

Spatial Extremes and High Quantile Estimation of Aggregated Rainfall

Ana Ferreira

Abstract We give a brief overview of some basic concepts of Extreme Value Theory in Infinite Dimensional Space, and present a limiting tail probability result for the integral of a stochastic process. This result leads to ways of estimating high quantiles (or similarly return periods, or tail probabilities) related to that integral. The theory is applied to daily rainfall data over the Venice lagoon region (Italy). We estimate the 100-year quantile pertaining to the daily total amount of rain for that region.

Key words: Extreme Value Theory, Spatial Data Analysis, Quantile and Tail Probability Estimation

1 Introduction

Let $X := \{X(s)\}_{s \in S}$ be some stochastic process on $C(S)$, the space of continuous functions on S , with S some compact subset of $\mathbb{R}^d$ ($d \geq 1$). For example the process X may represent daily rainfall over the space S . A quantity of interest is then the daily total amount of rain over that region, $\int_S X(s) ds$, and we are interested in studying extremes related to it. For example in studying a high quantile, i.e. the quantity x such that,

$$p = P\left(\int_S X(s) ds > x\right),$$

for a given small value p . Or, similarly one defines p as the tail probability associated to a given high value x . Note that in applications in extreme value theory, we are dealing with a value x such that none or very few of the observed processes X verify the requirement $\int_S X(s) ds > x$. Since we want to keep this feature in the asymptotic analysis, we impose conditions such that this holds in the model. This leads us to assume that $p = p_n$ and $p_n = O(1/n)$, where n is the number of available

observations or, similarly, that $x = x_n$ and x_n converges to the right endpoint of the distribution of $\int_S X(s) ds$.

In this paper we start by giving some basic concepts of Extreme Value Theory in Infinite Dimensional Space and, afterwards, present a limiting tail probability for $\int_S X(s) ds$ (Section 2). From this result one obtains estimators for high quantiles or tail probabilities related to $\int_S X(s) ds$ (Section 3). Finally we apply the results to daily rainfall data over the Venice lagoon region, Italy (Section 4). The rainfall data is available at some fixed locations in space, the monitoring stations. Based on this observations we estimate the 100-year quantile related to the daily total amount of rain for that region.

Original ideas for aggregating the original process X are from Coles and Tawn (1996), with recent developments from Ferreira, Chen and de Haan (2010).

2 Some Theory: Max-Stable Processes, Domain of Attraction and Tail Probability of $\int_S X(s) ds$

Max-Stable Processes

A process $Y := \{Y(s)\}_{s \in \mathbb{R}^d}$ with non-degenerate marginals is called *max-stable* if, there are continuous functions $c_s(n)$ positive and $d_s(n)$ real and if $Y_1, Y_2, \dots$, are independent and identically distributed (i.i.d.) copies of Y , then

$$\bigvee_{i=1}^n \frac{Y_i - d_n}{c_n} \stackrel{d}{=} Y, \quad n = 1, 2, \dots \quad (1)$$

Of course that if (1) holds, then it holds marginally, and we shall call a process *simple max-stable* when its marginal distributions are standard Fréchet.

A max-stable process Y can be represented via a *simple max-stable* process, η say (hence $P(\eta(s) \leq x) = \exp(-1/x)$, $x > 0$, for all s), by

$$\{Y(s)\}_{s \in \mathbb{R}^d} = \left\{ \frac{\eta(s)^{\gamma(s)} - 1}{\gamma(s)} \right\}_{s \in \mathbb{R}^d}, \quad (2)$$

with $\gamma(s) \in \mathbb{R}$, the extreme value index function, known to be a continuous function.

Different ways to characterize and represent max-stable processes are known, some more abstract or more developed than others, a subject under current investigation, though it can go back to e.g. de Haan (1984), de Haan and Pickands (1986), Smith (1990) and Resnick and Roy (1991).

Just to give an example, a subclass of stationary max-stable processes with a sufficiently simple representation (at least for applications) is the class of *moving max-ima processes* (e.g. de Haan and Pereira (2006) and note that this includes Smith's (1990) model)

$$Y(s) = \bigvee_{i=1}^{\infty} \frac{\phi(T_i - s)}{U_i}, \quad s \in \mathbb{R}^d$$

with ϕ a unimodal continuous probability density and $\{(T_i, U_i)\}_{i=1}^{\infty}$ a realization of a homogeneous Poisson process on $\mathbb{R}^d \times \mathbb{R}_+$.

