

# A Non-stationary Dependence Model for Extreme European Windstorms

Author:

**Paul Castelain**

Supervision:

**Prof. Anthony Davison**  
**Dr Raphaël de Fondeville**

Chair of Statistics

EPFL, Swiss Federal Institute of Technology in Lausanne  
Switzerland

**EPFL**

January 2021



## Abstract

European windstorms are major extratropical cyclones that impact Western Europe during winter. Extreme value models can be used to provide policymakers and insurance companies with probabilistic characterization of these meteorological phenomena. In this context, we represent windstorms by their footprints and define a risk functional to identify extreme events on a univariate risk scale. We then use this risk to apply the recently developed functional peaks-over-threshold model to windstorm footprints.

As windstorm behaviour is highly influenced by topology, our dataset displays important non-stationary extremal dependence. We propose to model the windstorm footprints by a generalized  $r$ -Pareto process with non-stationary dependence structure. We achieve this by using locally anisotropic Stochastic Partial Differential Equations to generate non-stationary Matérn-like Gaussian fields used as the underlying process for the generalized  $r$ -Pareto random field.

Different non-stationary models are estimated and then compared using visual analysis and gradient scoring. Our results show that non-stationary models significantly outperform isotropic ones and produce coherent anisotropic vector fields with a high level of confidence. The use of a simulation algorithm for storm footprints is briefly demonstrated and results are compared with empirical data. More flexibility in the sub-asymptotic dependence model is highlighted as the main lead towards future improved models.

**Keywords:** Extreme value analysis, Anisotropy, Functional peaks-over-threshold,  $r$ -Pareto process, Gaussian Markov random fields, SPDE.

# Acknowledgements

Two years and a half ago I was attending my first lecture of Extreme Value Analysis given by Prof. Anthony Davison; it was a “match”, one could say and I have since then never let this topic down. I would like to express my deep gratitude to Prof. Anthony Davison, for his amazing teachings and his considerate expectations, allowing me to surpass myself.

I would also like to express my deepest gratitude to Dr Raphaël de Fondeville for his outstanding moral and academic support; during this semester, he has been the main person I have been working with, and I am glad I had such a talented and mindful supervisor watching over me.

I would like to thank my family and my flatmates for their support during this project and all along my studies. I have special thoughts for my friends from the climate movement to whom I own my mental health today, their presence helped me develop my pessimistic vision of the world prior to the pandemic; this was a key in allowing me to stay strong during this lockdown semester.

Finally, and overall, thank you to Ella-Mona Chevalley, for your unconditional love and support.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Peaks-over-threshold exceedances</b>	<b>5</b>
2.1	Univariate model . . . . .	5
2.2	Functional exceedances . . . . .	5
2.2.1	Functional regular variation . . . . .	6
2.2.2	Generalized $r$ -Pareto process . . . . .	7
2.2.3	Simulation algorithm . . . . .	9
2.3	Statistical inference of functional exceedances . . . . .	10
<b>3</b>	<b>Extreme European Windstorms</b>	<b>12</b>
3.1	Storm and risk definition . . . . .	12
3.2	Data transformation and storm selection . . . . .	13
3.3	Dependence exploratory analysis . . . . .	14
3.4	Tail marginal modelling . . . . .	16
<b>4</b>	<b>Non-stationary extremal dependence modelling</b>	<b>21</b>
4.1	Generalities on Gaussian Markov random fields . . . . .	21
4.2	Parametric models for vector fields . . . . .	22
4.2.1	Polynomial . . . . .	22
4.2.2	Bidimensional interpolation . . . . .	23
4.3	Inference and implementation . . . . .	25
4.3.1	Score estimation . . . . .	25
4.3.2	Composite inference . . . . .	26
4.3.3	Computational considerations on the SPDE method . . . . .	26
4.3.4	Optimization algorithm . . . . .	27
<b>5</b>	<b>Estimated dependence models</b>	<b>29</b>
5.1	Extremal dependence . . . . .	31
5.2	Example of simulated footprints . . . . .	33
<b>6</b>	<b>Discussion</b>	<b>35</b>
<b>A</b>	<b>Detailed dependence results</b>	<b>36</b>

# 1 Introduction

On the morning of January 25, 1990, Cyclone Daria was hitting the coast of Ireland; the storm would then go through United Kingdom, France, Belgium and the Netherlands before finishing its journey in Denmark. Daria is one of the strongest European windstorms on record; it caused 95 deaths and resulted in an approximated EUR 6 billion worth of damage. European windstorms are powerful extratropical cyclones forming in the Atlantic Ocean and hitting western Europe, often between October and March. They represent a high threat to people’s lives and public infrastructure, so their characterization is of major interest for appropriate infrastructure design and insurance forecasting.

In this project we propose a model to describe the behaviour of these windstorms using extreme value theory, more specifically the recently developed functional peaks-over-threshold approach (Dombry & Ribatet, 2015). We build upon de Fondeville & Davison (2020), who investigated the same dataset, but we introduce non-stationary dependence structure. Our model uses storm footprints, a common methodology in storm characterization (Roberts et al., 2014, Allen, 2017) that involves taking local wind gust maxima and plotting them on a heatmap, thus highlighting the path taken by the storm. Figure 1.1 illustrates the footprint process for Daria.

The study of tail behaviour of random variables is the core of extreme value theory. Extreme events were initially studied using maxima on a regular time frame, e.g., monthly maxima, whose distribution can be approximated by so-called max-stable distributions. The domain of attraction of max-stable distributions, i.e., the domain of random variables whose maxima converge to these distributions, was shown to be wide thus allowing for use of this theory in a large range of applications, e.g., insurance, hydrology or climatology (Beirlant et al., 1994, Katz et al., 2002, Ni et al., 2020).

Another approach to univariate extreme values is the peaks-over-threshold (POT) method, which defines extreme events as events exceeding some threshold — often defined to be a high quantile. This approach is often preferred as it allows for the extraction and use of all the events in the tail distribution contained in the dataset, whereas in max-stable inference observations that are not the maxima of their time frame are discarded.

Studying extremes in the spatial domain requires the extension of the notion of extremes to the functional domain. To this end, recent developments (Dombry & Ribatet, 2015, de Fondeville & Davison, 2018) on the notion of risk functionals are of much interest. Risk functionals are functions from the sample path domain to  $\mathbb{R}$ ; they allow for extreme events to be defined on the univariate risk scale in  $\mathbb{R}$ . The main theory used in this project is from de Fondeville & Davison (2020), where risk functionals are used to define a new process called the generalized  $r$ -Pareto process. This spatial process is separable into a risk part, responsible for the intensity of events, and an angular part, responsible for the spatial path. More details on those notions are given in the coming chapters. One limitation to models implemented thus far is the stationarity of the extremal dependence, i.e., the probability of an event being extreme, at two points on the domain at the same time, only depends on the distance between those two points.

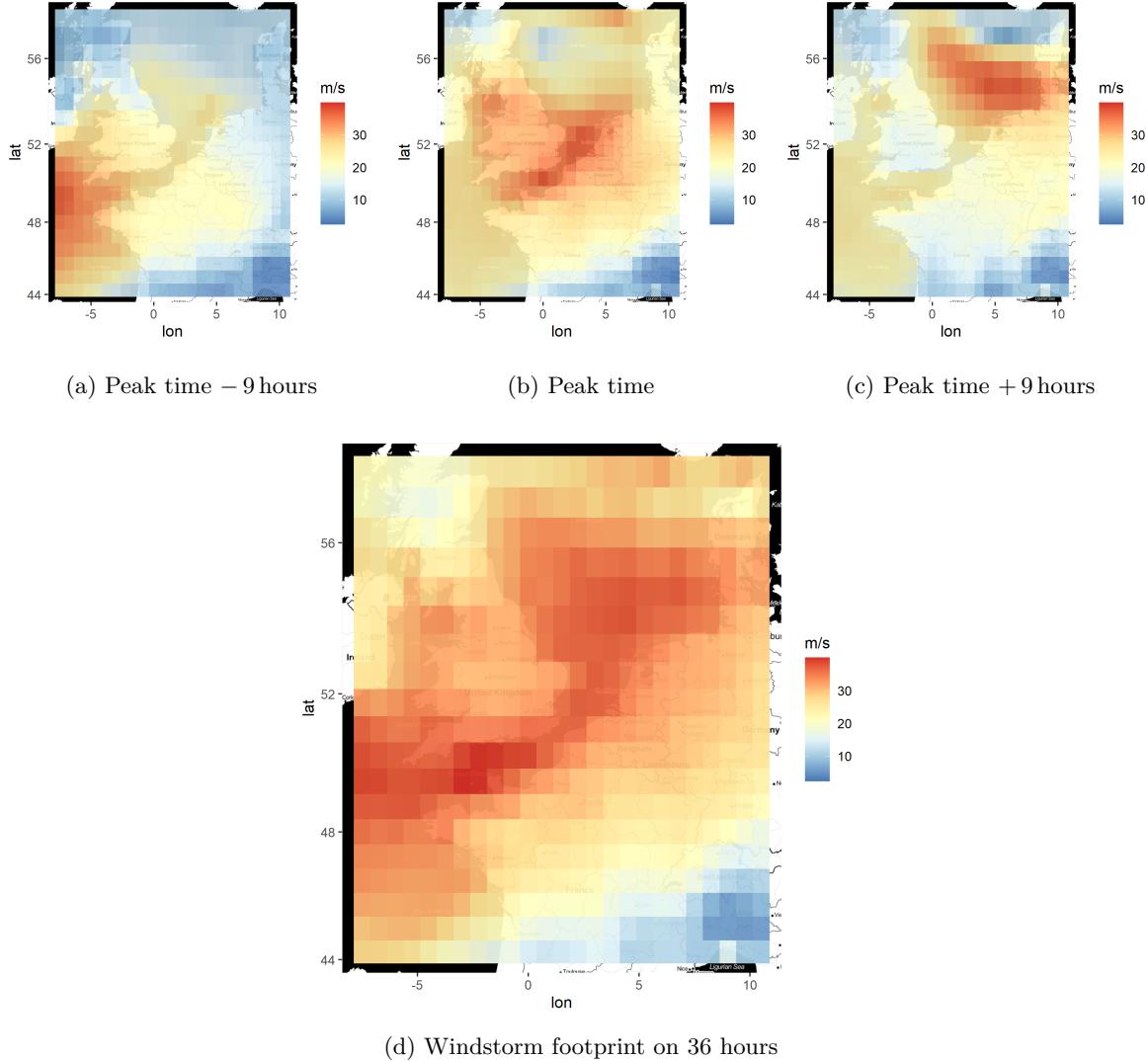


Figure 1.1: Wind gust speeds in  $\text{m s}^{-1}$  during windstorm Daria at different times relative to the peak risk time (top). The resulting 36-hour footprint of Daria (bottom).

A practical example of the limitation of this assumption in our application is the English Channel; windstorms tend to follow the Channel rather than crossing it. If the wind is extremely high at one location  $A$  in the Channel, the probability of another point  $B$  at distance  $d$  having extreme winds at the same time is much higher if  $B$  is along the Channel than if  $B$  is inland.

In de Fondeville & Davison (2020), the restriction to stationary dependence is due to stationary Gaussian fields being used as underlying processes in the angular part of the  $r$ -Pareto variable. To account for non-stationary dependence we use a more complex method allowing for the generation of non-stationary Gaussian fields.

This method is the Stochastic Partial Differential Equation (SPDE) approach, it uses an SPDE to generate complex Gaussian fields. It allows for fast computation of the fields' covariance matrix using sparsity in the precision matrix (Lindgren et al., 2011). The drawback to this method is the reduced interpretability on the covariance parameters of the generated field. Its use has, however, been growing rapidly in recent years with the development of the R-INLA package (Rue et al., 2009) allowing for fast and technicality-free Bayesian inference on stationary Matérn Gaussian fields. In this project we use an

extension of the stationary field SPDE method to allow non-stationarity; developed in Fuglstad et al. (2015), this extension uses locally anisotropic SPDEs and generates non-stationary Matérn-like Gaussian fields.

This report is structured as follows. Chapter 2 details the theoretical background behind generalized  $r$ -Pareto processes. The European windstorm case is presented in Chapter 3 where exploratory analysis and marginal parameter estimations are performed. Specification of the dependence models is made in Chapter 4 and details are given on the implementation methods and computational considerations. Dependence estimation results are reported and discussed in Chapter 5. Finally, a review of the strengths and limitations of our approach is provided in Chapter 6.

## 2 Peaks-over-threshold exceedances

This chapter is mainly a synthesis of the theoretical notions developed in de Fondeville & Davison (2020).

### 2.1 Univariate model

A random variable  $X$  with distribution function  $F$  is said to belong to the maximum domain of attraction of a non-degenerate distribution function  $G$  if there exist sequences of constants  $\{a_n\}_{n=1}^\infty > 0$  and  $\{b_n\}_{n=1}^\infty$  such that

$$n \Pr \left( \frac{X - b_n}{a_n} > x \right) \rightarrow -\log G(x), \quad n \rightarrow \infty. \quad (2.1)$$

For such an  $X$  and a high threshold  $u$  such that  $u < \inf\{x : F(x) = 1\}$ , the distribution of  $X - u$  conditional on  $X > u$  follows approximately a generalized Pareto distribution,

$$\Pr(X - u > x | X > u) \approx H_{\xi, \sigma}(x) = \begin{cases} (1 + \xi x / \sigma)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x / \sigma), & \xi = 0, \end{cases} \quad (2.2)$$

where  $(\cdot)_+ = \max(\cdot, 0)$  and  $\sigma = \sigma(u) > 0$ . If  $\xi < 0$  then the support of  $H_{\xi, \sigma}$  has an upper bound  $\{x : H_{\xi, \sigma}(x) > 0\} \leq -\sigma / \xi$ , otherwise, if  $\xi \geq 0$ ,  $X - u$  can take any positive value.

The distribution function approximation for  $x$  above  $u$  used in such cases is

$$F(x) \approx 1 - \zeta_u H_{\xi, \sigma}(x - u), \quad x > u, \quad (2.3)$$

where  $\zeta_u = \Pr(X > u)$ .

### 2.2 Functional exceedances

In the univariate case, the definition of a threshold for tail events is straightforward. When approaching multivariate problems, tail event selection is more complex, and to this end, we here develop on a few notions regarding risk functionals and  $r$ -exceedances.

Let  $S \subset \mathbb{R}^D$  ( $D \geq 1$ ), be a compact metric space and let  $\mathcal{F}$  denote the Banach space of real-valued continuous functions on  $S$  associated with the norm  $\|\cdot\|$ . To prevent the appearance of degenerate measures in limiting procedures we define  $\mathcal{F}_+ \subset \mathcal{F}$  to be the set of non-negative functions that are not everywhere zero and write  $\mathcal{F}_0 = \mathcal{F}_+ \cup \{0\}$ .

A risk functional  $r$  is a continuous mapping from  $\mathcal{F}$  to  $\mathbb{R}_+$  and an  $r$ -exceedance event is defined as  $\{r(X) > u\}$ , with  $u$  being the threshold. This approach transforms the notion of exceedance from  $\mathbb{R}^D$  to an univariate risk exceedance.

The definition of risk functionals was introduced by Dombry & Ribatet (2015) under the name “cost functionals” with a homogeneity constraint, i.e., functionals for which there exists a constant  $\kappa > 0$  such that  $r(ay) = a^\kappa r(y)$  with  $y \in \mathcal{F}_+$  and  $a > 0$ .

In the following developments, we will only focus on linear risk functionals, as they are simpler to handle and provide the desired risks for this study (de Fondeville & Davison, 2020). Indeed we will focus on risk functionals that average the random field on some domain, such as  $r(X) = \int_A X(s) ds$ .

### 2.2.1 Functional regular variation

To connect univariate peaks-over-threshold theory and the risk functional approach we give a few notions on functional regular variation.

