Advances in Geosciences

Bayesian trend analysis for daily rainfall series of Barcelona

M. I. Ortego[1], **J. Gibergans-Báguena**[1], **R. Tolosana-Delgado**[2], **J. J. Egozcue**[1], **and M. C. Llasat**[3]

[1]Universitat Politècnica de Catalunya, Departament de Matemàtica Aplicada III, Barcelona, Spain
[2]Universitat Politècnica de Catalunya, Laboratori d'Enginyeria Marítima, Barcelona, Spain
[3]Universitat de Barcelona, Departament d'Astronomia i Meteorologia, Barcelona, Spain

Abstract. A Point-Over-Threshold approach using a reparameterization of the Generalized Pareto Distribution (GPD) has been used to assess changes in the daily rainfall Barcelona series (1854–2006). A Bayesian approach, considering the suitable scale and physical features of the phenomenon, has been used to look for these alterations. Two different models have been assessed: existence of abrupt changes in the new GPD parameters due to modifications of the observatory locations and trends in these GPD parameters, pointing to a climate change scenario.

1 Introduction

Frequency analysis of hydrological series is a key point to acquire an in-depth understanding of the behavior of hydrologic events. The occurrence of extreme hydrologic events in an area may imply great social and economical impacts. Therefore, modelling extreme values is of great interest in all environmental studies and civil engineering.

This work adopts a Bayesian version of the commonly used point over threshold (POT) or exceedance methods. The occurrence of events in time are assumed to be Poisson distributed and the excesses over a selected high threshold are assumed to follow a Generalized Pareto Distribution (GPD).

During the last decades, a common concern has been to know whether magnitudes of hazardous events have changed. This issue has been assessed from different points of view, both Bayesian and frequentist. Ideas of extreme value theory have been applied to enviromental studies, e.g.

Smith (1989), where the possibility of a trend was considered. Within the same study, the possibility of an abrupt change due to an improvement in measurement technology is discussed (Raftery, 1989). Other climatic variables such as temperatures (Parey et al., 2007) or rainfall have been also assessed for trends, through nonparametric tests (Pujol et al., 2007) or linear regression type tests (Bodini and Cossu, 2010). However, some authors (Katz et al., 2002) consider that the evidence of trends based on least squares regression does not necessarily reflect properties of extremes like heavy tails. A Bayesian approach may be suitable (Katz et al., 2002; Coles et al., 2003).

The first part of the work focuses on the possible existence of abrupt changes in the parameters of the GPD. These abrupt changes may be due to changes in the location of the observatories and/or technological advances introduced in the measuring instruments. A new parameterization of the GPD distribution is suggested. Using this reparameterization, the second part of the work examines the possible existence of trends.

The paper is organized as follows: the Poisson-GPD occurrence model and the Bayesian estimation of parameters are described in Sect. 2. This methodology is applied to the Barcelona rainfall series and results are shown in Sect. 3.

2 Methodology

A standard hazard analysis approach, Point-Over-Threshold (POT), has been used to model events occurring in time and their magnitude. Rainfall events are modelled as points in time. For each event, a magnitude X (natural logarithm of daily precipitation) is taken as a random variable. The assumptions are: (a) magnitudes are independent from the point process itself; (b) they are independent from event to

event; (c) all magnitudes have the same distribution F_X; and (d) the occurrence of the events in time follows a Poisson process. Thus, the number of events, $N(u)$, occurring in a given arbitrary time interval, t, is assumed to be Poisson distributed, $N(u) \sim \text{Poisson}(\lambda)$,

$$P[N(u) = n | \lambda(u), t] = \frac{(\lambda(u)t)^n}{n!} e^{-\lambda(u)t}, \ n = 0, 1, 2, \dots . \quad (1)$$

There are difficulties in identifying a model for F_X, but the Generalized Pareto Distribution (GPD) provides a suitable model for excesses of X over a high enough threshold x_0, $Y = X - x_0$ given $X > x_0$, (Pickands, 1975). This GPD distribution is parametrized by a shape parameter, ξ, and a scale parameter β:

$$F_Y(y | \xi, \beta) = \left(1 + \frac{\xi y}{\beta}\right)^{-\frac{1}{\xi}}, \quad 0 < y < y_{\text{sup}} . \quad (2)$$