Domain of Attraction Condition

Let S be some compact set of $\mathbb{R}^d$ ($d \geq 1$), and $C(S)$ the space of continuous functions on S , equipped with the supremum norm, $|f|_{\infty} = \sup_{s \in S} |f(s)|$. Let $X := \{X(s)\}_{s \in S}$ be a stochastic process on $C(S)$ with non-degenerate marginals.

The following condition means that X is in the domain of attraction of the max-stable process Y : if for $X_1, X_2, \dots$, i.i.d. copies of X and continuous functions $a_s(n)$ positive and $b_s(n)$ real,

$$\left\{ \max_{1 \leq i \leq n} \frac{X_i(s) - b_s(n)}{a_s(n)} \right\}_{s \in S} \xrightarrow{d} \{Y(s)\}_{s \in S}, \quad n \rightarrow \infty, \quad (3)$$

for some limiting stochastic process $Y := \{Y(s)\}_{s \in S}$, in the $C(S)$ space. Of course that if (1) holds, then it holds marginally, i.e. the maximum domain of attraction condition holds in the finite dimensional framework.

Some implications of (3) which will be useful in the sequel are the following:

1. The existence of the *exponent measure*, ν say, on the space

$$C^+(S) := \{f \in C(S) : f \geq 0\}.$$

It is such that, for each Borel subset E of $C^+(S)$ with $\inf\{|f|_{\infty} : f \in E\} > 0$ and $\nu(\partial E) = 0$,

$$\lim_{t \rightarrow \infty} tP \left(\left\{ \left(1 + \gamma(s) \frac{X(s) - b_s(t)}{a_s(t)} \right)_+^{1/\gamma(s)} \right\}_{s \in S} \in E \right) = \nu(E), \quad (4)$$

the limit being finite (notation $x_+ := \max(0, x)$).

2. The existence of the *spectral measure*, ρ say, on the space

$$\bar{C}_1^+(S) := \{g \in C^+(S) : |g|_{\infty} = 1\}.$$

It is such that,

$$\nu(E) = \int \int_{rg \in E} \frac{dr}{r^2} d\rho(g) \quad \text{and} \quad \int_{\bar{C}_1^+(S)} g(s) d\rho(g) = 1, \quad \text{for all } s \in S. \quad (5)$$

A more comprehensive exposition of these results may be found in de Haan and Lin (2001) and de Haan and Ferreira (2006).

Tail Probability of $\int_S X(s) ds$

Next we give the main result characterizing the tail distribution of $\int_S X(s) ds$, which will be the basis for the estimation procedure to analyse extreme aggregated rainfall over the entire region.

We need two kinds of homogeneity over the space, in order that the result holds:

1. On the shape function: the max-domain of attraction condition must hold with a constant index function, i.e. $\gamma(s) = \gamma \in \mathbb{R}$ for all $s \in S$.
2. The scaling function $a_s(t)$ must verify: for some positive functions $a(t)$ and $A(s)$,

$$\sup_{s \in S} \left| \frac{a_s(t)}{a(t)} - A(s) \right| \rightarrow 0, \quad \text{as } t \rightarrow \infty. \quad (6)$$

Recall: S is a compact set of R^d ($d \geq 1$), $C(S)$ the space of continuous functions on S and $X := \{X(s)\}_{s \in S}$ on $C(S)$ has non-degenerate marginals.

