

# A Non-stationary Dependence Model for Extreme European Windstorms

Master thesis oral exam  
Paul Castelain

---

Supervision: Prof. Anthony C. Davison and Dr Raphaël de Fondeville

February 5, 2021

Chair of Statistics, EPFL

# Table of contents

1. Introduction
2. Peaks-over-threshold exceedances
3. Data preprocessing
4. Marginal estimates
5. Non-stationarity
6. Dependence results
7. Discussion

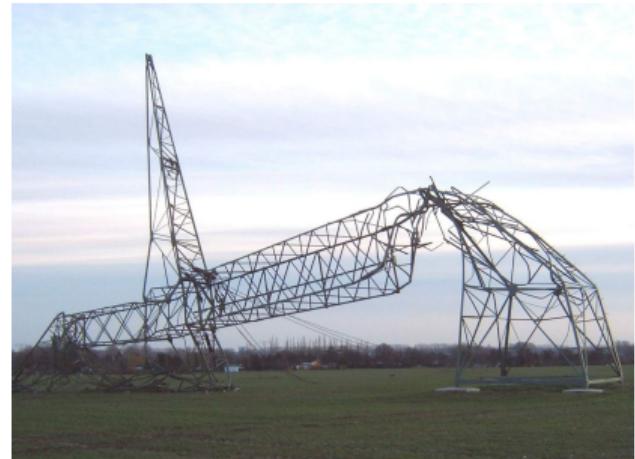
# Introduction

---

# European Windstorms

European windstorms occur during winter and hit Europe injuring the population and damaging the infrastructure.

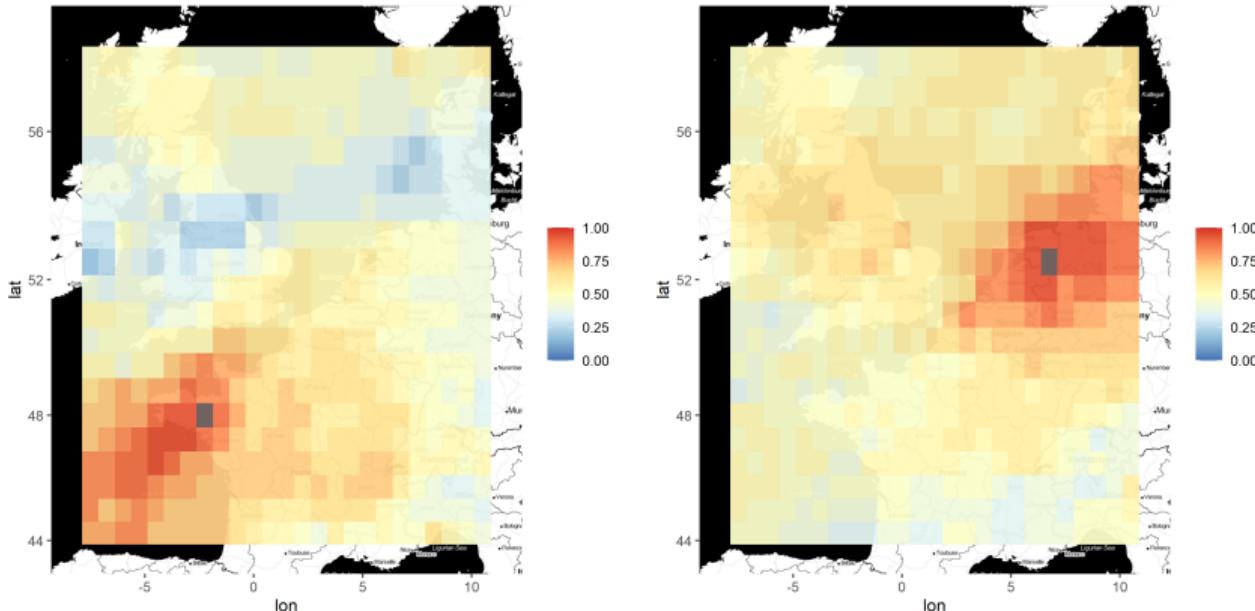
Insurance and policy makers need good understanding of extreme Windstorms patterns.



Broken pylon in Germany, Windstorm Kyrill, 23 January 2007

# What is non-stationary dependence ?

# What is non-stationary dependence ?



Two empirical extremograms at quantile  $q = 0.52$

# Goals

We use the generalized  $r$ -Pareto framework from de Fondeville & Davison (2020) combined with locally anisotropic Stochastic Partial Differential Equation (SPDE) to generate non-stationary Gaussian fields (Fuglstad et al., 2015).

The project is focused on two main goals:

- Implement non-stationary dependence with generalized r-Pareto model
- Benchmark different non-stationary models

# Goals

We use the generalized  $r$ -Pareto framework from de Fondeville & Davison (2020) combined with locally anisotropic Stochastic Partial Differential Equation (SPDE) to generate non-stationary Gaussian fields (Fuglstad et al., 2015).

The project is focused on two main goals:

- Implement non-stationary dependence with generalized r-Pareto model
- Benchmark different non-stationary models

## Peaks-over-threshold exceedances

---

## Univariate exceedances

For a random variable  $X$  in the maximum domain of attraction of a non-degenerate distribution function,

$$\Pr(X > x \mid X > u) \approx H_{\xi, \sigma, u}(x) = \begin{cases} \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}, & \xi \neq 0, \\ \exp\left(-\frac{x-u}{\sigma}\right), & \xi = 0. \end{cases}$$

## Risk functional

Let  $\mathcal{F}$  be the space of continuous function on  $S \in \mathbb{R}^2$  and  $X$  be a spatial process taking samples in  $\mathcal{F}$

We use a risk functional  $r : \mathcal{F} \rightarrow \mathbb{R}_+$ ,

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

with  $E_r$  a “risk” region in  $S$ .

