

Climate change in a Point-Over-Threshold model: an example on ocean-wave-storm hazard in NE Spain

R. Tolosana-Delgado¹, M. I. Ortego², J. J. Egozcue², and A. Sánchez-Arcilla¹

¹Universitat Politècnica de Catalunya, Laboratori d'Enginyeria Marítima, Barcelona, Spain

²Universitat Politècnica de Catalunya, Departament de Matemàtica Aplicada III, Barcelona, Spain

Received: 15 February 2010 – Revised: 31 Mai 2010 – Accepted: 18 August 2010 – Published: 27 September 2010

Abstract. A reparametrization of the Generalized Pareto Distribution is here proposed. It is suitable to parsimoniously check trend assumptions within a Point-Over-Threshold model of hazardous events. This is based on considerations about the scale of both the excesses of the event magnitudes and the distribution parameters. The usefulness of this approach is illustrated with a data set from two buoys, where hypotheses about the homogeneity of climate conditions and lack of trends are assessed.

1 Introduction

Climatic change is a problem of general concern. When dealing with hazardous events such as wind-storms, heavy rainfall or wave storms this concern becomes even more serious. Climate change might mean an increase of human and material losses, and therefore efforts to detect it from limited data sets should be taken.

In this contribution, a hazard assessment of storm events in the northern Mediterranean Spanish coast is carried out, following a standard model for extremes such as heavy rainfall or wave storms. An event is defined as the period during which a certain magnitude of the phenomenon (significant wave height in this case) exceeds a given reference threshold. For this reason, this model is typically called Point-over-Threshold (POT) model (Embrechts et al., 1997): time-occurrence of these events is assumed to be Poisson distributed, and the magnitude exceeding the threshold for each event is modelled as a random variable with a Generalized Pareto Distribution (GPD). Independence is assumed, both between this magnitude and occurrence in time, and from event to event. For this contribution, we focus on assess-

ing the presence of a change on the magnitude parameters: the independence assumed ensures us that the occurrence and magnitude estimation can be done separately.

Scarcity of data arises as an additional difficulty, as hazardous events are usually rare. Estimation of hazard parameters such as return periods may imply a great amount of uncertainty. Bayesian methods (e.g. Gelman et al., 1995) have been used successfully to deal with this unavoidable uncertainty of the results, and therefore a Bayesian estimation of GPD models (Egozcue and Tolosana-Delgado, 2002) seems appropriate.

The selection of proper scales for the description of phenomena also arises as an important issue. A handful of phenomena are better described by a relative scale (e.g. positive data where the null value is unattainable) and are thus suitably treated in a logarithmic scale: logarithmic scales has been used successfully for daily rainfall data and wave-height (Egozcue and Ramis, 2001; Pawlowsky-Glahn et al., 2005; Egozcue et al., 2005; Sánchez-Arcilla et al., 2008).

2 The Generalized Pareto Distribution

2.1 Classical parametrization

The Generalised Pareto Distribution (GPD) models excesses over a threshold (Pickands, 1975). If X is the magnitude of an event and x_0 a value of the support of X , the excess over the threshold x_0 is $Y = X - x_0$, conditioned to $X > x_0$. Therefore, the support of Y is either an interval $[0, y_{\text{sup}}]$ or, the positive real line. In our case, X will be the natural logarithm of significant wave height measurements from buoys, and the threshold $x_0 = 5.2$, as explained in the application section.



Correspondence to:
R. Tolosana-Delgado
(raimon.tolosana@upc.edu)

