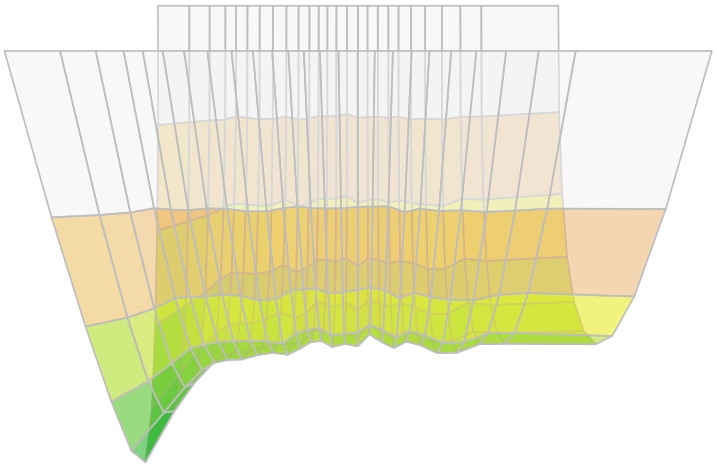
A User's Guide to the SpatialExtremes Package

Mathieu Ribatet

Chair of Statistics
École Polytechnique Fédérale de Lausanne
Switzerland



Contents

Introduction **1**

1 An Introduction to Max-Stable Processes **3**
1.1 The Smith Model . . . 3
1.2 The Schlather Model . . . 6

2 Spatial Dependence of Max-Stable Random Fields **11**
2.1 The Extremal Coefficient . . . 11
2.2 Madogram-based approaches . . . 14
2.2.1 Madogram . . . 14
2.2.2 F-Madogram . . . 16
2.2.3 λ-Madogram . . . 18

3 Fitting a Unit Fréchet Max-Stable Process to Data **21**
3.1 Least Squares . . . 21
3.2 Pairwise Likelihood . . . 22
3.2.1 Misspecification . . . 22
3.2.2 Assessing Uncertainties . . . 25

4 Model Selection **27**
4.1 Takeuchi Information Criterion . . . 27
4.2 Likelihood Ratio Statistic . . . 28
4.2.1 Adjusting the χ^2 distribution . . . 29
4.2.2 Adjusting the composite likelihood . . . 30

5 Fitting a Max-Stable Process to Data **33**
5.1 Response Surfaces . . . 34
5.1.1 Linear Regression Models . . . 35
5.1.2 Semiparametric Regression Models . . . 36
5.2 Building Response Surfaces for the GEV Parameters . . . 40

6 Conclusion **43**

A P-splines with radial basis functions **45**
A.1 Model definition . . . 45
A.2 Fast computation of p-splines . . . 46

List of Figures

1.1 Two simulations of the Smith model with different Σ matrices. Left panel: $\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 0$. Right panel: $\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 1$. 6

1.2 Plots of the Whittle–Matérn (left panel), the powered exponential (middle panel) and the Cauchy (right panel) correlation functions. The sill and the range parameters are $c_1 = c_2 = 1$ while the smooth parameters are given in the legend. . . 7

1.3 Two simulations of the Schlather model with different correlation functions having approximately the same practical range. Left panel: Whittle–Matérn with $c_1 = c_2 = \nu = 1$. Right panel: Powered exponential with $c_1 = \nu = 1$ and $c_2 = 1.5$. 8

1.4 Contour plots of an isotropic (left panel) and anisotropic (right panel) correlation function. Powered exponential family with $c_1 = c_2 = \nu = 1$. The anisotropy parameters are: $\varphi = \pi/4$, $r = 0.5$. 9

2.1 Extremal coefficient functions for different max-stable models. Σ is the 2×2 identity matrix. Correlation function: Whittle–Matérn with $c_1 = c_2 = \nu = 1$. . . 12

2.2 Pairwise extremal coefficient estimates from the Schlather and Tawn (left panel) and Smith (right panel) estimators from a simulated max-stable random field having a Whittle–Matérn correlation function - $c_1 = c_2 = \nu = 1$. The red lines are the theoretical extremal coefficient function. 13

2.3 Madogram (left panel) and binned madogram (right panel) with unit Gumbel margins for the Schlather model with the Whittle–Matérn correlation functions. The red lines are the theoretical madograms. Points are pairwise estimates. . . . 15

2.4 Pairwise madograms (left panel) and extremal coefficients (right panel) estimates from a simulated max-stable random field having a Whittle–Matérn correlation function - $c_1 = c_2 = \nu = 1$. The red lines are the theoretical madogram and extremal coefficient function. 16

2.5 Pairwise F-madogram (left panel) and extremal coefficient (right panel) estimates for the Smith model with Σ equals to the identity matrix. The red lines are the theoretical F-madogram and extremal coefficient function. 17

2.6 Binned λ-madogram estimates for two Schlather models having a powered exponential (right panel) and a Cauchy covariance (left panel) functions. $c_1 = c_2 = \nu = 1$. 20

5.1 Impact of the number of knots in the fitted p-spline. Left panel: $q = 2$, middle panel: $q = 10$, right panel: $q = 50$. The small vertical lines corresponds to the location of each knot. 37

5.2 Impact of the smoothing parameter λ on the fit. Left panel: $\lambda = 0$, middle panel: $\lambda = 0.1$ and right panel: $\lambda = 10$. 38
5.3 Cross-validation and generalized cross validation curves and corresponding fitted curves. 39

Introduction

What is the SpatialExtremes package?

The **SpatialExtremes** package is an add-on package for the R [R Development Core Team, 2007] statistical computing system. It provides functions for the analysis of spatial extremes using (currently) max-stable processes.

All comments, criticisms and queries on the package or associated documentation are gratefully received.

Obtaining the package/guide

The package can be downloaded from CRAN (The Comprehensive R Archive Network) at `http://cran.r-project.org/`. This guide (in pdf) will be in the directory `SpatialExtremes/doc/` underneath wherever the package is installed. You can get it by invoking

```
> vignette("SpatialExtremesGuide")
```

To have a quick overview of what the package does, you might want to have a look at its own web page `http://spatialextremes.r-forge.r-project.org/`.

Contents

To help users to use properly the *SpatialExtremes* packages, this report introduces all the theory and references needed. Some practical examples are inserted directly within the text to show how it works in practice. Chapter 1 is an introduction to max-stable processes and introduces several models that might be useful in concrete applications. Statistics and tools to analyze the spatial dependence of extremes are presented in Chapter 2. Chapter 3 tackles the problem of fitting max-stable process to data that are assumed to be unit Fréchet distributed. Chapter 4 is devoted to model selection. Chapter 5 presents models and procedures on how to fit max-stable processes to data that do not have unit Fréchet margins. Chapter **??** is devoted to model checking while Chapter **??** is devoted to inferential procedures and predictions. Lastly, Chapter 6 draws conclusions on spatial extremes. Note that several computations are reported in the Annex part.

Caveat

I have checked these functions as best I can but they may contain bugs. If you find a bug or suspected bug in the code or the documentation please report it to me at mathieu.ribatet@epfl.ch. Please include an appropriate subject line.

Legalese

Acknowledgements

This work has been supported by the Competence Center Environment and Sustainability `http://www.cces.ethz.ch/index` within the EXTREMES project `http://www.cces.ethz.ch/projects/hazri/EXTREMES`. I would like to thank S. A. Padoan for his support when we were both calculating the tedious partial derivatives required for the estimation of the asymptotic sandwich covariance matrix.

1

An Introduction to Max-Stable Processes

A max-stable process $Z(\cdot)$ is the limit process of maxima of independent identically distributed random fields $Y_i(x)$, $x \in \mathbb{R}^d$. Namely, for suitable $a_n(x) > 0$ and $b_n(x) \in \mathbb{R}$,

$$Z(x) = \lim_{n \to +\infty} \frac{\max_{i=1}^n Y_i(x) - b_n(x)}{a_n(x)}, \qquad x \in \mathbb{R}^d \tag{1.1}$$

Note that (1.1) does not ensure that the limit exists. However, provided it does and from (1.1), we can see that max-stable processes might be appropriate models for modelling annual maxima of spatial data, for example.

Theoretically speaking, there is no loss of generality in transforming the margins to have a unit Fréchet scale i.e.

$$\Pr\left[Z(x) \leq z\right] = \exp\left(-\frac{1}{z}\right), \qquad \forall x \in \mathbb{R}^d, \qquad z > 0 \tag{1.2}$$

and we will first assume that the unit Fréchet assumption holds for which we have $a_n(x) = n$ and $b_n(x) = 0$.

Currently, there are two different characterisations of a max-stable process. The first one, often referred to as the *rainfall-storm* model, was first introduced by Smith [1991]. More recently, Schlather [2002] introduced a new characterisation of a max-stable process allowing for a random shape. In this Chapter, we will present several max-stable models that might be relevant in studying spatial extremes.

1.1 The Smith Model

Haan [1984] was the first to proposed a characterisation of max-stable processes. Later, Smith [1991] use this characterisation to provide a parametric model for spatial extremes. The construction was the following. Let $\{(\xi_i, y_i), i \geq 1\}$ denote the points of a Poisson process on $(0, +\infty) \times \mathbb{R}^d$ with intensity measure $\xi^{-2} d\xi \nu(dy)$, where $\nu(dy)$ is a positive measure on $\mathbb{R}^d$. Then one characterisation of a max-stable process with unit Fréchet margins is

$$Z(x) = \max_i \left\{\xi_i f(y_i, x)\right\}, \qquad x \in \mathbb{R}^d \tag{1.3}$$

where $\{f(y, x), x, y \in \mathbb{R}^d\}$ is a non-negative function such that

$$\int_{\mathbb{R}^d} f(x, y)\nu(dy) = 1, \qquad \forall x \in \mathbb{R}^d$$

To see that equation (1.3) defines a stationary max-stable process with unit Fréchet margins, we have to check that the margins are indeed unit Fréchet and $Z(x)$ has the max-stable property. Following Smith, consider the set defined by:

$$E = \left\{(\xi, y) \in \mathbb{R}_*^+ \times \mathbb{R}^d : \xi f(y, x) > z\right\}$$

for a fixed location $x \in \mathbb{R}^d$ and $z > 0$. Then

$$\begin{aligned} \Pr\left[Z(x) \leq z\right] &= \Pr\left[\text{no points in } E\right] = \exp\left[-\int_{\mathbb{R}^d}\int_{z/f(y,x)}^{+\infty} \xi^{-2} d\xi \nu(\mathrm{dy})\right] \\ &= \exp\left[-\int_{\mathbb{R}^d} z^{-1} f(x, y)\nu(\mathrm{dy})\right] = \exp\left(-\frac{1}{z}\right) \end{aligned}$$

and the margins are unit Fréchet.

The max-stable property of $Z(\cdot)$ follows because the superposition of n independent, identical Poisson processes is a Poisson process with its intensity multiplied by n. More precisely, we have:

$$\left\{\max_{i=1}^{n} Z_i(x_1), \ldots, \max_{i=1}^{n} Z_i(x_k)\right\} \dot{\sim} n\left\{Z(x_1), \ldots, Z(x_k)\right\}, \qquad k \in \mathbb{N}.$$

The process defined by (1.3) is often referred to as the rainfall-storm process, as one can have a more physical interpretation of the above construction. Think of y_i as realisations of rainfall storm centres in $\mathbb{R}^d$ and $\nu(dy)$ as the spatial distribution of these storm centres over $\mathbb{R}^d$ - usually $d = 2$. Each ξ_i represents the intensity of the i-th storm and therefore $\xi_i f(y_i, x)$ represents the amount of rainfall for this specific event at location x. In other words, $f(y_i, \cdot)$ drives how the i-th storm centred at y_i diffuses in space.

Definition (1.3) is rather general and Smith considered a particular setting where $\nu(dy)$ is the Lebesgue measure and $f(y, x) = f_0(y - x)$, where $f_0(y - x)$ is a multivariate normal density with mean y and covariance matrix Σ[1]. With these additional assumptions, it can be shown that the bivariate CDF is

$$\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2] = \exp\left[-\frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_1}{z_2}\right)\right] \tag{1.4}$$

where Φ is the standard normal cumulative distribution function and, for two given locations 1 and 2

$$a^2 = \Delta x^T \Sigma^{-1} \Delta x, \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \quad \text{or} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \quad \text{and so forth}$$

where Δx is the distance vector between location 1 and location 2. Figure 1.1 plots two simulations of Smith's model with different covariance matrices.

[1] Another form of Smith's model that uses a Student distribution instead of the normal one. However, it is not currently implemented.

Proof.

$$\begin{aligned}
-\log \Pr\left[Z(x_1) \leq z_1, Z(x_2) \leq z_2\right] &= \iint_{\min\{z_1/f_0(s-x_1), z_2/f_0(s-x_2)\}} \xi^{-2} d\xi \mathrm{ds} \\
&= \int \max\left\{\frac{f_0(s-x_1)}{z_1}, \frac{f_0(s-x_2)}{z_2}\right\} \mathrm{ds} \\
&= \int \frac{f_0(s-x_1)}{z_1} \mathbb{I}\left(\frac{f_0(s-x_1)}{z_1} > \frac{f_0(s-x_2)}{z_2}\right) \mathrm{ds} \\
&+ \int \frac{f_0(s-x_2)}{z_2} \mathbb{I}\left(\frac{f_0(s-x_1)}{z_1} \leq \frac{f_0(s-x_2)}{z_2}\right) \mathrm{ds} \\
&= \int \frac{f_0(s)}{z_1} \mathbb{I}\left(\frac{f_0(s)}{z_1} > \frac{f_0(s-x_2+x_1)}{z_2}\right) \mathrm{ds} \\
&+ \int \frac{f_0(s)}{z_2} \mathbb{I}\left(\frac{f_0(s-x_1+x_2)}{z_1} \leq \frac{f_0(s)}{z_2}\right) \mathrm{ds} \\
&= \frac{1}{z_1} \mathbb{E}\left[\mathbb{I}\left(\frac{f_0(X)}{z_1} > \frac{f_0(X-x_2+x_1)}{z_2}\right)\right] \\
&+ \frac{1}{z_2} \mathbb{E}\left[\mathbb{I}\left(\frac{f_0(X-x_1+x_2)}{z_1} \leq \frac{f_0(X)}{z_2}\right)\right]
\end{aligned}$$

where X is a r.v. having f_0 as density. To get the closed form of the bivariate distribution, it remains to compute the probabilities of the event $\{f_0(X)/z_1 > f_0(X-x_2+x_1)/z_2\}$.

$$\begin{aligned}
\frac{f_0(X)}{z_1} > \frac{f_0(X-x_2+x_1)}{z_2} &\iff 2\log z_1 + X^T\Sigma^{-1}X < 2\log z_2 + (X-x_2+x_1)^T \Sigma^{-1} (X-x_2+x_1) \\
&\iff X^T\Sigma^{-1}(x_1-x_2) > \log\frac{z_1}{z_2} - \frac{1}{2}(x_1-x_2)^T\Sigma^{-1}(x_1-x_2)
\end{aligned}$$

As X has density f_0, $X^T\Sigma^{-1}(x_1-x_2)$ is normal with mean 0 and variance $a^2 = (x_1-x_2)^T\Sigma^{-1}(x_1-x_2)$. And finally, we get

$$\begin{aligned}
\frac{1}{z_1}\mathbb{E}\left[\mathbb{I}\left(\frac{f_0(X)}{z_1} > \frac{f_0(X-x_2+x_1)}{z_2}\right)\right] &= \frac{1}{z_1}\left\{1 - \Phi\left(\frac{\log z_1/z_2}{a} - \frac{a}{2}\right)\right\} \\
&= \frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{\log z_2/z_1}{a}\right)
\end{aligned}$$

and

$$\frac{1}{z_2}\mathbb{E}\left[\mathbb{I}\left(\frac{f_0(X-x_1+x_2)}{z_1} \leq \frac{f_0(X)}{z_2}\right)\right] = \frac{1}{z_2}\Phi\left(\frac{a}{2} + \frac{\log z_1/z_2}{a}\right)$$

This proves equation (1.4). □

In equation (1.4), a is the Mahanalobis distance and is similar to the Euclidean distance except that it gives different weights to each component of Δx. It is positive and the limiting cases $a \to 0^+$ and $a \to +\infty$ correspond respectively to perfect dependence and independence. Therefore, for Σ fixed, the dependence decreases monotically and continuously as $||\Delta x||$ increases. On the contrary, if Δx is fixed, the dependence decreases monotically as a increases.

