



**HAL**  
open science

# Simulation and assessment of multivariate extreme models for environmental data

Juliette Legrand

► **To cite this version:**

Juliette Legrand. Simulation and assessment of multivariate extreme models for environmental data. Statistics [math.ST]. Université Paris-Saclay, 2022. English. NNT : 2022UPASJ015 . tel-03827820

**HAL Id: tel-03827820**

**<https://theses.hal.science/tel-03827820v1>**

Submitted on 24 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Simulation and assessment of multivariate extreme models for environmental data

*Simulation et évaluation de modèles d'extrêmes  
multivariés pour des données environnementales*

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 129 : Sciences de l'Environnement d'Ile-de-France (SEIF)

Spécialité de doctorat: Sciences du climat, de l'atmosphère et des  
océans, terrestres et planétaire

Graduate School : Géosciences, climat, environnement et planètes

Réfèrent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche LSCE (Université Paris-Saclay, CNRS,  
CEA, UVSQ), sous la direction de **Philippe Naveau**, Directeur de Recherche  
CNRS, le co-encadrement de **Pierre Ailliot**, Maître de conférences, et le  
co-encadrement de **Nicolas Raillard**, Ingénieur de Recherche.

Thèse soutenue à Paris-Saclay, le 2 septembre 2022, par

**Juliette LEGRAND**

### Composition du jury

<b>Liliane Bel</b> Professeure, AgroParisTech	Présidente
<b>Anthony C. Davison</b> Professeur, Ecole Polytechnique Fédérale de Lausanne	Rapporteur et examinateur
<b>Carlo Gaetan</b> Professeur, Università Ca'Foscari Venezia	Rapporteur et examinateur
<b>Jean-François Filipot</b> Directeur scientifique, France Energies Marines	Examinateur
<b>Anne-Laure Fougères</b> Professeure, Université de Lyon 1	Examinatrice
<b>Valérie Monbet</b> Professeure, Université Rennes 1	Examinatrice
<b>Jérémy Rohmer</b> Ingénieur de Recherche, BRGM	Examinateur
<b>Philippe Naveau</b> Directeur de Recherche CNRS, LSCE	Directeur de thèse



**Titre:** Simulation et évaluation de modèles d'extrêmes multivariés pour des données environnementales

**Mots clés:** Extrêmes multivariés; Simulation d'extrêmes cooccurrents; Variation régulière multivariée et cachée; Évaluation de classifieurs extrêmes; Hauteur significative de vagues; Débits de rivière

**Résumé:** L'estimation précise des probabilités d'occurrence des événements extrêmes environnementaux est une préoccupation majeure dans l'évaluation des risques. Pour l'ingénierie côtière par exemple, le dimensionnement de structures implantées sur ou à proximité des côtes doit être tel qu'elles résistent aux événements les plus sévères qu'elles puissent rencontrer au cours de leur vie. Cette thèse porte sur la simulation d'événements extrêmes multivariés, motivée par des applications aux hauteurs significatives de vagues, et sur l'évaluation de modèles de prédiction d'occurrence d'événements extrêmes.

Dans la première partie du manuscrit, nous proposons et étudions un simulateur stochastique qui génère conjointement, en fonction de certaines conditions d'état de mer au large, des extrêmes de hauteur significative de vagues ( $H_s$ ) au large et à la côte. Pour cela, nous nous appuyons sur l'approche par dépassements de seuils bivariés et nous développons un algorithme de simulation non-paramétrique de lois de Pareto généralisées bivariées. À partir de ce simulateur d'événements cooccurrents, nous dérivons un modèle de simulation condition-

nel. Les deux algorithmes de simulation sont mis en oeuvre sur des expériences numériques et appliqués aux extrêmes de  $H_s$  près des côtes bretonnes françaises. Un autre développement est traité quant à la modélisation des lois marginales des  $H_s$ . Afin de prendre en compte leur non-stationnarité, nous adaptons une extension de la loi de Pareto généralisée, en considérant l'effet de la période et de la direction pic sur ses paramètres.

La deuxième partie de cette thèse apporte un développement plus théorique. Pour évaluer différents modèles de prédiction d'extrêmes, nous étudions le cas spécifique des classifieurs binaires, qui constituent la forme la plus simple de prévision et de processus décisionnel : un événement extrême s'est produit ou ne s'est pas produit. Des fonctions de risque adaptées à la classification binaire d'événements extrêmes sont développées, ce qui nous permet de répondre à notre deuxième question. Leurs propriétés sont établies dans le cadre de la variation régulière multivariée et de la variation régulière cachée, permettant de considérer des formes plus fines d'indépendance asymptotique. Ces développements sont ensuite appliqués aux débits de rivière extrêmes.

**Title:** Simulation and assessment of multivariate extreme models for environmental data

**Keywords:** Multivariate extremes; Simulation of joint extremes; Multivariate and hidden regular variation; Evaluation of extreme classifiers; Significant wave height; River discharges

**Abstract:** Accurate estimation of the occurrence probabilities of extreme environmental events is a major issue for risk assessment. For example, in coastal engineering, the design of structures installed at or near the coasts must be such that they can withstand the most severe events they may encounter in their lifetime. This thesis focuses on the simulation of multivariate extremes, motivated by applications to significant wave height, and on the evaluation of models predicting the occurrences of extreme events.

In the first part of the manuscript, we propose and study a stochastic simulator that, given offshore conditions, produces jointly offshore and coastal extreme significant wave heights ( $H_s$ ). We rely on bivariate Peaks over Threshold and develop a non-parametric simulation scheme of bivariate generalised Pareto distributions. From such joint simulator, we derive a conditional simulation model. Both simulation algorithms are applied

to numerical experiments and to extreme  $H_s$  near the French Brittany coast. A further development is addressed regarding the marginal modelling of  $H_s$ . To take into account non-stationarities, we adapt the extended generalised Pareto model, letting the marginal parameters vary with the peak period and the peak direction.

The second part of this thesis provides a more theoretical development. To evaluate different prediction models for extremes, we study the specific case of binary classifiers, which are the simplest type of forecasting and decision-making situation: an extreme event did or did not occur. Risk functions adapted to binary classifiers of extreme events are developed, answering our second question. Their properties are derived under the framework of multivariate regular variation and hidden regular variation, allowing to handle finer types of asymptotic independence. This framework is applied to extreme river discharges.



# Remerciements

Je souhaite tout d'abord remercier chaleureusement les membres de mon jury de thèse. Anthony Davison et Carlo Gaetan, j'ai été très honorée que vous ayez accepté de relire ce manuscrit. Je vous remercie pour vos retours pertinents et vos conseils avisés. Merci également à Liliane Bel, Jean-François Filipot, Anne-Laure Fougères, Valérie Monbet et Jérémy Rohmer pour votre bienveillance mais également pour les riches discussions et questions soulevées lors de la soutenance qui sont autant de pistes possibles pour poursuivre cette recherche.

Merci beaucoup Pierre, Philippe et Nicolas pour votre encadrement. Ces trois années n'auraient pas pu mieux se passer. J'ai évidemment appris énormément à vos côtés mais j'ai également passé de bons moments à rire et à échanger avec vous autour d'une bière brestoise.

Merci infiniment à Valérie Monbet, qui depuis le Master a été, et est encore, d'un soutien immense. Merci pour les nombreuses discussions que l'on a pu avoir, c'est toujours un réel plaisir d'échanger avec toi.

*A special thanks* à Julie Bessac, merci de m'avoir si bien accueillie à Chicago, de m'avoir aidée pour faire mes courses, de m'avoir prêté ton vélo et pour les pizzas chez toi ! Mais surtout merci de m'avoir introduite au monde de la recherche et de la modélisation.

À Jérémy Rohmer, merci pour ta gentillesse et tes idées enthousiastes sur mes travaux. Grâce à toi les projections de vagues n'ont plus de secret pour moi ! J'en profite pour remercier les membres du BRGM pour leur accueil chaleureux à Orléans.

Un très grand merci à Gwladys Toulemonde qui a accepté avec beaucoup de bienveillance de participer au suivi de cette thèse. Merci énormément pour ta disponibilité, tes encouragements et tes conseils tout au long de ces trois années.

Many thanks to Marco Oesting, I am very honoured to have been one of your co-authors. Thank you so much for your kindness and advice. Thank you also for introducing me to the particularities of German cuisine in Stuttgart!

Un immense merci à Soulivanh Thao. Ton aide précieuse a permis que cette soutenance se déroule au mieux et sans souci technique.

Les amis c'est chouette ! Pour cela merci à tous les copains de thèse, de Rennes, de Brest, de Gif ou d'ailleurs. Je pense notamment à ceux de la communauté des valeurs extrêmes, que j'ai la chance de revoir chaque année en par-

---

ticulier à Aussois. Philomène, Pauline, Romain, Gloria, Bastien, Paula, Nicolas, Jonathan (et tous ceux que j'oublie) merci pour toutes ces parties de baby-foot et ces karaoké d'anthologie. Un immense merci aux organisateurs de VALPRED Clément Dombry, Olivier Wintenberger et Philippe, de rendre ces rencontres possibles.

Si cette thèse a pu si bien se dérouler, c'est notamment grâce au soutien financier de l'ENS de Rennes que j'ai reçu, par le biais des Contrats Doctoraux Spécifiques Normaliens. Je remercie également l'IFREMER pour son soutien matériel et pour l'accès aux données d'états de mer.

Je tiens évidemment à remercier mes parents pour leur soutien mais aussi pour leur envie constante d'essayer de comprendre ce que font leurs enfants afin de pouvoir expliquer à d'autres ce qu'ils "cherchent".

Enfin, simplement merci pour tout Camille. Après deux agreg et deux thèses, je pense qu'on a eu notre quota pour les prochaines années.

# Résumé en Français

Les zones côtières sont fréquemment soumises à des événements maritimes extrêmes. Ces événements sont, par définition, ceux ayant le plus d'impact sur les populations ou les activités économiques. En France, l'exemple le plus dramatique fût la tempête Xynthia qui s'est abattue en février 2010 sur la façade atlantique, engendrant un nombre élevé de victimes et de dommages matériels (Genovese and Przyluski, 2013). De part sa dimension catastrophique, cet événement est très certainement la tempête maritime qui a le plus marqué les mémoires.

Chaque année, plusieurs tempêtes touchent le littoral français, fort heureusement sans avoir de conséquences humaines aussi dramatiques que la tempête Xynthia, mais qui ont cependant de nombreux impacts que ce soit sur la fiabilité des structures installées en mer ou sur les côtes, ou encore sur l'érosion des littoraux.

Les hauteurs de vagues extrêmes sont une des caractéristiques de ces tempêtes maritimes. Afin de déterminer les potentiels impacts des phénomènes maritimes extrêmes, il est donc crucial de pouvoir caractériser et prédire les vagues les plus extrêmes pouvant survenir près des littoraux. Dans ce contexte, les travaux présentés dans ce manuscrit visent à répondre aux deux questions suivantes :

- Comment simuler de tels événements extrêmes ? (**Problématique n°1**)
- Comment comparer différents modèles de simulation, ou de prédiction, d'événements extrêmes ? (**Problématique n°2**)

Une première partie des travaux porte sur la modélisation des extrêmes de la hauteur significative des vagues, notée  $H_s$ . Cette quantité permet de mesurer l'énergie des vagues et donc leur sévérité. La **problématique n°1** est traitée en considérant les conditions d'états de mer au large pouvant générer des hauteurs de vagues extrêmes près des côtes.

Une fois le modèle de simulation construit, une question naturelle est d'évaluer ses performances en termes de prédiction d'extrêmes. La seconde partie des travaux de cette thèse vise à répondre à cette **problématique n°2**, en considérant le cas particulier des classifieurs binaires, qui peuvent être vus comme des encodeurs d'événements extrêmes (un extrême a eu lieu ou n'a pas eu lieu).

Un résumé de ces différents travaux est donné ci-dessous.



### Simulation jointe de $H_s$ à la côte et au large

Pour répondre à la **problématique n°1**, nous considérons deux points spécifiques un point au large (représenté en rouge sur la Figure 1) et un point à la côte (en vert sur la Figure 1). Nous souhaitons modéliser conjointement les extrêmes de hauteur significative de vagues en ces deux points.

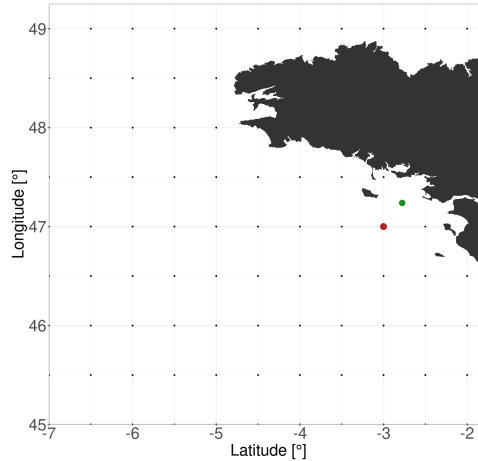


Figure 1: Carte représentant le point au large (en rouge) et le point près des côtes (point vert).

Afin de modéliser simultanément les  $H_s$  au large et à la côte, l'outil adéquat est la théorie des extrêmes multivariés, qui permet de modéliser la structure de dépendance entre les valeurs les plus extrêmes d'un vecteur aléatoire.

Dans le cadre univarié, un extrême peut être défini comme un événement dépassant un certain seuil, fixé au préalable (approche par dépassements de seuil). Dans ce contexte, pour un seuil  $u$  suffisamment grand, les dépassements de seuils, définis par  $[X - u \mid X > u]$  où  $X$  est une variable aléatoire réelle, peuvent être approximés par une loi de Pareto généralisée, notée GPD (Coles, 2001). Cette caractérisation a été ensuite étendue au cadre multivarié. Nous suivons ici la définition développée par Rootzén and Tajvidi (2006), en se restreignant au cadre bivarié (correspondant à notre application) : si l'on considère un vecteur  $\mathbf{X} = (X_1, X_2)$  dans  $\mathbb{R}^2$ , alors les dépassements de seuils bivariés

$$[\mathbf{X} - \mathbf{u} \mid \mathbf{X} \not\leq \mathbf{u}],$$

peuvent être approximés par une GPD bivariée, où  $\mathbf{X} \not\leq \mathbf{u}$  signifie que  $X_1 \geq u_1$  ou  $X_2 \geq u_2$ , si  $\mathbf{u} = (u_1, u_2) \in \mathbb{R}^2$ . Une illustration de ces dépassements de seuils est présentée sur la Figure 2 avec les données de  $H_s$  au large et à la côte.

Afin de construire de tels vecteurs GPD, Rootzén et al. (2018a) ont établi la représentation suivante

$$\mathbf{Z} = \mathbf{E} + \mathbf{T} - \max(\mathbf{T}), \quad (1)$$

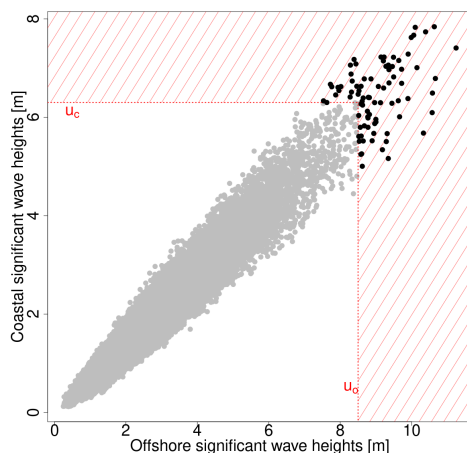


Figure 2:  $H_s$  à la côte en fonction des  $H_s$  au large, la zone hachurée rouge correspond aux dépassements bivariés des hauteurs significatives de vagues.

où  $\mathbf{T}$  est un vecteur bivarié et  $E$  une variable exponentielle, indépendante de  $\mathbf{T}$ . Le vecteur  $\mathbf{Z}$  ainsi défini suit alors une loi GPD standard.

A partir de la représentation (1), nous développons un générateur stochastique permettant de simuler conjointement des  $H_s$  extrêmes au large et à la côte, et cela en fonction de certaines conditions au large, qui sont notamment des  $H_s$  au large modérément élevées.

Pour cela, un algorithme de simulation non-paramétrique de vecteurs GPD est présenté. Cet algorithme s'appuie sur la ré-écriture suivante de l'Equation (1) :

$$\begin{cases} Z_1 = E + \Delta \mathbb{1}_{\Delta < 0}, \\ Z_2 = E - \Delta \mathbb{1}_{\Delta \geq 0}, \end{cases} \quad (2)$$

où  $\Delta := Z_1 - Z_2 = T_1 - T_2$  et  $\mathbb{1}$  correspond à la fonction indicatrice. Le point clef de l'algorithme est l'utilisation d'un bootstrap non-paramétrique pour simuler des valeurs de  $\Delta$ . Les performances de l'algorithme sont illustrées sur plusieurs simulations numériques.

Avant d'appliquer ce générateur stochastique de vecteurs GPD aux données de  $H_s$ , l'Equation (1) requiert que les variables soient standardisées à la même échelle. Pour cela, les distributions marginales des  $H_s$  au large et à la côte sont modélisées par une EGPD (Naveau et al., 2016), qui permet d'éviter la sélection parfois difficile du seuil  $u$  de la GPD. L'EGPD autorise la modélisation de toute la plage de données et pas uniquement la queue de distribution. Cela nous permet notamment de modéliser comment des  $H_s$  au large modérément élevées peuvent produire des  $H_s$  extrêmes à la côte.

Une spécificité des  $H_s$  est leur caractère non-stationnaire. En effet, Jonathan and Ewans (2013) ont montré que la modélisation des extrêmes de hauteur significative de vagues nécessite de prendre en compte l'effet de certaines covariables,

telles que la direction des vagues ou la saisonnalité. Dans notre étude, nous considérons l'effet de la période pic (notée  $T_p$ ) et de la direction pic (notée  $D_p$ ) sur la relation entre les  $H_s$  au large et celles à la côte. Tout comme la hauteur significative de vagues, ces deux variables contribuent à caractériser un état de mer. Elles correspondent à la période et la direction de propagation du maximum - ou "pic" - d'énergie des vagues. La dépendance en  $T_p$  et  $D_p$  des modèles sur les lois marginales est portée, dans les deux cas, par le paramètre d'échelle de l'EGPD à l'aide de splines de lissage (Le Carrer, 2022).

Enfin, un second algorithme est développé, permettant de simuler des  $H_s$  à la côte conditionnellement aux  $H_s$  au large. Ce second générateur peut être vu comme une méthode de dowsaling statistique des  $H_s$  au large (considérées comme des données provenant d'un modèle global de vagues) vers les  $H_s$  à la côte (issues d'un modèle local).

Les deux algorithmes de simulation sont ensuite validés sur des expériences numériques et appliqués aux extrêmes de  $H_s$  près des côtes bretonnes françaises.

### Evaluation et comparaison de classifieurs d'événements extrêmes

Pour répondre à la **problématique n°2**, nous étudions le cas particulier des classifieurs binaires. Pour cela, nous considérons que l'on dispose d'une variable binaire  $Y \in \{-1, +1\}$  qui encode l'occurrence d'un événement extrême :

$$Y := Y^{(u)} = \begin{cases} +1, & \text{si } H > u, \\ -1, & \text{sinon,} \end{cases}$$

où  $H \in [0, \infty)$  est une certaine variable aléatoire (par exemple un débit de rivière) et  $u > 0$  un seuil critique élevé.

A partir d'un ensemble d'observations multivariées  $\mathbf{X} \in [0, \infty)^d$ , nous cherchons à déterminer, parmi différents classifieurs binaires  $g(\mathbf{X}; u) \in \{-1, +1\}$  de ces observations, le meilleur en terme de prédictions d'occurrence d'événements extrêmes, i.e. correspondant à  $Y^{(u)} = +1$ . Pour fixer les idées, on peut considérer que les observations  $\mathbf{X}$  représentent des débits de rivière en différentes stations de mesure (points rouges sur la carte 3), et que  $Y^{(u)}$  décrit l'occurrence d'un débit de rivière extrême en une station différente de  $\mathbf{X}$  (triangle jaune sur la carte 3).

Afin de comparer différents classifieurs, nous définissons la fonction de risque suivante

$$R^{(u)}(g) = \frac{\mathbb{P}(g(\mathbf{X}; u) \neq Y^{(u)})}{\mathbb{P}(Y^{(u)} = 1 \text{ ou } g(\mathbf{X}; u) = 1)} \in [0, 1].$$

Si l'on regarde la Figure 4 représentant des prédictions binaires obtenues via un arbre de décision, cette fonction de risque revient à compter le nombre moyen de points se trouvant dans les zones jaune et rouge, pondéré par la probabilité d'être dans les zones jaune, verte ou rouge.

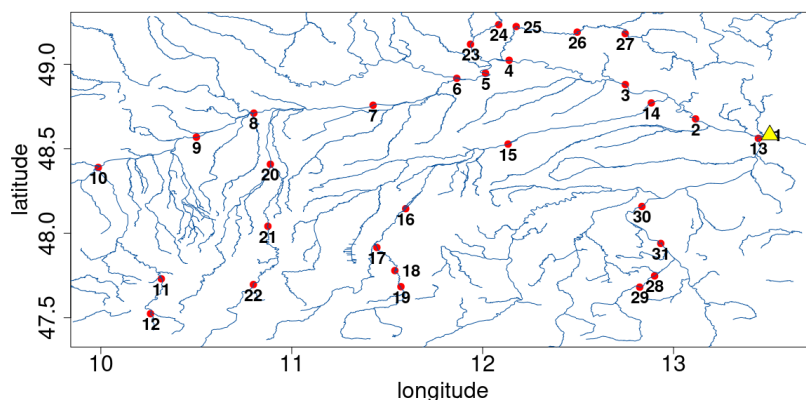


Figure 3: Carte du bassin versant du Danube, montrant les 31 stations de jaugeage le long du Danube et de ses affluents.

Comme nous considérons les occurrences d'événements extrêmes, en pratique, le seuil critique  $u$  est très grand et il y a très peu de points dans la zone verte. Nous définissons alors le risque extrême d'un classifieur par  $R(g) := \lim_{u \rightarrow \infty} R^{(u)}(g)$ . Et nous considérons que pour deux classifieurs binaires  $g_1$  et  $g_2$ , si  $R(g_1) < R(g_2)$ , alors  $g_1$  est un meilleur classifieur en terme de prédictions d'occurrence d'événements extrêmes que  $g_2$ .

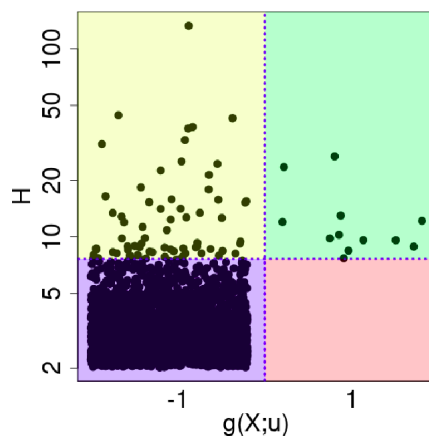


Figure 4: Exemple de prédictions binaires obtenues avec un arbre de décision, contre les observations  $H$ . La ligne horizontale bleue représente le seuil critique  $u$ , la ligne verticale sépare les deux classes prédites  $-1$  et  $+1$ .

Un second développement est porté sur le comportement de dépendance asymptotique entre la variable d'intérêt  $H$  et les classifieurs binaires. Pour cela, nous considérons des classifieurs de la forme

$$g(\mathbf{X}; u) = \begin{cases} +1, & \text{si } \bar{g}(\mathbf{X}) > u, \\ -1, & \text{sinon,} \end{cases}$$

pour une certaine fonction  $\bar{g} : \mathbb{R}^d \rightarrow (0, \infty)$ . Dans ce cas-là, nous montrons que la mesure de risque extrême  $R(g)$  telle que définie précédemment ne permet pas de discriminer des classifieurs dans le cas où  $\bar{g}(\mathbf{X})$  et  $H$  sont asymptotiquement indépendants.

Pour contourner ce problème, nous définissons une seconde fonction de risque, adaptée aux cas de l'indépendance asymptotique, en considérant le conditionnement suivant pour  $\varepsilon \in [0, 1)$

$$R_\varepsilon(g) := \lim_{u \rightarrow \infty} \frac{\mathbb{P}(g(X; u) \neq Y^{(u)} \mid Y^{(\varepsilon u)} = g(X; \varepsilon u) = 1)}{\mathbb{P}(Y^{(u)} = 1 \text{ ou } g(X; u) = 1 \mid Y^{(\varepsilon u)} = g(X; \varepsilon u) = 1)}.$$

Nous établissons alors un estimateur empirique de  $R_\varepsilon(g)$ , dont les propriétés asymptotiques sont dérivées dans le cadre de la variation régulière multivariée et de la variation régulière cachée, à l'aide principalement du modèle de [Ramos and Ledford \(2009\)](#), dont une adaptation est la suivante

$$\mathbb{P}[\bar{g}(X) > u, H > v] = L(u, v)(u^{-\alpha_g} v^{-\alpha_H})^{1/2\eta},$$

où  $\eta \in (0, 1]$  indique le vitesse de décroissance de la queue bivariée,  $L(\cdot, \cdot)$  est une fonction à variation lente et  $\alpha_H, \alpha_g > 0$  sont les indices de queue respectifs de  $H$  et  $\bar{g}(\mathbf{X})$ .

Cette seconde fonction de risque est appliquée à l'analyse des débits extrêmes de rivière dans le bassin versant du Danube. L'application compare différents algorithmes de prédiction (classifieur linéaire, régression logistique, arbres de décision, ...) et teste leur capacité à prédire les débits de rivière extrêmes en une station donnée à partir des observations en d'autres stations (voir Figure 3).

# Table of contents

<b>Résumé en Français</b>	<b>7</b>
<b>List of Figures</b>	<b>17</b>
<b>List of Tables</b>	<b>19</b>
<b>Introduction</b>	<b>21</b>
<b>1 Elements of extreme value theory</b>	<b>23</b>
1.1 Univariate extremes . . . . .	24
1.1.1 Possible limits of block maxima . . . . .	24
1.1.2 Peaks over thresholds approach . . . . .	27
1.1.3 Regular variation and domain of attraction . . . . .	31
1.2 Multivariate extremes . . . . .	32
1.2.1 Domain of attraction and characterisation . . . . .	34
1.2.2 Bivariate threshold excess model . . . . .	38
1.2.3 Asymptotic dependence and independence: Definitions and characterisations. . . . .	43
1.2.4 Limitations of bivariate EVT . . . . .	50
1.2.5 Models coping with asymptotic independence . . . . .	51
1.3 Tools considered in the following . . . . .	53
<b>2 Joint stochastic simulation of extreme coastal and offshore significant wave heights</b>	<b>55</b>
2.1 Preamble to Paper I . . . . .	56
2.1.1 Wave data . . . . .	57
2.1.2 On the choice of appropriate covariates for modelling extreme coastal sea states . . . . .	60
2.1.3 Extended generalised Pareto distribution . . . . .	62
2.2 Introduction . . . . .	66
2.3 Sea state data . . . . .	69
2.4 Marginal regression analysis . . . . .	70
2.4.1 Marginal regression . . . . .	70

TABLE OF CONTENTS

---

2.4.2	Covariates effects . . . . .	72
2.5	Multivariate Pareto model . . . . .	73
2.5.1	Simulation of bivariate standard generalised Pareto distributed vectors . . . . .	74
2.5.2	Numerical experiments . . . . .	74
2.5.3	Conditional simulation within the MGP class . . . . .	75
2.5.4	Numerical experiment continued . . . . .	78
2.6	Application to extreme significant wave height . . . . .	78
2.6.1	Joint simulation of significant wave heights . . . . .	79
2.6.2	Conditional simulation of coastal significant wave heights . . . . .	81
2.7	Discussion and conclusions . . . . .	82
	Appendix 2.A Marginal regression modelling . . . . .	84
	Appendix 2.B Proof of Algorithm 1 . . . . .	85
2.3	Summary of Paper I . . . . .	86
<b>3</b>	<b>Evaluation of binary classifiers for asymptotically dependent and in- dependent extremes</b>	<b>87</b>
3.1	Preamble to Paper II . . . . .	88
3.1.1	Binary classifiers and extremal risk . . . . .	89
3.1.2	Refinement of asymptotic independence . . . . .	90
3.2	Introduction . . . . .	94
3.3	Risk, upper tail equivalence and extremal dependence . . . . .	97
3.3.1	Conditional risk and hidden regular variation . . . . .	100
3.3.2	Risk function inference . . . . .	102
3.4	Simulations . . . . .	103
3.4.1	A simple linear setup . . . . .	103
3.4.2	Classifiers descriptions . . . . .	103
3.4.3	Implementation and results . . . . .	105
3.5	Danube river discharges . . . . .	107
3.6	Supplementary Materials . . . . .	110
3.7	Supplements to Paper II . . . . .	110
3.7.1	Linear Classifiers: Definition, Basic Properties and Inference . . . . .	111
3.7.2	Feature Selection . . . . .	113
3.7.3	River network . . . . .	114
3.7.4	Possible extension to hidden regular variation, a simulation study . . . . .	118
3.8	Summary of Paper II . . . . .	120
<b>4</b>	<b>Some directions for future work</b>	<b>123</b>
	<b>Appendix A Technical proofs of Paper II</b>	<b>127</b>
A.1	Proofs . . . . .	127
A.1.1	Proof of Lemma 3.4: . . . . .	127

A.1.2	Proof of Lemma 3.6: . . . . .	128
A.1.3	Proof of Proposition 3.7: . . . . .	128
A.1.4	Proof of Proposition 3.8: . . . . .	129
A.2	Proofs of the appendix . . . . .	130
A.2.1	Proof of Proposition 3.9 . . . . .	130
A.2.2	Proof of Lemma A.1 . . . . .	131
<b>Bibliography</b>		<b>133</b>





# List of Figures

1	Carte des données de vagues . . . . .	8
2	$H_s$ à la côte en fonction des $H_s$ au large . . . . .	9
3	Carte du bassin versant du Danube . . . . .	11
4	Prédictions issues d'un arbre de décision . . . . .	11
1.1	Densities of the three generalised extreme value distributions . . . . .	25
1.2	Densities of the GP distribution for different values of the shape parameter $\xi$ . . . . .	28
1.3	Upper tail distribution of significant wave heights . . . . .	29
1.4	Block maxima versus peaks over thresholds . . . . .	30
1.5	Two possible definitions of bivariate extremes . . . . .	33
1.6	Asymptotic independence of a bivariate Gaussian vector . . . . .	34
1.7	Bivariate logistic extreme value model . . . . .	37
1.8	Scatter plot of significant wave heights . . . . .	38
1.9	Illustration of $\mathbf{u}(t)$ . . . . .	40
1.10	Bivariate generalised Pareto distributions . . . . .	42
1.11	Asymptotic dependence behaviour . . . . .	44
1.12	Spectral measure and Pickands dependence function . . . . .	45
1.13	Dependence measure $\chi$ for wave data . . . . .	47
1.14	Asymptotic dependence comparison on Pareto scale . . . . .	48
2.1	Spectral decomposition of a sea state . . . . .	57
2.2	Wave spectrum . . . . .	58
2.3	Hindcast databases grid . . . . .	60
2.4	Illustration of a wind field . . . . .	61
2.5	Measure of asymptotic dependence between wave data . . . . .	62
2.6	Densities of the extended GP distribution for different values of the shape parameter $\kappa$ . . . . .	64
2.7	Grid of the wave data and scatter plot of significant wave heights . . . . .	67
2.8	Empirical histograms of $H_s$ threshold exceedances . . . . .	68
2.9	Dependence of $H_s$ with respect to $T_p$ and $D_p$ . . . . .	69
2.10	Maps illustrating the covariates effects . . . . .	72
2.11	Simulation results for Algorithm 1 . . . . .	76
2.12	Simulation results for Algorithm 2 . . . . .	79

LIST OF FIGURES

---

2.13	Sampled values of coastal versus offshore $H_s$ from Algorithm 1 . . .	80
2.14	Predicted $H_c$ conditionally on $(H_o, T_p, D_p)$ using Algorithm 2. . . . .	82
2.15	Non-stationary illustration of the predicted $H_c$ using Algorithm 2 . . .	83
2.16	Conditional marginal EGPD fits . . . . .	84
3.1	Space considered for multivariate regular variation . . . . .	92
3.2	Example of hidden regular variation . . . . .	93
3.3	Simulated example for Lemma 3.6 . . . . .	101
3.4	Predicted binary output for the decision tree classifier . . . . .	105
3.5	Estimation of $R_\epsilon(g)$ for different classifiers . . . . .	106
3.6	River map of the upper Danube basin . . . . .	107
3.7	Summer daily measurements of river discharges from station 1 (y-axis) against station 23 (left x-axis) and station 24 (right x-axis) . . . . .	108
3.8	Estimation of $R_\epsilon(g)$ for five different classifiers with the Danube river discharges . . . . .	109
3.9	Correlation plot between Station 1 and the stations considered to contribute to the extremes of Station 1 . . . . .	116
3.10	Estimation of $R_0(g)$ for five different classifiers . . . . .	116
3.11	Predictive power of the linear classifier . . . . .	117
4.1	Possible applications of a multivariate model . . . . .	125

# List of Tables

2.1	Summary of available data . . . . .	67
2.2	Estimated parameters for the regression marginal models . . . . .	71
2.3	Simulation experiment settings . . . . .	75
2.4	Empirical joint survival probability of exceeding the observed extremes $h_c$ and $h_o$ . . . . .	80
3.1	Summary and key features of different classifiers . . . . .	104
3.2	Estimates of the shape parameter with the river network . . . . .	115
3.3	Empirical estimates of the feature selection coefficients . . . . .	115
3.4	Empirical estimates of the limits $c_{i,\varepsilon}^*$ . . . . .	120



# Introduction

## Motivations

Coastal areas are frequently subject to extreme maritime events. These events are, by definition, those that have the greatest impact on populations or economic activities. In France, the most dramatic example was the Xynthia storm which hit the Atlantic coast in February 2010, causing a high number of fatalities and material damages (Genovese and Przulski, 2013). Due to its catastrophic dimension, this event is undoubtedly the maritime storm that has left the greatest mark on people's memories.

Every year, numerous storms hit the French coastline, fortunately without having such dramatic human consequences as the Xynthia storm, but which nevertheless have multiple impacts, whether on the reliability of structures installed at sea or on the coasts, or on coastal erosion.

Extreme wave heights are one of the characteristics of these maritime storms. In order to determine the potential impacts of extreme maritime events, it is therefore crucial to be able to characterise and predict the most extreme waves that can occur near the coast. In this context, the work presented in this manuscript aims to answer the following questions:

- How to simulate such extreme events? (**Problematic n°1**)
- How to compare different simulation, or prediction, models of extreme events? (**Problematic n°2**)

The first part of the work concerns the modelling of extreme significant wave heights, denoted  $H_s$ . This quantity measures the energy of the waves and, consequently, their severity. The **problematic n°1** is addressed by considering the offshore sea state conditions that can generate extreme wave heights near the coast.

Once the simulation model is built, a natural question is to evaluate its performance in terms of prediction of extremes. The second part of the work of this thesis aims at answering this **problematic n°2**, by considering the particular case of binary classifiers, which can be seen as extreme event encoders (an extreme has or has not occurred).

## Outline of the thesis

The outline of this manuscript is as follows. In Chapter 1, the main results of extreme value theory (EVT) are discussed. Since this theory entails many concepts and questions, a focus is made on the key elements of univariate and multivariate EVT, that are useful for the rest of this thesis.

The main goal of Chapter 2 is to answer the **problematic n°1**. For that, we propose and study a stochastic simulator that, given offshore conditions, produces jointly offshore and coastal extreme significant wave heights. We rely on bivariate Peaks over Threshold (Rootzén and Tajvidi, 2006) and develop a non-parametric simulation scheme of bivariate GPD. To take into account non-stationarities, we also adapt the extended generalised Pareto model (Naveau et al., 2016), letting the parameters vary with specific sea state parameters.

Chapter 3 addresses the **problematic n°2**. We study the specific case of binary classifiers, which are the simplest type of forecasting and decision-making situation: an extreme event did or did not occur. Such classifier tailored for extremes will be called an extremal classifier and risk functions that answer our initial question will be developed. Their properties will be derived under the framework of multivariate regular variation and hidden regular variation, allowing to handle finer types of asymptotic independence.

All the codes used in this manuscript are produced with R (R Core Team, 2022).

# Chapter 1

## Elements of extreme value theory

### Overview of Chapter 1

In this Chapter, the main results of extreme value theory (EVT) are discussed. Since this theory entails many concepts and questions, a focus is made on the key elements that are useful for the rest of this thesis. The proofs are omitted and can be found in many great reference books such as [Coles \(2001\)](#), [Beirlant et al. \(2004\)](#) or [de Haan and Ferreira \(2007\)](#). The outline of the present chapter is as follows.

In Section 1.1, we give the foundations of univariate EVT. Starting from its key result that the generalised extreme value distributions are the only possible limits that can be obtained for the maximum of a random sample, under a suitable renormalisation, we introduce two modelling approaches for extremes: block maxima and peaks over thresholds. Then, a brief note is made on regular variation which is a powerful tool in EVT.

Section 1.2 is devoted to multivariate EVT. A focus is made on bivariate extremes, this enables us to avoid complex notations. Similarly to the univariate case, the bivariate extreme value distributions are introduced as the limit of component-wise maxima. Characterisations of this family are given with its spectral representation. Bivariate peaks over thresholds models are discussed, following the representation of [Rootzén and Tajvidi \(2006\)](#). Then a discussion is made on the dependence structure of bivariate extreme value distributions and we show that extreme value models fail for asymptotically independent variables. Therefore, the last section addresses this issue by looking at other types of models apart from the class of extreme value distributions.



This chapter was written with the help of the above-mentioned textbooks. Other references that have proved very valuable include [Fougères \(2004\)](#), [Reiss and Thomas \(2007\)](#) and the recent review of [Davison and Huser \(2015\)](#) which contains helpful animations that illustrate some of the following concepts.

In this chapter, we do not address in details the question of statistical inference methods. A comprehensive account can be found in [Beirlant et al. \(2004\)](#) or [Dey and Yan \(2016\)](#) among others.

## 1.1 Univariate extremes

As mentioned in the introduction of this dissertation, univariate extreme value theory (EVT) provides a solid theoretical basis and framework for studying the largest values of a sample. In this context different modelling methods coexist: block maxima,  $r$  largest order statistics, peaks over threshold and point process. In the following, we focus on the maxima and the threshold exceedances of random variables, see, e.g. [Coles \(2001\)](#) or [Dey and Yan \(2016\)](#) for details on the other approaches.

### 1.1.1 Possible limits of block maxima

The strength of EVT began in 1928 with the [Fisher and Tippett \(1928\)](#) Theorem (and later with [Gnedenko \(1943\)](#)). The idea was that if we consider a sequence of independent and identically distributed random variables  $X_1, \dots, X_n$  with common cumulative distribution function  $F$ , then there are only three possible limiting distributions for the maximum  $M_n := \max\{X_1, \dots, X_n\}$  after proper renormalisation. These three families are known as GEV (Generalised Extreme Value distributions).

**Theorem 1.1** ([Fisher and Tippett \(1928\)](#); [Gnedenko \(1943\)](#)). *If there exists sequences  $(a_n > 0)$  and  $(b_n)$  such that*

$$\lim_{n \rightarrow \infty} \mathbb{P}[(M_n - b_n)/a_n \leq x] = G(x), \quad (1.1)$$

where  $G$  is a non-degenerate distribution function. Then  $G$  belongs to the GEV family i.e. is of the form:

$$G_\xi(x) = \begin{cases} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} & \text{if } \xi \neq 0, x \in \{x; 1 + \xi(x - \mu)/\sigma > 0\}, \\ \exp \left\{ - \exp \left[ \left( \frac{x - \mu}{\sigma} \right) \right] \right\} & \text{if } \xi = 0, x \in \mathbb{R}, \end{cases}$$

for parameters  $\mu \in \mathbb{R}$  (location),  $\sigma > 0$  (scale) and  $\xi \in \mathbb{R}$  (shape).

In EVT, an important issue concerns the shape parameter  $\xi$ , which governs the behaviour of the upper-tail distribution (see Figure 1.1):

- If  $\xi > 0$ , the upper end-point of  $G_\xi$  is infinite and the distribution is heavy-tailed, i.e. the tail of the distribution decreases polynomially (Fréchet family).
- If  $\xi = 0$ , the upper end-point of  $G_\xi$  is also infinite but the distribution is rather light-tailed, i.e. the tail decreases exponentially (Gumbel family).
- If  $\xi < 0$ , the upper end-point of  $G_\xi$  is finite, the distribution tail is bounded (Weibull family). Furthermore, the upper end-point  $x_F$  can be linked to the parameters of the GEV through  $x_F = \mu - \sigma/\xi$ .

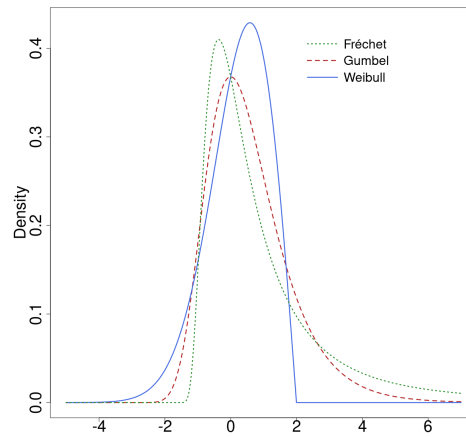


Figure 1.1: Densities of the three generalised extreme value distributions. For the Fréchet distribution  $\xi = 0.5$  and for the Weibull distribution  $\xi = -0.5$ . The shape parameter  $\sigma$  is equal to 1 and the location parameter  $\mu$  is equal to 0.

If equation (1.1) is satisfied then  $F$  (or  $X$  with distribution function  $F$ ) is said to be in the domain of attraction of the GEV distribution  $G_\xi$ , often denoted  $F \in \mathcal{D}(G_\xi)$ .

**Example 1.2.** For example, the Uniform distribution is in the Weibull domain of attraction, with  $\xi = -1$ . Indeed, let  $X_i \sim U(0, 1)$ , setting  $a_n = n^{-1}$  and  $b_n = 1$  we have:

$$\mathbb{P}[(M_n - 1)/n^{-1} \leq x] = F^n(1 + n^{-1}x) = (1 + n^{-1}x)^n \xrightarrow[n \rightarrow \infty]{} \exp(x).$$

Note that the Weibull family and the Fréchet family can be linked through

$$X \in \mathcal{D}(G_{\xi < 0}) \Leftrightarrow Y := (x_F - X)^{-1} \in \mathcal{D}(G_{\xi > 0}), \quad (1.2)$$

where  $G_{\xi < 0}$  (resp.  $G_{\xi > 0}$ ) is the Weibull (resp. Fréchet) extreme value distribution and  $x_F$  denotes the upper end-point of the distribution function  $F$  of  $X$  (see [Beirlant et al., 2004](#)). This correspondence is one of the reason why the EVT literature often focuses on the heavy-tail case. It is also justified from an application perspective. Indeed, there is an historical tradition between hydrology and EVT, and it is widely agreed that the distributions of hydrologic variables are heavy-tailed (e.g. [Katz et al., 2002](#)).

From a statistical point of view, however, Equation (1.2) is of no use since, in general, no prior information is available on the sign of  $\xi$ .

A convenient related concept to the extreme value distributions is *max-stability*. A distribution function  $F$  is *max-stable* if, for every  $n \in \mathbb{N}$ ,

$$F^n(b_n + a_n x) = F(x), \quad (1.3)$$

for a suitable choice of constants  $b_n$  and  $a_n > 0$ . And it can be shown that the classes of generalised extreme value distributions and max-stable distributions coincide (see, for example, [Beirlant et al. \(2004\)](#)).

A classification of the standard distributions according to their domain of attraction, with their respective normalising constants, can be found for example in [Embrechts et al. \(2013\)](#).

In practice, the normalising constants  $a_n$  and  $b_n$  are unknown and the maxima are modelled directly using the following approximation for a large enough block size  $n$

$$\mathbb{P}(M_n \leq x) \approx G[(x - b_n)/a_n] = G^*(x),$$

where  $G^*$  also belongs to the GEV family and differs from  $G$  only in the scale and location parameters.

The max-stability property (1.3) implies in particular that if block maxima with fixed block size  $n$  are well approximated by a GEV distribution  $G$ , then block maxima over a larger block size should also be well approximated by the GEV  $G$ , up to a change of scale and location. Indeed, consider  $m = \alpha n \geq n$ ,  $\alpha \in \mathbb{N}$ , then

$$\mathbb{P}(M_m \leq x) = F^m(x) = [F^n(x)]^\alpha = [\mathbb{P}(M_n \leq x)]^\alpha.$$

And from this simple equation, if  $F \in \mathcal{D}(G)$ , using the max-stability property (1.3), one can write

$$\mathbb{P}(M_m \leq x) \approx \{G[(x - b_n)/a_n]\}^\alpha = G[(x - b_{\alpha n})/a_{\alpha n}].$$

From Theorem 1.1, and the preceding remark, one can derive directly the first approach for modelling extreme events which is the so-called *block maxima* approach (see Figure 1.4). For that, the data will be divided into blocks of equal sizes (typically blocks of one year or one month), we then compute the maxima for each block and we can then fit a GEV distribution to the set of block maxima. Note that

the size of the block implies a bias-variance trade-off: if the block size is too small, the approximation from the limit in Theorem 1.1 will be bad inducing higher bias; if the block size is too big, we will have too few datapoints, leading to higher variance.

### 1.1.2 Peaks over thresholds approach

One limitation of the block maxima approach is if many extremes occur in the same block, only one event will be kept, which may lead to a significant loss of data. The second approach, named *peaks over thresholds*, is an alternative where instead of considering maxima per block, we will look at all events, or *peaks*, above a given - usually high - threshold. This approach relies on the following theorem where, with a similar emphasis, it states that the only possible limiting distributions of the threshold excesses is the generalised Pareto family (GP).