The parameter ξ is non-dimensional, and β has units that are proportional to the units of the excesses (Y). The value of ξ defines three so-called domains of attraction (DA) (Davison and Smith, 1990; Castillo et al., 2004). The Weibull DA corresponds to GPD's such that $\xi < 0$ and $y_{\text{sup}} = -\beta/\xi$, i.e. the support of Y, and hence of X, is limited with a finite upper tail. The Fréchet DA is characterized by $\xi > 0$ and $y_{\text{sup}} = +\infty$, and the support of Y becomes infinite. Finally, the Gumbel DA corresponds to GPD's such that $\xi = 0$ and $y_{\text{sup}} = +\infty$, the GPD takes the limit form

$$F_Y(y | \xi = 0, \beta) = 1 - \exp\left[-\frac{y}{\beta}\right], \quad 0 < y < +\infty , \quad (3)$$

which is an exponential distribution. Densities of GPD distributions of different domains of attraction show distinct features (Tolosana-Delgado et al., 2010).

The first step in the estimation of the GPD distribution for the excesses, Y, is the selection of an appropriate reference threshold x_0. A graphical technique (Davison and Smith, 1990; Embrechts et al., 1997; Dupuis, 1998) can be applied taking into account that, for GPD distributed excesses and for $x_1 \geq x_0$, the mean excess is linear with respect to x_1, ($\xi < 1$),

$$E[X - x_1 | X > x_1 \geq x_0] = \frac{\beta + \xi x_1}{1 - \xi} . \quad (4)$$

Once a suitable threshold x_0 has been selected, and considering that this work examines the possible existence of trends, the parameters ξ and β of the GPD are considered as a function of time. In order to parsimoniously deal with this climate variation a new parameterization of the GPD distribution is suggested. The classical scale and shape parameters of the GPD (ξ, β) are reformulated as a location parameter μ, linked to the upper limit of the distribution, and a shape parameter ν,

$$\mu = \log\left(\frac{-\beta}{\xi}\right); \quad \nu = \log(-\xi \cdot \beta) . \quad (5)$$

In order to take account for uncertainty of the estimation of these parameters, a Bayesian approach for the joint estimation of μ and ν is chosen. One of the basis of Bayesian estimation of parameters is to assume that they are random variables and, then, they are described by their joint probability distribution (Gelman et al., 1995), before and after the data acquisition. The prior density $f_{\lambda, \mu, \nu}(\lambda, \mu, \nu)$ represents our knowledge about the parameters before the data sample. As independence of the magnitudes from the point process has been assumed, the prior density can be written as:

$$f_{\lambda, \mu, \nu}(\lambda, \mu, \nu) = f_\lambda(\lambda) f_{\mu, \nu}(\mu, \nu) .$$

Using Bayes' theorem, the posterior distribution of the parameters is obtained:

$$f_{\lambda, \mu, \nu}(\lambda, \mu, \nu \mid y_1, \dots, y_n) =$$

$$C \cdot f_\lambda(\lambda) f_{\mu, \nu}(\mu, \nu) \cdot L(\lambda, \mu, \nu \mid y_1, \dots, y_n) ,$$

where $L(\lambda, \mu, \nu | y_1, \dots, y_n)$ is the likelihood function of the excesses and $y_1, \dots, y_n$ is a sample of excesses over threshold x_0. Therefore, the joint posterior density $f_{\lambda, \mu, \nu}(\lambda, \mu, \nu | y_1, \dots, y_n)$ accounts for the uncertainty of the parameters. Not only information about the parameters can be obtained from this distribution, but it is also the basis to obtain the distribution of quantities of interest such as hazard parameters, usually through simulated samples of λ, μ and ν. Monte Carlo methods can be used in the cases where this joint posterior probability distribution is not easily tractable (Robert and Casella, 2000). As attention is set to the existence of changes in the GPD parameters, this work approaches only the characterization of the GPD distribution, leaving apart the estimation of the Poisson occurrence parameter. Taking into account the independence between parameters due to model assumptions, only the marginal joint posterior density $f_{\mu, \nu}(\mu, \nu | y_1, \dots, y_n)$ is considered:

$$f_{\mu, \nu}(\mu, \nu | y_1, \dots, y_n) = C \cdot f_{\mu, \nu}(\mu, \nu) \cdot L(\mu, \nu | y_1, \dots, y_n) ,$$

where $L(\mu, \nu | y_1, \dots, y_n)$ is the likelihood function of the excesses

$$L(\mu, \nu | y_1, \dots, y_n) = \prod_{i=1}^n f_Y(y_i | \mu, \nu) .$$

However, once the estimation of the Poisson and the GPD parameters has been performed, hazard parameters can be computed easily using marked Poisson processes properties (Grandell, 1997).

