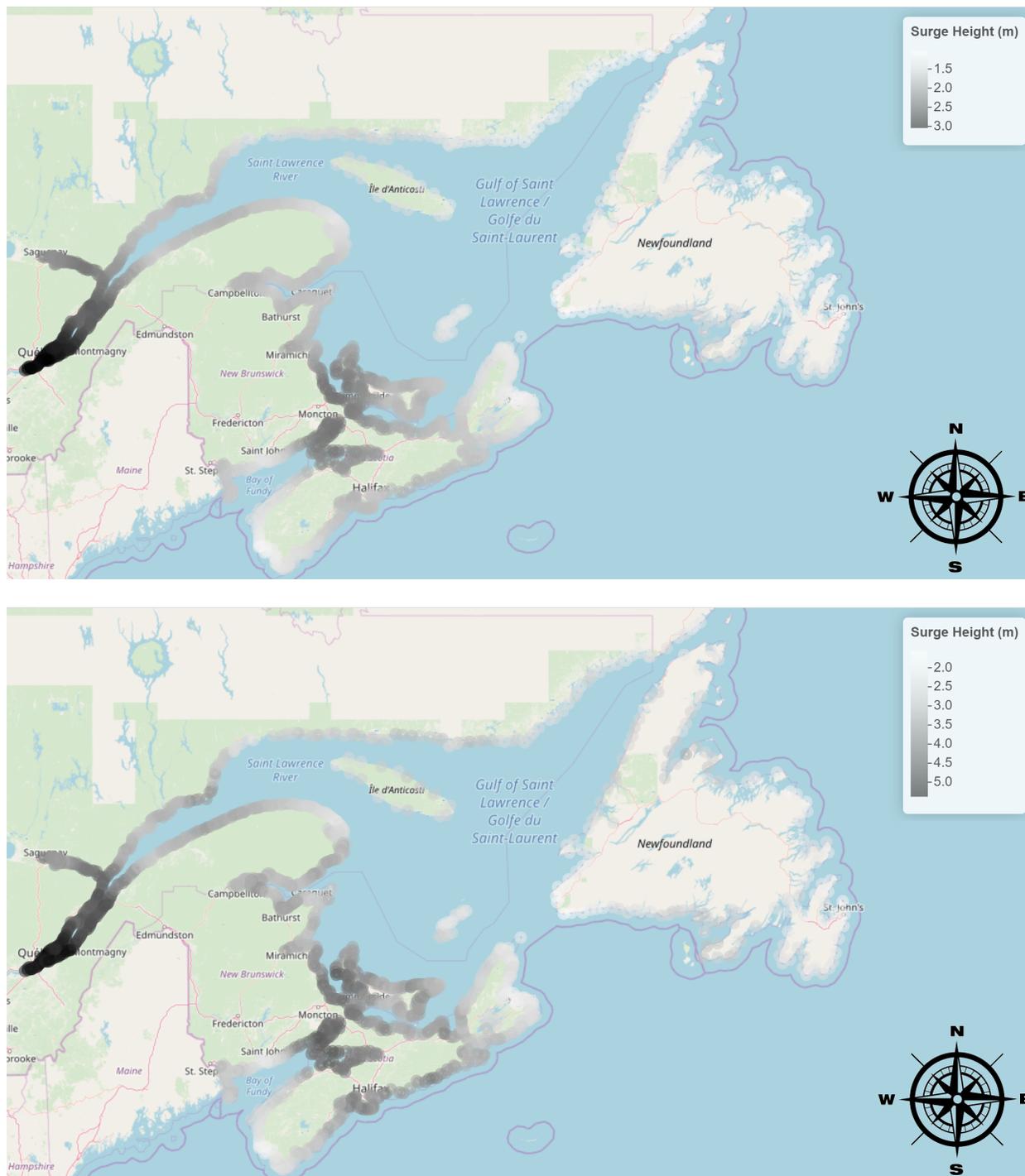


Figure 4.13: Maps showing the 97.5% (upper panel) and 99.9% (lower panel) quantiles of the posterior distribution for the surge heights at 1361 locations along the Atlantic Coast of Canada.

1-in-1000-year extreme annual surge to occur at low tide, the resulting water-level would only be about 3 m, i.e., half the height of the normal deterministic high-tide. Such a surge would be very

4.3 Flood risk in Atlantic Canada

6 Simulation algorithm.

- (1) Consider a realization $\Delta = (\Theta, \eta, \xi)$ of the parameters from the posterior distribution.
 - (2) Using Algorithm 5, interpolate the location parameters at the unmonitored stations to obtain $\mu_{\mathcal{S}^*}$ and $\sigma_{\mathcal{S}^*}$.
 - (3) Generate a realization from the copula $C_{\nu, \Xi}$, denoted $\mathbf{u} = (u_1, \dots, u_d)$.
 - (4) Invert \mathbf{u} using the relationship $z_i = F^{-1}(u_i | \mu_i, \sigma_i, \xi)$ for all $i \in \{1, \dots, d\}$. This yields a single realization $\mathbf{z} = (z_1, \dots, z_d)$.
 - (5) Repeat as necessary, once for each of the n realizations from the posterior distribution.
-

unlikely to be of any consequence to the people of Lauzon. In principle, therefore, it is important to consider when surges occur.

To assess the effect of the simultaneity (or not) of an extremal surge and an extremal tide, the following two scenarios were considered at three stations, namely 65 (St. John, NB), 490 (Halifax, NS), and 2985 (Rimouski, QC):

- (i) random-case scenario: the maximal surges occur randomly throughout the tidal process;
- (ii) worst-case scenario: the maximal surges occur at the maximum of the tide distribution.

For each station, 10,000 samples were generated from the fitted Hierarchical Bayesian model. Each of these surges was then turned into a water level by adding to it a tide observed at that location. For scenario (i), the 10,000 tides were selected randomly; for scenario (ii) all tides were taken to be equal to the largest possible value observed at the given station over the entire study period.

Based on our analysis, scenario (i) is the more realistic given that at the 5% level, a test of independence based on the Kendall's tau statistic could only detect dependence between the annual maximum surges and their corresponding tides at one location, namely station 3250 (Lauzon, QC). Nevertheless, scenario (ii) is interesting to consider as it provides a global upper bound on water levels across the entire domain. The simultaneous flooding resulting from these levels would be relevant in assessing the worst-case scenario of damages for which insurance companies might be liable.

