Advances in Geosciences, 2, 117–121, 2005
SRef-ID: 1680-7359/adgeo/2005-2-117
European Geosciences Union

Advances in Geosciences

Scale effect in hazard assessment – application to daily rainfall

V. Pawlowsky-Glahn[1], **R. Tolosana-Delgado**[1], **and J. J. Egozcue**[2]

[1]Dept. Informàtica i Matemàtica Aplicada, U. de Girona, Spain
[2]Dept. Matemàtica Aplicada III, U. Politècnica de Catalunya, Spain

Received: 24 October 2004 – Revised: 2 March 2005 – Accepted: 4 March 2005 – Published: 9 May 2005

Abstract. Daily precipitation is recorded as the total amount of water collected by a rain-gauge in 24 h. Events are modelled as a Poisson process and the 24 h precipitation by a Generalized Pareto Distribution (GPD) of excesses. Hazard assessment is complete when estimates of the Poisson rate and the distribution parameters, together with a measure of their uncertainty, are obtained. The shape parameter of the GPD determines the support of the variable: Weibull domain of attraction (DA) corresponds to finite support variables, as should be for natural phenomena. However, Fréchet DA has been reported for daily precipitation, which implies an infinite support and a heavy-tailed distribution. We use the fact that a log-scale is better suited to the type of variable analyzed to overcome this inconsistency, thus showing that using the appropriate natural scale can be extremely important for proper hazard assessment. The approach is illustrated with precipitation data from the Eastern coast of the Iberian Peninsula affected by severe convective precipitation. The estimation is carried out by using Bayesian techniques.

1 Introduction

The goal of hazard assessment is to estimate the probability of occurrence of large events in a given lifetime. Hazardous events due to natural or anthropogenic phenomena (precipitation, earthquakes, wind, eruptions, floods, fires, etc.) are often modelled by marked Poisson processes: events are assumed to occur as a point Poisson process in time, and intensity of events is assumed to be random, independent from the time-occurrence process and from event to event.

This simple model may be useful in situations when one is interested in rare but dangerous events. However, the scarcity of data leads to highly uncertain parameter estimates, a problem which can be overcome using Bayesian estimation to account for uncertainty. A standard model for large intensity events is the Generalized Pareto Distribution (GPD), leading to a global model with four parameters: the rate of the Poisson process; the scale and shape for the GPD; and a reference threshold. The reference threshold is assessed empirically and afterwards validated. This assessment is a key point of the analysis because a trade-off must be made between a high threshold, guaranteeing a better model fit, and the number of available data with intensity over it. The other three parameters are considered jointly distributed and estimated using Bayesian techniques. Prior information is obtained from expert opinions or physical knowledge.

This approach was used by Egozcue and Ramis (2001) to analyze precipitation in Eastern Spain using a database covering 30 years (Romero et al., 1998). Heavy precipitation is a serious weather hazard in the Valencia region, especially in autumn. Every year several rainfall events exceeding 100 mm daily precipitation occur. Strong convective systems are responsible for it, and rainfall tends to discharge over short periods. For example, in Gandía, on 3 November 1987, more than 800 mm were recorded in 24 h. Some of these events produce floods and severe damage to properties, infrastructure and agriculture, like the one that destroyed the Tous dam (Valencia) on 20 October 1982. The main problem in the study performed by Egozcue and Ramis (2001) appeared to be that excesses exhibit a heavy, unbounded upper tail, something contradictory with the naturally bounded character of precipitation. Precipitation in 24 h must be limited due to several physical reasons: water content in the atmosphere is limited, the movement of convective cells is limited, and also the time of precipitation. These facts have been neglected by most authors, e.g. Coles and Tawn (1996) or Egozcue and Ramis (2001), for the sake of model simplicity, because actual physical upper limit is not known.

While analyzing the underlying reasons for heavy tail behavior of precipitation, we realized that 3 mm and 6 mm precipitations are one double of the other, whereas 100 mm and 103 mm are approximately the same. In other words, the natural scale is relative and, thus, it claims for a logarithmic transformation. To study the effect of scale transformation,

Correspondence to: V. Pawlowsky-Glahn
(vera.pawlowsky@udg.es)