## Risk functional

Let  $\mathcal{F}$  be the space of continuous function on  $S \in \mathbb{R}^2$  and  $X$  be a spatial process taking samples in  $\mathcal{F}$

We use a risk functional  $r : \mathcal{F} \rightarrow \mathbb{R}_+$ ,

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

with  $E_r$  a “risk” region in  $S$ .

## Generalized $r$ -Pareto Process

We suppose that  $X$  is regularly varying and use the approximation

$$\Pr\{X \in \cdot \mid r(X) \geq r(b)\} \approx \Pr\{P \in \cdot\},$$

where  $b$  is continuous function and  $P$  denotes a generalized  $r$ -Pareto process with tail index  $\xi$ , scale function  $a > 0$ , location  $b$  and a measure  $\Lambda$ .

## Generalized $r$ -Pareto Process

Let  $a > 0$  and  $b$  be continuous functions on  $S$  and  $\xi$  a tail index,

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

where  $Y_r$  is a stochastic process with probability measure that is function of  $\Lambda(\cdot)$ .

The reference process  $Y_r$  can be decomposed as

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

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  $S$ .

## Generalized $r$ -Pareto Process

Let  $a > 0$  and  $b$  be continuous functions on  $S$  and  $\xi$  a tail index,

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

where  $Y_r$  is a stochastic process with probability measure that is function of  $\Lambda(\cdot)$ .

The reference process  $Y_r$  can be decomposed as

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

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  $S$ .

## Generalized $r$ -Pareto Process - Marginal properties

Marginaly at location  $s \in S$ , the exceedances of  $P(s)$  follow a Generalised Pareto Distribution (GPD)

$$\Pr\{P(s) > x \mid P(s) > b(s)\} = \left\{ 1 + \xi \frac{x - b(s)}{a(s)} \right\}_+^{-1/\xi}, \quad x \geq b(s).$$

## Key Points

- The extreme events are defined using a risk functional.
- In the generalized  $r$ -Pareto process the spatial dependence is independent of the intensity of events.

The inference is done in two steps:

1. Find marginal parameters  $\xi, a, b$  using marginal properties of the process
2. Fit the dependence model for  $W_1$

## Key Points

- The extreme events are defined using a risk functional.
- In the generalized  $r$ -Pareto process the spatial dependence is independent of the intensity of events.

The inference is done in two steps:

1. Find marginal parameters  $\xi, a, b$  using marginal properties of the process
2. Fit the dependence model for  $W_1$

## Key Points

- The extreme events are defined using a risk functional.
- In the generalized  $r$ -Pareto process the spatial dependence is independent of the intensity of events.

The inference is done in two steps:

1. Find marginal parameters  $\xi, a, b$  using marginal properties of the process
2. Fit the dependence model for  $W_1$

## Key Points

- The extreme events are defined using a risk functional.
- In the generalized  $r$ -Pareto process the spatial dependence is independent of the intensity of events.

The inference is done in two steps:

1. Find marginal parameters  $\xi, a, b$  using marginal properties of the process
2. Fit the dependence model for  $W_1$

## Data preprocessing

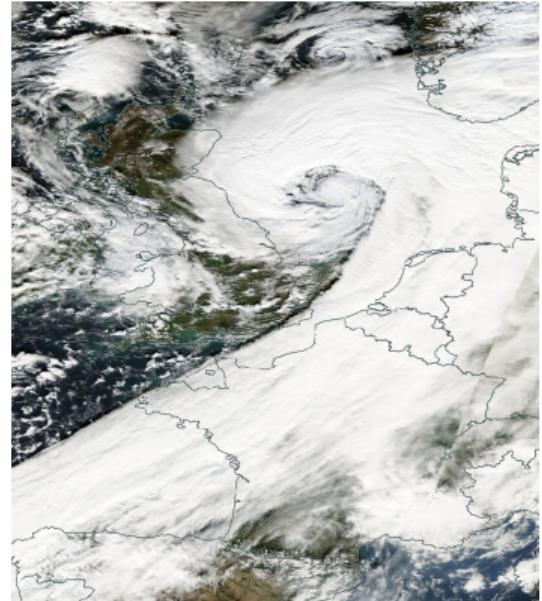
---

# Dataset

We use wind gust data from the ERA-Interim model.

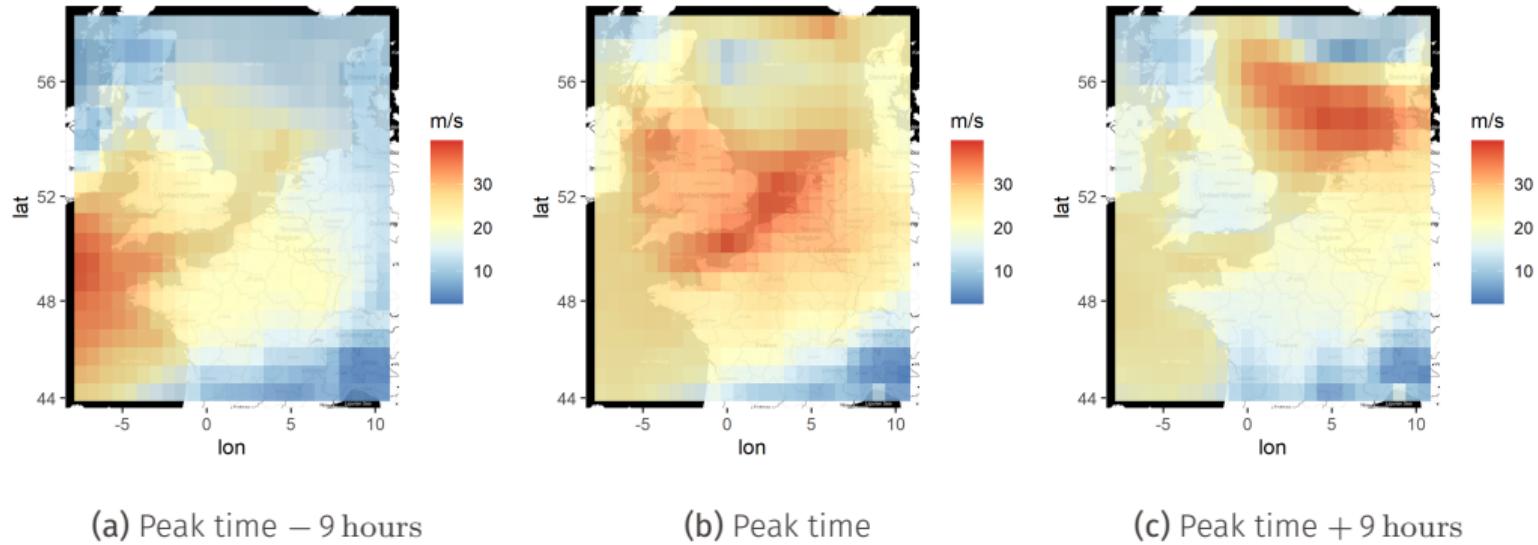
The time and space resolution is 3h and  $0.75^\circ$  latitude-longitude over 34.5 years.

