Estimation of extreme events from spatial rainfall data

Ana Ferreira[1,4], Rita M. Cardoso[2], Pedro M.M. Soares[2]
and Margarida Belo-Pereira[3]
[1]ISA, Univ Tecn Lisboa and CEAUL, Portugal
[2] Instituto Dom Luiz, CGUL, University of Lisbon, Portugal
[3] Instituto Português do Mar e Atmosfera, Portugal
[4]Corresponding author: Ana Ferreira, e-mail: anafh@isa.utl.pt

Abstract

Let daily rainfall over the space be represented by a stochastic process, assumed continuous on some compact space S. We discuss estimation of extreme events on the basis of observations of such processes. Of special concern are quantities related to the tail probability of the process spatially aggregated, i.e. the tail probability of the integral of the stochastic process over S, like high quantiles or return values. We apply the methodology to daily rainfall data over the northwest region of Portugal.

Keywords: Extreme value theory, northwest-Portugal precipitation, Pareto distribution, return values, spatial dependence

1 Introduction

The precipitation regime in Iberia is characterized by large intra and inter-annual variability and high spatial heterogeneity (Soares et al. 2012a, Soares et al. 2012b, Cardoso et al. 2012, Esteban-Parra et al. 1998) and is known for its sensitivity to climate change in what concerns precipitation (Giorgi, 2002). Iberian variability is partially due to the Mediterranean climate properties, but it is substantially enhanced by complex topography and coastal processes. Portugal and Spain, side by side on Iberia, share most atmospheric systems. However, being at the western edge, Portugal has significantly higher mean annual precipitation, with a large mean precipitation gradient. While the northwest region of Portugal is one of the wettest spots in Europe, with recorded mean annual accumulated precipitation in excess of 3000 mm, average rain amounts in the SE are of the order of 400 mm. Likewise other Mediterranean regions, climate is characterized by large inter-annual variability, with recurrence of precipitation extremes, drought and vulnerability to floods and desertification.

This study uses surface daily precipitation collected by the Institute of Meteorology and the Portuguese Water Institute (SNIRH, 2010). We focus on the northwest region of Portugal (cf. basin 1 in Fig. 1), and use the data collected at precipitation stations for the period 1950-2008, with maximum value of missing values of 5%. In order to overcome difficulties regarding non-stationarity we only consider the months September-October-November in each year.

We want to apply extreme value theory. Let daily precipitation over the space be represented by a stochastic process $X = \{X(s)\}_{s\in S}$ defined in the space of continuous functions, $C(S)$ with S some compact subset of $\mathbb{R}^2$. We assume the basic domain of attraction condition of max-stable processes:

Denote by $X_1, X_2, \ldots$, independent and identically distributed copies of X. Suppose there are continuous functions $a_n(s) > 0$ and $b_n(s) \in \mathbb{R}$, for all n, such that the sequence of stochastic processes

$$\left\{ \max_{1\le i\le n} \frac{X_i(s) - b_n(s)}{a_n(s)} \right\}_{s\in S} \tag{1.1}$$

converges weakly (or in distribution) in $C(S)$ to a stochastic process with non-degenerate marginals. Then it is well known that the limiting process is necessarily a max-stable

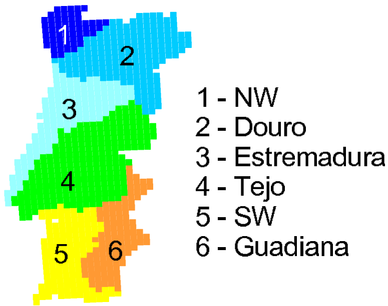

Figure 1: Main Portuguese river basins.

process (de Haan and Lin, 2001; cf. de Haan and Ferreira, 2006). Loosely speaking max-stable processes are those processes that are distribution-invariant under maxima operation, with proper linear normalization. Hence, max-stable processes are naturally used to model extreme events.

Return values are commonly used when evaluating precipitation extremes. Loosely, the $N-$year return value is that value that is expected to occur once in $N-$years; hence for N large it is basically an extreme quantile of the precipitation distribution. One of the important applications of extreme value theory is the estimation of high quantiles, where high means that we are specially interested in high values beyond the sample information. We aim to bring to discussion results on estimation of extreme events like return values over the region and for the entire region, on the basis of the daily precipitation data in the northwest region of Portugal.

In Section 2 we present some estimation results along with the used methodology. In Section 3 is a brief discussion with some extensions.