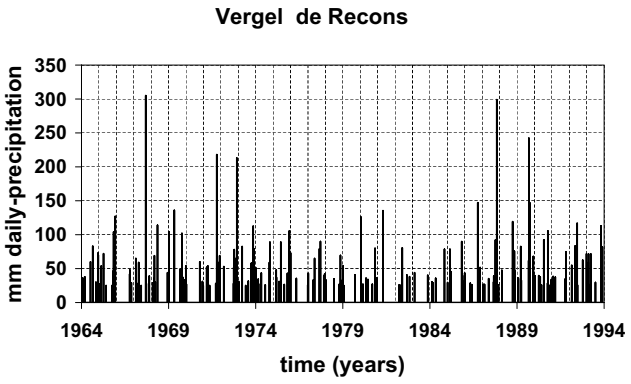

Fig. 1. Observed rainfall at Vergel de Racons (Alicante, Spain), 1964–1993.

we selected a single rain-gauge from the data base by Romero et al. (1998), located at Vergel de Racons (Alicante). Only daily rainfall over 25 mm was extracted and, to ensure independence between events, the maximum daily rainfall in 7 days was considered and consecutive maxima must be separated more than three days. The obtained set of events is represented in Fig. 1.

2 Peak-over-threshold hazard model

The Poisson point-process model used for natural event occurrences is assumed stationary, something contradictory with the seasonal character of precipitation. However, if attention is restricted to high intensity events, the periodic component can be neglected. Rain events exceeding a certain threshold u are modelled as points in time. The event size X (daily precipitation in our case) is usually taken as a random variable, independent from the point process itself and from event to event. Thus, the number of events N_u occurring in a given arbitrary time t is governed by the Poisson probability (n=0, 1, 2, ...)

$$P[N_u = n|\lambda_u, t] = \frac{1}{n!}(\lambda_u t)^n e^{-\lambda_u t}\,, \tag{1}$$

where λ_u stands for the rate of the Poisson process given the threshold u. The event size X is modelled only in the upper tail of the distribution using the peak-over-threshold method (Embrechts et al., 1997). It defines the excess as Y=$\{X-u|X>u\}$ and uses the relationship (y=$x-u>0$)

$$1 - F_Y(y) = P[Y > y|X > u] = \frac{1 - F_X(x)}{1 - F_X(u)}\,, \tag{2}$$

linking the distributions of X and Y. GPD is a simple and parsimonious model for excesses, as it is the limiting distribution for excesses whenever u is high enough (Pickands, 1975). GPD is ($\beta>0$, $y>0$)

$$F_Y(y|\xi, \beta) = 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-1/\xi}\,, \tag{3}$$

with ξ and β the shape and scale parameters. The support of Y is the positive real line $\mathbb{R}^+$ for ξ=0, while it is bounded in the interval $[0, -\beta/\xi]$ for $\xi<0$. For ξ=0, Eq. (3) takes an exponential form. Asymptotically, it approaches upper tails of distributions with exponentially decaying upper tails, a case known as the Gumbel domain of attraction (DA). The DA defined by $\xi<0$ corresponds to bounded variables and is known as the Weibull DA. The Fréchet DA is defined by $\xi>0$ and contains distributions with heavy upper tails. Natural phenomena are physically bounded; thus, their intensity should be in the Weibull DA. However, heavy tail distributions have been reported, in particular for intense rainfall in different climates (Coles and Tawn, 1996; Egozcue and Ramis, 2001). This fact is usually considered to be due to lack of data, but the reason might be simply an inappropriate scale. In our case study, a logarithmic scale reveals a clear Weibull DA for precipitation events.

The first step in the estimation is the selection of an appropriate reference threshold u. A graphical technique (Embrechts et al., 1997) can be applied attending to the fact that, for any $u'>u$, the mean excess is linear with respect to u' ($\xi<1$)

$$\mathrm{E}[X - u'|X > u' \geq u] = \frac{\beta + \xi u'}{1 - \xi}\,. \tag{4}$$

Inspection of the mean excess function was performed for the raw data (original scale in mm rainfall) and the log_{10}-scale. Figures 2 and 3 show two estimates of the mean excess function. The lines with circle markers correspond to the sample average excess over each threshold. Also a preliminar bayesian estimate of the mean excess has been carried out (Egozcue and Tolosana-Delgado, 2002). The median of the posterior mean excess is the thick line, whereas thin lines give some quantiles of the posterior, figuring out the uncertainty of the estimate. From Eq. (4), positive slopes of the mean excess function indicate that the data set corresponds to a Fréchet DA distribution (heavy and unlimited tails), whereas negative slopes suggest data from GPD's of the Weibull DA (limited support). In order to guess a reference threshold, we look for a value such that the mean excess function can be assumed linear from this point on, according to (4). We have selected the threshold in both cases trying to fit a linear segment to the estimated mean excess function taking into account the uncertainty of the estimates. For raw data, the absolute threshold selected was u=45 mm, and for log-transformed data it was u=1.8, that corresponds to 63 mm.