Theorem 1 (Ferreira, de Haan and Chen (2010)). *Let the stochastic process X satisfy the basic convergence (3) with $\gamma(s) \equiv \gamma \in \mathbb{R}$ and $a_s(t)$ verifying (6). Require: for $\gamma = 0$, $a_s(t) = a(t)A(s)$ for all t and s ; for $\gamma \leq 0$, $\rho\{g \in \bar{C}_1^+(S) : \inf_{s \in S} g(s) = 0\} = 0$; for $\gamma > 0$, that X is non-negative.*

Then,

$$\lim_{t \rightarrow \infty} tP \left(\frac{\int_S X(s) ds - \int_S b_s(t) ds}{\int_S a_s(t) ds} > x \right) = \theta_\gamma (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0, \quad (7)$$

with

$$\theta_\gamma := \int_{\bar{C}_1^+(S)} \left(\int_S A(s) g^\gamma(s) ds \right)^{1/\gamma} d\rho(g); \quad (8)$$

for $\gamma = 0$ the right-hand side of (7) should be read as $\theta_0 e^{-x}$ and the right-hand side of (8) as $\int_{\bar{C}_1^+(S)} \exp(\int_S A(s) \log g(s) ds) d\rho(g)$. We have $\theta_\gamma > 0$ for all γ .

Note the resemblance between (7) and the max-domain of attraction condition on the one dimensional extreme value framework in the Generalized Pareto form (in the following W is a random variable): if for some real functions $a(t) > 0$ and $b(t)$

$$\lim_{t \rightarrow \infty} tP \left(\frac{W - b(t)}{a(t)} > x \right) = (1 + \gamma x)^{-1/\gamma}, \quad \text{for all } x \text{ with } 1 + \gamma x > 0, \quad \gamma \in \mathbb{R}. \quad (9)$$

The extra factor θ_γ that appears in (7) was named in Coles and Tawn (1996), the *areal coefficient* and interpreted as the effect of spatial dependence.

3 High Quantile and Tail Probability Estimation

On the basis of an i.i.d. sample of size n from X , we suggest the following estimators for a high quantile x_n (given $p = p_n$) and for a tail probability p_n (given $x = x_n$):

$$\hat{x}_n = \int_S \hat{b}_s\left(\frac{n}{k}\right) ds + \int_S \hat{a}_s\left(\frac{n}{k}\right) ds \frac{\left(\frac{np_n}{k\hat{\theta}}\right)^{-\hat{\gamma}} - 1}{\hat{\gamma}} \quad (10)$$

$$\hat{p}_n = \frac{k}{n} \hat{\theta} \left(1 + \hat{\gamma} \frac{x_n - \int_S \hat{b}_s\left(\frac{n}{k}\right) ds}{\int_S \hat{a}_s\left(\frac{n}{k}\right) ds} \right)^{-1/\hat{\gamma}} \quad (11)$$

with

$$\hat{\theta} = \int_{\bar{C}_1^+(s)} \left(\int_S \hat{A}(s) g^{\hat{\gamma}}(s) ds \right)^{1/\hat{\gamma}} d\hat{p}(g), \quad (12)$$

and $k = k(n)$ an intermediate sequence, i.e. $k \rightarrow \infty$ and $k/n \rightarrow 0$, as $n \rightarrow \infty$. The ‘intermediate’ estimators $\hat{a}_s(\frac{n}{k})$ and $\hat{b}_s(\frac{n}{k})$ come from marginal estimation and, $\hat{\gamma}$ and $\hat{A}(s)$ are obtained by combining appropriately over the space marginal estimation of shape and scale respectively, as explained next. For $\hat{p}$ we use an ‘empirical-type’ estimator,

$$\hat{p}(A) := \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{\left\{ \sup_{s \in S} R(X_i(s)) > n+1-k \text{ and } \left\{ \frac{n+1 - \sup_{s \in S} R(X_i(s))}{n+1 - R(X_i(s))} \right\}_{s \in S} \in A \right\}}$$

where $R(X_i(s))$ is the rank of $X_i(s)$ among $(X_1(s), \dots, X_n(s))$.

Under univariate extreme value theory, many estimators have been proposed for the scale, location and shape parameters, with somehow comparative properties. In the application we will have to estimate these parameters at each monitoring station, on the basis of each data set of each station. Recall that if (3) holds then it also holds marginally. In the application we use the well-known moment-type estimators (Dekkers, Einmahl and de Haan (1989) and more recently Einmahl and Lin (2006) under the infinite dimensional setting):

$$\hat{\gamma}(s) = M_n^{(1)}(s) + 1 - \frac{1}{2} \left\{ 1 - \frac{\left(M_n^{(1)}(s) \right)^2}{M_n^{(2)}(s)} \right\}^{-1} \quad (13)$$