The covariance matrix Σ plays a major role in equation (1.4) as it determines the shape of the storm events. Indeed, due the use of a multivariate normal distribution, the storm events have an elliptical shape. Considering the eigen-decomposition of Σ, we can write

$$\Sigma = U\Lambda U^T, \tag{1.5}$$

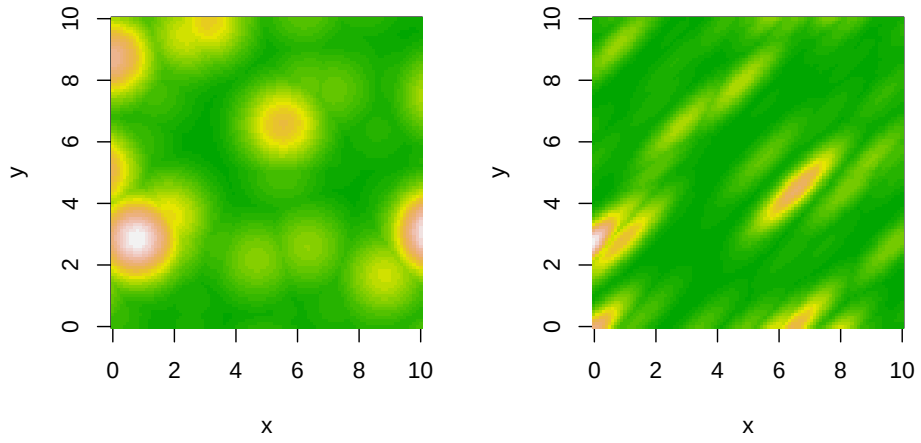


Figure 1.1: *Two simulations of the Smith model with different* Σ *matrices. Left panel:* $\sigma_{11} = \sigma_{22} = 9/8$ *and* $\sigma_{12} = 0$. *Right panel:* $\sigma_{11} = \sigma_{22} = 9/8$ *and* $\sigma_{12} = 1$.

where U is a rotation matrix and Λ is a diagonal matrix of the eigenvalues. Thus, U controls the direction of the principal axes and Λ controls their lengths.

If Σ is diagonal and all the diagonal terms are identical, then $\Sigma = \Lambda$, so that the ellipsoids change to circles and model (1.4) becomes isotropic. Figure 1.1 is a nice illustration of this. The left panel corresponds to an isotropic random field while the right one depicts a clear anisotropy for which we have

$$\Sigma = \begin{bmatrix} 9/8 & 1 \\ 1 & 9/8 \end{bmatrix} = \begin{bmatrix} \cos(-3\pi/4) & \sin(-3\pi/4) \\ -\sin(-3\pi/4) & \cos(-3\pi/4) \end{bmatrix} \begin{bmatrix} 1/8 & 0 \\ 0 & 17/8 \end{bmatrix} \begin{bmatrix} \cos(-3\pi/4) & -\sin(-3\pi/4) \\ \sin(-3\pi/4) & \cos(-3\pi/4) \end{bmatrix},$$

so that the main direction of the major principal axis is $\pi/4$ and a one unit move along the direction $-\pi/4$ yields the same decrease in dependence as 17 unit moves along the direction $\pi/4$.

1.2 The Schlather Model

More recently, Schlather [2002] introduced a second characterisation of max-stable processes. Let $Y(\cdot)$ be a stationary process on $\mathbb{R}^d$ such that $\mathbb{E}[\max\{0, Y(x)\}] = 1$ and $\{\xi_i, i \geq 1\}$ be the points of a Poisson process on $\mathbb{R}_*^+$ with intensity measure $\xi^{-2}d\xi$. Then Schlather shows that a stationary max-stable process with unit Fréchet margins can be defined by:

$$Z(x) = \max_i \xi_i \max\{0, Y_i(x)\} \tag{1.6}$$

where the $Y_i(\cdot)$ are i.i.d copies of $Y(\cdot)$.

As before, the max-stable property of $Z(\cdot)$ stems from the superposition of n independent, identical Poisson processes, while the unit Fréchet margins holds by the same argument as for the Smith model. Indeed, let consider the following set:

$$E = \left\{(\xi, y(x)) \in \mathbb{R}_*^+ \times \mathbb{R}^d : \xi \max(0, y(x)) > z\right\}$$

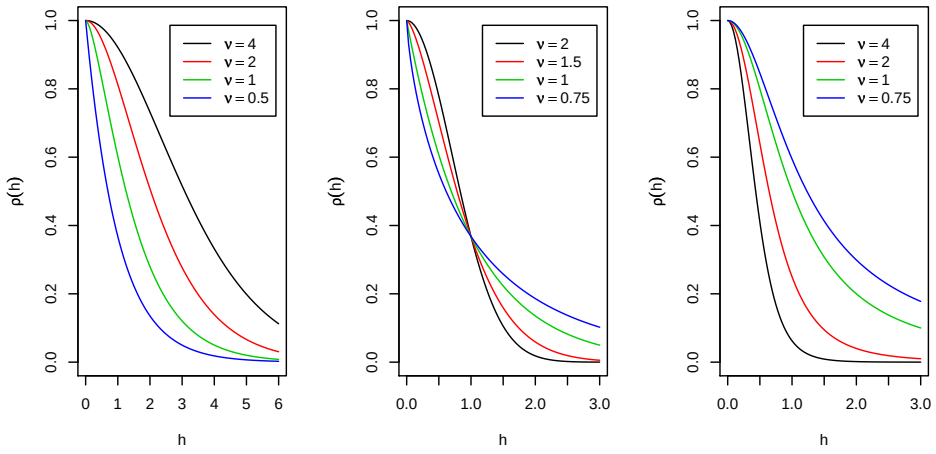


Figure 1.2: *Plots of the Whittle–Matérn (left panel), the powered exponential (middle panel) and the Cauchy (right panel) correlation functions. The sill and the range parameters are $c_1 = c_2 = 1$ while the smooth parameters are given in the legend.*

for a fixed location $x \in \mathbb{R}^d$ and $z > 0$. Then

$$\begin{aligned}\Pr[Z(x) \leq z] &= \Pr[\text{no points in } E] = \exp\left[-\int_{\mathbb{R}^d}\int_{z/\max(0,y(x))}^{+\infty} \xi^{-2} d\xi \nu(dy(x))\right] \\ &= \exp\left[-\int_{\mathbb{R}^d} z^{-1}\max\{0, y(x)\}\nu(dy(x))\right] = \exp\left(-\frac{1}{z}\right)\end{aligned}$$

As with the Smith model, the process defined in equation (1.6) has a practical interpretation. Think of $\xi_i Y_i(\cdot)$ as the daily spatial rainfall events so that all these events have the same spatial dependence structure but differ only in their magnitude ξ_i. This model differs slightly from Smith's one as we now have no deterministic shape such as a multivariate normal density for the storms but a random shape driven by the process $Y(\cdot)$.

The Schlather and Smith characterisations have strong connections. To see this, let consider the case for which $Y_i(x) = f_0(x - X_i)$ where f_0 is a probability density function and $\{X_i\}$ is a homogeneous Poisson process both in $\mathbb{R}^d$. With this particular setting, model (1.6) is identical to model (1.3).

Equation (1.6) is very general and we need additional assumptions to get practical models. Schlather proposed to take $Y_i(\cdot)$ to be a stationary Gaussian process with correlation function $\rho(x)$, scaled so that $\mathbb{E}[\max\{0, Y_i(x)\}] = 1$. With these new assumtions, it can be shown that the bivariate CDF of process (1.6) is

$$\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2] = \exp\left[-\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \sqrt{1 - 2(\rho(h)+1)\frac{z_1 z_2}{(z_1+z_2)^2}}\right)\right] \quad (1.7)$$

where $h \in \mathbb{R}^+$ is the Euclidean distance between location 1 and location 2. Usually, $\rho(h)$ is chosen from one of the valid parametric families, such as

Whittle–Matérn	$\rho(h) = c_1 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{h}{c_2}\right)^{\nu} K_\nu\left(\frac{h}{c_2}\right)$,	$0 \leq c_1 \leq 1,\ c_2 > 0,\ \nu > 0$
Cauchy	$\rho(h) = c_1\left[1 + \left(\frac{h}{c_2}\right)^2\right]^{-\nu}$,	$0 \leq c_1 \leq 1,\ c_2 > 0,\ \nu > 0$
Powered Exponential	$\rho(h) = c_1 \exp\left[-\left(\frac{h}{c_2}\right)^{\nu}\right]$,	$0 \leq c_1 \leq 1,\ c_2 > 0,\ 0 < \nu \leq 2$

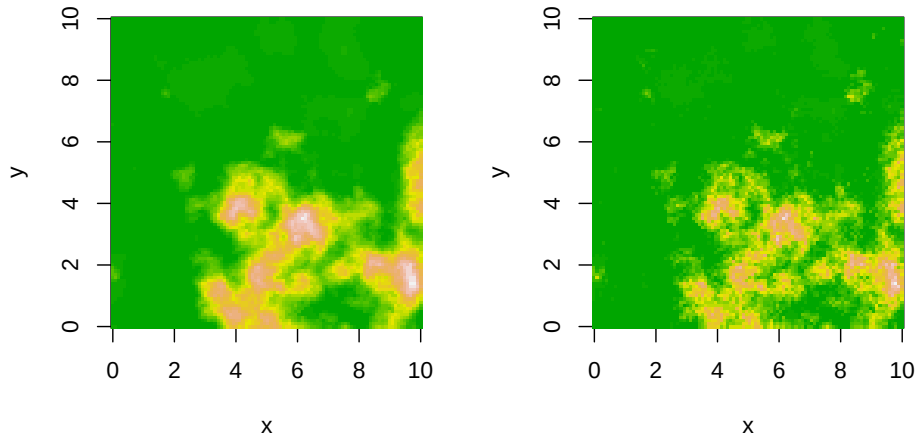


Figure 1.3: *Two simulations of the Schlather model with different correlation functions having approximately the same practical range. Left panel: Whittle–Matérn with* $c_1 = c_2 = \nu = 1$. *Right panel: Powered exponential with* $c_1 = \nu = 1$ *and* $c_2 = 1.5$.

where c_1, c_2 and ν are the sill, the range and the smooth parameters of the correlation function, Γ is the gamma function and K_ν is the modified Bessel function of the third kind with order ν. Figure 1.2 plots the correlation functions for the parametric families introduced above. The left panel was generated with the following lines

```
> covariance(sill = 1, range = 1, smooth = 4, cov.mod = "whitmat",
+     xlim = c(0, 6), ylim = c(0, 1))
> covariance(sill = 1, range = 1, smooth = 2, cov.mod = "whitmat",
+     add = TRUE, col = 2, xlim = c(0, 6))
> covariance(sill = 1, range = 1, smooth = 1, cov.mod = "whitmat",
+     add = TRUE, col = 3, xlim = c(0, 6))
> covariance(sill = 1, range = 1, smooth = 0.5, cov.mod = "whitmat",
+     col = 4, add = TRUE, xlim = c(0, 6))
> legend("topright", c(expression(nu == 4), expression(nu == 2),
+     expression(nu == 1), expression(nu == 0.5)), col = 1:4, lty = 1,
+     inset = 0.05)
```

Figure 1.3 plots two realisations of the Schlather model with the powered exponential and Whittle–Matérn correlation functions. It can be seen that the powered exponential model leads to more rough random fields as, with this particular setting for the covariance parameters, the slope of the powered exponential correlation function near the origin is steeper than the Whittle–Matérn.

The correlation functions introduced above are all isotropic, but model (1.7) doesn't require this assumption. From a valid correlation function ρ it is always possible to get an elliptical correlation function ρ_e by using the following transformation:

$$\rho_e(\Delta x) = \rho\left(\sqrt{\Delta x^T A \Delta x}\right) \tag{1.8}$$

where Δx is the distance vector between two stations, A is any positive semi-definite matrix that may involve additional parameters. For example, if the spatial domain belongs to $\mathbb{R}^2$, a convenient parametrization for A is given by

$$A = \begin{bmatrix} \cos\varphi & r^2 \sin\varphi \\ r^2 \sin\varphi & \cos\varphi \end{bmatrix}$$

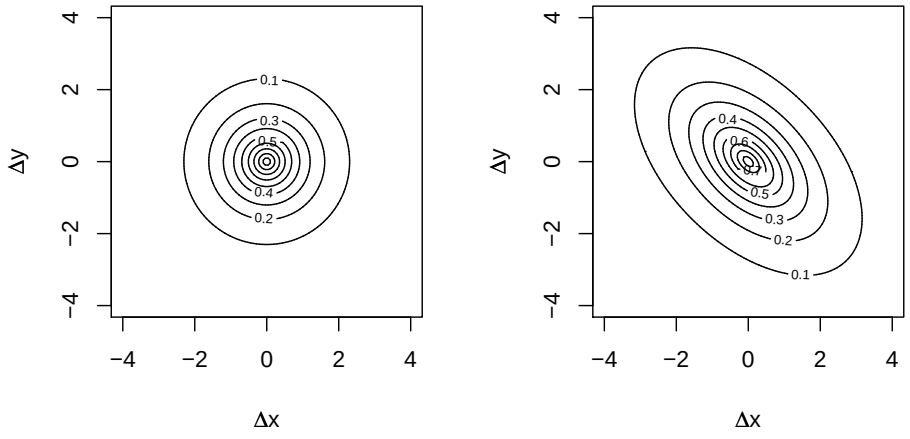


Figure 1.4: *Contour plots of an isotropic (left panel) and anisotropic (right panel) correlation function. Powered exponential family with* $c_1 = c_2 = \nu = 1$. *The anisotropy parameters are:* $\varphi = \pi/4$, $r = 0.5$.

where $\varphi \in [0, \pi)$ is the rotation angle and $0 < r < 1$ is the ratio between the minor and major principal axes of the ellipse. Figure 1.4 plots the contour of an isotropic correlation function and an anistropic one derived from equation (1.8).

The correlation coefficient $\rho(h)$ can take any value in $[-1, -1]$. Complete dependence is reached when $\rho(h) = 1$ while independence occurs when $\rho(h) = -1$. However, most parametric correlation families don't allow negative values so that independence is never reached. The next two sections propose two alternatives based on characterisation (1.6) that bypass this hurdle.

2

Spatial Dependence of Max-Stable Random Fields

Sooner or later, statistical modellers will be interested in knowing how evolves dependence in space. When dealing with non-extremal data, a common tools is the (semi-)variogram γ [Cressie, 1993]. Let $Y(\cdot)$ be a stationary gaussian process with correlation function ρ and variance σ^2. It is well known that $Y(\cdot)$ is fully characterized by its mean and its covariance. Consequently, the variogram defined as

$$\gamma(x_1 - x_2) = \frac{1}{2}\text{Var}\left[Y(x_1) - Y(x_2)\right] = \sigma^2\left\{1 - \rho(x_1 - x_2)\right\} \tag{2.1}$$

determines the degree of spatial dependence of $Y(\cdot)$.

