Spatial modelling of extreme precipitation in the Basque country, Spain

Kate Lee and Paul Cowpertwait

School of Computing and Mathematical Science, AUT

Bayes on the beach 2012

Outline

1. Introduction
2. Univariate/multivariate extremes
3. Inference
4. Application to precipitation data : Basque country, Spain
5. Conclusion and issues

Introduction

Basic problem :
Simplest case, $X_1, X_2, \cdots, X_n \sim F$. Require accurate inference on tail F.

- Few observations in the tail of the distribution.
- Standard density estimation technique is often biased in estimating tail probabilities.
- Base tail models on asymptotically-motivated distributions : Statistics of extremes

Applications :

- Environmental : sea levels, pollution concentration, precipitation levels (rainfall, snow), river flow
- Reliability modelling : finance, insurance, telecommunication, ...

Univariate extremes

Framework for block maxima distribution Let $Y_1, \cdots, Y_n \sim F$ and define

$$M_n = \max\{Y_1, \cdots, Y_n\}\ .$$

The distribution of M_n is

$$Pr\{M_n \leq x\} = Pr\{Y_1 \leq y, \cdots, Y_n \leq y\} = F(y)^n$$

Approximate F^n by limit distribution as $n \to \infty$.

Extremal types theorem

For an i.i.d. sample $Y_1, \ldots, Y_n \in \mathbb{R}$.
If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

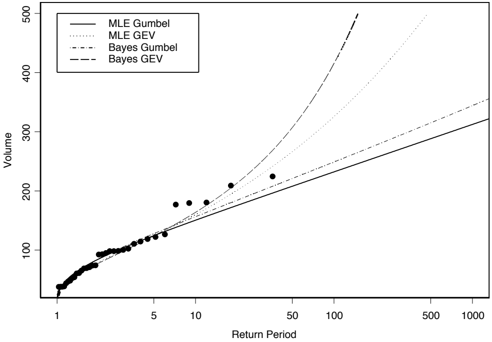
$$\lim_{n\to\infty} \Pr\left(\frac{\max\{Y_i\} - b_n}{a_n} \leq z\right) \to G(z)$$

is a non degenerate distribution function, then G is a generalised extreme value distribution (GEV)

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}.$$

- In analogy with central limit theorem for mean of samples

- In practice: fit GEV to sample maxima & predict levels of future extremes
- E.g. prob of maximum daily (n=365.25) rainfall exceeding a given level in the next 100 years



- Extension to spatial data via hierachical models
    - Spatially/temporally varying parameters
    - Spatial dependence through parameters only
    - No dependence at data level
- Annual maxima of daily spatial rainfall might be approximated by a stationary max-stable process
    - Still spatially/temporally varying parameters
    - Dependence at data level retained

Multivariate extremes

Max stable process (de Haan 1984)

A max-stable process $Z(x)$ is the limit process of maxima of i.i.d. random fields $Y_i(x)$, $x \in \mathbb{R}^d$

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

for two sequences of functions $a_n(\cdot) > 0$, $b_n(\cdot)$ (where they exist).

- Max-stable processes generalize multivariate extreme value distributions to the infinite dimensional case
- Incorporates dependence between locations x
- Can obtain GEV margins for each $x \in \mathbb{R}^d$ (assume unit Frećhet margins for the moment ...)

Max stable process (de Haan 1984)

A max-stable process $Z(x)$ is the limit process of maxima of i.i.d. random fields $Y_i(x)$, $x \in \mathbb{R}^d$

$$Z(x) = \lim_{n\to\infty} \frac{\max\{Y_i(x)\} - b_n(x)}{a_n(x)}, \qquad x \in \mathbb{R}^d \tag{2}$$

for two sequences of functions $a_n(\cdot) > 0$, $b_n(\cdot)$ (where they exist).

- Without loss of generality, if $a_n(x) = n, b_n(x) = 0$, $\{Z(x)\}_{x\in\mathbb{R}^d}$ has a unit Fréchet margins with distribution function $F(z) = \exp(-1/z)$, $z > 0$.
- If $\{Z(x)\}_{x\in\mathbb{R}^d}$ is a stationary process, it can be expressed through spectral representation (de Haan and Pickands 1986).