475 locations on our domain.



Cyclone Oratia, 28 October 2000,  
NASA

# Storm Footprints



(a) Peak time – 9 hours

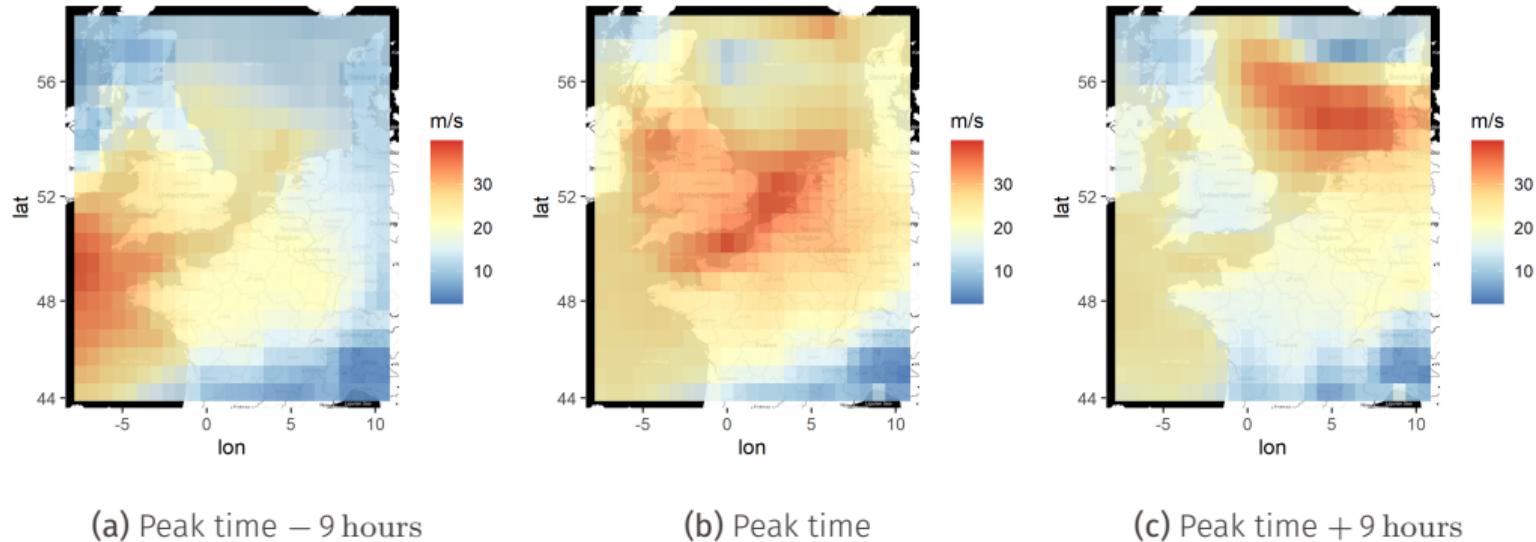
(b) Peak time

(c) Peak time + 9 hours

Wind gust speeds during windstorm Daria at different times relative to the peak risk time

We want to represent the storms using one footprint per storm.

# Storm Footprints



(a) Peak time – 9 hours

(b) Peak time

(c) Peak time + 9 hours

Wind gust speeds during windstorm Daria at different times relative to the peak risk time

We want to represent the storms using one footprint per storm.

## Storm Footprints

Wind footprint at time  $t$  is defined as

$$x(t, s_l) = \max_{|t' - t| < 18\text{h}} W_i(s_l, t'),$$

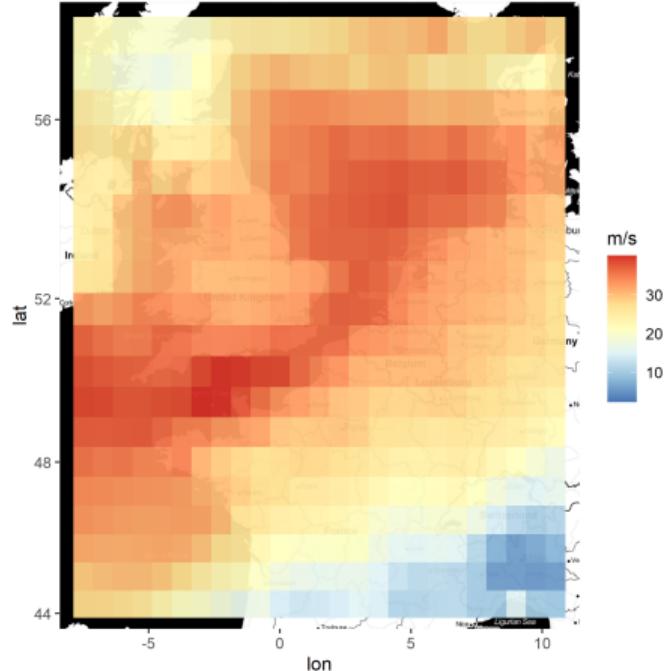
$W_i$  standing for the “instantaneous” wind data,  $s_l$  being any location in our domain.