When extreme values are of interest, the variogram is no longer a useful tool, as it may not even exist. Therefore, there is a need to develop more suitable tools for analyzing the spatial dependence of max-stable fields. In this chapter, we will present the extremal coefficient as a measure of the degree of dependence for extreme values, and variogram-based approaches that are especially well adapted to extremes.

2.1 The Extremal Coefficient

Let $Z(\cdot)$ be a stationary max-stable random field with unit Fréchet margins. The extremal dependence among N fixed locations in $\mathbb{R}^d$ can be summarised by the extremal coefficient, which is defined as:

$$\Pr\left[Z(x_1) \leq z, \ldots, Z(x_N) \leq z\right] = \exp\left(-\frac{\theta_N}{z}\right) \tag{2.2}$$

where $1 \leq \theta_N \leq N$ with the lower and upper bounds corresponding to complete dependence and independence and thus provides a measure of the degree of spatial dependence between stations. Following this idea, θ_N can be thought as the effective number of independent stations.

Given the properties of the max-stable process with unit Fréchet marings, the finite-dimensional CDF belongs to the class of multivariate extreme value distributions

$$\Pr\left[Z(x_1) \leq z_1, \ldots, Z(x_N) \leq z_N\right] = \exp\left\{-V(z_1, \ldots, z_N)\right\} \tag{2.3}$$

where V is a homogeneous function of order -1 called the exponent measure [Pickands, 1981; Coles, 2001]. As a consequence, the homogeneity property of V implies a strong relationship between the exponent measure and the extremal coefficient

$$\theta_N = V(1, \ldots, 1) \tag{2.4}$$

An important special case of equation (2.2) is to consider pairwise extremal coefficients, that is

$$\Pr\left[Z(x_1) \leq z, Z(x_2) \leq z\right] = \exp\left\{-\frac{\theta(x_1 - x_2)}{z}\right\} \tag{2.5}$$

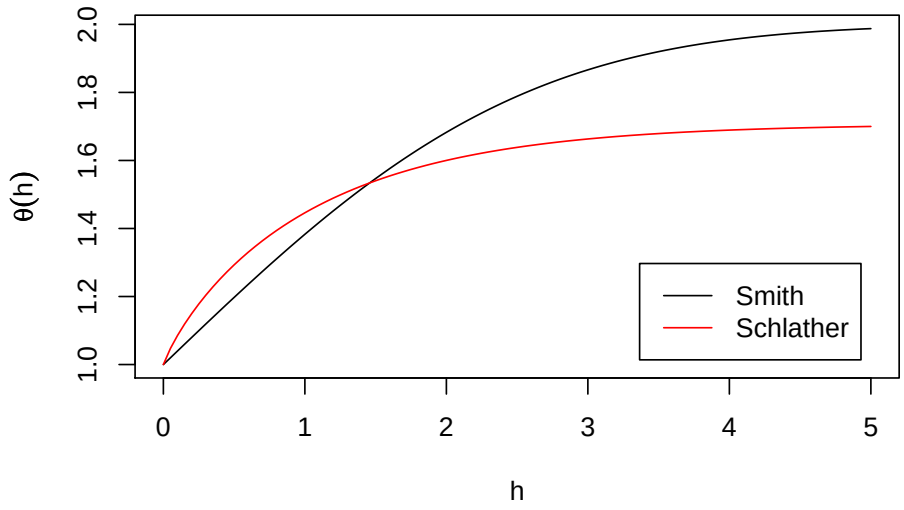


Figure 2.1: *Extremal coefficient functions for different max-stable models. Σ is the 2×2 identity matrix. Correlation function: Whittle–Matérn with $c_1 = c_2 = \nu = 1$.*

Following Schlather and Tawn [2003], $\theta(\cdot)$ is called the extremal coefficient function; it provides sufficient information about extremal dependence for many problems although it does not characterise the full distribution.

The extremal coefficient functions for max-stable models presented in Chapter 1 can be derived directly from their bivariate distribution by letting $z_1 = z_2 = z$. More precisely, we have:

Smith $\quad \theta(x_1 - x_2) = 2\Phi\left(\frac{\sqrt{(x_1-x_2)^T\Sigma^{-1}(x_1-x_2)}}{2}\right)$

Schlather $\quad \theta(x_1 - x_2) = 1 + \sqrt{\frac{1-\rho(x_1-x_2)}{2}}$

Figure 2.1 plots the extremal coefficient function for the different max-stable models introduced. The Smith model covers the whole range of dependence while Schlather's model has an upper bound of $1 + \sqrt{1/2}$.

Schlather and Tawn [2003] prove several interesting properties for $\theta(\cdot)$:

1. $2 - \theta(\cdot)$ is a semi-definite positive function;

2. $\theta(\cdot)$ isn't differentiable at 0 unless $\theta \equiv 1$;

3. if $d \geq 2$ and $Z(\cdot)$ is isotropic, $\theta(\cdot)$ has at most one jump at 0 and is continuous elsewhere.

These properties have strong consequences. The first indicates that the spatial dependence structure of a stationary max-stable process can be characterised by a correlation function. The second states that a valid correlation functions does not always lead to a valid extremal coefficient function. For instance, the Gaussian correlation model, $\rho(h) = \exp(-h^2)$, $h \geq 0$, is not allowed since it is differentiable at $h = 0$.

Equation (2.2) forms the basis for a simple maximum likelihood estimator. Suppose we have $Z_1(\cdot), \ldots, Z_n(\cdot)$, independent replications of $Z(\cdot)$ observed at a set $A = \{x_1, \ldots, x_{|A|}\}$ of

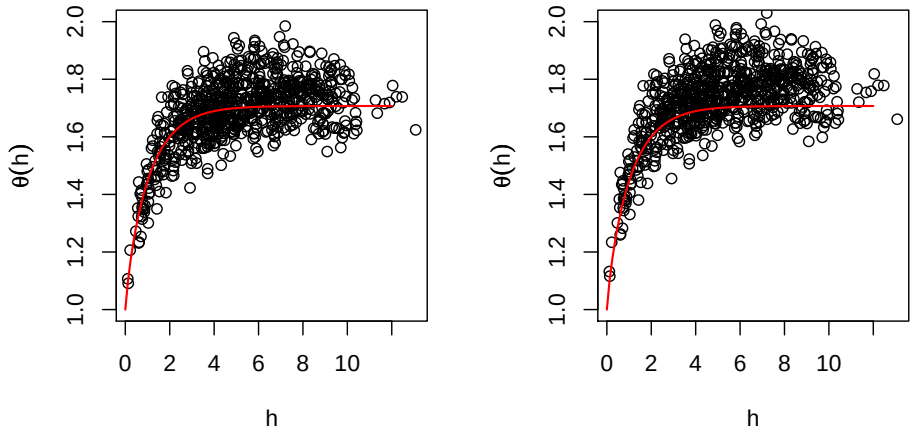


Figure 2.2: *Pairwise extremal coefficient estimates from the Schlather and Tawn (left panel) and Smith (right panel) estimators from a simulated max-stable random field having a Whittle–Matérn correlation function - $c_1 = c_2 = \nu = 1$. The red lines are the theoretical extremal coefficient function.*

locations. The log-likelihood based on equation (2.2) is given by:

$$\ell_A(\theta_A) = n \log \theta_A - \theta_A \sum_{i=1}^{n} \frac{1}{\max_{j \in A} \left\{ Z_i(x_j) \overline{Z(x_j)} \right\}} \tag{2.6}$$

where the terms of the form $\log \max_{j \in A}\{Z_i(x_j)\}$ were omitted and $\overline{Z(x_j)} = n^{-1} \sum_{i=1}^{n} 1/Z_i(x_j)$. The scalings by $\overline{Z(x_j)}$ are here to ensure that $\hat{\theta}_A = 1$ when $|A| = 1$.

The problem with the MLE based on equation (2.6) is that the extremal coefficient estimates may not have the properties on the extremal coefficient function stated above. To solve this, Schlather and Tawn [2003] propose self consistent estimators for θ_A by either sequentially correcting the estimates obtained by equation (2.6) or by modifying the log-likelihood to ensure these properties.

Smith [1991] proposed another estimator for the pairwise extremal coefficients. As $Z(x)$ is unit Fréchet for all $x \in \mathbb{R}^d$, $1/Z(x)$ has a unit exponential distribution and, according to equation (2.5), $1/\max\{Z(x_1), Z(x_2)\}$ has an exponential distribution with rate $\theta(x_1 - x_2)$. By the law of large numbers $\sum_{i=1}^{n} 1/Z_i(x_1) = \sum_{i=1}^{n} 1/Z_i(x_2) \approx n$[1], this suggests a simple estimator for the extremal coefficient between locations x_1 and x_2:

$$\hat{\theta}(x_1 - x_2) = \frac{n}{\sum_{i=1}^{n} \min\{Z_i(x_1)^{-1}, Z_i(x_2)^{-1}\}} \tag{2.7}$$

Figure 2.2 plots the pairwise extremal coefficient estimates from a simulated Schlather model having a Whittle–Matérn correlation function using equations (2.6) and (2.7). This figure was generated using the following code:

```
> n.site <- 40
> n.obs <- 100
> x <- runif(n.site, 0, 10)
> y <- runif(n.site, 0, 10)
```

[1] In fact, these relations are exact if the margins were transformed to unit Fréchet by using the maximum likelihood estimates.

```
> set.seed(12)
> ms0 <- MaxStableRF(x, y, grid = FALSE, model = "wh", param = c(0,
+     1, 0, 1, 1), maxstable = "extr", n = n.obs)
> ms0 <- t(ms0)
> par(mfrow = c(1, 2))
> fitextcoeff(ms0, cbind(x, y), loess = FALSE)
> fitextcoeff(ms0, cbind(x, y), estim = "Smith", loess = FALSE)
```

2.2 Madogram-based approaches

As we already stated, variograms are useful quantities to assess the degree of spatial dependence for Gaussian processes. However their use for extreme observations is limited as variograms may not exist. To see this, consider a stationary max-stable process with unit Fréchet margins. For such random processes, both mean and variance are not finite and the variogram does not exist theoretically, so we need extra work to get reliable variogram-based tools.

2.2.1 Madogram

A standard tool in geostatistics, similar to the variogram, is the madogram [Mathéron, 1987]. The madogram is

$$\nu(x_1 - x_2) = \frac{1}{2}\mathbb{E}\left[|Z(x_1) - Z(x_2)|\right], \tag{2.8}$$

where $Z(\cdot)$ is a stationary random field with mean assumed finite.

The problem with the madogram is almost identical to the one we emphasized with the variogram as unit Fréchet random variables have no finite mean. This led Cooley et al. [2006] to consider identical GEV margins with shape parameter $\xi < 1$ to ensure that the mean, and even the variance, are finite. It is possible to use the same strategy to ensure that the variogram exists theoretically but, as we will show later, we will see that working with the madogram is particularly suited for extreme values and has strong links with the extremal coefficient.

By using simple arguments and some results obtained by Hosking et al. [1985] on probability weighted moments, Cooley et al. [2006] show that

$$\theta(x_1 - x_2) = \begin{cases} u_\beta\left(\mu + \frac{\nu(x_1-x_2)}{\Gamma(1-\xi)}\right), & \xi < 1, \xi \neq 0, \\ \exp\left(\frac{\nu(x_1-x_2)}{\sigma}\right), & \xi = 0, \end{cases} \tag{2.9}$$

where μ, σ and ξ are the location, scale and shape parameters for the GEV distribution, $\Gamma(\cdot)$ is the Gamma function and

$$u_\beta(x) = \left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{1/\xi}$$

equation (2.8) suggests a simple estimator

$$\hat{\nu}(x_1 - x_2) = \frac{1}{2n}\sum_{i=1}^{n}|z_i(x_1) - z_i(x_2)| \tag{2.10}$$

where $z_i(x_1)$ and $z_i(x_2)$ are the i-th observations of the random field at location x_1 and x_2 and n is the total number of observations. If isotropy is assumed, then it might be preferable to use a "binned" version of estimator (2.10)

$$\hat{\nu}(h) = \frac{1}{2n|B_h|}\sum_{(x_1,x_2)\in B_h}\sum_{i=1}^{n}|z_i(x_1) - z_i(x_2)| \tag{2.11}$$

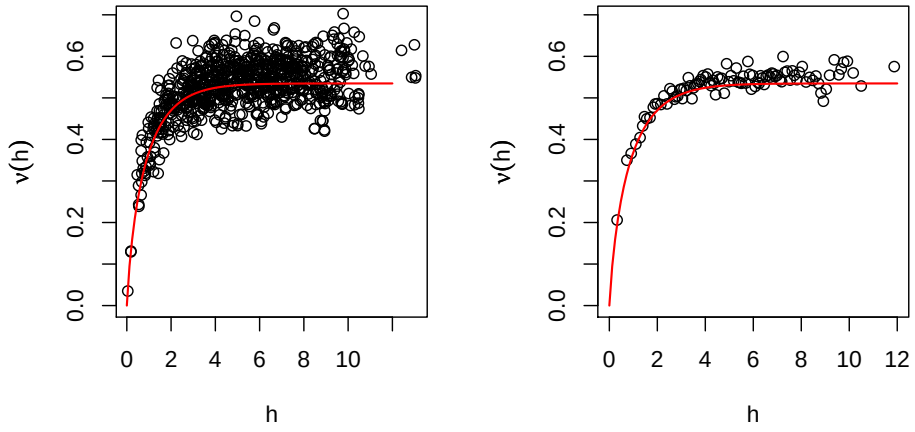


Figure 2.3: *Madogram (left panel) and binned madogram (right panel) with unit Gumbel margins for the Schlather model with the Whittle–Matérn correlation functions. The red lines are the theoretical madograms. Points are pairwise estimates.*

where B_h is the set of pair of locations whose pairwise distances belong to $[h-\epsilon, h+\epsilon[$, for $\epsilon > 0$.