where $M_n^{(j)}(s) = k^{-1} \sum_{i=0}^{k-1} (\log X_{n-i,n}(s) - \log X_{n-k,n}(s))^j$, $j = 1, 2$, with $X_{i,n}$ representing the order statistics,

$$\hat{a}_s\left(\frac{n}{k}\right) = X_{n-k,n}(s) M_n^{(1)}(s) \frac{1}{2} \left\{ 1 - \frac{\left(M_n^{(1)}(s)\right)^2}{M_n^{(2)}(s)} \right\}^{-1},$$

$$\hat{b}_s\left(\frac{n}{k}\right) = X_{n-k,n}(s).$$

For the "global" scale and shape parameters we use

$$\hat{\gamma} = \int_S \hat{\gamma}(s) ds / |S|, \quad (14)$$

$$\hat{A}(s) = \hat{a}_s\left(\frac{n}{k}\right) / \int_S \hat{a}_s\left(\frac{n}{k}\right) ds, \quad (15)$$

with $|S|$ representing the total area.

The estimator $\hat{p}_n$ in (11) is known to be consistent in the sense that $\hat{p}_n/p_n \rightarrow 1$, as $n \rightarrow \infty$, under some conditions (cf. Ferreira, de Haan and Chen (2010)): in particular under the conditions of Theorem 1, $\gamma > -0.5$ and the 'usual' growth conditions on the sequence k from the univariate estimation of high quantiles.

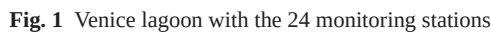
4 Application to Rainfall Data

The estimators will be applied to high quantile estimation pertaining to aggregated daily rainfall over the drainage basin of the Venice lagoon (Figure 1). The analysis is based on daily rainfall data at 24 monitoring stations from that region, over a time period of 55 years (1940–1994), from the months September–November, i.e. 91 days over 55 years giving a total of 5005 observations for each station (though still with missing values). We shall estimate the 100-year quantile i.e., that value that is exceeded once in 100 years. As we have 91 observations per year it corresponds to a given value of $p_n = 1/9100$.

The considered stations are (with the respective number): (47) Botti Barbarighe, (49) Cavanella Motte, (50) Chioggia, (53) Stanghella, (56) Bagnili di Sopra, (57) Battaglia Terme, (59) Este, (63) Bovolenta, (67) S. Margherita di Codevigo, (69) Curtarolo, (70) Gambarare, (71) Mirano, (72) Padova, (92) Castelfranco Veneto, (93) Cittadella, (97) Massanzago, (98) Mestre, (99) Mogliano Veneto, (100) Piombino Dese, (101) San Nicolo' di Lido, (102) Stra, (106) Portesine, (107) Treviso, (108) Zuccarello.

Since we only have observations of the process at some fixed (and finite) points in the space (the monitoring stations), for the overall estimation one has to extrapolate 'information' over the space. In the calculations, the stations are connected with lines, the area is partitioned in triangles, and 'information' is extrapolated linearly over the space (for more details see the Application's section in Ferreira, de Haan and Chen (2010)).

One has first to proceed with marginal data analysis. That is, we first perform 'standard' univariate extreme value analysis at each station separately. A typical issue then is the choice of the threshold, above which generalized Pareto distributions



Proceeding with the estimation, the intermediate results of estimates of the global shape (cf. 14) and of the dependence coefficient θ (cf. (15)) are shown respectively in Figures 4 and 5.

Acknowledgements I thank Chen Zhou for providing me part of the R-code for the application.