2.1 Models

A common concern is to know whether magnitudes of hazardous events have changed in the last decades. Long data series are very appreciated in order to properly study these issues. The series of daily rainfall in Barcelona (1854–2006) seems suitable to address these issues, due to its length. As
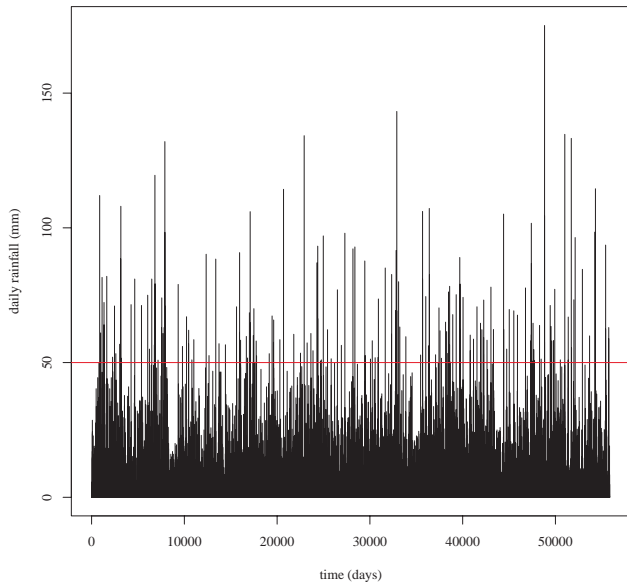
Fig. 1. Barcelona daily rainfall series (1854–2006).

Fig. 2. Graphical estimation of log-daily-rainfall reference threshold: $\exp(3.9) \approx 50\,\mathrm{mm}$.

previously stated, rainfall is taken in a log-scale as a relative scale appears to be natural. For strictly positive data, not only for precipitation, taking log's might be not unimportant. When using a natural scale, some modelling requirements, such as goodness-of-fit or compatibility with physical assumptions, appear to be fulfilled in an easier way than when using the usual raw scale in real space (Egozcue et al., 2006; Pawlowsky-Glahn et al., 2005). On the other hand, excesses over a threshold are modelled by a GPD with a limited maximum value ($y_{\mathrm{sup}} = -\beta/\xi$ finite) in order to be consistent with physical limitations.
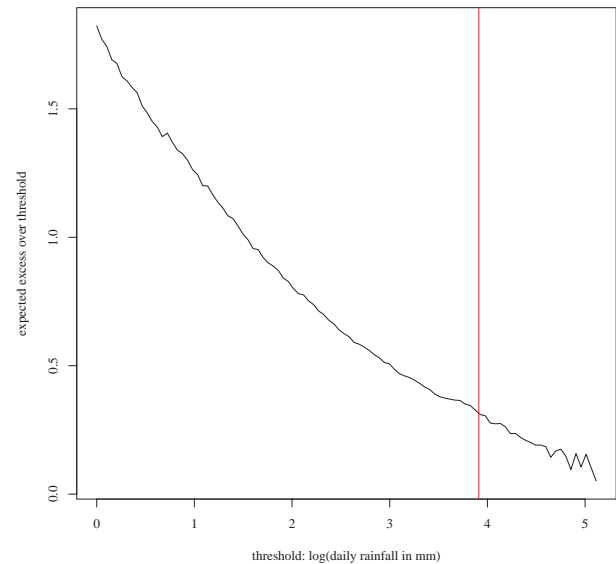
First, the possible existence of abrupt changes in the parameters of the GPD is treated. The parameters of the model are considered as a function of time. Using the suggested new parameterization of the GPD distribution ($\mu = \log(-\beta/\xi)$ and $\nu = \log(-\xi \cdot \beta)$), the location parameter μ is fixed, due to the limited character of the magnitude, and the possible abrupt change of the parameters of the GPD is then translated into a possible abrupt change in the ν parameter. Therefore, the breakpoint t_0 is a point in time such that

$$\begin{cases} \nu = \nu_0, & \{t < t_0\} \cup \{t > t_0 + L\}, \\ \nu = \nu_0 + \Delta\nu, & t_0 \leq t \leq t_0 + L, \end{cases} \qquad (6)$$

where L is a fixed length.