The boxplots in Figure 4.14 summarize the water level distribution for each of these locations and scenarios. For stations 65 and 2985, the distinction between scenarios (i) and (ii) is easily noticeable, both in terms of median level and spread. Clearly, the coincidence of a maximum surge with a maximum tide would have disastrous consequences at those locations, and the proposed model is helpful in assessing its extent. By contrast, the change of scenario appears to be less critical on the water levels at station 490, e.g., at the lower tail of the distribution. Nevertheless, the occurrence of such high water levels could have devastating effects, depending on the location.

4.3 Flood risk in Atlantic Canada

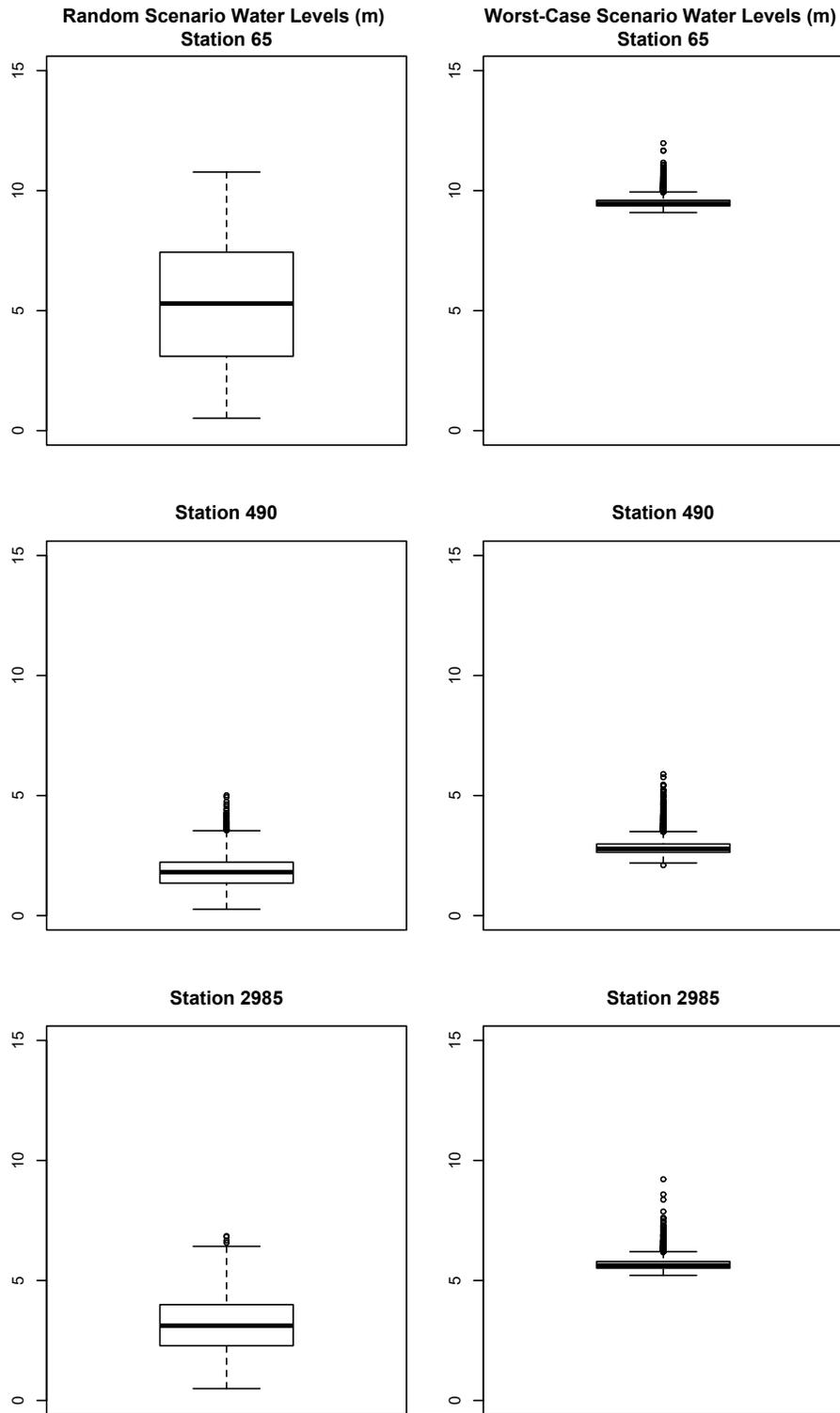


Figure 4.14: Boxplots representing the distribution of water levels at three locations under scenarios (a) and (b), based on 10,000 realizations of the surge model.

4.3.7 Discussion

In Section 4.3.6, the generation of the maximal surges was completed using the full Bayesian hierarchical model, taking in the dependence at both the data and process-level. In contrast, the generation of the corresponding tides was done on a site-by-site basis. As there is likely some spatial interaction across tidal processes, inference could be improved by developing an analogous model for this phenomenon. Another possible extension of the approach would be to take temporal dependence into account. Given that the maximum surge process appears to be stationary in time and independent of tide, this would involve including a temporal component in the tidal process.

Another consideration is the computational efficiency of the model. While the HBM is quite flexible and can easily incorporate residual dependence through the inclusion of a copula, operations on large correlation matrices can be cumbersome. In the future it would be interesting to evaluate the risk of maximal surges using Integrated Nested Laplacian Approximations (INLA) as described, e.g., by [Rue et al. \(2009\)](#) and [Blangiardo et al. \(2013\)](#). A second potential modeling approach would be the use of intrinsic Gaussian Markov random fields (iGMRF) as described in [Paciorek \(2013\)](#) and applied in [Jalbert et al. \(2017\)](#). Methods such as these would allow one to fit GEVs to both the observed and unobserved locations simultaneously as opposed to sequentially. Moreover, the computation time needed to generate samples from the posterior distributions would be vastly decreased by employing sparse precision matrices. Inclusion of a copula in either approach might also be considered. The risk of coastal flooding in Atlantic Canada could then be compared using the HBM, INLA methods and iGMRF fitting.

Finally, an important factor in the modeling process is how we quantify the dependence between stations. Given the physical nature of the problem, the great circle distance was a natural choice for this purpose. Unfortunately, while appropriate, we believe that only considering this notion of dissimilarity might limit the inference potential of the model as it ignores some of the geographical properties of the domain we study. As we are dealing with locations within a complex body of water, it would make sense to consider distances which factor in this irregular geography. However, one must be careful when considering such non-Euclidean distance measures. If not, the resulting covariance matrix might not possess positive definiteness.