3 Bayesian estimation of parameters

Given the absolute threshold, the remaining parameters are estimated using a Bayesian approach (Egozcue and Ramis, 2001; Egozcue and Tolosana-Delgado, 2002). Bayesian estimation techniques allow combining two sources of information: prior knowledge about the parameters ξ, β and λ_u (coded in a joint prior probability density) and the observed

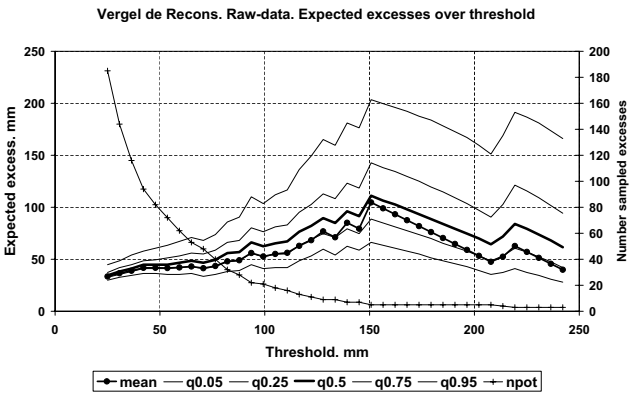

Fig. 2. Estimated expected excess over several thresholds for raw data. Sample average excess, line with circle markers. Median of posterior estimate, thick continuous line; quantiles 0.05, 0.25, 0.75, 0.95, thin lines. Number of excesses used in the estimation, line with plus markers, labelled in the secondary axis.

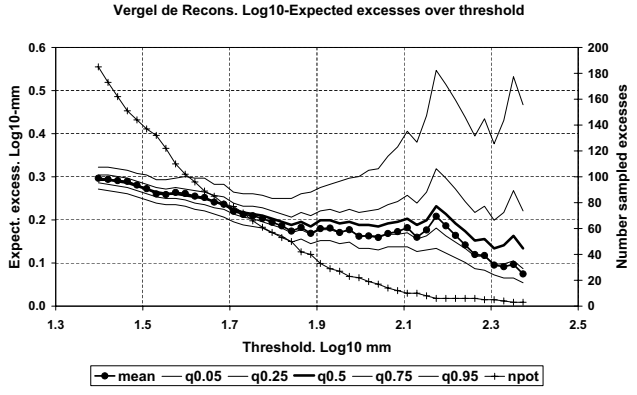

Fig. 3. Estimated expected excess over several thresholds for log_{10}-data. Sample average excess, line with circle markers. Median of posterior estimate, thick continuous line; quantiles 0.05, 0.25, 0.75, 0.95, thin lines. Number of excesses used in the estimation, line with plus markers, labelled in the secondary axis.

values of the excess variable Y (coded as the likelihood of the data). Bayesian estimation consists in multiplying both prior and likelihood to obtain the posterior probability density: the joint density of the three parameters after taking into account the data. In this approach λ_u was assumed a priori independent from ξ, β. As a consequence, the posterior density, conditioned to data, can be factorized into two factors

$$f_{\xi\beta\lambda_u}(\xi, \beta, \xi) = f_{\xi\beta}(\xi, \beta)\, f_{\lambda_u}(\xi)\,. \tag{5}$$

Details of prior assessment were developed in Egozcue and Ramis (2001) and Egozcue and Tolosana-Delgado (2002). Prior information is mainly used to give bounds of the admissible domain for the GPD-parameters ξ and β. We now restrict our attention to the marginal joint posterior density $f_{\xi\beta}(\xi, \beta)$. Figures 4 and 5 show contours of this density when analyzing raw data and log_{10} data respectively. Each point in the plane (ξ, β) represents a GPD, and the value of

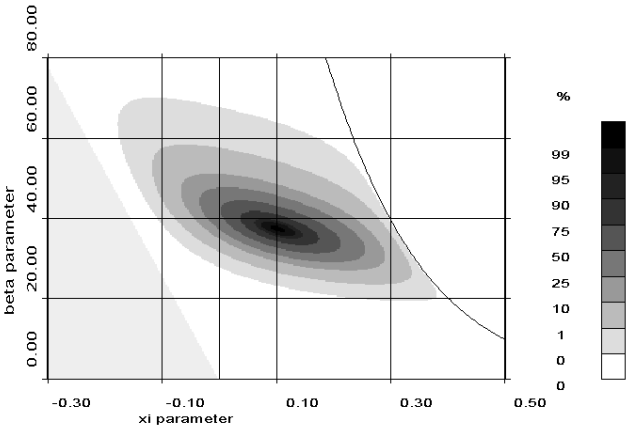

Fig. 4. Joint posterior density for ξ and β for raw data.