Figure 2.3 plots the theoretical madograms for the Schlather's model having a Whittle–Matérn correlation function. Pairwise and binned pairwise estimates as defined by equations (2.10) ans (2.11) are also reported. The code used to generate these madogram estimates was

```
> n.site <- 50
> n.obs <- 100
> x <- runif(n.site, 0, 10)
> y <- runif(n.site, 0, 10)
> set.seed(12)
> ms0 <- MaxStableRF(x, y, grid = FALSE, model = "wh", param = c(0,
+     1, 0, 1, 1), maxstable = "extr", n = n.obs)
> ms0 <- t(ms0)
> par(mfrow = c(1, 2))
> madogram(ms0, cbind(x, y), which = "mado")
> madogram(ms0, cbind(x, y), which = "mado", n.bins = 100)
```

Using a plugin estimate in equation (2.9) leads to a simple estimator for $\theta(\cdot)$:

$$\hat{\theta}(x_1 - x_2) = \begin{cases} u_\beta \left(\mu + \frac{\hat{\nu}(x_1 - x_2)}{\Gamma(1-\xi)} \right), & \xi < 1, \xi \neq 0, \\ \exp\left(\frac{\hat{\nu}(x_1 - x_2)}{\sigma} \right), & \xi = 0, \end{cases} \tag{2.12}$$

Figure 2.4 plots the madogram and extremal coefficient functions estimated from a simulated max-stable process with unit Fréchet margins and having a Whittle–Matérn correlation function. These estimates were obtained by using equations (2.10) and (2.12) respectively. The figure was generated using the code below.

```
> n.site <- 40
> n.obs <- 100
> x <- runif(n.site, 0, 10)
```

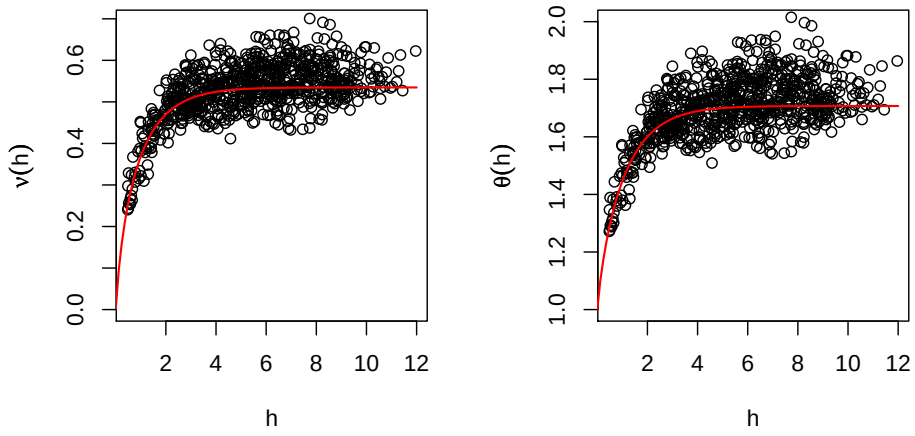


Figure 2.4: *Pairwise madograms (left panel) and extremal coefficients (right panel) estimates from a simulated max-stable random field having a Whittle–Matérn correlation function - $c_1 = c_2 = \nu = 1$. The red lines are the theoretical madogram and extremal coefficient function.*

```
> y <- runif(n.site, 0, 10)
> set.seed(12)
> ms0 <- MaxStableRF(x, y, grid = FALSE, model = "wh", param = c(0,
+     1, 0, 1, 1), maxstable = "extr", n = n.obs)
> ms0 <- t(ms0)
> madogram(ms0, cbind(x, y))
```

2.2.2 F-Madogram

In the previous subsection, we introduced the madogram as a summary statistic for the spatial dependence structure. However, the choice of the GEV parameters to compute this madogram is somewhat arbitrary. Instead, Cooley et al. [2006] propose a modified madogram called the F-madogram

$$\nu_F(x_1 - x_2) = \frac{1}{2}\mathbb{E}\left[\left|F\left(Z(x_1)\right) - F\left(Z(x_2)\right)\right|\right] \tag{2.13}$$

where $Z(\cdot)$ is a stationary max-stable random field with unit Fréchet margins and $F(z) = \exp(-1/z)$.

The F-madogram is well defined even in the presence of unit Fréchet margins as we work with $F(Z(x_1))$ instead of $Z(x_1)$. Obviously, equation (2.13) suggests a simple estimator:

$$\hat{\nu}_F(x_1 - x_2) = \frac{1}{2n}\sum_{i=1}^{n}\left|\hat{F}(z_i(x_1)) - \hat{F}(z_i(x_2))\right| \tag{2.14}$$

where $z_i(x_1)$ and $z_i(x_2)$ are the i-th observations of the random field at location x_1 and x_2 and n is the total number of observations.

The F-madogram has strong connections with the extremal coefficient introduced in Section 2.1. Indeed, we have

$$2\nu_F(x_1 - x_2) = \frac{\theta(x_1 - x_2) - 1}{\theta(x_1 - x_2) + 1} \tag{2.15}$$

or conversely

$$\theta(x_1 - x_2) = \frac{1 + 2\nu_F(x_1 - x_2)}{1 - 2\nu_F(x_1 - x_2)} \tag{2.16}$$

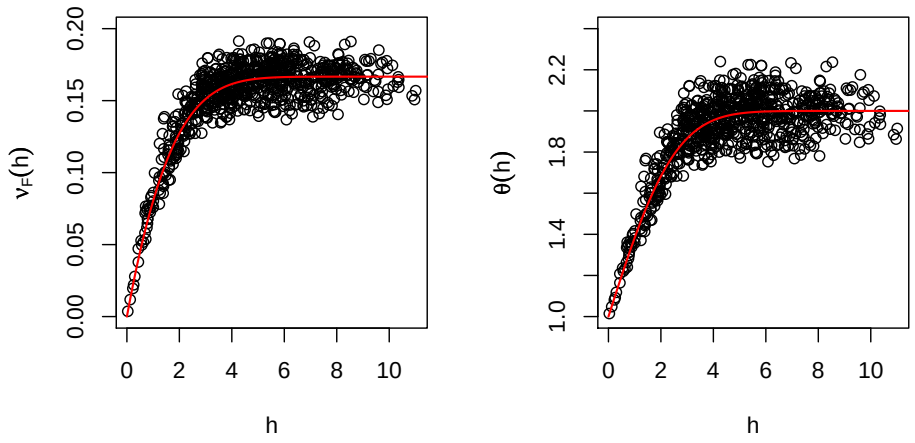


Figure 2.5: *Pairwise F-madogram (left panel) and extremal coefficient (right panel) estimates for the Smith model with Σ equals to the identity matrix. The red lines are the theoretical F-madogram and extremal coefficient function.*

Proof. Let first note that

$$|x - y| = 2\max\{x, y\} - (x + y) \tag{2.17}$$

Using equation (2.17) in equation (2.13), we have:

$$\begin{aligned}
\nu_F(x_1 - x_2) &= \frac{1}{2}\mathbb{E}\left[\left|F\left(Z(x_1)\right) - F\left(Z(x_2)\right)\right|\right] \\
&= \mathbb{E}\left[\max\left\{F\left(Z(x_1)\right), F\left(Z(x_2)\right)\right\}\right] - \mathbb{E}\left[F\left(Z(x_1)\right)\right] \\
&= \mathbb{E}\left[F\left(\max\{Z(x_1), Z(x_2)\}\right)\right] - \frac{1}{2}
\end{aligned}$$

where we used the fact that $F(Z(x_1))$ is uniformly distributed on $[0, 1]$ and F is monotone increasing. Now, from Section 2.1, we know that

$$\Pr\left[\max\left\{Z(x_1), Z(x_2)\right\} \leq z\right] = \exp\left(-\frac{\theta(x_1 - x_2)}{z}\right)$$

so that

$$\begin{aligned}
\mathbb{E}\left[F\left(\max\{Z(x_1), Z(x_2)\}\right)\right] &= \theta(x_1 - x_2)\int_0^{+\infty} \exp\left(-\frac{1}{z}\right)\exp\left(-\frac{\theta(x_1 - x_2)}{z}\right) z^{-2}\mathrm{dz} \\
&= \frac{\theta(x_1 - x_2)}{\theta(x_1 - x_2) + 1}
\end{aligned}$$

This proves equations (2.15) and (2.16). □

As we can see from equation (2.16), there is a one-to-one relationship between the extremal coefficient and the F-madogram. This suggests a simple estimator for $\theta(x_1 - x_2)$

$$\hat{\theta}(x_1 - x_2) = \frac{1 + 2\hat{\nu}_F(x_1 - x_2)}{1 - 2\hat{\nu}_F(x_1 - x_2)} \tag{2.18}$$

Figure 2.5 plots the pairwise F-madogram and extremal coefficient estimates from 100 replications of the isotropic Smith model. The code used to produce this figure was:

```
> n.site <- 40
> n.obs <- 100
> x <- runif(n.site, 0, 10)
> y <- runif(n.site, 0, 10)
> set.seed(12)
> sigma <- matrix(c(1, 0, 0, 1), ncol = 2)
> sigma.inv <- solve(sigma)
> sqrtCinv <- t(chol(sigma.inv))
> model <- list(list(model = "gauss", var = 1, aniso = sqrtCinv/2))
> set.seed(12)
> ms0 <- MaxStableRF(x, y, grid = FALSE, model = model, maxstable = "Bool",
+     n = n.obs)
> ms0 <- t(ms0)
> par(mfrow = c(1, 2))
> fmadogram(ms0, cbind(x, y))
```

As with the madogram presented in the previous Section, it is also possible to have binned estimates of the F-madogram by passing the argument `n.bins` into the `fmadogram` function.

2.2.3 λ-Madogram

The extremal coefficient, and thus the (F-)madogram, does not fully characterize the spatial dependence of a random field. Indeed, from equation (2.2), it only considers $\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2]$ where $z_1 = z_2 = z$. To solve this issue, Naveau et al. [2009] introduce the λ-madogram defined as

$$\nu_\lambda(x_1 - x_2) = \frac{1}{2}\mathbb{E}\left[\left|F^\lambda\{Z(x_1)\} - F^{1-\lambda}\{Z(x_2)\}\right|\right] \tag{2.19}$$

for any $\lambda \in [0, 1]$.

The idea beyond this is that by varying λ, we will focus on $\Pr[Z(x_1) \leq z_1, Z(x_2) \leq z_2]$ where $z_1 = \lambda z$ and $z_2 = (1-\lambda)z$ and thus explore the whole space. The λ-madogram is related to the exponent measure V, namely

$$\nu_\lambda(x_1 - x_2) = \frac{V_{x_1,x_2}(\lambda, 1-\lambda)}{1 + V_{x_1,x_2}(\lambda, 1-\lambda)} - c(\lambda) \tag{2.20}$$

where $c(\lambda) = 3/\{2(1+\lambda)(2-\lambda)\}$.

Proof. Applying equation (2.17) with $x = F^\lambda\{Z(x_1)\}$ and $y = F^{1-\lambda}\{Z(x_2)\}$, we have

$$\begin{aligned}
\nu_\lambda(x_1 - x_2) &= \mathbb{E}\left[\max\{F^\lambda\{Z(x_1)\}, F^{1-\lambda}\{Z(x_2)\}\}\right] - \mathbb{E}\left[F^\lambda\{Z(x_1)\}\right] - \mathbb{E}\left[F^{1-\lambda}\{Z(x_2)\}\right] \\
&= \mathbb{E}\left[\max\{F^\lambda\{Z(x_1)\}, F^{1-\lambda}\{Z(x_2)\}\}\right] - \frac{1}{2(1+\lambda)} - \frac{1}{2(2-\lambda)}
\end{aligned}$$

where we used the fact that $\mathbb{E}[X^k] = 1/(1+k)$, $X \sim U(0,1)$, $k > 0$. From Section 2.1 we know that

$$\begin{aligned}
\Pr\left[\max\{F^\lambda\{Z(x_1)\}, F^{1-\lambda}\{Z(x_2)\}\} \leq z\right] &= \Pr\left[Z(x_1) \leq -\frac{\lambda}{\log z}, Z(x_2) \leq -\frac{1-\lambda}{\log z}\right] \\
&= \exp\left\{-\log(z) V_{x_1,x_2}(\lambda, 1-\lambda)\right\}
\end{aligned}$$

where V_{x_1,x_2} is the homogeneous function of order -1 introduced in equation (2.3). Differentiating this distribution with respect to z gives the probability density function of the random variable $\max\{F^\lambda\{Z(x_1)\}, F^{1-\lambda}\{Z(x_2)\}\}$, so that we have

$$\begin{aligned}\mathbb{E}\left[\max\{F^\lambda\{Z(x_1)\}, F^{1-\lambda}\{Z(x_2)\}\}\right] &= \int_0^1 -V_{x_1,x_2}(\lambda, 1-\lambda)\exp\left\{-\log(z)V_{x_1,x_2}(\lambda, 1-\lambda)\right\}\mathrm{d}z \\ &= \frac{V_{x_1,x_2}(\lambda, 1-\lambda)}{1+V_{x_1,x_2}(\lambda, 1-\lambda)}\end{aligned}$$

and finally

$$\nu_\lambda(x_1 - x_2) = \frac{V_{x_1,x_2}(\lambda, 1-\lambda)}{1+V_{x_1,x_2}(\lambda, 1-\lambda)} - c(\lambda)$$

where $c(\lambda) = 3/\{2(1+\lambda)(2-\lambda)\}$. □

Again there is a one-to-one relationship between ν_λ and the dependence measure, so that we can express V_{x_1,x_2} as a function of ν_λ

$$V_{x_1,x_2}(\lambda, 1-\lambda) = \frac{c(\lambda) + \nu_\lambda(x_1 - x_2)}{1 - c(\lambda) - \nu_\lambda(x_1 - x_2)} \tag{2.21}$$

As $V_{x_1,x_2}(0.5, 0.5) = 2\theta(x_1 - x_2)$, the previous equation induces that the madogram and the F-madogram are special cases of the λ-madogram when $\lambda = 0.5$. For instance, we have

$$\nu_{0.5}(x_1 - x_2) = \frac{8\nu_F(x_1 - x_2)}{3\{3 + 2\nu_F(x_1 - x_2)\}}$$

Equation (2.19) suggests a simple estimator

$$\hat{\nu}_\lambda(x_1 - x_2) = \frac{1}{2n}\sum_{i=1}^{n} |\hat{F}^\lambda\{z_i(x_1)\} - \hat{F}^{1-\lambda}\{z_i(x_2)\}| \tag{2.22}$$

where $z_i(x_1)$ and $z_i(x_2)$ are the i-th observations of the random field at location x_1 and x_2, n is the total number of observations and $\hat{F}$ is an estimate of the CDF at the specified location.

There is an issue with the previous estimator. Indeed, the boundary conditions for the λ-madogram when $\lambda = 0$ or $\lambda = 1$ are not always fulfilled. From equation (2.19), if $\lambda = 0$ or $\lambda = 1$, $\nu_\lambda(x_1 - x_2) = 1/4$. If F is estimated by MLE, then this condition fails while if $\hat{F}(x_{i:n}) = i/(n+1)$ we get the required conditions. To bypass this hurdle, Naveau et al. [2009] propose the following adjusted estimator

$$\begin{aligned}\hat{\nu}_\lambda(x_1 - x_2) &= \frac{1}{2n}\sum_{i=1}^{n} |\hat{F}^\lambda\{z_i(x_1)\} - \hat{F}^{1-\lambda}\{z_i(x_2)\}| - \frac{\lambda}{2n}\sum_{i=1}^{n}\left[1 - \hat{F}^\lambda\{z_i(x_1)\}\right] \\ &- \frac{1-\lambda}{2n}\sum_{i=1}^{n}\left[1 - \hat{F}^{1-\lambda}\{z_i(x_2)\}\right] + \frac{1 - \lambda + \lambda^2}{2(2-\lambda)(1+\lambda)}\end{aligned}$$

By plug in this estimator into equation (2.21) we get an estimator for the dependence measure

$$\hat{V}_{x_1,x_2}(\lambda, 1-\lambda) = \frac{c(\lambda) + \hat{\nu}_\lambda(x_1 - x_2)}{1 - c(\lambda) - \hat{\nu}_\lambda(x_1 - x_2)} \tag{2.23}$$

Figure 2.6 plots the binned λ-madogram estimates from 100 replications of the Schlather model having a powered exponential and a Cauchy correlation functions. The code used to produce this figure was:

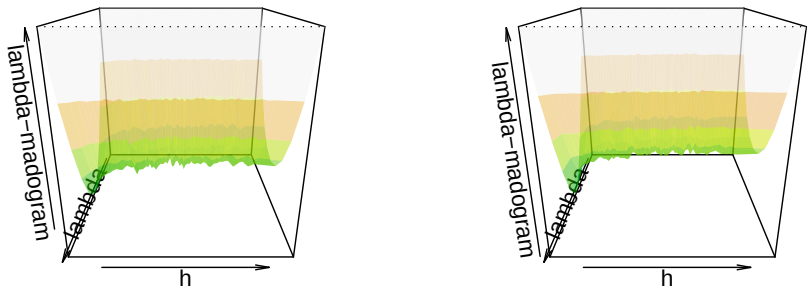


Figure 2.6: *Binned λ-madogram estimates for two Schlather models having a powered exponential (right panel) and a Cauchy covariance (left panel) functions.* $c_1 = c_2 = \nu = 1$.

```
> n.site <- 40
> n.obs <- 100
> x <- runif(n.site, 0, 10)
> y <- runif(n.site, 0, 10)
> set.seed(12)
> ms0 <- MaxStableRF(x, y, grid = FALSE, model = "stable", param = c(0,
+     1, 0, 1, 1), maxstable = "extr", n = n.obs)
> ms0 <- t(ms0)
> set.seed(12)
> ms1 <- MaxStableRF(x, y, grid = FALSE, model = "cauchy", param = c(0,
+     1, 0, 1, 1), maxstable = "extr", n = n.obs)
> ms1 <- t(ms1)
> par(mfrow = c(1, 2))
> lmadogram(ms0, cbind(x, y), n.bins = 60)
> lmadogram(ms1, cbind(x, y), n.bins = 60)
```

It might be useful to use the excellent *persp3d* function provided by the contributed *rgl* R package to explore dynamically the λ-madogram.