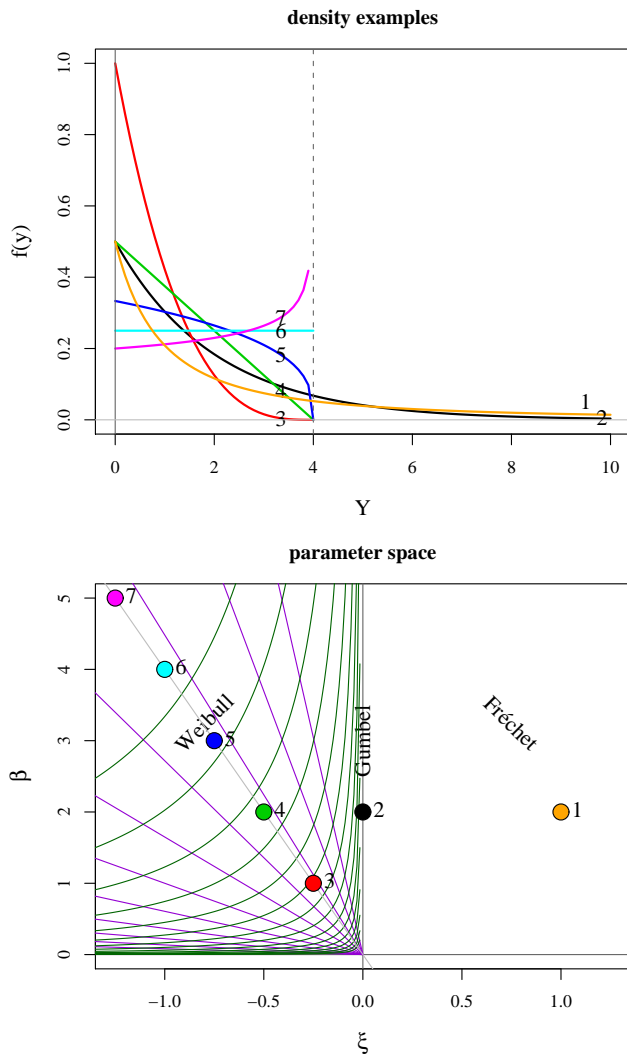


Fig. 1. Examples of GPD densities (upper diagram) covering all domains of attraction, and their representation in the parameter space (lower diagram), numbered correspondingly. This lower diagram shows the classical parametrization, the domains of attraction, and the proposed reparametrization: the rays are iso- μ lines (increasing μ values clockwise), and the hyperbolas are iso- ν lines (increasing ν values upwards; thus Gumbel domain corresponds to $\nu \rightarrow -\infty$).

The GPD cumulative function function is

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

with exponential limit form

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

for $\xi = 0$.

The associated probability density functions for the first case is

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

The scale parameter of the distribution is β , a positive value. The shape parameter, ξ , is real-valued, and it defines three different sub-families of distributions. GPD distributions with $\xi < 0$ have limited support, with expectation and upper bound

$$y_{\text{sup}} = -\frac{\beta}{\xi}, \tag{3}$$

$$E[y] = \frac{\beta}{1-\xi}. \tag{4}$$

These distributions belong to the Weibull domain of attraction. For values $\xi > 0$, $y_{\text{sup}} = +\infty$, distributions belong to the Fréchet domain of attraction, and for $\xi = 0$ (exponential case), distributions have an infinite support and belong to the Gumbel domain of attraction. Figure 1 displays several representations of GPD, as densities (upper diagram) and in the parameter space (lower diagram).

2.2 A new parametrization

One can reasonably assume that, in a climate change scenario, the description model of the variable of interest should not change, but the model parameters could reflect the change: thus, modeling excesses of magnitudes over a threshold by a Generalised Pareto Distribution, using a relative scale, admitting a physical limitation features, ... are elements which will be considered constant. On the other hand, if climate change has occurred (is occurring), there should be a change of the parameters of the GPD distribution (ξ and β), maintaining the physical sense of the described phenomena. In particular, if limited phenomena are described, a GPD-Weibull domain of attraction should be chosen as a priori statement, in order to include this limited feature to the model (Egozcue et al., 2006). This may be easily controlled by a new parameterisation of the GPD distribution:

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

where μ is a new location parameter, informing about the upper bound of the distribution (Eq. 3), and ν is a shape parameter. The classical parameters can be retrieved with

$$\beta = \exp\left(\frac{\nu + \mu}{2}\right); \quad -\xi = \exp\left(\frac{\nu - \mu}{2}\right).$$

The lower diagram of Fig. 1 displays the parameter space of the GPD with the two families of parameters: the classical parameters are represented as a cartesian coordinate system, whereas the proposed parameters form a hyperbolic coordinate system.