2 Results

For simplicity we shall mainly present the methodology in a univariate context. The domain of attraction condition (1.1) implies that each marginal distribution, $P(X(s) \leq x)$ for $x \in \mathbb{R}$, is in the domain of attraction of some generalized extreme value distribution. This is equivalent to the following condition on the tail of the marginal distribution and with the generalized Pareto distribution as the related limiting distribution. Suppose for some normalizing functions $a_t(s) > 0$ and $b_t(s)$,

$$\lim_{t\to\infty} tP\left(\frac{X(s) - b_t(s)}{a_t(s)} > y\right) = (1 + \gamma(s)y)^{-1/\gamma(s)}, \tag{2.1}$$

for all y with $1 + \gamma y > 0$, $\gamma \in \mathbb{R}$, $s \in S$. The function $\gamma(s)$, $s \in S$, is the extreme value index function. The extreme value index gives an indication of the heaviness of the tail of the distribution, the larger the indice the heavier the tail. Negative values refer to short tails whereas positive values refer to heavy tail distributions, with a slow decay as a power function.

We have daily data from 31 monitoring stations and over 59 years (giving a sample of size $n = 91 \times 59 = 5369$). Under the given framework the $N-$year return value, rv_N say, at each location s, is approximately calculated from the relation

$$P(X(s) > rv_N) = (91 \times N)^{-1}.$$

From (2.1) and on the basis of an independent and identically distributed sample of size n, the following is a known estimator for a high quantile with a given exceedance probability p_n (de Haan and Ferreira, 2006),

$$\hat{x}_n(s) = \hat{b}_{n/k}(s) + \hat{a}_{n/k}(s)\,\frac{\left(\frac{k}{np_n}\right)^{\hat{\gamma}(s)} - 1}{\hat{\gamma}(s)} \tag{2.2}$$

where $k = k(n)$ is an intermediate sequence, i.e. $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. In our case it will be applied to rv_N with $p_n = (91 \times N)^{-1}$. It is important to remark that one

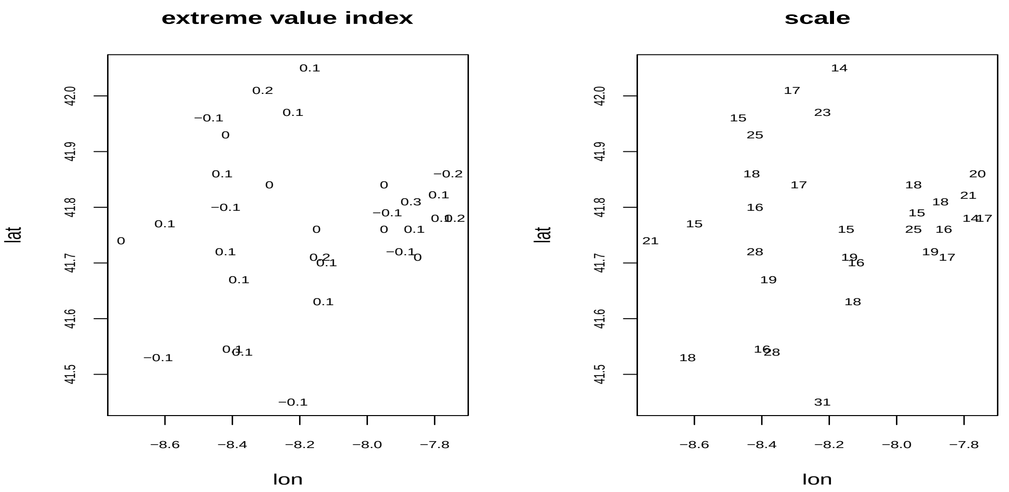

Figure 2: Local estimates over the space, with $k = 200$. Each estimate is located at the latitude-longitude coordinate of the corresponding station.

has to deal with the 'intermediate' estimators $\hat{\gamma}(s)$, $\hat{a}_{n/k}(s)$ and $\hat{b}_{n/k}(s)$, for which we use moment-type estimators (Dekkers, Einmahl and de Haan (1989) and more recently under the infinite dimensional setting Einmahl and Lin (2006)).

Consequently, for each set of n observations (i.e. at each station) we calculate $\hat{\gamma}(s)$ (i.e. the extreme value index estimates), $\hat{a}_{n/k}(s)$ (the scale estimates) and $\hat{b}_{n/k}(s)$ (the shift estimates). In Figure 2 are shown the estimates of the extreme value index and the scale over the space. In Figure 3 are the corresponding 100-year return values. In particular, the magnitudes of the return value estimates agree with the significant precipitation observed in the northwest region of Portugal. The choice of k in all the estimates involves the well-known variance-bias trade-off problem. In this case it was choosen via a simple graphic inspection of the more stable values.

3 Discussion

Another interesting measurement is the aggregation of precipitation over space. In our example this corresponds to the total amount of daily rain over the northwest basin 1. Then, instead of considering local estimates as in the previous section, it is more natural to concentrate on the total amount of rain over the whole region, represented by $\int_S X(s)ds$ with S standing for basin 1.