The possible existence of trends in the GPD parameters is also treated. The parameters of the model are also considered as a function of time. Using the suggested new parameterization of the GPD distribution the parsimonious choice is to consider shape as a linear function of time, $\nu(t) = \nu_0 + t\,\Delta\nu$ while keeping location fixed $\mu(t) = \mu_0$. Then, the climate change is assessed by checking the hypothesis $\Delta\nu \neq 0$.
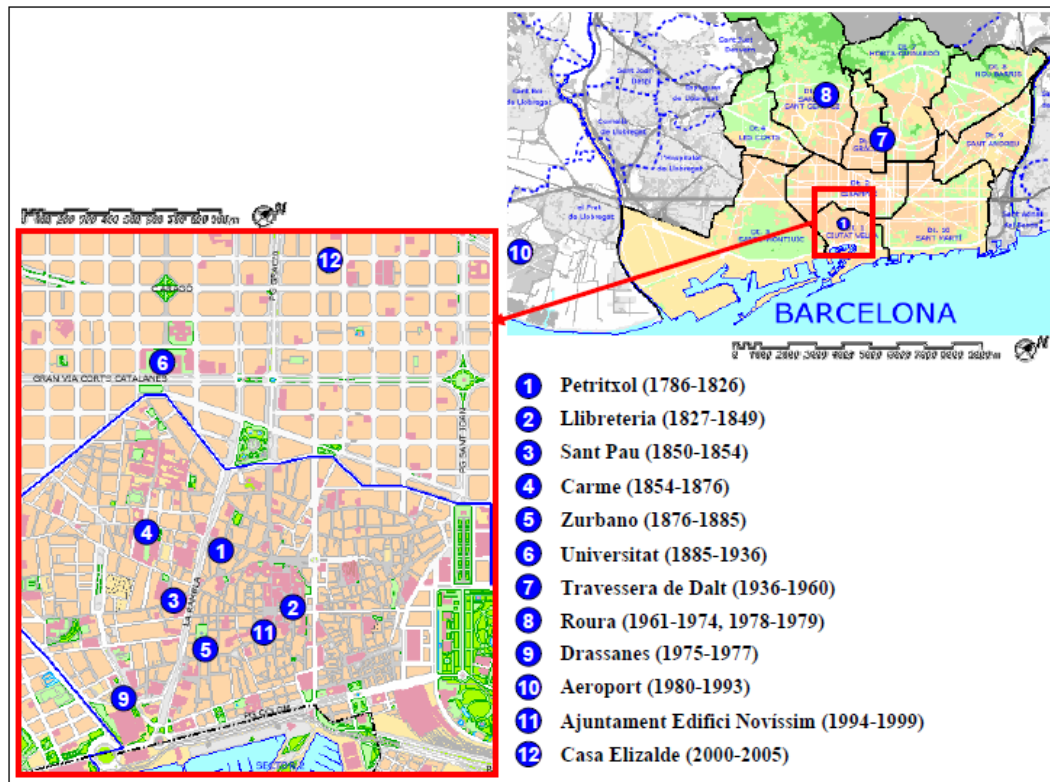
3 Barcelona daily rainfall series

The series of daily rainfall in Barcelona is one of the longer continuous european instrumental daily rainfall series available. This data base contains records of daily precipitation along more than 150 years (1854–2006) (Fig. 1), although a longer data series of monthly precipitation (1786–2006) sharing the same name is also available, (Barrera-Escoda, 2008).

The POT method has been used, and a GPD distribution has been fit to excesses over threshold x_0. In our case, applying the mean excess graphical technique (Fig. 2), it seems that a good fit for a reference threshold is $x_0 = 3.9$, which corresponds to a raw daily rainfall of 50 mm approximately.