# Storm Footprints

Wind footprint at time  $t$  is defined as

$$x(t, s_l) = \max_{|t' - t| < 18\text{h}} W_i(s_l, t'),$$

$W_i$  standing for the “instantaneous” wind data,  $s_l$  being any location in our domain.



The 36-hour footprint of Daria.

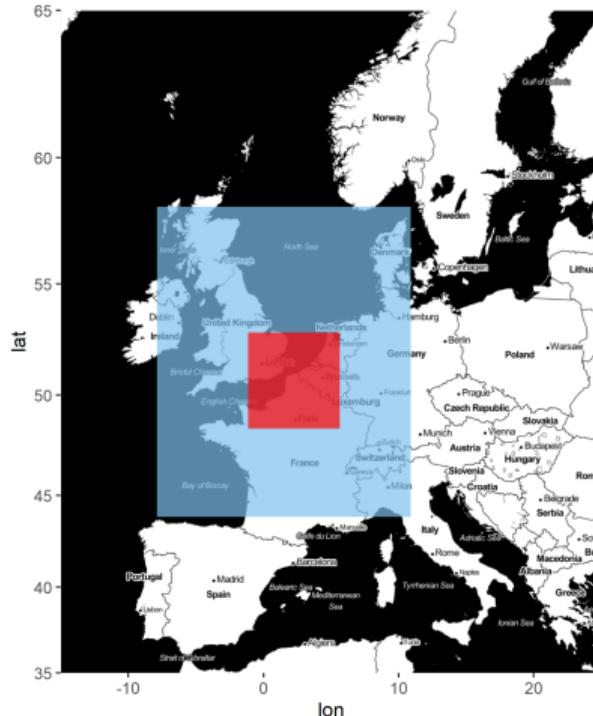
## Storm selection

The most important storms are identified using a risk functional

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

for  $x(s)$  a footprint and with  $E_r$  the “risk” region.

We select the 35 most intense footprints at time  $t_i$ , ( $i = 1, \dots, 35$ ), these are the samples we model.



Area of study in blue with the risk region in red.

## Marginal estimates

---

## Marginal estimates - $\xi, a, b$

The marginal location parameter is set as

$$b(s) = u_q(s), \quad s \in S,$$

where  $u_q(s)$  is the  $q$  quantile of footprint database  $x_i(s)$ , ( $i = 1, \dots, 35$ ).

$q = 0.19$  is chosen such that  $r(b) \leq r(x_i)$ , ( $i = 1, \dots, 35$ ).

$\xi$  and  $a(s)$  are found by fitting  $\text{GPD}(\xi, a(s), b(s))$  marginally where  $\xi$  is common to all locations.

## Marginal estimates - $\xi, a, b$

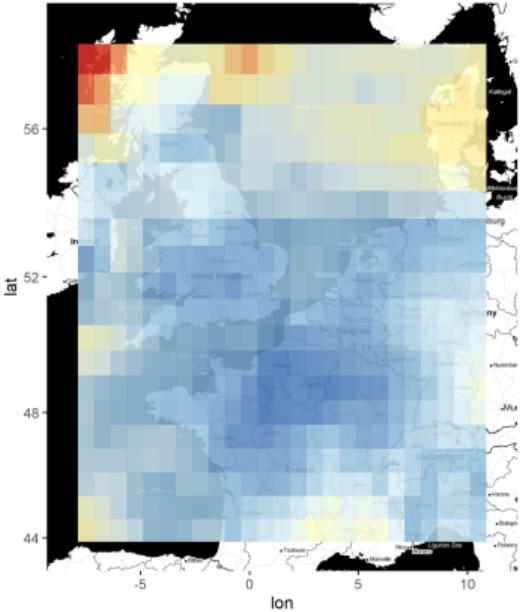
The marginal location parameter is set as

$$b(s) = u_q(s), \quad s \in S,$$

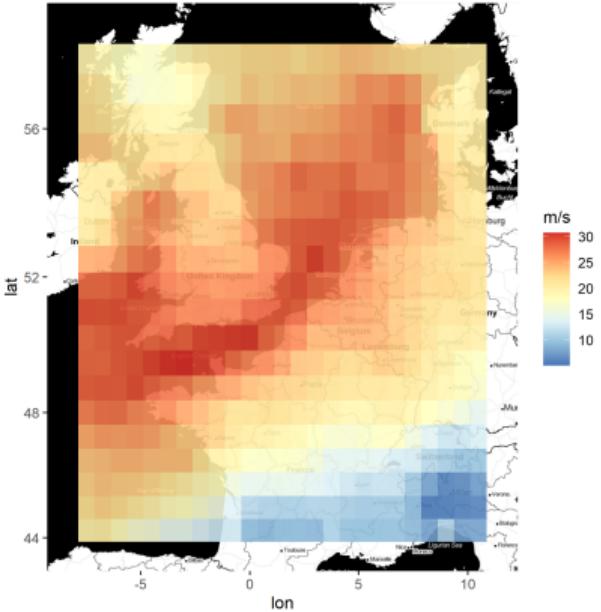
where  $u_q(s)$  is the  $q$  quantile of footprint database  $x_i(s)$ , ( $i = 1, \dots, 35$ ).

$q = 0.19$  is chosen such that  $r(b) \leq r(x_i)$ , ( $i = 1, \dots, 35$ ).