3

Fitting a Unit Fréchet Max-Stable Process to Data

In Chapter 1, we described max-stable processes and gave two different characterisations of them. However, we mentioned that only the bivariate distributions are analytically known so that the fitting of max-stable processes to data is not straightforward. In this Chapter, we will present two different approaches to fitting max-stable processes to data. The first one is based on least squares and was first introduced by Smith [1991]. The second one uses the maximum composite likelihood estimator [Lindsay, 1988], more precisely the maximum pairwise likelihood estimator. We will consider these two different approaches separately.

3.1 Least Squares

As stated by equations (1.4) and (1.7), the density of a max-stable process is analytically known only for the bivariate case so that maximum likelihood estimators are not available. This observation led Smith [1991] to propose an estimator based on least squares. This fitting procedure consists in minimizing the objective function

$$C(\psi) = \sum_{i<j} \left(\frac{\theta_{i,j} - \tilde{\theta}_{i,j}}{s(\tilde{\theta}_{i,j})} \right)^2 \tag{3.1}$$

where ψ is the vector parameter of the max-stable process, $\theta_{i,j}$ is the extremal coefficient predicted from the max-stable model for stations i and j , $\tilde{\theta}_{i,j}$ is a semi-parametric estimator defined by equation (2.7) for stations i and j and $s(\tilde{\theta}_{i,j})$ is the standard deviation related to the estimation of $\tilde{\theta}_{i,j}$, estimated by the jacknife estimator [Efron, 1982].

S.J. Neil, in his M.Phil. thesis, suggests the use of this weighted sum of squares criterion to avoid unsatisfactory fits in regions of high dependence i.e. when $\theta_{i,j}$ is close to 1.

Although Smith [1991] proposed this estimator for his own max-stable model, there is no restriction in applying it to any of the max-stable models introduced in Chapter 1. An illustration of this fitting procedure is given by the following lines:

```
> n.site <- 40
> n.obs <- 80
> locations <- matrix(runif(2 * n.site, 0, 10), ncol = 2)
> colnames(locations) <- c("lon", "lat")
> sigma <- matrix(c(9/8, 1, 1, 9/8), ncol = 2)
> sigma.inv <- solve(sigma)
> sqrtCinv <- t(chol(sigma.inv))
> model <- list(list(model = "gauss", var = 1, aniso = sqrtCinv/2))
> set.seed(12)
```

```
> ms0 <- MaxStableRF(locations[, 1], locations[, 2], grid = FALSE,
+     model = model, maxstable = "Bool", n = n.obs)
> ms0 <- t(ms0)
> fitcovmat(ms0, locations, marge = "emp")

        Estimator: Least Square
            Model: Smith
  Objective Value: 753.475
Covariance Family: Gaussian

Estimates
  Marginal Parameters:
  Not estimated.
  Dependence Parameters:
 cov11   cov12   cov22
0.8122  0.7810  0.9580

Optimization Information
  Convergence: successful
  Function Evaluations: 168
```

This approach suffers from two major drawbacks. First, unless we use Monte-Carlo techniques, standard errors are not available and because the observations are far from being normal, the least squares estimator should be far from efficiency in the way given by the Cramér-Rao lower bound [Cramér, 1946] and the Gauss-Markov theorem. Secondly, for concrete analysis, it is hopeless that the observed (spatial) annual maxima have unit Fréchet margins so that we need first to transform the data to the unit Fréchet scale. This suggests the use of a more flexible estimator.

3.2 Pairwise Likelihood

As already stated, the "full" likelihood for the max-stable model presented in Chapter 1 is not analytically known if the number of stations under consideration is greater or equal to three. However, as the bivariate density is analytically known, this suggests the use of the pairwise likelihood in place of the full likelihood. The log pairwise-likelihood is given by

$$\ell_p(\mathbf{z};\psi) = \sum_{i<j}\sum_{k=1}^{n_{i,j}} \log f(z_k^{(i)}, z_k^{(j)};\psi) \tag{3.2}$$

where $\mathbf{z}$ is the data available on the whole region, $n_{i,j}$ is the number of common observations between sites i and j, $y_k^{(i)}$ is the k-th observation of the i-th site and $f(\cdot,\cdot)$ is the bivariate density of the unit Fréchet max-stable process.

3.2.1 Misspecification

Properties of the maximum composite likelihood estimator are well known [Lindsay, 1988; Cox and Reid, 2004] and belong to the class of maximum likelihood estimation under misspecification[1].

[1] More rigorously we should say *partially specified*.

A statistical model $\{f(y;\psi), \psi \in \mathbb{R}^p\}$ is said *misspecified* if the observations y_i, $i = 1, \ldots, n$ are drawn from a unknown true density g instead of f. We say that the model $\{f(y;\psi), \psi \in \mathbb{R}^p\}$ is *correct* if there exists $\psi_* \in \mathbb{R}^d$ such that $f(y;\psi_*) = g(y)$, for all y.

Let $\hat{\psi}$ be the maximum likelihood estimator. Because of the law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n} \log f\left(y_i;\hat{\psi}\right) \longrightarrow \int \log f\left(y;\psi_g\right) g(y)\mathrm{d}y, \qquad n \to +\infty \tag{3.3}$$

where ψ_g is the value that minimizes the Kullback–Leibler discrepancy defined as

$$D\left(f_\psi, g\right) = \int \log\left(\frac{g(y)}{f(y;\psi)}\right) g(y)\mathrm{d}y \tag{3.4}$$

By definition of $\hat{\psi}$, we have:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(y_i;\hat{\psi})}{\partial \psi} = 0$$

so that, provided the log-likelihood is regular enough, a Taylor expansion about ψ_g yields

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(y_i;\psi_g)}{\partial \psi} + \left(\hat{\psi} - \psi_g\right)^T \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log f(y_i;\psi_g)}{\partial \psi \partial \psi^T} \doteq 0 \\ \iff \quad &\hat{\psi} \doteq \psi_g - \left\{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log f(y_i;\psi_g)}{\partial \psi \partial \psi^T}\right\}^{-1} \left\{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(y_i;\psi_g)}{\partial \psi}\right\} \end{aligned}$$

It can be shown using the central limit theorem and the weak law of large numbers [Davison, 2003, p. 124] that the previous equation implies that

$$\hat{\psi} \dot{\sim} N\left(\psi_g, H(\psi_g)^{-1} J(\psi_g) H(\psi_g)^{-1}\right) \tag{3.5}$$

where

$$H(\psi_g) = n\int \frac{\partial^2 \log f(y;\psi)}{\partial \psi \partial \psi^T} g(y)\mathrm{d}y \tag{3.6}$$

$$J(\psi_g) = n\int \frac{\partial \log f(y;\psi)}{\partial \psi} \frac{\partial \log f(y;\psi)}{\partial \psi^T} g(y)\mathrm{d}y \tag{3.7}$$

Note that if by a lucky chance our candidate model $f(y;\psi)$ contains the true one, then $\psi_g = \psi$ and $H(\psi_g) = -J(\psi_g)$ so that equation (3.5) reduces to the usual asymptotic distribution for the MLE.

The use of pairwise likelihood, as a specific case of composite likelihood, leads to further simplifications. To see this, we consider the gradient of the log-pairwise likelihood

$$\nabla \ell_p(\mathbf{y};\psi) = \sum_{i<j}\sum_{k=1}^{n_{i,j}} \nabla \log f(y_k^{(i)}, y_k^{(j)};\psi) \tag{3.8}$$

For each fixed i and j,

$$\sum_{k=1}^{n_{i,j}} \nabla \log f(y_k^{(i)}, y_k^{(j)};\psi) = 0$$

is an unbiased estimating equation so that $\nabla \ell_p(\mathbf{y}; \psi) = 0$ is unbiased too as a linear combination of unbiased estimating equations. This leads to a modification of equation (3.5)

$$\hat{\psi}_p \dot{\sim} N\left(\psi, H(\psi)^{-1} J(\psi) H(\psi)^{-1}\right) \tag{3.9}$$

where $H(\psi)$ and $J(\psi)$ are given by equations (3.6) and (3.7).

Let consider a simple case study to see how it works in practice. Here we simulate independent replications of the Schlather model with a Whittle–Matérn correlation function having its sill, range and shape parameters equal to 0.8, 3 and 1.2 respectively.

```
> n.site <- 40
> set.seed(12)
> locations <- matrix(runif(2 * n.site, 0, 10), ncol = 2)
> colnames(locations) <- c("lon", "lat")
> ms0 <- MaxStableRF(locations[, 1], locations[, 2], grid = FALSE,
+     model = "wh", param = c(0, 1, 0.2, 3, 1.2), maxstable = "extr",
+     n = 80)
> ms0 <- t(ms0)
> fitmaxstab(ms0, locations, cov.mod = "whitmat", std.err.type = "none")

        Estimator: MPLE
            Model: Schlather
   Pair. Deviance: 569912.3
              TIC: NA
Covariance Family: Whittle-Matern

Estimates
  Marginal Parameters:
  Assuming unit Frechet.
  Dependence Parameters:
  sill   range  smooth
0.8353  4.0825  0.7345

Optimization Information
  Convergence: successful
  Function Evaluations: 164
```

From this output, we can see that we indeed use the Schlather's representation with a Whittle–Matérn correlation function. The convergence was successful and the parameter estimates for the covariance function as well as the pairwise deviance are accessible. Large deviations from the theoretical values may be expected as the parameters of the Whittle–Matérn covariance function are far from orthogonal. Thus, the range and smooth estimates may be totally different while leading (approximately) to the same covariance function.

The `fitmaxstab` function provides a powerful option that can fix any parameter of the model under consideration. For instance, this could be especially useful when using the Whittle–Matérn correlation function as it is sometimes preferable to fix the smooth parameter using prior knowledge on the process smoothness [Diggle et al., 2007]. Obviously, this feature is not restricted to this specific case and one can consider more exotic models. Fixing parameters of the model is illustrated by the following lines

```
> fitmaxstab(ms0, locations, cov.mod = "whitmat", smooth = 1.2,
+     std.err.type = "none")
```

```
> fitmaxstab(ms0, locations, cov.mod = "whitmat", sill = 1)
> fitmaxstab(ms0, locations, cov.mod = "whitmat", range = 3)
```

Although the Whittle–Matérn model is flexible, one may want to consider other families of covariance functions. This is achieved by invoking:

```
> fitmaxstab(ms0, locations, cov.mod = "cauchy")
> fitmaxstab(ms0, locations, cov.mod = "powexp")
```

One may also consider the Smith model, this could be done as follows

```
> fitmaxstab(ms0, locations, cov.mod = "gauss")
```

It is also possible to used different optimization routines to fit the model to data by passing the `method` argument. For instance, if one wants to use the `BFGS` method:

```
> fitmaxstab(ms0, locations, cov.mod = "gauss", cov12 = 0, method = "BFGS")
```

Instead of using the `optim` function, one may want to use the `nlm` or `nlminb` functions. This is done as before using the `method = "nlm"` or `method = "nlminb"` option.

Finally, it is important to note that maximizing the pairwise likelihood can be expensive. The choice of the numerical optimizer is important. In particular, optimizers that use the gradient of the pairwise likelihood might be clumsy. Indeed, if the Whittle–Matérn covariance function is considered and the smooth parameter has to be estimated, then the pairwise likelihood is not differentiable with respect to this specific parameter. In general, the Nelder–Mead [Nelder and Mead, 1965] approach seems to perform better even if the convergence is sometimes slow.

3.2.2 Assessing Uncertainties

Recall that the maximum pairwise likelihood estimator $\hat{\psi}_p$ satisfies

$$\hat{\psi}_p \dot{\sim} \mathcal{N}\left(\psi, H(\psi)^{-1} J(\psi) H(\psi)^{-1}\right), \qquad n \to +\infty,$$

where $H(\psi) = \mathbb{E}[\nabla^2 \ell_p(\psi; \mathbf{Y})]$ (the Hessian matrix) and $J(\psi) = \mathrm{Var}[\nabla \ell_p(\psi; \mathbf{Y})]$, where the expectations are with respect to the "full" density.

In practice, to get the standard errors we need estimates of $H(\psi)$ and $J(\psi)$. The estimation of $H(\psi)$ is straightforward and is $\hat{H}(\hat{\psi}_p) = \nabla^2 \ell_p(\hat{\psi}_p; \mathbf{y})$; that is, the Hessian matrix evaluated at $\hat{\psi}_p$. Usually, standard optimizers are able to get finite-difference based estimates for $H(\hat{\psi}_p)$ so that no extra work is needed to get $\hat{H}(\hat{\psi}_p)$.

The estimation of $J(\hat{\psi}_p)$ is a little bit more difficult and can be done in two different ways [Varin and Vidoni, 2005]. First, $J(\hat{\psi}_p)$ can be estimated using the "naive" estimator $\hat{J}(\hat{\psi}_p) = \nabla \ell_p(\hat{\psi}_p; \mathbf{y}) \nabla \ell_p(\hat{\psi}_p; \mathbf{y})^T$. This estimator is tagged `grad` as it uses the gradient of the log pairwise likelihood. Another estimator is given by noticing that $J(\psi)$ corresponds to the variance of the pairwise score equations $\ell_p(\psi; \mathbf{Y}) = 0$. Consequently, a second estimator, tagged `score`, is given by the sample variance of each contribution to the pairwise score function. The second estimator is only accessible if independent replications of $\mathbf{Y}$ are available[2].

These two types of standard errors are available by invoking the following two lines:

```
> fitmaxstab(ms0, locations, cov.mod = "gauss", std.err.type = "score")
> fitmaxstab(ms0, locations, cov.mod = "gauss", std.err.type = "grad")
```

[2] which will mostly be the case for spatial extremes.

4

Model Selection

Model selection plays an important role in statistical modelling. According to Ockham's razor, given several models that fit the data equally well, we should focus on simple models rather than more complex ones. Depending on the models to be compared, several approaches exist for model selection. In this section, we will present theory on information criteria such as the Akaike Information Criterion (AIC) [Akaike, 1974] and the likelihood ratio statistic [Davison, 2003, Sec. 4.5]. We present these two approaches in turn.

4.1 Takeuchi Information Criterion

Having two different models, we want to know which one we should prefer for modelling our data. If two models have exactly the same maximized log-likelihoods, we should prefer the one which has fewer parameters because it will have a smaller variance. However, if these two maximized log-likelihoods only differ by a small amount, does this small increase worth the price of having additional parameters? To answer this question, we resort to the Kullback–Leibler discrepancy.