Several works have addressed such issues, e.g., see [Ver Hoef and Peterson \(2010\)](#), [Asadi et al. \(2015\)](#) or [Ver Hoef et al. \(2006\)](#) for a discussion on the notion of river/stream, or hydrological, distance. However, these works rely on the assumption that the network is made up of (width-less) lines, essentially creating a projection into \mathbb{R} , a luxury the irregular shape of Atlantic Canada does not harmonize with.

A more general framework has been developed in [Curriero \(2006\)](#) to explain the viability of alternative distance metrics. Here, situations in which non-Euclidean distance matrix can be

4.4 Conclusion

considered in a geostatistics framework are described. The key concept is *isometric embedding*. Consider a distance metric $d(\cdot, \cdot)$ for a spatial domain \mathcal{D} . If there exists points $\mathbf{s}_i^*, \mathbf{s}_j^*$ in a spatial domain \mathcal{D}^* and a function $\eta : \mathcal{D} \mapsto \mathcal{D}^*$ such that

$$d(\mathbf{s}_i, \mathbf{s}_j) = \|\mathbf{s}_i^* - \mathbf{s}_j^*\| \quad (4.9)$$

for all $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{D}$ and $\eta(\mathbf{s}) = \mathbf{s}^*$ then d is said to be embeddable in a Euclidean space of dimension $d^* = \dim(\mathcal{D}^*)$. If a distance metric is embeddable, then the corresponding covariance matrix will be positive-definite. A more flexible technique for addressing non-Euclidean distance matrices is through the use of multidimensional scaling (MDS). See for instance [Bibby et al. \(1979\)](#), [Løland and Høst \(2003\)](#) or [Schabenberger and Gotway \(2017\)](#). In short, MDS creates locations (in an augmented space) whose Euclidean distances approximate the original distance matrix, providing a type of approximation to (4.9).

Using these concepts, it would be interesting to consider different distance metrics for the process and data layers of the model. For example, if the correlation structure associated to the residual dependence structure could consider some type of water-based distance between stations, this could provide additional insight into the physical properties of the bodies of water in Atlantic Canada and could further accentuate the interplay between the large- and small-scale variation described by these two layers.

4.4 Conclusion

In this chapter, we have provided an overview of modeling spatial extremes with applications towards natural disasters. In particular, in Section 4.3, we used several of these techniques to model the prevalence of extreme surges on the Atlantic Coast of Canada. Model fitting over this large and sparse domain was done using a well established HBM on GEV margins. Latent Gaussian fields were used at the process level to introduce spatiality into the model. Having the location and (log) scale parameters vary spatially, as opposed to just location, substantially improved model performance. Moreover, the use of a copula allowed for possible dependence between the annual maximum surges.

The employed model also allowed us to generate simultaneous realizations of annual maximum surges at a large collection of both observed and unobserved locations of interest. Two different scenarios for calculating water levels resulting from extremal surges could then be computed. Quantitative information about the water levels across the entire domain under such scenarios could be helpful from an infrastructure or insurance perspective. Moreover, getting a sense of when these catastrophes might occur relative to each other would also be useful for organizations in charge of public safety, e.g., in terms of resource allocation for disaster response.

4.4 Conclusion

Turning our focus to this work's initial motivation, i.e., the assessment of overland flooding risk for insurance regulatory purposes, the next step would be to quantify the damages caused inland as a result of extreme water levels. One way to ascertain this information would be through the use of water propagation models included in, e.g., Geographic Information Systems (GIS) software. Using the coastal water levels, one could then calculate the water propagation inland and use damage curves to evaluate the resulting costs at each site of the entire domain for a given year. From here, the problem would be simplified to a univariate one and assessing risk frequency would be straightforward.

More specifically, consider realizations $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ of the d -dimensional surge process \mathbf{Z} . For each year $i \in \{1, \dots, n\}$, find the vector \mathbf{W}_i of water levels corresponding to \mathbf{Z}_i . Through GIS, inundation levels $\mathbf{f}_i = (f_{i1}, \dots, f_{id_\ell})$ can be obtained at d_ℓ locations of interest. Using damage curves, the total cost $\mathbf{D}_i = (D_{i1}, \dots, D_{id_\ell})$ can then be determined. Thus, for each year $i \in \{1, \dots, n\}$, we have

$$\mathbf{Z}_i \xrightarrow{\text{Find Tide}} \mathbf{W}_i \xrightarrow{\text{GIS}} \mathbf{f}_i \xrightarrow{\text{Damage Curves}} \mathbf{D}_i.$$

The total cost across the domain in year i is then $S_i = D_{i1} + \dots + D_{id_\ell}$. Using these total damages will drastically simplify the ensuing analyses as the remaining calculations will be done in a univariate context. With these values, the amount of capital needed to protect an insurer at a prescribed solvency level can be easily determined. Moreover, if we wish to consider the individual flood locations, we can better assess the specific risk, and subsequent premium necessary, for each location. A method for working with such random sums is described in [Mailhot and Mesfioui \(2016\)](#). In a multivariate context, it would also be interesting to consider how the Σ -expectiles (3.5) introduced in Chapter 3 could be used on either of the vectors \mathbf{f} or \mathbf{D} , in particular, if one were to use a distance based correlation matrix. Some examples of applications of multivariate risk measures applied to (extreme) environmental phenomena can be found in [Torres et al. \(2017\)](#), [Di Bernardino and Palacios-Rodríguez \(2017\)](#), and [Di Bernardino and Prieur \(2018\)](#).

Appendix

4.A List of Stations by ID

Table A: List of monitored stations and their locations, along with the number of annual maxima available at each site.

ID	Station Name	#	ID	Station Name	#
65	St. John, NB	49	2000	Lower Escuminac, NB	40
365	Yarmouth, NS	48	2145	Belledune, NB	15
490	Halifax, NS	48	2330	Rivière-au-Renard, QC	43
491	Bedford Institute, NS	17	2780	Sept-Îles, QC	43
612	North Sydney, NS	46	2985	Rimouski, QC	31
665	Port-aux-Basques, NL	37	3057	Saint-Joseph-de-la-Rive, QC	44
835	Argentia, NL	45	3100	Saint-François de l'Île d'Orléans, QC	49
905	St. John's, NL	50	3248	Vieux Québec, QC	6
1700	Charlottetown, PE	47	3250	Lauzon, QC	45
1805	Shediac Bay, NB	32	3460	Port-Alfred (Saguenay), QC	32
1970	Cap-aux-Meules, QC	18			

4.B Justification of 15% rule for including maxima

This appendix provides a rationale for keeping an annual maximum at a given site only when at most 15% of the hourly measurements are missing, and to treat it as the true maximum. Because the tidal process is a deterministic phenomenon which is available throughout the study period, one can restrict the discussion to the stationary series X_1, \dots, X_n of hourly surges for a given site in a given year. For a non-leap year, $n = 8760$.