3 Bayesian estimation

In a Bayesian estimation process (e.g. Gelman et al., 1995; Egozcue and Tolosana-Delgado, 2002), the observable variable is assumed to follow a parametric model of unknown parameters. In the case presented, the event magnitudes above the threshold follow a GPD with the proposed reparametrization, $Y \sim GPD(\mu, \nu)$. These parameters are given a prior probability distribution $\pi^0(\mu, \nu)$, encoding the knowledge available before looking at the data. In practical computer applications, this is typically a uniform distribution on a discrete grid spanning the range of a priori credible values of the parameters. Then, the data set of excesses $\{y_i, i = 1, \dots, M\}$ comes into the playground: a posterior distribution for the parameters is derived by perturbing the prior distribution by the data likelihood (Eq. 2) according to the parametric model,

$$\pi(\mu, \nu) \propto \pi^0(\mu, \nu) \times \prod_{i=1}^M f_Y(y_i | \mu, \nu)$$

Finally, estimation of the parameters is derived from $\pi(\mu, \nu)$, either as the most likely value (maximum posterior estimation), as the expected value or as any other desired statistic. These are computed directly from the estimated grid posterior probabilities.

4 Assessing the climate change hypothesis at local scale

Several models about parameter changes can be assessed within this framework: abrupt change in a point of time, change as a function of time (linear, logistic or other), etc. For hazardous phenomena with a physical upper limit, the parsimonious choice is to consider a linear change on ν with time, whilst μ remains constant,

$$\mu(t) = \mu_0 + \Delta\mu \cdot t, \quad \nu(t) = \nu_0 + \Delta\nu \cdot t,$$

Then, the climate change hypothesis can be checked by assessing the change on ν :

$$H_0 : \{\Delta\mu = 0, \Delta\nu = 0\}, \quad H_1 : \{\Delta\mu = 0, \Delta\nu \neq 0\}.$$

5 Application

These issues are illustrated using a set of 18 years of significant wave-height data (Sánchez-Arcilla et al., 2008) in log-scale, simultaneously at two stations, the buoys of Roses and Tortosa. Figure 2 shows their location along the Catalan Coast and the sample of events and intensities. The same figure also shows the diagram of expected excess over a threshold, used for identifying x_0 , the threshold of analysis. Note that there are some occurrence gaps in the Roses series (before 1994 and between 1997 and 2001 approximately), but this does not affect computations regarding event magnitude.

If the possible trends were due to a global climate change, one should expect them to be consistently reflected at several