Let consider a random sample $Y_1, \ldots, Y_n$ drawn from an unknown density g. Ignoring g, we fit a statistical model $f(y;\psi)$ by maximizing the log-likelihood. The Kullback–Leibler discrepancy measures the discrepancy of our fitted model f from the true one g

$$D\left(f_\psi, g\right) = \int \log\left(\frac{g(y)}{f(y;\psi)}\right) g(y)\mathrm{dy} \tag{4.1}$$

The Kullback–Leibler discrepancy is always positive. Indeed, as $x \mapsto -\log(x)$ is a convex function, Jensen's inequality states

$$D\left(f_\psi, g\right) = \int \log\left(\frac{g(y)}{f(y;\psi)}\right) g(y)\mathrm{dy} \geq -\log\left(\int \frac{f(y;\psi)}{g(y)} g(y)\mathrm{dy}\right) = 0$$

where we used the fact that $f(y;\psi)$ is a probability density function. Furthermore, the lower bound is reached if and only if the convex function is not strictly convex which only occurs iff $f(y;\psi) = g(y)$.

Consequently, for model selection, we aim to choose models that minimize $D(f_\psi, g)$. However, $D(f_\psi, g)$ is not enough discriminant as several models may satisfy $D(f_\psi, g) = 0$. To solve this issue, suppose we have an estimate $\hat{\psi}$, we need to average $D(f_{\hat{\psi}}, g)$ over the distribution of $\hat{\psi}$. Intuitively, because of their larger sampling variance, averaging over $\hat{\psi}$ will penalized much more models having a larger number of parameters than those with fewer parameters.

Let ψ_g be the value that minimizes $D(f_\psi, g)$. A Taylor expansion of $\log f(y;\hat{\psi})$ around ψ_g gives

$$\log f(y;\hat{\psi}) \approx \log f(y;\psi_g) + \frac{1}{2}\left(\hat{\psi} - \psi_g\right)^T \frac{\partial^2 \log f(y;\psi_g)}{\partial\psi\partial\psi^T}\left(\hat{\psi} - \psi_g\right)$$

where we use the fact that ψ_g minimise the Kullback–Leibler discrepancy so its partial derivative with respect to ψ vanishes. By putting this into equation (4.1) we get

$$D\left(f_{\hat{\psi}}, g\right) \;\dot{=}\; D\left(f_{\psi_g}\right) - \frac{1}{2n}\mathrm{tr}\left\{(\hat{\psi}-\psi_g)(\hat{\psi}-\psi_g)^T J(\psi_g)\right\}$$

where $J(\psi_g)$ is given by equation (3.7). Finaly, we have

$$n\mathbb{E}_g\left[D\left(f_{\hat{\psi}}, g\right)\right] \dot{=} nD\left(f_{\psi_g}, g\right) - \frac{1}{2}\mathrm{tr}\left\{J(\psi_g)^{-1}H(\psi_g)\right\} \tag{4.2}$$

where $H(\psi_g)$ is given by equation (3.6).

The AIC is, up to constant, an estimator of equation (4.2) and is defined as

$$\mathrm{AIC} = -2\ell(\hat{\psi}) + 2p \tag{4.3}$$

The simplification $\mathrm{tr}\{J(\psi_g)^{-1}H(\psi_g)\} = -p$ arises as the AIC supposes that there is no misspecification.

Because we see in Section 3.2.1 that by using the maximum pairwise likelihood estimator we work under misspecification, the AIC is not appropriate. Another estimator of equation (4.2), which allows for misspecification, is the Takeuchi information criterion (TIC)

$$\mathrm{TIC} = -2\ell(\hat{\psi}) - 2\mathrm{tr}\left\{\hat{J}\hat{H}^{-1}\right\} \tag{4.4}$$

In accordance with the AIC, the best model will correspond to the one that minimizes equation (4.4). Recently, Varin and Vidoni [2005] rediscover this information criterion and demonstrate its use for model selection when composite likelihood is involved.

In practice, one can have a look at the output of the `fitmaxstab` function or use the `TIC` function.

```
> set.seed(1)
> n.site <- 40
> locations <- matrix(runif(2 * n.site, 0, 10), ncol = 2)
> colnames(locations) <- c("lon", "lat")
> ms0 <- MaxStableRF(locations[, 1], locations[, 2], grid = FALSE,
+     model = "cauchy", param = c(0, 1, 0.2, 3, 1.2), maxstable = "extr",
+     n = 80)
> ms0 <- t(ms0)
> model1 <- fitmaxstab(ms0, locations, cov.mod = "powexp")
> model2 <- fitmaxstab(ms0, locations, cov.mod = "cauchy")
> TIC(model1, model2)

  model2   model1
583940.2 584021.1
```

The TIC for `model2` is lower that the one for `model1` so that we might prefer using `model2`.

4.2 Likelihood Ratio Statistic

TIC is useful when comparing different models. When dealing with nested models, one may prefer using the likelihood ratio statistic because of the Neyman–Pearson lemma [Neyman and Pearson, 1933].

Suppose we are interested in a statistical model $\{f(x;\psi)\}$ where $\psi^T = (\kappa^T, \phi^T)$ and more especially wether a particular value κ_0 of κ is relevant with our data. Let $(\hat{\kappa}^T, \hat{\phi}^T)$ be the maximum likelihood estimate for ψ and $\hat{\phi}_{\kappa_0}$ the maximum likelihood estimate under the restriction $\kappa = \kappa_0$. A common statistic to check if κ_0 is relevant or not is the likelihood ratio statistic $W(\kappa_0)$ which satisfies, under mild regularity conditions,

$$W(\kappa_0) = 2\left\{\ell(\hat{\kappa}, \hat{\phi}) - \ell(\kappa_0, \hat{\phi}_{\kappa_0})\right\} \longrightarrow \chi^2_p, \qquad n \to +\infty \tag{4.5}$$

where p is the dimension of κ_0.

Unfortunately, when working under misspecification, the usual asymptotic χ^2_p distribution does not hold anymore. There's two ways to solve this issue: (a) adjusting the χ^2 distribution [Rotnitzki and Jewell, 1990] or (b) adjusting the composite likelihood so that the usual χ^2_p holds [Chandler and Bate, 2007]. We will consider these two strategies in turn.

4.2.1 Adjusting the χ^2 distribution

If the model is misspecified, equation (4.5) has to be adjusted. More precisely, as stated by [Kent, 1982], we have

$$W(\kappa_0) = 2\left\{\ell(\hat{\kappa}, \hat{\phi}) - \ell(\kappa_0, \hat{\phi}_{\kappa_0})\right\} \longrightarrow \sum_{i=1}^{p} \lambda_i X_i, \qquad n \to +\infty \tag{4.6}$$

where λ_i are the eigenvalues of $((H^{-1}JH^{-1})_\kappa\{-(H^{-1})_\kappa\}^{-1}$, the X_i are independent χ^2_1 random variables and $(H^{-1}JH^{-1})_\kappa$ and $(H^{-1})_\kappa$ are the submatrices of $H^{-1}JH^{-1}$ and H^{-1} corresponding to the elements of κ and where the matrices H and J are given by equations (3.6) and (3.7). For practical purposes, the matrices H and J are substituted for their respective estimates as described in Section 3.2.2.

The problem with equation (4.6) is that generally the distribution of $\sum_{i=1}^p \lambda_i X_i$ is not known exactly. This led Rotnitzki and Jewell [1990] to consider $pW(\kappa_0)/\sum_{i=1}^p \lambda_i$ as a χ^2_p random variable. However, a better approximation uses results on quadratic forms of normal random distribution.

An application of this approach is given by the following lines:

```
> set.seed(7)
> n.site <- 30
> locations <- matrix(rnorm(2 * n.site, sd = sqrt(0.2)), ncol = 2)
> colnames(locations) <- c("lon", "lat")
> sigma <- matrix(c(100, 25, 25, 220), ncol = 2)
> sigma.inv <- solve(sigma)
> sqrtCinv <- t(chol(sigma.inv))
> model <- list(list(model = "gauss", var = 1, aniso = sqrtCinv/2))
> ms0 <- MaxStableRF(locations[, 1], locations[, 2], grid = FALSE,
+     model = model, maxstable = "Bool", n = 50)
> ms0 <- t(ms0)
> M0 <- fitmaxstab(ms0, locations, "gauss", , cov11 = 100)
> M1 <- fitmaxstab(ms0, locations, "gauss")
> anova(M0, M1)

Eigenvalue(s):
281.74
```

```
Analysis of Variance Table
   MDf Deviance Df  Chisq Pr(> sum lambda Chisq)
M0   2    47716
M1   3    47631  1 84.585                 0.5837
```

From this ouput, we can see that the p-value is approximately 0.58 which turns out to be in favour of H_0 i.e. $\sigma_{11} = 100$ in Σ. Note that the eigenvalue estimate was also reported.

4.2.2 Adjusting the composite likelihood

Contrary to the approach of Rotnitzki and Jewell [1990], Chandler and Bate [2007] propose to adjust the composite likelihood instead of adjusting the asymptotic likelihood ratio statistic distribution. The starting point is that, under misspecification, the log-composite likelihood evaluated at its maximum likelihood estimate $\hat{\psi}$ has curvature $-\hat{H}^{-1}$ while it should be $\hat{H}^{-1}\hat{J}\hat{H}^{-1}$. This forms the basis for adjusting the log-composite likelihood is the following way,

$$\ell_{adj}(\psi) = \ell(\psi^*), \qquad \psi^* = \hat{\psi} + C(\psi - \hat{\psi}) \tag{4.7}$$

for some $p \times p$ matrix C.

It is straightforward to see that $\hat{\psi}$ maximizes ℓ_A with zero derivative. The key point is that its curvature at $\hat{\theta}$ is $C^T\hat{H}^{-1}C$. By choosing an appropriate C matrix, it is possible to ensure that ℓ_A has the right curvature for applying (4.5) directly. More precisely, by letting

$$C = M^{-1}M_A \tag{4.8}$$

where $M^TM = \hat{H}$ and $M_A^TM_A = \hat{H}^{-1}\hat{J}\hat{H}^{-1}$, we ensure that ℓ_A has curvature $\hat{H}^{-1}\hat{J}\hat{H}^{-1}$. If $p > 1$, the matrix square roots M and M_A are not unique and one may use the Cholesky or the singular value decompostions.

With this setting, we can apply (4.5) directly i.e.

$$W_A(\kappa_0) = 2\left\{\ell_A(\hat{\kappa}, \hat{\phi}) - \ell_A(\kappa_0, \hat{\phi}_A)\right\} \longrightarrow \chi^2_p$$

where $\hat{\phi}_A$ is the maximum adjusted likelihood estimated for the restricted model.

The problem with the above equation is that it requires the estimation of $\hat{\phi}_A$ which could be CPU prohibitive. Unfortunately, substituting $\hat{\phi}$ for $\hat{\phi}_A$ doesn't solve the problem as $\ell_A(\hat{\phi}) \leq \ell_A(\hat{\phi}_A)$ so that this substitution will inflate $W_A(\kappa_0)$ and thus $\Pr_{H_0}[W_a(\kappa_0) > w_a(\alpha)] > 1 - \alpha$, where $w_a(\alpha)$ is the $1 - \alpha$ quantile for the χ^2_p distribution and α the significance level of the hypothesis test.

To solve these problems, Chandler and Bate [2007] propose to compensate for the use of $\hat{\phi}$ instead of $\hat{\phi}_A$ by considering a modified likelihood ratio statistic

$$W_A^*(\kappa_0) = 2c\left\{\ell_A(\hat{\kappa}, \hat{\phi}) - \ell_A(\kappa_0, \hat{\phi}_A)\right\} \longrightarrow \chi^2_p \tag{4.9}$$

where

$$c = \frac{(\hat{\kappa} - \kappa_0)^T\{-(\hat{H}^{-1}\hat{J}\hat{H}^{-1})_{\kappa_0}\}^{-1}(\hat{\kappa} - \kappa_0)}{\{(\hat{\kappa}^T, \hat{\phi}^T)\}^T\{-(\hat{H}^{-1}\hat{J}\hat{H}^{-1})_{\kappa_0}\}\{(\hat{\kappa}^T, \hat{\phi}^T)\}}$$

The next lines performs the same hypothesis test as that in Section 4.2.1.

```
> anova(M0, M1, method = "CB")
```

```
Analysis of Variance Table
   MDf Deviance Df  Chisq Pr(> sum lambda Chisq)
M0   2    29661
M1   3    29661  1 0.2629                0.6081
```

By using the Chandler and Bate methodology, we draw the same conclusion in the previous section, i.e. the observations are consistent with the null hypothesis $\sigma_{11} = 100$.

5

Fitting a Max-Stable Process to Data With Unknown GEV Margins

In practice, the observations will never be drawn from a unit Fréchet distribution so that Chapter 3 won't help much with concrete applications. One way to avoid this problem is to fit a GEV to each location and then transform all data to the unit Fréchet scale. Given a continuous random variable Y whose cumulative distribution function is F, one can define a new random variable Z such as Z is unit Fréchet distributed

$$Z = -\frac{1}{\log F(Y)} \tag{5.1}$$

More precisely, if Y is a random variable distributed as a GEV with location, scale and shape parameters equal to μ, σ and ξ respectively, it turns out that equation (5.2) becomes

$$Z = \left(1 + \xi \frac{Y - \mu}{\sigma}\right)^{1/\xi} \tag{5.2}$$

The above transformation can be done by using the `gev2frech` function

```
> x <- c(2.2975896, 1.6448808, 1.3323833, -0.4464904, 2.2737603,
+     -0.2581876, 9.5184398, -0.5899699, 0.4974283, -0.8152157)
> z <- gev2frech(x, 1, 2, 0.2)
```

or conversely if Z is a unit Fréchet random variable, then the random variable Y defined as

$$Y = \mu + \sigma \frac{Z^{\xi} - 1}{\xi} \tag{5.3}$$

is GEV distributed with location, scale and shape parameter equal to $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_*^+$ and $\xi \in \mathbb{R}$ respectively.

```
> frech2gev(z, 1, 2, 0.2)

 [1]  2.2975896  1.6448808  1.3323833 -0.4464904  2.2737603 -0.2581876
 [7]  9.5184398 -0.5899699  0.4974283 -0.8152157
```

The drawback of this approach is that standard errors are incorrect as the margins are fitted separately from the spatial dependence structure. Consequently, the standard errors related to the spatial dependence parameters are underestimated as we suppose that data were originally unit Fréchet.