Let  $\xi$  be a real-valued shape parameter, let  $a \in \mathcal{F}_+$  and let  $b \in \mathcal{F}$ . We define the sets of functions

$$\mathcal{F}^{\xi,a,b} = \begin{cases} \mathcal{F}_+ - \{b - \xi^{-1}a\}, & \xi > 0, \\ \mathcal{F}, & \xi = 0, \\ -\mathcal{F}_+ + \{b - \xi^{-1}a\}, & \xi < 0, \end{cases} \quad (2.4)$$

i.e., the shift of  $\mathcal{F}_+$  by  $-\{b - \xi^{-1}a\}$  and its reflection if  $\xi < 0$ .

Given functions  $a > 0$  and  $b$ , consider a risk functional  $r : \mathcal{F}^{\xi,a,b} \rightarrow \mathbb{R}$  then the range of  $r$  is  $0 \leq r(\cdot) < \infty$  if  $\xi \geq 0$  and  $0 \leq r(\cdot) < r(b) - \xi^{-1}r(a)$  if  $\xi < 0$ . We write  $\mathcal{U}_r^\xi(u)$  for the restriction of the range of  $r$  above a threshold  $u \geq 0$ , i.e.,

$$\mathcal{U}_r^\xi(u) = \begin{cases} [u, \infty), & \xi \geq 0, \\ [u, r(b) - \xi^{-1}r(a)), & \xi < 0. \end{cases} \quad (2.5)$$

A stochastic process  $X$  taking sample paths in  $\mathcal{F}$  is said to be regularly varying if there exist  $\xi > 0$ , sequences of functions in  $\mathcal{F}$ ,  $\{a_n\}_{n=1}^\infty > 0$ ,  $\{b_n\}_{n=1}^\infty$  and a measure  $\Lambda$  on  $\mathcal{F}_+$  such that

$$\left. \begin{aligned} n \Pr \left[ \left\{ 1 + \xi \left( \frac{X - b_n}{a_n} \right) \right\}_+^{1/\xi} \in \cdot \right], \quad \xi \neq 0 \\ n \Pr \left\{ \exp \left( \frac{X - b_n}{a_n} \right) \in \cdot \right\}, \quad \xi = 0 \end{aligned} \right\} \rightarrow \Lambda(\cdot), \quad n \rightarrow \infty, \quad (2.6)$$

where  $\{\cdot\}_+ = \max(\cdot, 0)$  with the maxima taken pointwise for each  $s \in S$  and  $a_n, b_n$  chosen such that

$$\lim_{n \rightarrow \infty} n \Pr \left\{ \frac{X(s) - b_n(s)}{a_n(s)} > x \right\} = \begin{cases} (1 + \xi x)_+^{-1/\xi}, & 1 + \xi x > 0, \xi \neq 0, \\ \exp(-x), & x > 0, \xi = 0. \end{cases} \quad (2.7)$$

If  $X$  is regularly varying, we write  $X \in \text{GRV}(\xi, a_n, b_n, \Lambda)$ .

Equations (2.6) and (2.7) are the natural functional equivalents of the univariate convergence to a non-degenerate measure introduced in (2.1). While allowing for a pointwise rescaling with  $a_n(s)$  and  $b_n(s)$ , the tail parameter  $\xi$  is fixed for the whole domain  $S$ . As the risk will be computed directly on  $X$  without reshaping, a stochastic process with varying tail index  $\xi(s)$  on  $S$  would induce the limiting distribution of the risk to be drawn to locations with the heaviest tail and this is to be avoided.

The main constraint on the extension to functionals with GRV is the restriction in (2.6) of the measure  $\Lambda$  to the domain of positive continuous functions, as it excludes asymptotic independence for

the process  $X$ . For an asymptotically independent process  $X'$  to converge in (2.6), it would require  $\Lambda$  to have zero weight on the set of continuous functions and convergence for (2.6) in this case has not yet been extensively studied. Our model is thus limited to asymptotically dependent process.

In environmental science when focusing on a specific phenomenon the stationarity of  $\xi$  is not a strong hypothesis, as the underlying tail behaviour is induced everywhere by the same physical phenomenon thus  $\xi$  should not be different from site to site. Furthermore, the use of  $r$ -functionals to select the events studied, should allow us to focus on events characteristic of a certain underlying physical process, thus making the stationary  $\xi$  a realistic assumption. The asymptotic dependence assumption, however, is less straightforward to justify, as extreme data in environmental applications often display asymptotic independence; this point is discussed in different chapters later in this report.

### 2.2.2 Generalized $r$ -Pareto process

Let us suppose that for a linear risk functional  $r$ , there exists  $A(s) \in \mathcal{F}_+$  such that the sequence  $\{a_n\}_{n=1}^\infty$  satisfies

$$\lim_{n \rightarrow \infty} \sup_{s \in S} \left| \frac{a_n(s)}{r(a_n)} - A(s) \right| = 0; \quad (2.8)$$

this allows us to write  $a_n(s) \approx r(a_n)A(s)$ . This assumption is also used in Ferreira et al. (2012), Engelke et al. (2019) and de Fondeville & Davison (2020); it seems reasonable in most applications. The assumptions made yet,  $\xi$  constant and the decomposition of  $a_n$ , are similar to the fairly general assumption that the marginal distribution of  $X(s)$  belongs to a location-scale family,

$$\Pr\{X(s) \leq x\} = F\left\{\frac{x - B(s)}{A(s)}\right\}, \quad s \in S.$$

This assumption would also imply  $\xi$  being constant over the domain and induce the existence of a similar decomposition for  $a_n$ .

**Theorem 1.** *Let  $X$  be a stochastic process with sample paths in  $\mathcal{F}$  and let  $r$  be a linear risk functional if  $\xi \neq 0$  and an evaluation functional if  $\xi = 0$ , i.e.,  $r(x) = x(s)$  for some  $s \in S$ . If  $X \in \text{GRV}(\xi, a_n, b_n, \Lambda)$  and the functions  $a_n(s)$  satisfy (2.8), then*

$$\Pr\left\{\left\lfloor \frac{X - b_n}{r(a_n)} \right\rfloor \in \cdot \mid r(X) \geq u_n\right\} \rightarrow \Pr\{P \in \cdot\}, \quad n \rightarrow \infty, \quad (2.9)$$

where  $u_n = r(a_n)u + r(b_n)$ ,  $u \in \mathcal{U}_r^\xi(0)$ , and  $P$  denotes a generalized  $r$ -Pareto process with tail index  $\xi$ , scale function  $A$ , zero location and measure  $\Lambda$ . When  $\xi > 0$ , we define  $[x] = \max\{x, -\xi^{-1}A\}$  with the maxima taken pointwise and  $[x] = x$  when  $\xi \leq 0$ , for  $x \in \mathcal{F}$ .

The generalized  $r$ -Pareto process appears as the limit of a rescaled regularly varying process  $X$  conditionally on  $r$ -exceedance over an increasingly high threshold  $u_n$ .

Before stating the generalized  $r$ -Pareto definition, we define two useful sets. For a linear risk functional  $r$ , a tail index  $\xi \in \mathbb{R}$ , a positive function  $a \in \mathcal{F}_+$  and  $A = a/r(a)$ , the set

$$\mathcal{A}_r = \begin{cases} \{y \in \mathcal{F}_+ : r(Ay^\xi) \geq 1\}, & \xi > 0, \\ \{y \in \mathcal{F}_+ : r(A \log y) \geq 0\}, & \xi = 0, \\ \{y \in \mathcal{F}_+ : r(Ay^\xi) \leq 1\}, & \xi < 0, \end{cases} \quad (2.10)$$

is important in the definition of the  $r$ -Pareto process  $P$ ; it is the sample space of the process  $Y_r$ , a rescaled version of  $P$ . We also define the restriction of  $\mathcal{A}_r$  to functions with unit norm and denote it by  $\mathcal{S}_r^{\xi,A}$ ; it is used as a set of angular functions to decompose  $Y_r$ ;

$$\mathcal{S}_r^{\xi,A} = \begin{cases} \{y \in \mathcal{F}_+ : r(Ay^\xi) \geq 1, \|y\| = 1\}, & \xi > 0, \\ \{y \in \mathcal{F}_+ : r(A \log y) = 0\}, & \xi = 0, \\ \{y \in \mathcal{F}_+ : r(Ay^\xi) \leq 1, \|y\| = 1\}, & \xi < 0. \end{cases} \quad (2.11)$$

**Definition 1.** Let  $a > 0$  and  $b$  be continuous functions on  $S$ , let  $r : \mathcal{F}^{\xi,a,b} \rightarrow \mathcal{U}_r^\xi$  be a linear risk functional and let  $\Lambda$  be a  $(-1)$ -homogeneous measure on  $\mathcal{F}_+$ . The generalized  $r$ -Pareto process  $P$  associated to the measure  $\Lambda$  and tail index  $\xi$  is the stochastic process on  $\{x \in \mathcal{F}^{\xi,a,b} : r(x) \geq r(b)\}$  defined by

$$P = \begin{cases} a(Y_r^\xi - 1)/\xi + b, & \xi \neq 0, \\ a \log Y_r + b, & \xi = 0, \end{cases} \quad (2.12)$$

where  $Y_r$  is the stochastic process on  $\mathcal{A}_r$  with probability measure  $\Lambda(\cdot)/\Lambda(\mathcal{A}_r)$ .

In Definition 1,  $Y_r$ , the process of reference inducing  $P$ , has marginal unit tail index; it contains the dependence structure of  $P$  but not the marginal behaviour.

A pseudo-polar decomposition of the reference process  $Y_r$  exists (Dombry & Ribatet, 2015, de Fondeville & Davison, 2018)

$$Y_r = R_1 \frac{W_1}{r(W_1)}, \quad (2.13)$$

where  $R_1$  and  $W_1$  are independent, the scalar  $R_1$  is a unit Pareto random variable and  $W_1$  is a stochastic process with state space  $S$ , paths in  $\mathcal{S} = \{y \in \mathcal{F}_+ : \|y\|_1 = 1\}$  and probability measure

$$\sigma_0(\cdot) = \frac{\Lambda\{y \in \mathcal{A}_r : y/\|y\|_1 \in \cdot\}}{\Lambda(\mathcal{A}_r)}, \quad (2.14)$$

where  $\|\cdot\|_1$  is the 1-norm on  $\mathcal{F}_+$ .

Finally, the  $r$ -Pareto process  $P$  has nice marginal properties, as it behaves according to a generalized Pareto distribution. Let  $s_0 \in S$ , then providing  $u_0 \geq 0$  a threshold sufficiently high such that

$$\{x \in \mathcal{F}^{\xi,a,b} : x(s_0) > u_0\} \subset \{x \in \mathcal{F}^{\xi,a,b} : r(x) \geq r(b)\}, \quad (2.15)$$

then  $P$  conditionally on exceeding threshold  $u_0$  behaves as

$$\Pr\{P(s_0) > x \mid P(s_0) > u_0\} = \left\{ 1 + \xi \frac{x - u_0}{\sigma(s_0)} \right\}_+^{-1/\xi}, \quad x \geq u_0, \quad (2.16)$$

where  $\sigma(s_0) = r(a)A(s_0) + \xi\{u_0 - b(s_0)\}$ . The distribution of  $r(P) - r(b)$  also follows a generalized Pareto distribution with shape  $\xi$  and scale  $r(a)$  (de Fondeville & Davison, 2020).

The decomposition of  $P$  with  $R_1$  and  $W_1$ , presented in Definition 1 and (2.13), highlights the independence between the intensity of the process  $P$  and its spatial dependence structure. The spatial part of the process, corresponding to angular random variable  $W_1$ , is independent of the intensity of the process, modelled by  $R_1$ . In our model, the spatial dependence pattern is thus the same whatever

$r(P)$  is. This modelling assumption is often not true, especially in environmental applications, where dependence often decreases at high intensities. This implies that in some applications our model will tend to overestimate the dependence and thus the risk at high intensities. While an appropriate model with exact estimates is naturally always preferred, in a risk approach the overestimation drawback of our model might not be a major issue, as overestimating risk leads to safer policies than its underestimation.

### 2.2.3 Simulation algorithm

In the case of linear risk functionals de Fondeville & Davison (2020) provide another definition of  $r$ -Pareto processes allowing for more sophisticated simulations.

**Definition 2.** Let  $\xi$  be a tail index, let  $a(s) > 0$  and  $b(s)$  be continuous functions on  $S$ , let  $r : \mathcal{F}^{\xi,a,b} \rightarrow \mathcal{U}_r^\xi$  be a linear risk functional, let  $\Lambda$  be a  $(-1)$ -homogeneous measure on  $\mathcal{F}_+$  and let  $A = a/r(a)$ . When  $\xi = 0$ , we also suppose that  $r$  is an evaluation functional. The generalized  $r$ -Pareto process  $P$  associated to the measure  $\Lambda$  and tail index  $\xi$  is the stochastic process on  $\{x \in \mathcal{F}^{\xi,a,b} : r(x) \geq r(b)\}$  defined by

$$P = \begin{cases} \frac{r(a)}{\xi} R_2^\xi \frac{W_{\xi,A}}{r(W_{\xi,A})} + b - \xi^{-1}a, & \xi \neq 0, \\ r(a) \log(R_2 W_{\xi,A}) + b, & \xi = 0, \end{cases} \quad (2.17)$$

where  $R_2$  is a scalar unit Pareto random variable independent of  $W_{\xi,A}$ , a stochastic process with state space  $S$  that takes values in  $\mathcal{S}_r^{\xi,A}$  and has probability measure

$$\sigma_r(\cdot) = \begin{cases} \frac{\Lambda\{y \in \mathcal{A}_r : Ay^\xi / \|Ay^\xi\| \in \cdot\}}{\Lambda(\mathcal{A}_r)}, & \xi \neq 0, \\ \frac{\Lambda\{y \in \mathcal{A}_r : A \log y - r(A \log y) \in \cdot\}}{\Lambda(\mathcal{A}_r)}, & \xi = 0. \end{cases} \quad (2.18)$$

Instead of  $Y_r$  this definition uses  $R_2$  and  $W_{\xi,A}$  to explicitly highlight the intensity random variable  $R_2$ . The angular process  $W_{\xi,A}$  is closely tied to  $Y_r$  and can be seen as an  $A$  rescaling and  $\xi$  reshaping of  $Y_r$ . The link is

$$W_{\xi,A} = \begin{cases} \frac{AY_r^\xi}{\|AY_r^\xi\|}, & \xi \neq 0, \\ \exp\{A \log Y_r - r(A \log Y_r)\}, & \xi = 0. \end{cases} \quad (2.19)$$

This more detailed decomposition of  $P$  allows the use of Algorithm 1 to generate events of a specific return level.

Provided a threshold  $u > 0$  such that

$$\mathcal{A}_r \subset \{y \in \mathcal{F}_+ : \|y\|_1 \geq u\}, \quad (2.20)$$

then Algorithm 1 allows the simulation of  $P$ ; a simulation algorithm must, however, be available for  $W_1$ . The efficiency of the algorithm is mainly determined by the threshold  $u$ , the closer the set on the right-hand side of (2.20) is to  $\mathcal{A}_r$  the more efficient the algorithm is.

A key feature of this algorithm is allowing for the simulation of  $P$  with specified risk return level. This is made possible by replacing, in Algorithm 1,  $R_2$  by the desired quantile of a unit Pareto distribution.