During the studied period of time (more than 150 years) the location of rainfall observatories where data were collected has changed as shown in Fig. 3. Most of these sites lied within the Barcelona built-up area, but the existence of parts of the series recorded at the El Prat Airport, ca. 15 km. from Barcelona city center, points to a possible change in the parameters of the selected distribution of daily rainfall. As technological advances have been introduced in the measuring instruments during this period as well, the first part of the work focuses on the possible existence of abrupt changes in the parameters of the GPD (μ, ν) due to these changes of location and/or instruments. Therefore, as previously stated, three parameters need to be estimated: ν_0, t_0 and $\Delta\nu$.

According to the Bayesian paradigm, previous knowledge about the parameters has been included in its prior density. Prior density for ν_0, t_0 and $\Delta\nu$ has been assumed uniform for a wide range of values. As log-daily-rainfall is considered

Fig. 3. Barcelona daily rainfall series (1786–2006). Location of rainfall observatories during this period.

a bounded magnitude, the domain for the ξ parameter has been completely placed in the Weibull DA, and this character has been transmitted to the prior assumptions for the ν and μ parameters. Through the likelihood of the log-daily-rainfall excesses over $x_0 = 3.9$ this prior assumptions have been updated, obtaining the posterior density of the parameters. Once the posterior density is obtained, a simulated sample of 10 000 values $(\nu_0, t_0, \Delta\nu)$ is generated using the Gibbs sampler (Robert and Casella, 2000).

There is no strong evidence that the migration of the rain gauge positions through the city and the airport imply important shape differences (Fig. 4). Only mild "significance" can be attached to the changes from locations 6 to 7 (~ 0.12), the return from 7 to 8 (~ 0.1) and from the airport (10) back to the city (11) (~ 0.07). Other suspicious transitions can be rejected, with Bayesian p-values as large as 0.2.

Next, the possible existence of trends is examined. In fact, as previously presented, the interest is set in the possible slope of the spread parameter ν. As in the first model, a uniform prior density of the parameters for a wide range of values has been considered. Posterior joint density of GPD parameters in the new and classical parameterizations are shown in Fig. 5 and Fig. 6. In order to assess the existence of a trend, the marginal posterior density of the slope in the spread parameter, $\Delta\nu$ has been obtained, Fig. 7. This density shows that no clear slope is significantly different from zero,

and therefore a linear trend for the spread parameter is not significant. Furthermore, if there were a trend, it would have positive slope.