$\xi$  and  $a(s)$  are found by fitting  $\text{GPD}(\xi, a(s), b(s))$  marginally where  $\xi$  is common to all locations.



(a) Scale parameter  $a(s)$



(b) Location parameter  $b(s)$

Estimated marginal parameters  $a(s)$  and  $b(s)$ . The associated shape parameter is  $\xi = -0.59$ .

## Non-stationarity

---

## Angular process modelling

The angular process  $W_1$  is modelled using a log-Gaussian process and the covariance of the underlying Gaussian process is key to the extremal dependence structure of  $P$ .

Based on Fuglstad et al. (2015) we use a SPDE to produce a non-stationary Matérn-like Gaussian field  $u(s)$  with

$$(\kappa^2 - \nabla \cdot H(s)\nabla)u(s) = \mathcal{W}(s), \quad s \in S \subset \mathbb{R}^2$$

where  $\kappa > 0$  is a scale parameter,  $\mathcal{W}(s)$  is a spatial Gaussian white noise and  $H(s)$  is a  $2 \times 2$  matrix positive definite function.

For interpretability purpose we decompose it as

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

with  $v(s)$  a vector field on  $S$ .

## Angular process modelling

The angular process  $W_1$  is modelled using a log-Gaussian process and the covariance of the underlying Gaussian process is key to the extremal dependence structure of  $P$ .

Based on Fuglstad et al. (2015) we use a SPDE to produce a non-stationary Matérn-like Gaussian field  $u(s)$  with

$$(\kappa^2 - \nabla \cdot H(s)\nabla)u(s) = \mathcal{W}(s), \quad s \in S \subset \mathbb{R}^2$$

where  $\kappa > 0$  is a scale parameter,  $\mathcal{W}(s)$  is a spatial Gaussian white noise and  $H(s)$  is a  $2 \times 2$  matrix positive definite function.

For interpretability purpose we decompose it as

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

with  $v(s)$  a vector field on  $S$ .

## Angular process modelling

The angular process  $W_1$  is modelled using a log-Gaussian process and the covariance of the underlying Gaussian process is key to the extremal dependence structure of  $P$ .

Based on Fuglstad et al. (2015) we use a SPDE to produce a non-stationary Matérn-like Gaussian field  $u(s)$  with

$$(\kappa^2 - \nabla \cdot H(s)\nabla)u(s) = \mathcal{W}(s), \quad s \in S \subset \mathbb{R}^2$$

where  $\kappa > 0$  is a scale parameter,  $\mathcal{W}(s)$  is a spatial Gaussian white noise and  $H(s)$  is a  $2 \times 2$  matrix positive definite function.

For interpretability purpose we decompose it as

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

with  $v(s)$  a vector field on  $S$ .

## Implementation

The non-stationary implementation is based on Fuglstad et al. (2015) using sparse matrices. We use gradient scoring rule for inference on  $W_1$  (de Fondeville & Davison, 2018). The optimization is based on a Hessian method with recent parallel implementation in R, the Marquardt–Levenberg Algorithm (Philipps et al., 2020).

## Implementation

The non-stationary implementation is based on Fuglstad et al. (2015) using sparse matrices. We use gradient scoring rule for inference on  $W_1$  (de Fondeville & Davison, 2018). The optimization is based on a Hessian method with recent parallel implementation in R, the Marquardt–Levenberg Algorithm (Philipps et al., 2020).

## Implementation

The non-stationary implementation is based on Fuglstad et al. (2015) using sparse matrices. We use gradient scoring rule for inference on  $W_1$  (de Fondeville & Davison, 2018). The optimization is based on a Hessian method with recent parallel implementation in R, the Marquardt–Levenberg Algorithm (Philipps et al., 2020).

# Stationary models

## Isotropic model

The vector field  $v(s)$  is null, only the  $\gamma$  parameter is fluctuating.

## Stationary anisotropic model

The vector field  $v(s)$  is constant. This induces a constant anisotropic dependence on the whole domain.

## Polynomial model

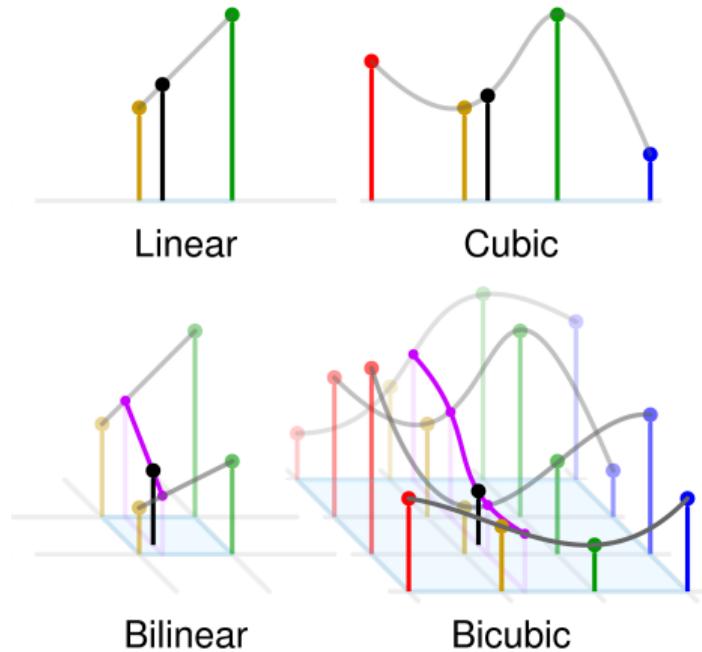
Each dimension of  $v(x, y)$  is modelled by a  $(x, y)$  polynomial of degree 2.