---

**Algorithm 1:** Simulation of generalized  $r$ -Pareto process  $P$ , for a linear risk functional  $r$ 


---

```

Set  $Y_r = 0$  ;
while  $Y_r \notin \mathcal{A}_r$  do
    generate a unit Pareto random variable  $R_1$ ;
    generate  $W_1$  with probability measure  $\sigma_0$  defined in (2.14) ;
    set  $Y_r = R_1 W_1 / u$ ;
end
generate a unit Pareto random variable  $R_2$ ;
if  $\xi \neq 0$  then
    set  $W_{\xi,A} = A Y_r^\xi / \|AY_r^\xi\|$ ;
    return  $P = r(a)\xi^{-1}R_2^\xi W_{\xi,A}/r(W_{\xi,A}) + b - \xi^{-1}a$ ;;
else if  $\xi = 0$  then
    set  $W_{\xi,A} = \exp\{A \log Y_r - r(A \log Y_r)\}$ ;
    return  $P = r(a) \log(R_2 W_{\xi,A}) + b$ ;
end

```

---

## 2.3 Statistical inference of functional exceedances

Inference for  $r$ -exceedances of a stochastic process  $X \in \text{GRV}(\xi, a_n, b_n, \Lambda)$  is based on the approximation

$$\begin{aligned} \Pr(X \in \mathcal{R}) &= \Pr\{r(X) \geq u_n\} \times \Pr\{X \in \mathcal{R} \mid r(X) \geq u_n\} \\ &\approx \Pr[r(X) \geq u_n] \times \Pr(P \in \mathcal{R}), \end{aligned} \quad (2.21)$$

where  $\mathcal{R} \subset \mathcal{R}_r(u_n) = \{x \in \mathcal{F}^{\xi, a_n, b_n} : r(x) \geq u_n\}$ ,  $u_n = r(b_n) + ur(a_n)$  for some  $u \geq 0$  is a high quantile of  $r(X)$ .

For a set of independent realizations  $x_j \in \mathbb{R}_+^L (j = 1, \dots, n)$  of a regularly varying process  $X$  observed at locations  $s_1, \dots, s_L \in S$ , the log-likelihood function for the model described in (2.21) is

$$\mathcal{L}_{\text{Thres}}(\theta) = \log \Pr(N_{u_n} = n_{u_n}) + \sum_{j \in K_{u_n}} \log f^r(x_j, \theta), \quad (2.22)$$

where  $N_{u_n}$  is the random number of  $r$ -exceedances,  $K_{u_n} = \{j \in 1, \dots, n : r(x_j) \geq u_n\}$  is the set of  $n_{u_n}$   $r$ -exceedances over  $u_n$ ,  $f^r$  is the finite-dimensional density function of a generalized  $r$ -Pareto process at locations  $s_1, \dots, s_L \in S$ , and equals

$$\begin{aligned} f^r(x, \theta) &= \frac{\lambda_\theta \left( [1 + \xi\{x(s_{1:L}) - b_n(s_{1:L})\}/a_n(s_{1:L})]_+^{1/\xi} \right)}{\Lambda_\theta(\mathcal{A}_r)} \\ &\times \prod_{l=1}^L a_n(s_l)^{-1} \left\{ 1 + \xi \frac{x(s_l) - b_n(s_l)}{a_n(s_l)} \right\}_+^{1/\xi-1}, \quad x \in \mathcal{R}_r(u_n), \end{aligned} \quad (2.23)$$

where  $x(s_{1:L}) = x(s_1), \dots, x(s_L)$ ,  $\theta \in \Theta \subset \mathbb{R}^{d_\Theta}$  is the parametrization of the measure  $\Lambda_\theta$  within the parameter domain  $\Theta$  and the  $L$ -dimensional intensity  $\lambda_\theta$  is defined by

$$\Lambda_\theta\{\mathcal{A}_{\max}(z)\} = \int_{\mathbb{R}^L \setminus (0, z]^L} \lambda_\theta(y) dy$$

with  $\mathcal{A}_{\max}(z) = \{y \in \mathcal{A}_r : \max_{l=1,\dots,L} y(s_l)/z(s_l) \geq 1\}$  (Wadsworth & Tawn, 2014).

Joint inference on dependence parameter  $\theta$  and marginal parameters  $\xi, a_n, b_n$  can be difficult, convergence tends to be slow, and this leads to poor confidence in estimation results. To circumvent this we proceed in two steps, we first estimate the marginal parameters and then proceed with dependence estimation.

The marginal estimation is performed by maximizing the independence log-likelihood

$$\ell_{\text{indep}}(\xi, a_n, b_n) = \sum_{j=1}^n \sum_{l=1}^L \mathbb{1}\{x_j(s_l) \geq b_n(s_l)\} \log \left[ \frac{1}{a_n(s_l)} \left\{ 1 + \xi \frac{x_j(s_l) - b_n(s_l)}{a_n(s_l)} \right\}^{-1/\xi-1} \right], \quad (2.24)$$

which is equivalent to fitting marginal GPD distribution locally under the constraint of the shape parameter  $\xi$  being common to all locations.

### 3 Extreme European Windstorms

As introduced earlier, our application focuses on extreme windstorms hitting Western Europe from the Atlantic Ocean and aims to provide return levels and dependence patterns for these events. The dataset we use is composed of maximum wind gusts extracted from the ERA-Interim reanalysis (Dee et al., 2011), where maximum wind gust speed is defined for each time and spatial point as the maximum speed of wind gusts sustained for at least 3 seconds at that point.

The area of interest has boundaries  $57.75^{\circ}\text{N}$ ,  $44.25^{\circ}\text{S}$ ,  $-8^{\circ}\text{E}$ ,  $10.5^{\circ}\text{W}$ . It includes densely populated areas, four important capitals (London, Paris, Brussels and Amsterdam) and is divided by the English Channel. The time and spatial resolutions are respectively 3 hours and  $0.75^{\circ}$  longitude-latitude and the data span from October 1979 until March 2014. A map of the domain of study (in blue) is in Figure 3.1.

This dataset was used in de Fondeville & Davison (2020), who implemented a spatiotemporal dependence structure, we re-use it with a space-only dependence approach but allowing for non-stationarity.

#### 3.1 Storm and risk definition

Working within the  $r$ -functional framework, extreme events, that we will call storms, are defined using the risk functional

$$r(X) = \frac{1}{|E_r|} \int_{E_r} X(s) ds, \quad (3.1)$$

for  $X \in A_r$  a functional and with  $E_r$  being the risk area of interest (in red in Figure 3.1). The risk functional and the region  $E_r$  are reused from de Fondeville & Davison (2020). This functional acts as an average on the risk region  $E_r$ .

Storms are defined as wind events with intense risk functional values; they are centred on a peak time and are considered to spread out over 36 hours, 18 hours before and 18 hours after the peak risk time. This approach is similar to the methodology used in the European Extreme Wind Storms (XWS) catalogue but we use 36 hours when XWS uses 72 hours. This reduction in the timespan definition of storms is motivated by the smaller size of our region of interest compared to that used by XWS.

Storms are chosen using the risk functional by selecting the times in the dataset with the highest risk. These times are considered as the “peak” of the storm and storms are then de-clustered by ensuring at least 36 hours between two peak risks.

The number of storm events selected has been tuned to 35 and stands approximately for one storm event per year, as we have 34.5 years of records in our dataset. Risk and marginal qq-plots were used to choose the appropriate number of storms. We initially worked with 70 storms before raising the threshold to 35 to improve the risk and the marginal fits; more details on marginal parameter qq-plot assessments are given later in this chapter.

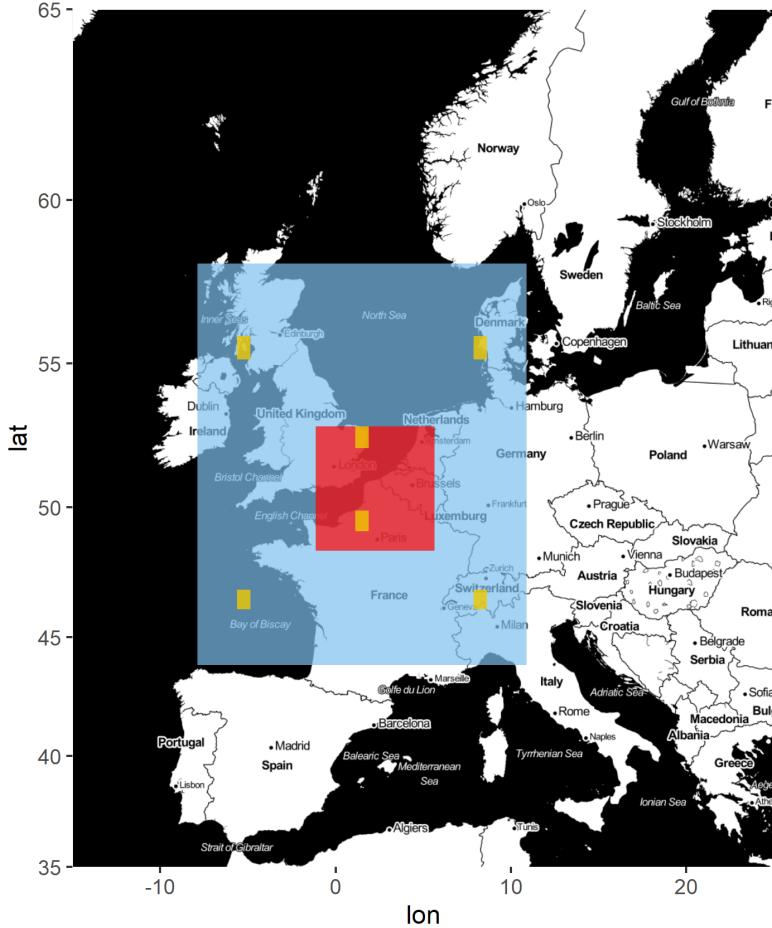


Figure 3.1: Area of study in blue with the risk region in red. Sampling points used in model assessment are indicated in yellow.

From now on, we will only focus on the selected 35 storms, so we let  $n$  denotes the number of storms selected instead of the previous notation  $n_{u_n}$ , for the sake of simplicity.

## 3.2 Data transformation and storm selection

Our model is a spatial extreme model only and is not suited to account for the time evolution of extreme events. As each storm happens during several time frames, the data are preprocessed into storm footprints representing local maxima over a 36-hour period. This methodology is largely based on the one used by XWS but is less sophisticated, as in XWS the spatial centre of the storm is tracked and only spatial locations nearby are included in the storm footprint. In our case a storm footprint always contains all locations in the study domain. Wind footprints  $x_i(s_l)$  are created as

$$x_i(s_l) = \max_{|t' - t_i| < 18} W_i(s_l, t'),$$

$W_i$  standing for the “instantaneous” wind data,  $s_l$  being a location index in the region of interest and  $t_i$ , ( $i = 1, \dots, n$ ) being storm dates.

This newly created  $r$ -exceedance dataset  $x_i(s_l)$ , ( $i = 1, \dots, n$ ,  $l = 1, \dots, L$ ) with  $n = 35$  and  $L = 475$

in our case, comprises the observations we will perform our analysis on. The lowest risk in the selected storms defines the  $r$ -exceedance threshold used; in the 35-storm case, it is  $u_n = 25.88 \text{ m s}^{-1}$ . Note that while the risk functional is exceptionally used, at the very beginning, on instantaneous wind gust data to find the peak time for each storm, the risk functional is only meaningful and supposed to be used with storm footprints.

### 3.3 Dependence exploratory analysis

To quantify the spatial extremal dependence of our process we use the extremal dependence measure (de Fondeville & Davison, 2018)

$$\pi(s', s) = \lim_{q \rightarrow 1} \Pr[X(s') > u_q(s') \mid \{X(s) > u_q(s)\} \cap \{r(X) \geq u\}], \quad u \geq 0, \quad s, s' \in S, \quad (3.2)$$

where  $u_q(s)$  denotes the  $q$  quantile of  $X(s)$ . This extremal dependence coefficient is an extension to the  $r$ -extremal process of the usual extremal pairwise dependence coefficient  $\chi$ .

The empirical version of  $\pi$  is

$$\hat{\pi}_q(s_j, s_l) = \frac{\sum_{i=1}^n 1\{x_i(s_j) \geq u_q(s_j), x_i(s_l) \geq u_q(s_l), r(x_i) \geq u\}}{\sum_{i=1}^n 1\{x_i(s_l) \geq u_q(s_l), r(x_i) \geq u\}}, \quad (3.3)$$

where  $u_q(s_l)$  denotes the empirical quantile of  $x(s_l)$  and  $u$  a risk threshold.

Plots of the empirical  $\hat{\pi}$  are displayed in Figure 3.2. The top left figure shows the extremal dependence summarized for the whole domain  $S$ , a slight south-west to north-east anisotropy is visible but the overall dependence is balanced and is close to being isotropic. On the top right, extremal dependence for the point in grey is plotted on a map; it shows a strong anisotropy along the Channel. These two figures on top means that while dependence is strongly anisotropic at some locations, when summarized at a domain level, no strong pattern is visible. This indicates that the anisotropy direction is changing on the domain and result in the summarized view being balanced.

The bottom figure in Figure 3.2 shows the cloud of  $\hat{\pi}$  as function of the distance between pairs of points. The thickness of the cloud along the ordinate indicates that pairs of points at the same distance can have significantly different dependence coefficient; this gives weight to the dependence structure being non-stationary. The slight upward trend at the end of the cubic spline, in blue, is not mathematically coherent with the behaviour expected for an extremogram, it should be decreasing. It is believed to result from the small number of points in this region of the plot, and the wider confidence region there, in grey, does include the horizontal line.

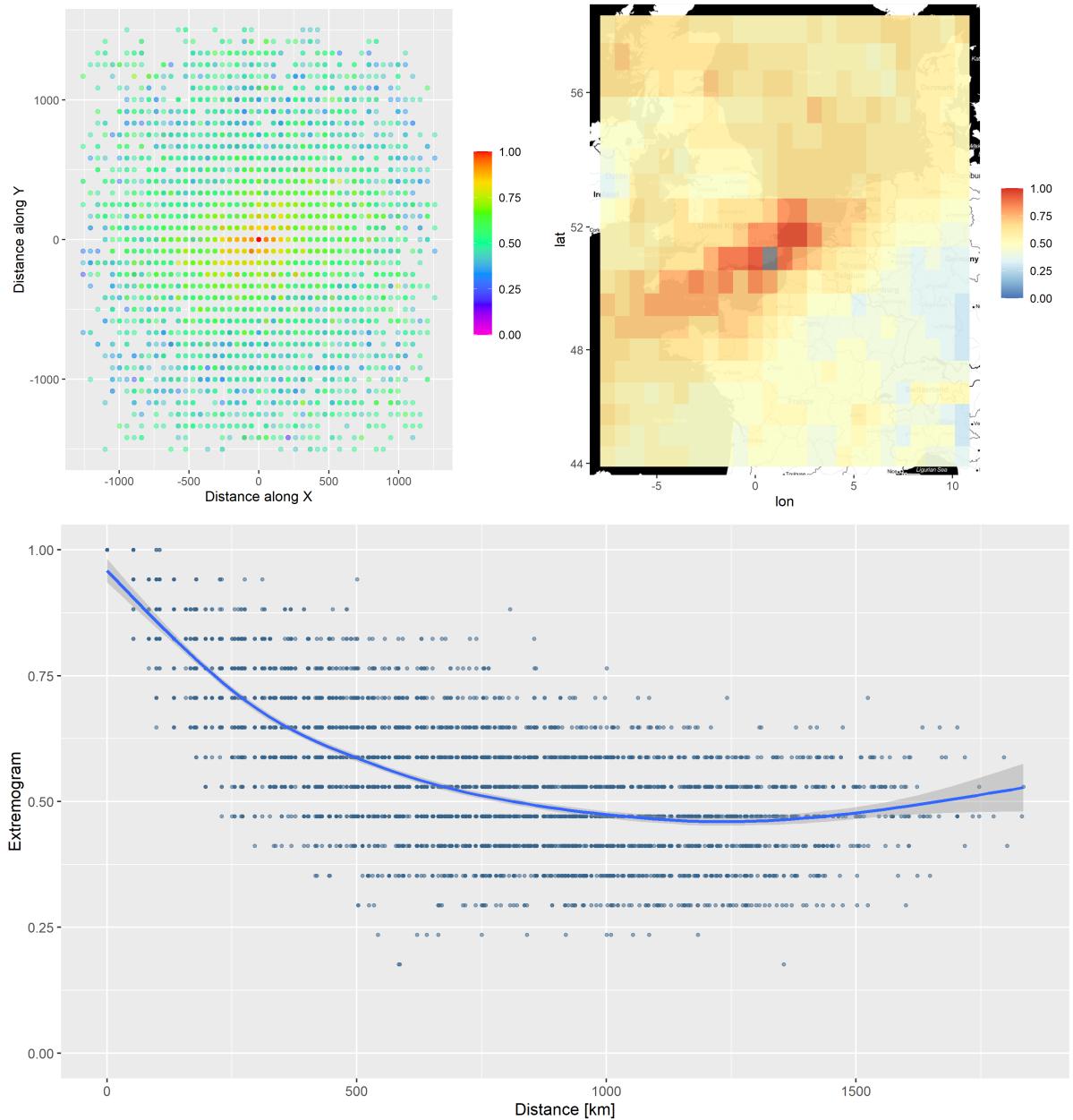


Figure 3.2: Three plots of the empirical extremogram  $\hat{\pi}_{0.52}$ . On top left is a plot of  $\hat{\pi}_{0.52}$  for the whole domain  $S$  as a function of the  $x$  and  $y$  distance between pairs of points. On bottom is a plot of  $\hat{\pi}_{0.52}$  as a function of distances between pairs of points; the blue line is a cubic smooth of the cloud, and the region in grey is its 95% confidence interval. Lastly, on the right, is a heatmap of the extremal coefficient  $\hat{\pi}_{0.52}(s, \cdot)$  for a fixed point  $s$ , in grey on the map.