Suppose for simplicity that the observations are exchangeable and hence, without further loss of generality, that one has observed only X_1, \dots, X_k with $k \geq 0.85 \times n$. For a non-leap year, one then has $k \geq 7446$ observations. If $M_k = \max(X_1, \dots, X_k)$, then $M_k \geq X_{(k)}$, where $X_{(1)} < \dots < X_{(n)}$ are the order statistics associated with the series X_1, \dots, X_n , which are almost certainly distinct if the base distribution is continuous.

A simple combinatorial argument then implies that, for all $\ell \in \{k, \dots, n\}$,

$$\Pr\{M_k = X_{(\ell)}\} = p_\ell \equiv \binom{\ell-1}{k-1} / \binom{n}{k}.$$

This probability is non-decreasing in ℓ so that if $k \geq 0.85n$, one has $\Pr\{M_k = X_{(n)}\} = k/n \geq 0.85$. Therefore, the 15% rule guarantees that the probability that the observed maximum is the true one is at least 85%.

4.B Justification of 15% rule for including maxima

As mentioned in Section 2.1, the distribution of $X_{(n)}$ can be suitably approximated by a GEV distribution when n is large enough. Under the additional assumption that X_1, \dots, X_n form a random sample, Theorem 3.4 in Coles (2001) states that, as $n \rightarrow \infty$, and for sufficiently large integer $\ell \leq n$,

$$\Pr\{X_{(\ell)} \leq x\} = H_\ell(x) \equiv \exp \left[- \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi} \right] \sum_{s=0}^{n-\ell} \{1 + \xi(x - \mu)/\sigma\}^{-s/\xi} / s!,$$

whenever $\xi > 0$ and $1 + \xi(x - \mu)/\sigma > 0$. When $\ell = n$, one recovers the expression of the GEV distribution given in Algorithm 4. This is the distribution of M_k when $M_k = X_{(n)}$. More generally, the asymptotic distribution of M_k is the mixture $p_k H_k + \dots + p_n H_n$ with the weights p_k, \dots, p_n introduced above. This mixture has the same three location, scale, and shape parameter as the GEV distribution.

Given that for $n = 8760$ and $k = 7446$, one has $p_{n-1} = k(n - k)/(n^2 - n) \approx 0.128$, $p_{n-2} \approx 0.019$, and $p_{n-3} \approx 0.003$, it seems reasonable to approximate the distribution of M_k by H_n . This comes at the cost of a small bias. A more accurate approximation would be given by $(p_{n-1}H_{n-1} + p_n H_n)/(p_{n-1} + p_n) = (0.128H_{n-1} + 0.85H_n)/0.978$ but this refinement was not considered here because of the presence of autocorrelation in the hourly (and even in the daily) measurement series X_1, \dots, X_n . From the results reported in Leadbetter et al. (1983), this dependence is likely such that the large-sample distribution of $X_{(n)}$ can still be approximated by a GEV with the same shape parameter as before but it is no longer clear that the asymptotic distribution of M_k is a mixture. Only the fact that $M_k = X_{(n)}$ with probability 0.85 remains.

While other values than $k/n = 85\%$ could have been considered, this choice seemed adequate on balance. On one hand, any significant reduction in k/n would imperil the approximation of the distribution of M_k by H_n , even in the absence of autocorrelation. On the other hand, too many years would have been discarded if k/n had been made larger, causing a large increase in uncertainty.

4.C Effect on shape parameter given different cutoffs for data augmenting

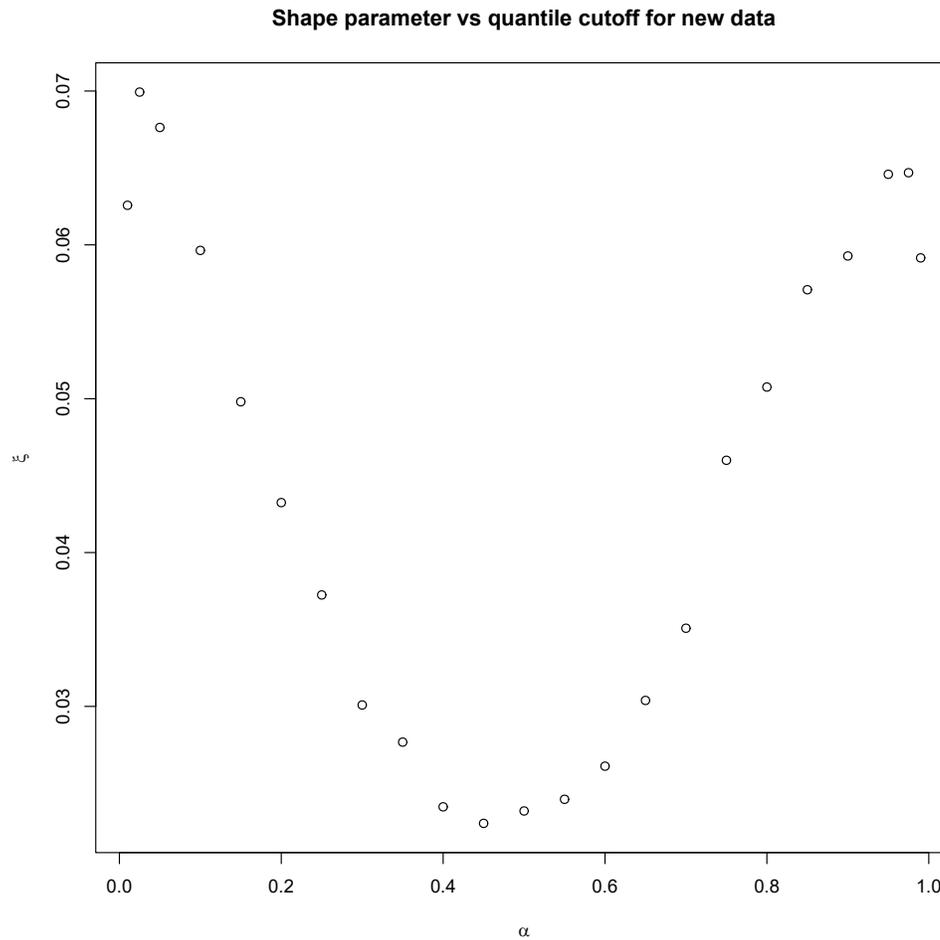


Figure 4.15: Variability in the fitted constrained shape parameter for different quantiles q_α acting as the cutoff for new data.