## Interpolation model

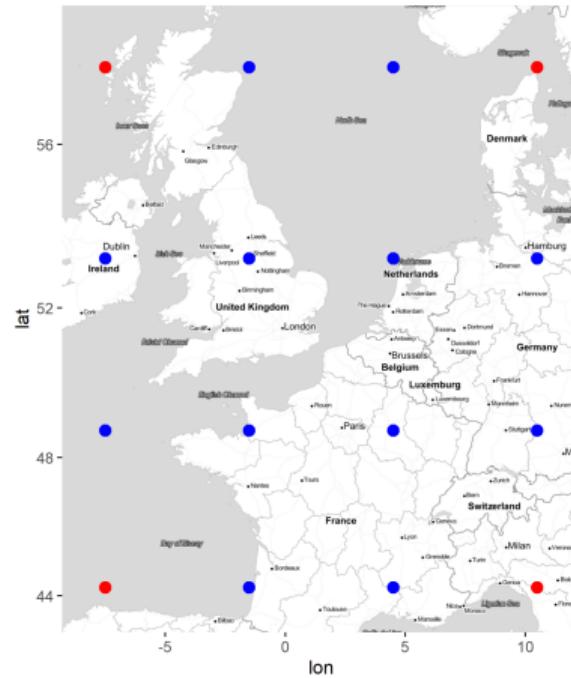
Each dimension of  $v(x, y)$  is the result of an interpolation with fixed knots on the domain  $S$ ; two interpolation methods are used:

- **Bilinear:** linear interpolation in 2D, 4 knots per dimension
- **Bicubic:** cubic interpolation in 2D, 16 knots per dimension

# Interpolation



Representation of interpolation methods.

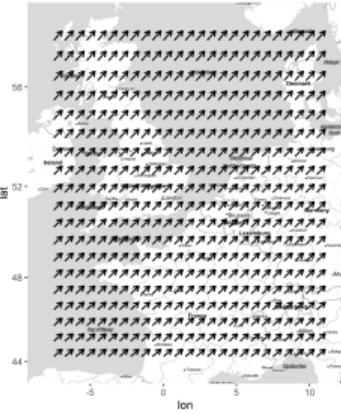


Location of the interpolation knots.

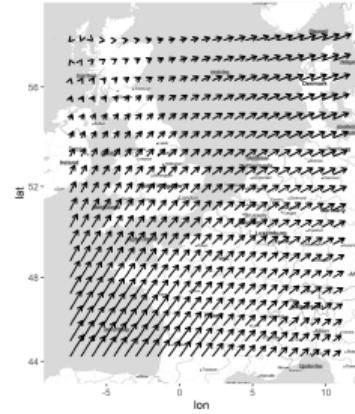
## Dependence results

---

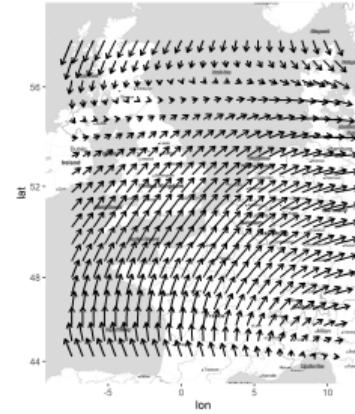
# Vector fields



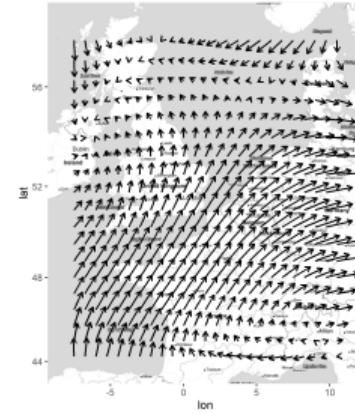
(a) Stationary anisotropy



(b) Bilinear interpolation



(c) Polynomial of order 2



(d) Bicubic interpolation

Anisotropy fields for different models

## Performance

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

**Table 1:** Specification and gradient score values for fitted models. **Lower gradient score is better.**

## Discussion

---

## Performance

- Significant improvement in the gradient score between stationary and non-stationary models.
- Very good performance of the polynomial model.
- The bicubic model seems overspecified.

## Strengths and weaknesses

- The independence of intensity and angular component does not suits the dataset.
  - The SPDE is currently constraint to a regular grid.
- + The dependence inference is robust.
  - + The combination of the SPDE approach, the gradient scoring and the optimization method is efficient.

## Strengths and weaknesses

- The independence of intensity and angular component does not suits the dataset.
  - The SPDE is currently constraint to a regular grid.
- + The dependence inference is robust.
  - + The combination of the SPDE approach, the gradient scoring and the optimization method is efficient.

## Strengths and weaknesses

- The independence of intensity and angular component does not suits the dataset.
- The SPDE is currently constraint to a regular grid.
- + The dependence inference is robust.
- + The combination of the SPDE approach, the gradient scoring and the optimization method is efficient.

## Strengths and weaknesses

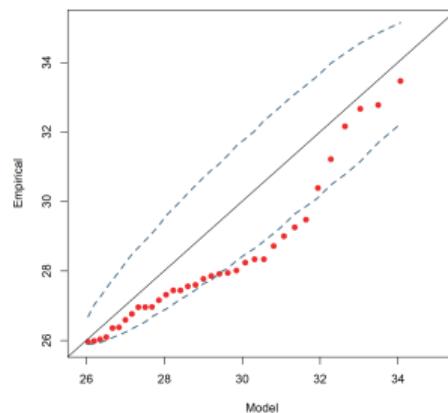
- The independence of intensity and angular component does not suits the dataset.
- The SPDE is currently constraint to a regular grid.
- + The dependence inference is robust.
- + The combination of the SPDE approach, the gradient scoring and the optimization method is efficient.

Questions?