**Theorem 1.3** (Balkema and de Haan (1974); Pickands (1975)). *For every  $\xi \in \mathbb{R}$ ,  $X$  is in the domain of attraction of a GEV distribution  $G_\xi$  if, and only if, the distribution function of the excesses  $X - u$ , conditionally on  $X > u$ , converges as follows*

$$\lim_{u \rightarrow x_F} \sup_{0 < x < x_F - u} |\mathbb{P}(X - u \leq x \mid X > u) - H_{\xi, \tilde{\sigma}(u)}(x)| = 0, \quad (1.4)$$

for some positive function  $\tilde{\sigma}$ , where  $x_F$  is the upper end-point of  $F$  and  $H_{\xi, \tilde{\sigma}(u)}$  is the generalised Pareto distribution function and defined as

$$H_{\xi, \tilde{\sigma}(u)}(x) := 1 - (1 + \xi x / \tilde{\sigma}(u))^{-1/\xi}.$$

If Theorem 1.3 holds true, then the shape parameter  $\xi$  is the same for the GEV distribution  $G_\xi$  and the GP distribution  $H_{\xi, \tilde{\sigma}(u)}$ , and  $\tilde{\sigma}(u) = \sigma + \xi(u - \mu)$ .

The different behaviours of the tail distribution given the shape parameter  $\xi$  are the same as for the GEV family (see Figure 1.2), where for  $\xi = 0$ , taking the limit  $\xi \rightarrow 0$  in (1.4), we have

$$H_{0, \tilde{\sigma}(u)}(x) = 1 - \exp(-x / \tilde{\sigma}(u)).$$

Similarly to the notion of max-stability, regarding peak over thresholds a useful property is the *threshold stability* (e.g. Falk et al., 2010). A distribution  $F$  is *threshold stable* (sometimes called POT-stable) if, for every  $u \in [0, x_F]$ , with  $x_F$  the upper-end point of  $F$ ,

$$F^{[u]}(b_u + a_u x) = F(x), \quad (1.5)$$

for a suitable choice of constants  $b_u$  and  $a_u > 0$ .  $F^{[u]}$  denotes the distribution of exceedances above the threshold  $u$  and is defined by

$$F^{[u]}(x) = \frac{F(x) - F(u)}{1 - F(u)}, \quad x \geq u.$$

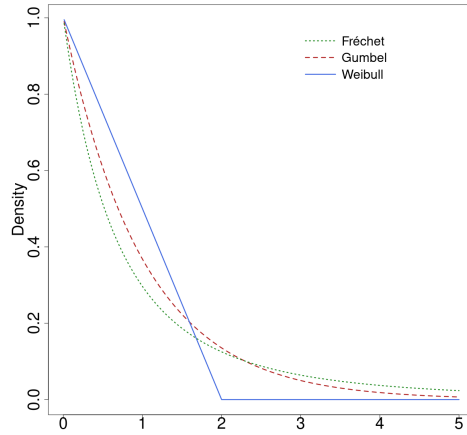


Figure 1.2: Densities of the GP distribution for different values of the shape parameter  $\xi$ . For the Fréchet family  $\xi = 0.5$  and for the Weibull family  $\xi = -0.5$ . The scale parameter  $\tilde{\sigma}$  is equal to 1.

Note that if  $X$  has distribution function  $F$ , then the threshold exceedances  $X - u \mid X > u$  have distribution function  $F^{[u]}$ :

$$\mathbb{P}(X - u \leq x \mid X > u) = F^{[u]}(x + u).$$

The following proposition shows that the class of threshold stable distributions coincides with the class of GP distributions (see, e.g. [Falk et al. \(2010\)](#)).

**Proposition 1.4.** *Let  $F$  be a distribution function with upper-end point  $x_F$ .*

1. *If there exists functions  $a(u) > 0$  and  $b(u)$  such that*

$$\lim_{u \rightarrow x_F} F^{[u]}(b(u) + a(u)x) = L(x), \quad (1.6)$$

*where  $L$  is a non-degenerate distribution function, then  $L$  is threshold stable.*

2. *Conversely, if  $L$  is a threshold stable distribution, then  $L$  satisfies (1.6).*

*If, in addition to being threshold stable,  $L$  is continuous, then  $L$  is a GP distribution, up to a change of scale and location.*

**Example 1.5.** *Considering the Uniform distribution as in Example 1.2, its conditional threshold exceedances correspond to the generalised Pareto distribution with  $\xi = -1$  and  $\tilde{\sigma}(u) = 1 - u$*

$$\mathbb{P}(X - u \leq x \mid X > u) = \frac{F(x + u) - F(u)}{1 - F(u)} = \frac{(x + u) - u}{1 - u} = \frac{x}{1 - u}.$$

From Theorem 1.3 we can derive the second approach to model extreme events, the so-called *peaks over thresholds* model (see Fig. 1.4). Instead of splitting data into blocks, we will choose a high enough threshold  $u$  and then fit a GP distribution to the thresholds exceedances. As for the *block maxima* approach with the choice of the block length, the choice of the threshold  $u$  implies a bias-variance trade-off. For many applications,  $u$  will typically be chosen as a high quantile of the data of interest.

Similarly to the max-stability for the block maxima approach, the threshold stability property (1.5) implies in particular that if threshold exceedances over a given threshold  $u$  are well approximated by a GP distribution  $H$ , then the threshold exceedances over a higher threshold will also be well approximated by the GP distribution  $H$ , up to a change of scale and location. To see this, consider a higher threshold  $v = \alpha + u > u$ ,  $\alpha > 0$ , then

$$\mathbb{P}(X - v > x \mid X > v) = \frac{\mathbb{P}(X > x + v)}{\mathbb{P}(X > u)} \times \frac{\mathbb{P}(X > u)}{\mathbb{P}(X > v)} = \frac{\mathbb{P}(X - u > x + \alpha \mid X > u)}{\mathbb{P}(X - u > \alpha \mid X > u)}.$$

Then, if the exceedances  $X - u \mid X > u$  are approximated by a GP distribution  $H$ , using the threshold stability of  $H$ , one can write

$$\mathbb{P}(X - v \leq x \mid X > v) \approx \frac{H(x + \alpha) - H(\alpha)}{1 - H(\alpha)} = H((x + \alpha - b_u)/a_u).$$

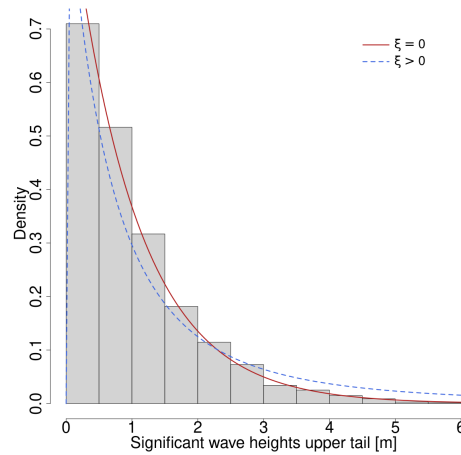


Figure 1.3: Upper tail distribution of significant wave heights ( $H_s$ ), corresponding to the threshold exceedances above the 0.98 quantile of  $H_s$ . GP density with shape parameter  $\xi = 0$  (resp.  $\xi > 0$ ) is superimposed as a red solid line (resp. blue dashed line).

An example of peaks over threshold modelling is given in Figure 1.3, with the wave data used in Chapter 2. As mentioned in the Introduction, in this chapter,

we will consider significant wave height, denoted  $H_s$ . This quantity basically give information about the ocean surface roughness, more details are provided in Section 2.1.1. Figure 1.3 depicts the distribution of the threshold exceedances of  $H_s$  given that  $H_s > u$ , where  $u$  is the 0.98 quantile of  $H_s$ . Two GP densities are superimposed, showing that the  $H_s$  exceedances should be rather well approximated by a light-tailed distribution. This *a priori* analysis will be confirmed in Chapter 2, where the estimated shape parameters for  $H_s$  data in two locations will be slightly negative (see Table 2.2).

To avoid a careful threshold selection, [Naveau et al. \(2016\)](#) introduced an extension of the GP distribution, the so-called extended generalised Pareto distribution (EGPD). In a nutshell, the EGPD allows to model the entire range of a distribution with a smooth transition between the lower and the upper tails. This model is introduced in Section 2.1.3 and it has been applied to the wave data in [Legrand et al. \(2022\)](#), see Chapter 2.

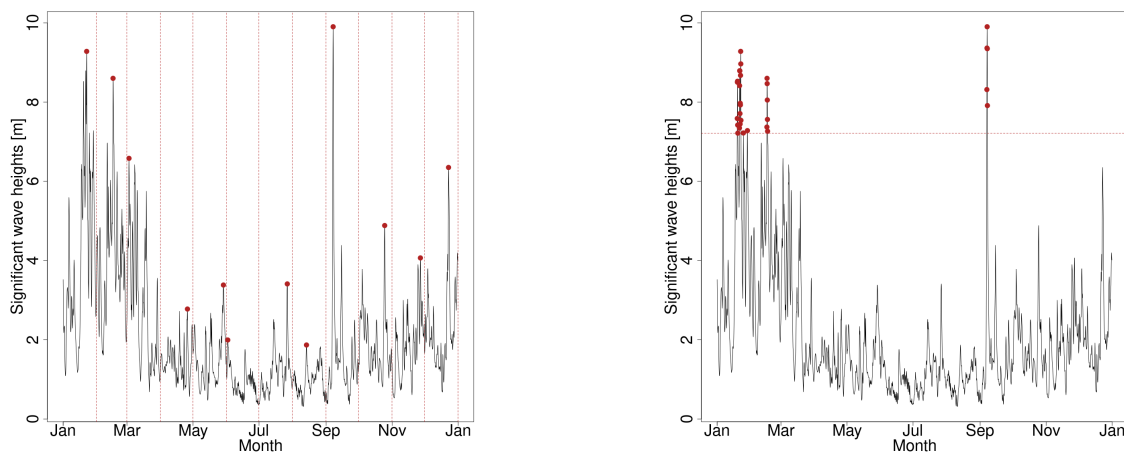


Figure 1.4: Illustration of the *block maxima* versus the *peaks over thresholds* approaches with one year of the wave data described in Chapter 2, Section 2.1.1.

### To sum up

To model extreme events, two main strategies can be used:

1. The *block maxima* approach which relies on the approximation of renormalised maxima by a GEV distribution (Theorem 1.1).  
 $\rightsquigarrow$  Needs to choose the size of the blocks.
2. The *peaks over thresholds* approach which relies on the approximation of the thresholds exceedances by a GP distribution (Theorem 1.3).  
 $\rightsquigarrow$  Needs to choose the threshold.

### 1.1.3 Regular variation and domain of attraction

When studying EVT, one theoretical tool that quickly comes into play is *regular variation*. Here we briefly introduce the key concepts and show how regular variation can be used to determine domains of attraction. See, e.g. [Bingham et al. \(1987\)](#) for more details.

**Definition 1.6.** A positive measurable function  $f$  is regularly varying (r.v.) with index  $\alpha \in \mathbb{R}$  if

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha, \text{ for all } x > 0.$$

If  $\alpha = 0$ ,  $f$  is called a slowly varying function.

**Example 1.7.** Examples of regularly varying functions are: all the functions converging to a positive constant,  $x \mapsto x^\alpha$  for  $\alpha \in \mathbb{R}$ ,  $x \mapsto \log^\beta(x)$  for  $\beta \in \mathbb{R}$ .

In the following, we consider regularly varying survival functions  $1 - F$  of a random variable  $X$ . For this, an additional constraint is required by imposing  $\alpha > 0$ . This can be seen through the following convergence, assuming that  $1 - F$  is r.v. with index  $\alpha$ ,

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t \rightarrow \infty} \frac{\mathbb{P}(X > tx)}{\mathbb{P}(X > t)} = x^{-\alpha}, \quad (1.7)$$

then necessarily,  $\alpha$  must be strictly positive.

Thanks to the following result, due to [Gnedenko \(1943\)](#), the domains of attraction of the Fréchet family and the Weibull family can then be characterised in terms of regular variation.

**Theorem 1.8 (Gnedenko (1943)).** In the following,  $G_{\xi > 0}$  (resp.  $G_{\xi < 0}$ ) denotes the Fréchet (resp. Weibull) family distribution and  $x_F$  denotes the upper end-point of  $F$ .



- $F \in \mathcal{D}(G_{\xi>0})$  iff  $1 - F$  is regularly varying (r.v.) with index  $1/\xi$ .
- $F \in \mathcal{D}(G_{\xi<0})$  iff  $x_F < \infty$  and  $1 - F(x_F - 1/x)$  is r.v. with index  $-1/\xi$ .

Note that using the link between the Weibull and the Fréchet families provided in Equation (1.2), only the first statement of Theorem 1.8 is needed.

**Example 1.9.** Every Pareto-like distributions, i.e. such that  $1 - F(x) \sim Cx^{-\alpha}$  for large  $x$  and for some  $C, \alpha > 0$ , are in the domain of attraction of the Fréchet family. For example the Pareto, Cauchy, and Burr distributions.

From this characterisation theorem, we see that the theory of regular variation is an appropriate framework to study heavy-tailed distributions, bringing many valuable theoretical tools (e.g. [Resnick, 2007](#)). In Chapter 3, we will provide the multivariate extension of regular variation and this theoretical framework will be applied in particular to extreme river discharges, which are heavy-tailed data.

## 1.2 Multivariate extremes

The extension from univariate extremes to multivariate extremes is not straightforward, partly because it implies a point of view choice. For example, considering temperature data in different cities, one could say that an extreme (multivariate) event occurs if large values of temperature in one of the cities is observed, but it could also be that large values in all cities occur simultaneously. In the bivariate case, an illustration of these two approaches is given in Figure 1.5 with daily river discharges from the Danube river network (more details on this dataset are given in Chapter 3). The underlying question here could then be the following: Do we want to assess the risk of flooding at one (or more) location or are we more interested in assessing the overall risk of flooding - which increases if extreme river discharges occur at different locations simultaneously?

The second key aspect of multivariate extremes is the dependence structure between the different components in the asymptotic. For large values, the dependence structure between the components of a multivariate distribution can be disrupted. An example developed by [Sibuya \(1960\)](#) is the bivariate Gaussian distribution, for which, whenever the correlation coefficient  $\rho$  is strictly smaller than 1, its components are independent in the extremes (see Figure 1.6).

To focus on the dependence structure, the standard method is to transform the data to common margins which removes marginal effects. This is what is done in Figure 1.6, where the data are transformed to common Pareto margins (for example). Note that this transformation does not alter the dependence structure. Once the marginal effects are removed, it is then easier to seek for dependence or independence. This two step study can be elegantly illustrated through the copula theory and the following theorem:

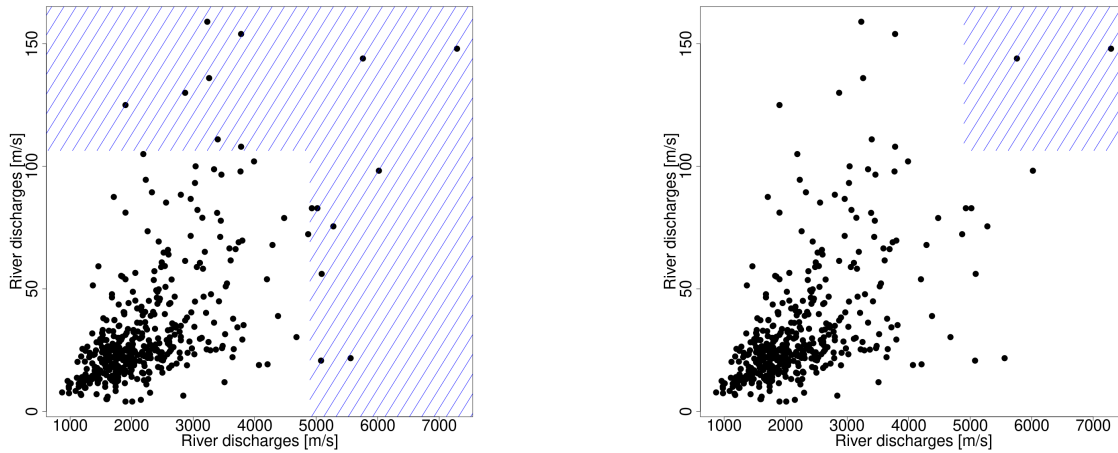


Figure 1.5: Scatter plots of daily river discharges of two stations in the Danube river network. The two possible regions of extremes are depicted in blue: when an extreme event is defined as a large value observed in at least one of the components (left) and when it is defined as a large value in both components simultaneously (right).

**Theorem 1.10 (Sklar (1959)).** *Let  $F$  be a joint distribution function with margins  $F_X$  and  $F_Y$ . Then there exists a copula  $C^1$  such that, for all  $x, y \in \mathbb{R}$ ,*

$$F(x, y) = C \{F_X(x), F_Y(y)\}. \quad (1.8)$$

*If  $F_X, F_Y$  are continuous, then  $C$  is unique. Conversely, if  $C$  is a copula and  $F_X, F_Y$  are distribution functions, then the function  $F$  defined by (1.8) is a joint distribution function with margins  $F_X$  and  $F_Y$ .*

From Theorem 1.10 we see that the study of the joint distribution  $F$  can be decomposed into the study of its margins and of the dependence structure, via the copula  $C$ . More details on copula theory, and its links with EVT, can be found for example in [Nelsen \(2006\)](#).

On this basis and derived from theoretical and practical perspectives, the modelling of multivariate extreme events is in general carried out in two steps: the first step concerns the treatment of the margins and the second step the description of the dependence behaviour (e.g. [Cai et al., 2013](#); [Tendijck et al., 2021](#); [Rohmer et al., 2021](#)).

The marginal modelling can involve classic univariate EVT (see Section 1.1). Given the application, some models will be more adapted than others. For example, in Chapter 2 the margins are modelled using a specific type of extreme value distribution: the extended generalised Pareto distribution ([Naveau et al., 2016](#)). If

<sup>1</sup>A copula (in dimension 2) is a bivariate distribution function on  $[0, 1] \times [0, 1]$  with uniform margins in  $[0, 1]$  ([Nelsen, 2006](#)).

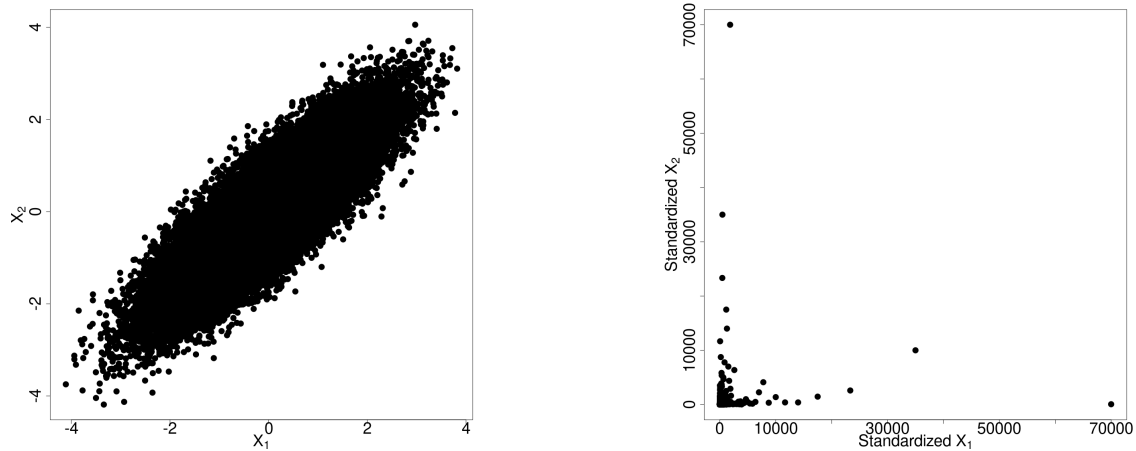


Figure 1.6: Illustration of asymptotic independence with a bivariate Gaussian vector on the original scale (left) and transformed to common Pareto scale (right).

there is no specific constraint on the marginal modelling, to facilitate the choice, [Beirlant et al. \(2004\)](#) suggest that when the main interest lies in the dependence structure, a convenient choice is to transform into standard Fréchet margins as there are good theoretical properties in this case.

Regardless of the chosen standardisation distribution, once the data are transformed to common margins, different theoretical objects can be defined to model and to measure the dependence at high levels. In the remaining of this section we will briefly introduce these tools and provide the links between them. We will restrict our attention to the bivariate case, generalisation to higher dimensions can be found in the references mentioned earlier.

To lighten the notations, in the following bold face symbols will denote bivariate vectors:  $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ ,  $\mathbf{0} = (0, 0), \dots$ . Operations and relations involving such vectors will be considered component-wise. For example  $\min(\boldsymbol{x}, \mathbf{0})$  will correspond to  $(\min(x_1, 0), \min(x_2, 0)) \in \mathbb{R}^2$ .

### 1.2.1 Domain of attraction and characterisation

The definition commonly adopted in the literature for the notion of maximum in the multivariate case is the *component-wise maxima*. If we assume that we have a set of bivariate vectors  $(X_1, Y_1), \dots, (X_n, Y_n)$  that are independent and identically distributed with common bivariate distribution function  $F$ , then the sample maximum  $\boldsymbol{M}_n$  is defined by

$$\boldsymbol{M}_n := \left( \max_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} Y_i \right).$$

Note that the sample maximum does not necessarily correspond to a sample point. Then, similarly to the univariate case, we can define the family of multivariate extreme value distributions as the only possible limit for the sample maxima after proper renormalisation. Recall that operations involving vectors are meant component-wise.

**Definition 1.11.** *If there exists sequences in  $\mathbb{R}^2$  ( $\mathbf{a}_n$ ) and ( $\mathbf{b}_n$ ) with  $a_{n,j} > 0$  for  $j = 1, 2$ , such that*

$$\lim_{n \rightarrow \infty} \mathbb{P}[(\mathbf{M}_n - \mathbf{b}_n)/\mathbf{a}_n \leq (x, y)] = G(x, y), \quad (1.9)$$

where  $G$  is a non-degenerate distribution function, then  $G$  belongs to the bivariate extreme value family and  $F$  is said to be in the (bivariate) domain of attraction of  $G$ , denoted  $F \in \mathcal{D}(G)$ .

This definition implies that, if  $\mathbf{X} = (X, Y) \in \mathcal{D}(G)$ , the marginal distribution  $F_X$  (resp.  $F_Y$ ) of  $X$  (resp.  $Y$ ) must be in the univariate domain of attraction of the GEV distribution  $G_1$  (resp.  $G_2$ ), where  $G_j$  denotes the  $j$ th marginal distribution of  $G$ , for  $j = 1, 2$ .

As in the univariate case, there is a multivariate version of max-stability:  $G$  is max-stable if

$$G^n(\mathbf{b}_n + \mathbf{a}_n \mathbf{x}) = G(\mathbf{x}),$$

for suitable vectors  $\mathbf{b}_n$  and  $\mathbf{a}_n > 0$ . And it can be proved that  $G$  is a bivariate extreme value distribution if, and only if,  $G$  is max-stable.

**Example 1.12.** *Consider a bivariate vector with unit Fréchet margins. Then in this case, taking  $\mathbf{b}_n = \mathbf{0}$  and  $\mathbf{a}_n = (n, n)$  for all  $n > 0$ ,*

$$\mathbb{P}[\mathbf{M}_n/\mathbf{a}_n \leq (x, y)] = F^n(n(x, y)) = (\exp(-1/(n(x, y)))^n = \exp(-1/(x, y)).$$

The standard Fréchet distribution is then in its own domain of attraction.

We now assume that both margins are transformed to standard Fréchet distributions (whose distribution function is given by  $x \mapsto \exp(-1/x)$  for  $x > 0$ ). Then the following theorem provides a characterisation of the bivariate extreme value family defined in Equation (1.9) (see [Fougères \(2004\)](#) or more specifically [Resnick \(1987\)](#) for the multivariate version).

**Theorem 1.13.** *The following statements are equivalent:*

- (i)  $G$  is a bivariate extreme value distribution with unit Fréchet margins.
- (ii) There exists a finite measure  $H$  on  $[0, 1]$  such that for each  $x, y > 0$ , one has that

$$G(x, y) = \exp \left\{ -2 \int_0^1 \max \left( \frac{w}{x}, \frac{1-w}{y} \right) dH(w) \right\}$$

with the mean constraints  $\int_0^1 dH(w) = 1$  and  $\int_0^1 w dH(w) = 1/2$ .

If we are in the framework of Theorem 1.13, the following measure is also described

$$V(x, y) := 2 \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w), \text{ for any } x, y > 0. \quad (1.10)$$

The measure  $V$  can be any positive function such that for  $t > 0$  and  $x, y > 0$ ,  $V(tx, ty) = t^{-1}V(x, y)$  (homogeneity property) and  $V(z, \infty) = V(\infty, z) = 1/z$  for any  $z > 0$  (marginal constraints).

**Definition 1.14.** *The measure  $H$  is known as the spectral measure and  $V$  is sometimes called the exponent measure of the extreme value distribution  $G$ .*

Instead of  $V$  defined in (1.10), the *stable tail dependence function* (e.g. [Beirlant et al., 2004](#)) is widely considered in the literature and is defined by

$$l(x, y) := V\left(\frac{1}{x}, \frac{1}{y}\right), \quad x, y \geq 0.$$

Other alternative and equivalent characterisations can be found. For example, the Pickands dependence function ([Pickands, 1981](#)) is defined as

$$A(t) := V\left(\frac{1}{1-t}, \frac{1}{t}\right), \text{ for any } t \in [0, 1]. \quad (1.11)$$

The *Pickands dependence function* can be linked to the spectral measure through

$$A(t) = 2 \int_0^1 \max\{w(1-t), (1-w)t\} dH(w).$$

From the marginal constraints on  $V$ ,  $A(0) = A(1) = 1$ . Moreover,  $A$  lies within the set  $\max(t, 1-t) \leq A(t) \leq 1$  for any  $t \in [0, 1]$  and is convex within this region.

The measure  $V$  is completely determined by the Pickands dependence function  $A$  through

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

for  $x, y > 0$  such that  $x + y > 0$ .

The spectral measure  $H$  can also be recovered from  $V$ , or equivalently  $A$ , see [Beirlant et al. \(2004\)](#) for more details, so that there are one-to-one mappings between  $H$ ,  $V$  and  $A$ .

**Example 1.15.** *Historically, the first parametric family of bivariate extremes is the logistic model ([Beirlant et al., 2004](#)). Its distribution  $G$  is given by*

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^\alpha\right\}, \quad x, y > 0, \quad (1.12)$$

for a parameter  $0 < \alpha \leq 1$ . The corresponding exponent measure is therefore

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

and it can be shown that the corresponding spectral measure  $H$  has associated spectral density  $h$  given by

$$h(w) = \frac{1-\alpha}{\alpha} \{w(1-w)\}^{1/\alpha-2} \{(1-w)^{1/\alpha} + w^{1/\alpha}\}^{\alpha-2}, \quad w \in (0, 1).$$

The behaviour of this family given  $\alpha$  will be discussed later in Section 1.2.3. Other parametric families can be found in Section 9.2.2 of [Beirlant et al. \(2004\)](#). Simulations from this model are displayed in Figure 1.7 for different values of the parameter  $\alpha$ .

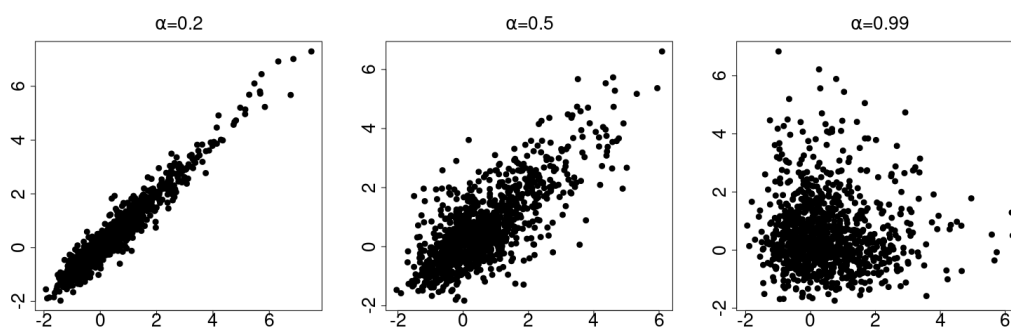


Figure 1.7: Simulations of  $n = 1000$  points from the bivariate logistic extreme value model (1.12) for different values of  $\alpha \in \{0.2, 0.5, 0.99\}$  (left to right).

Alternative ways of describing multivariate extremes to the ones presented here exist, see for instance [Segers \(2012\)](#) for further details.

### To sum up

Assuming that the bivariate extreme value distribution  $G$  has unit Fréchet margins, it can be equivalently characterised by:

1. the spectral measure  $H$

$$G(x, y) = \exp \left\{ -2 \int_0^1 \max \left( \frac{w}{x}, \frac{1-w}{y} \right) dH(w) \right\},$$

2. the exponent measure  $V$

$$G(x, y) = \exp(-V(x, y)),$$

3. the Pickands dependence function  $A$

$$G(x, y) = \exp \left[ - \left\{ \left( \frac{1}{x} + \frac{1}{y} \right) A \left( \frac{x}{x+y} \right) \right\} \right].$$

### 1.2.2 Bivariate threshold excess model

As for the univariate case, instead of considering maxima one can consider threshold exceedances. However, the extension to the multivariate case is not straightforward since it requires to define the conditioning event  $\mathbf{X} > \mathbf{u}$  in a multivariate way. For that, there is no universal choice on how to define multivariate exceedances. In this work, we follow the definition of [Rootzén and Tajvidi \(2006\)](#), saying that an observation is extreme if it is extreme in at least one of its components, but other definitions exist (e.g. [Falk and Guillou, 2008](#)).

Such event is denoted  $\mathbf{X} \not\leq \mathbf{u}$  and means that at least one component of  $\mathbf{X}$  is above a given (high) threshold, i.e.

$$\{\mathbf{X} \not\leq \mathbf{u}\} := \{X_1 > u_1 \text{ or } X_2 > u_2\}, \quad (1.13)$$

where  $\mathbf{u} = (u_1, u_2)$  and  $\mathbf{X} = (X_1, X_2)$ . The subset defined in (1.13) corresponds to the blue L-shaped region in the left-hand side of Figure 1.5 or the red L-shaped region in Figure 1.8.

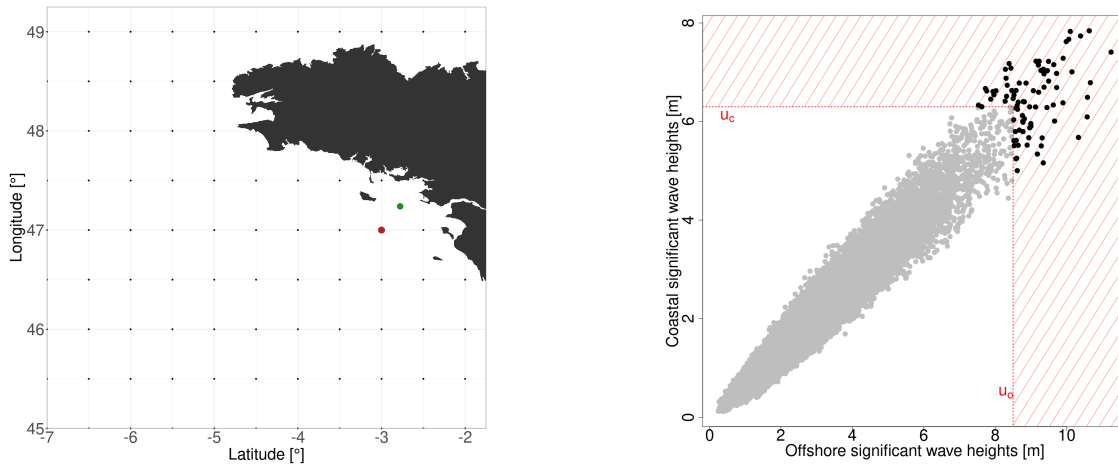


Figure 1.8: (left) Map illustrating the locations considered in Chapter 2. The red dot corresponds to the "offshore" point and the green dot corresponds to the "coastal" point. (right) Scatter plot of the significant wave heights at the two locations considered, more details on the data are provided in Chapter 2. The black dots belong to the region where the multivariate model will be fitted.

In Chapter 2, a joint simulation model of significant wave heights ( $H_s$ ) in two locations is developed. An application of such stochastic simulator could be on the

installation of wind farms in two relatively distant sites and the study of the extreme wave conditions that these sites could face simultaneously. In this context, the definition of [Rootzén and Tajvidi \(2006\)](#) appears to be the most appropriate since it allows for the possibility that the  $H_s$  are not necessarily extreme on both sites simultaneously (see Figure 1.8).

[Rootzén and Tajvidi \(2006\)](#) provide the following Definition 1.16 for the multivariate generalised Pareto distribution (stated here only in the bivariate case), which fulfils the two following motivating conditions for the extension from univariate to multivariate GP ([Rootzén and Tajvidi, 2006](#), p.919):

- (M1) *exceedances (of suitably coordinated levels) asymptotically have a multivariate GP distribution if and only if componentwise maxima asymptotically are EV distributed;*
- (M2) *the multivariate GP distribution is the only one which is preserved under (a suitably coordinated) change of exceedance levels.*

**Definition 1.16.** *A distribution function  $H$  is a bivariate generalised Pareto distribution if*

$$H(\mathbf{x}) := \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})}, \quad (1.14)$$

*for some bivariate extreme value distribution  $G$  with non-degenerate margins and with  $0 < G(\mathbf{0}) < 1$ . In particular,  $H(\mathbf{x}) = 0$  for  $\mathbf{x} < \mathbf{0}$  and  $H(\mathbf{x}) = 1 - \log G(\mathbf{x}) / \log G(\mathbf{0})$  for  $\mathbf{x} > \mathbf{0}$ .*

Considering the spectral representation of Theorem 1.13, when  $\mathbf{x} > \mathbf{0}$ ,  $H(\mathbf{x})$  can easily be linked to the spectral measure of the associated GEV distribution  $G$ , or similarly, to the Pickands dependence function  $A$  (see Propositions 5.3 and 6.1 of [Tajvidi \(1996\)](#)). In the following, we will provide the explicit link between the exponent measure  $V$  and the GP distribution in a specific case.

[Rootzén and Tajvidi \(2006\)](#) showed that the properties (M1) and (M2) are fulfilled with this definition of bivariate GP distributions thanks to the two following theorems. Theorem 1.17 answers the motivation (M1), and can be viewed as a multivariate extension of Theorem 1.3. The second theorem, Theorem 1.18, is the analogue of the threshold stability property in the univariate case (see Proposition 1.4), and answers the above motivation (M2).

Hereinafter we consider a threshold function defined as an increasing continuous curve  $\mathbf{u}(t)$  parameterised by  $t \in [1, \infty)$ , i.e. the thresholds are defined from the set  $\{\mathbf{u}(t) = (u_1(t), u_2(t)) : t \in [1, \infty)\}$ . Here "increasing" means that if  $t > s$ , with  $t, s \in [1, \infty)$ ,  $\mathbf{u}(t) > \mathbf{u}(s)$ . Recall that this inequality is meant component-wise, i.e. it corresponds to  $u_1(t) > u_1(s)$  and  $u_2(t) > u_2(s)$ . An illustration is given in Figure 1.9. This parametrisation is needed to specify how the 2-dimensional levels increase as we look further into the tail of the bivariate distribution  $F$  of  $\mathbf{X}$  (see [Tajvidi, 1996](#), Chapter 4).



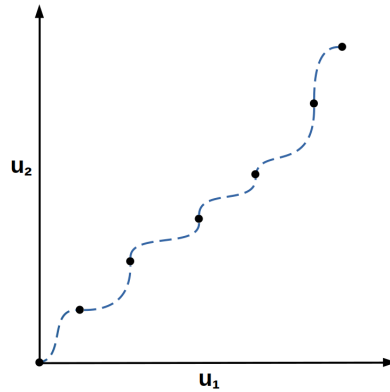


Figure 1.9: Illustration of the curve  $\mathbf{u}(t) = (u_1(t), u_2(t))$ .

**Theorem 1.17** (Rootzén and Tajvidi (2006)). *Let consider  $\mathbf{X} = (X_1, X_2)$  a bivariate random vector with distribution function  $F$ .*

1. *If  $F$  is in the domain of attraction of a bivariate extreme value distribution  $G$  with  $0 < G(\mathbf{0}) < 1$ , then there exists an increasing continuous curve  $\mathbf{u}(t)$ , starting at  $\mathbf{u}(1) = \mathbf{0}$ , with  $\lim_{t \rightarrow \infty} F(\mathbf{u}(t)) = 1$  and a function  $\sigma(\mathbf{u}(t)) > \mathbf{0}$ , such that*

$$\lim_{t \rightarrow \infty} \mathbb{P} \left\{ \sigma(\mathbf{u}(t))^{-1} (\mathbf{X} - \mathbf{u}(t)) \leq \mathbf{x} \mid \mathbf{X} \not\leq \mathbf{u}(t) \right\} = \frac{1}{-\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})},$$

for all  $\mathbf{x}$ .

2. *Conversely, if there exists an increasing continuous curve  $\mathbf{u}$  with  $\lim_{t \rightarrow \infty} F(\mathbf{u}(t)) = 1$  and a function  $\sigma(\mathbf{u}(t)) > \mathbf{0}$  such that, for  $\mathbf{x} > \mathbf{0}$ ,*

$$\lim_{t \rightarrow \infty} \mathbb{P} \left\{ \sigma(\mathbf{u}(t))^{-1} (\mathbf{X} - \mathbf{u}(t)) \leq \mathbf{x} \mid \mathbf{X} \not\leq \mathbf{u}(t) \right\} = H(\mathbf{x}), \quad (1.15)$$

for some function  $H$ , where the marginals of  $H$  on  $\mathbb{R}_+$  are non-degenerate. Then the left-hand side of (1.15) converges to a limit  $H(\mathbf{x})$  for all  $\mathbf{x}$  and there is a unique bivariate extreme value distribution  $G$  with  $G(\mathbf{0}) = e^{-1}$  such that

$$H(\mathbf{x}) = \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})}, \quad (1.16)$$

$G(\mathbf{x}) = \exp \{H(\mathbf{x}) - 1\}$  for  $\mathbf{x} > \mathbf{0}$ , and  $F \in \mathcal{D}(G)$ .

Note that the two representations (1.14) and (1.16) of the bivariate GP distributions coincide: letting  $t = 1/(-\log G(\mathbf{0}))$  in (1.14),

$$H(\mathbf{x}) = \log \frac{G(\mathbf{x})^t}{G(\mathbf{x} \wedge \mathbf{0})^t}$$

and by max-stability of  $G$ ,  $G^t$  is again an extreme value distribution with  $G(\mathbf{0})^t = \exp(-1)$ .

**Theorem 1.18** (Rootzén and Tajvidi (2006)). *Again, let consider  $\mathbf{X} = (X_1, X_2)$  a bivariate random vector.*

1. *If  $\mathbf{X}$  has a bivariate GP distribution  $H$ , then there exists an increasing continuous curve  $\mathbf{u}(t)$  with  $\mathbf{u}(1) = \mathbf{0}$  and  $\lim_{t \rightarrow \infty} H(\mathbf{u}(t)) = 1$ , and a function  $\boldsymbol{\sigma}(\mathbf{u}(t)) > \mathbf{0}$  such that*

$$\mathbb{P}[\boldsymbol{\sigma}(\mathbf{u}(t))^{-1}(\mathbf{X} - \mathbf{u}(t)) \leq \mathbf{x} \mid \mathbf{X} \not\leq \mathbf{u}(t)] = H(\mathbf{x}), \quad (1.17)$$

*for  $t \in [1, \infty)$  and all  $\mathbf{x}$ .*

2. *Conversely, if there exists an increasing continuous curve  $\mathbf{u}(t)$  with  $\mathbf{u}(1) = \mathbf{0}$  and  $\lim_{t \rightarrow \infty} \mathbb{P}[\mathbf{X} < \mathbf{u}(t)] = 1$ , and a function  $\boldsymbol{\sigma}(\mathbf{u}(t)) > \mathbf{0}$  such that (1.17) holds for  $\mathbf{x} > \mathbf{0}$ , and  $\mathbf{X}$  has non-degenerate margins, then  $\mathbf{X}$  has a bivariate GP distribution.*

The marginals of a bivariate GP distribution are typically not univariate GP. However, their restriction to the positive subset are:

$$\mathbb{P}(X_j \leq x \mid X_j > 0) = 1 - (1 + \xi_j x / \sigma_j)^{-1/\xi_j},$$

for  $x \geq 0$  such that  $\sigma_j + \xi_j x > 0$  and where  $\xi_j$  and  $\sigma_j$ , for  $j = 1, 2$ , are the marginal shape and scale parameters. For any bivariate GP distribution  $H$ , we therefore consider its associated shape parameter  $\boldsymbol{\xi}$  as the vector of the marginal parameters  $(\xi_1, \xi_2)$ , and, similarly, its scale parameter  $\boldsymbol{\sigma} := (\sigma_1, \sigma_2)$ . We write  $GP(\boldsymbol{\sigma}, \boldsymbol{\xi})$  the GP distribution  $H$  with parameters  $\boldsymbol{\sigma}$  and  $\boldsymbol{\xi}$ , and the *standard* GP distribution is defined for  $\boldsymbol{\xi} = \mathbf{1}$  and  $\boldsymbol{\sigma} = \mathbf{1}$ .

Any bivariate vector  $\mathbf{X} = (X_1, X_2) \sim GP(\boldsymbol{\sigma}, \boldsymbol{\xi})$  can be characterised through the *standard* GP distribution, as follows

$$\mathbf{X} = \boldsymbol{\sigma} \frac{e^{\boldsymbol{\xi} \mathbf{Z}} - 1}{\boldsymbol{\xi}}, \quad (1.18)$$

where  $\mathbf{Z} \sim GP(\mathbf{1}, \mathbf{0})$ . From Equation (1.18), the study of bivariate GP distributions can then be reduced to the study of standard form GP distributions. Rootzén et al. (2018b,a) introduce a number of stochastic representations of such standardised GP random vectors. One of them is as follows.

Let  $\mathbf{T}$  be a bivariate random vector and let  $E$  be a unit exponential random variable independent of  $\mathbf{T}$ . Then

$$E + \mathbf{T} - \max(\mathbf{T}) \sim GP(\mathbf{1}, \mathbf{0}). \quad (1.19)$$

Conversely, any standard GP vector can be expressed in this way. To emphasise the dependence on the bivariate vector  $\mathbf{T}$ , it may be better to write  $GP(\mathbf{1}, \mathbf{0}, \mathbf{T})$  instead of  $GP(\mathbf{1}, \mathbf{0})$ .

Using Definition 1.16, we can consider the exponent measure  $V$ , as defined in (1.10), of the GEV  $G$  associated to the GP distribution  $GP(\mathbf{1}, \mathbf{0}, \mathbf{T})$ . Then Rootzén et al. (2018a) showed that  $V$  and  $\mathbf{T} = (T_1, T_2)$  can be linked through

$$V(x_1, x_2) = \mathbb{E} \left\{ \max \left[ \frac{1}{x_1} \frac{e^{T_1 - \max(\mathbf{T})}}{\mathbb{E}(e^{T_1 - \max(\mathbf{T})})}, \frac{1}{x_2} \frac{e^{T_2 - \max(\mathbf{T})}}{\mathbb{E}(e^{T_2 - \max(\mathbf{T})})} \right] \right\}.$$

From (1.19), it appears that to construct bivariate GP models, a possibility is to assume a parametric distribution on  $\mathbf{T}$ . This question was addressed by Kirilouk et al. (2019) where different explicit density formulas for  $\mathbf{T}$  are derived. Two examples of such models are given in Figure 1.10 with  $\mathbf{T}$  that follows a bivariate Gaussian distribution and, on the other hand,  $T_1$  and  $T_2$  being two independent Gumbel-distributed variables.

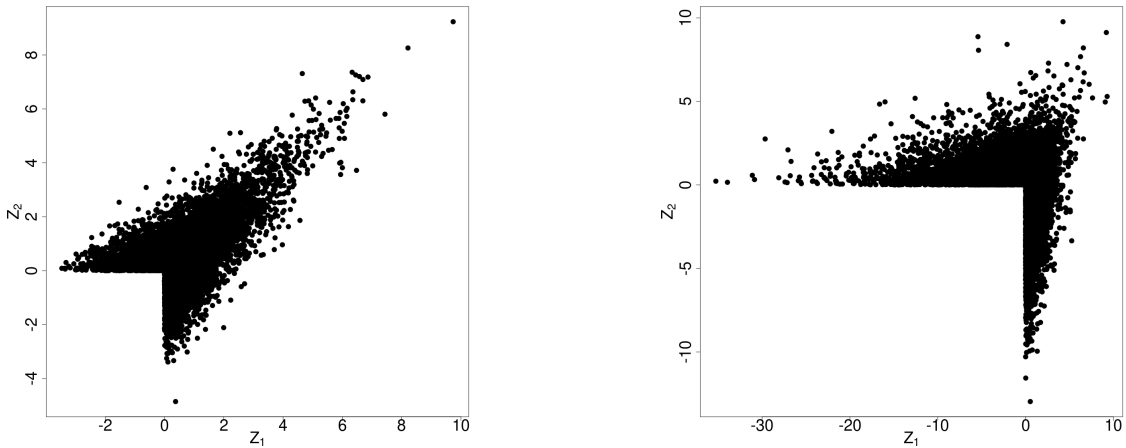


Figure 1.10: Simulations of bivariate GP models from the representation (1.19) with (left)  $\mathbf{T}$  following a Gaussian distribution with zero-mean and covariance matrix  $\Sigma = [(1, 0.4), (0.4, 1)]$  and (right)  $\mathbf{T} = (T_1, T_2)$  where  $T_1$  and  $T_2$  are independent variables with  $T_1 \sim \text{Gumbel}(0, 1)$  and  $T_2 \sim \text{Gumbel}(0, 4)$ .