5

Final Conclusion and Future Work

The contribution of this thesis lies at the intersection of extreme value theory, risk analysis, and environmental statistics. Extreme value theory is an important branch of statistics which has a plethora of applications to real-world problems. Specifically, we have examined the behavior of extreme events through two separate lenses: risk measure estimation and statistical modelling. Risk measures are a fundamental tool for quantifying the risk of a variety of processes, including stocks, insurance claims, and water levels. Moreover, modeling these events can provide valuable information about their behavior. Both of these projects were considered in the multivariate framework where dependence between random variables, especially for tail events, is a key factor.

In the first case, we developed an estimation procedure for multivariate L^1 -expectiles when the level of risk $\alpha \approx 1$. In particular, we followed a semi-parametric approach that, under the assumption that the tails of each margin behaved similarly, consistently estimated extreme L^1 -expectiles. Furthermore, the methodology presented here can account for any underlying tail dependence structure. This poses a tangible improvement over prior methods in the literature where classic methods performed poorly for elevated levels of risk and methods tailored to extremes were limited by pre-imposed assumptions of asymptotic independence or comonotonicity between random variables. Using the BFGS algorithm for unconstrained optimization, the results of our simulation study showed satisfactory performance for a wide range of sample sizes.

In the future, it would be interesting to consider alternative optimization algorithms. For instance, L-BFGS-B is a natural choice as it provides an extension of BFGS which can reduce computation time by limiting matrix operations and improve efficiency by incorporating simple box constraints. This latter modification would allow us to utilize some of the properties of the underlying loss function. Considering stochastic optimization, as was done in [Maume-Deschamps et al. \(2017\)](#), for the alternative problem would present another interesting avenue. A further consideration in this project is that we focused on L^1 -expectiles, which are in fact a special case of Σ -expectiles. It would be interesting to examine the impact that different matrices Σ could impart on the estimation

Final Conclusion and Future Work

procedure, in particular for real-world examples. For instance, treating Σ as a distance matrix for spatially defined problems could allow the risk measure to naturally incorporate the properties of the region of study. It would be interesting to see how this measure could be used for our study on Atlantic water surges.

In the aforementioned study of Atlantic Canada, we modelled annual maximum surges at 21 locations in a large, irregular domain using a hierarchical Bayesian model with GEV margins. The model was then fitted with a variety of residual dependence structures. In the end, a Student t copula with 9 degrees of freedom was considered the best choice. This model provides a sense of continuity throughout the entirety of the domain, a factor which has been stressed recently (Towe et al., 2018; Tawn et al., 2018) while still allowing for the existence of tail dependence between extreme events.

Using the simulations generated from the model, it would be interesting to see the impacts that these events would have inland. One way to accomplish this would be to make use of GIS software which could provide physical simulations of inland water propagation given extreme water levels on the coast. These subsequent simulations could then be used to estimate inundation levels across communities which could in turn inform potential damages from extreme flooding events. The information from this process could be used at both a private and public level. Flood levels could help inform future iterations of flood plain mapping, thereby ensuring that citizens and businesses are housed in safe areas, whereas the estimated damages could aid insurance companies for premium pricing as well as capital allocation.

In fact, the research presented was inspired from a collaboration with a prominent Canadian insurance company, where a prior version of the work helped contribute to the launching of a comprehensive flood insurance product in Canada in March of 2017, a Canadian first. This product ensured that victims of the 2017 floods in Ontario and Québec were adequately covered, resulting in claims payouts totalling over \$220 million.

List of Publications

Published:

N. Beck, C. Genest, J. Jalbert & M. Mailhot (2020). Predicting extreme surges from sparse data using a copula-based hierarchical Bayesian spatial model. *Environmetrics*, e2616.

Submitted:

N. Beck, E. Di Bernardino, M. Mailhot (2020). Semi-parametric estimation of multivariate extreme expectiles. *Journal of Multivariate Analysis*, submitted.

Bibliography

- Acerbi, C. and Tasche, D. (2002a). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388.
- Acerbi, C. and Tasche, D. (2002b). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.
- Ahn, J. Y. and Shyamalkumar, N. D. (2011). Large sample behavior of the CTE and VaR estimators under importance sampling. *North American Actuarial Journal*, 15(3):393–416.
- Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). *Reinsurance: Actuarial and Statistical Aspects*. Wiley, Hoboken, NJ.
- Albrecher, H., Kortschak, D., and Prettenthaler, F. (2020). Spatial dependence modeling of flood risk using max-stable processes: The example of Austria. *Water*, 12(6):1805.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023–2050.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, New York.
- Basel Committee on Banking Supervision (2006). Basel II: International convergence of capital measurement and capital standards: a revised framework. <https://www.bis.org/publ/bcbs128.htm>. Last visited August 27th, 2020.
- Basel Committee on Banking Supervision (2017). Basel III: Finalising post-crisis reforms. <https://www.bis.org/bcbs/publ/d424.htm>. Last visited August 27th, 2020.
- Beck, N. and Mailhot, M. (2018). A consistent estimator to the orthant-based tail value-at-risk. *ESAIM: Probability and Statistics*, 22:163–177.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, UK.

BIBLIOGRAPHY

- Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, 15(5):725–733.
- Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506.
- Bibby, J., Kent, J., and Mardia, K. (1979). *Multivariate Analysis*. Academic Press, London.
- Billingsley, P. (2008). *Probability and Measure*. Wiley, New York.
- Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 4:33–49.
- Bowman, F. (1958). *Introduction to Bessel Functions*. Dover Publications Inc., New York.
- Bracken, C., Rajagopalan, B., Cheng, L., Kleiber, W., and Gangopadhyay, S. (2016). Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resources Research*, 52:6643–6655.
- Bresson, É., Laprise, R., Paquin, D., Thériault, J., and de Elía, R. (2017). Evaluating the ability of CRCM5 to simulate mixed precipitation. *Atmosphere-Ocean*, 55(2):79–93.
- Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.
- Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.
- Buhl, S. and Klüppelberg, C. (2019). Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes. *Extremes*, 22(2):223–269.
- Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.
- Cai, J.-J., Einmahl, J. H., Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: The mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):417–442.
- Casson, E. and Coles, S. (1999). Spatial regression models for extremes. *Extremes*, 1(4):449–468.
- Castruccio, S., Huser, R., and Genton, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, 25(4):1212–1229.