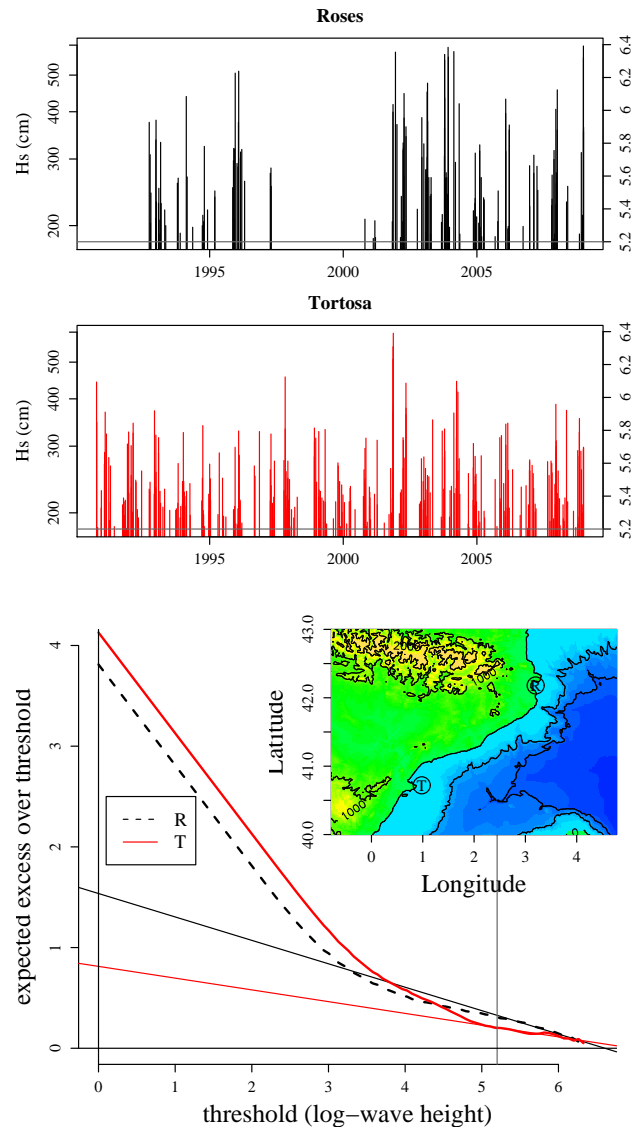


Fig. 2. Significant wave height data series (upper plots), location of the buoys (middle, right plot) and diagram of expected excesses as a function of the threshold (lower plot). This is used to choose the threshold, as (under the hypothesis that excesses are GPD distributed) the function should be a line above it. Dashed/black lines denote Roses, and solid/red ones Tortosa.

nearby, homogeneous locations. For this reason, we analyse simultaneously two locations, selected because they are both prone to the same kind of storms, mostly N-NW or E dominated (*Mestral* and *Llevant* regimes, respectively). We assume that the parameters might have a different value at both stations, but that they should evolve consistently,

$$\begin{aligned} \mu^R(t) &= \mu_0, & \mu^T(t) &= \mu_0 + b_\mu, \\ \nu^R(t) &= \nu_0 + a_\nu \cdot t, & \nu^T(t) &= \nu_0 + a_\nu \cdot t + b_\nu, \end{aligned}$$

Table 1. Prior and posterior characterization. The prior distribution is uniform on a 5-dimensional grid, with 21 equally-spaced nodes along each axis between the minimum and the maximum values reported. The time increment a_ν is measured in ν units per year. The posterior distribution is computed on the same support, and has its maximum value at the vector indicated as “maxpost”.

	μ_0	ν_0	a_ν	b_μ	b_ν
minimum	0.15	-9.00	-0.079	0.15	-5.00
maximum	0.50	+9.00	+0.079	0.70	+5.00
maxpost	0.225	-1.667	0.0197	0.306	-1.250

A Bayesian joint estimation of all these parameters (initial values μ_0, ν_0 , common time trend in shape only a_ν , and the local differences b_μ, b_ν between Tortosa and Roses) is carried out using simple R routines, with flat prior distributions within grids defined in Table 1. The maximum posterior estimates (most likely value of the vector of parameters according to the joint posterior distribution) are included in the same table. The marginal posterior distributions of the parameters are shown in Fig. 3, together with a visual assessment of the hypotheses of zero parameter according to the position of the posterior with respect to the zero value. For instance, regarding the time trend, we can conclude that the hypothesis of no trend ($a_\nu = 0$) is strongly likely, thus there is no evidence in favor of a change in the shape of the GPD (i.e. in the relative likelihood of strong vs. medium storms). However, if there is a change in time, it is more probably a positive one, of the order of +0.02 units ν /year.