In Chapter 2, we will built on model (1.19) to develop a non-parametric generator of such standard GP vectors, without assuming any specific parametric shape on  $\mathbf{T}$ .

Finally, from a practical modelling perspective, a similar approach to the univariate case can be used as motivated by Theorem 1.17. Indeed, for a sufficiently large threshold  $\mathbf{u}$  (i.e. large in each component),  $\mathbf{X} - \mathbf{u} \mid \mathbf{X} \not\leq \mathbf{u}$  can be well approximated by a bivariate GP distribution. As this statistical modelling approach

is still at an early stage, few papers deal with practical applications of multivariate GP models (e.g. [Brodin and Rootzén, 2009](#); [Kiriliouk et al., 2019](#); [Kiriliouk and Naveau, 2020](#), ...) following the representation of [Rootzén and Tajvidi \(2006\)](#).

### To sum up

Bivariate GP distributions emerge as the limit distributions of bivariate threshold excesses, defined as  $\mathbf{X} - \mathbf{u} \mid \mathbf{X} \not\leq \mathbf{u}$ .

One construction of standard forms for such bivariate GP vectors is given by

$$E + \mathbf{T} - \max(\mathbf{T}),$$

where  $E \sim \text{Exp}(1)$  and  $\mathbf{T}$  is any bivariate vector independent of  $E$ .

### 1.2.3 Asymptotic dependence and independence: Definitions and characterisations.

As already mentioned, an important feature in multivariate EVT is to describe the dependence behaviour of two variables as they become larger and larger. This question has many applications, among which we mention the following two.

- Among the components of a multivariate random vector, identifying subgroups such that large values occur simultaneously allows to reduce the dimension of the model by removing all the other components that do not contribute to the joint extremal behaviour. This is related to the notion of sparsity for extreme value, a very active research field (e.g. [Lehtomaa and Resnick, 2020](#); [Engelke and Ivanovs, 2020](#); [Meyer and Wintenberger, 2021](#)). This issue will also be addressed in a more humble way in Section 3.7.2.
- Another typical application of multivariate EVT is the probability estimation of failure regions, i.e.  $\mathbb{P}[(X, Y) \in \mathcal{D}]$ , where  $\mathcal{D}$  is a remote set (with no observations). This can simply correspond to the subsets  $\{X > x, Y > y\}$  or  $\{X > x \text{ or } Y > y\}$  for large  $x, y$  (see Figure 1.5), or it can also correspond to more sophisticated sets given some specific constraints such as environmental loads (e.g. [de Haan and de Ronde, 1998](#); [Beersma and Buishand, 2004](#); [Cai et al., 2013](#); [Ewans and Jonathan, 2014](#)). In this context, wrongly assuming the dependence structure between the large values of  $X$  and  $Y$  will lead to an overestimation or an underestimation of the probability of failure region (see the aforementioned references).

Given the data at hand, the dependence structure in the extremes can be very different. For example, looking at Figure 1.11, which compares the data used in Chapter 2 with those used in Chapter 3, the dependence behaviour between the

largest values in the extreme region, symbolised by the upper blue hatched quadrant, appears to be quite different. Regarding the wave data, large values mainly occur simultaneously in both variables, whereas in the case of the river data, a large value in one component does not always corresponds to a large value in the other component.

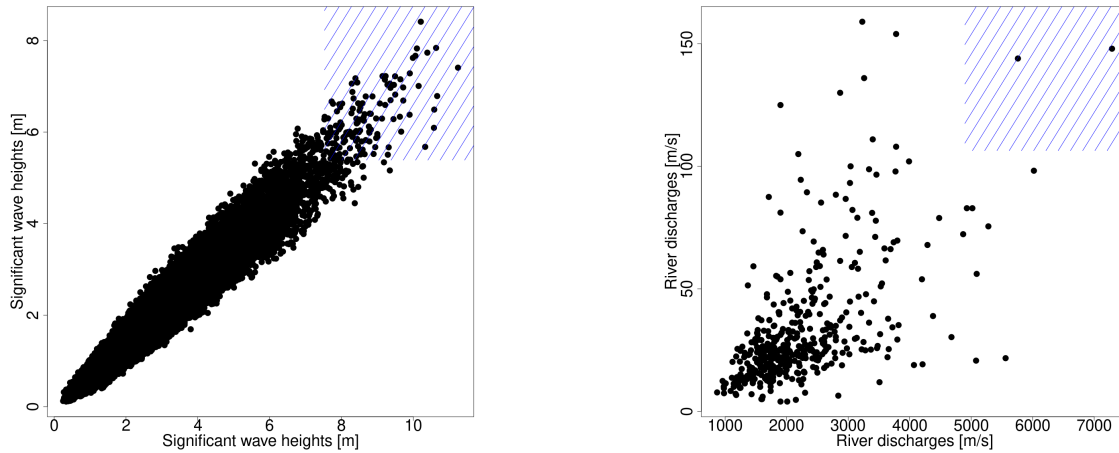


Figure 1.11: Asymptotic dependence behaviour for the different application data sets considered in this thesis. (left) Scatter plot between the significant wave heights in the two locations of interest (see Chapter 2). (right) Scatter plot between the river discharges in two distant stations (see Chapter 3). The blue hatched region is just a visual indication of where both components are large simultaneously.

The limiting extreme value distribution  $G$  provides this type of information, i.e. the dependence structure of the variables  $X$  and  $Y$ . In the following, we keep the previous notations, i.e. the bivariate distribution function  $F$  of a vector  $(X, Y)$  is in the domain of attraction of a bivariate extreme value distribution  $G$  characterised by a spectral measure  $H$ , an exponent measure  $V$  or, equivalently, a Pickands dependence function  $A$ .

The random variables  $X$  and  $Y$  are said to be *asymptotically independent* if, equivalently,

- $H$  puts mass  $1/2$  on the boundaries  $w = 0$  and  $w = 1$ ,
- $V(x, y) = 1/x + 1/y$  for any  $x, y > 0$ ,
- $A(t) = 1$  for any  $t \in [0, 1]$ .

These properties correspond to be in the domain of attraction of the independence, i.e. the limiting extreme value distribution decomposes into

$$G(x, y) = G_1(x) \times G_2(y),$$

where  $G_1, G_2$  are the marginal distributions of  $G$ .

Conversely,  $X$  and  $Y$  are said to be *completely dependent* if, equivalently,

- $H$  puts mass 1 at the barycenter  $w = 1/2$ ,
- $V(x, y) = \max(1/x, 1/y)$  for any  $x, y > 0$ ,
- $A(t) = \max(t, 1 - t)$  for any  $t \in [0, 1]$ .

**Example 1.19.** Continuing with Example 1.15, Figure 1.12 depicts the spectral measure  $H$  and the Pickands dependence function  $A$  for the bivariate logistic model with different values of  $\alpha$ . Figure 1.12 shows that  $\alpha$  measures the strength of dependence between the two components. In particular, when  $\alpha \rightarrow 0$  in (1.12),

$$G(x, y) \rightarrow \exp \left\{ - \max \left( x^{-1}, y^{-1} \right) \right\},$$

which corresponds to complete dependence, and when  $\alpha \rightarrow 1$ ,

$$G(x, y) \rightarrow \exp \left\{ - \left( x^{-1} + y^{-1} \right) \right\},$$

and this corresponds to asymptotic independence.

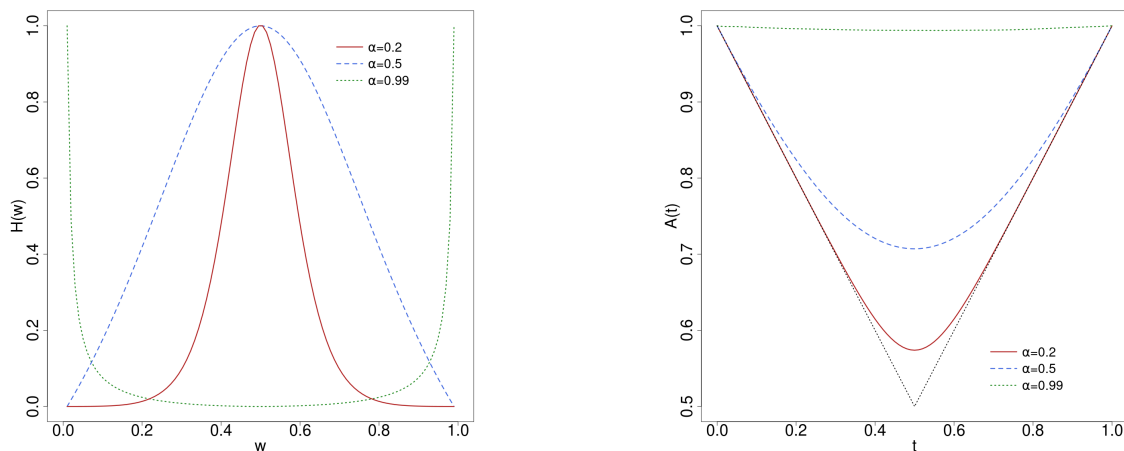


Figure 1.12: (left) Spectral measure  $H$  and (right) Pickands dependence function  $A$  for the logistic model defined in (1.12) (Coles, 2001), with  $\alpha = 0.2$  (red solid line),  $\alpha = 0.5$  (blue dashed line) and  $\alpha = 0.99$  (green dotted line). The spectral measures on the left panel have been normalised between 0 and 1 for visual purposes. On the right panel, the black dotted line represents the lower bound of  $A$ . Estimations are performed using the R packages `evd` (Stephenson, 2002) for  $H$  and `copula` (Hofert et al., 2020) for  $A$ .

In practice, asymptotic independence means that the largest values between the components of a random vector are unlikely to occur simultaneously. If this

is not the case, i.e. large values occur simultaneously with a high probability, we say that  $X$  and  $Y$  are asymptotically dependent. The goal is then to determine the strength of this dependence.

From Theorem 1.13 and the spectral representation, all the dependence structure of the limiting distribution is contained in the spectral measure  $H$ . From this, it can be very useful to have a summary of this information. Different coefficients have been proposed to measure the strength of asymptotic dependence between two variables. The best-known is probably the *dependence measure*  $\chi$  (Coles et al., 1999), defined as

$$\chi := \lim_{x \rightarrow x_F} \mathbb{P}(X > x \mid Y > x) \quad (1.20)$$

A generalisation in the case where the marginal distributions of  $X$  and  $Y$  are non-identical is given by

$$\chi := \lim_{u \rightarrow 1} \mathbb{P}(F_X(X) > u \mid F_Y(Y) > u).$$

The coefficient  $\chi$  defined in (1.20) can be written as the limit of a function  $\chi(u)$  such that  $\chi = \lim_{u \rightarrow 1} \chi(u)$  with

$$\chi(u) := 2 - \frac{\log \{\mathbb{P}(F_X(X) < u, F_Y(Y) < u)\}}{\log(u)}.$$

The different properties for  $\chi$  and  $\chi(u)$  are the following:

- $0 \leq \chi \leq 1$ .
- $\chi = 0$  if, and only if,  $X$  and  $Y$  are asymptotically independent. This characterisation corresponds to the definition given by Sibuya (1960) for asymptotically independent variables.
- If  $\chi > 0$ ,  $X$  and  $Y$  are asymptotically dependent, and the value of  $\chi$  increases with the strength of asymptotic dependence.
- The sign of  $\chi(u)$  determines whether  $X$  and  $Y$  are positively or negatively associated at the quantile level  $u$ .

**Example 1.20.** Looking at the wave data used in Chapter 2, we can consider the strength of asymptotic dependence between the  $H_s$  at the two locations of interest. From Figure 1.13, it seems that  $\chi = \lim_{u \rightarrow 1} \chi(u) > 0$ , indicating that the two variables might be asymptotically dependent.

**Remark 1.21.** Looking at the right-hand side of Figure 1.13, as  $u$  gets closer to 1, the point-wise confidence interval increases, up to contain 0, due to the smaller number of points in the extreme. It is therefore not so evident to be in favour of asymptotic

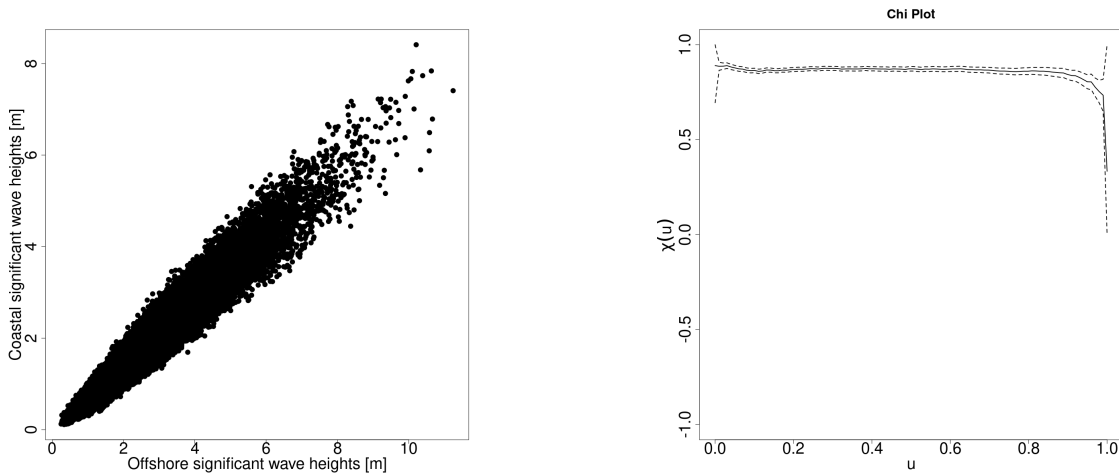


Figure 1.13: (left) Scatter plot of significant wave heights in two nearby locations off the coast of Brittany, more details are given in Chapter 2. (right) Plot of the corresponding estimated dependence measure  $\chi$  for increasing values of quantile levels  $u$  with 95% point-wise confidence intervals (estimation is performed using the R package `evd` (Stephenson, 2002)).

*dependence.*

*In this context, one could use other graphical tools to check for extremal dependence. For example, and as already mentioned, it is possible to standardised data in common margins in order to focus on the extremal dependence structure. This is shown in Figure 1.14, where we compare the asymptotic dependence behaviour of the  $H_s$  data and the simulated Gaussian vector of Figure 1.6. Unlike the Gaussian vector, for the  $H_s$  data some points are lying in the interior of the positive quadrant, supporting asymptotic dependence.*

*Note that more precise diagnostics can be achieved with the help of statistical tests. There is an extensive literature on this subject, a review can be found in Bacro and Toulemonde (2013). For example Draisma et al. (2004) proposed a test based on an estimator of the coefficient of tail dependence  $\eta$  (see Equation (1.21)). With an extension to the spatial context, Bacro et al. (2010) developed a test linked to the F-madogram, a quantity often considered in spatial extremes and defined by  $\frac{1}{2}\mathbb{E}|F_X(X) - F_Y(Y)|$  (see the references therein).*

If we assume that  $X$  and  $Y$  have common standard Fréchet margins,  $\chi$  can also be written as  $\chi = \lim_{x \rightarrow +\infty} \chi(x)$  where

$$\chi(x) := 2 + x \log \{ \mathbb{P}(X < x, Y < x) \}.$$

From this, one can easily see that for the bivariate extreme value distribution given by  $G(x, y) = \exp(-V(x, y))$ ,  $\chi(x) \equiv \chi = 2 - V(1, 1)$ .



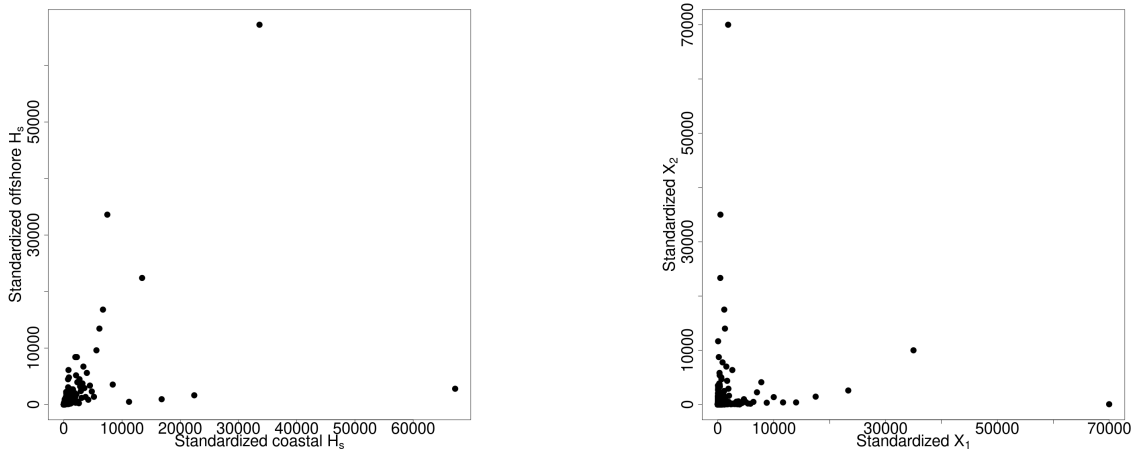


Figure 1.14: Asymptotic dependence comparison on Pareto scale between  $H_s$  data (left) and a Gaussian vector (right).

More generally, if  $(X, Y)$  is in the domain of attraction of a bivariate extreme value distribution  $G$  with exponent measure  $V$ , then  $\chi = 2 - V(1, 1)$ . In terms of the Pickands dependence function this can be written as  $\chi = 2(1 - A(1/2))$ . From this, we retrieve that asymptotic independence is equivalent to  $\chi = 0$ .

**Example 1.22.** Following Examples 1.15 and 1.19, the dependence measure for the logistic model is given by

$$\chi = 2 - 2^\alpha.$$

The role of  $\alpha$  is once again highlighted: when  $\alpha$  gets closer to 0,  $\chi$  tends to 1 (i.e. stronger asymptotic dependence) and when  $\alpha = 1$ ,  $\chi = 0$  (i.e. asymptotic independence).

Yet, in the case of asymptotic independence (i.e.  $\chi = 0$ ), no information is available on the strength of dependence at large but finite levels. In this case, an alternative measure of dependence, denoted  $\bar{\chi}$ , is often introduced (e.g. Coles et al., 1999) and provides information on the strength of dependence in case of asymptotic independence. This coefficient is defined through the following limit, provided that it exists,  $\bar{\chi} := \lim_{u \rightarrow 1} \bar{\chi}(u)$ , where

$$\bar{\chi}(u) := \frac{2 \log(1 - u)}{\log \mathbb{P}(F_X(X) > u, F_Y(Y) > u)} - 1, \quad u \in (0, 1).$$

Its properties are the following:

- $-1 \leq \bar{\chi} \leq 1$ .
- $\bar{\chi} = 1$  if, and only if,  $X$  and  $Y$  are asymptotically dependent, and in this case  $\chi$  informs us on the strength of asymptotic dependence between the variables.

- If  $\bar{\chi} < 1$ ,  $X$  and  $Y$  are asymptotically independent, and the value of  $\bar{\chi}$  provides a measure that increases with dependence strength.

All in all, the pair  $(\chi, \bar{\chi})$  provides complete information on the extreme dependence behaviour between two variables. This has been applied for example in [Towe et al. \(2013\)](#) to study the dependence structure between extreme significant wave height and extreme wind speed (such data will be considered in Chapter 2).

The disadvantage of  $\bar{\chi}$  is that, unlike  $\chi$ , it is not easily interpreted. A third dependence coefficient was introduced by [Ledford and Tawn \(1996\)](#), which is perhaps easier to interpret as it is directly related to the way the joint tail decreases. Assuming that  $X$  and  $Y$  have unit Fréchet margins, they suggested the following tail model

$$\mathbb{P}(X > x, Y > x) \sim \frac{l(x)}{x^{1/\eta}}, \quad \text{for large } x, \quad (1.21)$$

where  $\eta \in (0, 1]$  and  $l$  is a slowly varying function (see Definition 1.6). The coefficient  $\eta$  in (1.21) is called the *coefficient of tail dependence* and constitute another measure of extremal dependence.

Using (1.20), one can show that, provided that the limit exists,

$$\chi = \lim_{x \rightarrow \infty} x^{1-1/\eta} l(x). \quad (1.22)$$

From (1.22), if  $\eta = 1$  and  $\lim_{x \rightarrow \infty} l(x) = c \in (0, 1]$ ,  $X$  and  $Y$  are asymptotically dependent with strength of dependence given by  $\chi = c$ . On the other hand, if  $\eta < 1$  or if  $\eta = 1$  and  $\lim_{x \rightarrow \infty} l(x) = 0$ ,  $X$  and  $Y$  are asymptotically independent. Within the class of asymptotic independence, three types of dependence can be distinguished:

1. If  $1/2 < \eta \leq 1$ ,  $X$  and  $Y$  are positively associated: large observations for both  $X$  and  $Y$  occur more frequently than under exact independence.
2. If  $\eta = 1/2$ , extremes of  $X$  and  $Y$  are near independent. If additionally  $l(x) = 1$ , the variables are exactly independent.
3. If  $0 < \eta < 1/2$ ,  $X$  and  $Y$  are negatively associated: large observations for both  $X$  and  $Y$  occur less frequently than if  $X$  and  $Y$  were independent.

In [Heffernan \(2000\)](#) values of the coefficient of tail dependence  $\eta$  and slowly varying function  $l$  can be found for a wide range of bivariate distributions, including both asymptotic dependence and asymptotic independence. Finally,  $\bar{\chi}$  and  $\eta$  can be linked through  $\bar{\chi} = 2\eta - 1$ .

**To sum up**

The coefficient of extremal dependence  $\chi$  defined in (1.20) measures the strength of asymptotic dependence between the components of a random vector. Unlike the measures introduced in Section 1.2.1,  $\chi$  is a finite number, easier to handle.

If  $\mathbf{X} \in \mathcal{D}(G)$ ,  $\chi$  can be linked to the previous measures through

$$\chi = 2 - V(1, 1) = 2(1 - A(1/2)).$$

In case of asymptotic independence (i.e.  $\chi = 0$ ), other measures have been developed to quantify the dependence structure at high but finite levels. Namely the coefficient of extremal dependence  $\bar{\chi}$  and the coefficient of tail dependence  $\eta$ . The relation between the two is given by  $\bar{\chi} = 2\eta - 1$ .

**1.2.4 Limitations of bivariate EVT**

From Theorem 1.17, the bivariate threshold excess model presented in Section 1.2.2, and used in Chapter 2, requires that the underlying distribution  $F$  belongs to the domain of attraction of a bivariate EV distribution. For asymptotically dependent variables, such models are fully adequate. Therefore, looking at Figure 1.13, a bivariate threshold excess model seems well adapted to the wave data discussed in Chapter 2.

For some applications, asymptotically independent data can also be encountered (e.g. [Heffernan and Resnick, 2005](#); [Wadsworth and Tawn, 2012](#); [Towe et al., 2017](#), ...). For example, in Chapter 3, we will consider asymptotically independent measurements of river discharges from distant stations.

However, if asymptotic independence arises, the class of bivariate extreme value distributions is no longer appropriate. This can be illustrated in several ways. For example, assume that we are in the framework of Section 1.2.1, i.e. we have a set of bivariate vectors  $\{(X_i, Y_i)\}_{1 \leq i \leq n}$ , independent and identically distributed, with unit Fréchet margins and common distribution  $F$ . Moreover, assume that  $F \in \mathcal{D}(G)$ , i.e.

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{M}_n/\mathbf{n} \leq (x, y)] = G(x, y).$$

Then, using that  $\log x \sim x - 1$  for  $x \rightarrow 1$ ,

$$\log \mathbb{P}[\mathbf{M}_n/\mathbf{n} \leq (x, y)] = n \log F(nx, ny) \sim n(F(nx, ny) - 1).$$

Therefore, the above convergence can then be rewritten as

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X_1/n > x \text{ or } Y_1/n > y) = -\log G(x, y).$$

This leads to

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X_1/n > x \text{ and } Y_1/n > y) = \log G(x, y) - \log G_1(x) - \log G_2(y),$$

where  $G_1$  and  $G_2$  are the marginals of  $G$  (see the remark following Definition 1.11). Now going back to our question, assume that we are in the domain of attraction of the asymptotic independence. Then in this case

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X_1 > nx \text{ and } Y_1 > ny) = 0.$$

This shows that in the case of asymptotic independence, we can not extrapolate joint probabilities at high but finite levels. Thus, if we only rely on bivariate EVT, asymptotic independence is similar to perfect independence and this is certainly an excessive assumption.

To fill this gap, models adapted to the case of asymptotic independence have been developed. We present two of them in the following section.

### 1.2.5 Models coping with asymptotic independence

To overcome the above mentioned issue, models that accommodate both asymptotic dependence and asymptotic independence have been introduced. In the following,  $X$  and  $Y$  are assumed to have unit Fréchet margins.

Note that compared to the bivariate GP model introduced in Section 1.2.2, the following models focus on regions where the two components are simultaneously large. This corresponds for example to the situation in the right-hand side of Figure 1.5.

Building on their first study (Ledford and Tawn, 1996), Ledford and Tawn (1997) introduced an extension of model (1.21) with a more flexible model given by

$$\mathbb{P}(X > x, Y > y) \sim \frac{\mathcal{L}(x, y)}{x^{c_1} y^{c_2}}, \text{ for large } x \text{ and } y, \quad (1.23)$$

where  $c_1, c_2 > 0$  are such that  $c_1 + c_2 \geq 1$  and  $\mathcal{L}$  is a bivariate slowly varying function, that is, there exists a function  $g$  such that, for all  $x, y > 0$  and  $c > 0$ ,

$$g(x, y) = \lim_{r \rightarrow \infty} \frac{\mathcal{L}(rx, ry)}{\mathcal{L}(r, r)} \text{ and } g(cx, cy) = g(x, y). \quad (1.24)$$

The homogeneity property in (1.24) implies that  $g$  is constant along any ray  $y = ax$  for  $a > 0$  and one can introduce a ray dependence function  $g_*$  defined as  $g(x, y) = g_*(x/(x+y)) = g_*(w)$  for  $w = x/(x+y) \in (0, 1)$ . Setting  $\eta = 1/(c_1 + c_2)$  shows that this latter model (1.23) contains the previous model (1.21).

**Example 1.23.** Assume that  $X$  and  $Y$  are two independent Fréchet-distributed variables. Then

$$\mathbb{P}(X > x, Y > y) = [1 - \exp(-1/x)][1 - \exp(-1/y)] \sim \mathcal{L}(x, y)(xy)^{-1},$$

for large  $x$  and  $y$ , where  $\mathcal{L}(x, y) \equiv 1$  is a bivariate slowly varying function. In particular  $c_1 = c_2 = 1$  and  $\eta = 1/2$  (exact independence).

Finally, building on a simplified version of model (1.23) by considering  $c_1 = c_2$ , Ramos and Ledford (2009) proposed the following joint tail model

$$\mathbb{P}(X > x, Y > y) = \frac{\mathcal{L}(x, y)}{(xy)^{1/(2\eta)}}, \quad (1.25)$$

with the same assumptions on  $\mathcal{L}$  and  $\eta$  as before, i.e.  $\eta \in (0, 1]$  and  $\mathcal{L}$  is a bivariate slowly varying function as in (1.24). This new model aimed to overcome some issues of model 1.23 for which, in some special cases, the joint density obtained was valid only when  $x/(x + y) \in (0, 1)$ .

The main result of Ramos and Ledford (2009) is an alternative spectral representation as in Theorem 1.13 which provides a unified framework covering both asymptotic dependence and asymptotic independence, unlike the classical multivariate extreme value case.

**Theorem 1.24** (Ramos and Ledford (2009)). *If (1.25) holds and if*

$$\lim_{u \rightarrow \infty} \frac{\mathbb{P}(X > us, Y > ut)}{\mathbb{P}(X > u, Y > u)} = \frac{g_*(s/(s + t))}{(st)^{1/(2\eta)}},$$

for all  $s, t \geq 1$  where  $g_*$  is the limit function that is defined following equation (1.24) and  $\eta$  is the coefficient of tail dependence. Then  $g_*(w)$  satisfies

$$\eta^{-1} g_*(w) = \left( \frac{1-w}{w} \right)^{1/(2\eta)} \int_0^w z^{1/\eta} dH_\eta(z) + \left( \frac{w}{1-w} \right)^{1/(2\eta)} \int_w^1 (1-z)^{1/\eta} dH_\eta(z),$$

where  $H_\eta$  is a (non-negative) measure on  $[0, 1]$  satisfying the normalisation condition

$$\eta^{-1} = \int_0^{1/2} w^{1/\eta} dH_\eta(w) + \int_{1/2}^1 (1-w)^{1/\eta} dH_\eta(w).$$

The measure  $H_\eta$  introduced in Theorem 1.24 is the analogue of the spectral measure  $H$  defined in Theorem 1.13 and a relationship between the two measures can be derived (Ramos, 2003; Ramos and Ledford, 2009). For example, if  $(X, Y)$  follows a bivariate extreme value distribution with standard Fréchet margins and spectral measure  $H$ , assuming that  $X$  and  $Y$  are not exactly independent, so that  $\eta = 1$  (or equivalently  $\chi > 0$ ), then one can show that  $H = \{2 - V(1, 1)\} H_1$  (see Ramos and Ledford, 2009).

In Chapter 3, model (1.25) of Ramos and Ledford (2009) is used to develop a risk measure adapted to a wide range of dependence models, including both

asymptotic dependence and asymptotic independence. In particular, links between model (1.25) and *hidden regular variation* are discussed in Section 3.1.2.

In this section, only joint tail models based on multivariate regular variation have been introduced. But many other models for asymptotically independent data have been developed in the literature such as the conditional tail model of [Heffernan and Tawn \(2004\)](#), which avoids the restriction to regions where both components are simultaneously large. [de Carvalho and Ramos \(2012\)](#) provide a review on the existing methods for the statistical modelling of asymptotically independent data.

We give here some possible limitations of the aforementioned appealing models that deal with both asymptotic dependence and asymptotic independence. Firstly, and as already mentioned, [Ramos and Ledford \(2009\)](#) model focuses on regions where both variables are large, this can be a significant limitation for some applications, such as the application considered in Chapter 2 with the wave data. On the other hand, for the flexible model of [Heffernan and Tawn \(2004\)](#), which removes this constraint, some authors have raised some theoretical issues (e.g. [Drees and Janßen, 2017](#)).

Finding models that deal with both asymptotic dependence and asymptotic independence is a current intensive research area, and recent advances in this direction have been made (e.g [Wadsworth et al., 2017](#)).

A final remark is that, in our case, the data at hand used in Chapter 2, are asymptotically dependent. This is one of the reasons why we choose to focus and develop bivariate threshold excess models, which are theoretically valid models, relying on EVT.

### 1.3 Tools considered in the following

#### To sum up

Among the different tools presented in this chapter, the ones that we will use in the continuation of this manuscript are the following.

- In Chapter 2, a non-parametric simulator of bivariate GP vectors (see [Rootzén and Tajvidi \(2006\)](#) and Section 1.2.2) will be developed to simulate extreme wave events.
- In Chapter 3, a risk measure adapted to binary classification of extreme events in case of both asymptotic dependence and independence will be considered, relying on [Ramos and Ledford \(2009\)](#) model (see also Section 1.2.5).



## Chapter 2

# Joint stochastic simulation of extreme coastal and offshore significant wave heights

### Overview of Chapter 2

The main goal of the following chapter is to propose and study a stochastic simulator that, given offshore conditions, produces jointly offshore and coastal extreme significant wave heights. For that, we rely on bivariate Peaks over Threshold (see Section 1.2.2 and [Rootzén and Tajvidi \(2006\)](#)) and develop a non-parametric simulation scheme of bivariate GPD. To take into account non-stationarities, we also adapt the extended generalised Pareto model ([Naveau et al., 2016](#)), letting the parameters vary with specific sea state parameters.

Section 2.2 gives some background motivations for the study of extreme coastal wave heights. More specifically, precise definitions of the parameters used to describe a sea state are detailed in Section 2.1.1, along with the databases used in Paper I ([Legrand et al., 2022](#)). Section 2.1.2 briefly describe the evolution of the conducted study, motivating the modelling developed in Paper I. This preliminaries are ended in Section 2.1.3, with a brief introduction on the extended generalised Pareto distribution (EGPD), which is adapted in Paper I for a non-stationary modelling.

From Section 2.2 to Appendix 2.B, Paper I is reproduced as it is. In Section 2.3, the sea state data considered in this study are presented. We then develop in Section 2.4 the marginal regression models, incorporating the effects of the peak period and the peak direction on the EGPD parameters. In Section 2.5, our non-parametric method to simulate MGP



vectors is presented, along with some numerical experiments. Two algorithms are outlined, one for bivariate simulations and a second one for conditional simulations. Finally, both algorithms are applied in Section 2.6 to the sea state data. The results show that these algorithms successfully simulate new realistic extreme  $H_s$  events.

## 2.1 Preamble to Paper I

The characterisation of future extreme wave events is crucial because of their multiple impacts, covering a broad range of topics such as coastal flood hazard, coastal erosion, reliability of offshore and coastal structures . . .

For instance, for coastal flood risk assessment, [Idier et al. \(2020\)](#) combined historical data, statistical and physical models to reconstruct past coastal flood events in a specific site on the French coastline. They showed that the return period associated to historical flood events tends to decrease due to sea-level rise. On a global scale (i.e. along the world coastlines), [Marcos et al. \(2019\)](#) considered the compound effects of extreme wave heights and storm surges. As mentioned in Section 1.2.3, [Marcos et al. \(2019\)](#) found that return periods of coastal extreme water levels were significantly overestimated if the dependency structure between wave heights and storm surges was not considered.

Characterisation of wave climatology is also of great concern regarding the implementation of Marine Renewable Energy structures (often shortened to MRE). For example, [Stopa et al. \(2013\)](#) used a numerical wave model (see Section 2.1.1) to characterise the wave climate in Hawaii for potential implementation sites of wave energy converters.

Concerning the variability of the intensity of the extremes, authors have shown an increase in extreme significant wave heights in the last decades over the North Atlantic Ocean ([Bertin et al., 2013](#); [Young and Ribal, 2019](#)), and which could be attributed to climate change ([Rohmer et al., 2020](#)). Regarding future wave climate projections, under different emission scenarios<sup>1</sup> (RCP4.5 and RCP8.5), there is a consensus on a decrease of the mean significant wave heights in the North Atlantic Ocean ([Aarnes et al., 2017](#); [Bricheno and Wolf, 2018](#); [Lemos et al., 2021](#)), but it has also been shown a slight increase in the most extreme wave events ([Bricheno and Wolf, 2018](#)). Caution should be exercised here as there is no general consensus on this subject (e.g. [Lobeto et al., 2021](#)). Note that in order to address such issues, initiatives such as COWCLIP (Coordinated Ocean Wave Climate Project) have been carried out with the aim, among others, of producing a unified database of global wave climate projections ([Morim et al., 2020](#)).

---

<sup>1</sup>IPCC (2014) defined different scenarios given the evolution of greenhouse gas concentrations in the atmosphere, named RCP, according to the climate policies adopted.

All these studies show the importance of predicting extreme wave heights accurately, and more particularly in coastal regions. In Paper I ([Legrand et al., 2022](#)), we propose a first attempt in this direction by looking at how moderately high offshore wave heights can produce large coastal significant wave heights. For this, we will propose a joint stochastic simulator which, given offshore sea states conditions, produces extreme offshore and coastal significant wave heights.

### 2.1.1 Wave data

#### Spectral definition of sea state parameters

In this section, we provide a precise definition of significant wave height, denoted  $H_s$ . More details on ocean waves can be found, for example, in [Holthuijsen \(2007\)](#). Recall that in Chapter 1, we said that  $H_s$  gave a measure of the ocean surface roughness. More precisely, this quantity can be defined from the energy of a sea state.

A sea state is the characterisation, in a given area and for a limited period, of the sea surface elevation, which is assumed to be a stationary process. Because waves are generated from various weather systems, mainly wind sea (waves generated locally from the wind) and swell (that have travelled from a remote generating area), the description of this sea surface can be complex.

To overcome this issue, a sea state can be decomposed into several simple wave trains (or harmonic waves), each having a particular direction of propagation and wave period, the so called *spectral decomposition* (see Figure 2.1).

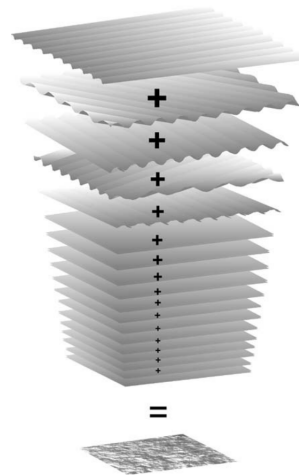


Figure 2.1: Spectral decomposition of a sea state as a sum of harmonic waves (taken from [Holthuijsen \(2007\)](#)).

From this spectral decomposition, one can consider the spectral energy density function of each simple wave train involved, denoted  $E(f, \theta)$ , where  $\theta$  is the

direction of propagation and  $f$  is the wave frequency (i.e. the inverse of the wave period). Note that it is a convention to consider the frequency rather than the period. A representation of the total spectral energy of a sea state is then given by the so-called *wave spectrum*, which gives the spectral energy of each wave train given its direction  $\theta$  and frequency  $f$ . An example is given in Figure 2.2. From this representation, one can identify the different wave regimes:

- swells that correspond to a relatively narrow directional dispersion and to low frequencies: they are visually rather regular and long-crested,
- wind sea corresponding to a much broader directional dispersion and to higher frequencies: corresponding to irregular and short-crested waves.

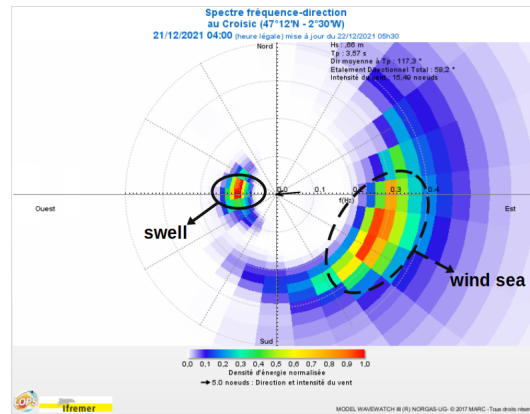


Figure 2.2: Example of wave spectrum at the coastal location considered in Paper I (see Figure 2.7). The corresponding date is 21/12/2021 04:00. The wave spectrum is provided by <https://marc.ifremer.fr/>.

To synthesise the information contained in this 2-dimensional spectrum, several statistics are defined. These parameters are then used to describe a sea state. We give here the formal definitions of the spectral parameters used in Paper I. More details are given, for example, in Chapter 4 of [Holthuijsen \(2007\)](#).

- The significant wave height  $H_s$  (in meters) is estimated from the total variance of the moving sea surface elevation, denoted  $\langle \eta^2 \rangle$ , which is shown to be the integrated spectral density in both frequency and direction:

$$H_s = 4\sqrt{\langle \eta^2 \rangle} = 4\sqrt{\int_0^\infty \int_0^{2\pi} E(f, \theta) df d\theta}.$$

The significant wave height is usually interpreted as a measure of the wave energy of the sea state. Historically, it is also defined as the average of the highest one-third of wave heights.

- The peak period  $T_p$  (in seconds) is the period associated to the maximum of the omnidirectional spectrum, it corresponds to the period of the dominant waves in the sea state and is defined as

$$T_p = \left[ \arg \max_f E(f) \right]^{-1} = \left[ \arg \max_f \int_0^{2\pi} E(f, \theta) d\theta \right]^{-1}.$$

- The peak direction  $D_p$  (in degrees) is the direction associated to the maximum of energy of the wave spectrum, defined as

$$D_p = \arg \max_{\theta} \int_0^{\infty} E(f, \theta) df.$$

As an example, for the wave spectrum in Figure 2.2,  $H_s = 0.66m$ ,  $T_p = 3.57s$  and  $D_p = 117^\circ$ , indicating that the corresponding sea state is rather dominated by a wind sea regime and with dominant waves coming from the North-West.

These statistics are limited as they give a limited picture of a sea state, as explained by [Holthuijsen \(2007\)](#) (p.25):

*For instance, wave conditions may well be similar in the sense that the significant wave height and period are equal, but they may still be very different in detail: a mixed sea state of wind sea (short, irregular, locally generated waves) and swell (long, smooth waves, generated in a distant storm) may have the same significant wave height and period as a slightly higher wind sea without swell.*

However, for a long-term description of the wave climate, these statistics are sufficient and are conventionally used, especially for coastal studies.

### Wave data sources

Significant wave height is obtained either from direct measurements of the sea surface elevation, using *in situ* measurements (such as buoys) or remote-sensing techniques (e.g. satellites), or from numerical wave models. Numerical wave models produce simulated wave spectrum based on the physical equation of wave energy balance ([Filipot and Ardhuin, 2012](#)), taking into account different forcing (such as winds, currents, sea ice, bathymetry, ...).

Note that *in situ* measurements generally provide estimates of the omnidirectional spectrum  $E(f)$ , whereas numerical models can estimate the full wave spectrum  $E(f, \theta)$ , but they are considered to be closer to the truth than satellite or numerical data and therefore used often as reference (e.g. [Ailliot et al., 2011](#)).

Historical outputs from numerical wave models are often called *hindcast* data (in contrast to forecast data), since it consist of numerical predictions in the past. In this thesis, two hindcast databases based on the numerical wave model

WAVEWATCH-III (Tolman et al., 2014) are considered (more details are given later, Section 2.3). Both have been developed by IFREMER (French Research Institute for Exploitation of the Sea) and provide sea state statistics (as  $H_s$ ,  $T_p$  and  $D_p$ ). For the two databases, the numerical wave model is forced by the CFSR wind reanalysis<sup>2</sup> (Saha et al., 2010). Their main characteristics are the following:

- IOWAGA (Ardhuin and Accensi, 2014): covers the whole globe with a regular  $0.5^\circ$  resolution grid in latitude and longitude (see Figure 2.3) and 3-hourly temporal resolution from 01/01/1990 to 31/03/2020.
- HOMERE (Accensi and Maisondieu, 2015): covers the Channel and the Bay of Biscay on an unstructured grid, refined close to the coast (see Figure 2.3). It takes into account a high resolution bathymetry and is forced by IOWAGA on the wet boundaries. The temporal resolution is one hour time step, data are available from 01/01/1994 to 30/04/2021.

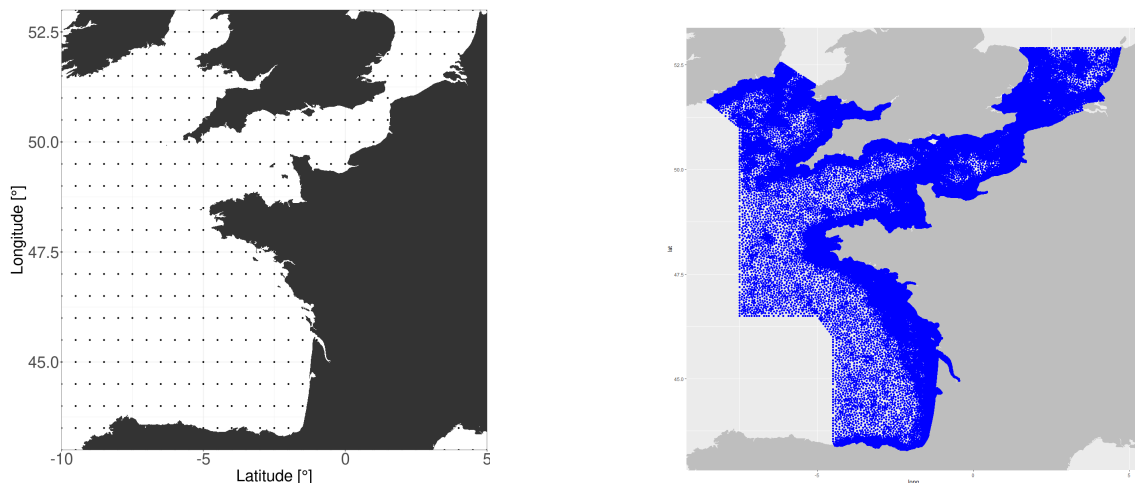


Figure 2.3: Grids of the hindcast databases considered. (left) Portion of the IOWAGA grid. (right) Total HOMERE grid.

## 2.1.2 On the choice of appropriate covariates for modelling extreme coastal sea states

The starting point of this study was to characterise the conditional distribution of extreme significant wave height, considered as exceedances over a high threshold, close to the French coast.

<sup>2</sup>Reanalysis data are obtained by combining historical observations and numerical model outputs through data assimilation (e.g. Evensen, 2009), providing more consistent databases.

We chose to focus on one specific point from the HOMERE database, represented in yellow in Figure 2.4, which corresponds to the SEM-REV sea test site (Mousslim et al., 2009). The main advantage of this specific location is that it is well-documented and is used for the development and optimisation of renewable energy structures, it is therefore of great interest to describe extreme sea states at this specific site.

Since waves are generated by winds, a first part of the study was devoted to find a conditional model for extreme  $H_s$  given wind fields. Indeed, waves observed at the point of interest depends on wind conditions over the North Atlantic Ocean. The large spatial extent of wind fields implied the need to identify the characteristics of the spatio-temporal wind processes associated with extreme  $H_s$  and to derive a representation in a finite dimensional space, in order to facilitate the calculation of the above mentioned model.