BIBLIOGRAPHY

- Chaouch, M. and Goga, C. (2010). Design-based estimation for geometric quantiles with application to outlier detection. *Computational Statistics & Data Analysis*, 54(10):2214–2229.
- Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.
- Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.
- Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606.
- Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.
- Cossette, H., Mailhot, M., Marceau, É., and Mesfioui, M. (2013). Bivariate lower and upper orthant value-at-risk. *European Actuarial Journal*, 3(2):321–357.
- Cossette, H., Mailhot, M., Marceau, E., and Mesfioui, M. (2016). Vector-valued tail value-at-risk and capital allocation. *Methodology and Computing in Applied Probability*, 18(3):653–674.
- Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of value-at-risk. *Journal of Multivariate Analysis*, 119:32–46.
- Cousin, A. and Di Bernardino, E. (2014). On multivariate extensions of conditional-tail-expectation. *Insurance: Mathematics and Economics*, 55:272–282.
- Cressie, N. A. (1993). *Statistics for Spatial Data*. Wiley, New York.
- Cuevas, A., González-Manteiga, W., and Rodríguez-Casal, A. (2006). Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19.
- Cuevas, A. and Rodríguez-Casal, A. (2004). On boundary estimation. *Advances in Applied Probability*, 36(2):340–354.
- Curriero, F. C. (2006). On the use of non-euclidean distance measures in geostatistics. *Mathematical Geology*, 38(8):907–926.
- Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 19(5B):2557–2589.

BIBLIOGRAPHY

- Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(2):263–292.
- Daouia, A., Girard, S., and Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli*, 25(1):264–309.
- Davison, A. C. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and its Application*, 2:203–235.
- Davison, A. C., Huser, R., and Thibaud, E. (2019). Spatial extremes. In *Handbook of Environmental and Ecological Statistics*, pages 711–744. CRC Press, Boca Raton, FL.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.*, 27(2):161–186.
- de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.
- de Haan, L. and Pereira, T. T. (2006). Spatial extremes: Models for the stationary case. *The Annals of Statistics*, 34(1):146–168.
- Deheuvels, P., Haeusler, E., and Mason, D. M. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, 104(2):371–381.
- Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, 73(1):111–129.
- Deprez, O. and Gerber, H. U. (1985). On convex principles of premium calculation. *Insurance: Mathematics and Economics*, 4(3):179–189.
- Dhaene, J., Laeven, R. J., Vanduffel, S., Darkiewicz, G., and Goovaerts, M. J. (2008). Can a coherent risk measure be too subadditive? *Journal of Risk and Insurance*, 75(2):365–386.
- Di Bernardino, E., Fernández-Ponce, J., Palacios-Rodríguez, F., and Rodríguez-Griñolo, M. (2015). On multivariate extensions of the conditional value-at-risk measure. *Insurance: Mathematics and Economics*, 61:1–16.
- Di Bernardino, E., Laloë, T., Maume-Deschamps, V., and Prieur, C. (2013). Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory. *ESAIM: Probability and Statistics*, 17:236–256.

BIBLIOGRAPHY

- Di Bernardino, E. and Palacios-Rodríguez, F. (2017). Estimation of extreme component-wise excess design realization: a hydrological application. *Stochastic Environmental Research and Risk Assessment*, 31(10):2675–2689.
- Di Bernardino, E. and Prieur, C. (2018). Estimation of the multivariate conditional tail expectation for extreme risk levels: Illustration on environmental data sets. *Environmetrics*, 29(7):e2510.
- Diggle, P. J. and Ribeiro, P. J. (2007). *Model-Based Geostatistics*. Springer, New York.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.
- Dupont, F., Hannah, C., Greenberg, D., Cherniawsky, J., and Naimie, C. (2002). *Modelling System for Tides for the Northwest Atlantic Coastal Ocean*, volume 221. Fisheries & Oceans Canada, Maritimes Region, Ocean Sciences Division, Coastal Ocean Science Section, Bedford Institute of Oceanography.
- Dupuis, D. J. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1(3):251–261.
- Egbert, G. D., Bennett, A. F., and Foreman, M. G. (1994). Topex/poseidon tides estimated using a global inverse model. *Journal of Geophysical Research: Oceans*, 99(C12):24821–24852.
- Egbert, G. D. and Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic technology*, 19(2):183–204.
- Einmahl, J. H., Kiriliouk, A., Krajina, A., and Segers, J. (2016). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1(78):275–298.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*. Springer, Berlin.
- Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, 97(2):526–547.
- European Commission (2016). Solvency II. https://ec.europa.eu/info/business-economy-euro/banking-and-finance/insurance-and-pensions/risk-management-and-supervision-insurance-companies-solvency-2_en. Last visited Septmeber 14th, 2020.

BIBLIOGRAPHY

- Fawcett, L. and Walshaw, D. (2006). A hierarchical model for extreme wind speeds. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(5):631–646.
- Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190.
- Fissler, T. and Ziegel, J. F. (2016). Higher order elicibility and Osband’s principle. *The Annals of Statistics*, 44(4):1680–1707.
- Fogarty, C. (2004). Hurricane Juan storm summary. http://www.novaweather.net/Hurricane_Juan_files/Juan_Summary.pdf. Last visited September 14, 2020.
- Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447.
- Frittelli, M. and Gianin, E. R. (2002). Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486.
- Frittelli, M. and Gianin, E. R. (2005). Law invariant convex risk measures. In *Advances in Mathematical Economics*, volume 7, pages 33–46. Springer, Tokyo.
- García, J., Martín, J., Naranjo, L., and Acero, F. (2018). A Bayesian hierarchical spatio-temporal model for extreme rainfall in extremadura (Spain). *Hydrological Sciences Journal*, 63(6):878–894.
- Gardes, L. and Girard, S. (2015). Nonparametric estimation of the conditional tail copula. *Journal of Multivariate Analysis*, 137:1–16.
- Gardes, L., Girard, S., and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2):419–433.
- Gardes, L. and Stupfler, G. (2019). An integrated functional Weissman estimator for conditional extreme quantiles. *REVSTAT Statistical Journal*, 17(1):109–144.
- Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press, Boca Raton, FL.