Spectral representation

- Smith's storm model (Smith 1990)
- Schlather's model (Schlather 2002)

Smith's model

Let $\{\tau_i, k_i\}$ denote a non-homogeneous Poisson process on $\mathbb{R}^2 \times \mathbb{R}_+$ with intensity measure $k^{-2}dk \times \mu(d\tau)$ (μ is a +ive measure) then

$$Z(x) = \max_i k_i f(x - \tau_i) \qquad x \in \mathbb{R}^2$$

where f is a unimodal continuous pdf.
"Storm profile model" interpretation:
k_i=storm magnitude, τ_i=storm location, f=storm shape.

For $f = N_d(0, \Sigma)$, the bivariate CDF is
$Pr(Z(x_1) \leq z_1, Z(x_2) \leq z_2) =$
$$\exp\left[-\frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_2}{z_1}\right) - \frac{-1}{z_2}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_1}{z_2}\right)\right]$$
where Φ is a standard normal CDF and $a^2 = \Delta x^T \Sigma^{-1} \Delta x$, $h = (x_2 - x_1)$.

Schlather's model

Let $\{Y_i\}$ denote a stationary process on $\mathbb{R}^2$ such that $\mathbb{E}[\max\{0, Y(x)\}] = 1$ and $\{k_i\}$ be the points of a Poisson process on $\mathbb{R}_+$ with intensity measure $k^{-2}dk$ then

$$Z(x) = \max_i k_i \max\{0, Y_i(x)\} \ .$$

Taking Y_i to be a stationary standard Gaussian process with correlation function $\rho(h)$, the bivariate CDF is

$$Pr(Z(0) \le z_1, Z(h) \le z_2) =$$

$$\exp\left[\frac{-1}{2}\left(\frac{1}{z_1}+\frac{1}{z_2}\right)\left(1+\sqrt{1-2(\rho(h)+1)\frac{z_1 z_2}{(z_1+z_2)^2}}\right)\right] \ .$$

Correlation functions : Three common parametric families for ρ are

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

where c_1, c_2 and ν are the sill, range and smooth parameters.

Inference

- Estimator : probability-weighted moments, likelihood-based techniques
  For max-stable processes, as only the bivariate densities are known we will consider the pairwise likelihood

  $$logL_p(y;\theta) = \sum_{j} \sum_{i<j} \sum_{k=1}^{n} logf(y_k^i, y_k^j)$$

  (Lindsay 1998, Varin 2008)
- Model comparison : graphical model checking, composite likelihood information criterion (CLIC)
- Return levels and return periods
- Spatial dependent measures : the extremal coefficient function, geostatistics based approaches (F-madogram, λ-madogram)

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

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

where $1 \leq \theta_N \leq N$ with the lower and upper bounds corresponding to complete dependence and independence.

- θ_N can be regarded as the effective number of independent stations.

- We will especially focus on pairwise extremal coefficients

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

  and $\theta(\cdot)$ is the extremal coefficient function.
- The extremal coefficient functions for the two models are

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

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

  NOTE : Schlather's model has an upper bound of $1 + \sqrt{1/2}$ for a positive ρ.

Case study : Basque country, Spain

- Daily precipitation records for 97 years, 1914-2010, over 234 catchments.
- Unbalanced records are infilled using a spacial-temporal model (Cowpertwait 206).

- Fréchet margin transformation : spatial GEV model (longitude, latitude, and altitude), AIC

$$g : Y(x) \rightarrow \left(1 + \xi(x)\frac{Y(x) - \mu(x)}{\sigma(x)}\right)^{1/\xi(x)}$$

  where $\mu(x), \sigma(x)$ and $\xi(x)$ are linear dependent to x. Hence, the bivariate distribution is rewritten as

$$Pr[Y(x_1) \leq y_1, Y(x_2) \leq y_2] = Pr[Z(x_1) \leq g(y_1), Z(x_2) \leq g(y_2)] .$$

- Spatial extreme model selection : composite likelihood information criteria (CLIC)