Using a low quantile  $q$  in  $\hat{\pi}_q$  induces a bias in the results and care should be taken in the interpretation

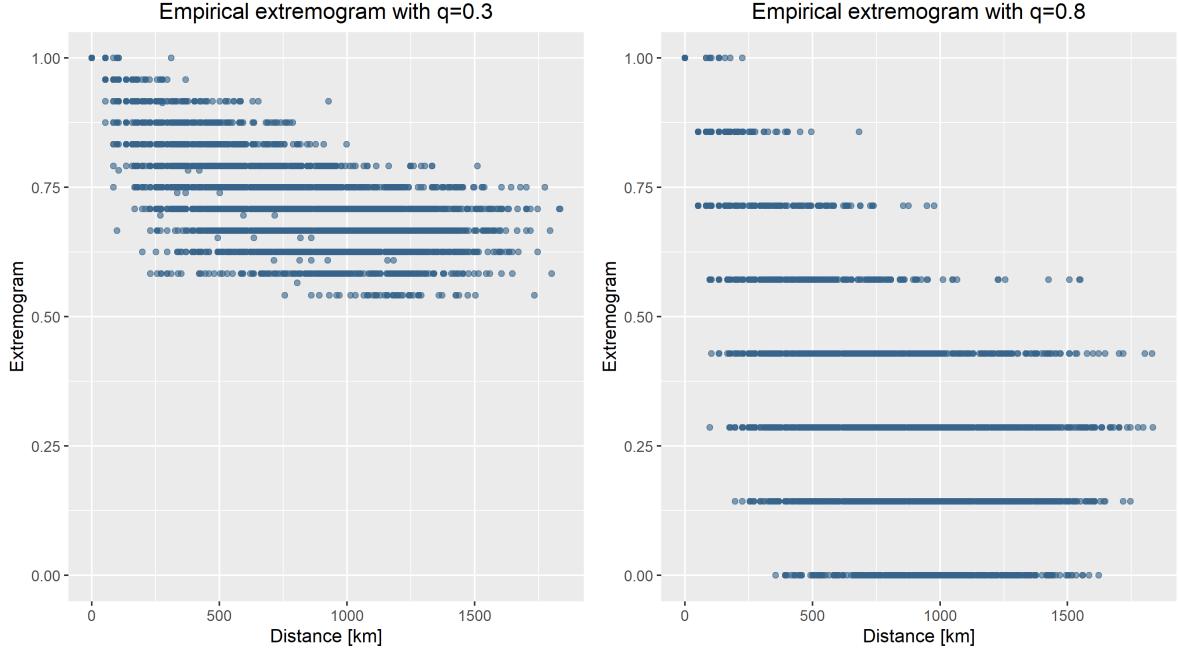


Figure 3.3: Cloud of  $\hat{\pi}_q$ , with  $q = 0.3$  and  $0.8$ , for pairs of points in the domain  $S$  as function of the distance. Illustration of the bias of  $\hat{\pi}$  as presented in (3.4)

of the results, as two locations  $s_j, s'_j$  that are independent will have

$$\begin{aligned}
 E[\hat{\pi}_q(s_{j'}, s_j)] &= \frac{\sum_{i=1}^n \Pr\{x_i(s_{j'}) \geq u_q(s_{j'}), x_i(s_j) \geq u_q(s_j)\}}{\sum_{i=1}^n \Pr\{x_i(s_j) \geq u_q(s_j)\}}, \\
 &= \frac{\sum_{i=1}^n \Pr[\{x_i(s_{j'}) \geq u_q(s_{j'})\} \cap \{x_i(s_j) \geq u_q(s_j)\}]}{\sum_{i=1}^n \Pr\{x_i(s_j) \geq u_q(s_j)\}}, \\
 &= \Pr\{x(s_{j'}) \geq u_q(s_{j'})\} \frac{\sum_{i=1}^n \Pr\{x_i(s_j) \geq u_q(s_j)\}}{\sum_{i=1}^n \Pr\{x_i(s_j) \geq u_q(s_j)\}}, \\
 &\approx 1 - q,
 \end{aligned} \tag{3.4}$$

if  $r(x_i) > u$  for all  $i = 1, \dots, n$ .

The implication of this bias in our application is that  $\hat{\pi}_q(s_j, s'_j)$  tends to take values close to  $1 - q$  for  $s_j, s'_j$  sufficiently distant. Figure 3.3 provides an illustration for this phenomenon with  $q = 0.3$  and  $q = 0.8$ ; one can see Figure 3.2 for an in between case with  $q = 0.52$ . These three plots all display an asymptotic extremal coefficient of approximately  $1 - q$ , as described in (3.4).

### 3.4 Tail marginal modelling

The vector  $b_n$  defining the local excess is defined using local quantiles of the 35-storm database, with  $u_q(s_l)$  the local empirical quantile of  $r$ -exceedances at location  $s_l$ . The vector  $b_n$  is defined as

$$b_n(s_l) = u_q(s_l), \quad s_l \in S, \tag{3.5}$$

where  $q \in [0, 1]$  a shared quantile level for all locations defined such that  $r(b_n) = u_n$ . The constraint  $r(b_n) = u_n$  comes from Theorem 1 by setting the variable  $u = 0$ .

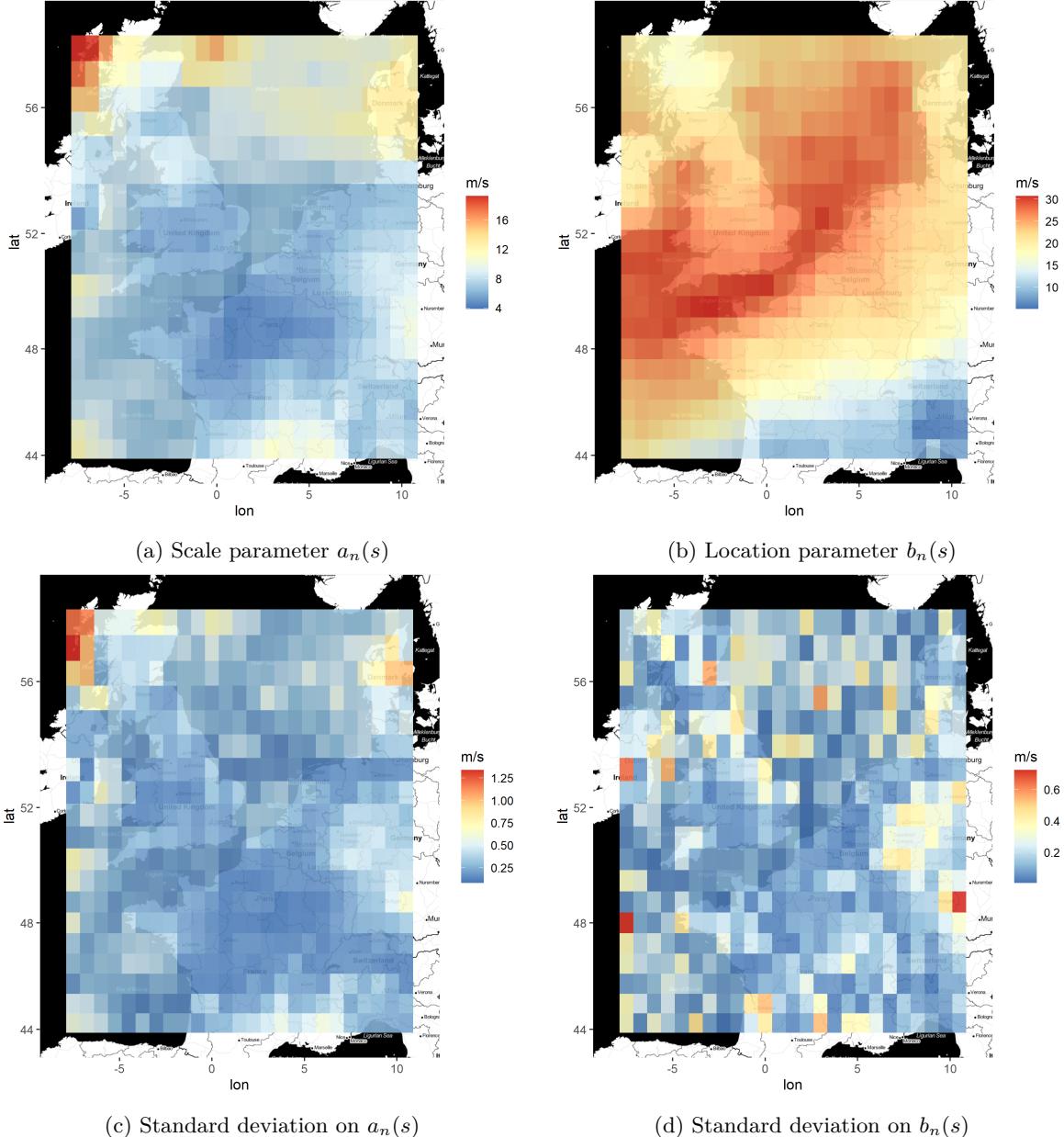


Figure 3.4: Estimated marginal parameters  $a_n(s)$  and  $b_n(s)$  with associated standard deviations. The associated shape parameter is  $\xi = -0.59_{0.02}$ . The standard deviations, on the maps and indicated by subscripts, are obtained with jackknife leave-one-out method on the set of 35 exceedance observations.

The estimation of marginal scales  $a_n$  and common shape  $\xi$  is done simultaneously using the independence log-likelihood from (2.24).

The threshold functional  $b_n$  resulting from solving (3.5) is  $b_n(s_l) = u_q(s_l)$  for  $q = 0.19$  and is displayed in Figure 3.4. The common scale coefficient is  $\xi = -0.59_{0.02}$  and the  $a_n$  estimate is shown in Figure 3.4 as well. These estimates display strong geographical pattern and the land versus sea difference is clearly visible, though they show reasonable spatial continuity. On top left corner of the domain, the scale function  $a_n$  and its jackknife errors are very high compared to the rest of the domain, this is due to one footprint in our dataset containing very high wind gust in this top left region. A more sophisticated process to transform instantaneous wind gust data in storm footprints could probably provide better

results, for example by omitting points far from the centre of the storm in the footprint, as done in the XWS catalogue (Roberts et al., 2014).

Assessing the estimates of the marginal model, composed of parameters  $a_n$ ,  $b_n$  and  $\xi$ , can be done with qq-plots. Equation (2.16) implies that the marginal model at location  $s_l$  for exceedance of  $x(s_l)$  over the threshold  $b_n(s_l)$ , is a GPD with parameters  $(\xi, a_n(s_l), b_n(s_l))$ . Confidence intervals are generated using resampling on the marginal GPD distribution. At all locations, the number of observations  $x_i(s_l)$  exceeding  $b_n(s_l)$  is equal, this is induced by our definition of  $b_n$  in (3.5); let us denote this number of locally exceeding observations by  $n_q$ .

We draw  $m = 10\,000$  samples of size  $n_q$   $(Z_1^1, \dots, Z_{n_q}^1), \dots, (Z_1^m, \dots, Z_{n_q}^m)$  from the modelled GPD distribution. We order each sample as  $(Z_{(1)}^1, \dots, Z_{(n_q)}^1)$  where  $Z_{(j)}^k$  denotes the  $j$ th order statistic of sample  $k$ . The confidence interval for the  $j$ th order statistic in the qq-plot is then defined taking the 0.5% and 99.5% quantiles from the set  $Z_{(j)}^1, \dots, Z_{(j)}^m$ .

Generating our confidence intervals this way, we suppose the errors come from the finite dimension of our sample and neglect the errors on our parameters  $a_n$ ,  $b_n$  and  $\xi$ . An improvement would be for each sample  $k$  to sample  $a_n^k$ ,  $b_n^k$  and  $\xi^k$  from normal distributions centred on  $a_n$ ,  $b_n$  and  $\xi$ , and with variance equal to the jackknife estimated variance. We could then sample each set  $k$  using a different GPD distribution with parameters  $a_n^k(s_l)$ ,  $b_n^k(s_l)$  and  $\xi^k$ .

The distribution of  $r(X)$  also only depends on the marginal parameters  $\xi$ ,  $a_n(s_l)$ ,  $b_n(s_l)$ , as it follows a generalized Pareto distribution with shape  $\xi$ , scale  $r(a_n)$  and threshold  $r(b_n) = u_n$ .

Figure 3.5 displays the storm risk qq-plot with the empirical distribution, the risk functional evaluated for each storm footprint, and the theoretical distribution, a GPD with shape parameter  $\xi$  and scale parameter  $r(a_n)$ . The result is not very good, as points are outside the confidence interval, but we were unable to obtain a better fit, even increasing the threshold  $u_n$ . This qq-plot shows that our model overestimates the risk distribution, but we rather be on the safe side than fitting an underestimated risk distribution.

The marginal qq-plots for 6 sample points, the yellow points in Figure 3.1, are displayed in Figure 3.6. They show very good results except for the upper right point, where the risk is significantly overestimated, but the overall marginal model performs well.

Tests on marginal parameter estimates were performed to choose the appropriate number of storms to keep for our study. Too many events of low risk impacts the estimates of high quantiles' dependence negatively while too high a threshold results in poor precision in the estimates. The number of 35 storms and the associated threshold  $u_n = 25.88 \text{ m s}^{-1}$  was chosen as a compromise.

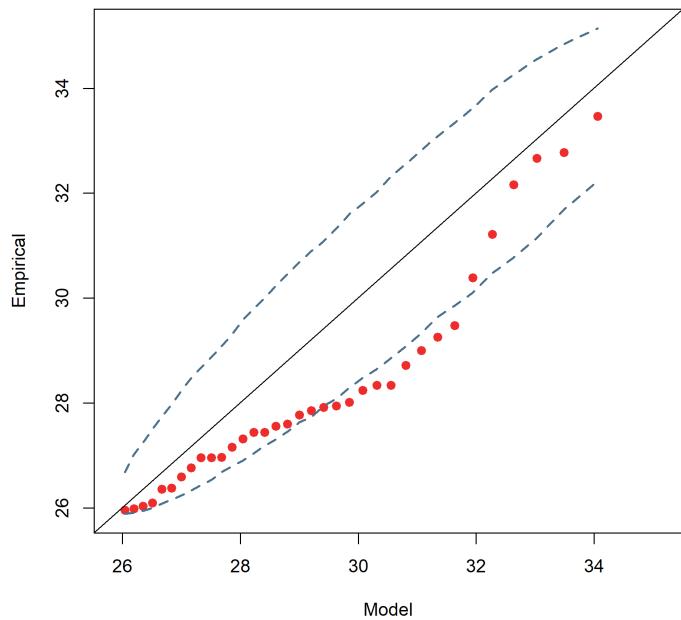


Figure 3.5: Storm dataset risk qq-plot using theoretical distribution obtained using marginal parameter estimates. The dashed lines are ninety-nine percent intervals computed using resampling.