As a secondary result, the method may also provide estimates of hazard-related parameters, like return periods, probabilities of exceedance and upper bounds of excesses (as we are fitting the data within the Weibull domain). One must nevertheless bear in mind that *these parameters are all extremely uncertain*, especially for data series so short as those used here. Figure 4 shows an example of this uncertainty, by depicting the data set together with kernel density estimates of the excess upper bound distribution. Note how in the case of Tortosa the spread of the upper bound may be comparable to the spread of the data itself. This happens because Tortosa measurements have more negative ν values, and thus fall nearer to the Gumbel domain (exponential distribution) than Roses measurements. Posed in other words, in Roses the observed excesses bear evidence of an upper boundary quite near to the data actually observed. On the contrary, Tortosa buoy measurements point to a larger upper bound, with more uncertainty, i.e. the fitted GPD is more similar to a distribution with no upper limit, like the exponential form for $\xi = 0$ of Eq. (1). This is in agreement with the fact that Roses buoy is placed on a quite sheltered bay, whereas Tortosa buoy is open to the Mestral and Llevant winds: thus one should expect *potentially* larger measurements in Tortosa than in Roses.

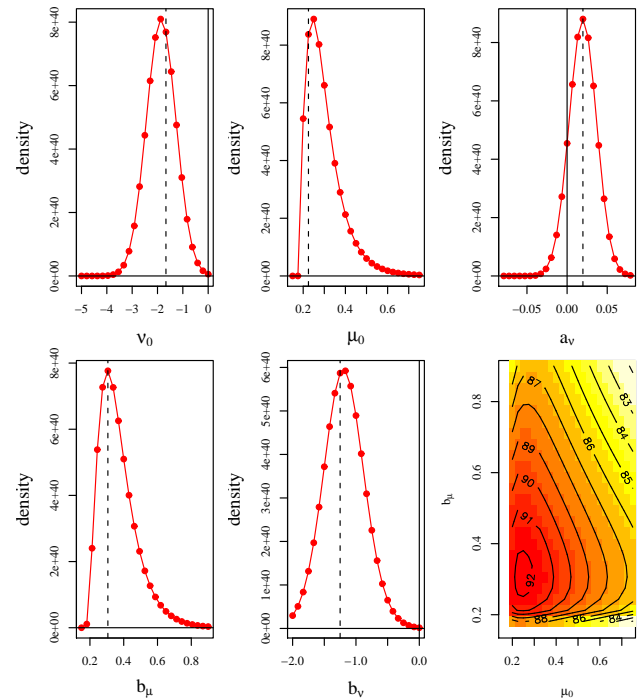


Fig. 3. Marginal posterior distributions for the model parameters, compared with the joint maximum posterior estimate (Table 1, dashed line) and the hypothesis of zero parameter (solid line). The posterior density map shows contour curves of $\log \pi(\mu_0, b_\mu)$. This is used to obtain estimates of μ^T . Note the white stripes in the lower and left margins of this figure: they correspond to zero posterior probability.

6 Conclusions

Assessing the *scale* of available data as well as model parameters allows to *parsimoniously* check models of evolution of these parameters with time. For point-over-threshold (POT) models of significant wave height, this general principle suggests to treat log-transformed data, and fit them a reparametrized Generalized Pareto Distribution *restricted to the Weibull domain*: the new parameters are the upper bound of the distribution as location parameter, and a shape parameter. This parameterization has two advantages: densities always have a bounded domain (as expected for *any* physical process), and checks on the evolution of the distribution shape can be done independent of the upper bound.

This is applied to an 18-year long data set of significant wave height from two different buoys, in the same region but sufficiently far away to consider them roughly independent. If a climate change is present, this should be reflected as a consistent trend in the shape parameter of both series. Results show *no significant trend* in extreme storm magnitudes during the last 18 years. Thus, there is no evidence in this (rather short) data set that climate change is recently modifying distributional properties of the magnitude of extreme