Figure: Spatial GEV model diagnostic plots

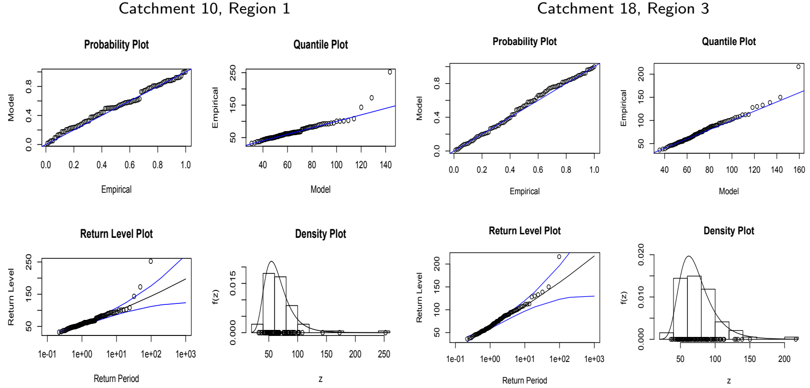
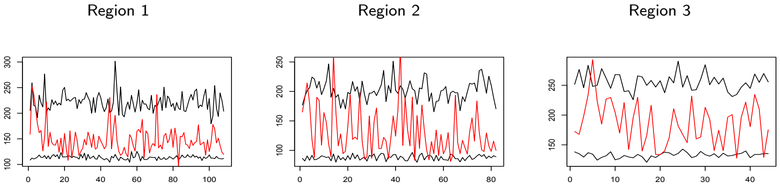


Figure: Spatial GEV model diagnostic plots; Return level of one in 97 years, 95% confidence intervals (red) and observations (black)



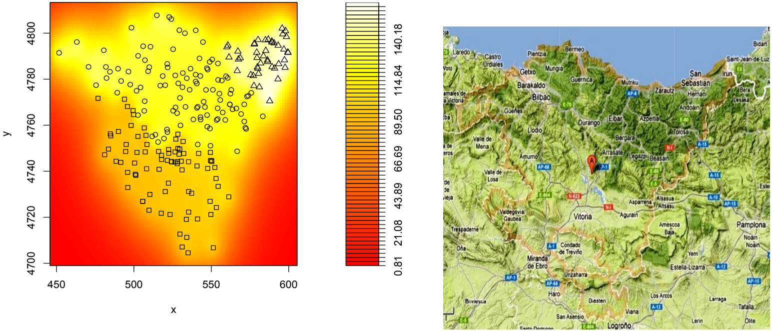
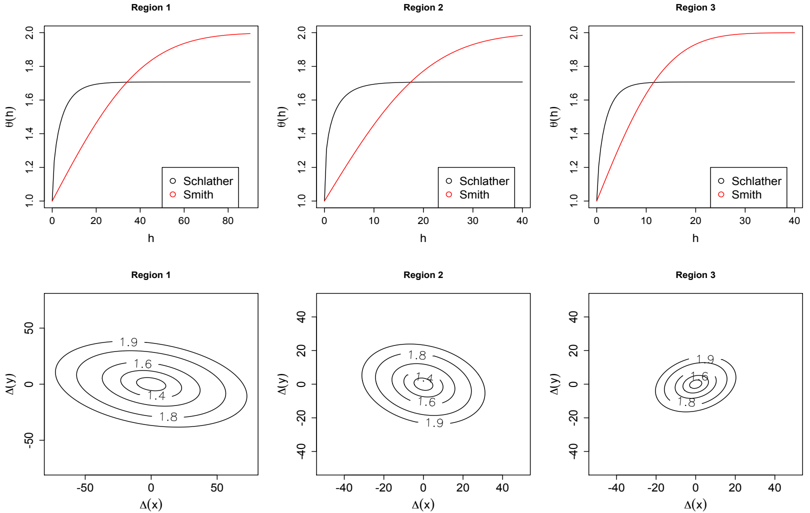


Figure: Left : Pointwise return levels of one in 50 years events. The sites located in the regions 1,2 and 3 are indicated by the circle, square, and triangle marks respectively. The units for x and y are in kilometers.

- Extremal coefficient estimates, θ_{Smith} and $\theta_{Schlather}$.