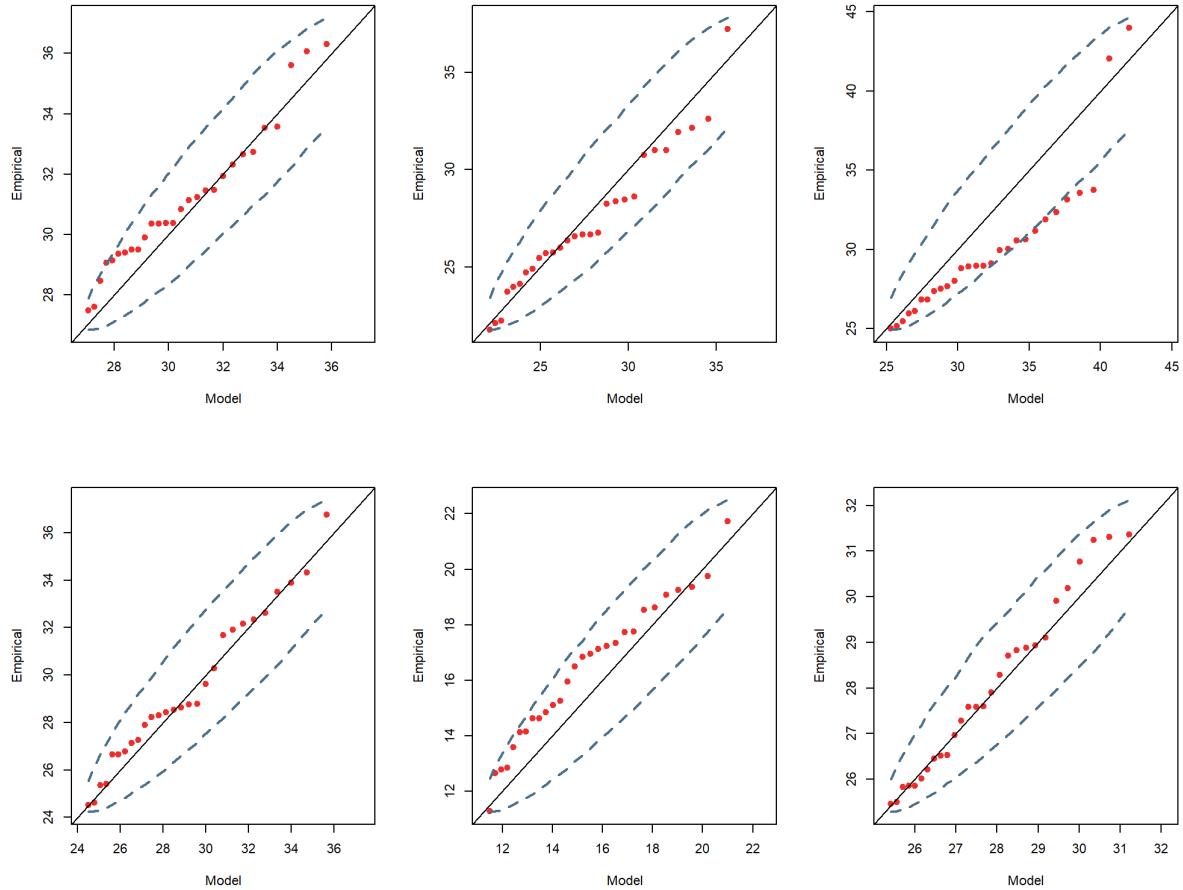


Figure 3.6: Marginal qq-plots of empirical distributions versus those modelled by parameters  $\xi$ ,  $a_n$  and  $b_n$ . Sample locations are indicated in Figure 3.1 in yellow; point locations on the map are matched with qq-plot position, i.e., the bottom left qq-plot relates to the bottom left location in Figure 3.1. Ninety-nine percent confidence intervals are computed by resampling and indicated by dashed blue lines.

# 4 Non-stationary extremal dependence modelling

This chapter focuses on the dependence part of the model, we here describe the family of processes used for  $W_1$  in (2.13). We give details on the methods used to generate non-stationary processes and on inference methodology.

## 4.1 Generalities on Gaussian Markov random fields

The process  $W_1$  in (2.13) contains the dependence structure of the  $r$ -Pareto process  $P$ , so one needs to choose a family of processes to model  $W_1$ . As in de Fondeville & Davison (2020), we model  $W_1$  as log-Gaussian; in our case, however, the underlying Gaussian field is non-stationary with modified Matérn covariance, as opposed to de Fondeville & Davison (2020) who used stationary Gaussian fields on space and time. We will introduce Gaussian fields and develop the stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011). We later give details on modification of the SPDE equation allowing non-stationary Gaussian fields (Fuglstad et al., 2015).

Let  $s_1, \dots, s_L \in S \subset \mathbb{R}^2$  be a finite subset of locations, with coordinates along two axes and unitary vectors  $(e_1, e_2)$ ,  $s_l = s_{l(1)}e_1 + s_{l(2)}e_2$ . A Gaussian Markov random field (GMRF) is a multivariate Gaussian random variable  $w \sim \mathcal{N}(\mu, Q^{-1})$ , where  $\mu$  is the mean vector and  $Q$  is the precision matrix. The Markov property emerges from the non-zero components of the  $Q$  corresponding to conditionally independent elements of  $w$ .

The conditional moments of a GMRF are

$$\mathbb{E}(w_i | w_{-i}) = \mu_i - \frac{1}{Q_{i,i}} \sum_{j \neq i} Q_{i,j}(w_j - \mu_j), \quad \text{Var}(w_i | w_{-i}) = \frac{1}{Q_{i,i}}, \quad (4.1)$$

where  $w_{-i}$  denotes the vector  $w$  with element  $i$  removed.

The connection between Matérn Gaussian fields and SPDEs is that with  $\mathcal{W}(s)$  spatial Gaussian white noise,  $\kappa > 0$  a scale coefficient and  $\Delta = \frac{\partial^2}{\partial e_1^2} + \frac{\partial^2}{\partial e_2^2}$ ,

$$(\kappa^2 - \Delta)w(s) = \mathcal{W}(s), \quad s \in S \subset \mathbb{R}^2, \quad (4.2)$$

gives rise to a Gaussian field  $w(s)$  with Matérn covariance function

$$\text{Cov}\{w(s_1), w(s_2)\} = \frac{1}{4\pi\kappa^2} (\kappa \|s_2 - s_1\|) K_1(\kappa \|s_2 - s_1\|), \quad s_1, s_2 \in S \subset \mathbb{R}^2, \quad (4.3)$$

where  $K_1$  is the modified Bessel function of the second kind of order 1. This is a Matérn covariance

function with a smoothness parameter  $\nu$  fixed to unity. This parameter is difficult to estimate and restricting it is not believed to be a major practical limitation. Its restriction allows the use of a locally anisotropic modification developed in Fuglstad et al. (2015).

Equation (4.2) can be reformulated in the case of a grid on  $\mathbb{R}^2$ , with  $w$  a GMRF, and a matrix  $B(\kappa)$  only involving the neighbours of each location, transforming (4.2) into

$$B(\kappa)w \sim \mathcal{N}(0, I). \quad (4.4)$$

The solution to the SPDE (4.4) is a GMRF  $w \sim \mathcal{N}(0, Q^{-1}(\kappa))$  with  $Q(\kappa) = B(\kappa)^\top B(\kappa)$ . The sparsity of the precision matrix is induced by the sparsity of  $B(\kappa)$ .

To introduce anisotropy, we modify the Laplacian operator  $\Delta$  with a  $2 \times 2$  positive definite matrix function  $H(s)$  obtaining

$$\begin{aligned} \nabla \cdot H(s) \nabla &= \frac{\partial}{\partial e_1} \left( h_{11}(s) \frac{\partial}{\partial e_1} \right) + \frac{\partial}{\partial e_1} \left( h_{12}(s) \frac{\partial}{\partial e_2} \right) \\ &\quad + \frac{\partial}{\partial e_2} \left( h_{21}(s) \frac{\partial}{\partial e_1} \right) + \frac{\partial}{\partial e_2} \left( h_{22}(s) \frac{\partial}{\partial e_2} \right). \end{aligned} \quad (4.5)$$

The modified SPDE is then

$$\{\kappa^2 - \nabla \cdot H(s) \nabla\} w(s) = \mathcal{W}(s), \quad s \in S = [A_1, B_1] \times [A_2, B_2] \subset \mathbb{R}^2. \quad (4.6)$$

Modifying the SPDE using (4.6) generates a random process corresponding to the combination of locally Matérn-like anisotropic fields at each location into a full process.

For interpretability we parametrize the anisotropy function  $H$  as

$$H(s) = \gamma I_2 + v(s)v(s)^\top, \quad (4.7)$$

where  $\gamma > 0$  specifies the standard isotropic effect and  $v$  is a two-dimensional vector field specifying the direction and magnitude of the local anisotropic effect added to the isotropic one.

Using the modified SPDE (4.6), the resulting field does not have an explicit covariance or marginal variance. However the anisotropic modification does affect the marginal variance; while not explicit in our case, the marginal variance has a closed form when  $H$  is constant (Fuglstad et al., 2015). If  $H(s)$  is constant on the domain  $S$ , then the marginal variance is

$$\sigma_{\text{cst}}^2 = \frac{1}{4\pi\kappa^2 \sqrt{\det(H)}}. \quad (4.8)$$

This implies that if  $v(s)$  is constant, the bigger  $\|v\|$  is, the smaller is the marginal variance of the resulting field. If  $H(s)$  is not constant, the impact is not explicit but is similar; in a region with increased  $v(s)$  amplitude, the marginal variance of the resulting Gaussian field is reduced.

## 4.2 Parametric models for vector fields

### 4.2.1 Polynomial

The parametric model for the anisotropy vector field suggested in Fuglstad et al. (2015) uses periodic functions on the domain. Periodic function bases are not appropriate for anisotropic vector fields in

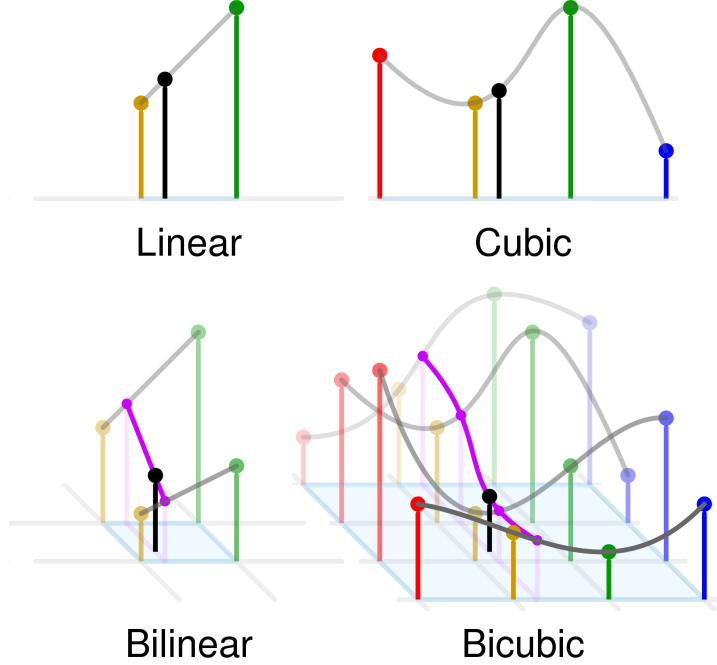


Figure 4.1: Representation of linear, cubic, bilinear and bicubic interpolation methods (Cmglee, 2020). The black dots correspond to the points being interpolated, and the red, yellow, green and blue dots correspond to the known field values called knots. Interpolation domain is in light blue.

environmental applications such as ours, as high numbers of functions would be required to fit anisotropic vector fields that are not periodic. Limiting the number of anisotropic parameters is important as every additional degree of freedom comes with high computational cost.

An obvious option is polynomial functions, so we write our anisotropic vector field as

$$v(s) = v(s_{(1)}, s_{(2)}) = v_{(1)}(s_{(1)}, s_{(2)})e_1 + v_{(2)}(s_{(1)}, s_{(2)})e_2$$

with

$$v_{(i)}(s) = \sum_{\substack{m=0 \\ j+k=m}}^M \beta_i^{j,k} s_{(1)}^j s_{(2)}^k, \quad (4.9)$$

where  $M$  is the maximum order of the polynomial functions, and  $\beta_i^{j,k}$  ( $i = 1, 2$ ) are coefficients.

#### 4.2.2 Bidimensional interpolation

Another approach is using bidimensional interpolation to generate a field on the domain  $S$ . We use two bivariate interpolation methods: the bilinear and the bicubic. As their name suggest, these methods are generalizations to fields of the linear and cubic interpolations. Figure 4.1 shows representations of both methods in one and two dimensions.

In one dimension, linear interpolation requires at least two known points and cubic interpolation requires at least four; logically their extensions in two dimensions require at least four and 16 points for bilinear and bicubic methods.

To model each coordinate of our vector field  $v$ , we fix knot points  $p_1, \dots, p_k \in \mathbb{R}^2$  regularly spaced

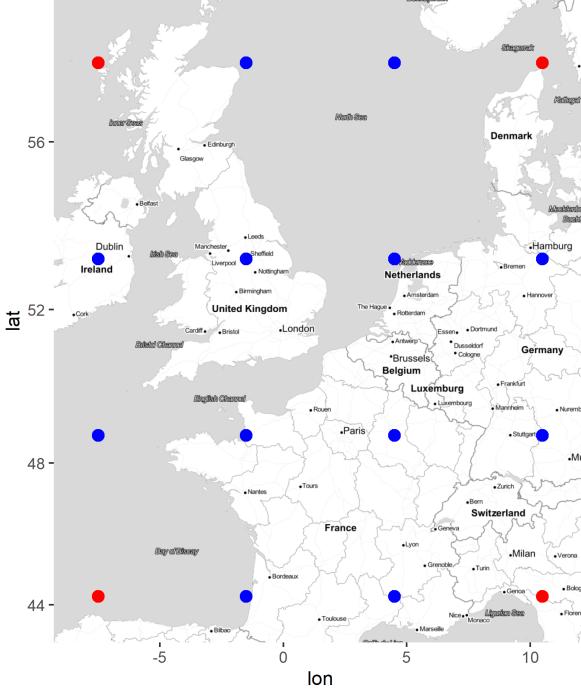


Figure 4.2: Location of the interpolation knots. The bilinear interpolation uses knots in red and the bicubic interpolation uses knots in both blue and red.

in a grid pattern such that our domain  $S$  is contained in the square region formed by the points. The number of points is  $k = 4$  in the bilinear case and  $k = 16$  in the bicubic case. We define the value of the field at each knot and let the rest of the domain be interpolated from the knots. The value of the vector field is

$$v_{(i)}(s) = \text{Interp}\{s, p_1, \dots, p_k, v_{(i)}(p_1), \dots, v_{(i)}(p_k)\}, \quad s \in S, \quad i = 1, 2, \quad (4.10)$$

where  $\text{Interp}$  denotes the interpolation function, bilinear or bicubic,  $s$  is the interpolated location and  $v_{(i)}(p_1), \dots, v_{(i)}(p_k)$  are the values of the field  $v_{(i)}$  at knot locations.

This modelling of the anisotropy field is simple and is efficient in optimization algorithms as each field parameter is independent of the others and only influence a local region on the domain, in opposition to the polynomial family where a slight change in the behaviour of the field in one region requires fine-tuning all the parameters.

The number of parameters describing the vector field is  $2 \times k = 8$  or 32 depending on the interpolation method. Although 32 is a high number of parameters to estimate in the bicubic model, one can use inference results of the bilinear model to interpolate reasonable values as starting parameters for the optimization of the bicubic model.

Finally, these interpolation methods allow for interpretability of the vector field parameters as they represent the different values of the field at knot locations on the domain.