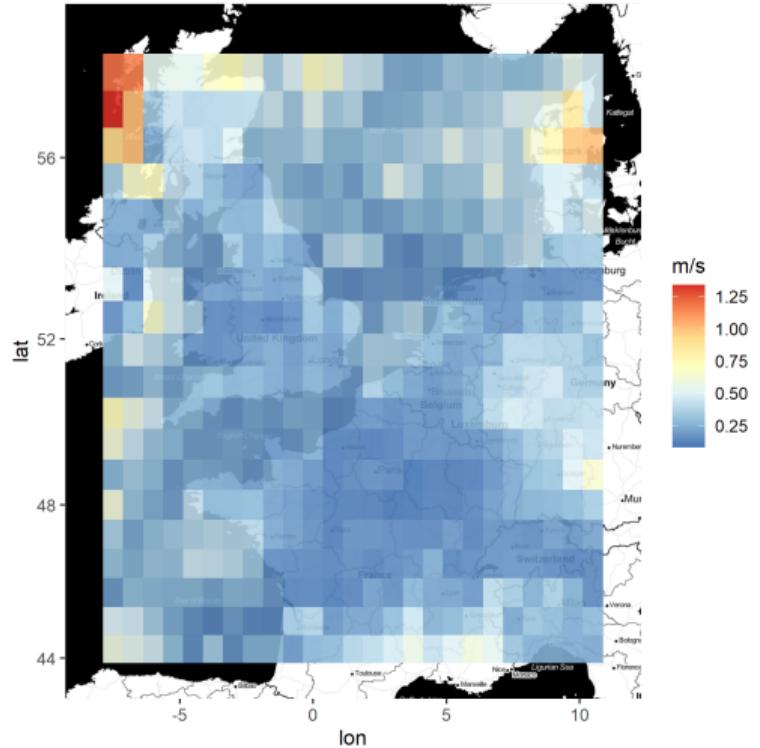
## Risk distribution

The distribution of the risk  $r(P)$  also follows a GPD

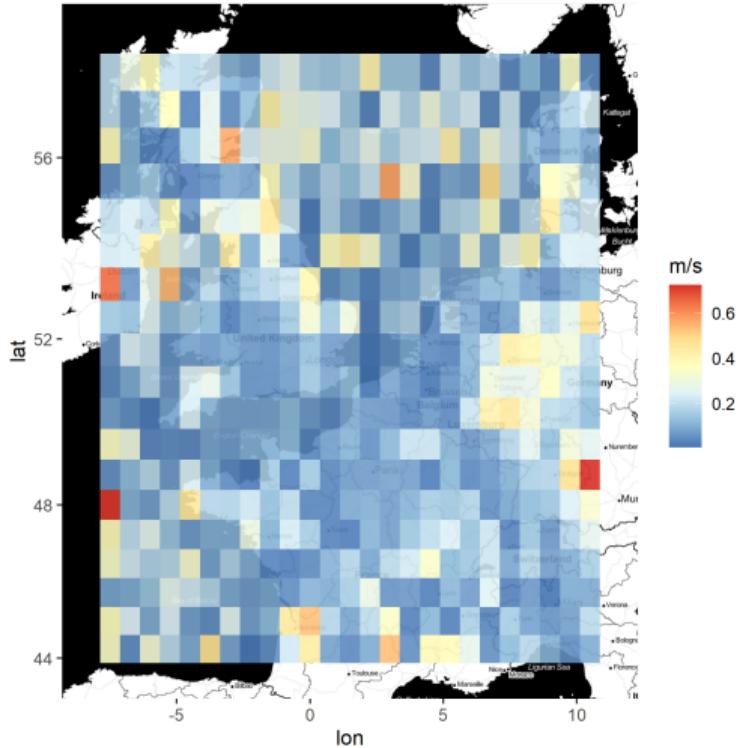
$$\Pr\{r(P) > x \mid r(P) > r(b)\} = \left\{1 + \xi \frac{x - r(b)}{r(a)}\right\}_+^{-1/\xi}, \quad x \geq r(b).$$



**Figure 6:** Storm dataset risk qq-plot using theoretical distribution obtained using marginal parameter estimates. The dashed lines are ninety-nine percent intervals computed using resampling.



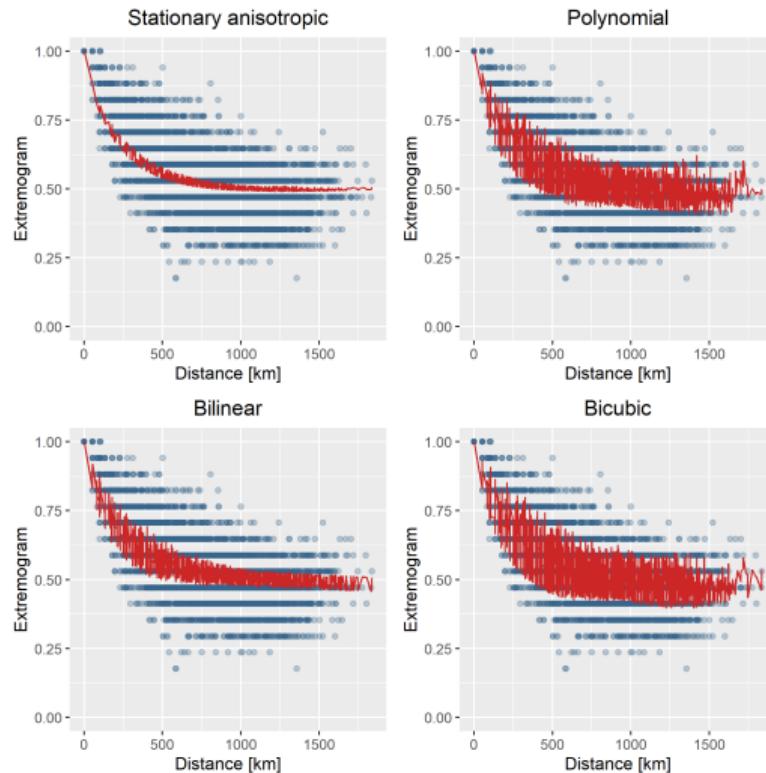
(a) Standard deviation on  $a(s)$



(b) Standard deviation on  $b(s)$

The standard deviations are obtained with jackknife leave-one-out method on the set of 35 exceedance observations. The associated shape parameter is  $\xi = -0.59_{0.02}$ .