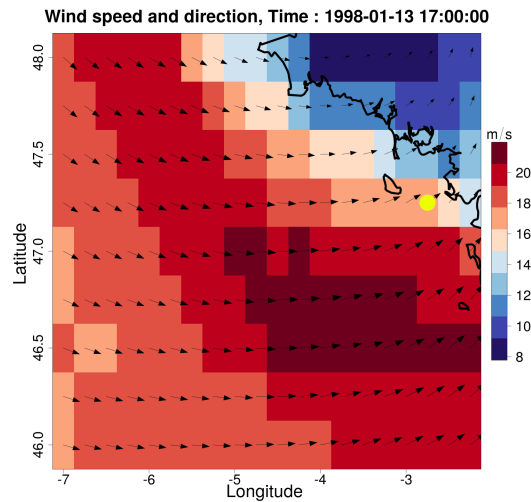


Figure 2.4: Wind field corresponding to the highest  $H_s$  at the location of interest, depicted by the yellow dot. Arrows represent the wind direction and colours its intensity. Wind data are extracted from the ERA5 database.

For that, we considered the 10m wind components  $U_{10}$  and  $V_{10}$  from the ERA5 global reanalysis database (Hersbach et al., 2020). This dataset was used in a preliminary approach, it was chosen mainly for accessibility reasons, although in a second approach it would have been more relevant to consider the CFRS reanalysis dataset since it corresponds to the wind forcing of the wave data.

Figure 2.4 illustrates the spatial footprint of the wind field corresponding to the maximum of  $H_s$  in the dataset. This preliminary work highlighted, for extreme  $H_s$ , dominant sectors of wind direction. But it also raised some limitations due to the large spatial extent of wind fields and the complex processes involved in the generation of extreme sea states, which imply the superposition of very different wave regimes (swells and wind sea). Obakrim et al. (2022) addressed this issue

using a downscaling approach based on weather-types. This approach was also investigated by Michel et al. (2022), relying on deep learning methods. One drawback of these studies is that it requires to take into account the whole wind field over the North Atlantic Ocean.

To avoid this, we decided to exploit the somewhat space-time Markovian behaviour of extreme  $H_s$ . Indeed, there is a strong spatial and temporal dependency, notably in the extremes, among the nearest data points due to the wave propagation (see Figure 2.5). Furthermore, the benefits of considering a wave field instead of a wind field is that the former contains most of the information resulting from the complex wave generation processes.

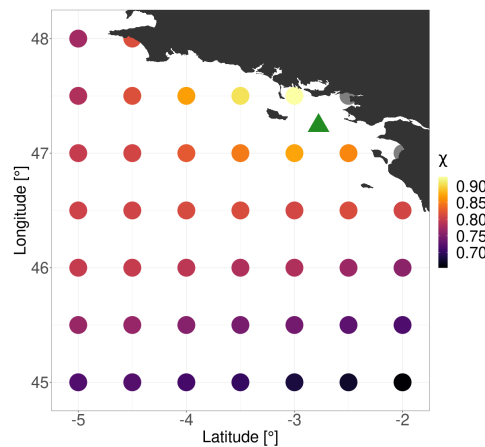


Figure 2.5: Measure of asymptotic dependence  $\chi$  between wave data points from IOWAGA and the location of interest extracted from HOMERE (green triangle point).

The advantage of the global IOWAGA database is that it can be run rapidly. Whereas HOMERE is slower to run, due to its strong refinement. Thus, finding a relationship between the large-scale model (represented by IOWAGA) and the small-scale model (represented by HOMERE) would allow rapid calculations of the probabilities of extreme  $H_s$  at the location of interest.

In Paper I (Legrand et al., 2022), we will built on the relation between a large-scale grid point and a small-scale grid point of interest, near the coast (see Figure 2.7). The proposed model will allow for a statistical downscaling of extreme  $H_s$  from the global wave model to the coastal model.

### 2.1.3 Extended generalised Pareto distribution

Among extreme value methodologies, GP models (see Section 1.1.2) are powerful tools for modelling extreme climate observations. However a major drawback of such methods is the threshold choice. To answer this problem, threshold se-

lection methods have been proposed (see, e.g. [Raymond-Belzile \(2019\)](#), Chapter 1).

Another drawback is that by definition of GP models, only values that are above the threshold are considered. But for some applications, one needs to model the entire range of the data. For instance, this is the case in Paper I, where we want to model  $H_s$  exceedances given moderately high values of  $H_s$ , i.e. we look at quantities of the following type

$$H_s - u \mid H_s > u,$$

but for moderate threshold  $u$ .

In this context, different models have been proposed to describe the range of observations. For example, studies have considered piece-wise models, for which one consider a GP model above the threshold and a parametric model for the bulk below the threshold, for wave applications people considered for instance a truncated Weibull ([Randell et al., 2016](#)) or a truncated Gamma distribution ([Ross et al., 2017](#)).

However such models are discontinuous at the threshold. To alleviate this, [Carreau and Bengio \(2009\)](#) developed a mixture GP model, with a smooth transition at the threshold, from an hybrid GP distribution that combines a Gaussian and a GP tail. Yet this model is rather complicated to use in practice due to the strong continuity constraints.

Recently, [Naveau et al. \(2016\)](#) (see also the references therein) proposed a model that bypasses the threshold selection and models the entire range of data, the so called extended generalised Pareto distribution (EGPD).

The EGPD is a distribution defined on  $(0, \infty)$  whose distribution function is given by

$$F(x) = G[H_{\xi, \sigma}(x)],$$

where  $G$  is a continuous distribution function on  $[0, 1]$  and  $H_{\xi, \sigma}$  is a GP distribution function as defined in Section 1.1.2 (Chapter 1). In [Naveau et al. \(2016\)](#), several constraints are required for  $G$  in order to ensure the lower and upper tails behaviours. Four different parametric models for  $G$  are derived.

In this study we will only use the first one, which in practice appears to be flexible enough ([Tencaliec et al., 2019](#); [de Carvalho et al., 2021](#); [Rivoire et al., 2021](#); [Le Gall et al., 2022](#)), and given by  $G(v) = v^\kappa$ . This model has therefore three parameters: a scale  $\sigma > 0$ , a shape  $\xi \in \mathbb{R}$  (controlling the upper-tail behaviour) and a second shape parameter  $\kappa > 0$  (which controls the shape of the lower tail). Figure 2.6 depicts the density of the EGPD for different values of  $\kappa$ .

In Paper I bellow, the EGPD will be used for marginal modelling, as described in Equation (2.2). The review of [Jonathan and Ewans \(2013\)](#) on the statistical modelling of extreme significant wave heights highlights the necessity of taking into account covariates effects. For this, Generalised Additive Models (GAMs) provide a nice flexible framework ([Hastie and Tibshirani, 1986](#)).



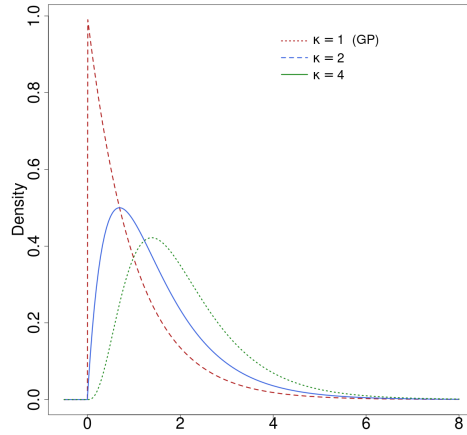


Figure 2.6: Densities of the extended GP distribution for different values of the shape parameter  $\kappa$ , with  $\xi = 0$  and  $\sigma = 1$ .

Recall that the simplest regression model, the linear model, can be written in compact form as

$$\begin{cases} Y \sim \mathcal{N}(\mu, \sigma^2), \\ \mu = X\beta, \end{cases}$$

where  $X$  is the design matrix and  $\beta$  the vector of parameters. A GAM is a generalisation of the linear model, that allows for response distributions other than Gaussian, for a certain degree of non-linearity in the model and with linear predictors specified as smooth function of the covariates. It can be written as

$$\begin{cases} Y \sim \text{some exponential family distribution}, \\ g(\mathbb{E}[Y]) = X\beta + \sum_{j=1}^J s_j(x_j), \end{cases}$$

where  $g$  is a smooth monotonic link function and  $s_j$  are smooth functions of some covariates  $(x_1, \dots, x_J)$ .

The smooth functions  $s_j$  are typically represented as linear combinations of basis functions  $b_{j,k}$

$$s_j(x) = \sum_{k=1}^K \alpha_{j,k} b_{j,k}(x), \quad 1 \leq j \leq J,$$

where  $\alpha_{j,k}$  are some coefficients to be estimated. The basis functions  $b_{j,k}$  are completely known, and different basis (also called *splines*) can be considered for a same model, such as natural cubic splines, thin plate splines, tensor product splines (see [Perperoglou et al. \(2019\)](#) for a comprehensive review, with applications in R). A complete introduction on GAMs can be found in [Wood \(2006\)](#).

The above GAM considers only single-parameter distributions with mean parameter that vary as a smooth function of some covariates. Such models have then been extended to distribution families with more than one parameter. For example, regarding extreme value models, [Chavez-Demoulin and Davison \(2005\)](#) proposed to model the behaviour of extremes with GAMs by allowing the GPD parameters to be represented as smooth functions of covariates. [Yee and Stephenson \(2007\)](#) linked an extended version of GAMs (named Vector Generalised Additive Models, VGAM<sup>3</sup>) with EVT. All these models are built for extreme value models. However, as discussed previously, we wish to model values that are not necessarily extreme, relying on the EGPD.

To model non-stationarity with EGPD models, [Le Carrer \(2022\)](#) developed an add-on to the `gamlss` package ([Stasinopoulos et al., 2008](#)). GAMLSS (Generalised Additive Models for Location, Scale and Shape) allows for GAM forms of (up to) four-parameter distributions. In the case of the EGPD, it can be written as

$$\begin{cases} Y \sim EGPD(\xi, \kappa, \sigma) \\ g_1(\xi) = X_1\beta_1 + s_{1,1}(x_{1,1}) + \cdots + s_{1,J_1}(x_{1,J_1}), \\ g_2(\kappa) = X_2\beta_2 + s_{2,1}(x_{2,1}) + \cdots + s_{2,J_2}(x_{2,J_2}), \\ g_3(\sigma) = X_3\beta_3 + s_{3,1}(x_{3,1}) + \cdots + s_{3,J_3}(x_{3,J_3}), \end{cases}$$

where, for  $i \in \{1, 2, 3\}$ ,  $g_i$  is a link function,  $X_i$  is the design matrix containing the linear additive terms,  $\beta_i$  is the vector of linear parameters and  $(s_{i,j})_{1 \leq j \leq J_i}$  are the smoothing functions of the explanatory variables  $(x_{i,1}, \dots, x_{i,J_i})$ . The implementation of [Le Carrer \(2022\)](#) allows to consider negative shape parameters  $\xi$  by setting  $g_1$  as an identity link, allowing an unconstrained search domain for the inference of  $\xi$ . Note that the model fitting is performed using maximum penalised likelihood estimation.

This is used in Paper I to let the scale parameter  $\sigma$  of the EGPD vary as a smooth function of the peak period and peak direction (see Equation (2.2)).

A final remark is that besides GAM related models, other recent approaches have been considered to let the parameters of the EGPD vary given some covariates. For example, [de Carvalho et al. \(2021\)](#) developed a Bayesian approach to learn the effect of covariates on an EGPD-based model.

\*\*\*

Below, Paper I ([Legrand et al., 2022](#)) is reproduced as it will be submitted. It was written and developed by Juliette L. under the supervision of Pierre Ailliot, Philippe Naveau and Nicolas Raillard.

---

<sup>3</sup>VGAMs are more general models than GAMs, that allow for multiple linear predictors and encompass models outside the exponential family distribution ([Yee, 2015](#)). These models encompass the cases where  $Y$  is a vector.

## 2.2 Introduction

French coastlines have been particularly affected by extreme maritime events in the past (Nicolae Lerma et al., 2015). Co-occurrence of high tidal coefficients, atmospheric surge conditions and specific sea states can lead to extreme maritime events. These events are particularly crucial for assessing flooding risks and their consequences (Genovese and Przulski, 2013; Bertin et al., 2012). According to the special IPCC report (see Collins et al. (2019)), extreme wave heights, which contribute to these extreme maritime events, have increased over the past few years. The recent IPCC report (Seneviratne et al., 2021) indicates, with high confidence, an increase in the occurrence and magnitude of such coastal events in the future. More specifically, Caires and Sterl (2005) showed that the most extreme wave conditions were expected to occur in the North Atlantic, which include the Bay of Biscay, our study area.

Sea surface elevation over a geographical area results from the superposition of waves generated by local winds and by remote swell (generated in distant regions). The characterisation of this complex surface is called a sea state and to describe it, various parameters are available. In this work, we will focus on three variables: the significant wave height denoted  $H_s$  [m], the peak period,  $T_p$  [s], and the peak direction,  $D_p$  [°] (see e.g. Holthuijsen (2007) for more details).

From a coastal risk point of view, a fundamental question is to determine how moderately high offshore significant wave heights can produce large coastal  $H_s$ . Peak direction and peak period influence the relationship between coastal and offshore  $H_s$ . In this context, our main goal is to propose and study a stochastic simulator that, given offshore conditions ( $T_p$ ,  $D_p$ ,  $H_o$  moderately high), produces jointly offshore and coastal extreme significant wave heights. The left-hand side map of Figure 2.7 shows the two locations of interest.

From such a stochastic generator (1st goal), many products can be derived. In particular our second objective is to propose a conditional simulation model (2nd goal). The framework for each step of this study is summarised in Table 2.1. In order to make the two simulation models as flexible as possible, non-parametric algorithms are derived using resampling techniques (or non-parametric bootstrap (Efron, 1979)). While this study focuses on simulation of extreme  $H_s$ , and as illustrated by the numerical simulations in Section 2.5, the two non-parametric algorithms developed could be applied to a broad range of data.

In multivariate extreme value analysis, one is often interested in the joint behaviour of the variables as they become large. As illustrated by the right-hand side of Figure 2.7, which depicts a scatter plot between  $H_o$  and  $H_c$ , large values tend to occur simultaneously. For this specific type of dependence, called asymptotic dependence (Coles, 2001), models from the class of multivariate Extreme Value Theory (EVT) can be used. To achieve the two objectives (1st Goal and 2nd Goal), we will therefore propose two simulation algorithms based on multivariate Peaks

		$H_c$	$H_o$	$D_p$	$T_p$
Inference		v	v	v	v
1st Goal:	Joint simulation	x	x	v	v
2nd Goal:	Conditional simulation	x	v	v	v

Table 2.1: Summary of available data for each step of this study. A tick v (resp. a cross x) indicates the availability (resp. non-availability) of the data.

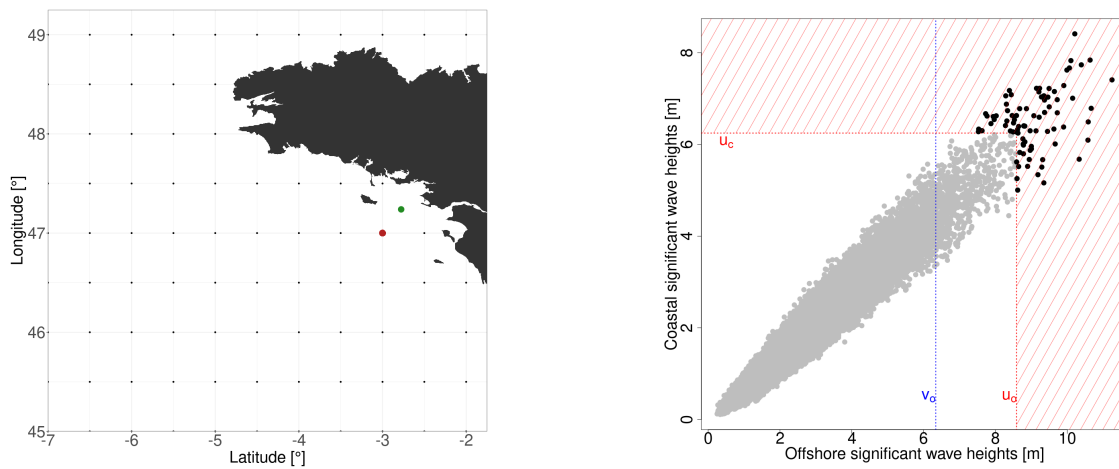


Figure 2.7: (left) Portion of IOWAGA hindcast database grid, the red dot corresponds to the "offshore" point (data extracted from the IOWAGA database) and the green dot corresponds to the "coastal" point (data extracted from the HOMERE database). (right) Scatter plot of the coastal significant wave heights versus the offshore significant wave heights with the different thresholds considered. The black dots belong to the region where the multivariate model is fitted.

over Thresholds (Sec. 8.3.1 [Beirlant et al., 2004](#)). Note that to assess whether data fall within the class of asymptotic dependence or not, summary statistics have been developed such as the dependence measures  $\chi$  and  $\bar{\chi}$  ([Coles et al., 1999](#)) and plots of these measures (not shown here) confirm strong dependence between large values of  $H_o$  and  $H_c$ .

For weakly dependent extremes, conditional models based on [Heffernan and Tawn \(2004\)](#) should be favoured to deal with our 2nd goal, see for example [Towe et al. \(2017\)](#); [Shooter et al. \(2019\)](#); [Tendijck et al. \(2021\)](#).

Before modelling the joint behaviour of large values, it is necessary to model margins ([Beirlant et al., 2004](#)). To visualise this task with respect to our data, the empirical histograms displayed in Figure 2.8 indicate that, given  $\{H_o > v_o\}$  (moderately high offshore significant wave heights), a traditional univariate extreme

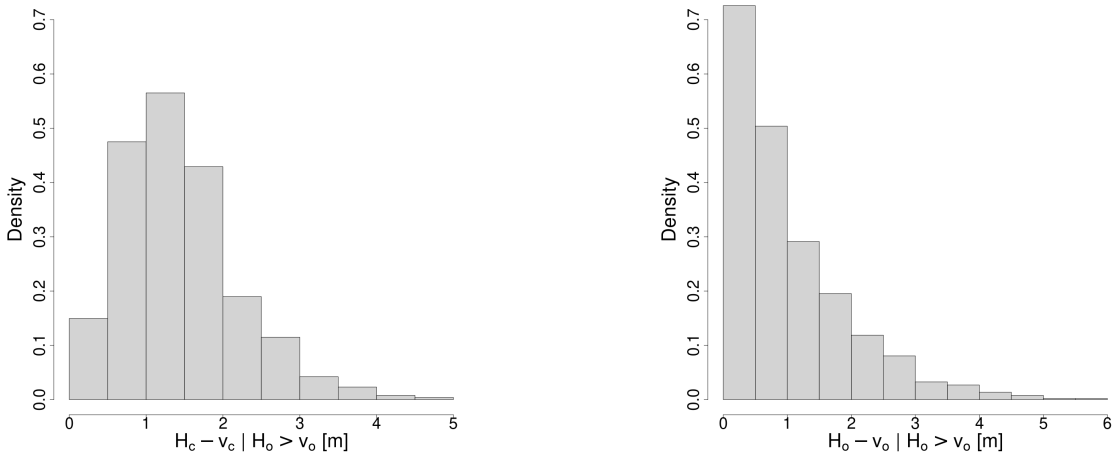


Figure 2.8: (left) Empirical histogram for the coastal significant wave height threshold exceedances and (right) similarly for the offshore significant wave heights threshold exceedances, illustrating that fitting a generalised Pareto distribution (Coles, 2001) is not suitable for the coastal data. The coastal marginal threshold is defined by  $v_c := \min(H_c; H_o > v_o)$ .

value approach based on fitting a generalised Pareto distribution (GPD) to the exceedances (Coles, 2001) is not appropriate. To tackle this issue, we use the extended generalised Pareto distribution (EGPD) introduced by Naveau et al. (2016) which handles this type of setting (see also Papastathopoulos and Tawn (2013)), more details are given in Section 2.4.

Furthermore, like many other environmental data, extreme  $H_s$  are non-stationary with respect to covariates (Jonathan and Ewans, 2013) and marginal models need to take into account this non-stationarity (e.g. Ewans and Jonathan, 2008; Méndez et al., 2008; Casas-Prat et al., 2014). To incorporate non-stationarities, Chavez-Demoulin and Davison (2005) proposed to let the parameters of an extreme value model vary as smooth functions of covariates. This has been intensively applied to oceanographic data (e.g. Feld et al., 2014; Jonathan et al., 2014; Ross et al., 2017, ...). However, there are only a few papers dealing with non-stationary EGPD (de Carvalho et al., 2021; Haruna et al., 2021). In this study, and as illustrated in Figure 2.9, the marginal EGPD models parameters will be described as smooth functions of the covariates  $T_p$  and  $D_p$ .

The key steps of our study are the following: (1) marginal regression modelling within the class of EGPD, (2) transformation of the data to common exponential margins, (3) modelling extremal dependence between the variables using multivariate generalised Pareto model (hereafter MGP models) (Rootzén and Tajvidi, 2006), (4) non-parametric simulation of bivariate extreme  $H_s$  within the class of MGP distributions. In our modelling scheme, different steps are novelties. To our knowledge, little attention has been paid in the literature to the modelling of mul-

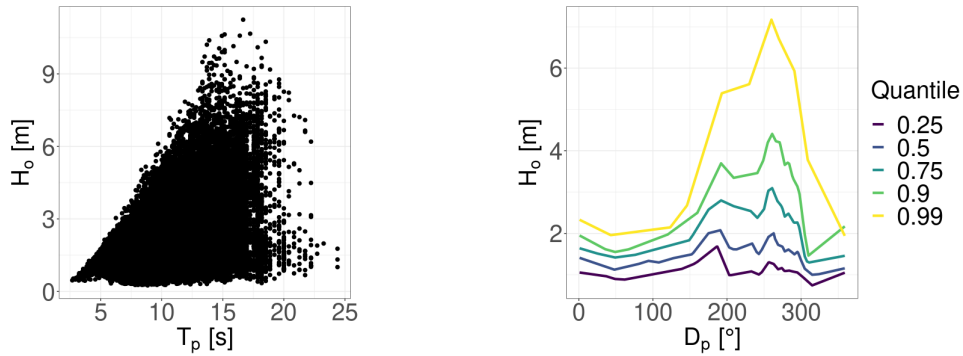


Figure 2.9: Dependence of  $H_s$  with respect to  $T_p$  and  $D_p$  (left) Offshore significant wave heights  $H_o$  given peak period  $T_p$ . (right) Estimated quantiles of  $H_o$  given the peak direction  $D_p$  for different quantile levels  $q \in \{0.25, 0.5, 0.75, 0.9, 0.99\}$ , estimation is performed using smoothed quantile regression (Koenker et al., 1994).

tivariate non-stationary extremes using EGPD and to the non-parametric simulation within the MGP class.

Our paper is organised as follows. In Section 2.3, the sea state data are presented and the marginal inference incorporating covariates in the EGPD modelling is described in Section 2.4. In Section 2.5, the non-parametric method to simulate MGP vectors is presented and some numerical experiments are shown. Two algorithms are outlined, one for bivariate simulations and a second one for conditional simulations. Both algorithms are applied in Section 2.6 to the sea state data.

## 2.3 Sea state data

Our study is carried out in the northern part of the Bay of Biscay in France. The specificity of this region is that it is exposed to the Atlantic Ocean and therefore subject to complex superpositions of wind generated waves and swell. The data are extracted from two different wave hindcasts provided by IFREMER and consist of simulations of sea states by a numerical model. First, the IOWAGA database (Ardhuin and Accensi, 2014) corresponds to sea states parameters that are generated by the wave model WAVEWATCH-III and forced by CFSR winds on a Global grid (0.5° resolution grid in latitude and longitude). HOMERE is the second database (Accensi and Maisondieu, 2015), also based on WAVEWATCH-III model and forced by IOWAGA on the wet boundaries, but on an unstructured grid covering only the English Channel and the Bay of Biscay, more refined close to the coast and with the inclusion of currents and water levels.

As mentioned in the introduction, we restrict our attention to two specific locations: an offshore grid point (47°N, 3°W) from the IOWAGA database and a coastal point from the HOMERE database, near the French coast (47°24N, 2°78W) cor-

responding to the SEM-REV sea test site (Mouslim et al., 2009) (see Figure 2.7). Among all the sea states parameters, the significant wave heights  $H_s[m]$  from the two locations, the peak period  $T_p[s]$  and the peak direction  $D_p[^\circ]$  only from the offshore location are used. Data are available at 3-hour intervals spanning from 1994 to 2016. More precisely, the original HOMERE database has a 1-hour resolution time step but the IOWAGA database is sampled every 3 hours, so to obtain data at the same time scale a sub-sampling of the HOMERE database every 3 hours is performed. In the following,  $H_c$  (resp.  $H_o$ ) denotes the coastal (resp. offshore) significant wave heights. A scatter plot between  $H_c$  and  $H_o$  can be found in Figure 2.7, highlighting a strong dependence structure between the variables. Data are then split into two sets:

- Set 1 contains the first 70% of the data and is used for the inference of the marginal regression models and the preliminary steps for the simulation of  $H_s$  (see Section 2.6);
- Set 2 contains the remaining 30% of the data and is used for the simulation of extreme  $H_s$ .

## 2.4 Marginal regression analysis

### 2.4.1 Marginal regression

In this section, only the data from Set 1 are considered. A regression model for  $H_c$  and  $H_o$  is chosen. We pre-select the extremes by considering, within the Set 1, data such that  $H_o > v_o$ , i.e. belonging to the right rectangular region in Figure 2.7, where  $v_o$  is defined as the 0.98 quantile of  $H_o$ . A common choice when someone is interested in extreme values is to work with the class of the generalised Pareto distributions (GPD) (e.g. Coles (2001)). However, this type of model always raises questions on the choice of the threshold and the GPD approximation holds true only for the very high values. In our case we want to model all the data that are above the blue line in Figure 2.7, this means that values are not necessarily extremes. To overcome such problems, Naveau et al. (2016) proposed a new class of extreme value distributions, called extended generalised Pareto distributions (EGPD). The EGPD class is suitable for modelling the entire range of data, not only the most extreme values, and avoids the need for careful threshold selection (see Figure 2.8). In Naveau et al. (2016), four parametric models are proposed. We restrict ourselves to the first and simplest one (corresponding to the EGP3 model introduced by Papastathopoulos and Tawn (2013)), which appears to be flexible enough, and whose cumulative distribution function is of the form

$$F(x) = \left( 1 - \left( 1 + \frac{\xi x}{\sigma} \right)^{-1/\xi} \right)^\kappa. \quad (2.1)$$

The model has three parameters: scale  $\sigma > 0$ , shape  $\xi \in \mathbb{R}$  and an additional parameter  $\kappa > 0$  which controls the shape of the lower tail. In [Naveau et al. \(2016\)](#), the authors developed the EGD for non-negative shape parameter  $\xi$ . Indeed, the main applications were the modelling of daily rainfall (e.g. [Naveau et al., 2016](#); [Tencaliec et al., 2019](#); [de Carvalho et al., 2021](#); [Rivoire et al., 2021](#)), which are heavy-tailed ( $\xi > 0$ ). In our case, as reported in [Jonathan and Ewans \(2013\)](#), extreme  $H_s$  data are generally described by upper-bounded tail distributions. Still, the case  $\xi < 0$  can be handled by model (2.2).

As mentioned in the Introduction,  $H_s$  data are non-stationary (see Figure 2.9), see also [Jonathan et al. \(2014\)](#); [Feld et al. \(2014\)](#); [De Leo et al. \(2021\)](#). Therefore, we regress  $H_s$  on the peak direction  $D_p$  and the peak period  $T_p$ . We choose here to put the dependency on the scale parameter. The regression marginal models can then be written as follows

$$\begin{cases} \mathbb{P}(H_o - v_o \leq x | H_o > v_o, T_p, D_p) = \left( 1 - \left( 1 + \frac{\xi_o x}{\sigma_o(T_p, D_p)} \right)^{-1/\xi_o} \right)^{\kappa_o}, \\ \mathbb{P}(H_c - v_c \leq x | H_o > v_o, T_p, D_p) = \left( 1 - \left( 1 + \frac{\xi_c x}{\sigma_c(T_p, D_p)} \right)^{-1/\xi_c} \right)^{\kappa_c}. \end{cases} \quad (2.2)$$

The coastal marginal threshold  $v_c$  introduced in (2.2) is defined by  $v_c := \min(H_c; H_o > v_o)$  so that the minimum of  $H_c - v_c$  is equal to zero. The regression marginal models (2.2) are estimated using the R package `gam1ss` ([Stasinopoulos et al., 2008](#)) with the EGD family ([Le Carrer, 2022](#)). The inference is performed using maximum penalised likelihood estimation (note that the model fitting is achieved with the CG algorithm ([Cole and Green, 1992](#))). In our model (2.2), we assume that the parameters  $\sigma_o$  and  $\sigma_c$  vary smoothly with  $T_p$  and  $D_p$ . This is achieved using tensor product of cubic regression splines. The parameter estimates for the regression marginal models are reported in Table 2.2. Asymptotic 95% confidence intervals for each parameter are given in brackets and derived from the asymptotic variance-covariance matrix of the fitted models.

	$\xi$	$\kappa$
Coast	-0.11[-0.15, -0.07]	4.11 [3.57, 4.64]
Offshore	-0.10[-0.16, -0.04]	1.16 [1.05, 1.26]

Table 2.2: Estimated parameters for the regression marginal models. 95% asymptotic confidence intervals are given in brackets.

Both estimated shape parameters are negative but close to zero, which suggests light-tailed or bounded distributions. This is in accordance with previous studies and the physical behaviour of wave heights in shallow waters ([Castillo and Sarabia, 1992](#); [Vanem and Fazerer-Ferradosa, 2022](#)). The goodness of fit of the model (2.2) are shown in Appendix 2.A.



## 2.4.2 Covariates effects

To visualise the effect of the covariates on the scale parameters, we choose to consider the theoretical expectation of the fitted EGPD models. As can be seen from Equation (2.3) (and similarly for  $H_o$ ), the theoretical expectation of model (2.2) is directly proportional to the scale parameter (see [Naveau et al. \(2016\)](#)):

$$\mathbb{E}(H_c | H_o > v_o, T_p, D_p) = \sigma_c(T_p, D_p) \frac{1}{\xi_c} [\kappa_c B(\kappa_c, 1 - \xi_c) - 1], \quad (2.3)$$

where  $B$  denotes the Beta function defined by

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt.$$

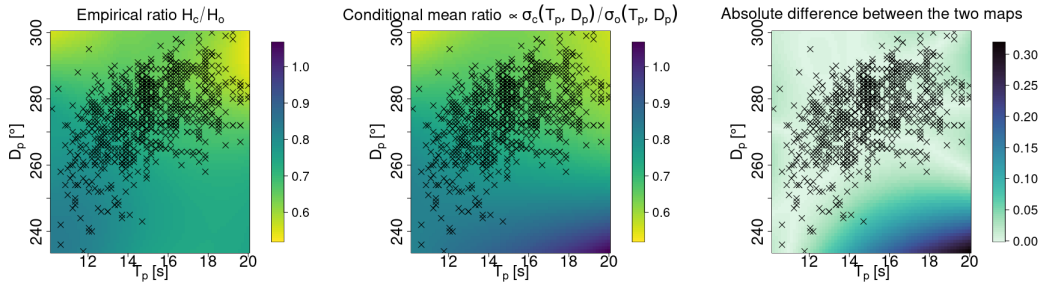


Figure 2.10: (left) Interpolated ratio of empirical extreme coastal significant wave heights  $H_c$  and extreme offshore significant wave heights  $H_o$ . The interpolated surface between the data points is performed using local polynomial interpolation of degree 2. (middle) Ratio of the predicted conditional expectations of extreme coastal significant wave heights  $H_c$  and extreme offshore significant wave heights  $H_o$ , conditionally to the offshore peak direction  $D_p$  and the offshore peak period  $T_p$ . (right) Absolute difference between the interpolated empirical ratio and the predicted ratio. On the three plots, observed data points are superimposed.

The first panel in Figure 2.10 represents the empirical ratio  $H_c/H_o$  given  $T_p$  and  $D_p$  values. Local polynomial interpolation (LOESS, [Cleveland and Devlin \(1988\)](#)) is performed between the observed data points to get values on a regular grid of  $D_p$  and  $T_p$ , ranging from 230 to 300 degrees for the peak direction and from 10 to 20 seconds for the peak period. This first panel is then compared to the estimated ratio of the two conditional expectations (second panel in Figure 2.10), which, from Equation (2.3), is proportional to

$$\hat{\sigma}_c(T_p, D_p) / \hat{\sigma}_o(T_p, D_p).$$

This ratio can give us an idea of the propagation of the wave energy from the offshore to the coast as a function of the covariates. These results can be physically interpreted: the loss of wave energy between the offshore and the coast is lower for small periods but also for waves coming from the SW rather than the NW due to the bathymetry (see the map on Figure 2.7). In the third panel of Figure 2.10, the dark blue region, corresponding to  $T_p > 16\text{s}$  and  $D_p < 255^\circ$ , has no data and consequently, produces large differences due to extrapolation issues.

Using the estimated  $\hat{\sigma}_c(T_p, D_p)$  and  $\hat{\sigma}_o(T_p, D_p)$ , the  $H_s$  data are then transformed to common exponential scale using the probability integral transform

$$\begin{aligned} H_o^E &:= -\log \left\{ 1 - \hat{F}_o[(H_o - v_o)/\hat{\sigma}_o(T_p, D_p)] \right\}, \\ H_c^E &:= -\log \left\{ 1 - \hat{F}_c[(H_c - v_c)/\hat{\sigma}_c(T_p, D_p)] \right\}, \end{aligned} \quad (2.4)$$

where  $\hat{F}_o$  (resp.  $\hat{F}_c$ ) is the fitted  $EGPD(\hat{\xi}_o, \hat{\kappa}_o, 1)$  cdf's (resp.  $EGPD(\hat{\xi}_c, \hat{\kappa}_c, 1)$ ) from Equation (2.2).

## 2.5 Multivariate Pareto model

In this section, the threshold exceedances of  $H_s$  transformed to common exponential margins are modelled. This vector is denoted  $\mathbf{H}^E := (H_c^E, H_o^E)$  in the following. For that, we adapt the definition of [Rootzén and Tajvidi \(2006\)](#) of bivariate threshold exceedances given as

$$[\mathbf{H}^E - \mathbf{u} | \mathbf{H}^E \not\leq \mathbf{u}] \quad (2.5)$$

where  $\mathbf{u} := (u_c, u_o) \in \mathbb{R}_+^2$  and  $\mathbf{H}^E \not\leq \mathbf{u}$  means that  $H_c^E > u_c$  and/or  $H_o^E > u_o$ , that is to say we are extreme in at least one of the two components. Then multivariate EVT theory states that (2.5) can be well approximated by a multivariate generalised Pareto (MGP) distribution ([Rootzén and Tajvidi, 2006](#)). Note that there are different equivalent definitions for multivariate threshold exceedances (see Section 8.3 of [Beirlant et al. \(2004\)](#)).

[Rootzén et al. \(2018a\)](#) derived a stochastic representation of standard MGP vectors considering that a bivariate random vector  $\mathbf{Z}$  follows a standard MGP distribution if, and only if,

$$\mathbf{Z} = E + \mathbf{T} - \max(\mathbf{T}), \quad (2.6)$$

with  $\mathbf{T}$  a random vector and  $E$  a unit exponential random variable independent of  $\mathbf{T}$ . Note that the standard MGP distribution is supported by the set  $L := \{\mathbf{x} \in \mathbb{R}^d; \mathbf{x} \not\leq \mathbf{0}\}$ .

In our study, Equation (2.6) is adapted to  $\mathbf{Z} = (Z_1, Z_2)$  defined as

$$\begin{cases} Z_1 := H_o^E - u_o | H_o^E > u_o \text{ or } H_c^E > u_c, \\ Z_2 := H_c^E - u_c | H_o^E > u_o \text{ or } H_c^E > u_c. \end{cases} \quad (2.7)$$

### 2.5.1 Simulation of bivariate standard generalised Pareto distributed vectors

Kiriliouk et al. (2019) established several parametric MGP models by setting explicit densities for  $\mathbf{T}$  in a multivariate setting. In the following, we consider only vectors of dimension 2, *i.e.*  $\mathbf{Z} = (Z_1, Z_2)$  and  $\mathbf{T} = (T_1, T_2)$ . To bypass the choice of the underlying distribution for  $\mathbf{T}$ , we start from the following rewriting of Equation (2.6),

$$\begin{cases} Z_1 = E + \Delta \mathbb{1}_{\Delta < 0}, \\ Z_2 = E - \Delta \mathbb{1}_{\Delta \geq 0}, \end{cases} \quad (2.8)$$

where  $\Delta := Z_1 - Z_2 = T_1 - T_2$  and  $\mathbb{1}_A$  denotes the indicator function, equals to 1 if  $A$  is true and 0 otherwise.

Equation (2.8) is the basis for our simulation algorithms. From this equation, we see that we need to simulate values of  $\Delta$  and  $E$  independently, instead of  $(T_1, T_2)$ . Generating independent, and identically distributed, unit exponential is trivial, so the main difficulty is to simulate  $\Delta$ . This can be achieved by bootstrapping (see Efron (1979)). Our approach is then described in Algorithm 1 and a theoretical proof can be found in Appendix 2.B.

---

#### Algorithm 1 Non-parametric bootstrap MGP simulation

---

- 1: **input** A sample  $(Z_{1,i}, Z_{2,i})_{1 \leq i \leq n}$  from a MGP distribution
  - 2: **output** A simulated sample  $(Z_{1,k}^{(m)}, Z_{2,k}^{(m)})_{1 \leq k \leq m}$ , potentially with  $m \neq n$
  - 3: **procedure**
  - 4:     Define  $\Delta_i := Z_{1,i} - Z_{2,i}$  for  $1 \leq i \leq n$
  - 5:     Generate  $m$  realisations  $E_k^{(m)} \sim \text{Exp}(1)$ , independently of  $(\Delta_i)_{1 \leq i \leq n}$ , for  $1 \leq k \leq m$
  - 6:     Bootstrap  $m$  realisations  $\Delta_k^{(m)}$ ,  $1 \leq k \leq m$ , from  $(\Delta_1, \dots, \Delta_n)$
  - 7: **end procedure**
  - 8: **return**  $Z_{1,k}^{(m)} := E_k^{(m)} + \Delta_k^{(m)} \mathbb{1}_{\Delta_k^{(m)} < 0}$  and  $Z_{2,k}^{(m)} := E_k^{(m)} - \Delta_k^{(m)} \mathbb{1}_{\Delta_k^{(m)} > 0}$ , for  $1 \leq k \leq m$
- 

### 2.5.2 Numerical experiments

In the following, we simulate MGP vectors  $\mathbf{Z} = (Z_1, Z_2)$  from the representation (2.6) with different parametric models on  $(T_1, T_2)$  and we compare with our simulation algorithm. The different experiments are reported in Table 2.3 and some graphical results are shown in Figure 2.11 which displays for each model a scatter plot of the data, the measure of extremal dependence  $\chi(u)$  for increasing values of  $u$ , and the marginal quantile-quantile plots. We use the measure  $\chi(u)$  which

gives a measure of asymptotic dependence between two variables  $X$  and  $Y$  (for more details see e.g. [Coles et al. \(1999\)](#)) and which is defined by

$$\chi(u) := \mathbb{P}(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)), \quad u \in (0, 1).$$

Bivariate model	Joint distribution of $\mathbf{T}$	Parameters
(a) Gaussian symmetric	$\mathcal{N}\left((\mu_1, \mu_2), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$	$\mu_1 = 0$ $\mu_2 = 0$ $\rho = 0.4$
(b) Gaussian asymmetric	$\mathcal{N}\left((\mu_1, \mu_2), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$	$\mu_1 = 0$ $\mu_2 = 2$ $\rho = 0.4$
(c) Logistic	$F(x_1, x_2) = (1 + e^{-x_1/\sigma_1} + e^{-x_2/\sigma_2})^{-1}$	$\sigma_1 = 1$ $\sigma_2 = 5$
(d) Gumbel	$F(x_1, x_2) = \exp[-\exp\{-x_1/\sigma_1\}] \exp[-\exp\{-x_2/\sigma_2\}]$	$\sigma_1 = 1$ $\sigma_2 = 4$
(e) Exponential	$S(x_1, x_2) = \exp\{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_3 \max(x_1, x_2)\}$	$\lambda_1 = 2$ $\lambda_2 = 10$ $\lambda_3 = 1$

Table 2.3: Overview of the different experiments carried-out. For each, we give the joint distribution  $F(x_1, x_2)$  when it writes easily or the survival function  $S(x_1, x_2)$ . In the third column, we give the different parameters values used in the numerical experiments.

In Table 2.3, (a) and (b) are two bivariate Gaussian models with same correlation coefficient  $\rho < 1$  but with  $\mu_1 \neq \mu_2$  for (b), leading to asymmetry. Model (c) corresponds to the Type I bivariate logistic distribution proposed by [Gumbel \(1961\)](#). For model (d) we consider two independent Gumbel distributed variables with different scale parameters. And lastly, (e) corresponds to a bivariate exponential distribution as defined in [Marshall and Olkin \(1967\)](#).

As seen in Figure 2.11, Algorithm 1 successfully simulates draws from the parametric simulations in terms of the marginals  $Z_1$  and  $Z_2$ , but also recovers well the dependence structure when looking at the measure of dependence  $\chi(u)$ .

### 2.5.3 Conditional simulation within the MGP class

From an application perspective, we also want to be able to simulate conditionally on one of the two variables. In this section we describe the conditional simulation algorithm for the MGP model.

CHAPTER 2. JOINT STOCHASTIC SIMULATION OF EXTREME COASTAL AND OFFSHORE SIGNIFICANT WAVE HEIGHTS

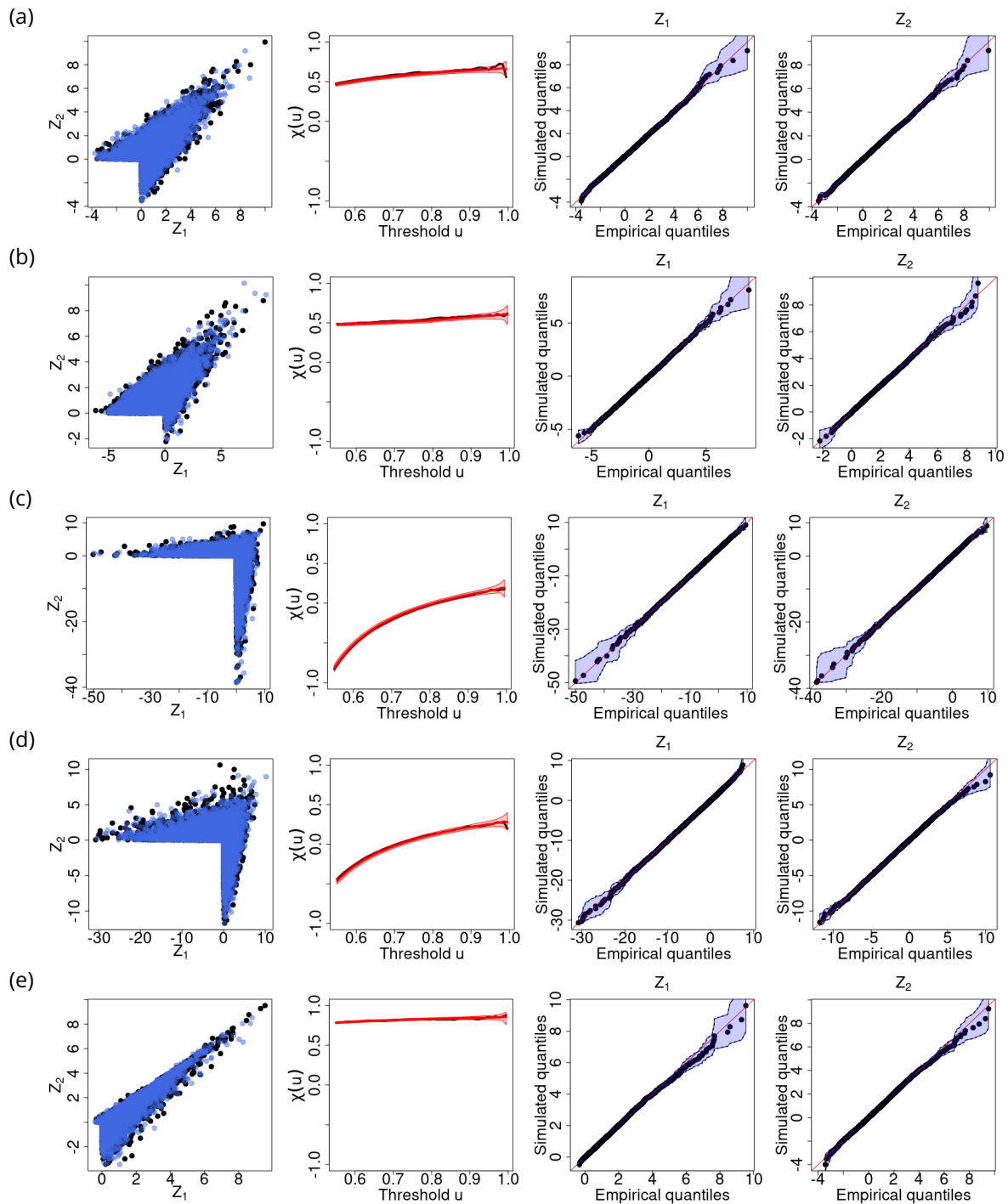


Figure 2.11: Each panel line corresponds to one of the parametric model (a) to (e), and shows from left to right: (1) Scatter plot of simulated data with the parametric model with sample size  $n = 10000$  (black dots) and sampled data from one simulation using Algorithm 1 with sample size  $m = 10000$  (blue dots); (2) Empirical estimates of the measure of asymptotic dependence  $\chi(u)$  for the simulated data with the parametric model (black line) and for the sampled data from Algorithm 1 (red line), with associated 95% pointwise confidence intervals based on 1000 bootstrap replications; (3) and (4) Quantile-quantile plots for  $Z_1$  and  $Z_2$  with associated 95% pointwise confidence intervals based on 1000 bootstrap replications.