4 Conclusions

A Point-Over-Threshold approach with a new parameterization of the GPD distribution has been used as a model for daily rainfall. The scale and physical features of the phenomenon have been considered, and therefore, daily rainfall has been treated in a log-scale and considered as a bounded magnitude. A long instrumental data series, the daily rainfall Barcelona series, has been considered and a Bayesian approach has been used to assess the existence of abrupt changes in the parameters of the reparameterized GPD distribution and/or trends in these parameters. No strong evidence of a linear trend or step change in the Barcelona rainfall series has been found. Although there are no strong evidences of heterogeneity, attention should be paid to the changes of location with higher frequency of possible breakpoint, i.e. changes from locations 6 to 7, 7 to 8 and 10 to 11 (Fig. 4). Under the parsimonious hypothesis of no change of the upper physical limit of rainfall, a climate change hypothesis is not significant.
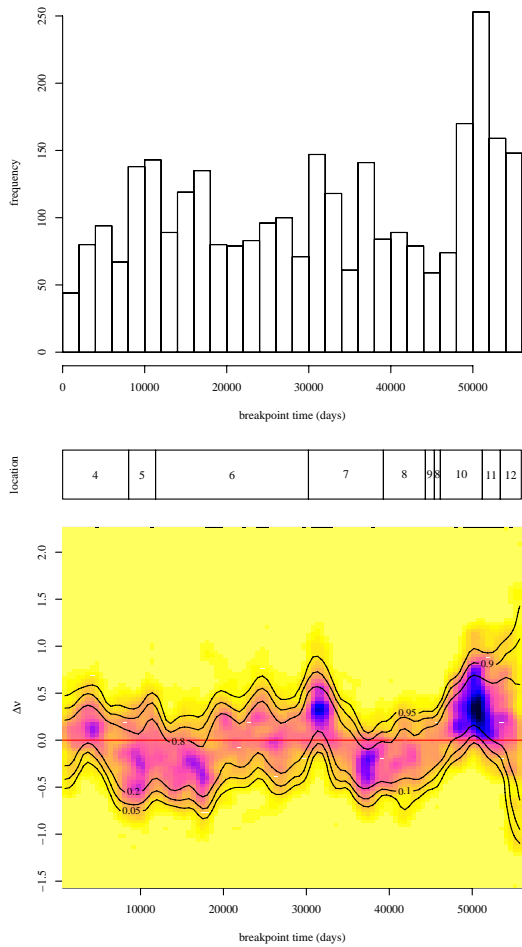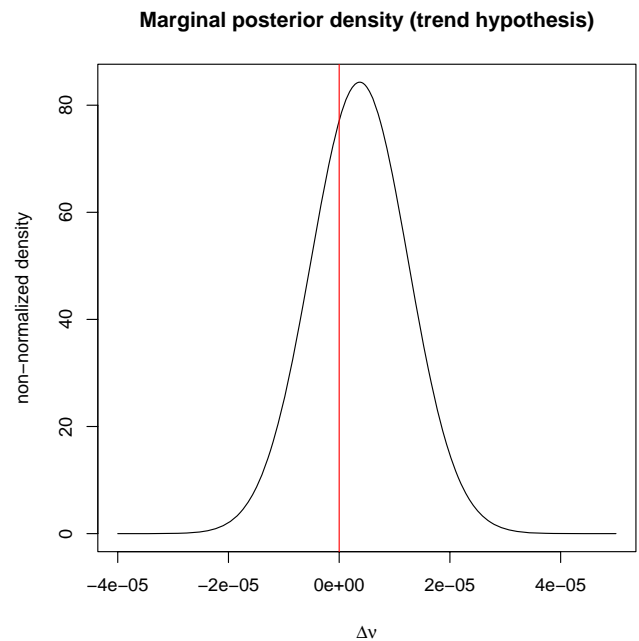
Fig. 4. Upper plot: histogram of posterior simulations of t_0. Middle plot: positions of the known rain gauge location changes. Lower plot: (color levels) kernel density estimate of the joint posterior distribution of $(t_0, \Delta \nu)$; (black lines) quantiles of the distribution of $\Delta \nu$ conditional on t_0.
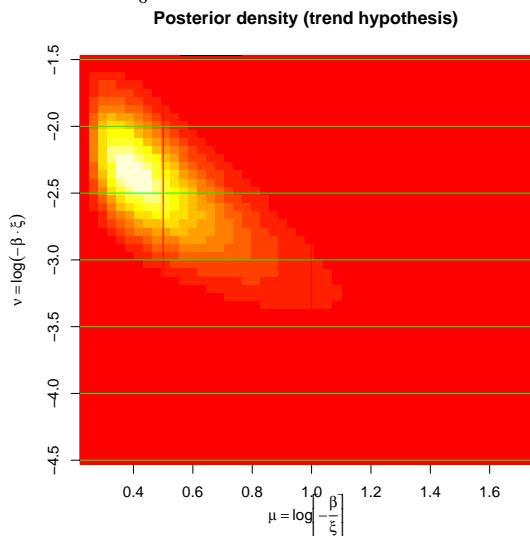
Fig. 5. Pixel map of posterior joint density for GPD parameters: new parameterization. Parameters μ and ν are non-dimensional.
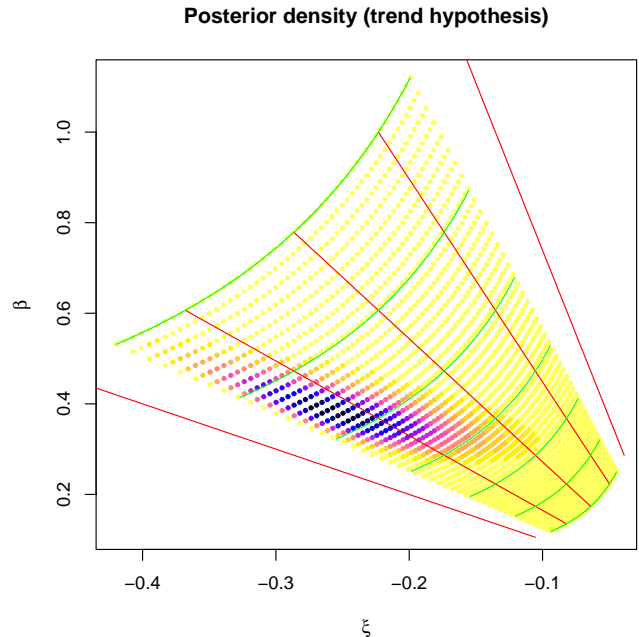
Fig. 6. Pixel map of posterior joint density for GPD parameters: usual parameterization. Parameter ξ is non-dimensional and β has units proportional to the units of the excesses (Y).