## 4.3 Inference and implementation

### 4.3.1 Score estimation

To allow the use of existing R packages, and to connect our work more easily with Wadsworth & Tawn (2014), the observations are rescaled marginally so that they have unit tail. This is done by transforming the initial  $r$ -exceeding events  $x_i$ ,  $i = 1, \dots, n$  to marginal unit Fréchet distributions by

$$\tilde{x}_i(s_l) = \left\{ 1 + \xi \frac{x_i(s_l) - b_n(s_l)}{a_n(s_l)} \right\}_+^{1/\xi}, \quad i = 1, \dots, n, \quad l = 1, \dots, L.. \quad (4.11)$$

This expression is that used as input for  $\lambda$  in (2.23).

Many different approaches are available for inference on the dependence structure of generalized  $r$ -Pareto processes. In this project we use the gradient score approach developed in de Fondeville & Davison (2018). In score inference, the likelihood function is replaced by a scoring rule; in our case we define our estimator as the minimization of the gradient scoring rule. For an  $L$ -dimensional rescaled vector of observations  $\tilde{x}$  with  $\tilde{x}_{(l)} = \tilde{x}(s_l)$ , the gradient score is

$$\delta(\lambda_\theta, \tilde{x}) = \sum_{l=1}^L \left( 2w_{(l)}(\tilde{x}) \frac{\partial w_{(l)}(\tilde{x})}{\partial \tilde{x}_{(l)}} \frac{\partial \log \lambda_\theta(\tilde{x})}{\partial \tilde{x}_{(l)}} + w_{(l)}(\tilde{x})^2 \left[ \frac{\partial^2 \log \lambda_\theta(\tilde{x})}{\partial \tilde{x}_{(l)}^2} + \frac{1}{2} \left\{ \frac{\partial \log \lambda_\theta(\tilde{x})}{\partial \tilde{x}_{(l)}} \right\}^2 \right] \right), \quad (4.12)$$

where  $w$  is a weight functional on  $\mathcal{A}_r$  taking values in  $[0, 1]$  and  $\tilde{x}$  is an  $L$ -dimensional vector of Fréchet rescaled observations. The main advantage of the gradient score inference compared to other methods (such as censored likelihood) is to avoid computing the scaling constant  $\Lambda(\mathcal{A}_r)$  in (2.23).

The weight function vector, for an observation vector  $x$ , is defined as

$$w(x) = x[1 - \exp(\zeta\{1 - r(x)\})], \quad (4.13)$$

where  $\zeta > 0$  is a scaling coefficient. The use of a weight function allows us to downweight small observations in the set, making the inference more robust to them. Any other weight function respecting the conditions given in de Fondeville & Davison (2018, Supplementary Material) could have been used. In this project we adjust  $\zeta$  so that 70% of the weights  $[1 - \exp(\zeta\{1 - r(x)\})]$  associated with our rescaled storms are above 0.9.

The formula used for intensity  $\lambda_\theta$  is from Wadsworth & Tawn (2014) and is discussed in de Fondeville & Davison (2018). It is appropriate in our case as it works for Brown–Resnick models, and even for non-stationary Gaussian fields:

$$\begin{aligned} \lambda_\theta(\tilde{x}) &= \frac{|\det \Sigma_\theta|^{-1/2} (1_L^\top \rho)^{-1/2}}{(2\pi)^{(D-1)/2} \tilde{x}_{(1)} \cdots \tilde{x}_{(L)}} \exp \left( -\frac{1}{2} \left[ \log \tilde{x}^\top \Gamma \log \tilde{x} + \log \tilde{x}^\top \left\{ \frac{2\rho}{1_L^\top \rho} + \Sigma_\theta^{-1} \sigma - \frac{\rho \rho^\top}{1_L^\top \rho} \right\} \right] \right) \\ &\times \exp \left( -\frac{1}{2} \left\{ \frac{1}{4} \sigma^\top \Sigma_\theta^{-1} \sigma - \frac{1}{4} \frac{\sigma^\top \rho \rho^\top}{1_L^\top \rho} + \frac{\sigma^\top}{1_L^\top \rho} - \frac{1}{1_L^\top \rho} \right\} \right), \end{aligned} \quad (4.14)$$

where  $\Sigma_\theta$  is the  $L$ -dimensional covariance matrix of a non-stationary Gaussian process,  $\rho = \Sigma_\theta^{-1} 1_L$ ,  $\Gamma = \Sigma_\theta^{-1} - \rho \rho^\top / 1_L^\top \rho$  and  $\sigma = \text{diag}(\Sigma_\theta)$ .

The R package `mvPot` (de Fondeville & Belzile, 2018, R Core Team, 2020) implements gradient score estimation for the Brown–Resnick model but using variograms and for intrinsic stationary Gaussian fields. We modified the function `mvPot::scoreEstimation` so that precision and covariance matrices generated

by the SPDE approach can be used instead of the variograms. The main computational complexity of the modified procedure results from the inversion of two matrices, though the precision matrix from the SPDE is sparse.

### 4.3.2 Composite inference

Often when the computation of the full joint likelihood is long, composite likelihood can be used, as a substitute, to reduce computation time. In our case, thanks to the SPDE approach, computing full joint likelihood on a set of  $L$  locations is as fast as computing a reduced likelihood on  $L' < L$  points. We noticed, however, that using gradient score with a number of spatial points much larger than the number of observations of the process results in the score function behaving badly, thus making the inference impossible. As a result, even though our approach would allow for fast full likelihood computation, we instead use a composite gradient score with likelihood limited to a number of spatial points of the same order of magnitude as our number of r-exceedance observations  $n = 35$ . We randomly sample  $n_c = 1000$  sets of  $L_c = 25$  spatial points and perform inference using this composite gradient score. To allow for comparison between the models, the same sets of points are used for composite score in all the models.

### 4.3.3 Computational considerations on the SPDE method

The implementation to generate the non-stationary Gaussian field precision matrix  $Q$  and the associated covariance matrix  $\Sigma = Q^{-1}$  is based on Fuglstad et al. (2015, Supplementary material S2) and unpublished preliminary implementation by R. de Fondeville. The methodology uses the divergence theorem and the finite volume method (Eymard et al., 2000) to discretize the SPDE to the boundary of small rectangular cells regularly spaced on the domain. While the implementation is written in a way allowing for fast matrix computation, it induces a constraint on the location of our sample points in  $\mathbb{R}^2$ . This restriction to regularly gridded locations is limiting in our application, it prevents us from using projected coordinates on earth surface (see Figure 4.3) allowing for physical interpretability on the covariance matrix.

To map latitude-longitude coordinates to metric coordinates while keeping the locations in a grid-like pattern, we perform a simple projection of the form

$$(x, y) = (R_E \cdot \cos(\overline{\text{lat}}) \cdot \text{lon}, R_E \cdot \text{lat}), \quad (4.15)$$

where  $R_E$  is Earth radius in km and  $\overline{\text{lat}}$  is the mean latitude of the study region. This mapping takes the x-y ratio at the mean latitude and uses it for all locations; it is an approximation to transform our latitude-longitude coordinates into usable and meaningful metric coordinates while keeping the locations in a grid pattern for the anisotropic SPDE to work. The final coordinates are then recentred to place the origin at the centre of the region of interest.

One of the drawbacks of the SPDE approach is the impact of the boundary of the domain on the generated Gaussian field. As a result, when using the SPDE approach, one should compute the Gaussian field on a wider domain and then keep only the results for the domain of interest. In the stationary case and in setups similar to ours, i.e., using Dirichlet boundary conditions, one can expect results to be impacted by boundary effect for points that are within the range distance of the boundary, the range being the distance at which correlation between two points is equal to 0.1 (Rue & Held, 2005, Section 5.1.4). In our case let us denote  $S_E$ , the extension to domain  $S$  we use to obtain the SPDE covariance matrix. All computations are performed on the wider domain  $S_E$  and the covariance matrix obtained is

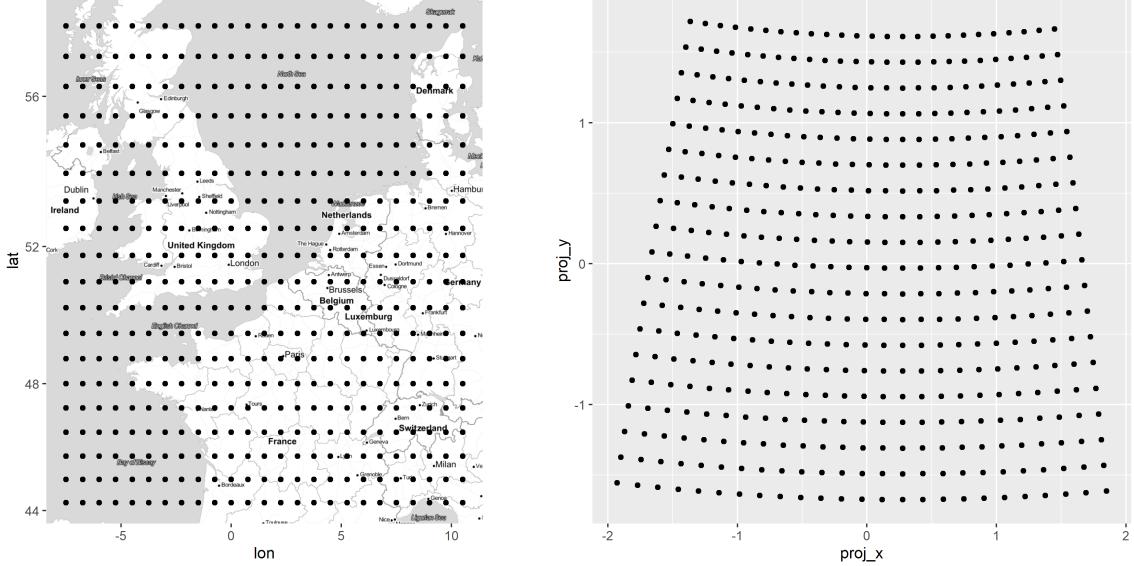


Figure 4.3: Gridded latitude-longitude of the locations used with a map of Europe (left) and their Universal Transverse Mercator rescaled projection on zone 31 (right).

then restricted to  $S$  to obtain  $\Sigma_\theta$ .

In all the computations, we use an extended domain  $S_E$  that is 160% the size of the initial domain  $S$  and is gridded in the same way, i.e., the restriction of the discrete grid of  $S_E$  to  $S$  gives the  $L$  gridded locations we have initially. A large extension domain is always preferable but as computation time increases quadratically with the size of the domain, one need to find a balance. In our case, tests on a few anisotropic fields showed that fluctuations of the amplitude of  $v(s)$  had higher impact than boundary effects on marginal variance at sample locations. Furthermore increasing the extension domain further would have made computation impossible on a personal laptop (6 cores, Intel i7-8750H, 32 GB RAM), requiring the use of computation clusters and all this for a very limited improvement in inference quality.

The implementation of the interpolation methods for bilinear and bicubic models is based on the **Akima** package (Akima & Gebhardt, 2020), which provides fast and flexible bilinear and bicubic interpolation functions.

#### 4.3.4 Optimization algorithm

Once the model and the cost function, here the gradient score, are established, parameter inference requires finding optimal parameters

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i \in K_{un}} \delta(\lambda_\theta, \tilde{x}_i).$$

Several methods are available to perform this optimization, ranging from gradient-free to Hessian methods; many trials with different methods were required before finding an appropriate one for our model. The complexity of  $\delta$ , especially the use of  $\sigma = \text{diag}(\Sigma_\theta)$  in  $\lambda_\theta$  (4.14), prevents straightforward analytical gradient or Hessian formulation. This implies that optimization algorithms using gradient or Hessian functions need to compute them numerically, inducing  $2 \times d_\theta$  respectively  $d_\theta \times (d_\theta + 1)/2$  evaluations of function under optimization,  $d_\theta$  being the dimension of the vector of dependence parameters  $\theta$ . While efficient in small dimension parameter domains, gradient-free methods, such as Nelder–Mead, face in-

creasing difficulty in converging as the parameter vector dimension increases.

To add to the difficulties induced by our model complexity, most currently implemented general optimization algorithms in R make no use of parallel computing. At a time when multicore processors are more and more accessible, for personal use as well as in clusters for research purpose, this limitation is a significant barrier in statistical model inference. Only two recently developed packages offer parallel implementation, both for Hessian methods, the packages `optimParallel` (Gerber & Furrer, 2019) and `marqLevAlg` (Philipps et al., 2020), implement the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS) and the Marquardt–Levenberg Algorithm (MLA) respectively.

The use of the Nelder–Mead method combined with L-BFGS parallel implementation, i.e., the output of each algorithm goes into the next one successively before reaching convergence, allows to benefit from the efficiency of gradient-free methods but also reaching convergence with hessian algorithms.

In use, the combination of these two algorithms showed better performance than Nelder–Mead alone. However, the estimation of model parameter errors using jackknife leave-one-out method requires the optimization process to be rerun  $n$  times. It leads to very long computation times especially for high dimension parameter models,  $d_\theta = 34$  or 14. At this time of the project, the use of the second parallel algorithm was introduced.

`marqLevAlg` is from an unpublished article (Philipps et al., 2020) uploaded on Arxiv in September 2020. They show that their parallel implementation of the Marquardt–Levenberg Algorithm is more robust and precise than the BFGS method but do not benchmark their package against the recent L-BFGS parallel implementation of `optimParallel`. Empirical tests on our inference function showed that `marqLevAlg` performs faster and more robustly than the combination of Nelder–Mead and L-BFGS. The convergence criteria used in `marqLevAlg` are parameter stability, objective function stability and Relative Distance to Maximum (RDM). At iteration  $k$  of the optimization algorithm, the first criterion is defined as  $\{\sum_{j=1}^{d_\theta} (\theta_j^k - \theta_j^{k+1})^2\}^{1/2}$  where  $\theta^k$  is the parameter vector at iteration  $k$ , the second criterion is equivalent to the first one but with the objective function instead of the parameter vector. The third criterion, RDM, is less intuitive and is defined as  $d_\theta^{-1} \{\nabla g(\theta)\}^\top H(\theta)^{-1} \nabla g(\theta)$ , for  $g$  the objective function,  $\theta$  the parameter vector,  $d_\theta$  its dimension and  $H(\theta)$  the Hessian of function  $g$  for parameters  $\theta$ . The Relative Distance to Maximum indicator represents a rescaled distance to the point in the parameter domain where the gradient is null. As the Hessian matrix is not invertible if evaluated at a saddlepoint, this criterion also prevents stopping at saddlepoint (Philipps et al., 2020). With the parameters scaled such that they have similar orders of magnitude, and with reasonable starting values, the algorithm converged in approximately five steps on most models used in this project.

While robust in most cases, if a model is badly specified or the parameter surface is too flat, the `marqLevAlg` algorithm can get stuck and have difficulties converging.

## 5 Estimated dependence models

The different dependence models we use and their parameters are the following:

- **Isotropic:** two parameters:  $\kappa$  and  $\gamma$ .
- **Stationary anisotropic:** four parameters:  $\kappa$ ,  $\gamma$  and constant anisotropy field parametrized with  $\beta_1^{0,0}$  and  $\beta_2^{0,0}$ .
- **Polynomial:** 14 parameters:  $\kappa$ ,  $\gamma$  and six parameters per vector field dimension, modelled by a polynomial of order two.
- **Bilinear:** ten parameters:  $\kappa$ ,  $\gamma$  and four parameters per vector field dimension, the vector field is the result of bilinear interpolation.
- **Bicubic:** 34 parameters:  $\kappa$ ,  $\gamma$  and 16 parameters per vector field dimension, the vector field is the result of bicubic interpolation.

A summary of the different models setup and gradient score results is displayed in Table 5.1. The gradient score improvement between the isotropic model and the stationary anisotropic one is quite low. This was expected as Figure 3.2 showed that while some locations display strong anisotropy, summed on the whole domain  $S$ , no anisotropy direction stands out clearly. This explains why adding a constant anisotropy on the domain  $S$  does not provide major improvement in the model performance.