From Equation (2.6) we can derive

$$Z_2 = Z_1 + T_2 - T_1 = Z_1 - \Delta. \quad (2.9)$$

From Equation (2.9), we can design a simulation strategy but caution is required because  $\Delta$  and  $Z_1$  are not necessarily independent. But for some values of  $Z_1$ , this will be the case. To see this, one can compute the conditional distribution of  $\Delta$  given  $Z_1 = z_1$  starting from the joint distribution function of  $(Z_1, Z_2)$  which is given by

$$f_{(Z_1, Z_2)}(z_1, z_2) = e^{-\max(z_1, z_2)} f_{\Delta}(z_1 - z_2) \mathbb{1}_{(z_1, z_2) \in L}, \text{ for } z_1, z_2 \in \mathbb{R},$$

where  $f_{\Delta}$  denotes the distribution function of  $\Delta$ .

*1st case:* If  $z_1 > 0$ . In this case, noting that if  $z_1 > 0$  then  $\mathbb{1}_{(z_1, z_2) \in L} = 1$ , the marginal distribution of  $Z_1$  is given as follows

$$f_{Z_1}(z_1) = e^{-z_1} K,$$

where  $K := \int_{-\infty}^0 e^u f_{\Delta}(u) du + \int_0^{\infty} f_{\Delta}(u) du$  and does not depend on  $z_1$ .

Therefore, the conditional distribution of  $Z_2$  given  $Z_1 = z_1$  when  $z_1 > 0$  is given by

$$f_{Z_2|Z_1}(z_2 | z_1) = \frac{1}{K} [f_{\Delta}(z_1 - z_2) \mathbb{1}_{z_1 \geq z_2} + e^{z_1 - z_2} f_{\Delta}(z_1 - z_2) \mathbb{1}_{z_1 < z_2}]$$

From this, the conditional distribution of  $\Delta$  given  $Z_1 = z_1 > 0$  can then be deduced:

$$f_{\Delta|Z_1}(\delta | z_1) = \frac{1}{K} [f_{\Delta}(\delta) \mathbb{1}_{\delta \geq 0} + e^{\delta} f_{\Delta}(\delta) \mathbb{1}_{\delta < 0}]. \quad (2.10)$$

This shows that, conditionally on  $Z_1 > 0$ ,  $\Delta$  does not depend on  $Z_1$ .

*2nd case:* If  $z_1 < 0$ . Then, noting that  $\mathbb{1}_{(z_1, z_2) \in L} = \mathbb{1}_{z_2 > 0}$  if  $z_1 < 0$ , we get

$$f_{Z_1}(z_1) = e^{-z_1} K(z_1),$$

where  $K(z_1) := \int_{-\infty}^{z_1} e^u f_{\Delta}(u) du$ . And we can derive the conditional distribution of  $Z_2$  given  $Z_1 = z_1 < 0$  as follows

$$f_{Z_2|Z_1}(z_2 | z_1) = \frac{1}{K(z_1)} e^{z_1 - z_2} f_{\Delta}(z_1 - z_2) \mathbb{1}_{z_2 > 0}.$$

The conditional distribution of  $\Delta$  given  $Z_1 = z_1 < 0$  is then given by

$$f_{\Delta|Z_1}(\delta | z_1) = \frac{1}{K(z_1)} e^{\delta} f_{\Delta}(\delta) \mathbb{1}_{\delta < z_1}. \quad (2.11)$$

From Equations (2.10) and (2.11) we derive the conditional simulation algorithm described in Algorithm 2, where the simulation procedure is split into two cases:

1. If  $z_1 > 0$ , we can sample values of  $\Delta$  independently of  $Z_1$ ,
2. otherwise, if  $z_1 < 0$ , we use a rejection sampling approach to approximate the targeted conditional density in Equation (2.11).

Similarly, we could also derive a simulation scheme of  $Z_1$  given  $Z_2 = z_2$ .

**Algorithm 2** Non-parametric conditional MGP simulation

---

```

1: input A sample  $(\Delta_i)_{1 \leq i \leq n}$ ; a realisation  $z_1$  of  $Z_1$ 
2: output A simulated sample  $(Z_{2,k}^{(m)})_{1 \leq k \leq m}$  conditionally on  $Z_1 = z_1$ , potentially
   with  $m \neq n$ 
3: procedure
4:   if  $z_1 > 0$  then
5:     Define  $\Delta_{|Z_1^+}$  the subset of  $(\Delta_i)_{1 \leq i \leq n}$  such that  $Z_1 > 0$ 
6:     Bootstrap  $m$  realisations  $\Delta_k^{(m)}$ ,  $1 \leq k \leq m$ , from  $\Delta_{|Z_1^+}$  independently of
    $Z_1$ 
7:   else
8:     for  $1 \leq k \leq m$  do
9:       Sample one realisation  $\Delta_k^{(m)}$  from  $(\Delta_i)_{1 \leq i \leq n}$  independently of  $Z_1$ 
10:      Generate a random number  $u \in [0, 1]$ 
11:      while  $u > \exp(\Delta_k^{(m)}) \mathbb{1}_{\Delta_k^{(m)} < z_1}$  do
12:        Repeat steps 9 and 10
13:      end for
14: end procedure
15: return  $Z_{2,k}^{(m)} := z_1 - \Delta_k^{(m)}$  for  $1 \leq k \leq m$ 

```

---

### 2.5.4 Numerical experiment continued

As for the bivariate simulations, we can illustrate Algorithm 2 with numerical experiments. We choose here to show the results only for Model (a) (Symmetric Gaussian) since for this specific model we have an explicit form for the theoretical distribution of  $\Delta$ . The results are presented in Figure 2.12 where we simulated the conditional distribution of  $Z_2$  for eight different conditioning values. The sampled and theoretical conditional distributions appear to be in close conformity.

## 2.6 Application to extreme significant wave height

The methodology presented in Section 2.5 is applied to the joint and the conditional simulations of extreme significant wave heights. For that, the sample of bivariate threshold exceedances  $(Z_1, Z_2)$  defined in Equation (2.7) is used as input data for Algorithm 1 or Algorithm 2. The thresholds  $u_o$  and  $u_c$  in (2.7) are defined as the 0.8 quantile of  $H_o^E$ , or equivalently of  $H_c^E$ .

Recall that in both cases, simulations are performed on the exponential scale. A final step of back transformation is then necessary to get simulations of  $H_s$  on the original scale. This final step corresponds to part 3 (resp. 5) in the following procedure for the joint (resp. conditional) simulation of  $H_s$ . For the sake of clarity we now divide the joint and the conditional simulation scheme of  $H_s$  in two

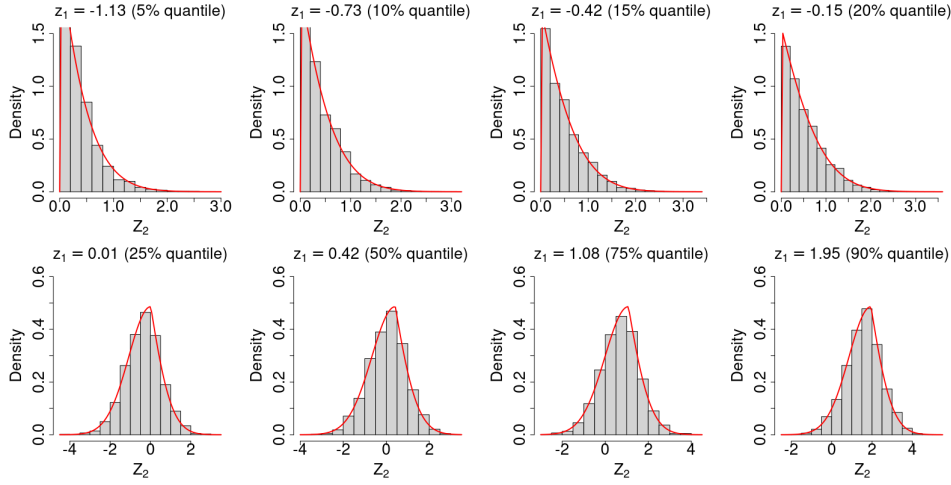


Figure 2.12: Sampled conditional distribution of  $Z_2$  given  $Z_1 = z_1$  using Algorithm 2 for the asymmetric Gaussian model with sample size  $n = 10000$ . Eight experiments are presented for different quantiles of  $Z_1$  whose values are reported in each panel title. The sample size for each simulation is  $m = 10000$ . The theoretical conditional density is superimposed in red.

separate sections.

### 2.6.1 Joint simulation of significant wave heights

The joint simulation scheme for extreme  $H_s$  is described hereafter. In the following we fix the pair value  $(t_p, d_p) \in \mathbb{R}^2$  which may be taken from Set 2.

1. Compute  $\hat{\sigma}_o(t_p, d_p)$  and  $\hat{\sigma}_c(t_p, d_p)$  from the marginal EGD models fitted on Set 1 (see Section 2.4).
2. Simulate  $m$  pairs of  $(z_1, z_2)$  applying Algorithm 1 with input data  $(Z_1, Z_2)$  as defined in (2.7). We therefore obtain  $m$  simulated pairs  $((z_{1,1}, z_{2,1}), \dots, (z_{1,m}, z_{2,m}))$  for a fixed value  $(t_p, d_p)$ .
3. Transform the simulated values to the original scale

$$\begin{aligned} h_{o,i} &:= \hat{\sigma}_o(t_p, d_p) \hat{F}_o^{-1}(1 - e^{-(z_{1,i} + u_o)}) + v_o \in \mathbb{R}^m, \\ h_{c,i} &:= \hat{\sigma}_c(t_p, d_p) \hat{F}_c^{-1}(1 - e^{-(z_{2,i} + u_c)}) + v_c \in \mathbb{R}^m, \end{aligned}$$

where  $\hat{F}_o^{-1}$  (resp.  $\hat{F}_c^{-1}$ ) is the inverse cdf of the  $EGPD(\hat{\xi}_o, \hat{\kappa}_o, 1)$  (resp.  $EGPD(\hat{\xi}_c, \hat{\kappa}_c, 1)$ ) estimated in Section 2.4.

This procedure is then applied to four selected pairs  $(t_p, d_p)$  from the Set 2 corresponding to the four largest  $H_c$  of this dataset. Figure 2.13 depicts simulated pairs of offshore and coastal  $H_s$  with simulation sample size  $m = 1000$ .



Figure 2.13 shows that for these specific conditions, Algorithm 1 successfully generates extreme  $H_c$  and  $H_o$ . Note that since the four points considered are the largest observations, they are expected to be among the extremes of the simulated distributions.

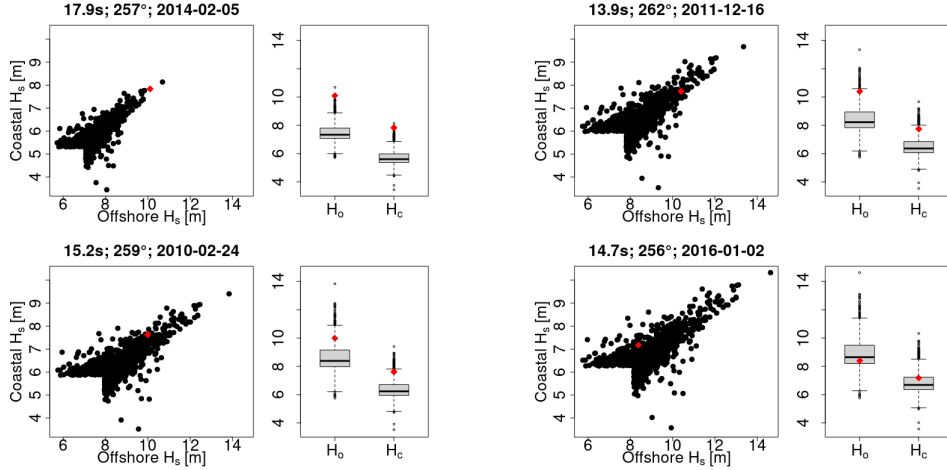


Figure 2.13: Sampled values of coastal versus offshore significant wave heights from Algorithm 1,  $m = 1000$  pairs of points are simulated. Each panel corresponds to a fixed value of  $(t_p, d_p)$  corresponding to the four largest  $H_c$  from Set 2 (in decreasing order). Next to each scatter plot, marginal distributions of simulated  $H_o$  and  $H_c$  are depicted with boxplots. On both the scatter plots and the boxplots, the red diamond-shaped points represent the true values of the coastal and offshore significant wave heights for the four pairs.

	Date-Time	$H_c$ [m]	$H_o$ [m]	$T_p$ [s]	$D_p$ [°]	Joint probability
1	2014-02-05 12:00:00 GMT	7.83	10.1	17.86	257	0.002
2	2011-12-16 03:00:00 GMT	7.74	10.4	13.89	262	0.045
3	2010-02-24 15:00:00 GMT	7.62	10.0	15.15	259	0.057
4	2016-01-02 06:00:00 GMT	7.18	8.4	14.71	256	0.277
5	2013-12-24 06:00:00 GMT	7.08	8.4	14.71	253	0.388
6	2014-02-14 21:00:00 GMT	7.06	8.3	14.08	248	0.691
7	2015-01-15 09:00:00 GMT	6.45	7.9	13.70	260	0.489
8	2016-03-28 03:00:00 GMT	6.30	7.6	14.08	256	0.762

Table 2.4: Empirical joint survival probability of exceeding the observed extreme significant wave heights  $h_c$  and  $h_o$  for the eight largest coastal significant wave heights of Set 2. Only events from different storms are given (i.e. events separated by more than 3 hours). Estimation is performed using  $m = 1.10^6$  simulated pairs  $(h_{o,i}, h_{c,i})_{1 \leq i \leq m}$  for each largest event with Algorithm 1.

Table 2.4 describes the eight largest  $h_c$  of Set 2, giving the time event and the corresponding  $h_o, t_p$  and  $d_p$  values. The last column of Table 2.4 gives the empirical estimate of the joint survival probability defined as  $1/m \sum_{i=1}^m \mathbb{1}\{h_{o,i} > h_o, h_{c,i} > h_c\}$ , applying the above procedure with simulation sample size  $m = 1.10^6$ . This estimated probability quantify the observations made with Figure 2.13: for the most extreme events, the associated joint probabilities are expected to be lower.

## 2.6.2 Conditional simulation of coastal significant wave heights

The conditional simulation scheme for extreme coastal  $H_s$  is described bellow. Note that the procedure is very symmetrical for simulating offshore  $H_s$ . In the following we fix the triplet value  $(h_o, t_p, d_p) \in \mathbb{R}^3$  which may be taken from Set 2.

1. Compute  $\hat{\sigma}_o(t_p, d_p)$  and  $\hat{\sigma}_c(t_p, d_p)$  from the marginal EGPD models fitted on Set 1 (see Section 2.4).
2. Transform  $h_o$  to the standardised space using the probability integral transform:

$$h_o^E = -\log \left\{ 1 - \hat{F}_o[(h_o - v_o)/\hat{\sigma}_o(t_p, d_p)] \right\}$$

where  $\hat{F}_o$  is the  $EGPD(\hat{\xi}_o, \hat{\kappa}_o, 1)$  cdf's.

3. Set  $z_1 := h_o^E - u_o$ , where  $u_o$  is the threshold on the offshore  $H_s$  on the exponential scale.
4. Simulate  $z_2$  applying Algorithm 2, only in the case  $z_1 > 0$ . Here  $\Delta_{|Z_1^+}$  is defined from Set 1 through  $\Delta_{|Z_1^+} := (Z_1 - Z_2)\mathbb{1}\{Z_1 > 0\}$ , with  $Z_1, Z_2$  as defined in (2.7), and bootstrapped  $m$  times. We therefore obtain  $m$  simulations of  $z_2 = (z_{2,1}, \dots, z_{2,m})$  for a fixed triplet  $(h_o, t_p, d_p)$ , given  $z_1 > 0$ .
5. Transform the predicted values to the original scale

$$h_{c,i} := \hat{\sigma}_c(t_p, d_p) \hat{F}_c^{-1}(1 - e^{-(z_{2,i} + u_c)}) + v_c$$

where  $\hat{F}_c^{-1}$  is the inverse cdf of the  $EGPD(\hat{\xi}_c, \hat{\kappa}_c, 1)$  (see Section 2.4).

Note that in Step 4 above, the simulation is restricted to the case when  $z_1 > 0$  for convenience, since our focus is on the simulation of extreme  $H_s$ .

The pseudo-algorithm described above is applied with all the triplet values  $(h_o, t_p, d_p)$  from Set 2, with simulation sample size  $m = 1000$ . These conditional simulations of coastal significant wave heights are then compared to the true values of  $H_c$  from Set 2. The overall coverage probability (*i.e.* the number of times the actual value of coastal  $H_s$  is within the 95% range of the predicted distribution) is equal to 95% and the simulations are shown in Figure 2.14. The simulations and

the true  $H_c$  values (red dots) are most of the time in good agreement. Since no declustering approach has been adopted, consecutive observations, which belong to the same storm event, are kept. They are depicted with identical colour, highlighting a temporal dependence structure between each storm cluster.

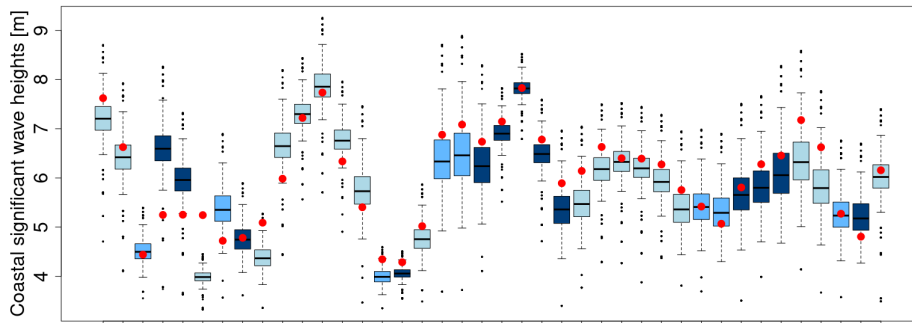


Figure 2.14: Boxplot of predicted  $H_c$  conditionally on  $(H_o, T_p, D_p)$  using Algorithm 2. The simulation sample size is equal to  $m = 1000$  for each observation. Red dots represents the observed  $H_c$  values from Set 2. The alternating colours depict different storms: consecutive boxes with same colour correspond to observations that belong to the same storm (i.e. separated by less than 3 hours).

The effect of the covariates  $T_p$  and  $D_p$  in the conditional simulations is depicted in Figure 2.15, showing that the simulation model is able to simulate both the most intense and the more moderate  $H_c$ . This plot also highlight for which sea state conditions the simulations are far from the observed values. It appears that the two predictions such that the observed  $H_c$  value does not fall within the 95% simulation range correspond to small  $H_c$  and  $H_o$ .

## 2.7 Discussion and conclusions

Simulation of extreme events in a multivariate setting is of great interest to capture not only the statistical behaviour of the extremes, but also the dependence between large values of complex processes. Based on the multivariate EVT, this work presents two non-parametric simulation algorithms of bivariate generalised Pareto distributed variables, without assuming any specific parametric shape for the MGP model. Thanks to Algorithm 1, one can simulate joint extremes. As for Algorithm 2, it allows the simulation of conditional extremes. Both methods have been validated with numerical simulations.

We would like to point out that in the context of bivariate extremes, other simulation algorithms have been developed. For example [Marcon et al. \(2017\)](#)

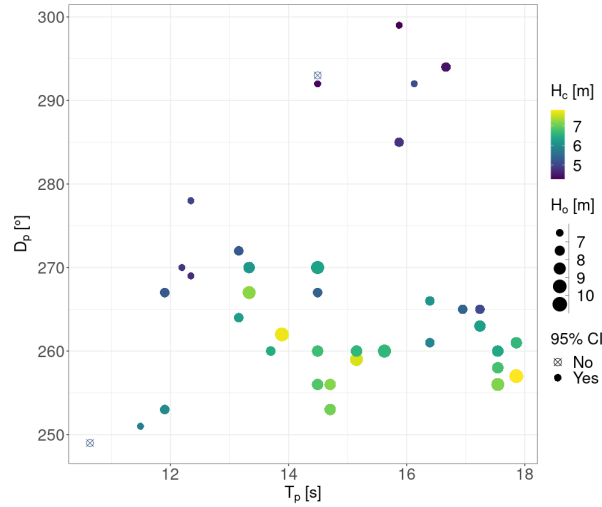


Figure 2.15: Scatter plot from Set 2 with the peak direction  $D_p$  on the y-axis and the peak period  $T_p$  on the x-axis. The dots' colour corresponds to the value of the coastal significant wave heights  $H_c$ . The size of the dots corresponds to the value of the offshore significant wave heights  $H_o$ . The shape indicates if the observed  $H_c$  falls within the 95% range of the predicted distribution from the conditional model where the simulation sample size is equal to  $m = 1000$  for each observation.

proposed a simulation method with a semi-parametric structure for the extremal dependence function, but it was not based on a MGP model and did not cover non-stationarities. Michel (2006) derived a non-parametric simulation framework of bivariate generalised Pareto variables using a different representation of MGP vectors than the one used in this paper.

For application purposes and as a byproduct, a non-stationary marginal modelling with the EGPD was also developed, adding covariate effects on the scale parameter of the EGPD using smoothing splines.

We applied this work to the simulation of extreme significant wave heights near the Brittany coast given specific offshore sea state conditions  $(T_p, D_p)$  with compelling results. In both joint and conditional settings, thanks to the presented algorithms, we are able to simulate realistic extreme  $H_s$  events.

Note that in possible extensions of this work to climate projections, it is assumed that data will not be available at the coastal location but only on a coarse grid, similar to the IOWAGA Global hindcast. This argument is illustrated in Table 2.1, and is in favour of the first pre-selection of the  $H_s$  data through  $\{H_o > v_o\}$  for the marginal regression analysis (see Section 2.4).

Extensions to the multivariate case will be the subject of future works. Considering more than two locations raises different modelling issues. It would also be interesting to apply this methodology to other locations in order to ensure the proper generalisation of the methodology.

## Acknowledgements

The authors would like to thank Noémie Le Carrer for helpful discussion on the implementation of the EGPD family within the `gam1ss` package.

The authors are indebted to the Laboratoire d’Océanographie Physique et Spatiale, IFREMER for providing the data used in this study.

Part of this work was supported by the DAMOCLES-COST-ACTION on compound events, the French national program (FRAISE-LEFE/INSU and 80 PRIME CNRS-INSU), and the European H2020 XAIDA (Grant agreement ID: 101003469). The authors also acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project), the ANR-Melody (ANR-19-CE46-0011).

## 2.A Marginal regression modelling

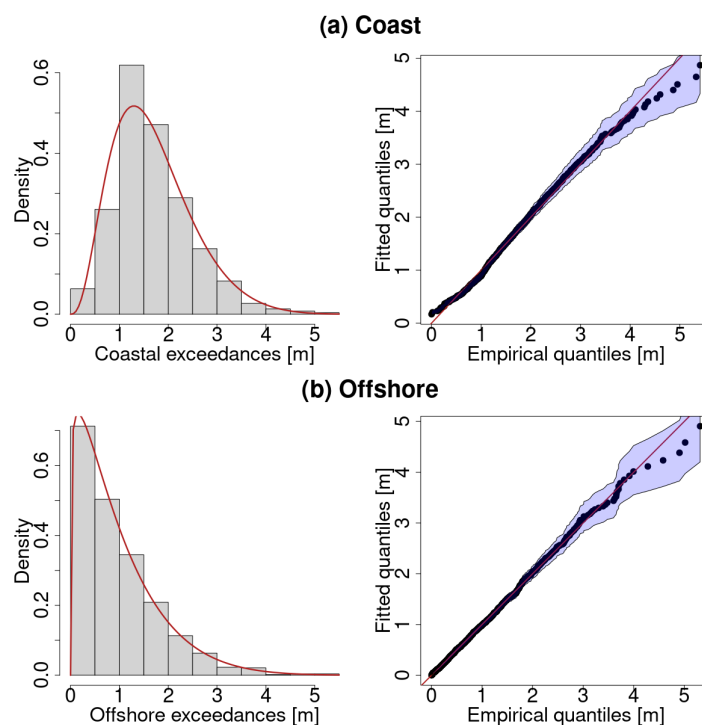


Figure 2.16: (left) Empirical histogram (grey) of the *standardised* extreme significant wave heights exceedances (a) at the coast and (b) offshore. The fitted EGPD density is superimposed. (right) The corresponding quantile-quantile plots with associated 95% pointwise confidence intervals computed using parametric bootstrap.

We show here the goodness of fit for the marginal regression models defined in Equation (2.2). As our models depend on some covariates, the diagnostic plots presented here are built for the *standardised*  $H_s$  exceedances which are defined as  $(H_c - v_c) / \sigma_c(T_p, D_p)$  (and similarly for  $H_o$ ). From Figure 2.16, one can see that the fits seem to be fairly good, a slight discrepancy in the lower values can be noticed for the coastal model but this is not a major issue as the interest lies mainly in the larger values.

## 2.B Proof of Algorithm 1

With the same notations as in Algorithm 1, let  $F$  be the common distribution function of  $\Delta_1, \dots, \Delta_n$  and  $F_{nm}^{(m)}$  be the empirical distribution function of the bootstrap sample  $\Delta_1^{(m)}, \dots, \Delta_m^{(m)}$ .

**Lemma 2.1.** *If  $F_{nm}^{(m)}$  converges in distribution to  $F$ , as  $n$  and  $m$  tend to infinity, then  $(Z_{1,k}^{(m)}, Z_{2,k}^{(m)})_{1 \leq k \leq m}$  converge in distribution to a bivariate GPD  $G$  where  $G$  is the common distribution function of the sample  $(Z_{1,i}, Z_{2,i})_{1 \leq i \leq n}$ .*

*Proof.* As  $P(E \leq u) = 1 - \min(1, \exp(-u))$  for any  $u \in \mathbb{R}$  if  $E \sim \text{Exp}(1)$ , the bivariate distribution function of  $(Z_1^{(m)}, Z_2^{(m)})$  is equal to

$$\begin{aligned} \mathbb{P} \left[ (Z_1^{(m)}, Z_2^{(m)}) \leq (z_1, z_2) \right] &= 1 - \mathbb{E} \left[ \min \left( 1, e^{-\min(z_1 - \Delta^{(m)} \mathbf{1}_{\Delta^{(m)} < 0}, z_2 + \Delta^{(m)} \mathbf{1}_{\Delta^{(m)} > 0})} \right) \right] \\ &= 1 - \mathbb{E} \left[ \min \left( 1, e^{-\min(z_1 - \Delta^{(m)}, z_2) - \max(\Delta^{(m)}, 0)} \right) \right], \end{aligned}$$

for any  $(z_1, z_2) \in \{\mathbf{x} \in \mathbb{R}^2; \mathbf{x} \not\leq 0\}$ .

Then, one can show that the function  $x \mapsto \min(1, x)$ , defined for  $x \geq 0$ , is Lipschitz and bounded by 1. And applying the Portmanteau theorem, we have, letting  $\min(n, m) \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P} \left[ (Z_1^{(m)}, Z_2^{(m)}) \leq (z_1, z_2) \mid \Delta_1, \dots, \Delta_n \right] &\rightarrow 1 - \mathbb{E} \left[ \min \left( 1, e^{-\min(z_1 - \Delta, z_2) - \max(\Delta, 0)} \right) \right] \\ &= 1 - \mathbb{E} \left[ \min \left( 1, e^{\max(T_1 - z_1, T_2 - z_2) - \max(T_1, T_2)} \right) \right]. \end{aligned}$$

Which is the cumulative distribution function of the MGP vector  $(Z_1, Z_2)$  as defined in [Rootzén et al. \(2018a\)](#) (Prop. 8).  $\square$

The assumption in Lemma 2.1 is linked to the bootstrap asymptotic theory (e.g. [Bickel and Freedman \(1981\)](#)).

## 2.3 Summary of Paper I

### To sum up

---

- Paper I addressed the specific issue of joint simulation of extreme environmental variables.
- In the context of asymptotic dependence, multivariate EVT provides a solid mathematical framework.
- Marginal treatment: non-stationary EGPD modelling, taking into account covariate effects on the scale parameter through smoothing splines.
- Dependence treatment: development of non-parametric simulation of bivariate GP vectors, based on a simple rewriting.
- Two algorithms for the simulation of extreme  $H_s$  given offshore conditions ( $T_p, D_p, H_o$  moderately high), are outlined:
  - A first stochastic simulator that produces jointly offshore and coastal extreme significant wave heights.
  - A second stochastic simulator that produces coastal extreme significant wave heights given offshore significant wave heights.
- ↗ Underlying hypothesis of asymptotic dependence between  $H_o$  and  $H_c$ , otherwise other models should be considered.
- ↗ Focus on a specific site: to ensure the generalisation of the method, other locations should be investigated.

# Chapter 3

## Evaluation of binary classifiers for asymptotically dependent and independent extremes

### Overview of Chapter 3

This chapter addresses another issue, that can be seen as a subsequent step to the work of Chapter 2. Instead of building one prediction model, assume that we have several prediction models available - or rather several forecasts - and we want to compare the performance of each model. Since this thesis focuses on EVT, this comparison should be made given their capacity to capture extreme occurrences.

In the following, we study the specific case of binary classifiers, which are the simplest type of forecasting and decision-making situation: an extreme event did or did not occur. Such classifier tailored for extremes will be called an *extremal classifier* and risk functions that answer our initial question will be developed. Their properties will be derived under the framework of multivariate regular variation and hidden regular variation, allowing to handle finer - or more specific - types of asymptotic independence.

Section 3.1 gives some preliminary notions on multivariate regular variations and the refinement to hidden regular variation. Their definition and some examples are provided.

Section 3.2 up to Section 3.6 reproduces Paper II ([Legrand et al., 2021](#)). First, the construction of our risk functions is developed by investigating some specific cases: beginning with two naive classifiers and then looking at a particular type of asymptotic independence (corresponding



to hidden regular variation). We then propose an empirical estimator of our risk functions, whose inferential properties are derived under the framework of multivariate regular variation and hidden regular variation. A simulation study compares different classifiers and indicates their performance with respect to our risk functions. To conclude, we apply our framework to the analysis of extreme river discharges in the Danube river basin. The application compares different predictive algorithms and tests their capacity at forecasting river discharges from other river stations.

Finally, Section 3.7 gives some further developments. We study in detail the special class of linear classifiers and show that the optimisation of our risk functions leads to a consistent solution. A tool to identify the explanatory variables that contribute the most to extremal behaviour is developed and applied to the river network data. To conclude, ideas for extending this pre-selection tool to the case of hidden regular variation are presented.

### 3.1 Preamble to Paper II

In the previous chapter, we proposed a stochastic simulator for bivariate extremes in the case of asymptotic dependence. Specifically, the conditional Algorithm 2 proposed in Paper I could be viewed as a statistical method to predict extreme values of a given variable (e.g. extreme coastal significant wave heights). Now imagine that we have a prediction model for the entire range of data, not only the largest values. Then, a natural question in EVT is whether the model performs well in predicting the most extreme values, and can we discriminate between different models.

This question is in some ways similar to the notion of scoring rules used in the forecasting literature where the aim is to provide summary measures of how well a probabilistic forecast performs compared to the reality, but also in comparison with other competing forecasts. A good introduction can be found, for example, in [Jolliffe and Stephenson \(2003\)](#). Evaluation of probabilistic forecasts is beyond the scope of this manuscript, but we can mention some recent advances in forecast of extreme events, reflecting various challenges that it raises. For example, considering rare events as binary outputs (which is actually analogous to our approach as explained below), [Stephenson et al. \(2008\)](#) discussed the advantages and drawbacks of various risk functions. To obtain accurate extreme wind gusts predictions, [Friederichs and Thorarinsdottir \(2012\)](#) derived closed-form expressions of the continuous ranked probability score (CRPS) (e.g. [Gneiting et al., 2007](#)) for extreme value-distributed data. A weighted CRPS ([Gneiting and Ranjan, 2011](#)) has been applied to compare spatio-temporal predictions of extreme sea surface temperatures during the EVA 2019 data competition ([Huser, 2021](#)). Finally,

Brehmer and Strokorb (2019) and Taillardat et al. (2019) showed the limitations of scoring functions to assess tail properties.

### 3.1.1 Binary classifiers and extremal risk

In Paper II (Legrand et al., 2021) hereunder, we will not consider forecast models directly but rather binary classifiers. In this context, the previous question now becomes how to assess and compare binary classifiers, based upon a set of multivariate observations  $\mathbf{X}$  and denoted  $g(\mathbf{X}; u)$ , with respect to their capacity to capture ‘extreme’ occurrences of a binary response  $Y^{(u)} \in \{-1, +1\}$ . To set the scene, one can imagine that the labels of  $Y^{(u)}$  encode the fact that a given random variable  $H$  is above (i.e. extreme) or below (i.e. non-extreme) a given threshold  $u$  (hence the above notations which depend on  $u$ ):

$$Y^{(u)} = \begin{cases} +1, & \text{if } H > u, \\ -1, & \text{otherwise.} \end{cases}$$

As is common practice in EVT, we will consider high thresholds  $u$  and call  $g(\mathbf{X}; u)$  an *extremal* classifier if it verifies some convergence conditions as  $u$  becomes large. But then, and due to the inherent rarities of extreme events, there will be many more  $-1$  labels than  $+1$ . This leads somehow to a binary classification problem with imbalanced classes and a wealth of models have been developed in the machine learning community to cope with imbalanced data classification (see, e.g. Haixiang et al. (2017) for a recent review).

In this context, our objective is not to construct such classifiers but rather to design an appropriate risk function allowing for their comparison. This risk function will be defined as

$$R^{(u)}(g) = \frac{\mathbb{P}(g(\mathbf{X}; u) \neq Y^{(u)})}{\mathbb{P}(Y^{(u)} = 1 \text{ or } g(\mathbf{X}; u) = 1)}$$

and will depend on the threshold  $u$  considered. Subsequently, and in the same vein as the tail dependence coefficient  $\chi(u)$  (Coles et al., 1999), one may want to look at  $R^{(u)}$  for high levels of  $u > 0$  to focus on the performance of  $g(\mathbf{X}; u)$  in the extremes.

This risk function  $R^{(u)}$  can then be used to compare different predictive models and test their capacity at forecasting extreme events, as for instance extreme river discharges (see Section 3.5). In Section 3.3, the statistical inference properties of our risk function  $R^{(u)}$  are derived under the framework of multivariate regular variation, which is an alternative way to describe the tail behaviour of multivariate extremes. The following section gives a brief overview of this concept, for more information see, e.g. Resnick (2007).

### 3.1.2 Refinement of asymptotic independence

As illustrated in Chapter 2, when working in the framework of asymptotic dependence, models based on bivariate extreme value theory usually suit well. However, and as discussed in the following, in the case of binary classifiers, it can be complex to find such dependent models. To overcome this problem, a finer risk measure adapted to the case of asymptotic independence is developed. For this purpose, we will rely on the framework of Ramos and Ledford (2009) with joint tail models adapted to the case of asymptotic independence (see Section 1.2.5). This specific framework handles the case of hidden regular variations which can be viewed as a second-order regular variation when asymptotic independence arises (Resnick, 2002). Hereinafter, we give more details on hidden regular variation, starting from the definition of multivariate regular variation, and its connection with the model of Ramos and Ledford (2009).

#### Multivariate regular variation

Formally, a  $d$ -dimensional non-negative random vector  $\mathbf{X}$ , with identical margins (for simplicity), is *multivariate regularly varying* with limit measure  $\nu$  if there exists a function  $b(t) \rightarrow \infty$  as  $t \rightarrow \infty$  and a non-negative Radon measure  $\nu \neq 0$  such that

$$t \mathbb{P} \left( \frac{\mathbf{X}}{b(t)} \in \cdot \right) \xrightarrow{v} \nu, \quad (3.1)$$

on  $[0, \infty]^d \setminus \{\mathbf{0}\}$ . The notation  $\xrightarrow{v}$  stands for *vague convergence*, a definition and its properties can be found in Chapter 3 of Resnick (2007) for example.

It can be shown that condition (3.1) implies that there exists a constant  $\alpha > 0$ , the *tail index*, such that, for all relatively compact sets  $A \subset [0, \infty]^d \setminus \{\mathbf{0}\}$  and  $t > 0$ ,  $\nu(tA) = t^{-\alpha} \nu(A)$ . This property is fundamental in order to obtain the polar decomposition of Equation (3.3). Moreover,  $b$  is a regularly varying function with index  $1/\alpha$  (see Resnick, 2007).

The link between the univariate case as discussed in Section 1.1.3 and the above definition of multivariate regular variation may not be directly evident. Yet, equivalent definitions of univariate and multivariate regular variation in terms of convergence of measures and tail probabilities can be derived (see Theorems 3.6 and 6.1 in Resnick (2007)).

**Theorem 3.1.** *Let  $\mathbf{X}$  be a non-negative  $d$ -dimensional random vector. The following statements are equivalent.*

1.  $\mathbf{X}$ , with distribution function  $F$ , is multivariate regularly varying with measure  $\nu$  and tail index  $\alpha > 0$ .
2. The following convergence holds

$$\lim_{t \rightarrow \infty} \frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = \lim_{t \rightarrow \infty} \frac{\mathbb{P}(\mathbf{X}/t \in [\mathbf{0}, \mathbf{x}]^c)}{\mathbb{P}(\mathbf{X}/t \in [\mathbf{0}, \mathbf{1}]^c)} = \nu([\mathbf{0}, \mathbf{x}]^c), \quad (3.2)$$

for all points  $\mathbf{x} > \mathbf{0}$  which are continuity points of the function  $\mathbf{x} \mapsto \nu([\mathbf{0}, \mathbf{x}]^c)$ , where  $[\mathbf{0}, \mathbf{x}]^c$  denotes the complement set taken with respect to  $[0, \infty]^d \setminus \{\mathbf{0}\}$  (see Figure 3.1 for an illustration in dimension 2).

3. There exists an  $\alpha$ -Pareto random variable  $P$  and, independently of  $P$ , a random vector  $\Theta \in [0, \infty)^d$  on the unit sphere  $\{\mathbf{x} \in \mathbb{R}^d; \|\mathbf{x}\| = 1\}$ , such that

$$\mathbb{P} \left[ \left( \frac{\|\mathbf{X}\|}{t}, \frac{\mathbf{X}}{\|\mathbf{X}\|} \right) \in \cdot \mid \|\mathbf{X}\| > t \right] \xrightarrow{v} \mathbb{P}[(P, \Theta) \in \cdot], \text{ as } t \rightarrow \infty, \quad (3.3)$$

for an arbitrary norm  $\|\cdot\|$  on  $\mathbb{R}^d$ .

Equation (3.2) is the analogue version of univariate regularly varying random variables as defined in (1.7). It implies that the survival function  $1 - F$  is *multivariate regularly varying* (see Resnick (2007), Chapter 6).

In the polar decomposition (3.3),  $\Theta$  is referred to as the *spectral tail vector* and its distribution as the *spectral measure*. A nice interpretation is that the spectral measure, or equivalently  $\nu$ , places mass in directions where large events occur (Meyer, 2020; Meyer and Wintenberger, 2021). This decomposition will be used in Section 3.7 to develop a special case of a linear classifier.

One further comment is that, as for the univariate case, multivariate regular variation and multivariate EVT are two strongly connected notions. Indeed, if  $\mathbf{X}$  is multivariate regularly varying as in (3.1) with limit measure  $\nu$  and tail index  $\alpha > 0$ , then  $\mathbf{X}$  is in the domain of attraction of an extreme value distribution  $G$  with

$$G(\mathbf{x}) = \exp \{-\nu([\mathbf{0}, \mathbf{x}]^c)\},$$

(see, for example Beirlant et al. (2004)). Thus, the limit measure  $\nu$  and the exponent measure  $V$  (defined in (1.10) for the bivariate case) are in the end the same quantity, i.e.  $V(\mathbf{x}) = \nu([\mathbf{0}, \mathbf{x}]^c)$ .

### Hidden regular variation

In view of the foregoing, *hidden regular variation* is defined as a refinement of regular variation on the following subset

$$\mathbb{E}^0 := \{\mathbf{s} \in [0, \infty]^d \setminus \{\mathbf{0}\} : \text{For some } 1 \leq i < j \leq d, \min(s_i, s_j) > 0\}.$$

This set  $\mathbb{E}^0$  corresponds to the points of  $[0, \infty]^d \setminus \{\mathbf{0}\}$  such that at most  $d - 2$  coordinates are 0 (see Figure 3.1).

Then, we say that  $\mathbf{X}$  is hidden regularly varying if in addition to (3.1), there exists a non-decreasing function  $b^0(t) \rightarrow \infty$  such that  $b(t)/b^0(t) \rightarrow \infty$ , as  $t \rightarrow \infty$ , and a Radon measure  $\nu^0 \neq 0$  on  $\mathbb{E}^0$  and such that

$$t \mathbb{P} \left( \frac{\mathbf{X}}{b^0(t)} \in \cdot \right) \xrightarrow{v} \nu^0 \quad (3.4)$$

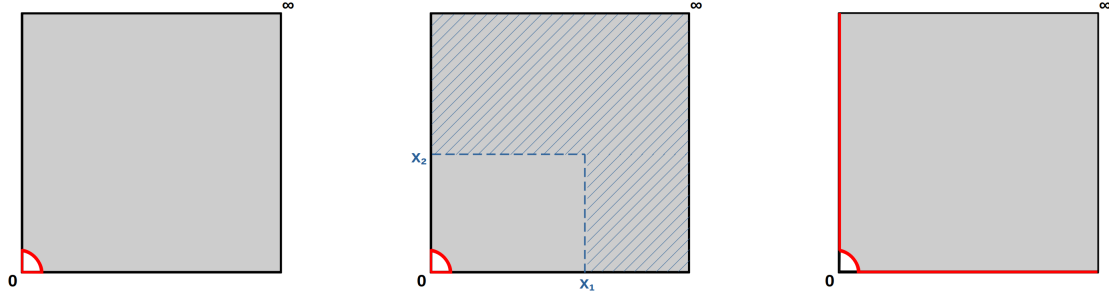


Figure 3.1: Illustration for  $d = 2$  of the different spaces considered. The red lines represent the excluded parts. (left)  $[0, \infty]^2 \setminus \{0, 0\}$  grey shaded area, (middle)  $[0, \mathbf{x}]^c$  with  $\mathbf{x} = (x_1, x_2)$  blue hatched area, (right)  $\mathbb{E}^0 = (0, \infty]^2$  grey shaded area.

on the set  $\mathbb{E}^0$ . Similarly to multivariate regular variation, if (3.4) holds, there exists  $\alpha_0 \geq \alpha$  such that  $b^0$  is regularly varying with index  $1/\alpha_0$  and  $\nu^0(t \cdot) = t^{-\alpha_0} \nu(\cdot)$ , for all  $t > 0$ .

Now, if  $\mathbf{X}$  is hidden regularly varying, then necessarily the components of  $\mathbf{X}$  are asymptotically independent, this is shown in Resnick (2002) and with Chapter 5 of Resnick (1987). For simplicity, we illustrate this for  $d = 2$ , considering  $\mathbf{X} = (X_1, X_2)$ . In this case,  $\mathbb{E}^0 = (0, \infty]^2$ . Assuming (3.1) and (3.4) hold, then, since  $b(t)/b^0(t) \rightarrow \infty$  as  $t \rightarrow \infty$ ,

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left( \frac{X_1}{b(t)} > \delta, \frac{X_2}{b(t)} > \delta \right) = \lim_{t \rightarrow \infty} t \mathbb{P} \left( \frac{\min(X_1, X_2)}{b^0(t)} > \frac{b(t)}{b^0(t)} \delta \right) = 0$$

for any  $\delta > 0$ . That is,  $\nu(\mathbb{E}^0) = 0$  and  $\nu$  concentrates on the axes, i.e. on the lines  $\{(x_1, 0); x_1 > 0\}$  and  $\{(0, x_2); x_2 > 0\}$ .

Therefore, the terminology *hidden* can be understood by the fact that the dependence structure is hidden by the mass put on the axes, hence a normalisation of smaller order, with  $b^0$ , is needed to capture the finer structure that may be present away from the axes.

**Example 3.2.** If  $\mathbf{X} := (X_1, X_2)$  with  $X_1$  and  $X_2$  two i.i.d. Pareto variables with parameter  $1/2$ , that is  $\mathbb{P}(X_i > x) = x^{-1/2}$ . Then, setting  $b^0(t) = 1/t$ ,  $\mathbf{X}$  is hidden regularly varying with  $\nu^0((x_1, \infty] \times (x_2, \infty]) = (x_1 x_2)^{-1/2}$ . This is illustrated in Figure 3.2.

**Example 3.3.** Consider  $\mathbf{X} := B\mathbf{Y} + (1 - B)\mathbf{U}$ , with  $B$  a Bernoulli variable,  $\mathbf{Y}$  having regularly varying marginals with index  $-1$ , and  $\mathbf{U}$  multivariate regularly varying with tail index  $1 < \alpha < 2$ ,  $B$ ,  $\mathbf{Y}$  and  $\mathbf{U}$  being independents. Then  $\mathbf{X}$  is hidden regularly varying (Resnick, 2002).

Hidden regular variation implies multivariate regular variation and asymptotic independence, but the converse is in general not true (see Resnick, 2007).