- Extremal coefficient estimates, θ_{Smith} and $\theta_{Schlather}$

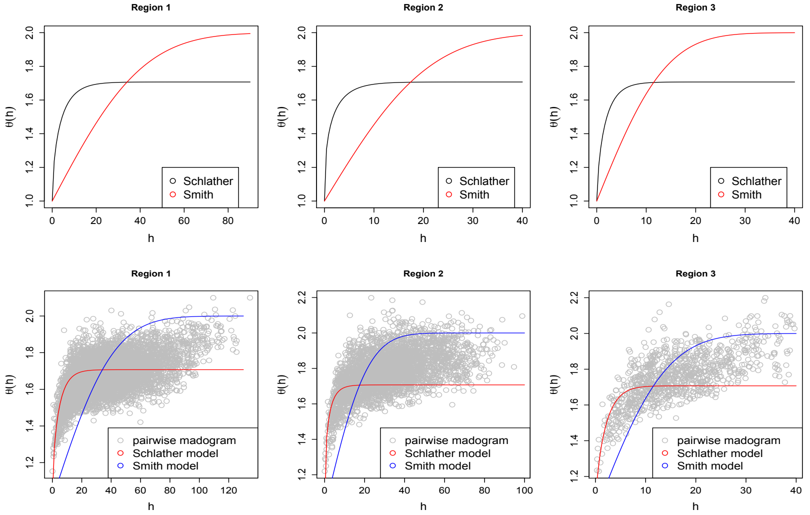
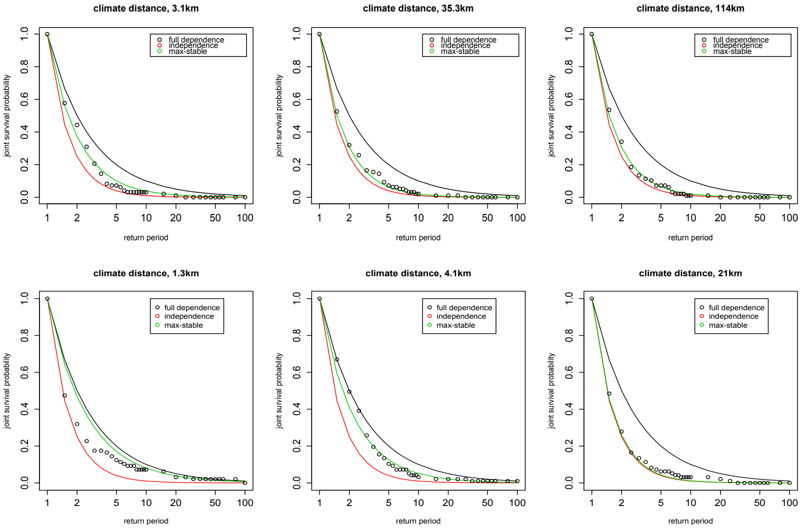


Figure: Risk analysis of pairwise annual maxima: joint survival probability versus return period.



Conclusion and issues

Conclusion :

- Max-stable models are powerful ad mathematically justified models of spatial extremes.
- More realistic predictions over hierachical models.
- Application to the precipitation in Basque country, Spain.
- Design storm : Average storm over the region (sort of).

Issues :

- Fréchet margin transformation techniques
- Infilled data technique (due to irregular observations) : splines (Neville, S. E. et al, 2011)

References

COLES, S. (2001). An Introduction to Statistical Modelling of Extreme Values. Springer, New York.

COOLEY, D., NAVEAU, P. and POUCET, P. (2006). Varlograms for spatial max-stable random fields. In Springer, editor, *Dependence in Probability and Statistics*, 187, 373-390. Springer, New York, lecture notes in statistics editions.

COWPERTWAIT, P. (2006). A spatial-temporal point process model of rainfall for the Thames catchment, UK. Journal of Hydrology, 330 586-595.

DE HAAN, L. (1984). A spectral representation for max-stable processes. Ann. Probab. 12 1194204.

DE HAAN, L. and PICKANDS, J. (1986). Stationary min-stable stochastic processes. Probability Theory and Related Fields, 74, 47792.

PADOAN, S., RIBATET, M. and SISSON, S. (2010). Likelihood-based inference for max-stable processes. J. Amer. Statist. Assoc. 105 26377.

RIBATET, M. (2009). A user's guide to the SpatialExtrees package. *Available as a part of the package or at the package web page.*

SCHLATHER, M. (2002). Models for stationary max-stable random fields. Extremes 5 334.

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

Thank you for your attention!

This work has been done by using the R package *SpatialExtremes*

`http://spatialextremes.r-forge.r-project.org/`