The non-stationary models, on the other hand, show significant improvement in gradient score compared to the stationary models, even with a reasonably low number of parameters such as in the bilinear model.

Table 5.1 also indicates the invertibility of the Hessian matrix at the estimated solutions. Having an invertible Hessian matrix at the estimated solution ensures that the solution is locally optimal. One can notice that all models have invertible Hessian matrices except the bicubic one. The most probable cause for this is the degrees of freedom of the model being too high compared to the fitted data; the model is overspecified.

	Isotropic	Stationary anisotropy	Polynomial	Bilinear	Bicubic
Nb parameters	2	4	14	10	34
Nb anisotropy parameters	0	2	12	8	32
Gradient score	-64 439	-65 513	-81 826	-73 249	-90 162
Hessian Matrix Singular	No	No	No	No	Yes

Table 5.1: Specification and gradient score values for fitted models. Lower gradient score is better.

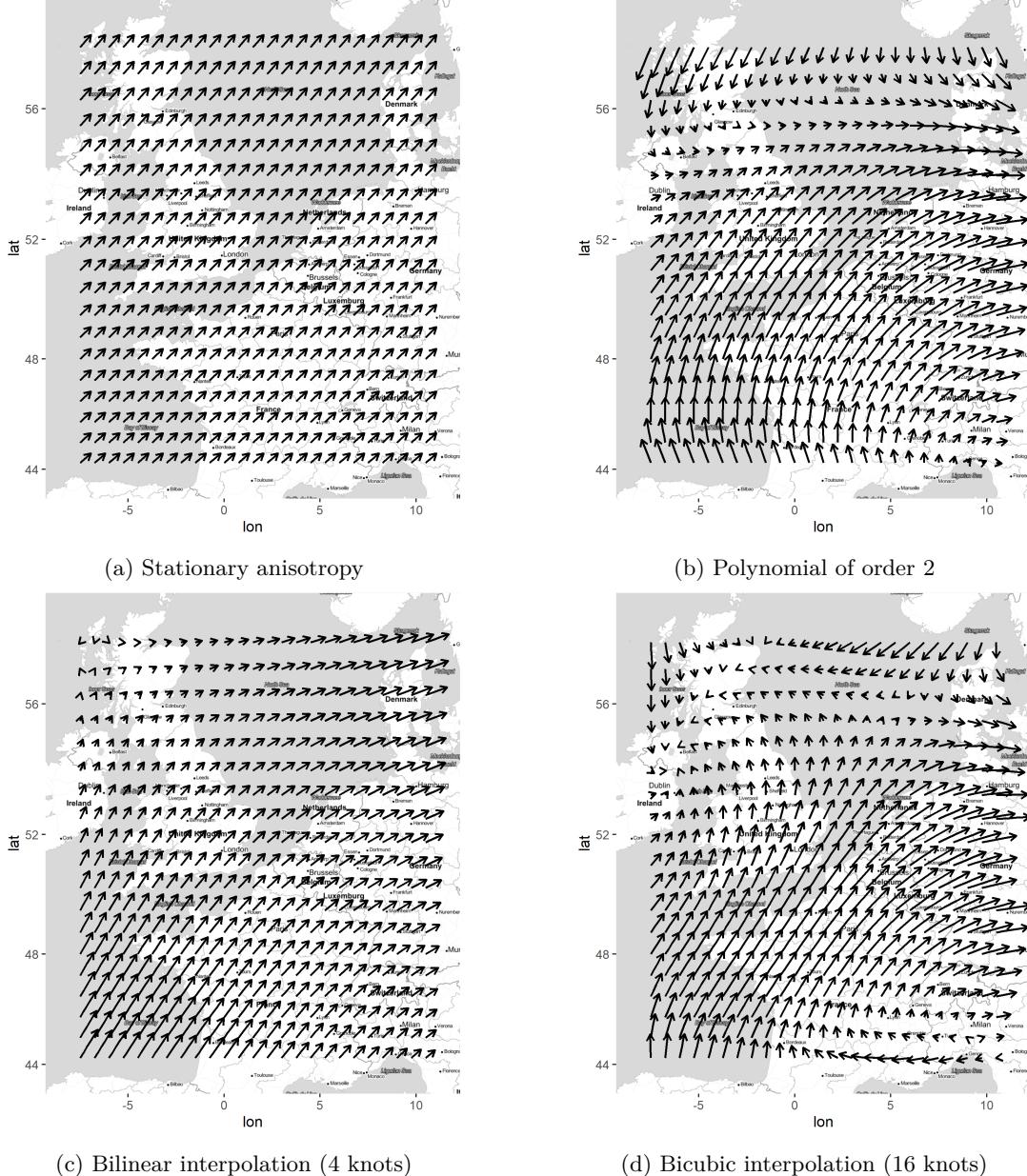


Figure 5.1: Anisotropy fields for different models

The resulting field estimates of the different models are shown on Figure 5.1. All the fields display strong anisotropy in the south-west to north-east direction matching the direction of the Channel. The vector fields in the polynomial model and in the bicubic models are extremely similar, giving weight to the bicubic model being over-parametrized.

In the two vector fields induced by high-flexibility models, the polynomial and the bicubic, the anisotropy in the Alps seems significantly different from the rest of the domain. There is, however, nothing surprising in windstorms having a different behaviour when hitting a mountain range. de Fondeville & Davison (2020) exclude the Alps as there is no reason to believe that physical phenomena are similar in high altitudes and at sea level. In our case, early implementation of the anisotropy field had restrictions on the rectangular shape of the domain  $S$  therefore the Alps region was kept.

The covariance of the Gaussian field induced by the vector field shows very strong anisotropy. Figure

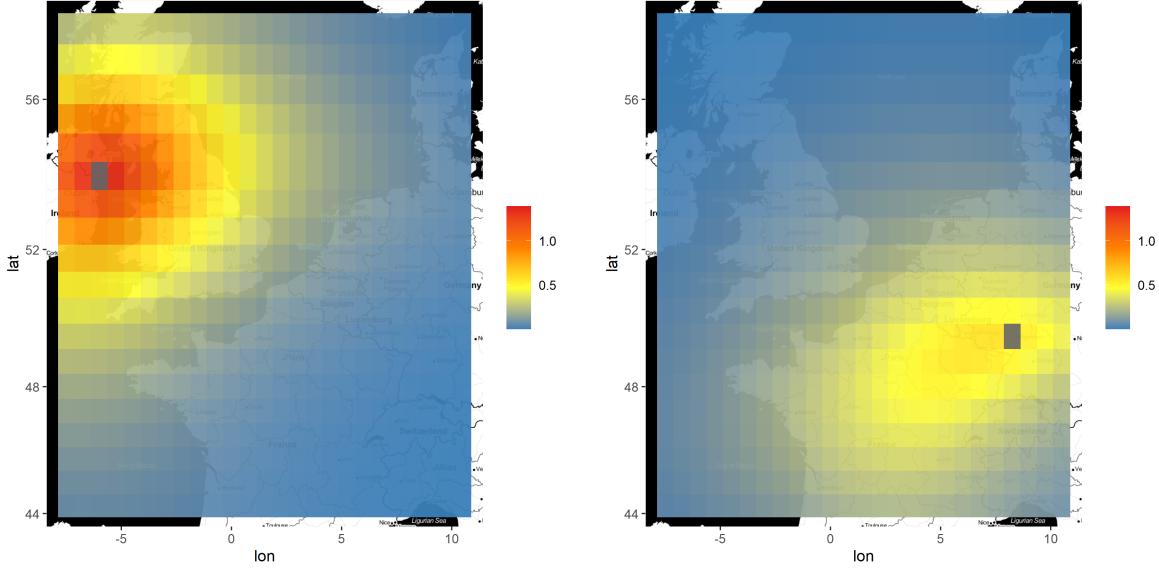


Figure 5.2: Covariance of the Gaussian field in the polynomial model, each plot represents a heatmap of the covariance  $\text{Cov}(s, \cdot)$  where  $s$  is the point in grey in each plot.

5.2 displays the covariance matrix obtained in the polynomial model; one can see the effect of the anisotropic field on the direction of the covariance and on the marginal variance, see equation (4.8). On the left side, the point is located in a low-amplitude region of the polynomial field; thus the covariance is mainly isotropic and the marginal variance is close to one. On the left side, the point is in a high amplitude region of the vector field and this induces a high anisotropy in the covariance matrix, the marginal variance is also much reduced compared to the plot on the left-hand side of the figure.

To quantify the variance of the estimated dependence parameters we use jackknife leave-one-out method on our 35 observation dataset. Jackknife was preferred over bootstrapping for its robustness and as the complexity of the optimization process would have made bootstrapping intractable. The resulting estimated errors are very low and indicate strong robustness of our method. The detailed parameter estimates for each model including jackknife standard errors on parameters are available in Appendix A.

## 5.1 Extremal dependence

To assess the extremal dependence structure resulting from our models, we use the  $\pi$  coefficient detailed in Section 3.3. Under the Brown–Resnick model, an explicit expression for  $\pi$  exists (de Fondeville & Davison, 2020). If  $\text{Cov}(\cdot, \cdot)$  is the covariance function of a non-stationary Gaussian field, its variogram is defined as

$$\gamma(s, s') = \frac{1}{2}[\text{Cov}(s, s) + \text{Cov}(s', s')] - \text{Cov}(s, s'), \quad s, s' \in S.$$

Using  $\gamma(s, s')$  the variogram of the Gaussian field underlying the log-Gaussian process  $W_1$ , then the extremal dependence coefficient  $\pi$  of our model reduces to

$$\pi(s, s') = 2 \left( 1 - \Phi \left[ \sqrt{\gamma(s, s')/2} \right] \right), \quad (5.1)$$

$\Phi$  being the normal cumulative distribution function. This explicit expression of  $\pi$  allow us to use it as the main indicator to compare the dependence structure of the fitted model with empirical data.

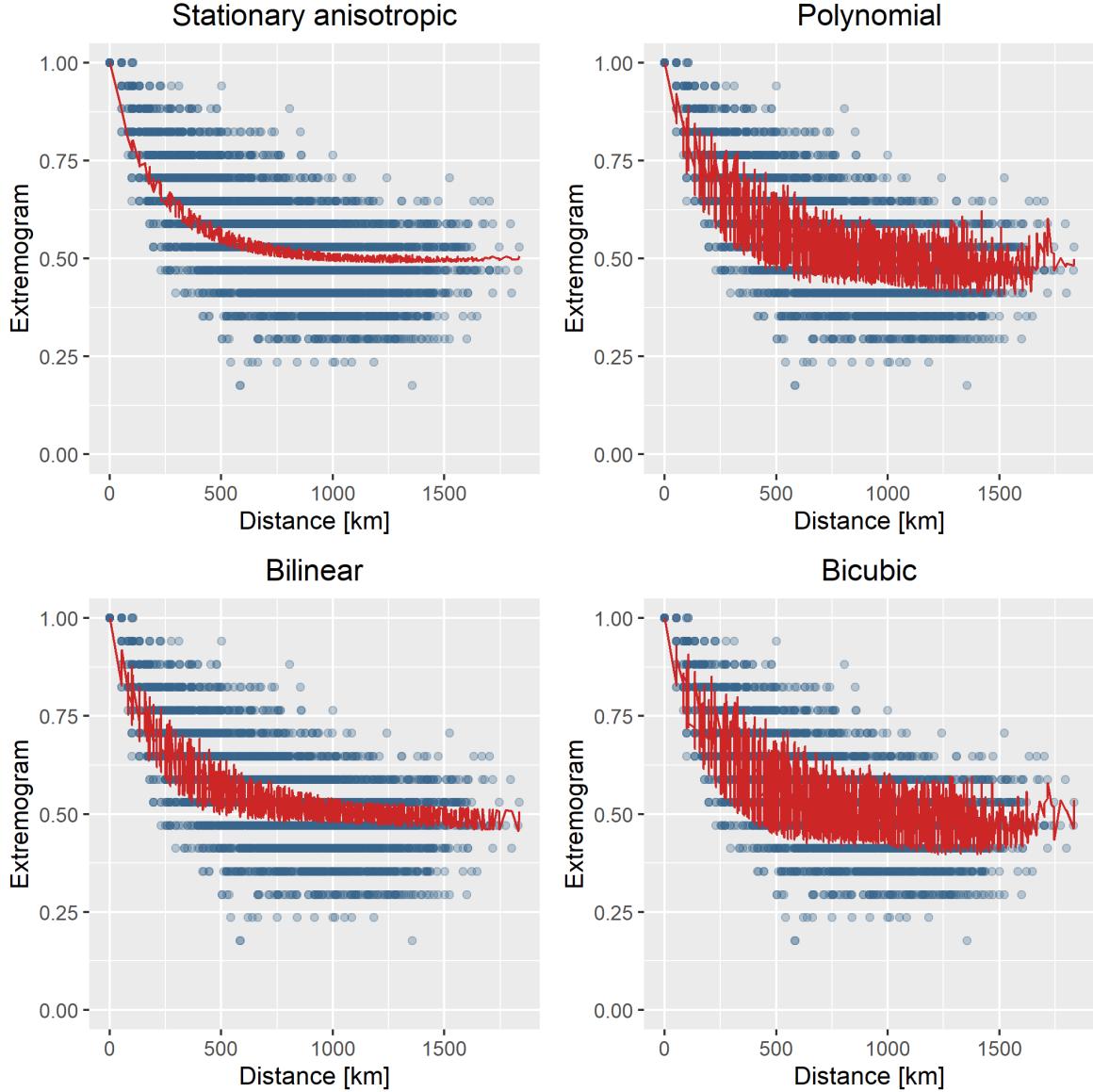


Figure 5.3: Extremogram comparison between different dependence models and empirical extremogram  $\hat{\pi}_{0.52}$ . The red line is the modelled extremogram obtained from (5.1) on pairs of points on the domain  $S$ . Blue points are the empirical extremogram for  $q = 0.52$ .

Figure 5.3 shows modelled extremograms as function of the distance. The vertical variation of the modelled extremograms in red is due to the anisotropy and the non-stationarity; in an isotropic model, the red line would be smooth and continuous as the associated variogram is only function of the distance.

We can see that the more complex our models are, the more they fill the cloud of empirical dependence  $\hat{\pi}$ , as they allow pairs of points at the same distance to have different extremal dependence depending on their location on the domain.

The choice of the quantile  $q = 0.52$  used for  $\hat{\pi}_q$  was made based on the results of the models. We choose  $q$  such that the behaviour of  $\hat{\pi}_q$  at very large distance is similar to that of the fitted extremograms. This is one big drawback of the methodology used; we have very limited control over the level of dependence fitted by our models. The generalized  $r$ -Pareto process dependence is invariant with the intensity of the

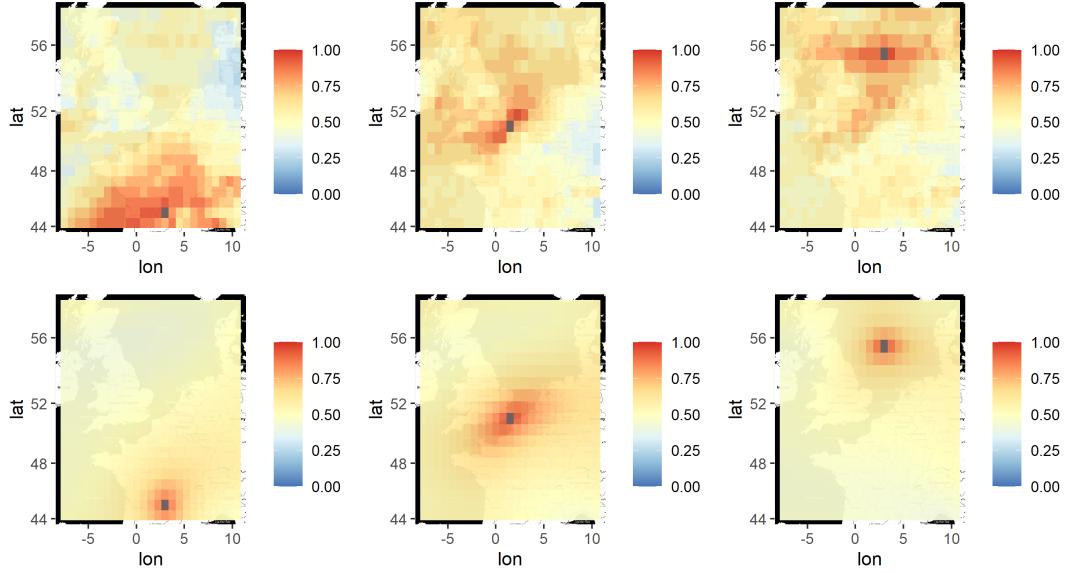


Figure 5.4: Extremogram comparison between  $\hat{\pi}_{0.52}(s, \cdot)$  (on top) and theoretical  $\pi(s, \cdot)$  in the polynomial model (at bottom), the three locations  $s$  are plotted in grey.

events but our data do not have the same property, one can see in Figure 3.3 that the dependence is function of the intensity of the events. The current model should be thus interpreted as “freezing” the dependence at the level  $q$  closest to the fitted model.