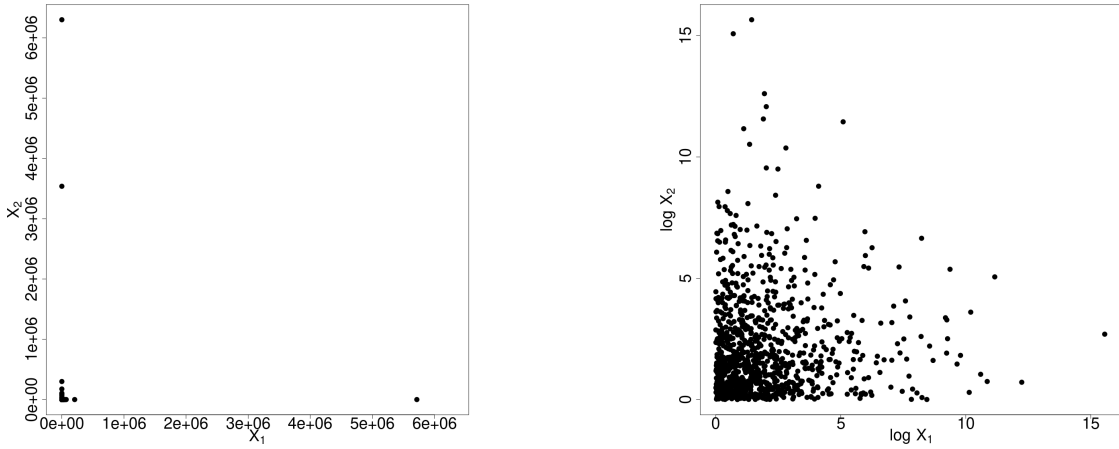


Figure 3.2: Example of hidden regular variation with  $\mathbf{X} := (X_1, X_2)$  as defined in Example 3.2. On the original scale (left) the variables clearly appear as asymptotically independent, but on the log-scale (right) second-order regular variation emerges.

For inference or simulation purposes, [Das and Resnick \(2015\)](#) review different generative models exhibiting hidden regular variation, namely the mixture model developed by [Maulik and Resnick \(2004\)](#) and the additive model of [Weller and Cooley \(2014\)](#). [Das and Resnick \(2015\)](#) also suggest diagnostic tools in order to detect regular variation and hidden regular variation in multivariate data.

### Link with [Ramos and Ledford \(2009\)](#) model

In dimension  $d = 2$ , it appears that the model of [Ramos and Ledford \(2009\)](#) discussed in Section 1.2.5 is a special case of hidden regular variation, where the coefficient of tail dependence  $\eta$  is given by  $\eta = 1/\alpha_0$  (see, e.g. [Heffernan and Resnick \(2005\)](#)).

Indeed, assume that  $\mathbf{X} = (X_1, X_2)$  satisfies (1.25). Consider  $b^0(t) := 2U^{\leftarrow}(t)$ <sup>1</sup> where  $U(t) := \frac{(2t)^{1/\eta}}{\mathcal{L}(2t, 2t)}$ . Then for any  $x_1, x_2 > 0$ ,

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left( \frac{X_1}{b^0(t)} > x_1, \frac{X_2}{b^0(t)} > x_2 \right) = g(x_1, x_2) (x_1 x_2)^{-1/(2\eta)} \sim (x_1 x_2)^{-1/(2\eta)}.$$

In the above, we assumed that the components of  $\mathbf{X}$  were identically distributed. However, with real data analysis, such behaviour is rarely observed. To overcome this issue, [Heffernan and Resnick \(2005\)](#) showed that the use of the rank transform method, resulting in equal marginals, preserved hidden regular

<sup>1</sup> $U^{\leftarrow}(t) := \inf\{s; U(s) \geq t\}$  denotes the left-continuous inverse of  $U$ , and  $U(U^{\leftarrow}(t)) \sim t$ .

variation. Yet, in the following Paper II (Legrand et al., 2021), equal marginal transformation is not desired as explained later. Instead an extension of Ramos and Ledford (2009) is developed, allowing for different tail indexes between the components.

\* \* \*

Below, Paper II (Legrand et al., 2021) is reproduced as submitted. It was written in collaboration with Philippe Naveau and Marco Oesting. In this paper, Juliette L. wrote the Sections 3.4 and 3.5. Juliette L. wrote the R code used and produced the figures in the article. The theoretical proofs were mainly derived by Marco Oesting. All three authors contributed equally to the writing of the remaining text.

The proofs of all lemmas and propositions are placed in Appendix A, as originally done in Paper II.

## 3.2 Introduction

In binary classification, one typically considers data of the form  $(X, Y)^\top$  where  $Y \in \{-1, 1\}$  represents a binary response to the input  $X \in [0, \infty)^d$ . In this paper, we focus on the case that  $Y = Y^{(u)}$  represents the occurrence of an extreme event,  $Y^{(u)} = 1$  indicating that a random quantity  $H$  crosses a level  $u$ , called the threshold, and  $Y^{(u)} = -1$  otherwise, that is

$$Y^{(u)} = \begin{cases} +1, & \text{if } H > u, \\ -1, & \text{otherwise.} \end{cases} \quad (3.5)$$

In the following, for simplicity, we focus on the case that  $H$  is a non-negative random variable such that  $\mathbb{P}(H > u) > 0$  for all  $u > 0$  and its upper end point is infinite.

In extreme value analysis, one is interested in the behaviour of  $Y^{(u)}$  for high levels, that is for  $u \rightarrow \infty$ , and, therefore, also any classifier  $g$  needs to be adapted to the threshold  $u$ . Thus, for every  $u > 0$ , let  $g(\cdot; u)$  be a measurable function from  $\mathbb{R}^d$  to  $\{-1, 1\}$ . In order to evaluate the quality of the classification at a certain level  $u$ , we consider a loss function  $l_u$  that assigns a cost to a classifier  $g(\cdot; u)$  and a realisation  $(x, y^{(u)})$ . Here, it is important to note that, by definition of rare events,  $\mathbb{P}(Y^{(u)} = 1)$  is very small and  $\mathbb{P}(Y^{(u)} = -1)$  is close to one as  $u$  gets large. This imbalance can lead to atypical and/or undesirable comparisons of classifiers. For example, the "always optimistic" classifier that never forecasts an extreme can be defined as  $g(X; u) \equiv -1$ , almost surely. To see how to handle this naive classifier, the classical risk function defined as the expectation of the indicator  $\mathbb{1}\{g(X; u) \neq Y^{(u)}\}$  can be written as

$$\mathbb{P}(g(X; u) \neq Y^{(u)}) = \mathbb{P}(H > u, g(X; u) = -1) + \mathbb{P}(H \leq u, g(X; u) = 1). \quad (3.6)$$

If  $g(X; u) \equiv -1$ , then  $\mathbb{P}(g(X; u) \neq Y^{(u)}) = \mathbb{P}(H > u)$  goes to zero as  $u$  gets large. Hence, the classical risk function  $\mathbb{E}(\mathbb{1}\{g(X; u) \neq Y^{(u)}\})$  will systematically favour the always optimistic classifier for extremes. To avoid this undesirable feature, the loss function has to be modified. One natural idea is to re-scale by  $\mathbb{P}(H > u)$  and introduce the loss function  $\mathbb{1}\{g(X; u) \neq Y^{(u)}\} / \mathbb{P}(H > u)$ . In this case, the risk, i.e., the expected loss  $\mathbb{E}(\mathbb{1}\{g(X; u) \neq Y^{(u)}\} / \mathbb{P}(H > u))$ , goes towards one as  $u$  gets large.

Another trivial but also interesting case is the “crying wolf” forecaster who always issues  $g(X; u) \equiv +1$ , see also the forecaster’s dilemma (e.g. [Lerch et al., 2017](#)). In this case, Equation (3.6) implies that  $\mathbb{P}(g(X; u) \neq Y^{(u)}) = \mathbb{P}(H \leq u)$  and, consequently, the risk  $\mathbb{E}(\mathbb{1}\{g(X; u) \neq Y^{(u)}\} / \mathbb{P}(H > u))$  goes towards infinity as  $u$  gets large. This limiting cost indicates that the “crying wolf” forecaster is much worse than the overly optimistic one. Both of them are unreasonable in practice and there is no reason to strongly favour one over the other one. For this reason, we propose to use a following weighted loss function

$$l_u(g; (x, y)) = \frac{1}{\mathbb{P}(Y^{(u)} = 1 \text{ or } g(X; u) = 1)} \mathbb{1}\{g(x; u) \neq y\}$$

and the associated risk

$$R^{(u)}(g) = \mathbb{E}(l_u(g; Z)) = \frac{\mathbb{P}(g(X; u) \neq Y^{(u)})}{\mathbb{P}(Y^{(u)} = 1 \text{ or } g(X; u) = 1)}. \quad (3.7)$$

By construction, the event  $\{g(X; u) \neq Y^{(u)}\}$  implies that  $\{Y^{(u)} = 1\}$  or  $\{g(X; u) = 1\}$  and therefore, necessarily,  $R^{(u)}(g) \in [0, 1]$ . In particular, the naive classifier  $g(X; u) \equiv -1$  possesses unit risk with  $R^{(u)}(g) = 1$  at each level  $u > 0$ . Similarly, the risk of the “crying wolf” classifier  $g(X; u) \equiv +1$  is then equal to  $R^{(u)}(g) = \mathbb{P}(H \leq u)$  and converges to one as  $u \rightarrow \infty$ .

Hence, the value of one is reached by the two worst cases scenarios in terms of classifiers. This unit value provides a clear benchmark that can be compared to any other classifier satisfying the existence of the limit

$$R(g) = \lim_{u \rightarrow \infty} R^{(u)}(g) \in [0, 1].$$

We call such classifiers *extremal*.

In the weather forecast literature (e.g. [Schaefer, 1990](#)), the definition of  $R^{(u)}(g)$  can be linked to the critical success index, also called the threat score. The critical success index computes the total number of correct event forecasts (hits) divided by the total number of forecasts plus the number of misses (hits + false alarms + misses). Hence,  $1 - R^{(u)}(g)$  can be understood as a critical success index for extremes. In the context of rare events forecasts, [Stephenson et al. \(2008\)](#) highlighted some advantages and drawbacks of various risk functions, including the



critical success index. In particular, these authors linked forecast scoring rules with two dependence indices used in EVT (see [Coles et al., 1999](#))

$$\chi = \lim_{u \rightarrow 1} \mathbb{P}(U > u \mid V > u) \text{ and } \bar{\chi} = \lim_{u \rightarrow 1} \left[ \frac{2 \log(\mathbb{P}(U > u))}{\log(\mathbb{P}(U > u, V > u))} - 1 \right],$$

where the two random variables  $U$  and  $V$  follow the same continuous uniform distribution on  $[0, 1]$ . The choice of uniform marginals can be made whenever the forecast can be assumed to be calibrated, i.e. observations and forecasts follow the same marginal distributions and can be transformed into uniforms. Concerning the extremal dependence strength between  $U$  and  $V$ , if  $\chi > 0$ , then the variables  $U$  and  $V$  are said to be asymptotically dependent and  $\bar{\chi} = 1$ . If  $\chi = 0$ , then the variables  $U$  and  $V$  are said to be asymptotically independent and  $\bar{\chi} < 1$  captures some second order extremal dependence information. [Stephenson et al. \(2008\)](#) advocated the use of  $\bar{\chi}$  and called it the extreme dependency score. Later on, [Ferro and Stephenson \(2011\)](#) proposed two different scores and studied their properties. But the link with the concept of asymptotic independence was not clear and the convergence results of their estimators were not fully developed. In contrast to  $\chi$ , one drawback of  $\bar{\chi}$  is that its formula is not easy to explain to practitioners. In comparison,  $R(g)$  as a type of the critical success index can be interpreted with ease. Hence, it is of interest to extend this definition to the asymptotic independent case.

In the machine learning literature, [Jalalzai et al. \(2018\)](#) also worked on binary classifiers for extremes. But they did not focus on  $R^{(u)}(g)$ . Instead, they studied a different setting where the object of interest was  $\mathbb{P}(g(X) \neq Y \mid \|X\| > u)$  where  $\|X\|$  represents a norm with  $u$  large. Hence, their conditioning event was  $\{\|X\| > u\}$ , while our conditioning depends on  $Y$  with the set  $\{Y^{(u)} = 1 \text{ or } g(X; u) = 1\}$ , see Equation (3.7). So, their interest was centred on the classifier performance when the norm of the explanatory vector  $X$  was large. Our focus is on large values of  $H$  in the production of extreme events of the type  $Y^{(u)} = 1$  when  $H > u$ , see Equation (3.5). [Jalalzai et al. \(2018\)](#) provided various theoretical results based on the main assumption that the conditional distribution of  $X$  given  $Y = \pm 1$  was regularly varying with an angular measure that depends on  $Y = \pm 1$ .

In this study, one part of our results is based on the concept of hidden regular variation (see, e.g. [Ledford and Tawn, 1996](#); [Heffernan and Resnick, 2005](#); [Ferro, 2007](#)). In particular, we take advantage of the model of [Ramos and Ledford \(2009\)](#) to derive the asymptotic properties of our estimators.

Our paper is organised as follows. In Section 3.3, we propose and study a risk function that can handle both the asymptotic dependent and independent cases. Estimators are also constructed and their asymptotic properties derived. Section 3.4 focuses on a simulation example that highlights the difficulty to compare common classifiers in the case of asymptotic independence. In Section 3.5, we revisit

the well studied example of the Danube river application and see how the choice of the metric can change the ranking of classifiers. Note that, besides the proofs of all propositions, the appendix addresses the questions of how to optimise the linear classifier for extremes and how to choose the relevant features, see Section 3.7.

### 3.3 Risk, upper tail equivalence and extremal dependence

The following lemma provides a flexible blueprint to link risk functions with probabilities based on general sets. We will apply it under different setups linked to extreme events.

**Lemma 3.4.** *Let  $A_\varepsilon$  be a sequence of measurable sets of increasing sizes with decreasing  $\varepsilon \in [0, 1]$ , in particular  $A_1 \subseteq A_\varepsilon \subseteq A_0$ . Let  $B_\varepsilon$  be the same type of set sequence such that  $\mathbb{P}(A_1 \cap B_1) > 0$ . The following ratio  $R(A_\varepsilon, B_\varepsilon)$  can be written as*

$$R(A_\varepsilon, B_\varepsilon) := \frac{\mathbb{P}(A_1 \Delta B_1 \mid A_\varepsilon \cap B_\varepsilon)}{\mathbb{P}(A_1 \cup B_1 \mid A_\varepsilon \cap B_\varepsilon)} = 1 - \left[ \frac{1}{\mathbb{P}(B_1 \mid A_1 \cap B_\varepsilon)} + \frac{1}{\mathbb{P}(A_1 \mid A_\varepsilon \cap B_1)} - 1 \right]^{-1}, \quad (3.8)$$

where  $A_1 \Delta B_1$  denotes the difference set. In addition, we have the three following properties for  $R_\varepsilon$ :

1.  $R(A_\varepsilon, B_\varepsilon)$  is non-increasing in  $\varepsilon$  with  $R(A_1, B_1) = 0$ .
2. Let  $A'_\varepsilon$  be another sequence of measurable sets of increasing sizes with decreasing  $\varepsilon$ .  
If  $A_1 = A'_1$  and  $A_\varepsilon \subseteq A'_\varepsilon$  for some  $\varepsilon \in [0, 1]$  then

$$R(A_\varepsilon, B_\varepsilon) \leq R(A'_\varepsilon, B_\varepsilon).$$

3. If for any  $\varepsilon \in [0, 1]$  and  $\varepsilon' \in [0, 1]$ , there exists some positive constants  $a$  and  $b$  and some positive function  $c_{\varepsilon, \varepsilon'}$  such that

$$\mathbb{P}(A_\varepsilon \cap B_{\varepsilon'}) = c_{\varepsilon, \varepsilon'} \mathbb{P}^a(A_\varepsilon) \mathbb{P}^b(B_{\varepsilon'}), \quad (3.9)$$

then

$$R(A_\varepsilon, B_\varepsilon) = 1 - \left[ \frac{c_{\varepsilon, 1}}{c_{1, 1}} (\mathbb{P}(A_1 \mid A_\varepsilon))^{-a} + \frac{c_{1, \varepsilon}}{c_{1, 1}} (\mathbb{P}(B_1 \mid B_\varepsilon))^{-b} - 1 \right]^{-1}. \quad (3.10)$$

We deduce from Equation (3.8) that  $R(A_\varepsilon, B_\varepsilon) = 0$  if and only if  $\mathbb{P}(B_1 \mid A_1 \cap B_\varepsilon) = \mathbb{P}(A_1 \mid A_\varepsilon \cap B_1) = 1$ . The second property of this lemma

indicates that, for a given  $\varepsilon$ , the risk function becomes smaller if the set  $A_\varepsilon$  is as small as possible.

Equation (3.9) can be viewed as a mixing condition that leads to a simple expression of  $R(A_\varepsilon, B_{\varepsilon'})$  based on disjoint events.

We want to use the ratio (3.8) in order to generalise the risk  $R^{(u)}(g)$  defined in (3.7). In the latter definition, adapted to extremes, the set  $\{H > u\}$  was considered and one could set  $B_1 = \{H > u\}$ , for instance, and consequently,  $B_\varepsilon = \{H > \varepsilon u\}$  would be a natural sequence if  $H$  is regularly varying. The choice of  $A_\varepsilon$  is open and it can play an important role<sup>2</sup>. In the next paragraphs, we set  $\varepsilon = 0$  and discuss the choice of  $A_1$  and  $A_0$ . In Section 3.3.1, we will work with  $\varepsilon > 0$ . In that case, we will call the corresponding risk the conditional risk.

In the rest of this paper, we restrict our attention on a particular form of classifiers

$$g(X; u) = \begin{cases} +1, & \text{if } \bar{g}(X) > u, \\ -1, & \text{otherwise,} \end{cases} \quad (3.11)$$

for some function  $\bar{g} : \mathbb{R}^d \rightarrow (0, \infty)$ . The function  $\bar{g}(\cdot)$  does not have to be a norm. It can be understood as any projection/summary of the explanatory variables  $X$  onto the positive real line (see, e.g. [Aghbalou et al., 2021](#), for projection techniques for extremes). This corresponds to the set  $A_1 = \{\bar{g}(X) > u\}$  in Lemma 3.4. A special case of this lemma is to set  $\varepsilon = 0$  and when  $A_0$  and  $B_0$  are equal to the full set, i.e.  $\mathbb{P}(A_0) = \mathbb{P}(B_0) = 1$ , and  $A_1 = \{\bar{g}(X) > u\}$  and  $B_1 = \{H > u\}$ . In this case, Equation (3.8) tells us that  $R^{(u)}(\bar{g})$  defined by (3.7) satisfies

$$R^{(u)}(\bar{g}) = 1 - \left[ \frac{1}{\mathbb{P}(H > u \mid \bar{g}(X) > u)} + \frac{1}{\mathbb{P}(\bar{g}(X) > u \mid H > u)} - 1 \right]^{-1}.$$

This leads to the following expression of  $R^{(u)}(\bar{g})$

$$R^{(u)}(\bar{g}) = 1 - \frac{\mathbb{P}(\bar{g}(X) > u \mid H > u)}{1 - \mathbb{P}(\bar{g}(X) > u \mid H > u) + \mathbb{P}(\bar{g}(X) > u) / \mathbb{P}(H > u)}. \quad (3.12)$$

Within the class defined by Equation (3.11), the effect of the marginal distributions of  $H$  and  $\bar{g}(X)$  on  $R(\bar{g})$ <sup>3</sup> can be explained. When  $\bar{g}(X)$  has a lighter tail than  $H$ , i.e.  $\mathbb{P}(\bar{g}(X) > u) / \mathbb{P}(H > u) \rightarrow 0$ , then we also have  $\mathbb{P}(\bar{g}(X) > u \mid H > u) \rightarrow 0$ , and consequently

$$R^{(u)}(\bar{g}) \rightarrow 1, \text{ as } u \rightarrow \infty.$$

In the case where  $\bar{g}(X)$  possesses a heavier tail than  $H$ , i.e.  $\mathbb{P}(\bar{g}(X) > u) / \mathbb{P}(H > u) \rightarrow \infty$ , we can also show that

$$R^{(u)}(\bar{g}) \rightarrow 1, \text{ as } u \rightarrow \infty.$$

<sup>2</sup>Although we will apply Lemma 3.4 to sets  $A_\varepsilon$  that are rare events, this is not a necessity.

<sup>3</sup>the shortcut notation  $R(\bar{g})$  corresponds to the case where  $g(X; u) = +1$  is built from the event  $\bar{g}(X) > u$  in the associated  $R(g)$ .

This indicates that, whenever  $\bar{g}(X)$  and  $H$  are not tail equivalent, the classifier  $\bar{g}(X)$  cannot outperform naive classifiers with respect to  $R^{(w)}(\bar{g})$  for large  $u$ . Hence, the marginal behaviour of  $\bar{g}(X)$  has a direct impact of its predicting capacity in terms of  $R(\bar{g})$ . This is a widely known fact in forecast verification. In particular, a paradigm promoted by Gneiting and his co-authors (see, e.g. [Gneiting et al., 2007](#)) is to work only with calibrated forecasts. In our case, calibration means that  $\bar{g}(X)$  and  $H$  have the same distributions, and consequently  $\mathbb{P}(\bar{g}(X) > u) / \mathbb{P}(H > u) = 1$  for all  $u$ . In practice, it may be difficult to ensure that this constraint holds for extremes (see, e.g. [Lerch et al., 2017](#); [Taillardat et al., 2019](#)). To illustrate this, suppose that  $\mathbb{P}(X > u) = u^{-1}$  for all  $u \geq 1$ , and the variable  $H$ , independently of the value of  $X$ , is either equal to  $\delta X$  or  $(2 - \delta)X$  with probability .5 and the constant  $\delta \in (0, 1)$ . Then, we have

$$\mathbb{P}(H > u) = \frac{1}{2} \mathbb{P}(\delta X > u) + \frac{1}{2} \mathbb{P}((2 - \delta)X > u) = \begin{cases} \mathbb{P}(X > u), & \text{if } u \geq 2 - \delta, \\ \frac{1}{2} \left(1 + \frac{\delta}{u}\right), & \text{if } \delta < u \leq 2 - \delta, \\ 1, & \text{if } \delta \geq u. \end{cases}$$

Hence,  $X$  and  $H$  are more than tail equivalent, they have identical tail behaviour for large  $u$ . Concerning classifiers, linear ones of the type  $\bar{g}_a(X) = aX$  with  $a > 0$  belong to the class defined by (3.11). They are tail equivalent to  $H$ , and  $R(\bar{g}_a) < 1$ . Although  $X$  and  $H$  have identical tail behaviours, the choice of  $a = 1$  is not optimal with respect to  $R(\bar{g}_a)$ . In particular, one can show that

$$R(\bar{g}_{2-\delta}) = \frac{1 - \delta}{2 - \delta} < R(\bar{g}_1) = \frac{2 - 2\delta}{3 - \delta}.$$

This is not surprising. By construction, the largest values of  $H$  are more likely to be produced by  $(2 - \delta)X$  than  $X$ , especially if  $\delta$  is small. In this context, the following lemma (that is a rewriting of Lemma 3.4) explains that the risk function  $R(\bar{g})$  both depend on the upper tail dependence between  $\bar{g}(X)$  and  $H$  and their marginal behaviours.

**Lemma 3.5.** *If*

$$c(\bar{g}) := \lim_{u \rightarrow \infty} \frac{\mathbb{P}(\bar{g}(X) > u)}{\mathbb{P}(H > u)} \in (0, \infty),$$

*then the limiting risk based on (3.7) has the following expression*

$$R(\bar{g}) = 1 - \frac{\chi^*(\bar{g})}{1 + c(\bar{g}) - \chi^*(\bar{g})},$$

*where  $\chi^*(\bar{g})$  denotes the limit of  $\mathbb{P}(\bar{g}(X) > u \mid H > u)$ . In particular,*

$$R(\bar{g}) = 0 \text{ if and only if } c(\bar{g}) = \chi^*(\bar{g}) = 1.$$

This lemma indicates that  $c(\bar{g}) = 1$ , i.e.  $\bar{g}(X)$  and  $H$  are “asymptotically calibrated”, is a necessary condition to have  $R(\bar{g}) = 0$ . In the above example with  $\delta = 0$ , we have  $c(\bar{g}_2) = 1$  and  $R(\bar{g}_2) = 1/2 \neq 0$ . This means that  $\chi^*(\bar{g}_2) < 1$ . Note that  $c(\bar{g}) = 1$  implies that the constant  $\chi^*(\bar{g})$  simply corresponds to the aforementioned tail dependence coefficient  $\chi(\bar{g})$ , and so, the case  $c(\bar{g}) = 1$  simplifies the expression of the risk

$$R(\bar{g}) = 1 - \frac{\chi(\bar{g})}{2 - \chi(\bar{g})}.$$

This equality tells us that any asymptotically calibrated classifier with  $\chi(\bar{g}) = 0$  always produces a risk function  $R(\bar{g}) = 1$ . Consequently, any asymptotically independent classifier is as uninformative as the two naive classifiers. A reasonable strategy will be to dismiss all asymptotically independent classifiers and find/construct new asymptotically dependent classifiers with positive  $\chi(\bar{g})$ . But, finding asymptotically dependent classifiers can be complex in practice, and in addition, in some not so exotic setups, this is not always possible. To see this, we consider the simple non-linear regression model in the following lemma.

**Lemma 3.6.** *Assume that the variable  $H$  in (3.5) is generated by the non-linear regression model*

$$H \stackrel{d}{=} f(X) + N,$$

where  $\stackrel{d}{=}$  represents the equality in distribution,  $N$  corresponds to a random noise and  $X$  corresponds to the explanatory variables, independent of  $N$ . If  $\mathbb{P}(f(X) > u) = o(\mathbb{P}(N > u))$ , then for any classifier of the type defined by (3.11), we always have

$$R(\bar{g}) = 1.$$

*Hence, no classifier can outperform naive classifiers for this regression model.*

Note that even if the forecaster knows exactly the function  $f(\cdot)$  and has drawn from the explanatory  $X$ , the “ideal” classifier  $\bar{g}(x) = f(x)$  will perform badly, i.e.  $R(f) = 1$ . In addition, the classical trick of using ranks to avoid the problem of marginals discrepancy cannot be applied here. For example, suppose that  $H$  is unit Fréchet distributed, then transforming the marginals of  $X$  into unit Fréchet random variables, say into  $\tilde{X}$ , does not remove the issue as the unobserved noise  $N$  has still have heavier tails than  $\tilde{f}(\tilde{X}) = f(X)$  for some function  $\tilde{f}(\cdot)$ . So, a finer risk measure is needed that is able to distinguish different classifiers in case of asymptotic independence.

### 3.3.1 Conditional risk and hidden regular variation

The choice of the conditioning set in Lemma 3.4 brings new possibilities to construct finer risk measures for extremes than  $R^{(u)}$ . To do so, we opted for the sets:

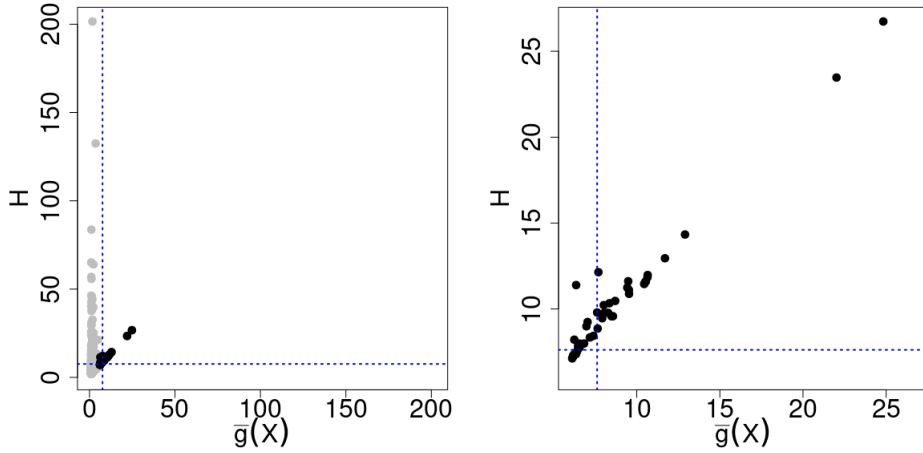


Figure 3.3: Simulated example for Lemma 3.6, for more details see equations (3.16) and (3.17).

$A_1 = \{\bar{g}(X) > u\}$  and  $B_1 = \{H > u\}$ , (when  $\varepsilon = 1$ ) and  $A_0 = \{\bar{g}(X) > 0\}$  and  $B_0 = \{H > 0\}$  (when  $\varepsilon = 0$ ). But, changing the size of the sets  $A_\varepsilon$  and  $B_\varepsilon$  to make them closer to  $A_1$  and  $B_1$ , will increase the conditional probabilities  $\mathbb{P}(B_1 | A_1 \cap B_\varepsilon)$  and  $\mathbb{P}(A_1 | A_\varepsilon \cap B_1)$ . A simple choice when  $\varepsilon \neq 0$  or  $\varepsilon \neq 1$  is to set  $A_\varepsilon = \{\bar{g}(X) > \varepsilon u\}$  and  $B_\varepsilon = \{H > \varepsilon u\}$  with  $\varepsilon > 0$ . This modelling strategy is at the core of hidden regular variation and asymptotic independent models. More precisely, we first need to fix marginal features. We assume that both  $\bar{g}(X)$  and  $H$  possess regularly varying tails with indices  $\alpha_g > 0$  and  $\alpha_H > 0$ , respectively. This means that for any  $\varepsilon \in (0, 1)$ ,

$$\lim_{u \rightarrow \infty} \mathbb{P}(\bar{g}(X) > u | \bar{g}(X) > \varepsilon u) = \varepsilon^{\alpha_g} \text{ and } \lim_{u \rightarrow \infty} \mathbb{P}(H > u | H > \varepsilon u) = \varepsilon^{\alpha_H}.$$

These limits have to be understood with respect to Equation (3.10), i.e. the terms  $\mathbb{P}(A_1 | A_\varepsilon)$  and  $\mathbb{P}(B_1 | B_\varepsilon)$ . To apply (3.10), the mixing condition (3.9) needs to be satisfied. To do so, we opt for an extended version of the framework of [Ramos and Ledford \(2009\)](#), i.e.

$$\mathbb{P}[\bar{g}(X) > u, H > v] = L(u, v)(u^{-\alpha_g} v^{-\alpha_H})^{1/2\eta}, \quad (3.13)$$

where  $\eta \in (0, 1]$  indicates the rate of decay of the joint survival function and  $L(\cdot, \cdot)$  is bivariate slowly varying function, i.e. there exists a limit function  $\ell : (0, \infty) \times (0, \infty) \rightarrow (0, \infty)$  defined as

$$\ell(s, t) = \lim_{u \rightarrow \infty} \frac{L(us, ut)}{L(u, u)}, \quad s, t > 0$$

and satisfying  $\ell(cs, ct) = \ell(s, t)$  for all  $c, s, t > 0$ . The parameter  $\eta$  measures the dependence strength. The case  $\eta = 1$  corresponds to the asymptotic dependence

while  $\eta < 1$  to the asymptotic independence case. In particular, if  $\eta = .5$ , then independence appears in the extremes. If  $.5 < \eta < 1$  ( $0 < \eta < .5$ ) the extremal dependence is said to be positively (negatively) associated.

Now, noticing that (3.13) corresponds to the mixing condition (3.9), we can apply Lemma 3.4, see Appendix A.1 for a proof.

**Proposition 3.7.** *Under the Ramos and Ledford model defined by (3.13), the following risk function*

$$R_\varepsilon(g) := \lim_{u \rightarrow \infty} \frac{\mathbb{P}(g(X; u) \neq Y^{(u)} \mid Y^{(\varepsilon u)} = g(X; \varepsilon u) = 1)}{\mathbb{P}(Y^{(u)} = 1 \text{ or } g(X; u) = 1 \mid Y^{(\varepsilon u)} = g(X; \varepsilon u) = 1)}, \quad (3.14)$$

which will henceforth also be called conditional risk, can be expressed as

$$R_\varepsilon(g) = 1 - \frac{1}{\ell(\varepsilon, 1)\varepsilon^{-\alpha_g/2\eta} + \ell(1, \varepsilon)\varepsilon^{-\alpha_H/2\eta} - 1}, \text{ for any } \varepsilon \in [0, 1).$$

Note that  $\eta \in (0, 1]$  takes a similar role as  $\chi$  in the case of the unconditional risk  $R$ .

For fixed  $\varepsilon \in [0, 1)$ , the risk function  $R_\varepsilon(g)$  decreases with increasing  $\eta$ . So, given all parameters are fixed but  $\eta$ , the forecaster should aim at maximising  $\eta$ . In practice, two forecasters, say  $g_1$  and  $g_2$ , may produce different  $\ell(\cdot, \cdot)$  and  $\alpha_g$ . Consequently, the minimisation of  $R_\varepsilon(g)$  can also depend, besides  $\eta$ , on other parameters.

### 3.3.2 Risk function inference

Concerning the estimation of  $R_\varepsilon(g)$  defined by (3.14), the empirical estimator can be easily computed from the sample  $(X_i, H_i)_{i=1, \dots, n}$ . The following proposition describes the asymptotic property of such an estimator.

**Proposition 3.8.** *Assume that the risk function  $R_\varepsilon(g)$  defined by (3.14) exists for a sequence of  $u_n \rightarrow \infty$  such that  $np_{g, \varepsilon}(u_n) \rightarrow \infty$  with*

$$p_{g, \varepsilon}(u_n) := \mathbb{P}(\max\{g(X; u_n), Y^{(u_n)}\} = 1, H > \varepsilon u_n, g(X; \varepsilon u_n) = 1).$$

If

$$\lim_{n \rightarrow \infty} \sqrt{np_{g, \varepsilon}(u_n)} \left( \frac{\mathbb{P}(g(X; u_n) \neq Y^{(u_n)}, H > \varepsilon u_n, g(X; \varepsilon u_n) = 1)}{p_{g, \varepsilon}(u_n)} - R_\varepsilon(g) \right) = 0,$$

then the empirical estimator based on a sample  $(g(X_i; u), \{H_i > u\})_{i=1, \dots, n}$  and defined by

$$\widehat{R}_{n, \varepsilon}(g) = \frac{\sum_{i=1}^n \mathbb{1}\{g(X_i; u_n) \neq Y_i^{(u_n)}, H_i > \varepsilon u_n, g(X_i; \varepsilon u_n) = 1\}}{\sum_{i=1}^n \mathbb{1}\{\max\{g(X_i; u_n), Y_i^{(u_n)}\} = 1, H_i > \varepsilon u_n, g(X_i; \varepsilon u_n) = 1\}}. \quad (3.15)$$

converges in distribution in the following way

$$\sqrt{np_{g, \varepsilon}(u_n)} \left( \widehat{R}_{n, \varepsilon}(g) - R_\varepsilon(g) \right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, R_\varepsilon(g)(1 - R_\varepsilon(g))).$$

## 3.4 Simulations

### 3.4.1 A simple linear setup

Our main simulated example is based on a simple linear regression model but with the feature that the explanatory variables  $X$  do not have the same tail behaviour and the noise is regularly varying, see Lemma 3.6. More precisely, the multivariate vector  $X$  is defined as follows

$$\begin{cases} X_1 \sim \text{Pareto}(3) \\ X_2 \sim \text{Pareto}(2) \\ X_3 \sim \text{Exp}(1) \\ X_4 \sim \text{Exp}(2) \end{cases}, \quad (3.16)$$

where all  $X_i$  are independent with  $X_1$  and  $X_2$  Pareto distributed with respective tail index 2 and 3, and  $X_3$  and  $X_4$  exponentially distributed with respective scale parameters 1 and 2. The variable of interest  $H$  is simply a linear transform of  $X_1$  tainted by an additive noise

$$H \stackrel{d}{=} X_1 + N, \quad (3.17)$$

where  $N \sim \text{Pareto}(2)$  represents an independent noise with heavier tail than  $X_1$ . So, given a sample  $(\{X_{j,i}\}_{1 \leq j \leq 4}, H_i)_{i=1, \dots, n}$  with  $n = 10000$ , our goal is to compare different classifiers in terms of predicting extreme occurrences, here defined as  $\{H > u\}$  with  $u$  equal to the 97th percentile of  $H$ . In this simulation setup, it is clear from (3.17) that all variables but  $X_1$  are useless to explain  $H$ . In addition, Lemma 3.6 tells us that the relevant information contained in the variable  $X_1$  is hidden by the heavier noise  $N$ , i.e. we are in the case of asymptotic independence. An example of such simulation is given in Figure 3.3. The left panel displays a scatter plot between  $H$  (left axis) and  $\bar{g}(X) = X_1$  (right axis). As expected, no sign of asymptotic dependence can be found in the upper corner. In the right panel, we remove the mass along the axis (grey points) by conditioning on the joint event  $A_\varepsilon \cap B_\varepsilon = \{\bar{g}(X) > \varepsilon u\} \cap \{H > \varepsilon u\}$  with  $\varepsilon = .7$ , see all dark points. The right panel zooms on these black points and highlights a clear dependence between  $H$  and  $X_1$  that was hidden by the heavier noise  $N$  in  $H = X_1 + N$ .

In practice, we do not know the optimal choice for  $\bar{g}(\cdot)$  and we need to introduce different classifiers and compare them.

### 3.4.2 Classifiers descriptions

Table 3.1 below provides the list of classifiers that we compare with our metric (3.15). This list contains some of the most standard classifiers found in the literature (see, e.g. [Hastie et al., 2009](#)): logistic regression (*Logistic*), decision tree (*Tree*), random forest (*RF*) and support vector machine (*SVM*).



Method	Main features
Linear classifier	Simple binary classifier, parameters estimation based on the minimisation of the risk function over the set of contributing variables, theoretical value of $R(g_\theta)$ inferred from spectral decomposition.
Logistic regression	Parametric linear model with a lasso penalty, coefficients of less contributing variables are set to zero.
Decision trees	Easy to interpret, gives relative importance of each variables, learns simple decision rules inferred from the input.
Random forests	Builds multiple decision trees combined by majority vote, better predictive power than decision trees.
Support vector machines	Finds the best hyperplane to separate two overlapping classes, generally performs better than the other classifiers.

Table 3.1: Summary and key features of the different classifiers studied. See for example [Hastie et al. \(2009\)](#) for a comprehensive review of the last four classification methods.

Except for the linear classifier, we apply them with their built-in cost function that is not necessarily fine-tuned to forecast extremes. This is not an issue because our main goal is to compare existing forecasters, and not to create new ones (see, e.g. [Jalalzai et al., 2018](#), for such developments). Still, to fix a baseline in terms of performance, the linear classifier defined as

$$g_\theta(X; u) = \begin{cases} +1, & \theta^\top X > u, \\ -1, & \theta^\top X \leq u, \end{cases} \quad \theta \in [0, \infty)^d.$$

should be optimal for the linear model (3.17), especially if the regression parameters are estimated by minimising our cost function (3.14). In such a context, we expect the linear classifier to be the best. In Appendix 3.7, Proposition 3.9 provides the condition of the consistency of the estimator  $\hat{\theta}_{n, u_n}$  based on minimising  $\hat{R}_n(g_\theta(\cdot; u_n))$  under a regularly varying framework.

The binary outputs from the decision tree classifier are explained in Figure 3.4. The light blue and light green regions represent the set of points that are well predicted by the classifier. On the contrary, wrongly classified points belong to the light yellow and light red regions: either an extreme is predicted when it is not (light red region), or an extreme event is missed (light yellow region). The difference between the left and right panels corresponds to the training set based

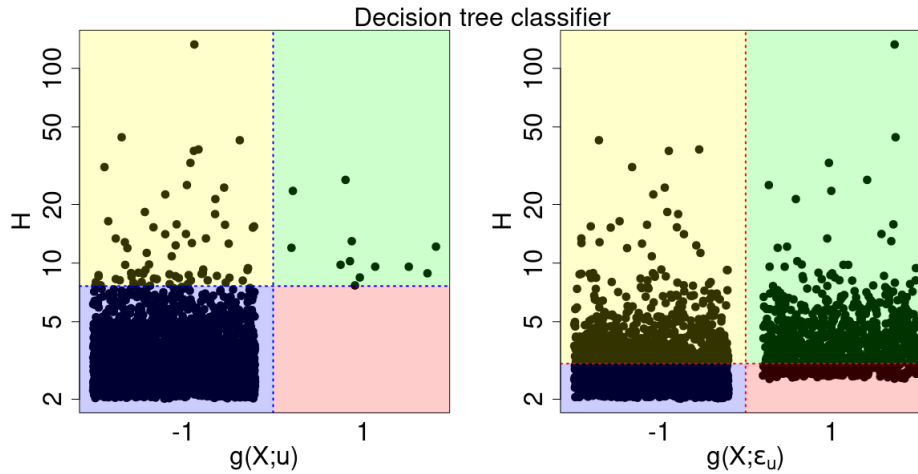


Figure 3.4: Example of predicted binary output for the decision tree classifier versus the true values of  $H$  (on a logarithmic scale). The classifier has been trained twice as explained below. On the left-hand side the classifier is trained with the subset  $\{H > u\}$  and the horizontal blue dashed line represents the threshold value  $u$ . On the right-hand side the classifier is trained with the subset  $\{H > \varepsilon_u\}$  and the horizontal red dashed line represents the threshold value  $\varepsilon_u = 0.4 \times u$ . The results shown are the predicted values made upon the testing set whose sample size is equal to 3000 (30% of the data).

on either  $\{H > u\}$  or on  $\{H > \varepsilon_u\}$ , i.e. mass removed in the latter case, see also (3.15).

### 3.4.3 Implementation and results

We split our simulated data set in two: 70% for a training part, over which we train our different classifiers to get good predictive power; 30% for a testing part, which we use to estimate the risks  $R_0(g)$  and  $R_\varepsilon(g)$ . Note that each algorithm has the same inputs, in particular the same binary sequence describing the events  $\{H > u\}$  with  $u$  set to be equal to the 97th percentile of  $H$ . This cross-validation procedure has been repeated 50 times. The sample used to compute our risk function is based on the bivariate binary vector  $(g(\{X_{j,i}\}_{1 \leq j \leq 4}; u), \{H_i > u\})_{i=1, \dots, n}$  where the output of the classifier  $g$  is binary. In addition, the binary outputs of the classifier  $g$  are obtained under the threshold  $u$  and the threshold  $\varepsilon u$ , so the training part has to be performed twice (once for each threshold). Then, the empirical risk estimator defined by (3.15) can be computed. Figure 3.5 shows the sensitivity of the classifier ranking with respect to the value of  $\varepsilon$ .

As expected from Lemma 3.6, the top-left panel, that corresponds to the case  $\varepsilon = 0$ , clearly indicates that our five classifiers cannot outperform naive classifiers as all classifiers have a risk near to one, the worst possible value. To start dis-

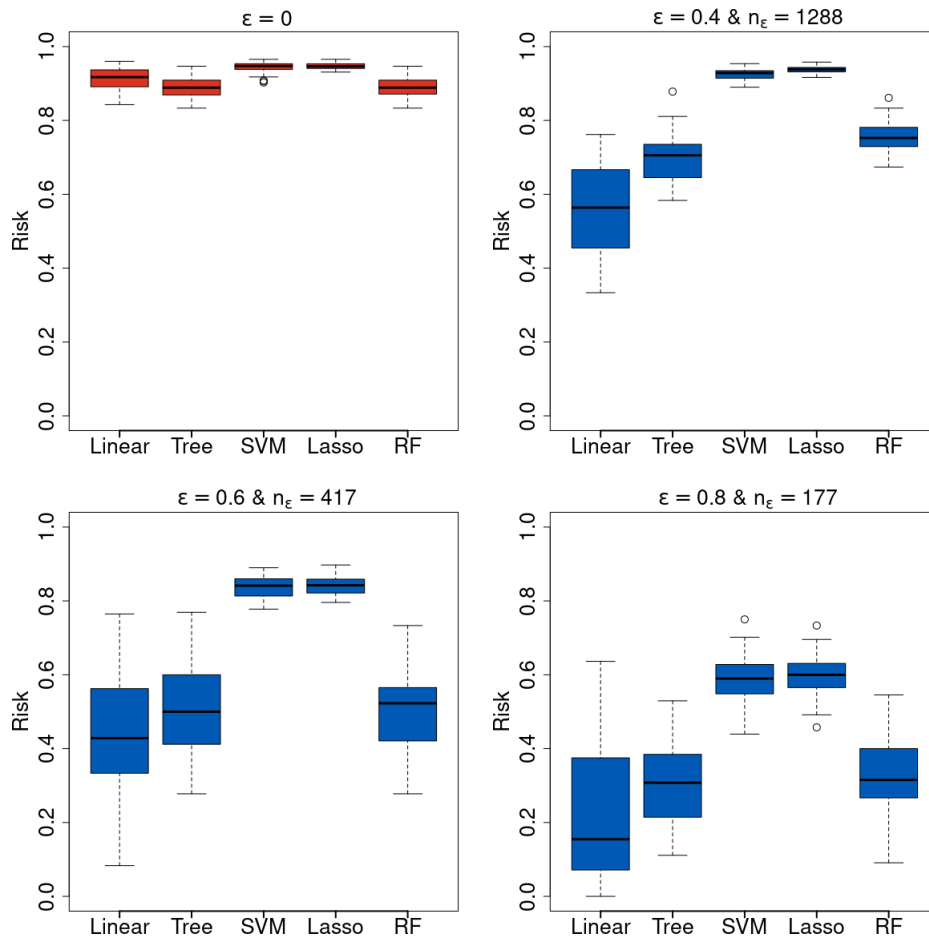


Figure 3.5: Estimation of  $R_\epsilon(g)$  for different classifiers (cross-validation with 50 repetitions). In red (top-left) are the estimates when  $\epsilon = 0$  and in blue for different values of  $\epsilon > 0$  ( $\epsilon \in \{0.4, 0.6, 0.8\}$ ). At the top of each plot, the value of  $\epsilon$  and the number of points such that  $H > \epsilon_u$  from the testing set are given.

criminating classifiers, we need to remove the masses along the axes by setting a value for  $\varepsilon$ . As we increase  $\varepsilon$ , the size of sets needed to compute (3.15) becomes smaller, see the values of  $n_\varepsilon$  in the legend of each panel. Hence, the blue box plots becomes wider as  $\varepsilon$  increases: a classical bias-variance trade-off. As we know from (3.17) that the true generative process is linear,  $\varepsilon = 0.4$  appears as a reasonable value to balance the bias-variance trade-off. More importantly, the overall ranking is not sensitive to the values of  $\varepsilon > 0$ . In all cases, our linear classifier tailored to handle linear asymptotic independence cases outperforms all the other classifiers. Among the other classifiers, decision tree appears to be the best, but it is still far from the optimal linear solution. Other simulations concerning the regular variation case are available upon request.