BIBLIOGRAPHY

- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.
- Genest, C., Favre, A.-C., Béliveau, J., and Jacques, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 43(9).
- Genest, C., Nešlehová, J. G., and Rémillard, B. (2014). On the empirical multilinear copula process for count data. *Bernoulli*, 20(3):1344–1371.
- Genest, C., Nešlehová, J. G., and Rémillard, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *Journal of Multivariate Analysis*, 159:82–110.
- Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369.
- Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199–213.
- Ghosh, S. and Mallick, B. K. (2011). A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22(2):192–204.
- Girard, S. and Stupfler, G. (2015). Extreme geometric quantiles in a multivariate regular variation framework. *Extremes*, 18(4):629–663.
- Girard, S. and Stupfler, G. (2017). Intriguing properties of extreme geometric quantiles. *REVSTAT Statistical Journal*, 15(1):107–139.
- Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2019). Nonparametric extreme conditional expectile estimation. working paper or preprint.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44(3):423–453.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.
- Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics [Editorial]. *Journal of Multivariate Analysis*, 146:1–6.

BIBLIOGRAPHY

- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- Herrmann, K., Hofert, M., and Mailhot, M. (2018). Multivariate geometric expectiles. *Scandinavian Actuarial Journal*, 2018(7):629–659.
- Herrmann, K., Hofert, M., and Mailhot, M. (2020). Multivariate geometric tail-and range-value-at-risk. *ASTIN Bulletin: The Journal of the IAA*, 50(1):265–292.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.
- Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.
- Hsing, T. (1989). Extreme value theory for multivariate stationary sequences. *Journal of Multivariate Analysis*, 29(2):274–291.
- Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518.
- Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103.
- Huser, R., Dombry, C., Ribatet, M., and Genton, M. G. (2019). Full likelihood inference for max-stable data. *Stat*, 8(1):e218.
- Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.
- Jalbert, J., Favre, A.-C., Bêlisle, C., and Angers, J.-F. (2017). A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(5):941–961.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press, London.
- Joe, H. (2014). *Dependence Modeling with Copulas*. CRC press, Boca Raton, FL.
- Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270.

BIBLIOGRAPHY

- Jones, B. L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, 7(4):44–54.
- Jouini, E., Meddeb, M., and Touzi, N. (2004). Vector-valued coherent risk measures. *Finance and Stochastics*, 8(4):531–552.
- Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065.
- Kiriliouk, A., Segers, J., and Tafakori, L. (2018). An estimator of the stable tail dependence function based on the empirical beta copula. *Extremes*, 21(4):581–600.
- Kou, S., Peng, X., and Heyde, C. C. (2013). External risk measures and basel accords. *Mathematics of Operations Research*, 38(3):393–417.
- Kusuoka, S. (2001). On law invariant coherent risk measures. In *Advances in Mathematical Economics*, volume 3, pages 83–95. Springer, Tokyo.
- Kuzmenko, V., Salam, R., and Uryasev, S. (2020). Checkerboard copula defined by sums of random variables. *Dependence Modeling*, 8(1):70–92.
- Lambert, N. S., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138.
- Lauer, A. and Zähle, H. (2016). Nonparametric estimation of risk measures of collective risks. *Statistics & Risk Modeling*, 32(2):89–102.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.
- Li, D.-H. and Fukushima, M. (2001a). A modified BFGS method and its global convergence in nonconvex minimization. *Journal of Computational and Applied Mathematics*, 129(1-2):15–35.
- Li, D.-H. and Fukushima, M. (2001b). On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. *SIAM Journal on Optimization*, 11(4):1054–1064.
- Lima, C. H., Lall, U., Troy, T., and Devineni, N. (2016). A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. *Journal of Hydrology*, 541(B):816–823.
- Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(3):503–528.

BIBLIOGRAPHY

- Løland, A. and Høst, G. (2003). Spatial covariance modelling in a complex coastal domain by multidimensional scaling. *Environmetrics*, 14(3):307–321.
- Löw, P. (2019). The natural disasters of 2018 in figures. <https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/natural-disasters/the-natural-disasters-of-2018-in-figures.html>. Last visited on September 14, 2020.
- Mailhot, M. and Mesfioui, M. (2016). Multivariate TVaR-based risk decomposition for vector-valued portfolios. *Risks*, 4(4):33.
- Mao, T., Ng, K., and Hu, T. (2015). Asymptotic expansions of generalized quantiles and expectiles for extreme risks. *Probability in the Engineering and Informational Sciences*, 29(3):309–327.
- Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., and Dugas, B. (2013). Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: Model performance evaluation. *Climate Dynamics*, 41(11-12):2973–3005.
- Matérn, B. (1960). *Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations*. Meddelanden Fran Statens Skogsforskningsinstitut.
- Maume-Deschamps, V., Rullière, D., and Said, K. (2017). Multivariate extensions of expectiles risk measures. *Dependence Modeling*, 5(1):20–44.
- Maume-Deschamps, V., Rullière, D., and Said, K. (2018). Extremes for multivariate expectiles. *Statistics & Risk Modeling*, 35(3-4):111–140.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ.
- McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d -monotone functions and l_1 -norm symmetric distributions. *The Annals of Statistics*, 37(5B):3059–3097.
- Mendes, J. M., de Zea Bermudez, P. C., Pereira, J., Turkman, K., and Vasconcelos, M. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. *Environmental and Ecological Statistics*, 17(1):1–28.
- Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C., and Ward, P. J. (2016). A global reanalysis of storm surges and extreme sea levels. *Nature Communications*, 7(1):1–12.

BIBLIOGRAPHY

- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847.
- Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148.
- Nocedal, J. and Wright, S. (1999). *Numerical Optimization*. Springer, New York.
- Oesting, M., Kabluchko, Z., and Schlather, M. (2012). Simulation of Brown–Resnick processes. *Extremes*, 15(1):89–107.
- Oesting, M., Schlather, M., and Friederichs, P. (2017). Statistical post-processing of forecasts for extremes using bivariate Brown–Resnick processes with an application to wind gusts. *Extremes*, 20(2):309–332.
- Oesting, M., Schlather, M., and Zhou, C. (2018). Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation. *Bernoulli*, 24(2):1497–1530.
- Office of the Superintendent of Financial Institutions (2019). Minimum capital test for federally regulated property and casualty insurance companies. <https://www.osfi-bsif.gc.ca/Eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/mct2019.aspx>. Last visited September 14th, 2020.
- Osband, K. (1985). *Providing Incentives for Better Cost Forecasting*. PhD thesis, University of California, Berkeley.
- Paciorek, C. J. (2013). Spatial models for point and areal data using markov random fields on a fine grid. *Electronic Journal of Statistics*, 7:946–972.
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.
- Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, Vol.2 (Buenos Aires, 1981)*, 49:859–878, 894–902.
- Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4):1430.
- Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