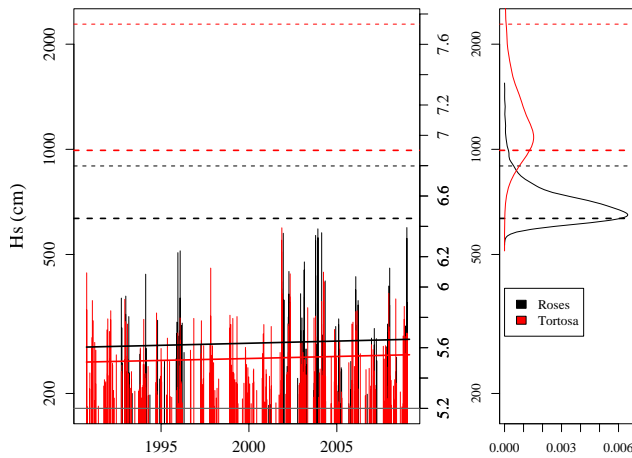


Fig. 4. Data set of events, compared with several results that share the same scale. The left panel shows the data (black: Roses, red: Tortosa), together with the time evolution of the expectation of the fitted GPD (Eq. 4, thick solid line), as derived from the maximum posterior parameter estimates (Table 1). The most likely posterior estimates (from the same Table, thick dashed line) and 95% confidence interval upper boundary (dashed line) for the upper bound of the excesses (Eq. 3) are also displayed. The marginal posterior density of these excess upper bound for Roses and Tortosa are displayed separately in the right panel. Note the higher uncertainty in Tortosa than in Roses.

storms in the Catalan coast. This does not deny climate change as a whole, given the shortness of the series and the inherent uncertainties of the GPD model.

A comparison of both stations suggest that the measurements in Tortosa are (relatively) more compatible with a Gumbel domain (i.e. an exponential law for the excesses of log-significant waveheight) than those in Roses: though both stations fall within the Weibull domain (bounded distributions), measurements from Tortosa show significantly larger, more uncertain estimates of the upper bound of the distribution. This is tentatively related to the sheltered position of the Roses buoy. The uncertainty on this upper bound estimates is extremely large. This would also happen with other hazard-related parameters like return periods and exceedance probabilities. The set of tools used (GPD with bounded domain for log-waveheight point-over-threshold exceedances within a Bayesian approach) has the additional advantage to fairly portray this uncertainty.

Acknowledgements. This research has been supported by the Spanish Ministry of Education and Science under two projects: “Ingenio Mathematica (i-MATH)” Ref. No. CSD2006-00032 and “CODA-RSS” Ref. MTM2009-13272; and by the *Agència de Gestió d’Ajuts Universitaris i de Recerca* of the *Generalitat de Catalunya* under the project Ref: 2009SGR424. The first author acknowledges also funding within the program “Juan de la Cierva” of the Spanish Ministry of Education and Science (ref. “JCI-2008-1835”).

Edited by: J. Salat

Reviewed by: one anonymous referee

References

- Egozcue, J. and Ramis, C.: Bayesian hazard analysis of heavy precipitation in eastern Spain, *Int. J. Climatol.*, 21, 1263–1279, 2001.
- Egozcue, J. and Tolosana-Delgado, R.: Program BGPE: Bayesian Generalized Pareto Estimation, Barcelona, Spain, CD-ROM, ISBN 84-69999125, 2002.
- Egozcue, J. J., Pawlowsky-Glahn, V., and Ortego, M. I.: Wave-height hazard analysis in eastern coast of Spain. Bayesian approach using generalized Pareto distribution. *Adv. Geosci.* 2, 25–30, 2005.
- 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*, Berlin, Germany, Springer-Verlag, 663 pp., 1997.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D.: *Bayesian data analysis*, New York, NY, USA, Wiley, 526 pp., 1995.
- Pawlowsky-Glahn, V., Tolosana-Delgado, R., and Egozcue, J. J.: Scale effect in hazard assessment application to daily rainfall, *Adv. Geosci.*, 2, 117–121, 2005.
- Pickands, J. I. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, 3, 119–131.
- Sánchez-Arcilla, A., Gómez-Aguar, J., Egozcue, J. J., Ortego, M. I., Galiatsatou, P., and Prinos, P.: Extremes from scarce data. The role of Bayesian and scaling techniques in reducing uncertainty. *J. Hydr. Res.*, 46(extra 2), 224–234, 2008.
- Sánchez-Arcilla, A., González-Marco, D., and Bolaños, R.: A review of wave climate and prediction along the Spanish Mediterranean coast, *Nat. Hazards Earth Syst. Sci.*, 8, 1217–1228, 2008.