### 3.5 Danube river discharges

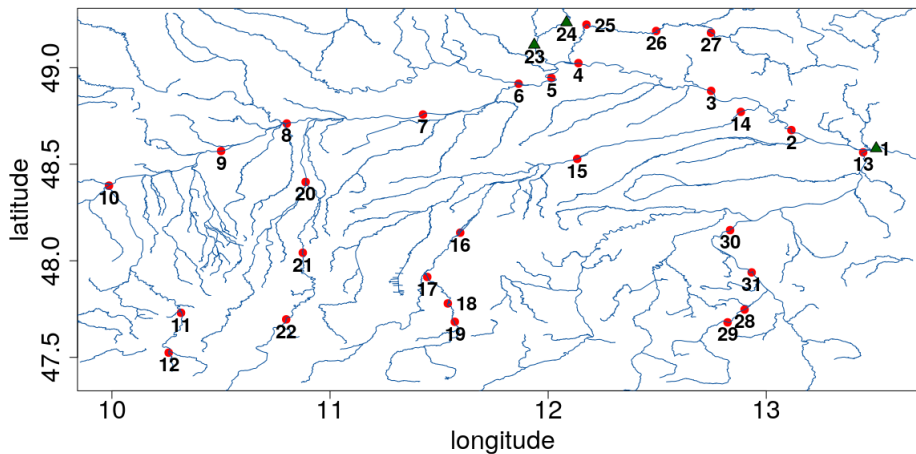


Figure 3.6: River map of the upper Danube basin, showing sites of the 31 gauging stations along the Danube and its tributaries. Water flows toward gauging station 1. The stations represented by a green triangle shaped dot are the three stations of interest as described in Section 3.5.

We now apply our assessment approach to summer daily river discharges (measured in  $m^3/s$ ) at 31 stations spread over the upper Danube basin, see Figure 3.6, and recorded over the time period 1960-2010 in June, July and August. These observations have been studied by the EVT community (see, e.g. [Asadi et al., 2015](#); [Mhalla et al., 2020](#); [Gnecco et al., 2021](#)). This dataset was made available by the Bavarian Environmental Agency (<http://www.gkd.bayern.de>). To remove temporal clustering in extreme river discharges, [Mhalla et al. \(2020\)](#) in their Section 5 implemented a declustering step. Each station then contains  $n = 428$  observations

that we will consider temporally independent. In order to reduce the large discrepancy in terms of discharges magnitude among stations, we force the starting value of all 31 time series to equal zero by subtracting to each station its minimum. Then, we re-normalise each time series by its range (i.e. the difference between the maximum and the minimum of each time series). These post processing treatments are useful to display and interpret the data at hand and do not impact the classifiers performance.

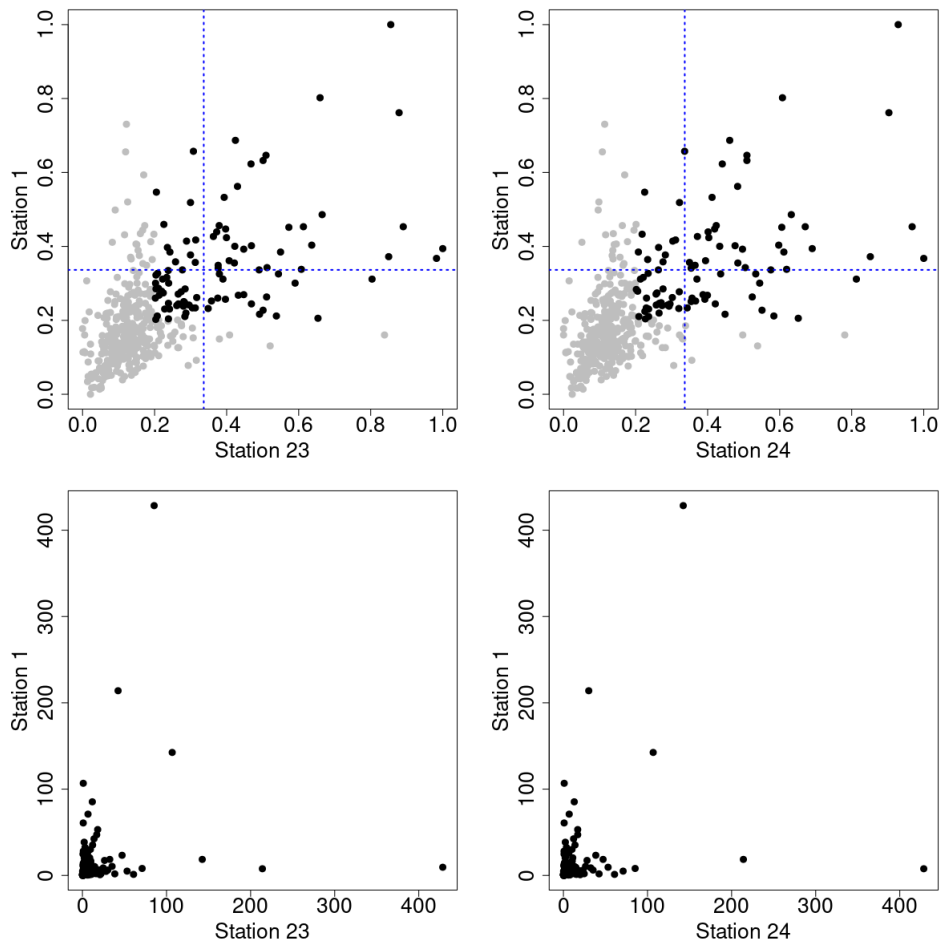


Figure 3.7: Summer daily measurements of river discharges from station 1 (y-axis) against station 23 (left x-axis) and station 24 (right x-axis) of Figure 3.6. The blue dotted lines are the threshold value  $u$  (the 85th percentile of station 1). The black dots on the graphs in the top row are the values such that  $\min(X_j, X_1) > \epsilon_u$  where  $\epsilon_u = 0.6u$ . The graphs in the bottom row corresponds to the same data plots but on unit Fréchet scale highlighting potential asymptotic independence in the data.

Although all 31 station recordings are available, we can artificially remove one station and try to predict its values from a given subset of other weather stations.

In this section, we remove station 1 (downstream) and try to predict its values from only stations 23 and 24, which are indirect tributaries to the main river flow. So, this setup is complex<sup>4</sup> for two reasons. First, station 1, as a downstream point that accumulates all discharges, has a much heavier tail than the two tributaries. Second, it is difficult to determine if we are in the asymptotic dependent or independent case, see Figure 3.7 that displays the scatter plot between the hidden station (y-axis with station 1) and the two tributaries (x-axis, stations 23 and 24). In this graph, the threshold  $u$  is taken to be equal to the 85th quantile of  $X_1$  and we choose  $\varepsilon = 0.6$ . Figure 3.8 summarises our findings. Removing the mass on the axes when thresholding by  $\varepsilon_u$  implies that only 190 points remain from the original length of 428 data points per station. This can explain why, looking at Figure 3.8, the uncertainty in the risk estimate increases when considering  $R_\varepsilon(\cdot)$  (blue boxes) instead of  $R_0(\cdot)$  (red boxes).

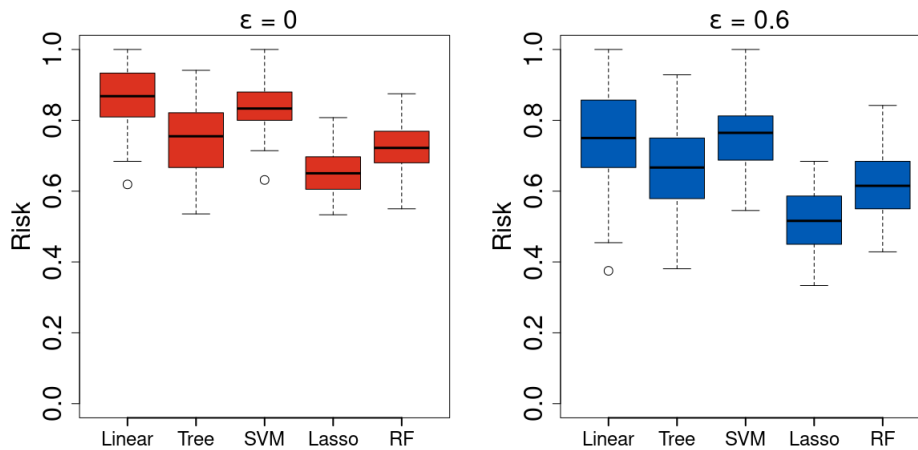


Figure 3.8: Estimation of  $R_\varepsilon(g)$  for five different classifiers (cross-validation with 50 repetitions, 70% train and 30% test), threshold is the 0.85 quantile of  $H$ . In red (left) are the estimates when  $\varepsilon = 0$  and in blue (right) for  $\varepsilon = 0.6$ . The length of the testing set is equal to 129, this leads to around 60 points such that  $H > \varepsilon_u$  and nearly 20 points such that  $H > u$ .

Unlike the simulation example in Section 3.4.1, it is not clear to assess whether our river discharges analysis of our three selected weather stations belongs to the framework of asymptotic independence or not. Still, it is reassuring that the ranking of the classifiers in Figure 3.8 appears to be insensitive to the values of  $R_0$  or  $R_\varepsilon$ . i.e. whether the data are asymptotically dependent or not. This hints that, among all the classifiers, the logistic regression with lasso penalty seems to

<sup>4</sup>Section 3.7.3 treats a simpler case where station 1 is predicted from the whole set of remaining stations. In this case, strong dependencies among station 1 and other stations can be observed. So, the main issue is to select these stations, a problem discussed in Section 3.7.

perform better than the four other classification methods. This ranking of classifiers is specific to this particular example. No general conclusions about lasso techniques for extremes should be drawn.

Besides this river example, we advocate practitioners to compute risk functions that can both handle the asymptotic dependent and independence cases. This also complements the recent tools used to discriminate between the two cases (see, e.g. [Ahmed et al., 2022](#)). In addition, the linear classifier could provide a simple benchmark with well understood properties with respect to  $R_0$ , see Proposition 3.9.

## 3.6 Supplementary Materials

A R package is available on GitHub that implements the empirical estimation of the risk function developed in this paper (<https://github.com/jlegrand35/ExtremesBinaryClassifier>) and can be used either to reproduce the results of the conducted classifier comparisons or to perform new comparisons using other binary classifiers. The data used in the application are available in the R package `graphicalExtremes` ([Engelke et al., 2019](#)).

## Acknowledgement

Part of this work was supported by the DAMOCLES-COST-ACTION on compound events, the French national program (FRAISE-LEFE/INSU and 80 PRIME CNRS-INSU), and the European H2020 XAIDA (Grant agreement ID: 101003469) and funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 - 390740016. The authors also acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project), the ANR-Melody and the Stuttgart Center for Simulation Science (SimTech).

\* \* \*

## 3.7 Supplements to Paper II

This section was originally part of the appendix of [Legrand et al. \(2021\)](#). We reproduce it below and give further details.

In the following, the focus is on linear classifiers. We first give its definition and basic properties, then provide a way to find such optimal classifiers with Proposition 3.9. It turns out that it is sufficient to consider only the components contributing to the extremes. This leads to the notion of sparsity and from this we

build a simple tool to detect relevant features. In Paper II, this methodology was then applied to the Danube river network, as reproduced below.

Hereinafter, we return to the bold notation for the vectors of  $\mathbb{R}^d$ , i.e.  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ .

### 3.7.1 Linear Classifiers: Definition, Basic Properties and Inference

In this section, we consider a specific type of classifiers which in Paper II is referred to as linear classifiers, i.e. classifiers of the form

$$g_{\boldsymbol{\theta}}(\mathbf{X}; u) = \begin{cases} +1, & \boldsymbol{\theta}^\top \mathbf{X} > u, \\ -1, & \boldsymbol{\theta}^\top \mathbf{X} \leq u, \end{cases} \quad \boldsymbol{\theta} \in [0, \infty)^d.$$

This is a simple binary classifier that is based on the idea that a large value along an appropriate linear combination could produce extremes in the hidden variable  $H$  of interest.

To obtain an optimal linear classifier of  $g_{\boldsymbol{\theta}}(\mathbf{X}; u)$ , i.e. some weight vector  $\boldsymbol{\theta}^*$  such that the classification risk  $R(g_{\boldsymbol{\theta}^*})$  gets minimal, we need to impose some joint extremal dependence structure on  $\mathbf{X}$  and  $H$  from (3.6).

Even though some of the results can also be obtained in a similar manner in a more general framework for the conditional risk  $R_\varepsilon$  (see Section 3.7.4), henceforth, we will focus on the asymptotically dependent case where we might find some optimal classifier with unconditional risk  $R(g_{\boldsymbol{\theta}^*}) < 1$ . As discussed before, in this case, at least one component of  $\mathbf{X}$  needs to have a similar tail behaviour as  $H$ . A natural assumption is therefore that  $(\mathbf{X}, H)$  is jointly regularly varying on  $[0, \infty)^{d+1}$  with index  $\alpha > 0$ , i.e. there exists an  $\alpha$ -Pareto random variable  $P$  and, independently of  $P$ , a random vector  $(\boldsymbol{\Gamma}, \Omega) \in [0, \infty)^d \times [0, \infty)$ , the so-called *spectral tail vector*, on the unit sphere  $\{x \in [0, \infty)^{d+1} : \|x\|_\infty = 1\}$  such that

$$\mathcal{L} \left( \left( \frac{\|(\mathbf{X}, H)\|_\infty}{u}, \frac{(\mathbf{X}, H)}{\|(\mathbf{X}, H)\|_\infty} \right) \middle| \|(\mathbf{X}, H)\|_\infty > u \right) \xrightarrow{v} \mathcal{L}(P, (\boldsymbol{\Gamma}, \Omega)), \text{ as } u \rightarrow \infty.$$

Note that assuming joint regular variation, implies in particular that  $H$  and all the relevant components of  $\mathbf{X}$  are heavy-tailed. For example, this can be the case for rainfall data (e.g. [Le Gall et al., 2022](#)) or river discharges (see Table 3.2). But other environmental data sometimes do not show such behaviour, as in the case of wind gusts ([Friederichs et al., 2009](#)), temperatures ([Toulemonde et al., 2015](#)) or significant wave heights ([Legrand et al., 2022](#)), which are more light-tailed or even upper-bounded. Some precautions should therefore be taken before considering a dataset, by applying for instance an univariate extreme value analysis (e.g. [Embrechts et al., 2013](#), Chapter 6).



Here, we additionally assume that  $\mathbb{P}(\|\Gamma\|_\infty > 0) > 0$  and  $\mathbb{P}(\Omega > 0) > 0$ . In this setup,  $\|\mathbf{X}\|_\infty$  and  $H$  are tail equivalent in the sense that

$$\begin{aligned} \lim_{u \rightarrow \infty} \frac{\mathbb{P}(\|\mathbf{X}\|_\infty > u)}{\mathbb{P}(H > u)} &= \lim_{u \rightarrow \infty} \frac{\mathbb{P}(\|\mathbf{X}\|_\infty > u \mid \|(\mathbf{X}, H)\|_\infty > u)}{\mathbb{P}(H > u \mid \|(\mathbf{X}, H)\|_\infty > u)} \\ &= \frac{\mathbb{P}(P \cdot \|\Gamma\|_\infty > 1)}{\mathbb{P}(P \cdot \Omega > 1)} = \frac{\mathbb{E}(\|\Gamma\|_\infty^\alpha)}{\mathbb{E}(\Omega^\alpha)} \in (0, \infty), \end{aligned}$$

This is the minimal requirement on the link between the covariates  $\mathbf{X}$  and the unobserved extremes of  $H$  essentially saying that at least one component of  $\mathbf{X}$  is tail-equivalent to  $H$ . It is important to highlight that we do not exclude the case that  $\Gamma_i = 0$  a.s. for some  $i \in \{1, \dots, d\}$  which means that  $X_i$  possesses a lighter tail than  $H$ . This property can be read off from the quantity

$$c_i = \lim_{u \rightarrow \infty} \frac{\mathbb{P}(X_i > u)}{\mathbb{P}(H > u)} = \frac{\mathbb{E}(\Gamma_i^\alpha)}{\mathbb{E}(\Omega^\alpha)}. \quad (3.18)$$

Thus,  $\Gamma_i = 0$  a.s. if and only if  $c_i = 0$ . By including this case, we therefore admit that most of the components of  $\mathbf{X}$  may not contribute to the extremes of the vector  $H$ . This feature is essential when the question of sparsity will be addressed.

Under these conditions, we obtain that, for all  $\boldsymbol{\theta} \in [0, \infty)^d$ , the classifier  $g_\boldsymbol{\theta}$  is an extremal classifier as

$$\begin{aligned} R(g_\boldsymbol{\theta}) &= \lim_{u \rightarrow \infty} \mathbb{P}(H > u \text{ or } \boldsymbol{\theta}^\top \mathbf{X} > u)^{-1} (\mathbb{P}[\max\{\boldsymbol{\theta}^\top \mathbf{X}, H\} > u] - \mathbb{P}[\min\{\boldsymbol{\theta}^\top \mathbf{X}, H\} > u]) \\ &= \frac{\mathbb{E}(\max\{\boldsymbol{\theta}^\top \Gamma, \Omega\}^\alpha) - \mathbb{E}(\min\{\boldsymbol{\theta}^\top \Gamma, \Omega\}^\alpha)}{\mathbb{E}(\max\{\boldsymbol{\theta}^\top \Gamma, \Omega\}^\alpha)} = 1 - \frac{\mathbb{E}(\min\{\boldsymbol{\theta}^\top \Gamma, \Omega\}^\alpha)}{\mathbb{E}(\max\{\boldsymbol{\theta}^\top \Gamma, \Omega\}^\alpha)} \in [0, 1]. \end{aligned} \quad (3.19)$$

Equation (3.19) implies that the function  $\boldsymbol{\theta} \mapsto R(g_\boldsymbol{\theta})$  is well-defined and continuous on  $[0, \infty)^d$ . Its value does not depend on those components  $\theta_i$  for which  $\Gamma_i = 0$  a.s., which is equivalent to  $c_i = 0$  as discussed above. Thus, in the following, we will consider this function only on the parameter set

$$C = \{\boldsymbol{\theta} \in [0, \infty)^d : \theta_i = 0 \text{ for all } i \text{ s.t. } c_i = 0\},$$

containing all the relevant information – here, note that, in practice, identifying the components  $i \in \{1, \dots, d\}$  such that  $c_i = 0$  a.s., from a given data set is a necessary step for the correct specification of the set  $C$ .

From the consideration in the introduction, it can be easily seen that  $R(g_0) = 1$  – the case  $\boldsymbol{\theta} = \mathbf{0}$  corresponds to the trivial always optimistic classifier. Furthermore, denoting the set of indices  $j$  with  $c_j > 0$  by  $J$ , we can see that

$$R(g_\boldsymbol{\theta}) \geq 1 - \frac{\mathbb{E}(\Omega^\alpha)}{\|\boldsymbol{\theta}\|_\infty^\alpha \cdot \min_{j \in J} \mathbb{E}(\Gamma_j^\alpha)} \rightarrow 1 \quad (3.20)$$

as  $\|\boldsymbol{\theta}\|_\infty \rightarrow \infty$ . By the continuity of  $\boldsymbol{\theta} \mapsto R(g_\boldsymbol{\theta})$ , we obtain that the function attains a global minimum on the domain  $C$ .

**Proposition 3.9.** *Additionally to the assumptions above on the joint distribution of  $(\mathbf{X}, H)$  with  $\alpha > 1$ , assume that there exists a function  $a(u)$  with  $a(u) \rightarrow 0$  as  $u \rightarrow \infty$  such that*

$$\mathbb{P}(u^{-1}\mathbf{X} \in A \mid \|(\mathbf{X}, H)\|_\infty > u) \leq (1 + a(u)) \mathbb{P}(P\mathbf{\Gamma} \in A) \quad (3.21)$$

for all  $A \subset [0, \infty)$ . Furthermore, let  $u_n \rightarrow \infty$  and  $n\mathbb{P}(H > u_n) \rightarrow \infty$  such that, for every compact subset  $K \subset C$ ,

$$\sup_{\boldsymbol{\theta} \in K} \sqrt{n\mathbb{P}(H > u_n)} \left| \frac{\mathbb{P}(g_{\boldsymbol{\theta}}(\mathbf{X}; u_n) \neq Y^{u_n})}{\mathbb{P}(H > u_n)} - \frac{\mathbb{E}(\max\{\boldsymbol{\theta}^\top \mathbf{\Gamma}, \Omega\}^\alpha) - \mathbb{E}(\min\{\boldsymbol{\theta}^\top \mathbf{\Gamma}, \Omega\}^\alpha)}{\mathbb{E}(\Omega^\alpha)} \right| = 0 \quad (3.22)$$

and

$$\sup_{\boldsymbol{\theta} \in K} \sqrt{n\mathbb{P}(H > u_n)} \left| \frac{\mathbb{P}(\max\{g_{\boldsymbol{\theta}}(\mathbf{X}; u_n), Y^{u_n}\} = 1)}{\mathbb{P}(H > u_n)} - \frac{\mathbb{E}(\max\{\boldsymbol{\theta}^\top \mathbf{\Gamma}, \Omega\}^\alpha)}{\mathbb{E}(\Omega^\alpha)} \right| = 0. \quad (3.23)$$

If the function  $\boldsymbol{\theta} \mapsto R(g_{\boldsymbol{\theta}})$  has a unique minimiser  $\boldsymbol{\theta}^*$  in  $C$ , then the estimator

$$\hat{\boldsymbol{\theta}}_{n, u_n} = \operatorname{argmin}_{\boldsymbol{\theta} \in C} \hat{R}_n(g_{\boldsymbol{\theta}}(\cdot; u_n)).$$

is consistent, i.e.  $\hat{\boldsymbol{\theta}}_{n, u_n} \rightarrow_p \boldsymbol{\theta}^*$ .

Given the set  $C$ , this result provides a strategy to find the optimal  $\boldsymbol{\theta}$ , i.e., the best linear classifier. Determining the set  $C$  requires the identification of the relevant features, i.e. the index set  $J$  such that  $c_j > 0$  if and only if  $j \in J$ . This is discussed in more detail in the following subsection.

### 3.7.2 Feature Selection

The notion of sparsity quickly comes into play when doing classification. This is all the more true when one is only interested in the extremes. Among the whole data set, only a small proportion will truly contribute to the extremal behaviour of the variable of interest. Here, we develop a method to identify the informative signals in terms of extremes among a large data set, assuming that  $(\mathbf{X}, H)$  is jointly regularly varying. For a comprehensive review of existing methods on sparsity and multivariate extremes we highly recommend the work of [Engelke and Ivanovs \(2021\)](#).

As we have seen above, for linear classifiers, all the relevant features  $X_i$  necessarily satisfy  $c_i > 0$ . Thus, feature selection can be based on estimation of the  $c_i$  which can be done according to the following proposition.

**Proposition 3.10.** *Assume that*

$$c_i = \lim_{u \rightarrow \infty} \frac{\mathbb{P}(X_i > u)}{\mathbb{P}(H > u)}$$

*exists. If  $u_n \rightarrow \infty$  and  $n \mathbb{P}(H > u_n) \rightarrow \infty$ , then*

$$\frac{\sum_{j=1}^n \mathbf{1}\{X_{j,i} > u_n\}}{\sum_{j=1}^n \mathbf{1}\{H_j > u_n\}} \xrightarrow{p} c_i.$$

*If, additionally,*

$$\sqrt{n \mathbb{P}(H > u_n)} \left( \frac{\mathbb{P}(X_i > u_n)}{\mathbb{P}(H > u_n)} - c_i \right) \rightarrow 0$$

*and*

$$\chi_i^* = \lim_{u \rightarrow \infty} \mathbb{P}(X_i > u \mid H > u) \in [0, 1]$$

*exists, then, we have*

$$\sqrt{n \mathbb{P}(H > u_n)} \left( \frac{\sum_{j=1}^n \mathbf{1}\{X_{j,i} > u_n\}}{\sum_{j=1}^n \mathbf{1}\{H_j > u_n\}} - c_i \right) \rightarrow \mathcal{N}(0, c_i \cdot [1 - 2\chi_i^* + c_i]).$$

The above Proposition 3.10 helps us to discriminate between informative coordinates and non-informative ones in terms of extremes.

### 3.7.3 River network

We return to the Danube river network. Contrary to what was done in Section 3.5, here the data are pre-processed differently: we only subtract the minimum to each station, no re-normalisation is performed. Thus some stations have a heavier tail than others.

Before applying the methodology outlined above, an a priori analysis is performed to check if the data are well regularly varying. From Table 3.2, we can not reject the hypothesis that the river discharges at the stations considered are heavy-tailed. Note that we show the results only for Station 1 (the station of interest in this study), Stations 2, 13 and 30 (the stations that appear to contribute the most to the extremes of Station 1, see below), and Stations 23 and 24 (the stations considered in Section 3.5).

This section deals with a simpler case than the application in Section 3.5. Here an application could be the following: we want to know which stations should continue to be maintained to prevent extreme floods and maybe some stations are not necessary.

As before, the goal is to predict the extreme events at Station 1, denoted  $X_1$ , where an extreme event is defined as an event exceeding the 85th quantile of

Station id	1	2	13	30	23	24
Shape	0.19	0.13	0.07	0.03	0.04	0.04

Table 3.2: Estimates of the shape parameter when fitting a GPD. The threshold was selected using mean residual life plot and parameter stability plots (see, e.g. [Coles \(2001\)](#)). Estimation is performed using the R package `ismev` ([Heffernan et al., 2018](#)).

$X_1$ . Unlike the study of Section 3.5, we assume that the whole set of remaining stations is available. In this case, strong dependencies among Station 1 and other stations can be observed. Therefore, the main issue is to identify and select these stations following the procedure presented above.

The stations that may not contribute to the extremes of  $X_1$  are identified through the estimation of the coefficients  $c_i$ . The estimation of the set  $C$  on all the data is presented in Table 3.3. Among the 30 stations, only three stations are relevant: Stations 2, 13 and 30. Looking at Figure 3.6, these stations correspond to the stations closest to  $X_1$ . Figure 3.9 shows the scatter plots between these stations and Station 1, reflecting strong dependencies between the variables.

Station id	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$c_i$	0.06	0	0	0	0	0	0	0	0	0	0	0.30	0	0	0	0
Station id	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
$c_i$	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0		

Table 3.3: Empirical estimates of  $\hat{c}_i$  (as defined in Proposition 3.10) for each station. The values different from zero are highlighted in red.

Once the contributing variables have been identified, we compare the performance of several classifiers, on the one hand keeping all the data and on the other hand keeping only the informative stations. Since there is a strong dependence between the data, we assume that it is sufficient here to look at the risk  $R_0$ . Comparison results are shown in Figure 3.10. The comparison is performed among the same classes of classifier as in Section 3.5, see Table 3.1 for a summary.

By definition of the linear classifier, the estimation is already done by keeping only the informative variables, which is why the estimates are identical for this specific classifier. As for the other classifiers, we see some improvements when keeping only the informative variables: the risk estimates are slightly smaller. This means that even if we remove a lot of information by going from 30 explanatory variables to 3, these 3 remaining stations contain all the information in terms of extremes of Station 1. However, caution must be exercised here, as no theoretical

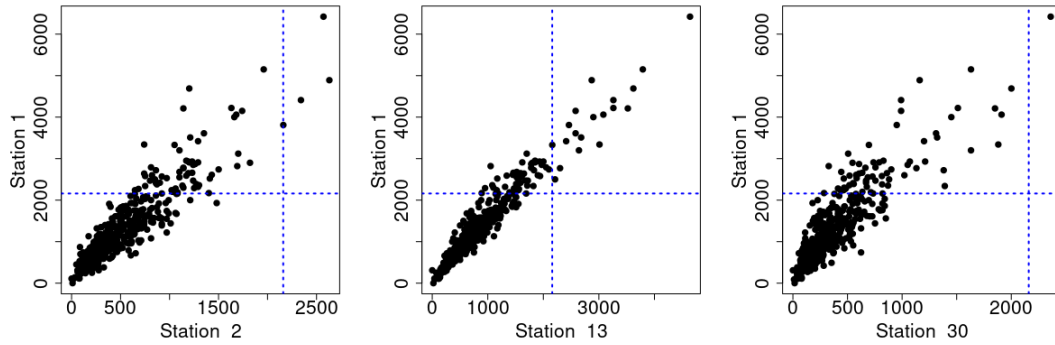


Figure 3.9: Correlation plot between  $X_1$  and the stations considered to contribute to the extremes of  $X_1$  according to Table 3.3 ( $i$ e for which  $c_i \neq 0$ ). The blue dotted lines represent the threshold  $u$  defined by the 85th percentile of  $X_1$ .

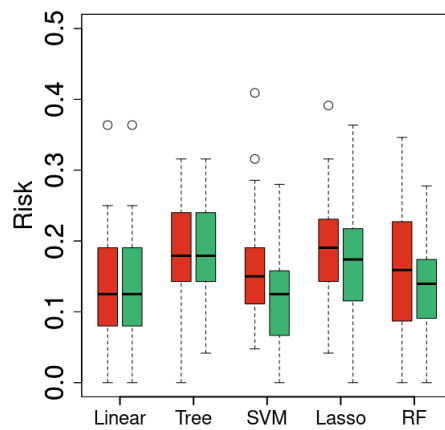


Figure 3.10: Estimation of  $R_0(g)$  for different classifiers  $g$  (cross-validation - 70% train, 30% test - with 50 repetitions). The red distributions come from the estimation with all the stations, and the green distributions represent the estimations with only the variables having  $\hat{c}_i \neq 0$ .

results have been given for the other classifiers.

### Predictive power of the linear classifier

We provide some additional remarks on the special case of the linear classifier. Of the foregoing, only Stations 2, 13 and 30 are relevant for the estimation of the optimal  $\theta$  (see Proposition 3.9). The left-hand side of Figure 3.11 depicts the distribution estimates of  $\hat{\theta}_2, \hat{\theta}_{13}$  and  $\hat{\theta}_{30}$ . The estimation is performed on 70% (training set) of the data drawn 50 times at random. From Figure 3.11,  $X_{30}$  has less effect than  $X_2$  and  $X_{13}$ . Looking at the river map (Fig. 3.6), Station 30 is the furthest station from Station 1 out of the three.

Then one of the 50 estimates is retained (at random, but sensitivity test showed that it does not impact our results). Let's denote it  $\hat{\theta} = (\hat{\theta}_2, \hat{\theta}_{13}, \hat{\theta}_{30})$ . And the predicted output of our linear classifier is computed through  $\hat{\theta}^T \mathbf{X}_{\text{test}}$ . The vector  $\mathbf{X}_{\text{test}}$  corresponds to the river discharges at Station 2, 13 and 30 from the remaining 30% data (that has not been used for the estimation of  $\hat{\theta}$ ). In the right-hand side of Figure 3.11, the predicted output  $\hat{\theta}^T \mathbf{X}_{\text{test}}$  is plotted against the observed river discharges at Station 1. From this, we find that the linear classifier predicts the extremes quite well (upper right quadrant), but also, it predicts well the river discharges that are not extreme (i.e. outside the extreme region). This is most likely due to the specificity of our data, which are highly correlated.

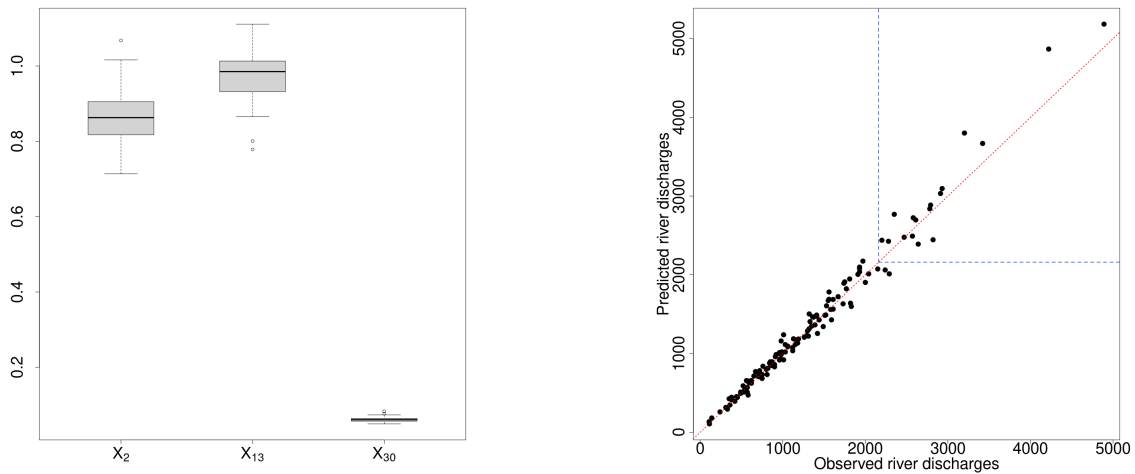


Figure 3.11: (left) Estimates of the optimal  $\theta_i$  of the three contributing variables  $X_2, X_{13}$  and  $X_{30}$  (see Table 3.3), with 50 replications of the optimisation procedure. (right) Predicted versus observed river discharges at Station 1. The predicted values are defined as  $\hat{\theta}^T \mathbf{X}$ , for one estimated triplet  $\hat{\theta}$  among the 50 estimations. The blue delimited upper region represents the region where both coordinates are above the 0.85 quantile of  $X_1$ . The red dotted line depicts the first diagonal.

### 3.7.4 Possible extension to hidden regular variation, a simulation study

Up to now, Section 3.7 only addressed the asymptotically dependent case. However, some similar results could be derived in the case of asymptotic independence.

For that, we consider a modified version of the  $c_i$  defined in (3.18). Let  $\varepsilon \in [0, 1]$  and define the following:

$$c_{i,\varepsilon} = \lim_{u \rightarrow \infty} \frac{\mathbb{P}(X_i > u \mid X_i > \varepsilon u)}{\mathbb{P}(H > u \mid H > \varepsilon u)}.$$

Then different cases arise according to the value of  $\varepsilon$ .

*Case  $\varepsilon = 0$ :*

We get  $c_{i,0} \in (0, \infty)$  if, and only if,  $X_i$  and  $H$  are tail equivalent. We get  $c_{i,0} = 0$  if the tail of  $X_i$  is less heavy than the tail of  $H$ . This corresponds to the previous coefficient defined in (3.18).

*Case  $\varepsilon > 0$ :*

If the tail function of  $H$  is regularly varying with index  $\alpha_H$  and the tail function of  $X_i$  is regularly varying with index  $\alpha_i$ , we obtain

$$c_{i,\varepsilon} = \varepsilon^{\alpha_i - \alpha_H}.$$

Thus,  $c_{i,\varepsilon} = 0$  only if  $X_i$  possesses light tails (assuming that  $H$  is heavy-tailed).

The following Lemma and the inference properties of  $c_{i,\varepsilon}$  from Proposition 3.12 are still an open question, but from the simulation results shown hereinafter, we are confident that this holds true.

**Lemma 3.11.** *Assume that the joint distribution of  $(\bar{g}(\mathbf{X}), H)$  satisfies the Ramos and Ledford model*

$$\mathbb{P}[\bar{g}(\mathbf{X}) > u, H > v] = L(u, v)(\bar{F}_g(u)v^{-\alpha_H})^{-1/2\eta}, \quad u, v > 0,$$

where  $\bar{F}_g$  denotes the tail distribution of  $\bar{g}(\mathbf{X})$ . Then, the following holds true:

1. If

$$\lim_{u \rightarrow \infty} \frac{\mathbb{P}(\bar{g}(\mathbf{X}) > u \mid \bar{g}(\mathbf{X}) > \varepsilon u)}{\mathbb{P}(H > u \mid H > \varepsilon u)} = 0,$$

we have that  $R_\varepsilon(g) = 1$  for all  $\varepsilon \in (0, 1]$ .

2. If  $R_\varepsilon(g) < 1$  and  $c_{i,\varepsilon} = 0$ , then, for all  $c > 0$ , we have that

$$R_\varepsilon(g) = R_\varepsilon(h_i(c)),$$

where  $\bar{h}_i(c) = \bar{g}(\mathbf{X}) + cX_i$ .

The second result in Lemma 3.11 emphasises that for partially linear classifiers, the components  $X_i$  for which  $c_{i,\varepsilon} = 0$ , do not have any effect, and can thus be removed as in the asymptotic dependent case.

Note that if we assume that the tail function of  $H$  is regularly varying with index  $\alpha_H$ , then,  $\lim_{u \rightarrow \infty} \mathbb{P}(H > u \mid H > \varepsilon u) = \varepsilon^{\alpha_H}$ . Thus, instead of estimating  $c_{i,\varepsilon}$ , we can equivalently estimate

$$c_{i,\varepsilon}^* := \varepsilon^{\alpha_H} c_{i,\varepsilon} = \lim_{u \rightarrow \infty} \mathbb{P}(X_i > u \mid X_i > \varepsilon u).$$

Therefore, a good candidate is the following empirical estimator

$$\hat{c}_{i,\varepsilon} := \frac{\sum_{j=1}^n \mathbf{1}\{X_{j,i} > u_n\}}{\sum_{j=1}^n \mathbf{1}\{X_{j,i} > \varepsilon u_n\}}, \text{ for large } u_n.$$

The inference results for the estimation of the  $c_{i,\varepsilon}^*$ , when  $\varepsilon > 0$ , are obtained through the following proposition, which is the analogue of Proposition 3.10 for  $\varepsilon > 0$ .

**Proposition 3.12.** *Assume that*

$$\lim_{u \rightarrow \infty} \mathbb{P}(X_i > u_n \mid X_i > \varepsilon u_n) \in [0, 1]$$

*exists. If  $u_n \rightarrow \infty$  and  $n \mathbb{P}(X_i > \varepsilon u_n) \rightarrow \infty$  and, additionally,*

$$\sqrt{n \mathbb{P}(X_i > \varepsilon u_n)} \left( \frac{\mathbb{P}(X_i > u_n)}{\mathbb{P}(X_i > \varepsilon u_n)} - c_{i,\varepsilon}^* \right) \rightarrow 0,$$

*then, we have*

$$\sqrt{n \mathbb{P}(X_i > \varepsilon u_n)} \left( \frac{\sum_{j=1}^n \mathbf{1}\{X_{j,i} > u_n\}}{\sum_{j=1}^n \mathbf{1}\{X_{j,i} > \varepsilon u_n\}} - c_{i,\varepsilon}^* \right) \rightarrow \mathcal{N} \left( 0, c_{i,\varepsilon}^* \cdot [1 - c_{i,\varepsilon}^*] \right).$$

The same simulation framework as in Section 3.4 is carried out, i.e.  $X_1 \sim \text{Pareto}(3)$ ,  $X_2 \sim \text{Pareto}(2)$ ,  $X_3 \sim \text{Exp}(1)$ ,  $X_4 \sim \text{Exp}(2)$ ,  $N \sim \text{Pareto}(2)$  and  $H = X_1 + N$ , where  $X_1, \dots, X_4$  and  $N$  are independent.

The theoretical values of the limits  $c_{i,\varepsilon}^*$  are then given by

$$c_{1,\varepsilon}^* = \varepsilon^3, \quad c_{2,\varepsilon}^* = \varepsilon^2, \quad c_{3,\varepsilon}^* = 0, \quad c_{4,\varepsilon}^* = 0.$$

Now to compare with the theoretical values, we generate 1000 samples with sample size  $n$  and threshold  $u$  defined as a given quantile of  $H$ . Different sample sizes are considered  $n \in \{5000, 1e4, 2e4, 1e5\}$  and  $u$  is chosen such that we get 300 exceedances each time. The 95% confidence intervals are obtained using the convergence result of Proposition 3.12.



True value	$n = 1000$ $n_x = 300$ $q = 70$	$n = 5000$ $n_x = 300$ $q = 94$	$n = 10000$ $n_x = 300$ $q = 97$	$n = 20000$ $n_x = 300$ $q = 98.5$	$n = 1e5$ $n_x = 300$ $q = 99.7$	
$c_{1,\varepsilon}^*$	0.34	0.34 <sub>94.8%</sub>	0.34 <sub>93.7%</sub>	0.34 <sub>93.7%</sub>	0.34 <sub>93.6%</sub>	0.35 <sub>94.2%</sub>
$c_{2,\varepsilon}^*$	0.49	0.49 <sub>95.8%</sub>	0.49 <sub>95.0%</sub>	0.49 <sub>93.8%</sub>	0.49 <sub>94.3%</sub>	0.49 <sub>95.0%</sub>
$c_{3,\varepsilon}^*$	0	0.36 <sub>0%</sub>	0.16 <sub>0%</sub>	0.09 <sub>41.7%</sub>	0.05 <sub>99%</sub>	0 <sub>100%</sub>
$c_{4,\varepsilon}^*$	0	0.12 <sub>92.8%</sub>	0.02 <sub>98.6%</sub>	0 <sub>99.7%</sub>	0 <sub>100%</sub>	0 <sub>100%</sub>

Table 3.4: Empirical mean estimations of the limits  $c_{i,\varepsilon}^*$  for different sample sizes  $n$  and 95% confidence interval coverage percentages computed among the 1000 replicates. The number of exceedances is denoted by  $n_x$  and the percentile chosen for the threshold is given by  $q$ . The value  $\varepsilon$  is equal to 0.7.

The simulation results are presented in Table 3.4. It appears that the sample size does not impact the estimates. We also see that  $X_1$  and  $X_2$  are detected as contributing to the extremes of  $H$  and that we could exclude  $X_3$  and  $X_4$  from our classification task but, in that case, confidence intervals appear to be difficult to find.

### 3.8 Summary of Paper II

#### To sum up

- Paper II addressed the specific issue of comparing binary classifiers in the case of rare event prediction.
  - Asymptotic dependent models: considering our first risk function  $R(\cdot)$  (which counts the number of mistakes), we have a benchmark with the naive classifiers and we can rank different classifiers.
  - Asymptotic independent models: for regression models of the form  $H = f(\mathbf{X}) + N$ , with tail of  $N$  heavier than tail of  $f(\mathbf{X})$ , any classifier will be as good as the naive classifiers.
- To tackle this issues, development of risk functions adapted to extremal classifiers.

- Flexible framework that allows to address both asymptotic dependence and asymptotic independence.
- Tool to identify the explanatory variables that contribute the most to extremal behaviour, with promising results in the case of hidden regular variation.
- ↗ Assumption of multivariate regular variation implies to work with heavy-tailed data.
- ↗ Transforming the data into identical margins would, in particular, defeat the optimisation scheme of the linear classifier.



# Chapter 4

## Some directions for future work

In three years, many avenues are explored, offering various topics for future research. Hereinafter we list some potential extensions of the work presented in this thesis.

### Extension of the simulation model to the multivariate case

Recall that in Chapter 2, we considered the stochastic representation of a standard GP vector through

$$\mathbf{Z} = E + \mathbf{T} - \max(\mathbf{T}), \quad (4.1)$$

with  $E$  a unit exponential variable and  $\mathbf{T}$  a multivariate vector.

For a bivariate vector  $\mathbf{Z} = (Z_1, Z_2)$ , defining  $\Delta = Z_1 - Z_2$  allowed to rewrite the representation (4.1) in a simple form, which was then easy to use for non-parametric simulations:

$$Z_i = E \pm \Delta, \text{ given the sign of } \Delta.$$

Now consider for example that  $\max(\mathbf{T}) = T_1$ . Then the above can be written as follows

$$Z_1 = E \text{ and } Z_2 = E - \Delta.$$

This gives hints for an extension to the multivariate case. Indeed, if we consider  $\mathbf{Z} = (Z_1, \dots, Z_d)$ ,  $d > 2$ , a first step would be to determine the index  $j$  such that  $\max(\mathbf{T}) = T_j$ , this could be achieved through a random sampling between  $\{1, \dots, d\}$ . Then defining  $\Delta_i = T_j - T_i$  for  $i \neq j$ , this leads to

$$Z_j = E \text{ and } Z_i = E - \Delta_i, \text{ for } i \neq j. \quad (4.2)$$

Simulating independent and identically distributed (iid) unit exponential variables  $E$  is costless. As for the bivariate case, the main point is the simulation of the  $\Delta_i$ . An avenue to explore is to bootstrap observations from the (d-1)-dimensional

set of points  $\Delta_{-j} := (\Delta_1, \dots, \Delta_{j-1}, \Delta_{j+1}, \dots, \Delta_d)$ . This could be achieved by re-sampling on the indexes: assuming that we have  $n$  iid replicates of  $\Delta_{-j}$ , random draws with replacements in  $\{1, \dots, n\}$  leads to  $m$  bootstrap samples

$$(\Delta_{1,k}^{(m)}, \dots, \Delta_{j-1,k}^{(m)}, \Delta_{j+1,k}^{(m)}, \dots, \Delta_{d,k}^{(m)}), \quad 1 \leq k \leq m.$$

From this, we could then perform non-parametric bootstrap MGP simulations in dimension greater than 2.