Few tools are available to tune the level of dependence of the model. Increasing the risk threshold  $u_n$  to select fewer events could help fitting a dependence structure closer to the one of very high intensity events. Tuning the  $\zeta$  parameter of the weight function in (4.13) can be used to modify the importance given to low-intensity events; tests with different  $\zeta$  parameters in our setup did not, however, result in significantly different fitted dependence structure. An interesting point to notice is that our five drastically different models for  $W_1$  all fit well with  $\hat{\pi}_{0.52}$ , suggesting that in our framework there is little to no influence of the dependence structure parametrization on the level of dependence fitted.

At a pointwise level, the resulting extremal dependence seems good for points within the risk region; for points outside this region, the resulting dependence is much narrower in the model than in the dataset and resulting extremogram compares poorly with empirical ones. A detailed view for three sample points on the domain is provided in Figure 5.4.

## 5.2 Example of simulated footprints

Using the dependence structure obtained from the polynomial model we can use Algorithm 1 to reproduce windstorm footprints.

Considering that we set the threshold  $u_n$  to select 35 windstorms and that our dataset span on 34.5 years, we can use  $u_n$  as the one-year return level. Based on Coles (2001, Section 4.3.3), obtaining simulated footprint for a specific return level is done by replacing the random unit Pareto variable  $R_2$  in Algorithm 1 by the  $1 - 1/m$  quantile of a unit Pareto distribution,  $m$  being the desired return level in years.

Figure 5.5 displays a comparative view of real footprints versus simulated ones. The simulated footprints reproduce the original footprints reasonably well, the path through Germany on the 17.5 years

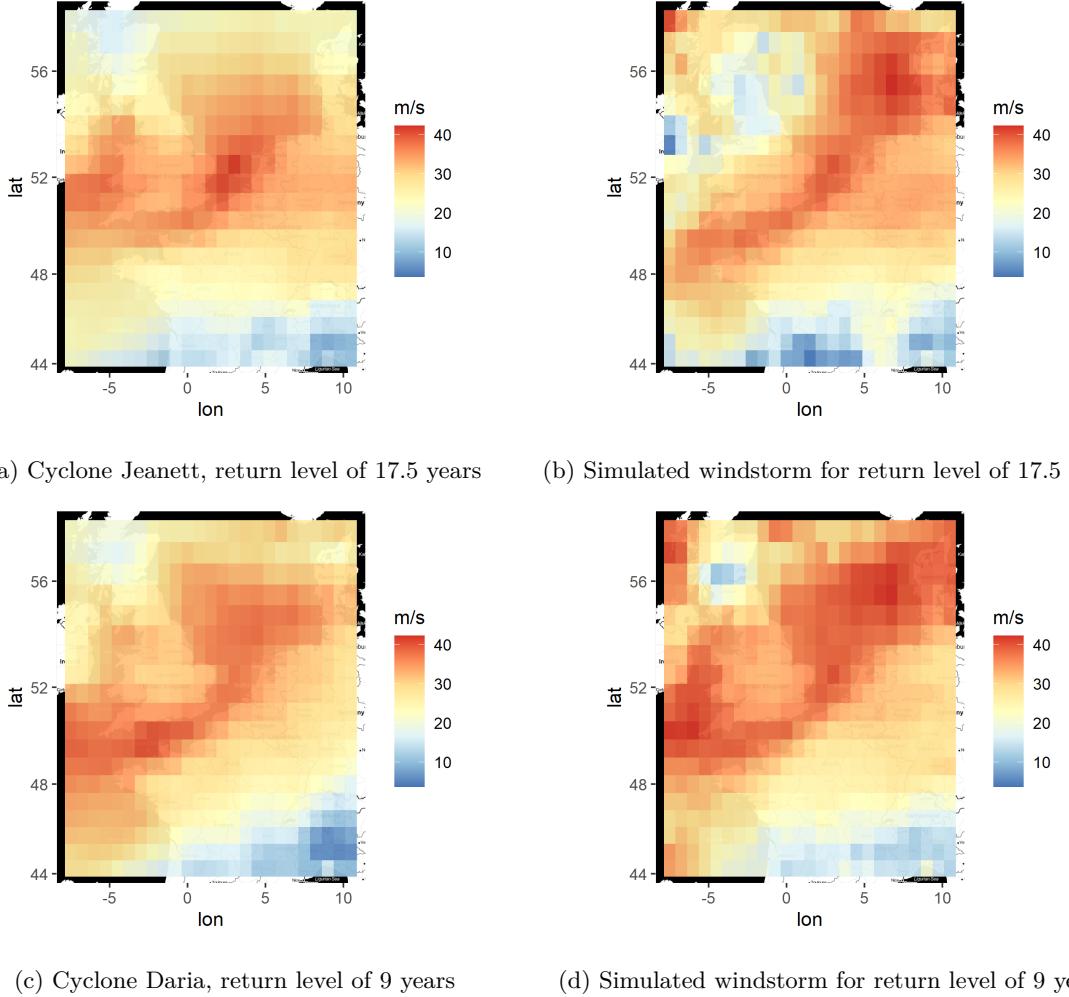


Figure 5.5: Comparison of real windstorm footprints with simulated ones for equivalent return levels using the polynomial dependence model.

simulated footprint is, for example, quite realistic. A slight overestimation of the wind gusts dependence is visible. As Daria and Jeanett are the two most intense events in our dataset, this dependence overestimation when simulating very intense events is not surprising; it is discussed in Section 5.1.

Our model also induces sharper footprints in the simulations, i.e., the real footprints are smoother than their simulated counterparts, this is induced by the smoothness parameter  $\nu$  in the Matérn covariance function. However, setting  $\nu = 1$  is a constraint inherent to the anisotropic SPDE methodology and it cannot be easily changed. Finally, one can notice very high wind gusts being simulated at the top left corner of the map, this is due to the very high marginal scale estimates  $a_n$  for these locations, see Section 3.4.

## 6 Discussion

The dependence inference part of the project shows very good results, the non-stationary models have significant improvement in the gradient score metric compared to stationary models. The direction of the anisotropic vector field is coherent between all models, and the standard errors on the fields' parameters show high precision in the estimates. While the dependence pattern is well estimated, our model gives very limited control on the level of dependence fitted and the modelled dependence is overestimated for intense events, as seen in the simulations in Figure 5.5. This restriction is mainly a consequence of our model being asymptotically dependent while our dataset does not display the same behaviour. Future work should implement sub-asymptotic models so that the spatial dependence can reduce with increasing event intensity, it would provide significant improvement in model fits and patch the main limitation of the model so far.

Regarding marginal behaviour, one small weak point of our current procedure is the low robustness to outliers of our pointwise scale estimation. While keeping the constraint  $u_n = r(b_n)$  some changes could include setting a continuity constraint on  $a_n$ , to prevent discontinuities such as in the top left corner of the domain. On the dataset, improvement easy to implement would be to rebuild the footprints database in a way such that are only added to footprints the locations on the domain that are involved in each storm, as the Extreme Wind Storm (XWS) catalogue is doing.

Finally, regarding the SPDE approach, it proved itself to be efficient and flexible to generate non-stationary Gaussian fields, however, a re-implementation of Fuglstad et al. (2015) allowing for non-gridded data, would be useful, allowing the projection of the studied locations on a physically more meaningful metric space.

## A Detailed dependence results

This appendix chapter contains the detailed dependence model estimates. Most model estimates are provided with jackknife standard errors as a subscript. The calculus of standard errors for the bicubic model was computationally too intensive to be performed; standard errors for this model is thus not provided.

	Polynomial	Stationary anisotropic		Isotropic
		$v_1$	$v_2$	
$\kappa^2 [10^{-3}]$		$0.84_{0.05}$		$1.15_{0.04}$
$\gamma [\text{km}^2]$		$48.5_{1.5}$		$55.1_{1.1}$
$\beta^{0,0} [\text{km}]$	$9.50_{0.24}$	$6.67_{0.27}$		$5.62_{0.20}$
$\beta^{1,0} [10^{-3}]$	$7.84_{0.80}$	$-2.13_{0.43}$		$3.47_{0.22}$
$\beta^{0,1} [10^{-3}]$	$0.03_{0.81}$	$-7.68_{0.37}$		
$\beta^{1,1} [10^{-6}\text{km}^{-1}]$	$2.44_{2.27}$	$6.47_{1.03}$		
$\beta^{2,0} [10^{-6}\text{km}^{-1}]$	$5.46_{0.99}$	$-7.70_{1.03}$		
$\beta^{0,2} [10^{-6}\text{km}^{-1}]$	$-22.4_{3.07}$	$-9.20_{0.82}$		

Table A.1: Parameter estimates for the isotropic model, the stationary isotropic model (equivalent to polynomial model of order 0) and the polynomial model of order 2

	Bicubic		Bilinear	
$\kappa^2 [10^{-3}]$	0.75		$1.27_{0.03}$	
$\gamma [\text{km}^2]$	52.4		$52.3_{1.45}$	
	$v_1$	$v_2$	$v_1$	$v_2$
$v(p_1)$ [km]	-0.31	8.76	$9.13_{0.44}$	$10.5_{0.43}$
$v(p_2)$ [km]	0.83	6.16	$4.97_{0.41}$	$2.86_{0.71}$
$v(p_3)$ [km]	-8.67	0.18	$-0.72_{0.60}$	$-0.39_{0.50}$
$v(p_4)$ [km]	-0.14	-0.01	$12.1_{0.85}$	$2.27_{1.40}$
$v(p_5)$ [km]	6.86	6.41		
$v(p_6)$ [km]	7.90	8.16		
$v(p_7)$ [km]	11.45	7.23		
$v(p_8)$ [km]	16.66	1.81		
$v(p_9)$ [km]	2.64	0.49		
$v(p_{10})$ [km]	0.00	4.38		
$v(p_{11})$ [km]	8.47	4.77		
$v(p_{12})$ [km]	16.38	0.15		
$v(p_{13})$ [km]	0.76	-7.38		
$v(p_{14})$ [km]	-0.61	-0.38		
$v(p_{15})$ [km]	-9.31	-3.34		
$v(p_{16})$ [km]	0.02	-4.27		

Table A.2: Parameter estimates for the bilinear and bicubic model

# References

- AKIMA, H. & GEBHARDT, A. (2020). Akima: Interpolation of Irregularly and Regularly Spaced Data. R package version 0.6-2.1.
- ALLEN, N. (2017). Press Release: PERILS releases new European Windstorm catalogue. PERILS. <https://www.perils.org/files/News/2017/Company-News/Press-Release-PERILS-UKMO-Storm-Catalogue-18-May-2017.pdf>.
- BEIRLANT, J., TEUGELS, J. L. & VYNCKIER, P. (1994). Extremes in non-life insurance. In *Extreme Value Theory and Applications: Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1 Gaithersburg Maryland 1993*, J. Galambos, J. Lechner & E. Simiu, eds. Boston, MA: Springer US, pp. 489–510.
- CMGLEE (2020). Illustration: Comparison of nearest-neighbour, linear, cubic, bilinear and bicubic interpolation methods. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Comparison\\_of\\_1D\\_and\\_2D\\_interpolation.svg](https://commons.wikimedia.org/wiki/File:Comparison_of_1D_and_2D_interpolation.svg). This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license.
- COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer London.
- DE FONDEVILLE, R. & BELZILE, L. (2018). mvPot: Multivariate Peaks-over-Threshold Modelling for Spatial Extreme Events. R package version 0.1.4.
- DE FONDEVILLE, R. & DAVISON, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika* **105**, 575–592.
- DE FONDEVILLE, R. & DAVISON, A. C. (2020). Functional Peaks-over-threshold Analysis. *arXiv:2002.02711 [stat]*.
- DEE, D. P., UPPALA, S. M., SIMMONS, A. J., BERRISFORD, P., POLI, P., KOBAYASHI, S., ANDRAE, U., BALMASEDA, M. A., BALSAMO, G., BAUER, P., BECHTOLD, P., BELJAARS, A. C. M., VAN DE BERG, L., BIDLOT, J., BORMANN, N., DELSOL, C., DRAGANI, R., FUENTES, M., GEER, A. J., HAIMBERGER, L., HEALY, S. B., HERSBACH, H., HÓLM, E. V., ISAKSEN, L., KÅLLBERG, P., KÖHLER, M., MATRICARDI, M., McNALLY, A. P., MONGE-SANZ, B. M., MORCRETTE, J.-J., PARK, B.-K., PEUBEY, C., DE ROSNAY, P., TAVOLATO, C., THÉPAUT, J.-N. & VITART, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**, 553–597.
- DOMBRY, C. & RIBATET, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface* **8**, 9–17.

- ENGELKE, S., DE FONDEVILLE, R. & OESTING, M. (2019). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika* **106**, 127–144.
- EYMARD, R., GALLOUËT, T. & HERBIN, R. (2000). Finite volume methods. In *Solution of Equation in  $\mathbb{R}^n$  (Part 3), Techniques of Scientific Computing (Part 3)*, vol. 7 of *Handbook of Numerical Analysis*. Elsevier, pp. 713–1018.
- FERREIRA, A., DE HAAN, L. & ZHOU, C. (2012). Exceedance probability of the integral of a stochastic process. *Journal of Multivariate Analysis* **105**, 241–257.
- FUGLSTAD, G.-A., LINDGREN, F., SIMPSON, D. & RUE, H. (2015). Exploring a New Class of Non-stationary Spatial Gaussian Random Fields with Varying Local Anisotropy. *Statistica Sinica* **25**, 115–133.
- GERBER, F. & FURRER, R. (2019). optimParallel: An R Package Providing a Parallel Version of the L-BFGS-B Optimization Method. *The R Journal* **11**, 352.
- KATZ, R. W., PARLANGE, M. B. & NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* **25**, 1287–1304.
- LINDGREN, F., RUE, H. & LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with Discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 423–498.
- NI, X., MUEHLBAUER, A., ALLEN, J. T., ZHANG, Q. & FAN, J. (2020). A Climatology and Extreme Value Analysis of Large Hail in China. *Monthly Weather Review* **148**, 1431–1447.
- PHILIPPS, V., HEJBLUM, B. P., PRAGUE, M., COMMENGES, D. & PROUST-LIMA, C. (2020). Robust and Efficient Optimization Using a Marquardt-Levenberg Algorithm with R Package marqLevAlg. *arXiv:2009.03840 [stat]*.
- R CORE TEAM (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- ROBERTS, J. F., CHAMPION, A. J., DAWKINS, L. C., HODGES, K. I., SHAFFREY, L. C., STEPHENSON, D. B., STRINGER, M. A., THORNTON, H. E. & YOUNGMAN, B. D. (2014). The XWS open access catalogue of extreme European windstorms from 1979 to 2012. *Natural Hazards and Earth System Sciences* **14**, 2487–2501.
- RUE, H. & HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman & Hall/CRC.
- RUE, H., MARTINO, S. & CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with Discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 319–392.
- WADSWORTH, J. L. & TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101**, 1–15.