Fig. 7. Marginal posterior density for the slope in ν (spread). Parameter ν is non-dimensional.

Edited by: G. Boni
Reviewed by: one anonymous referee

References

Barrera-Escoda, A.: Evolución de los extremos hídricos en Catalunya en los últimos 500 años y su modelización regional. (Evolution of hydric extrems in Catalonia during the last 500 years and its regional modelling), Ph.D. thesis, Universitat de Barcelona, Barcelona, Spain, 2008.

Bodini, A. and Cossu, Q. A.: Vulnerability assessment of Central-East Sardinia (Italy) to extreme rainfall events, Nat. Hazards Earth Syst. Sci., 10, 61–72, doi:10.5194/nhess-10-61-2010, 2010.

Castillo, E., Hadi, A., Balakrishnan, N., and Sarabia, J.: Extreme value and related models with Applications in Engineering and Science, Wiley, London, GB, 384 pp., 2004.

Coles, S., Pericchi, L. R., and Sisson, S.: A fully probabilistic approach to extreme rainfall modelling, J. Hydrol., 273, 35–50, 2003.

Davison, A. and Smith, R.: Models for exceedances over high thresholds (with discussion), J. Roy. Statist. Soc. B, 52, 393–442, 1990.

Dupuis, D.: Exceedances over high thresholds: a guide to threshold selection, Extremes, 1, 251–261, 1998.

Egozcue, J. J., Pawlowsky-Glahn, V., Ortego, M. I., and Tolosana-Delgado, R.: The effect of scale in daily precipitation hazard assessment, Nat. Hazards Earth Syst. Sci., 6, 459–470, doi:10.5194/nhess-6-459-2006, 2006.

Embrechts, P., Klüppelberg, C., and Mikosch, T.: Modelling extremal events for insurance and finance, Springer-Verlag, Berlin, Germany, 663 pp., 1997.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.: Bayesian data analysis, Wiley, New York, NY, USA, 1995.

Grandell, J.: Mixed Poisson processes, Chapman & Hall, London, GB, 268 pp., 1997.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Res., 25, 1287–1304, 2002.

Parey, S., Malek, F., Laurent, C., and Dacunha-Castelle, D.: Trends and climate evolution: Statistical approach for very high temperatures in France, Climatic Change, 81, 331–352, 2007.

Pawlowsky-Glahn, V., Tolosana-Delgado, R., and Egozcue, J. J.: Scale effect in hazard assessment – application to daily rainfall, Adv. Geosci., 2, 117–121, doi:10.5194/adgeo-2-117-2005, 2005.

Pickands, J. I.: Statistical inference using extreme order statistics, Ann. Stat., 3, 119–131, 1975.

Pujol, N., Neppel, L., and Sabatier, R.: Regional tests for trend detection in maximum precipitation series in the French Mediterranian region / Tests régionaux de détection de tendances dans des séries de maxima pluviométriques sour le pourtour Méditerranéen Français, Hydrol. Sci. J., 52, 956–973, 2007.

Raftery, A. E.: Extreme value analysis of enviromental time series: An application to trend detection in ground-level ozone. Comment: Are ozone rates decreasing?, Stat. Sci., 4, 378–381, 1989.

Robert, C. and Casella, G.: Monte Carlo Statistical Methods, Springer, New York, NY, USA, 507 pp., 2000.

Smith, R. L.: Extreme value analysis of enviromental time series: An application to trend detection in ground-level ozone., Stat. Sci., 4, 367–377, 1989.

Tolosana-Delgado, R., Ortego, M. I., Egozcue, J. J., and Sánchez-Arcilla, A.: Climate change in a Point-Over-Threshold model: an example on ocean-wave-storm hazard in NE Spain, Adv. Geosci., submitted, 2010.