It is known from Ferreira, de Haan and Zhou (2012) that the tail distribution of $\int_S X(s)ds$ satisfies a limiting relation of the same type of (2.1) with an extra 'spatial' parameter. This parameter was called the areal coefficient in Coles and Tawn (1996) and was interpreted as accounting for the effect of spatial dependence. The asymptotic theory suggests for estimating the $N-$year return value,

$$\hat{r}v_N = \int_S \hat{b}_{n/k}(s)ds + \int_S \hat{a}_{n/k}(s)ds \, \frac{\left(\frac{91Nk\hat{\theta}}{n}\right)^{\hat{\gamma}} - 1}{\hat{\gamma}}. \tag{3.1}$$

Regard that one has to additionally deal with the estimation of the areal coefficient θ. Moreover note that for (3.1) it is assumed $\gamma(s) \equiv \gamma$, $s \in S$, for some $\gamma \in \mathbb{R}$; further natural assumptions are also needed, cf. Ferreira, de Haan and Zhou (2012). The discussion of estimates under this framework is beyond the scope of this paper.

Acknowledgement. Research partially supported by FCT-PEst-OE/MAT/UI0006/2011 and FCT Project PTDC /MAT /112770 /2009.

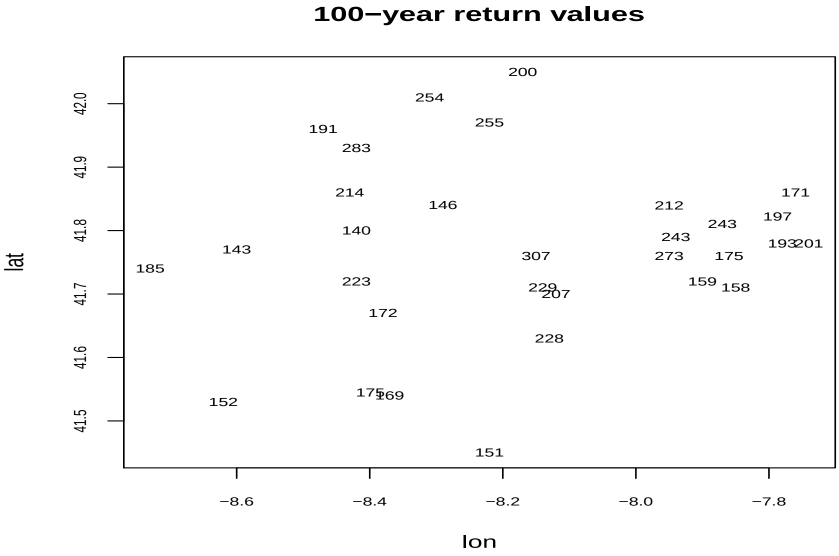

Figure 3: Local estimates (in mm) over the space, with $k = 300$. Each estimate is located at the latitude-longitude coordinate of the corresponding station.

References

[1] Cardoso, R.M., Soares, P.M.M., de Medeiros, J., Miranda, P.M.A. and Belo-Pereira, M. (2012). WRF high resolution simulation of Iberian mean and extreme precipitation climate. International Journal of Climatology, DOI: 10.1002/joc.3616.

[2] Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. J. R. Statist. Soc. B, 58, 329–347.

[3] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833–1855.

[4] Einmahl J.H.J. and Lin, T. (2006). Asymptotic normality of extreme value estimators on C[0,1]. Ann. Statist. 34, 469–492.

[5] Esteban-Parra, M. J., Rodrigo, F. S. and Castro-Dez, Y. (1998). Spatial and temporal patterns of precipitation in Spain for the period 1880-1992, Int. J. Climatol., 18, 1557-1574.

[6] Giorgi, F. (2002). Variability and trends of sub-continental scale surface climate in the twentieth century. Part I: Observations, Clim. Dyn., doi:10.1007/s00382-001-0204-x.

[7] de Haan, L. and Lin, T. (2001) On convergence toward an extreme value distribution in C[0,1]. Ann. Probab. 29, 467–483.

[8] de Haan, L. and Ferreira, A. (2006) Extreme Value Theory: An Introduction. Springer, Boston.

[9] Ferreira, A., de Haan, L. and Zhou, C. (2012). Exceedance probability of the integral of a stochastic process. Journal of Multivariate Analysis 105, 241–257.

[10] Sistema Nacional de Informacão de Recursos Hídricos (2010). Available at http://snirh.pt/

[11] Soares, P. M. M., Cardoso, R. M., Miranda, P. M. A., Viterbo, P. and Belo-Pereira, M. (2012) Assessment of the ENSEMBLES regional climate models in the representation of precipitation variability and extremes over Portugal, J. Geophys. Res., 117, D07114, doi:10.1029/2011JD016768.

[12] Soares, P.M.M., Cardoso, R.M., de Medeiros, J., Miranda, P.M.A., Belo-Pereira M. and Espirito-Santo, F. (2012) WRF high resolution dynamical downscaling of ERA-Interim for Portugal. Climate Dynamics, 39:2497-2522, DOI: 10.1007/s00382-012-1315-2.