One can solve this problem by fitting in *one step* both GEV and spatial dependence parameters [Padoan, 2008; Padoan et al., 2008; Gholam-Rezaee, 2009]. As the bivariate distributions

for the max-stable models introduced in Chapter 1 were imposing unit Fréchet margins, we need to rewrite them for unknown GEV margins. To this aim, let define the transformation t such that

$$t : Y(x) \mapsto \left(1 + \xi(x)\frac{Y(x) - \mu(x)}{\sigma(x)}\right)^{1/\xi(x)} \tag{5.4}$$

where $Y(\cdot)$ is supposed to be a max-stable random field having GEV margins such that $Y(x) \sim \text{GEV}(\mu(x), \sigma(x), \xi(x))$, $\sigma(x) > 0$ for all $x \in \mathbb{R}^d$. Consequently, the bivariate distribution of $(Y(x_1), Y(x_2))$ is

$$\Pr\left[Y(x_1) \leq y_1, Y(x_2) \leq y_2\right] = \Pr\left[Z(x_1) \leq z_2, Z(x_2) \leq z_2\right]$$

where $z_1 = t(y_1)$ and $z_2 = t(y_2)$. Thus, one can relate the bivariate density for $(Y(x_1), Y(x_2))$ to the one for $(Z(x_1), Z(x_2))$ that we introduced in Chapter 1 and the log pairwise likelihood becomes

$$\ell_p(\mathbf{y}; \psi) = \sum_{i<j}\sum_{k=1}^{n_{i,j}} \left\{\log f(z_k^{(i)}, z_k^{(j)}; \psi) + \log|J(y_k^{(i)})J(y_k^{(j)})|\right\} \tag{5.5}$$

where $n_{i,j}$ is the sample size of common observations between site i and j and

$$z_k^{(i)} = \left(1 + \xi_i\frac{y_k^{(i)} - \mu_i}{\sigma_i}\right)^{1/\xi_i - 1}$$

where μ_i, σ_i, ξ_i are the GEV parameters for the i-th site and $y_k^{(i)}$ is the k-th observation available at site i and $|J(t(y_k^{(i)}))|$ is the Jacobian of the mapping t evaluated at the $y_k^{(i)}$ observation i.e.

$$|J(t(y_k^{(i)}))| = \frac{1}{\sigma_i}\left(1 + \xi_i\frac{y_k^{(i)} - \mu_i}{\sigma_i}\right)^{1/\xi_i - 1}$$

Maximizing the log-pairwise likelihood given by equation (5.5) is possible by passing the option `fit.marge = TRUE` in the `fitmaxstab` function i.e.

```
> fitmaxstab(data, coord, "gauss", fit.marge = TRUE)
```

However, this will be really time consuming as such models will have $3n.site + p$ parameters to estimate, where p is the number of parameters related to the extremal spatial dependence structure. Another drawback is that prediction at unobserved locations won't be possible. Indeed, if no model is assumed for the evolution of the GEV parameters in space, it is therefore impossible to predict them where no data is available.

Another way may be to fit *response surfaces* for the GEV parameters. The next section aims to give an introduction to the use of response surfaces.

5.1 Response Surfaces

Response surfaces is a generic term when the problem under concern is to describe how a *response variable* y depends on *explanatory variables* $x_1, \ldots, x_k$. For instance, with our particular problem of spatial extremes, one may wonder how is it possible to predict the GEV parameters at a fixed location given the knowledge of extra covariables such as longitude, latitude, ... The goal of response surfaces is to get efficient predictions for the response variable while keeping, so far as we can, simple models.

In this section, we will first introduce the linear regression models. Next, we will increase in complexity and flexibility by introducing semiparametric regression models.

5.1.1 Linear Regression Models

Suppose we observe a response y through the $y_1, \ldots, y_n$ values. For each observed y_i, we also have p related explanatory variables denoted by $x_{1,i}, \ldots, x_{p,i}$. To predict y given the $x_{\cdot,\cdot}$ values, one might consider the following model:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i} + \epsilon_i$$

where $\beta_0, \ldots, \beta_p$ are the regression parameters to be estimated and ϵ_i is an unobserved error term.

It is possible to write the above equation in a more compact way by using matrix notation e.g.

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{5.6}$$

where $\mathbf{y}$ is a $n \times 1$ vector, $\mathbf{X}$ is a $n \times p$ matrix called the *design matrix* and ϵ is a $p \times 1$ vector.

Model (5.6) is called a *linear model* as it is linear in β but not necessarily in the covariates x. For example, the two following models are linear models

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_2 \log x_2 + \epsilon \end{aligned}$$

or equivalently in a matrix notation

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,1}^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{1,n}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \log x_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & \log x_{2,n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

Usually, β is estimated by minimizing least squares

$$SS(\beta) = \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \tag{5.7}$$

which amounts to solve the equations

$$\sum_{i=1}^{n} x_{j,i}(y_i - \beta^T x_i) = 0, \qquad j = 1, \ldots, p$$

or equivalently in a matrix notation

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

so that, provided that $\mathbf{X}^T\mathbf{X}$ is invertible, the least squares estimate for β is given by

$$\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{y} \tag{5.8}$$

and the fitted $\mathbf{y}$ values are given by

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{y} \tag{5.9}$$

The matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$ is called the *hat matrix* as it puts "hats" on $\mathbf{y}$. $\mathbf{H}$ is a projection matrix that orthogonally projects $\mathbf{y}$ onto the plane spanned by the columns of the design matrix $\mathbf{X}$.

The hat matrix plays an important role in parametric regression as it provides useful informations on the influence of some observations to the fitted values. Indeed, from equation (5.9), we have

$$\hat{y}_i = \sum_{j=1}^{n} H_{i,j} y_j$$

so that $H_{i,i}$ is the contribution of y_i to the estimate $\hat{y}_i$. Furthermore, if we consider the total influence of all the observations, we have

$$\begin{aligned} \sum_{i=1}^{n} H_{i,i} &= \operatorname{tr}(\mathbf{H}) = \operatorname{tr}\{\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\} \\ &= \operatorname{tr}\{\mathbf{X}^T\mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\} = \operatorname{tr}(\mathbf{I}_p) = p \end{aligned}$$

and the total influence of all observations is equal to the degrees of freedom of the model.

5.1.2 Semiparametric Regression Models

In the previous section, we talked about linear regression models for which the relationship between the explanatory variables and the response has a deterministic shape and is supposed to be known. However, it may happened applications for which the data have a complex behaviour. For such cases, we benefit from using semiparametric regression models defined as

$$y_i = f(x_i) + \epsilon_i \tag{5.10}$$

where f is a smooth function with unknown shape.

The idea of semiparametric regression models is to decompose g into an appropriate basis for which equation (5.10) simplifies to equation (5.6) e.g.

$$f(x) = \sum_{j=1}^{q} b_j(x)\beta_j \tag{5.11}$$

where $b_j(\cdot)$ is the j-th basis function and β_j is the j-th element of the regression parameter β.

Several basis functions exist such as the polynomial basis, the cubic spline basis, B-splines, ... It is beyond the scope of this document to introduce all of them in details but the interested reader should have a look at Ruppert et al. [2003]. Some details about the basis implemented in the package are reported in Annex A

Usually, the basis functions $b_j(\cdot)$ depends on *knots* κ so that equation (5.11) becomes

$$f(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{q} b_j(x - \kappa_j) \tag{5.12}$$

The problem with model (5.12) is that it is strongly affected by the number of knots. Figure 5.1 depicts this problem by fitting the same dataset to model (5.12) with $q = 2, 10$ and 50. Clearly, the first fit is not satisfactory and we need to increase the number of knots. The second one seems plausible while the last one clearly overfits. This figure was generated with the following lines

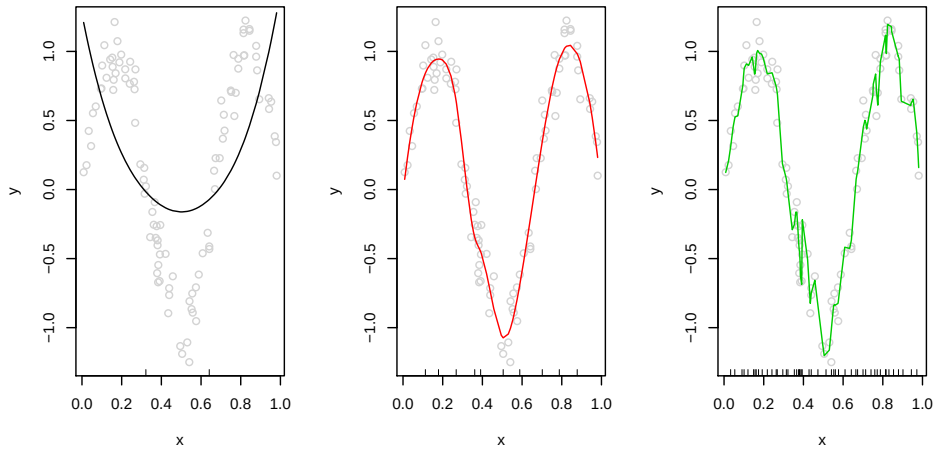


Figure 5.1: *Impact of the number of knots in the fitted p-spline. Left panel: $q = 2$, middle panel: $q = 10$, right panel: $q = 50$. The small vertical lines corresponds to the location of each knot.*

```
> set.seed(12)
> x <- runif(100)
> fun <- function(x) sin(3 * pi * x)
> y <- fun(x) + rnorm(100, 0, 0.15)
> knots1 <- quantile(x, prob = 1:2/3)
> knots2 <- quantile(x, prob = 1:10/11)
> knots3 <- quantile(x, prob = 1:50/51)
> M0 <- rbpspline(y, x, knots = knots1, degree = 3, penalty = 0)
> M1 <- rbpspline(y, x, knots = knots2, degree = 3, penalty = 0)
> M2 <- rbpspline(y, x, knots = knots3, degree = 3, penalty = 0)
> par(mfrow = c(1, 3))
> plot(x, y, col = "lightgrey")
> rug(knots1)
> lines(M0)
> plot(x, y, col = "lightgrey")
> rug(knots2)
> lines(M1, col = 2)
> plot(x, y, col = "lightgrey")
> rug(knots3)
> lines(M2, col = 3)
```

Consequently, there is a pressing need for a kind of "automatic knot selection". One common strategy to overcome this issue is to resort to *penalized splines* or *p-splines*. The idea beyond this is to consider a large number of knots but to constrain, in a sense to be defined, their influence.

To avoid overfitting, one wish to minimize the sum of square subject to some constraint on the β parameter i.e.

$$\text{minimize } ||\mathbf{y} - \mathbf{X}\beta||^2 \quad \text{subject to } \beta^T \mathbf{K} \beta \leq C$$

for a judicious choice of C and a given matrix $\mathbf{K}$. Using a Lagrange multiplier argument, this is equivalent to choosing β to minimize

$$||\mathbf{y} - \mathbf{X}\beta||^2 + \lambda \beta^T \mathbf{K} \beta \tag{5.13}$$

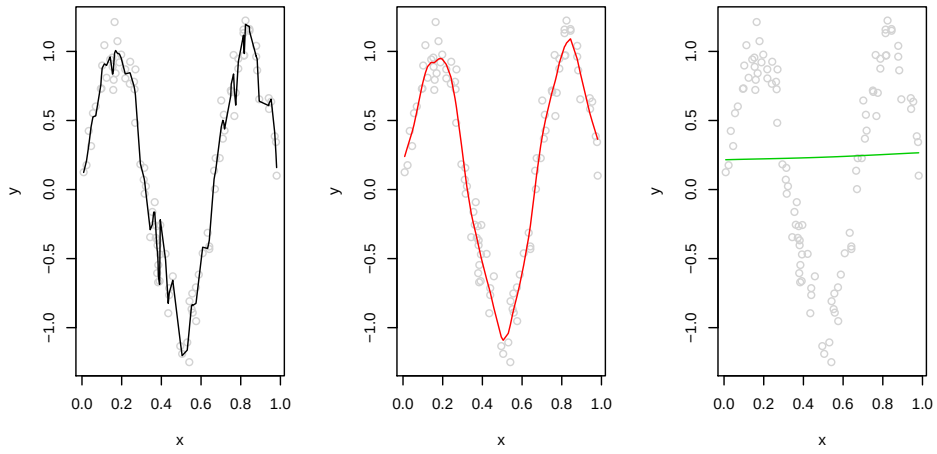


Figure 5.2: *Impact of the smoothing parameter λ on the fit. Left panel: $\lambda = 0$, middle panel: $\lambda = 0.1$ and right panel: $\lambda = 10$.*

for some $\lambda \geq 0$ called the *smoothing parameter* as it controls the amount of smoothing. Indeed, if $\lambda = 0$, then problem (5.13) is left unconstrained and leads to wiggly fits, while λ being large implies smoother fits. Figure 5.2 is a nice illustration of the impact of the smoothing parameter on the smoothness of the fitted curve. It was generated using the following code

```
> M0 <- rbpspline(y, x, knots = knots3, degree = 3, penalty = 0)
> M1 <- rbpspline(y, x, knots = knots3, degree = 3, penalty = 0.1)
> M2 <- rbpspline(y, x, knots = knots3, degree = 3, penalty = 10)
> par(mfrow = c(1, 3))
> plot(x, y, col = "lightgrey")
> lines(M0)
> plot(x, y, col = "lightgrey")
> lines(M1, col = 2)
> plot(x, y, col = "lightgrey")
> lines(M2, col = 3)
```

It can be shown that problem (5.13) has the solution

$$\hat{\beta}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T\mathbf{y} \tag{5.14}$$

and the corresponding fitted values for a penalized spline are given by

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T\mathbf{y} \tag{5.15}$$

In accordance with the hat matrix with linear models, one can define the *smoother matrix* $\mathbf{S}_\lambda$ such that

$$\hat{\mathbf{y}} = \mathbf{S}_\lambda\mathbf{y} \tag{5.16}$$

where $\mathbf{S}_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T$. Consequently, a kind of effective degrees of freedom is given by $\mathrm{tr}(\mathbf{S}_\lambda)$.

If the problem of knot selection seems to be resolved by using these constrained least squares minimisation, there is still some open questions: given our data and knots, what is the best value for λ? Would it be possible to get an "automatic selection" for λ?

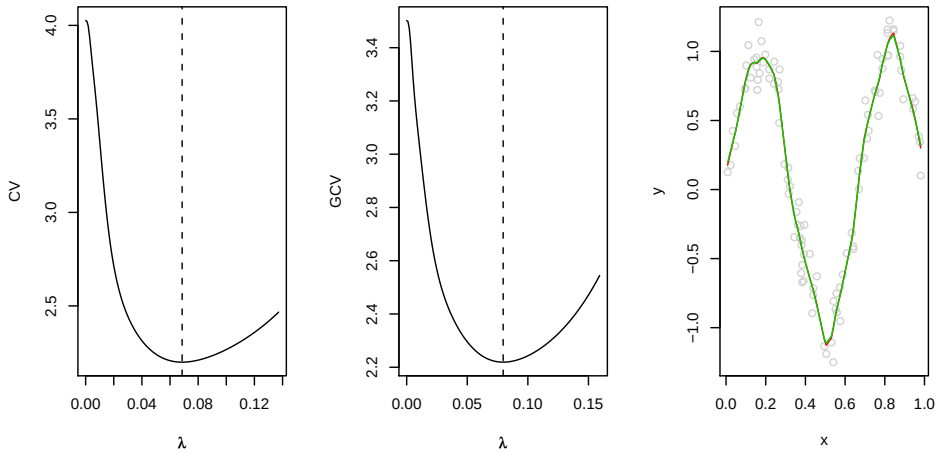


Figure 5.3: *Cross-validation and generalized cross validation curves and corresponding fitted curves.*

One common tool for answering these two questions is known as *cross-validation* (CV)

$$CV(\lambda) = \sum_{i=1}^{n} \{y_i - \hat{f}_{-i}(x_i; \lambda)\}^2 \tag{5.17}$$

where $\hat{f}_{-i}$ corresponds to the semiparametric estimator applied to the data but with (x_i, y_i) omitted. Intuitively, large values of $CV(\lambda)$ corresponds to models that are wiggly and/or have a large variance in the parameter estimates so that minimising $CV(\lambda)$ is a nice option for an "automatic selection" of λ.

Unfortunately, the computation of equation (5.17) directly is often too CPU demanding. However, it can be shown [Ruppert et al., 2003] that

$$CV(\lambda) = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - S_{\lambda,ii}} \right)^2 \tag{5.18}$$

where $S_{\lambda,ii}$ is the (i,i) element of $\mathbf{S}_\lambda$. Clearly, equation (5.18) does a better job than equation (5.17) as it only requires one fit to compute $CV(\lambda)$.