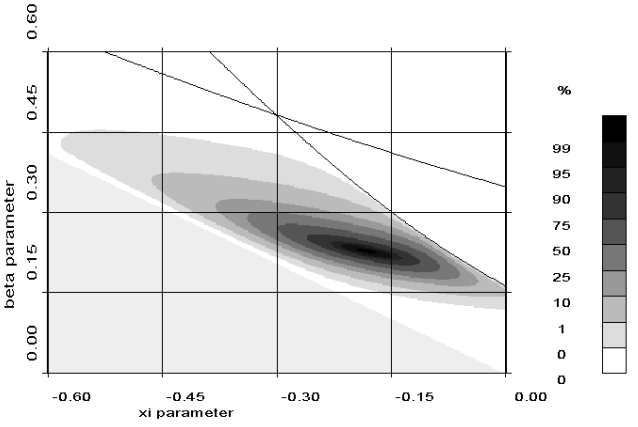

Fig. 5. Joint posterior density for ξ and β for log_{10}-data.

$f_{\xi\beta}(\xi, \beta)$ the relative likelihood of the parameters. Fig. 4, for raw data, shows the most likely points inside the Fréchet DA ($\xi > 0$), thus suggesting this DA for the data. A different conclusion is obtained for log_{10}-scale data (Fig. 5) where the most likely DA is clearly the Weibull DA ($\xi < 0$). There is a strong contrast with the theoretical situation: a log-transformation of a random variable does not change its DA. In fact, an infinite support remains infinite despite of the transformation, and the Fréchet-Gumbel DA is transformed into itself. Our data have been shifted from Fréchet to Weibull DA.

To validate the goodness-of-fit of the different possible parameter combinations for GPD, a Kolmogorov-Smirnov test was performed. Results are shown in Figs. 6 and 7.

These goodness of fit tests reveal that both raw-data and log_{10}-data fit quite well the GPD for GPD-parameters that are likely after the data sample. From the point of view of fitting to the GPD, there is no reason to prefer one of the two data scales.

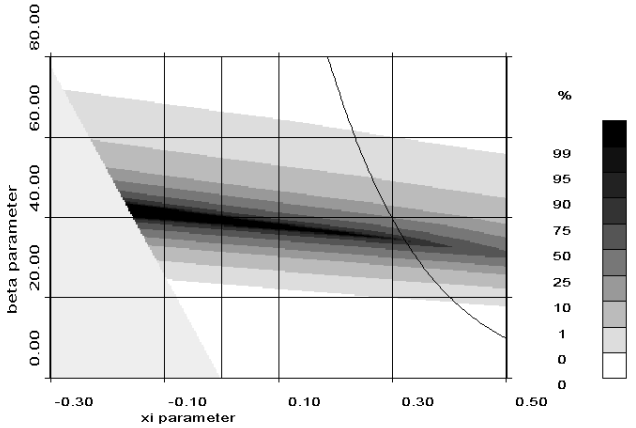

Fig. 6. Kolmogorov-Smirnov goodness-of-fit test p-value for raw data.

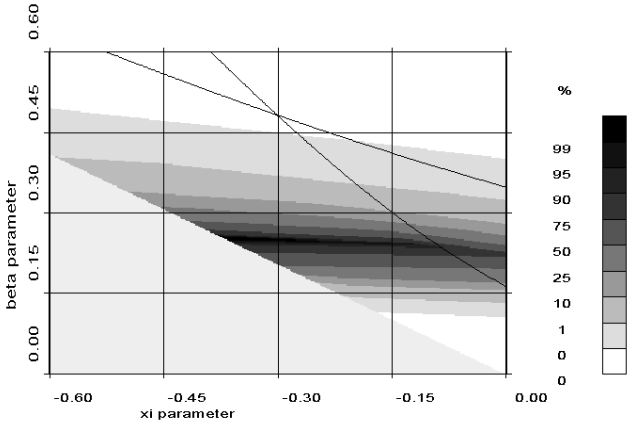

Fig. 7. Kolmogorov-Smirnov goodness-of-fit test p-value for log_{10}-data.