Perhaps the most challenging point for our application to extreme wave heights is the multivariate extension of the marginal regression models. Indeed, if we assume that instead of having only one offshore site, two or three sites are considered (represented in red in left-hand side of Figure 4.1). From an application perspective, this could certainly improve the coastal predictions as we would include more information in the covariates. But up to now, it is unclear how to define the conditioning part as in Equation (2.2). For the coastal marginal model, a possibility could be of the following

$$[H_c - v_c \mid H_o^1, T_p^1, D_p^1, H_o^2, T_p^2, D_p^2] \sim EGPD,$$

and to consider that the scale parameter of the EGPD vary as a smooth function of both  $(T_p^1, D_p^1)$  and  $(T_p^2, D_p^2)$  (assuming that we have two offshore grid point, site 1 with parameters  $(H_o^1, T_p^1, D_p^1)$  and site 2 with  $(H_o^2, T_p^2, D_p^2)$ ). This certainly leaves rooms for future developments.

In terms of applications of such multivariate simulation model, in addition to the above-mentioned one, which would allow for more refined simulations, another interesting application would be for the recovery of historical data. A possible approach is schematised in the right-hand side of Figure 4.1 and would be to consider two well documented sites such as buoys (represented by the two red dots in Figure 4.1), and to simulate data between this two measurement sites (represented by the blue dot). This idea was suggested by J er emy Rohmer during a visit at BRGM, and would deeply increase the understanding of past extreme wave events, for which in-situ measurements are rather sparse.

### Comparison of coastal extreme wave warning models

For coastal risk assessment, it is crucial to have efficient extreme sea state warning systems. From a simulation methods of extreme  $H_s$  as developed in this manuscript, one could then derive a warning (or forecast) model by looking at the probability of the event  $\{Y^u = +1\}$ , for a given high threshold  $u$ , where  $Y^u$  is defined as follows

$$Y^u = \begin{cases} +1, & \text{if } H_s > u, \\ -1, & \text{otherwise.} \end{cases}$$

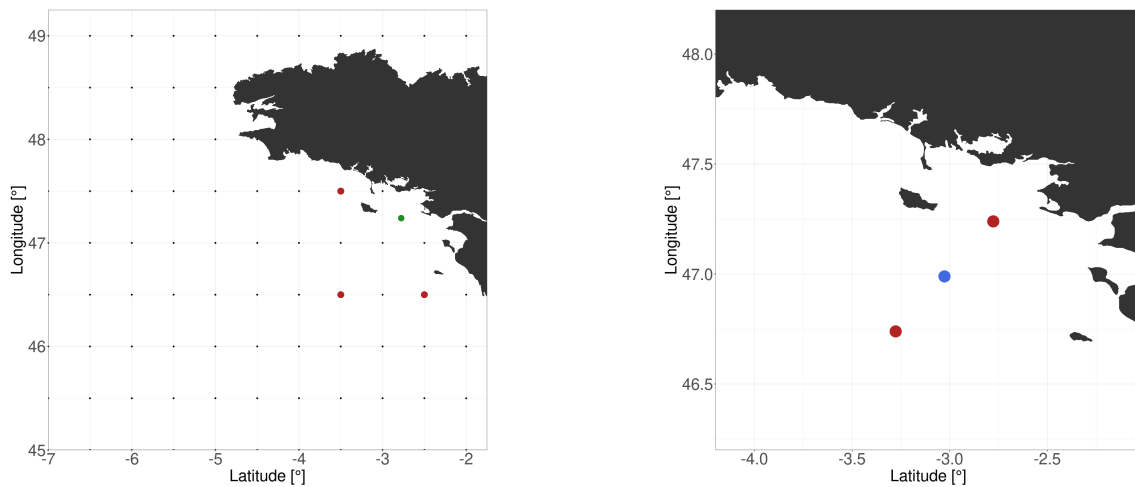


Figure 4.1: Two possible applications of a multivariate simulation model: (left) Refinement of coastal simulations; (right) Gap-filling in historical data.

In order to have a "good" warning model, one needs to ensure that the forecast does not produce too many false alarms or miss. In this context, the risk function developed in Paper II appears to be entirely relevant. The performances of several prediction models could be assessed on historical/hindcast data in order to select the most efficient in terms of extreme wave events forecast. Such future developments are definitely something worth trying.

### Application to climate projections

Finally, regarding adaptation strategies for future climate conditions, it is of fundamental importance to be able to derive information on the changes in the occurrence of extreme wave events between current and future climates, but also on the variability between different climate projection models. As discussed in the introduction of Chapter 2, such questions have been addressed on a global scale (e.g. [Aarnes et al., 2017](#)), but local variability has been less explored.

Assuming that the relation between the offshore point and the local point remains identical, some future extension of the presented work could then be the simulation of extreme coastal  $H_s$  given large scale wave climate projections.

This is somewhat linked to the idea of statistical downscaling, where it is assumed that there exists a transformation between the large scale and the local scale, and that this transformation will remain valid in the future climate. Such methods have been applied for instance in [Bechler et al. \(2015\)](#) in spatial context to downscale extreme rainfalls, or in [Towe et al. \(2017\)](#) to produce future extreme  $H_s$  from downscaled wind fields.



# Appendix A

## Technical proofs of Paper II

### A.1 Proofs

#### A.1.1 Proof of Lemma 3.4:

We can write that

$$\begin{aligned}\mathbb{P}(A_1 \cup B_1 | A_\varepsilon \cap B_\varepsilon) &= \frac{\mathbb{P}((A_1 \cup B_1) \cap (A_\varepsilon \cap B_\varepsilon))}{\mathbb{P}(A_\varepsilon \cap B_\varepsilon)}, \\ &= \frac{\mathbb{P}(A_1 \cap B_\varepsilon) + \mathbb{P}(A_\varepsilon \cap B_1) - \mathbb{P}(A_1 \cap B_1)}{\mathbb{P}(A_\varepsilon \cap B_\varepsilon)}.\end{aligned}$$

In the same way, we have

$$\mathbb{P}(A_1 \triangle B_1 | A_\varepsilon \cap B_\varepsilon) = \frac{\mathbb{P}(A_1 \cap B_\varepsilon) + \mathbb{P}(A_\varepsilon \cap B_1) - 2\mathbb{P}(A_1 \cap B_1)}{\mathbb{P}(A_\varepsilon \cap B_\varepsilon)}.$$

Hence, we deduce that

$$R(A_\varepsilon, B_\varepsilon) = \frac{\mathbb{P}(A_1 \triangle B_1 | A_\varepsilon \cap B_\varepsilon)}{\mathbb{P}(A_1 \cup B_1 | A_\varepsilon \cap B_\varepsilon)} = 1 - \left[ \frac{\mathbb{P}(A_1 \cap B_\varepsilon)}{\mathbb{P}(A_1 \cap B_1)} + \frac{\mathbb{P}(A_\varepsilon \cap B_1)}{\mathbb{P}(A_1 \cap B_1)} - 1 \right]^{-1}.$$

The expression given by (3.8) follows.

Item (a) of the lemma is based on the following inequality

$$\mathbb{P}(U | V) \geq \mathbb{P}(U | W), \text{ if the sets } U, V \text{ and } W \text{ satisfy } U \subset V \subset W.$$

For item (b), note that

$$\begin{aligned}R(A_\varepsilon, B_\varepsilon) &\leq R(A'_\varepsilon, B_\varepsilon) \\ \iff (1 - R(A_\varepsilon, B_\varepsilon))^{-1} &\leq (1 - R(A'_\varepsilon, B_\varepsilon))^{-1} \\ \iff \frac{\mathbb{P}(A_1 \cap B_\varepsilon)}{\mathbb{P}(A_1 \cap B_1)} + \frac{\mathbb{P}(A_\varepsilon \cap B_1)}{\mathbb{P}(A_1 \cap B_1)} &\leq \frac{\mathbb{P}(A'_1 \cap B_\varepsilon)}{\mathbb{P}(A'_1 \cap B_1)} + \frac{\mathbb{P}(A'_\varepsilon \cap B_1)}{\mathbb{P}(A'_1 \cap B_1)}, \\ \iff \mathbb{P}(A_1 \cap B_\varepsilon) + \mathbb{P}(A_\varepsilon \cap B_1) &\leq \frac{\mathbb{P}(A_1 \cap B_1)}{\mathbb{P}(A'_1 \cap B_1)} [\mathbb{P}(A'_1 \cap B_\varepsilon) + \mathbb{P}(A'_\varepsilon \cap B_1)].\end{aligned}$$



As  $\mathbb{P}(A_1 \cap B_1) = \mathbb{P}(A'_1 \cap B_1)$  and  $\mathbb{P}(A_1 \cap B_\varepsilon) = \mathbb{P}(A'_1 \cap B_\varepsilon)$  and  $\mathbb{P}(A_\varepsilon \cap B_1) \leq \mathbb{P}(A'_\varepsilon \cap B_1)$ , then

$$R(A_\varepsilon, B_\varepsilon) \leq R(A'_\varepsilon, B_\varepsilon).$$

This provides the second statement (b) since we assume that  $A_1 = A'_1$  and  $A_\varepsilon \subseteq A'_\varepsilon$ .

Item (c) is a direct consequence of (3.8).  $\square$

### A.1.2 Proof of Lemma 3.6:

From Eq. (3.12), we know that we can only get  $R(\bar{g}) < 1$ , only if  $H$  and  $\bar{g}(X)$  are tail equivalent. Thus, this will be assumed in the following. Let  $w$  such that  $\delta := \lim_{u \rightarrow \infty} w(u)/u \in (0, 1)$ , then for any positive  $u$  we can write that

$$\begin{aligned} \frac{\mathbb{P}(f(X) + N > u, \bar{g}(X) > u)}{\mathbb{P}(\bar{g}(X) > u)} &\leq \frac{\mathbb{P}(f(X) > w(u), \bar{g}(X) > u)}{\mathbb{P}(\bar{g}(X) > u)} \\ &\quad + \frac{\mathbb{P}(N > u - w(u), \bar{g}(X) > u)}{\mathbb{P}(\bar{g}(X) > u)}. \end{aligned}$$

Since  $\bar{g}(X)$  and  $N$  are independent, the second term reduces to  $\mathbb{P}(N > u - w(u))$  which converges to 0 since  $u - w(u) \sim (1 - \delta)u$  as  $u$  gets large.

For the first term, we rewrite the ratio as follows

$$\begin{aligned} \frac{\mathbb{P}(f(X) > w(u), \bar{g}(X) > u)}{\mathbb{P}(\bar{g}(X) > u)} &\leq \frac{\mathbb{P}(f(X) > w(u))}{\mathbb{P}(\bar{g}(X) > u)} \\ &\leq \frac{\mathbb{P}(f(X) > w(u))}{\mathbb{P}(N > w(u))} \frac{\mathbb{P}(N > w(u))}{\mathbb{P}(N > u)} \frac{\mathbb{P}(N > u)}{\mathbb{P}(\bar{g}(X) > u)}. \end{aligned}$$

Since  $w(u) \rightarrow \infty$  and  $\mathbb{P}(f(X) > u) = o(\mathbb{P}(N > u))$ , the ratio  $\frac{\mathbb{P}(f(X) > w(u))}{\mathbb{P}(N > w(u))}$  goes to 0 as  $u$  gets large. From the assumption  $w(u) \sim \delta u$ ,  $\frac{\mathbb{P}(N > w(u))}{\mathbb{P}(N > u)}$  behaves as a constant when  $u \rightarrow \infty$ .

The only remaining term is  $\frac{\mathbb{P}(N > u)}{\mathbb{P}(\bar{g}(X) > u)}$  which converges to a constant due to tail equivalence. So,  $\lim \mathbb{P}(\bar{g}(X) > u \mid H > u) = 0$ .  $\square$

### A.1.3 Proof of Proposition 3.7:

In Lemma 3.4 we fix  $A_\varepsilon = \{\bar{g}(X) > \varepsilon u\}$  and  $B_\varepsilon = \{H > \varepsilon u\}$  and  $A_1 = \{\bar{g}(X) > u\}$  and  $B_1 = \{H > u\}$ .

$$\mathbb{P}(H > u \text{ or } \bar{g}(X) > u \mid \min\{H, \bar{g}(X)\} > \varepsilon u) = \mathbb{P}(A_1 \cup B_1 \mid A_\varepsilon \cap B_\varepsilon),$$

and

$$\mathbb{P}(g(X; u) \neq Y^{(u)} \mid \min\{H, \bar{g}(X)\} > \varepsilon u) = \mathbb{P}(A_1 \Delta B_1 \mid A_\varepsilon \cap B_\varepsilon).$$

The Ramos and Ledford model corresponds to the special of case of (3.9)

$$\mathbb{P}[\bar{g}(X) > u, H > v] = L(u, v)(u^{-\alpha_g} v^{-\alpha_H})^{1/2\eta} = \mathbb{P}(A_\varepsilon \cap B_{\varepsilon'}) = c_{\varepsilon, \varepsilon'} \mathbb{P}^a(A_\varepsilon) \mathbb{P}^b(B_{\varepsilon'})$$

with

$$\begin{aligned} \mathbb{P}(A_\varepsilon) &= \mathbb{P}(\bar{g}(X) > \varepsilon u) = L_g(\varepsilon u) \varepsilon^{-\alpha_g} u^{-\alpha_g}, \\ \mathbb{P}(B_\varepsilon) &= \mathbb{P}(H > \varepsilon u) = L_H(\varepsilon u) \varepsilon^{-\alpha_H} u^{-\alpha_H} \end{aligned}$$

and

$$c_{\varepsilon, \varepsilon'} = L(\varepsilon u, \varepsilon u) / L(u, u), \quad a = \alpha_g / (2\eta) \text{ and } b = \alpha_H / (2\eta).$$

Then, from (3.10)

$$R(A_\varepsilon, B_\varepsilon) = 1 - \left[ \frac{c_{\varepsilon, 1}}{c_{1, 1}} (\mathbb{P}(A_1 | A_\varepsilon))^{-a} + \frac{c_{1, \varepsilon}}{c_{1, 1}} (\mathbb{P}(B_1 | B_\varepsilon))^{-b} - 1 \right]^{-1}.$$

Letting  $u$  gets large provides the required result.  $\square$

#### A.1.4 Proof of Proposition 3.8:

As we assume that  $R_\varepsilon(g)$  exists (for some  $\varepsilon \geq 0$ ), for  $u_n \rightarrow \infty$  such that  $np_{g, \varepsilon}(u_n) \rightarrow \infty$ , we obtain that

$$\begin{aligned} \sqrt{np_{g, \varepsilon}(u_n)} & \left[ \left( \frac{\sum_{i=1}^n \mathbb{1}\{g(\mathbf{X}_i, u_n) \neq Y_i^{(u_n)}, H_i > \varepsilon u_n, g(\mathbf{X}_i; \varepsilon u_n) = 1\}}{np_{g, \varepsilon}(u_n)} \right) \right. \\ & \left. - \left( \frac{\mathbb{P}(g(\mathbf{X}, u_n) \neq Y^{(u_n)}, H > \varepsilon u_n, g(\mathbf{X}; \varepsilon u_n) = 1)}{p_{g, \varepsilon}(u_n)} \right) \right] \\ & \xrightarrow{n \rightarrow \infty} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R_\varepsilon(g) & R_\varepsilon(g) \\ R_\varepsilon(g) & 1 \end{pmatrix} \right) \end{aligned}$$

Provided that the bias is negligible, i.e.

$$\lim_{n \rightarrow \infty} \sqrt{np_{g, \varepsilon}(u_n)} \left( \frac{\mathbb{P}(g(\mathbf{X}, u_n) \neq Y^{(u_n)}, H > \varepsilon u_n, g(\mathbf{X}; \varepsilon u_n) = 1)}{p_{g, \varepsilon}(u_n)} - R_\varepsilon(g) \right) = 0,$$

the Delta method yields

$$\sqrt{np_{g, \varepsilon}(u_n)} \left( \hat{R}_{n, \varepsilon}(g) - R_\varepsilon(g) \right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, R_\varepsilon(g)(1 - R_\varepsilon(g))).$$

$\square$

## A.2 Proofs of the appendix

### A.2.1 Proof of Proposition 3.9

The proof is based on the following lemma which is proven in Subsection A.2.2.

**Lemma A.1.** *Under the assumptions from Proposition 3.9, for every compact subset  $K \subset C$ , the sequences of processes  $\{A_n(\theta), \theta \in K\}$  and  $\{B_n(\theta), \theta \in K\}$  defined by*

$$A_n(\theta) = \sqrt{\frac{n}{\mathbb{P}(H > u_n)}} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\{\theta^\top \mathbf{X}_i > u_n\} \Delta \{H_i > u_n\}\} - \mathbb{P}(g_\theta(X; u_n) \neq Y^{(u_n)}) \right)$$

$$B_n(\theta) = \sqrt{\frac{n}{\mathbb{P}(H > u_n)}} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\{\theta^\top \mathbf{X}_i > u_n\} \cup \{H_i > u_n\}\} - \mathbb{P}(\max\{g_\theta(X; u_n), Y^{(u_n)}\} = 1) \right)$$

converge to centered Gaussian processes  $\{A(\theta), \theta \in K\}$  and  $\{B(\theta), \theta \in K\}$ , respectively, weakly in  $\ell^\infty(K)$ .

If the function  $\theta \mapsto R(g_\theta)$  has a unique minimizer  $\theta^*$ , then, necessarily,  $R(g_{\theta^*}) < 1$ .

Now, similarly to the notation above, let  $J$  denote the set of indices  $j$  with  $c_j > 0$ , and let us consider  $\theta \in C$  such that  $\|\theta\|_\infty > k_0$  for some constant  $k_0 > 0$ . Then,

$$\widehat{R}_n(g_\theta) = 1 - \frac{\sum_{i=1}^n \mathbb{1}\{\min(\theta^\top \mathbf{X}_i, H_i) > u_n\}}{\sum_{i=1}^n \mathbb{1}\{\max(\theta^\top \mathbf{X}_i, H_i) > u_n\}} \geq 1 - \frac{\sum_{i=1}^n \mathbb{1}\{H_i > u_n\}}{\min_{j \in J} \sum_{i=1}^n \mathbb{1}\{k_0 X_{ij} > u_n\}}$$

$$\xrightarrow[n \rightarrow \infty]{p} 1 - \max_{j \in J} \frac{\mathbb{E}(\Omega^\alpha)}{k_0^\alpha \mathbb{E}(\Gamma_j^\alpha)}$$

where the right-hand side goes to 1 as  $k_0 \rightarrow \infty$ . Thus, as  $\widehat{R}_n(g_{\theta^*}) \rightarrow_p R(g_{\theta^*}) < 1$ , we obtain that, for sufficiently large  $k_0 \gg \|\theta^*\|$ , with probability going to one,

$$\widehat{R}_n(g_{\theta^*}) \leq \min_{\theta \in C \setminus [0, k_0]^d} \widehat{R}_n(g_\theta)$$

and, consequently,

$$\operatorname{argmin}_{\theta \in C} \widehat{R}_n(g_\theta) = \operatorname{argmin}_{\theta \in C \setminus [0, k_0]^d} \widehat{R}_n(g_\theta).$$

Now, we note that, by Lemma A.1, the bias conditions (3.22) and (3.23) and the functional delta method,  $\widehat{R}_n(g_\theta)$  converges in probability to  $R(g_\theta)$  uniformly on every compact subset of  $C$ . In particular,

$$\sup_{\theta \in C \cap [0, K]^d} \left| \widehat{R}_n(g_\theta) - R(g_\theta) \right| \rightarrow_p 0.$$

Thus,

$$\operatorname{argmin}_{\theta \in C \cap [0, K]^d} \widehat{R}_n(g_\theta) \rightarrow_p \operatorname{argmin}_{\theta \in C \cap [0, K]^d} R(g_\theta) = \theta^*.$$

□

### A.2.2 Proof of Lemma A.1

We will proof the lemma by applying the Central Limit Theorem 2.11.9 in [Van der Vaart and Wellner \(1996\)](#). To this end, we define the function spaces  $\mathcal{A} = \{a_\theta, \theta \in K\}$  and  $\mathcal{B} = \{b_\theta, \theta \in K\}$  where

$$\begin{aligned} a_\theta &: (0, \infty)^d \times (0, \infty) \rightarrow \{0, 1\}, & a_\theta(\mathbf{x}, h) &= \mathbb{1}\{\{\theta^\top \mathbf{x} > 1\} \Delta \{h > 1\}\} \\ b_\theta &: (0, \infty)^d \times (0, \infty) \rightarrow \{0, 1\}, & b_\theta(\mathbf{x}, h) &= \mathbb{1}\{\{\theta^\top \mathbf{x} > 1\} \cup \{h > 1\}\}. \end{aligned}$$

Then, with

$$Z_{nl}(f) = \frac{1}{\sqrt{n \mathbb{P}(H > u_n)}} f(u_n^{-1} \mathbf{X}_l, u_n^{-1} H_l), \quad f \in \mathcal{A} \cup \mathcal{B},$$

for  $l = 1, \dots, n$ , we have that

$$\{A_n(\theta), \theta \in K\} = \left\{ \sum_{l=1}^n (Z_{nl}(f) - \mathbb{E} Z_{nl}(f)), f \in \mathcal{A} \right\}$$

and

$$\{B_n(\theta), \theta \in K\} = \left\{ \sum_{l=1}^n (Z_{nl}(f) - \mathbb{E} Z_{nl}(f)), f \in \mathcal{B} \right\}.$$

Now, we have that

$$\max \{ \|Z_{nl}\|_{\mathcal{A}}, \|Z_{nl}\|_{\mathcal{B}} \} = \sup_{f \in \mathcal{A} \cup \mathcal{B}} |Z_{nl}(f)| \leq \frac{1}{\sqrt{n \mathbb{P}(H > u_n)}} \text{ a.s.}$$

for all  $l = 1, \dots, n$  and  $n \in \mathbb{N}$ .

Consequently, we check the Lindeberg condition: For  $k \in \mathbb{N}$ , we have

$$\lim_{n \rightarrow \infty} \sum_{l=1}^n \mathbb{E} \left( \|Z_{nl}\|_{\mathcal{A} \cup \mathcal{B}}^k \mathbb{1}\{\|Z_{nl}\|_{\mathcal{A} \cup \mathcal{B}} > \eta\} \right) \leq \lim_{n \rightarrow \infty} \frac{n}{\sqrt{n \mathbb{P}(H > u_n)}^k} \mathbb{1}\{n \mathbb{P}(H > u_n) < 1/\eta^2\} = 0$$

as  $n \mathbb{P}(H > u_n) \rightarrow \infty$  by definition. For  $k = 2$ , we obtain a Lindeberg type condition that ensures convergence of  $A_n$  and  $B_n$  to  $A$  and  $B$ , respectively, in terms of finite-dimensional distributions. For  $k = 1$ , we obtain the Lindeberg type condition of Theorem 2.11.9 in [Van der Vaart and Wellner \(1996\)](#).

It remains to check the equi-continuity condition. In the following, to simplify notation, we assume that  $C = [0, \infty)^d$ . Then, for  $\theta^{(1)}, \theta^{(2)} \in [a, b] \subset K \subset C$ , we have that

$$|a_{\theta^{(1)}}(u_n^{-1} \mathbf{X}, u_n^{-1} H) - a_{\theta^{(2)}}(u_n^{-1} \mathbf{X}, u_n^{-1} H)| \in \{0, 1\}$$

and

$$|b_{\theta^{(1)}}(u_n^{-1} \mathbf{X}, u_n^{-1} H) - b_{\theta^{(2)}}(u_n^{-1} \mathbf{X}, u_n^{-1} H)| \in \{0, 1\}$$

and the probability that any of those two expressions is equal to one is bounded by the probability

$$\begin{aligned} & \mathbb{P}(\mathbb{1}\{a^\top \mathbf{X} > u_n\} \neq \mathbb{1}\{b^\top \mathbf{X} > u_n\}) = \mathbb{P}(a^\top \mathbf{X} \leq u_n, b^\top \mathbf{X} > u_n) \\ & = \mathbb{P}\left(\|(X, H)\| > \frac{u_n}{\|K\|d}\right) \mathbb{P}\left(a^\top \mathbf{X} \leq u_n, b^\top \mathbf{X} > u_n \mid \|(X, H)\| > \frac{u_n}{\|K\|d}\right) \end{aligned}$$

where we use that  $b^\top \mathbf{X} > u_n$  implies that  $\|X\| > u_n/(\|K\|d)$  with  $\|K\| = \sup_{x \in K} \|x\|_\infty$ . Making use of the fact that  $\mathbb{P}(\|(X, H)\| > u_n/(\|K\|d)) \leq C_0(\|K\|d)^\alpha \mathbb{P}(H > u_n)$  for some constant  $C_0 > 0$  and the bound given by Equation (3.21), we obtain that

$$\begin{aligned} & \mathbb{P}(\mathbb{1}\{a^\top \mathbf{X} > u_n\} \neq \mathbb{1}\{b^\top \mathbf{X} > u_n\}) \\ & \leq C_0(\|K\|d)^\alpha \mathbb{P}(H > u_n)[1 + a(u_n/(\|K\|d))] \mathbb{P}(Pa^\top \Gamma \leq \|K\|d, Pb^\top \Gamma > \|K\|d) \\ & = C_0(\|K\|d)^\alpha \mathbb{P}(H > u_n)[1 + a(u_n/(\|K\|d))] \mathbb{E}_\Gamma \left( \mathbb{P}\left(P \in \left[\frac{\|K\|d}{b^\top \Gamma}, \frac{\|K\|d}{a^\top \Gamma}\right]\right) \right) \\ & \leq C_0 \mathbb{P}(H > u_n)[1 + a(u_n/(\|K\|d))] \mathbb{E}((b^\top \Gamma)^\alpha - (a^\top \Gamma)^\alpha) \\ & \leq 2C_0 \mathbb{P}(H > u_n) \|a - b\| \end{aligned}$$

provided that  $u_n$  is sufficiently large as  $a(u_n/(\|K\|d)) \rightarrow 0$ .

Consequently,

$$\sup_{\|f-g\| < \delta} \sum_{l=1}^n \mathbb{E}[(Z_{nl}(f) - Z_{nl}(g))^2] = 2C_0\delta,$$

which tends to 0 as  $\delta \rightarrow 0$ . From this inequality, it can also be seen that any partition of  $K$  into hypercubes with length  $\varepsilon^2/(2C_0)$  leads to a valid  $\varepsilon$ -bracketing, i.e. the number  $N_\varepsilon \propto 1/\varepsilon^{2d}$  grows with a power rate and, so,  $\sqrt{\log(N_\varepsilon)}$  is integrable.

Thus, by Theorem 2.11.9, the processes  $A_n$  and  $B_n$  converge to Gaussian processes  $A$  and  $B$ , weakly in  $\ell^\infty(K)$ .  $\square$

# Bibliography

- Aarnes, O. J., Reistad, M., Breivik, , Bitner-Gregersen, E., Ingolf Eide, L., Gramstad, O., Magnusson, A. K., Natvig, B., and Vanem, E. (2017). Projected changes in significant wave height toward the end of the 21st century: Northeast atlantic. *Journal of Geophysical Research: Oceans*, 122(4):3394–3403.
- Accensi, M. and Maisondieu, C. (2015). HOMERE. <https://doi.org/10.12770/cf47e08d-1455-4254-955e-d66225c9dc90>.
- Aghbalou, A., Portier, F., Sabourin, A., and Zhou, C. (2021). Tail inverse regression for dimension reduction with extreme response. arXiv:2108.01432.
- Ahmed, M., Maume-Deschamps, V., and Ribereau, P. (2022). Recognizing a spatial extreme dependence structure: A deep learning approach. *Environmetrics*, 33(4):e2714.
- Ailliot, P., Baxevani, A., Cuzol, A., Monbet, V., and Raillard, N. (2011). Space-time models for moving fields with an application to significant wave height fields. *Environmetrics*, 22(3):354–369.
- Ardhuin, F. and Accensi, M. (2014). IOWAGA. <https://sextant.ifremer.fr/record/c87f6f24-63b4-46ec-b40e-f185a61dc672/>.
- Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023 – 2050.
- Bacro, J.-N., Bel, L., and Lantuéjoul, C. (2010). Testing the independence of maxima: from bivariate vectors to spatial extreme fields. *Extremes*, 13(2):155–175.
- Bacro, J.-N. and Toulemonde, G. (2013). Measuring and modelling multivariate and spatial dependence of extremes. *Journal de la société française de statistique*, 154(2):139–155.
- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792 – 804.

- Bechler, A., Vrac, M., and Bel, L. (2015). A spatial hybrid approach for downscaling of extreme precipitation fields. *Journal of Geophysical Research: Atmospheres*, 120(10):4534–4550.
- Beersma, J. J. and Buishand, T. A. (2004). Joint probability of precipitation and discharge deficits in the Netherlands. *Water resources research*, 40(12).
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley.
- Bertin, X., Bruneau, N., Breilh, J.-F., Fortunato, A. B., and Karpytchev, M. (2012). Importance of wave age and resonance in storm surges: The case Xynthia, Bay of Biscay. *Ocean Modelling*, 42:16–30.
- Bertin, X., Prouteau, E., and Letetrel, C. (2013). A significant increase in wave height in the North Atlantic Ocean over the 20th century. *Global and Planetary Change*, 106:77–83.
- Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196 – 1217.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press.
- Brehmer, J. R. and Stokorb, K. (2019). Why scoring functions cannot assess tail properties. *Electronic Journal of Statistics*, 13(2):4015–4034.
- Bricheno, L. M. and Wolf, J. (2018). Future wave conditions of Europe, in response to high-end climate change scenarios. *Journal of Geophysical Research: Oceans*, 123(12):8762–8791.
- Brodin, E. and Rootzén, H. (2009). Univariate and bivariate GPD methods for predicting extreme wind storm losses. *Insurance: Mathematics and Economics*, 44(3):345–356.
- Cai, J. J., Fougères, A.-L., and Mercadier, C. (2013). Environmental data: multivariate Extreme Value Theory in practice. *Journal de la société française de statistique*, 154(2):178–199.
- Caires, S. and Sterl, A. (2005). 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *Journal of Climate*, 18(7):1032–1048.
- Carreau, J. and Bengio, Y. (2009). A hybrid pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76.

- Casas-Prat, M., Wang, X. L., and Sierra, J. P. (2014). A physical-based statistical method for modeling ocean wave heights. *Ocean Modelling*, 73:59–75.
- Castillo, E. and Sarabia, J. M. (1992). Engineering analysis of extreme value data: Selection of models. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 118(2):129–146.
- Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 54(1):207–222.
- Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610.
- Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, 11(10):1305–1319.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag London.
- Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339–365.
- Collins, M., Sutherland, M., Bouwer, L., Cheong, S.-M., Frölicher, T., Combes, H. J. D., Roxy, M. K., Losada, I., McInnes, K., Ratter, B., Rivera-Arriaga, E., Susanto, R., Swingedouw, D., and Tibig, L. (2019). Extremes, abrupt changes and managing risk. In Pörtner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N., editors, *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. Cambridge University Press, Cambridge, UK and New York, NY, USA.
- Das, B. and Resnick, S. I. (2015). Models with hidden regular variation: Generation and detection. *Stochastic Systems*, 5(2):195 – 238.
- Davison, A. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, 2(1):203–235.
- de Carvalho, M., Pereira, S., Pereira, P., and de Zea Bermudez, P. (2021+). An extreme value bayesian lasso for the conditional left and right tails. *Journal of Agricultural, Biological and Environmental Statistics*.
- de Carvalho, M. and Ramos, A. (2012). Bivariate extreme statistics, II. *REVSTAT-Statistical Journal*, 10(ARTICLE):83–107.



- de Haan, L. and de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–45.
- de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York.
- De Leo, F., Besio, G., Briganti, R., and Vanem, E. (2021). Non-stationary extreme value analysis of sea states based on linear trends. Analysis of annual maxima series of significant wave height and peak period in the Mediterranean Sea. *Coastal Engineering*, 167:103896.
- Dey, D. and Yan, J. (2016). *Extreme Value Modeling and Risk Analysis: Methods and Applications*. CRC Press.
- Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251 – 280.
- Drees, H. and Janßen, A. (2017). Conditional extreme value models: fallacies and pitfalls. *Extremes*, 20(4):777–805.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1 – 26.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling Extremal Events: for Insurance and Finance*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.
- Engelke, S., Hitz, A. S., and Gnecco, N. (2019). *graphicalExtremes: Statistical Methodology for Graphical Extreme Value Models*. R package version 0.1.0.9000.
- Engelke, S. and Ivanovs, J. (2020). Sparse structures for multivariate extremes.
- Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8(1):241–270.
- Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer, Berlin.
- Ewans, K. and Jonathan, P. (2008). The effect of directionality on northern north sea extreme wave design criteria. *Journal of Offshore Mechanics and Arctic Engineering*, 130(4).
- Ewans, K. and Jonathan, P. (2014). Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. *Journal of Marine Systems*, 130:124–130.
- Falk, M. and Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized pareto distributions. *Journal of Multivariate Analysis*, 99(4):715–734.

- Falk, M., Hüsler, J., and Reiss, R.-D. (2010). *Laws of small numbers: extremes and rare events*. Springer Science & Business Media.
- Feld, G., Randell, D., Wu, Y., Ewans, K., and Jonathan, P. (2014). Estimation of storm peak and intra-storm directional-seasonal design conditions in the north sea. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, 4.
- Ferro, C. A. T. (2007). A probability model for verifying deterministic forecasts of extreme events. *Weather and Forecasting*, 22(5):1089 – 1100.
- Ferro, C. A. T. and Stephenson, D. B. (2011). Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, 26(5):699 – 713.
- Filipot, J.-F. and Ardhuin, F. (2012). A unified spectral parameterization for wave breaking: From the deep ocean to the surf zone. *Journal of Geophysical Research: Oceans*, 117(C11).
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.
- Fougères, A.-L. (2004). Multivariate extremes. *Extreme values in finance, telecommunications, and the environment*, 373:388.
- Friederichs, P., Göber, M., Bentzien, S., Lenz, A., and Krampitz, R. (2009). A probabilistic analysis of wind gusts using extreme value statistics. *Meteorologische Zeitschrift*, 18(6):615.
- Friederichs, P. and Thorarinsdottir, T. L. (2012). Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics*, 23.
- Genovese, E. and Przulski, V. (2013). Storm surge disaster risk management: the Xynthia case study in France. *Journal of Risk Research*, 16(7):825–841.
- Gnecco, N., Meinshausen, N., Peters, J., and Engelke, S. (2021). Causal discovery in heavy-tailed models. *The Annals of Statistics*, 49(3):1755 – 1778.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453.
- Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268.

- Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422.
- Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, 56(294):335–349.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239.
- Haruna, A., Blanchet, J., and Favre, A.-C. (2021+). Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation. *Hydrology and Earth System Sciences Discussions*.
- Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3):297 – 310.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd edition)*. Springer series in statistics. Springer, New York.
- Heffernan, J. and Resnick, S. (2005). Hidden regular variation and the rank transform. *Advances in Applied Probability*, 37(2):393–414.
- Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes*, 3(3):279–290.
- Heffernan, J. E., Stephenson, A. G., and Gilleland, E. (2018). *ismev: An Introduction to Statistical Modeling of Extreme Values*. R package version 1.42.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

- Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020). *copula: Multivariate Dependence with Copulas*. R package version 1.0-1.
- Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge university press.
- Huser, R. (2021). EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes. *Extremes*, 24(1):91–104.
- Idier, D., Rohmer, J., Pedreros, R., Le Roy, S., Lambert, J., Louisor, J., Le Cozannet, G., and Le Cornec, E. (2020). Coastal flood: a composite method for past events characterisation providing insights in past, present and future hazards—joining historical, statistical and modelling approaches. *Natural Hazards*, 101(2):465–501.
- IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland. Core Writing Team, Pachauri, R.K. and Meyer, L.A. (eds.).
- Jalalzai, H., Cléménçon, S., and Sabourin, A. (2018). On Binary Classification in Extreme Regions. In *Advances in Neural Information Processing Systems*, pages 3092–3100.
- Jolliffe, I. T. and Stephenson, D. B. (2003). *Forecast Verification: a Practitioner's Guide in Atmospheric Science*. Chichester, U.K.: Wiley.
- Jonathan, P. and Ewans, K. (2013). Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering*, 62:91–109.
- Jonathan, P., Ewans, K., and Randell, D. (2014). Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25(3):172–188.
- Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25(8-12):1287–1304.
- Kiriliouk, A. and Naveau, P. (2020). Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *The Annals of Applied Statistics*, 14(3):1342 – 1358.
- Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, 61(1):123–135.
- Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81(4):673–680.

- Le Carrer, N. (2022). egpd4gamlss. <https://github.com/noemielc/egpd4gamlss>.
- Le Gall, P., Favre, A.-C., Naveau, P., and Prieur, C. (2022). Improved regional frequency analysis of rainfall data. *Weather and Climate Extremes*, page 100456.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.
- Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.
- Legrand, J., Ailliot, P., Naveau, P., and Raillard, N. (2022+). Joint stochastic simulation of extreme coastal and offshore significant wave heights.
- Legrand, J., Naveau, P., and Oesting, M. (2021). Evaluation of binary classifiers for asymptotically dependent and independent extremes. arXiv:2112.13738.
- Lehtomaa, J. and Resnick, S. I. (2020). Asymptotic independence and support detection techniques for heavy-tailed multivariate data. *Insurance: Mathematics and Economics*, 93:262–277.
- Lemos, G., Menendez, M., Semedo, A., Miranda, P., and Hemer, M. (2021). On the decreases in North Atlantic significant wave heights from climate projections. *Climate Dynamics*, 57(9):2301–2324.
- Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F., and Gneiting, T. (2017). Forecaster’s Dilemma: Extreme Events and Forecast Evaluation. *Statistical Science*, 32(1):106–127.
- Lobeto, H., Menendez, M., and Losada, I. J. (2021). Future behavior of wind wave extremes due to climate change. *Scientific reports*, 11(1):1–12.
- Marcon, G., Naveau, P., and Padoan, S. (2017). A semi-parametric stochastic generator for bivariate extreme events. *Stat*, 6(1):184–201.
- Marcos, M., Rohmer, J., Vousdoukas, M. I., Mentaschi, L., Le Cozannet, G., and Amores, A. (2019). Increased extreme coastal water levels due to the combined action of storm surges and wind waves. *Geophysical Research Letters*, 46(8):4356–4364.
- Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.
- Maulik, K. and Resnick, S. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

- Meyer, N. (2020). *High-dimensional Learning for Extremes*. Theses, Sorbonne Université.
- Meyer, N. and Wintenberger, O. (2021). Sparse regular variation. *Advances in Applied Probability*, 53(4):1115–1148.
- Mhalla, L., Chavez-Demoulin, V., and Dupuis, D. J. (2020). Causal mechanism of extreme river discharges in the upper danube basin network. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(4):741–764.
- Michel, M., Obakrim, S., Raillard, N., Ailliot, P., and Monbet, V. (2022). Deep learning for statistical downscaling of sea states. *Advances in Statistical Climatology, Meteorology and Oceanography*, 8(1):83–95.
- Michel, R. (2006). *Simulation and estimation in multivariate generalized Pareto models*. PhD thesis, Universität Würzburg.
- Morim, J., Trenham, C., Hemer, M., Wang, X., Mori, N., Casas-Prat, M., Semedo, A., Shimura, T., Timmermans, B., Camus, P., Bricheno, L., Mentaschi, L., Dobrynin, M., Feng, Y., and Erikson, L. (2020). A global ensemble of ocean wave climate projections from CMIP5-driven models. *Scientific Data*, 7:105.
- Mousslim, H., Babarit, A., Clément, A., and Borgarino, B. (2009). Development of the french wave energy test site SEM-REV. In *Proceedings of the 8th European wave and tidal energy conference*, page 31 – 35, Uppsala, Sweden.
- Méndez, F. J., Menéndez, M., Luceño, A., Medina, R., and Graham, N. E. (2008). Seasonality and duration in extreme value distributions of significant wave height. *Ocean Engineering*, 35(1):131–138.
- Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.
- Nelsen, R. B. (2006). *An Introduction to Copulas (Springer Series in Statistics)*. Springer.
- Nicolae Lerma, A., Bulteau, T., Lecacheux, S., and Idier, D. (2015). Spatial variability of extreme wave height along the Atlantic and channel French coast. *Ocean Engineering*, 97:175–185.
- Obakrim, S., Ailliot, P., Monbet, V., and Raillard, N. (2022). Statistical modeling of the space-time relation between wind and significant wave height. *Earth and Space Science Open Archive*, page 20.
- Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

- Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., and Schmid, M. (2019). A review of spline function procedures in r. *BMC Medical Research Methodology*, 19.
- Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session*, page 859–878.
- Pickands, III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119 – 131.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramos, A. (2003). *Multivariate joint tail modelling and score tests of independence*. PhD thesis, Univ. of Surrey, (<https://openresearch.surrey.ac.uk/esploro/outputs/99515772002346>).
- Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):219–241.
- Randell, D., Turnbull, K., Ewans, K., and Jonathan, P. (2016). Bayesian inference for nonstationary marginal extremes. *Environmetrics*, 27(7):439–450.
- Raymond-Belzile, L. (2019). *Contributions to Likelihood-Based Modelling of Extreme Values*. PhD thesis, EPFL.
- Reiss, R. and Thomas, M. (2007). *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser Basel.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY.
- Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336.
- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer-Verlag New York.
- Rivoire, P., Martius, O., and Naveau, P. (2021). A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. *Earth and Space Science*, 8:e2020EA001633.
- Rohmer, J., Idier, D., Thieblemont, R., Le Cozannet, G., and Bachoc, F. (2021). Partitioning the uncertainty contributions of dependent offshore forcing conditions in the probabilistic assessment of future coastal flooding at a macrotidal site. *Natural Hazards and Earth System Sciences Discussions*, 2021:1–30.

- Rohmer, J., Louisor, J., Le Cozannet, G., Naveau, P., Thao, S., and Bertin, X. (2020). Attribution of Extreme Wave Height Records along the North Atlantic Coasts using Hindcast Data: Feasibility and Limitations. *Journal of Coastal Research*, 95(SI):1268–1272.
- Rootzén, H., Segers, J., and Wadsworth, J. L. (2018a). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.
- Rootzén, H., Segers, J., and Wadsworth, J. L. (2018b). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.
- Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917 – 930.
- Ross, E., Randell, D., Ewans, K., Feld, G., and Jonathan, P. (2017). Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. *Ocean Engineering*, 142:315–328.
- Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M. (2010). The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1058.
- Schaefer, J. T. (1990). The critical success index as an indicator of warning skill. *Weather and Forecasting*, 5(4):570 – 575.
- Segers, J. (2012). Max-stable models for multivariate extremes, *revstat-stat. J*, 10:61–82.
- Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., D. C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., and Wehner, M. (2021). Weather and climate extreme events in a changing climate. In Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.



- Shooter, R., Ross, E., Tawn, J. A., and Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, 30(6).
- Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.
- Sklar, M. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.
- Stasinopoulos, M., Rigby, B., and Akantziliotou, C. (2008). *Instructions on how to use the gamlss package in R Second Edition*.
- Stephenson, A. G. (2002). evd: Extreme value distributions. *R News*, 2(2):0.
- Stephenson, D. B., Casati, B., Ferro, C. A. T., and Wilson, C. A. (2008). The extreme dependency score: a non-vanishing measure for forecasts of rare events. *Meteorological Applications*, 15(1):41–50.
- Stopa, J. E., Filipot, J.-F., Li, N., Cheung, K. F., Chen, Y.-L., and Vega, L. (2013). Wave energy resources along the Hawaiian Island chain. *Renewable Energy*, 55:305–321.
- Taillardat, M., Fougères, A.-L., Naveau, P., and de Fondeville, R. (2019). Extreme events evaluation using CRPS distributions. arXiv:1905.04022.
- Tajvidi, N. (1996). *Characterisation and some statistical aspects of univariate and multivariate generalised Pareto distributions*. PhD thesis, Chalmers University of Technology.
- Tencaliec, P., Favre, A.-C., Naveau, P., Prieur, C., and Nicolet, G. (2019). Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, 31(2):e2582. e2582 env.2582.
- Tendijck, S., Eastoe, E. F., Tawn, J. A., Randell, D., and Jonathan, P. (2021). Modeling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*, 0(0):1–12.
- Tolman, H., Accensi, M., Alves, J.-H., Ardhuin, F., Bidlot, J., Booij, N., Bennis, A.-C., Campbell, T., Chalikov, D., Chawla, A., Filipot, J.-F., Foreman, M., Janssen, P., Leckler, F., Li, J.-G., Lind, K., Orzech, M., Padilla-Hernandez, R., Rogers, W., and Zieger, S. (2014). *User manual and system documentation of WAVEWATCH III version 4.18*.
- Toulemonde, G., Ribereau, P., and Naveau, P. (2015). Applications of extreme value theory to environmental data analysis. In *Extreme Events: Observations, Modeling, and Economics*, page 7–21. American Geophysical Union Washington DC.

- Towe, R., Eastoe, E., Tawn, J., Wu, Y., and Jonathan, P. (2013). The extremal dependence of storm severity, wind speed and surface level pressure in the northern North Sea. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 55324, page V02AT02A014. American Society of Mechanical Engineers.
- Towe, R., Eastoe, E. F., Tawn, J. A., and Jonathan, P. (2017). Statistical downscaling for future extreme wave heights in the North Sea. *The Annals of Applied Statistics*, 11(4):2375–2403.
- Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York.
- Vanem, E. and Fazeres-Ferradosa, T. (2022). A truncated, translated Weibull distribution for shallow water sea states. *Coastal Engineering*, 172:104077.
- Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):149–175.
- Weller, G. B. and Cooley, D. (2014). A sum characterization of hidden regular variation with likelihood inference via expectation-maximization. *Biometrika*, 101(1):17–36.
- Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.
- Yee, T. W. (2015). *Vector generalized linear and additive models: with an implementation in R*, volume 10. Springer.
- Yee, T. W. and Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, 10(1):1–19.
- Young, I. R. and Ribal, A. (2019). Multiplatform evaluation of global trends in wind speed and wave height. *Science*, 364(6440):548–552.