BIBLIOGRAPHY

- Resnick, S. I. (2013). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
- Reza Najafi, M. and Moradkhani, H. (2013). Analysis of runoff extremes using spatial hierarchical Bayesian modeling. *Water Resources Research*, 49(10):6656–6670.
- Ribatet, M., Dombry, C., and Oesting, M. (2016). Spatial extremes and max-stable processes. *Extreme Value Modeling and Risk Analysis, Methods and Applications*, pages 179–194.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22:400–407.
- Royden, H. L. (1988). *Real Analysis*. Macmillan Publishing Company, New York.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Rüschendorf, L. (2009). On the distributional transform, Sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927.
- Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.
- Sang, H. and Gelfand, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:49–65.
- Sang, H. and Genton, M. G. (2014). Tapered composite likelihood for spatial max-stable models. *Spatial Statistics*, 8:86–103.
- Schabenberger, O. and Gotway, C. A. (2017). *Statistical Methods for Spatial Data Analysis*. CRC Press, Boca Raton, FL.
- Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782.
- Segers, J., Sibuya, M., and Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis*, 155:35–51.
- Shin, Y., Lee, Y., Choi, J., and Park, J.-S. (2019). Integration of max-stable processes and bayesian model averaging to predict extreme climatic events in multi-model ensembles. *Stochastic Environmental Research and Risk Assessment*, 33(1):47–57.

BIBLIOGRAPHY

- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.
- Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*, 205.
- Song, Y. and Yan, J.-A. (2009). Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. *Insurance: Mathematics and Economics*, 45(3):459–465.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.
- Stephenson, A. G. (2016). Harmonic analysis of tides using tideharmonics. <https://CRAN.R-project.org/package=TideHarmonics>.
- Tasche, D. (2002). Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533.
- Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28:39–58.
- Teufel, B., Diro, G., Whan, K., Milrad, S., Jeong, D., Ganji, A., Huziy, O., Winger, K., Gyakum, J., de Elia, R., et al. (2017). Investigation of the 2013 Alberta flood from weather and climate perspectives. *Climate Dynamics*, 48(9-10):2881–2899.
- Teufel, B., Sushama, L., Huziy, O., Diro, G., Jeong, D., Winger, K., Garnaud, C., de Elia, R., Zwiers, F., Matthews, H., et al. (2019). Investigation of the mechanisms leading to the 2017 Montreal flood. *Climate Dynamics*, 52(7-8):4193–4206.
- Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *The Annals of Applied Statistics*, 10(4):2303–2324.
- Torres, R., De Michele, C., Laniado, H., and Lillo, R. E. (2017). Directional multivariate extremes in environmental phenomena. *Environmetrics*, 28(2):e2428.
- Torres, R., Lillo, R. E., and Laniado, H. (2015). A directional multivariate value at risk. *Insurance: Mathematics and Economics*, 65:111–123.
- Towe, R., Tawn, J., and Lamb, R. (2018). Why extreme floods are more common than you might think. *Significance*, 15(6):16–21.

BIBLIOGRAPHY

- Usseglio-Carleve, A. (2018). Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electronic Journal of Statistics*, 12(2):4057–4093.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- Ver Hoef, J. M., Peterson, E., and Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, 13(4):449–464.
- Ver Hoef, J. M. and Peterson, E. E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*, 105(489):6–18.
- Vettori, S., Huser, R., and Genton, M. G. (2018). A comparison of dependence function estimators in multivariate extremes. *Statistics and Computing*, 28(3):525–538.
- Vettori, S., Huser, R., Segers, J., and Genton, M. G. (2020). Bayesian model averaging over tree-based dependence structures for multivariate extremes. *Journal of Computational and Graphical Statistics*, 29(1):174–190.
- Von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Selected Papers*, 2:271–294.
- Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183.
- Watanabe, S. (2009). *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, Cambridge.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.
- Winsemius, H. C., Aerts, J. C., van Beek, L. P., Bierkens, M. F., Bouwman, A., Jongman, B., Kwadijk, J. C., Ligtoet, W., Lucas, P. L., Van Vuuren, D. P., et al. (2016). Global drivers of future river flood risk. *Nature Climate Change*, 6(4):381.
- Yamai, Y. and Yoshida, T. (2002). On the validity of value-at-risk: Comparative analyses with expected shortfall. *Monetary and Economic Studies*, 20(1):57–85.
- Yamai, Y. and Yoshida, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, 29(4):997–1015.

Acronyms and Notation

Yan, H. and Moradkhani, H. (2015). A regional Bayesian hierarchical model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, 29(3):1019–1036.

Yuan, G., Wei, Z., and Lu, X. (2017). Global convergence of BFGS and PRP methods under a modified weak Wolfe–Powell line search. *Applied Mathematical Modelling*, 47:811–825.

Ziegel, J. F. (2016). Coherence and elicibility. *Mathematical Finance*, 26(4):901–918.

Acronyms and Notation

CDF	Cumulative distribution function
CTE	Conditional tail expectation
ES	Expected shortfall
EVT	Extreme value theory
GPD	Generalized Pareto distribution
HBM	Hierarchical Bayesian model
IID	Independently and identically distributed
MCMC	Markov chain Monte Carlo
MDA	Maximum domain of attraction
MEE	Multivariate extreme expectiles
MHG	Metropolis–Hastings-within-Gibbs
MLE	Maximum likelihood estimate
POT	Peaks-over-threshold
QQ	Quantile-quantile
RMS	Root mean square error
RV	Random variable
STDF	Stable tail dependence function
SF	Survival function
SSE	Sum squared error
TVAR	Tail value-at-risk
UDF	Upper tail dependence function
VAR	Value-at-risk
WAIC	Watanabe-Akaike Information Criterion

$\mathcal{GEV}(\mu, \sigma, \xi)$	Univariate generalized extreme value distribution with location, scale and shape μ , σ and ξ respectively.
$\mathcal{N}(\mu, \sigma^2)$	Univariate Normal distribution with mean μ and variance σ^2
$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	d -dimensional Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathcal{U}(0, 1)$	Univariate (standard) uniform distribution on the interval $(0, 1)$
\rightsquigarrow	Weak convergence
$\xrightarrow[n \rightarrow \infty]{\mathbb{P}}$	Convergence in probability
$\xrightarrow[n \rightarrow \infty]{a.s.}$	Convergence almost surely
$\mathbb{1}(\cdot)$	Indicator function
\mathbb{R}^d	d -dimensional reals