4 Rain hazard estimation

Simulated samples of (λ_u, ξ, β) can be obtained according to their joint posterior distribution. These samples can be used to approximate the distribution of any desired hazard parameter. Here, attention is focused on the determination of the return period for each precipitation level. Figure 8 shows log_{10}-return periods obtained for different daily precipitations, together with a 90%-predictive interval for both raw-data and log_{10}-data. Note the high uncertainty: the 90%-predictive interval for the return period of a rainfall of 250 mm is $[10^1; 10^{1.8}] \approx [10; 65]$ years, and the median is $10^{1.25} \approx 18$ years. These results are similar independently of the scale used. However, when attention is paid to higher precipitation levels, there are some important differences. The most important one is the fact that over 400 mm daily rainfall the return period may be infinite within the 90%-predictive interval, which would mean that such an event is practically impossible. This is only seen when working in a log scale. Also, for high thresholds, say 300 mm, results obtained from raw data are rather conservative with respect to those obtained with log-data. Return periods, and their predictive curves, are assigned to higher thresholds of precipitation for raw scale results than for log_{10}-transformed data. This is a typical behavior of GPD when its parameters are in the Fréchet DA.

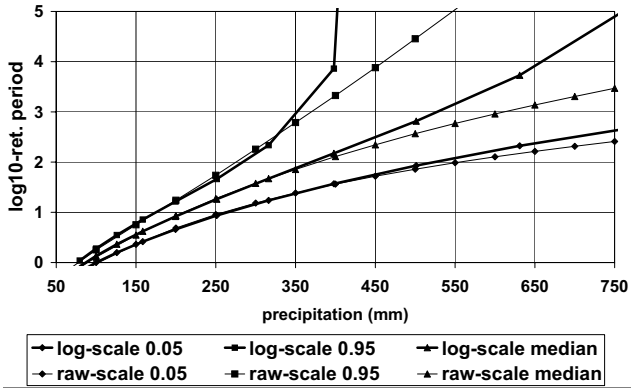

Fig. 8. Estimated quantiles of the posterior for the return period of rainfall

5 Conclusions

Estimation of hazard parameters is highly uncertain mainly due to lack of data. Using a Bayesian approach, this uncertainty can be monitored as is shown in the study of a 30-year rainfall series obtained from a rain-gauge at Vergel de Racons (Alicante). Events were modelled as a Poisson process and daily precipitation as Generalized Pareto distributed. Data suggests using a relative scale for rain intensity, calling for a log-transformation. Heavy tails for precipitation have been repeatedly reported, but it stands in contradiction with the plausible existence of a physical upper limit of the precipitation. Precipitation in log-scale shows a clear Weibull domain of attraction behavior as corresponds to an upper limited phenomenon.

Acknowledgements. The authors thank O. Serrano (Puertos del Estado, Spain) for his encouragement and advice. Data come from the Instituto Nacional de Meteorología of Spain. BGPE program and this study were supported under agreement of the Universidad Politécnica de Catalunya and Puertos del Estado (Spain). This research has also received financial support from the *Dirección General de Investigación* of the Spanish Ministry for Science and Technology through the project BFM2003-05640/MATE.

Edited by: L. Ferraris
Reviewed by: anonymous referees

References

Coles, S. and Tawn, J.: A Bayesian analysis of extreme rainfall data, Applied Statistics, 45, 463–478, 1996.

Egozcue, J. and Ramis, C.: Bayesian Hazard Analysis of Heavy Precipitation in Eastern Spain, International Journal of Climatology, 21, 1263–1279, 2001.

Egozcue, J. and Tolosana-Delgado, R.: Program BGPE: Bayesian Generalized Pareto Estimation, chap. CD-ROM, ISBN 84-69999125, Barcelona, Spain, 2002.

Embrechts, P. C. K. and Mikosh, T.: Modeling Extremal Events, Springer Verlag, Berlin, Germany, 1997.

Pickands, J.: Statistical inference using extreme order statistics, Annals of Statistics, 3, 119–131, 1975.

Romero, R., Guijarro, J., Ramis, C., and Alonso, S.: A 30-years (1964–93) daily rain-fall data base for the Spanish Mediterranean regions: First exploratory study, International Journal of Climatology, 18, 541–560, 1998.