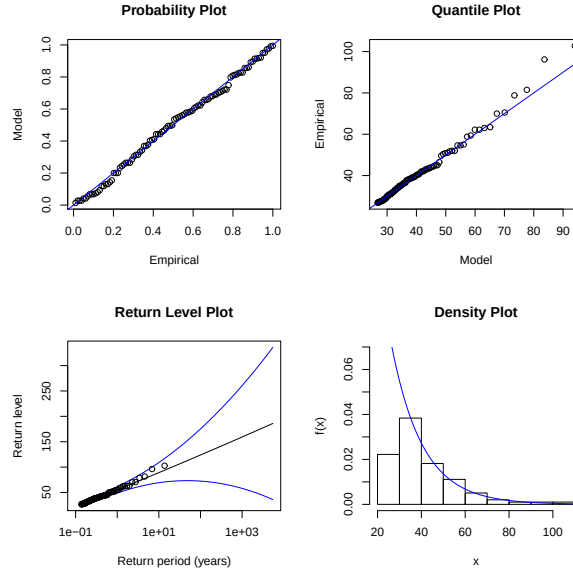


Fig. 2 Standard Generalized Pareto fit to the data from Station 49 (Cavanella Motte), with a threshold of 26.6 ($k = 100$); maximum likelihood estimation from the package ismev from R-software gives the marginal estimates: for the scale 14.32 (s.e.=2.10) and for the shape 0.01 (s.e.=0.11)

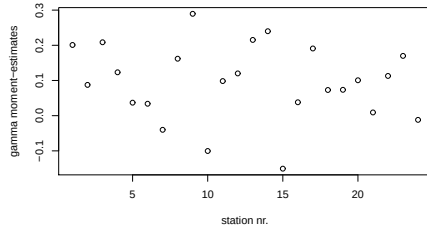


Fig. 3 Moment estimates of gamma versus the rank of station-numbers

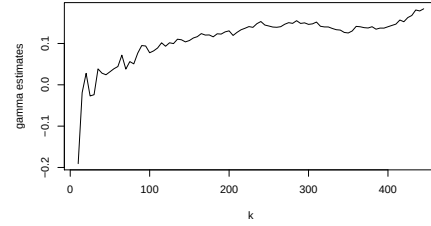


Fig. 4 Estimates of the ‘global’ shape parameter versus the number of upper order statistics k

References

1. Brown, B. and Resnick, S.: Extreme values of independent stochastic processes. J. Appl. Probab. **14**, 732–739 (1977)
2. Coles, S.G. and Tawn, J.A.: Modelling extremes of the areal rainfall process. J. R. Statist. Soc. B **58**, 329–347 (1996)
3. Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L.: A moment estimator for the index of an extreme-value distribution. Ann. Statist. **17**, 1833–1855 (1989)
4. Einmahl J.H.J. and Lin, T.: Asymptotic normality of extreme value estimators on $C[0,1]$. Ann. Statist. **34**, 469–492 (2006)

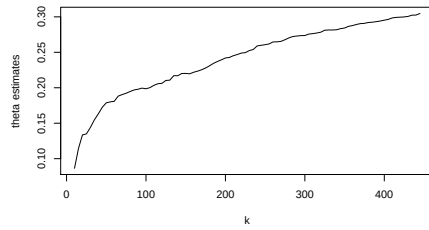


Fig. 5 Estimates of the areal dependence parameter

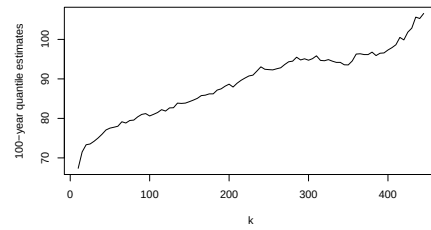


Fig. 6 100-year quantile estimates (scaled down by the total area) of the total daily rainfall over the area

5. de Haan, L.: A spectral representation for max-stable processes. *Ann. Prob.* **12**, 1194–1204 (1984)
6. de Haan, L. and Lin, T.: On convergence toward an extreme value distribution in $C[0,1]$. *Ann. Probab.* **29**, 467–483 (2001)
7. de Haan, L. and Pereira, T.T.: Spatial Extremes: the stationary case. *Ann. Statist.* **34**, 146–168 (2006)
8. de Haan, L. and Pickands, J.: Stationary min-stable stochastic processes. *Probab. Th. Rel. Fields* **72**, 477–492 (1986)
9. de Haan, L. and Ferreira, A.: *Extreme Value Theory: An Introduction*. Springer, Boston (2006)
10. Ferreira A., de Haan L. and Zhou C.: Exceedance probability of the integral of a stochastic process. Submitted (2009)
11. Smith, R.: Max-stable processes and spatial extremes. Unpublished manuscript (1990)