Sometimes, the weights $1 - S_{\lambda,ii}$ are replaced by the mean weight, $\text{tr}(\mathbf{Id} - \mathbf{S}_\lambda)/n$, where $\mathbf{Id}$ is the identity matrix, leading to the *generalized cross-validation* (GCV) score

$$GCV(\lambda) = n^2 \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{\text{tr}(\mathbf{Id} - \mathbf{S}_\lambda)} \right)^2 \tag{5.19}$$

GCV has computational advantages over CV, and it has also computational advantages in term of invariance [Wood, 2006].

Figure 5.3 plots the CV and GCV curves for the data plotted in Figure 5.2 and the corresponding fitted p-spline. The selection of λ using CV or GCV yield approximately to the same smoothing parameter value. These "best" λ values are in accordance with the values we held fixed in Figure 5.2. The fitted curves using eigher CV or GCV lead to indistinguishable curves. The code used to generate Figure 5.3 was

```
> par(mfrow = c(1, 3))
> lambda.cv <- cv(y, x, knots = knots3, degree = 3)$penalty
```

```
> abline(v = lambda.cv, lty = 2)
> lambda.gcv <- gcv(y, x, knots = knots3, degree = 3)$penalty
> abline(v = lambda.gcv, lty = 2)
> cv.fit <- rbpspline(y, x, knots3, degree = 3, penalty = "cv")
> gcv.fit <- rbpspline(y, x, knots3, degree = 3, penalty = "gcv")
> plot(x, y, col = "lightgrey")
> lines(cv.fit, col = 2)
> lines(gcv.fit, col = 3)
```

5.2 Building Response Surfaces for the GEV Parameters

In the previous section, we introduced the notion of response surfaces and we show that they should be used if one is interested in simultaneously fitting the GEV and the spatial dependence parameters of a max-stable process. However, one may wonder how to build accurate response surfaces for the GEV parameters. This is the aim of this section.

A first attempt could be to fit several max-stable models and identify the most promising ones by using the techniques on model selection introduced in Chapter 4. Although it is a legitimate approach, its use in practice is limited because the fitting procedure, due to the pairwise likelihood estimator, is CPU prohibitive.

A more pragmatic strategy is to consider only these response surfaces while omitting temporally the spatial dependence parameters. Although this strategy doesn't take into account all the uncertainties on the max-stable parameters, it should lead to accurate model selection as one expects the spatial dependence parameters and the GEV response surface parameters to be nearly orthogonal. The main asset of the latter approach is that fitting a (kind of) spatial GEV model to data is less CPU consuming.

This spatial GEV model is defined as follows:

$$Z(x) \sim GEV\left(\mu(x), \sigma(x), \xi(x)\right) \tag{5.20}$$

where the GEV parameters are defined through the following equations

$$\mu = X_\mu \beta_\mu, \qquad \sigma = X_\sigma \beta_\sigma, \qquad \xi = X_\xi \beta_\xi$$

where $X_.$ are design matrices and $\beta_.$ are parameters to be estimated.

The log-likelihood of the spatial GEV model is

$$\ell(\beta) = \sum_{i=1}^{n.site} \sum_{j=1}^{n.obs} \left\{ -\log \sigma_i - \left(1 + \xi_i \frac{z_{i,j} - \mu_i}{\sigma_i}\right)^{-1/\xi_i} - \left(1 + \frac{1}{\xi_i}\right) \log\left(1 + \xi_i \frac{z_{i,j} - \mu_i}{\sigma_i}\right) \right\} \tag{5.21}$$

where $\beta = (\beta_\mu, \beta_\sigma, \beta_\xi)$, μ_i, σ_i and ξ_i are the GEV parameters for the i-th site and $z_{i,j}$ is the j-th obvservation for the i-th site.

From equation (5.21), we can see that independence between stations is assumed. For most applications, this assumption is clearly incorrect and we require the use of the MLE asymptotic distribution under misspecification to get standard error estimates:

$$(\beta_\mu, \beta_\sigma, \beta_\xi) \dot{\sim} \mathcal{N}\left(\psi, H(\beta)^{-1} J(\beta) H(\beta)^{-1}\right), \qquad n \to +\infty \tag{5.22}$$

where $H(\beta) = \mathbb{E}[\nabla^2 \ell_p(\beta; \mathbf{Y})]$ (the Hessian matrix) and $J(\beta) = \mathrm{Var}[\nabla \ell_p(\beta; \mathbf{Y})]$.

In practice, the spatial GEV model is fitted to data through the `fitspatgev` function. The use of this function is similar to `fitmaxstab`.

Lets start by simulating a max-stable process with unit Fréchet margins and transform it to have a spatially structured GEV margins.

```
> n.site <- 20
> set.seed(15)
> locations <- matrix(runif(2 * n.site, 0, 10), ncol = 2)
> colnames(locations) <- c("lon", "lat")
> sigma <- matrix(c(100, 25, 25, 220), ncol = 2)
> sigma.inv <- solve(sigma)
> sqrtCinv <- t(chol(sigma.inv))
> model <- list(list(model = "gauss", var = 1, aniso = sqrtCinv/2))
> ms0 <- MaxStableRF(locations[, 1], locations[, 2], grid = FALSE,
+     model = model, maxstable = "Bool", n = 50)
> ms1 <- t(ms0)
> param.loc <- -10 + 2 * locations[, 2]
> param.scale <- 5 + 2 * locations[, 1] + locations[, 2]^2
> param.shape <- rep(0.2, n.site)
> for (i in 1:n.site) ms1[, i] <- frech2gev(ms1[, i], param.loc[i],
+     param.scale[i], param.shape[i])
```

Now we define appropriate response surfaces for our spatial GEV model and fit two different models.

```
> loc.form <- y ~ lat
> scale.form <- y ~ lon + I(lat^2)
> shape.form <- y ~ 1
> shape.form2 <- y ~ lon
> M1 <- fitspatgev(ms1, locations, loc.form, scale.form, shape.form)
> M2 <- fitspatgev(ms1, locations, loc.form, scale.form, shape.form2)
> M1

   Deviance: 10763.06
    Location Parameters:
locCoeff1  locCoeff2
   -9.918      2.797
       Scale Parameters:
scaleCoeff1  scaleCoeff2  scaleCoeff3
     4.5532       2.0988       0.9704
       Shape Parameters:
shapeCoeff
    0.3292

Standard Errors
  locCoeff1    locCoeff2  scaleCoeff1  scaleCoeff2  scaleCoeff3   shapeCoeff
     0.6911       0.6875       1.6082       0.2757       0.1459       0.1219

Asymptotic Variance Covariance
             locCoeff1  locCoeff2  scaleCoeff1  scaleCoeff2  scaleCoeff3
locCoeff1     0.477663   0.376199   0.905146     0.093538     0.070218
locCoeff2     0.376199   0.472642   0.690456     0.118999     0.092110
scaleCoeff1   0.905146   0.690456   2.586187     0.037776     0.147918
scaleCoeff2   0.093538   0.118999   0.037776     0.075997     0.028677
scaleCoeff3   0.070218   0.092110   0.147918     0.028677     0.021281
```

```
shapeCoeff   -0.055148  -0.065369  -0.077709    -0.008749    -0.008345
             shapeCoeff
locCoeff1    -0.055148
locCoeff2    -0.065369
scaleCoeff1  -0.077709
scaleCoeff2  -0.008749
scaleCoeff3  -0.008345
shapeCoeff    0.014859

Optimization Information
  Convergence: successful
  Function Evaluations: 387
```

The output of model `M1` is very similar to the one of a fitted max-stable process except the spatial dependence parameters are not present. As explained in Chapter 4, it is easy to perform model selection by inspecting the following output:

```
> anova(M1, M2)

Eigenvalue(s):
12.72
1.66

Analysis of Variance Table
   MDf Deviance Df  Chisq Pr(> sum lambda Chisq)
M1   6    10763
M2   7    10763  1 0.0671                0.9455

> TIC(M1, M2)

      M1       M2
10832.63 10836.21
```

From these two outputs, we can see that the p-value for the likelihood ratio test is around 0.95 which advocates the use of model `M1`. The TIC corroborates this conclusion.

6

Conclusion

A

P-splines with radial basis functions

A.1 Model definition

Let us recall that a general definition of a p-spline is given by

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^{q} b_j(x_i - \kappa_j) \tag{A.1}$$

for some basis functions b_j and knots κ_j.

As the purpose of this document is the modelling of spatial extremes, we benefit from using *radial basis* functions. Radial basis functions depend only on the distance $|x_i - \kappa_j|$ so that a generalisation to higher dimension, i.e. $||\mathbf{x}_i - \kappa_j||$, $\mathbf{x}_i, \kappa_j \in \mathbb{R}^d$, is straightforward.

The model for p-spline with radial basis function of order p, p being odd, is

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_{m-1} x^{m-1} + \sum_{j=1}^{q} \beta_{m+j} |x - \kappa_j|^{2m-1} \tag{A.2}$$

where $p = 2m - 1$.

The fitting criterion is

$$\text{minimize } ||\mathbf{y} - \mathbf{X}\beta||^2 + \lambda^{2m-1} \beta^T \mathbf{K} \beta \tag{A.3}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{m-1} & |x_1 - \kappa_1|^{2m-1} & \cdots & |x_1 - \kappa_q|^{2m-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{m-1} & |x_n - \kappa_1|^{2m-1} & \cdots & |x_n - \kappa_q|^{2m-1} \end{bmatrix}$$

and $\mathbf{K} = \mathbf{K}_*^T \mathbf{K}_*$ with

$$\mathbf{K}_* = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \dots & 0 & |\kappa_1 - \kappa_1|^{m-1/2} & \cdots & |\kappa_1 - \kappa_q|^{m-1/2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & |\kappa_q - \kappa_1|^{m-1/2} & \cdots & |\kappa_q - \kappa_q|^{m-1/2} \end{bmatrix}$$

where the m first rows and columns of $\mathbf{K}_*$ have zeros as elements.

A.2 Fast computation of p-splines

Although this section is included in the p-splines with radial basis functions, the methodology introduced here can be successfully applied to any other basis functions [Ruppert et al., 2003].

As we stated in Section 5.1.2, for a fixed smoothing parameter λ, the fitted values are given by

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}^T\mathbf{y}$$

for some symmetric matrix $\mathbf{K}$.

Consequently, to perform automatic selection for λ by minimising the CV or GCV criterion might be computationally demanding and numerically unstable. Fortunately, the Demmler–Reinsch orthogonalisation often overcomes these issues. The following lines describe how it works in practice.

1. Obtain the Cholesky decomposition of $\mathbf{X}^T\mathbf{X}$ i.e.

   $$\mathbf{X}^T\mathbf{X} = \mathbf{R}^T\mathbf{R}$$

   where $\mathbf{R}$ is a square matrix and invertible

2. Obtain the singular value decomposition of $\mathbf{R}^{-T}\mathbf{K}\mathbf{R}^{-\mathbf{1}}$ i.e.

   $$\mathbf{R}^{-T}\mathbf{K}\mathbf{R}^{-\mathbf{1}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

3. Define $\mathbf{A} \leftarrow \mathbf{X}\mathbf{R}^{-\mathbf{1}}\mathbf{U}$ and $\mathbf{b} \leftarrow \mathbf{A}^T\mathbf{y}$

4. The fitted values are

   $$\hat{\mathbf{y}} = \mathbf{A}\frac{\mathbf{b}}{\mathbf{1} + \lambda\mathbf{\Lambda}}$$

   with corresponding degrees of freedom

   $$df(\lambda) = \mathbf{1}^T\frac{\mathbf{1}}{\mathbf{1} + \lambda\mathbf{\Lambda}}$$

Once the matrices $\mathbf{A}$ and $\mathbf{\Lambda}$ and the vector $\mathbf{b}$ have been computed, the fitted values $\hat{\mathbf{y}}$ and $df(\lambda)$ are obtained through a simple matrix multiplication. This is appealing as now the automatic selection for λ will be cheaper.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. In *Automatic Control*, volume 19 of *IEEE Transactions on*.

Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.

Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer Series in Statistics. Springers Series in Statistics, London.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Springer, editor, *Dependence in Probability and Statistics*, volume 187, pages 373–390. Springer, New York, lecture notes in statistics edition.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.

Cramér, H. (1946). *Mathematical Methods of Statistics*, volume 9 of *Princeton Mathematical Series*. Princeton University Press.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. John Wiley & Sons inc., New York.

Davison, A. (2003). *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Diggle, P., Ribeiro, P., and Justiniano, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. In *CBMS-NSF Regional Conference Series in Applied Mathematics*, Philadelphia. SIAM.

Gholam-Rezaee, M. (2009). *Spatial extreme value: A composite likelihood*. PhD thesis, École Polytechnique Fédérale de Lausanne.

Haan, L. D. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

Hosking, J., Wallis, J., and Wood, E. (1985). Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. *Technometrics*, 27(3):251–261.

Kent, J. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69:19–27.

Lindsay, B. (1988). *Composite likelihood methods*. Statistical Inference from Stochastic Processes. American Mathematical Society, Providence.

Mathéron, G. (1987). Suffit-il, pour une covariance, d'être de type positif ? *Sciences de la Terre, série informatique géologique*, 26:51–66.

Naveau, P., Guillou, A., and Cooley, D. (2009). Modelling pairwise dependence of maxima in space. *Submitted to Biometrika*.

Nelder, J. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, 7:308–313.

Neyman, J. and Pearson, E. (1933). On the problem of the most efficient tests of statistical hypothesis. *Philosophical Transactions of the Royal Statistical Society of London. Series A*, 231:289–337.

Padoan, S. (2008). *Computational Methods for Complex Problems in Extreme Value Theory*. PhD thesis, University of Padova.

Padoan, S., Ribatet, M., and Sisson, S. (2008). Likelihood-based inference for max-stable processes. *Submitted to the Journal of the American Statistical Association (Theory & Methods)*.

Pickands, J. (1981). Multivariate Extreme Value Distributions. In *Proceedings 43rd Session International Statistical Institute*.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rotnitzki, A. and Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77:495–497.

Ruppert, D., Wand, M., and Caroll, R. (2003). *Semiparametric regression*. Cambridge University Press.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extremes: Properties and inference. *Biometrika*, 90(1):139–156.

Smith, R. (1991). Max-stable processes and spatial extreme. *Unpublished manuscript*.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Wood, S. (2006). *Generalized Additive Models*. Chapman & Hall.

Index

covariance function
    cauchy, 7
    elliptical, 8
    powered exponential, 7
    range, 8
    sill, 8
    smooth, 8
    Whittle–Matérn, 7
cross-validation, 39
    generalized, 39

degrees of freedom, 36, 38
design matrix, 35
distance
    Euclidean, 7
    Mahanalobis, 5

eigen-decomposition, 5
eigenvalues, 6
extremal coefficient, 11
    function, 12

Fréchet
    unit, 3

hat matrix, 36

information criterion
    AIC, 28
    TIC, 28
intensity measure, 3, 6

Jacobian, 34

Kullback–Leibler discrepancy, 23

least squares, 21, 35
likelihood ratio statistic, 29
linear model, 35

madogram, 14
    F-madogram, 16
    λ-madogram, 18
max-stable
    process, 3
    property, 4
misspecification, 22

p-splines, 37
pairwise-likelihood, 22
Poisson process, 3

rainfall-storm process, 3, 4

score equation, 25
smoother matrix, 38
smoothing parameter, 38
standard errors, 25

variogram, 11