# Extremogram comparison



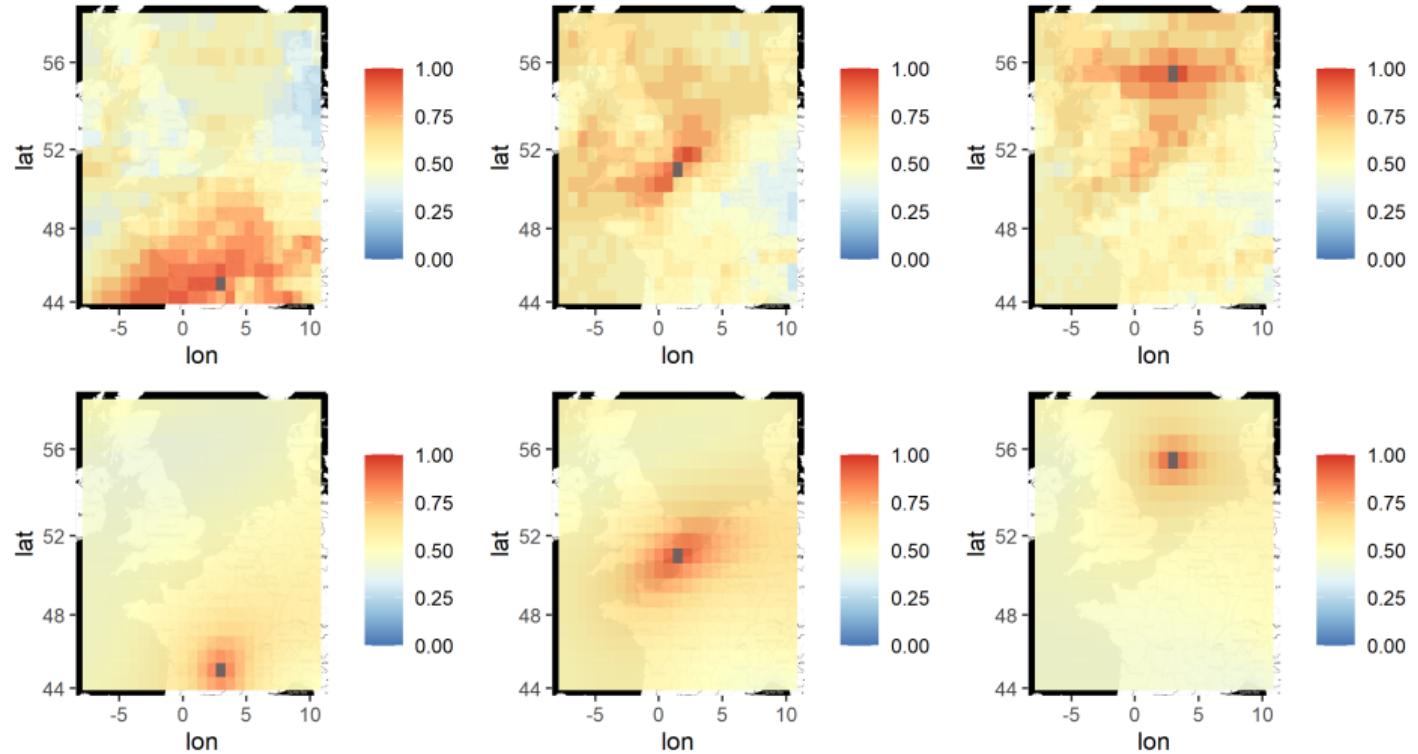
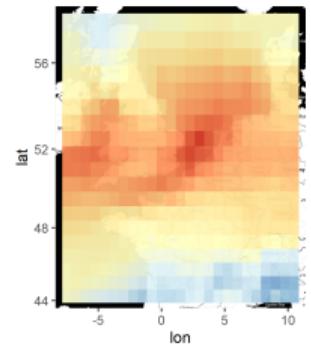
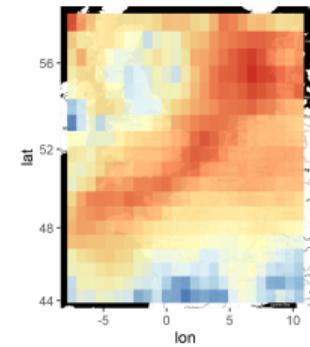


Figure 8: Extremogram comparison between  $\hat{\pi}_{0.52}(s, \cdot)$  (on top) and theoretical  $\pi(s, \cdot)$  in the polynomial model (at bottom), the three locations  $s$  are plotted in grey.

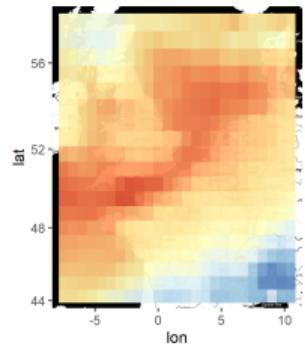
# Simulations



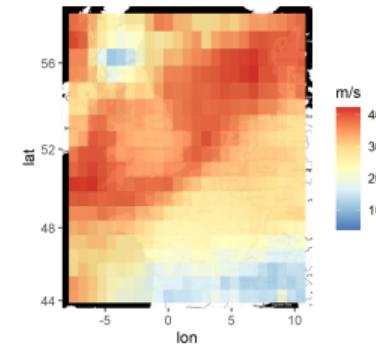
(a) Cyclone Jeanett



(b) Simulated return level  
of 17.5 years



(c) Cyclone Daria



(d) Simulated return level  
of 9 years

## References 1

- 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]* .
- 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.
- 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]* .