

**MAPPING THE HAZARD OF EXTREME RAINFALL
BY PEAKS OVER THRESHOLD EXTREME VALUE
ANALYSIS AND SPATIAL REGRESSION
TECHNIQUES**

Santiago Beguería¹

*UCEL - Utrecht Centre for Environment and Landscape Dynamics, Department of
Physical Geography - Utrecht University, PO Box 80115,
3584 CS Utrecht, The Netherlands*

*Instituto Pirenaico de Ecología – CSIC, Campus de Aula Dei, Zaragoza. PO Box 202,
50080 Zaragoza, Spain*

Sergio M. Vicente-Serrano

*Centre d'Etudes Spatiales de la Biosphère (CESBIO), 18 avenue. Edouard Belin,
bpi 2801, 31401 Toulouse cedex 9. France*

*Instituto Pirenaico de Ecología – CSIC, Campus de Aula Dei, Zaragoza. PO Box 202,
50080 Zaragoza, Spain*

¹ Corresponding author. Email: s.begueria@geog.uu.nl. Tel: +31-(0)30-2534014.
Fax: +31-(0)30-253.1145

Abstract: The occurrence of rainfalls of high magnitude constitutes a primary natural hazard in many parts of the world, and the elaboration of maps showing the hazard of extreme rainfalls has great theoretical and practical interest. In this work a procedure based on extreme value analysis and spatial interpolation techniques is described. The result is a probability model in which the distribution parameters vary smoothly in space. This methodology is applied to the middle Ebro valley (Spain), a climatically complex area with great contrasts due to relief and the exposure to different air masses. The database consists on 43 daily precipitation series from 1950 to 2000. Since rainfall tends to occur highly clustered in time in the area, a de-clustering process was applied to the data, and the series of daily cluster maxima were used hereinafter. The mean-excess plot and error minimizing were used to find an optimum threshold value to retain the highest records (peaks-over-threshold approach), and a Poisson-Generalised Pareto model were fitted to the resulting series. The at-site parameter estimates (location, scale and shape) were regressed upon a set of location and relief variables, enabling the construction of a distributed probability model. The advantages of this method to obtain maps of extreme precipitation hazard are discussed in depth.

Keywords: Extreme precipitation, Hazard mapping, Regionalisation, Spatial regression, Generalized Pareto distribution, Ebro valley, Spain.

1. Introduction

The analysis of extreme events is currently one of the lead research topics in climatology due to the potentially dangerous character of the phenomena (Obasi 1994; Bruce 1994). An increase on the interest on extreme events has been detected in the last years within the climatic change paradigm. One of the most accepted hypothesis is that of a future increase of extreme events due to the increase of climatic variability (Katz and Brown 1992; Groisman et al. 1999). In different regions a clear growth of climate extremes has been reported (Karl et al. 1995; Groisman et al. 2004), which have had great negative impacts on society and environment (Kunkel et al. 1999; Easterling et al. 2000).

Extreme precipitation events constitute a primary natural hazard since they are in the origin of degradation processes like severe erosion, landslide triggering or flash floods, which can have regional devastating power and pose a severe hazard to lives and property. Mapping the hazard of extreme precipitation allows assessing the spatial distribution of this climatic feature, but also permits the estimation of hazard at locations where no climatic record exists. Climatic hazard maps, in general, can also be useful as a part of decision support systems, especially in the fields of regional planning and environmental management. The main objective of this paper is to describe a method to obtain extreme rainfall hazard maps.

The problem of mapping the spatial distribution of extreme precipitation involves the need to translate the point information registered at different climatic stations in a region to a spatially continuous variable. Also, different variables have been used to describe the extreme precipitation. Thus, Prudhomme (1999) and Prudhomme and Reed (1999) used the median of the

annual maximum daily precipitation. Lorente and Beguería (2002) used the median of the annual maximum precipitation accumulated in 1, 3, 5 and 7 days. It can be objected that a median value is not the most adequate variable to express extreme events. Attempts using absolute maxima, however, have had very little success (García Ruiz et al. 2000).

Instead of using a single statistic like the median of the annual maxima to describe the occurrence of extreme rainfalls in a region, the extreme value theory provides a much more complete analysis of the statistical distribution of extreme rainfall events, allowing for the construction of magnitude-frequency curves. Derived statistics like quantile estimates (average magnitude of an event of given return period) have been widely used to express the degree of hazard related to extreme precipitation at a given location. Combined with spatial interpolation techniques, they have also been used to assess the spatial distribution of the hazard of extreme rainfalls. For example, Gajic-Capka (1991) and Lana et al. (1995) mapped quantile estimations obtained by fitting a Gumbel model to series of annual maxima by means of local interpolation methods. Beguería and Lorente (2000) used ordinary regression against relief parameters to provide estimations of the 100-year daily maximum rainfall also estimated by the Gumbel model at several points in the study area. Weisse and Bois (2001) compared krigging and ordinary regression against topography to model 10-year and 100-year rainfall estimates for rainfall duration of 1 to 24 hours. The use of the extreme value theory in these examples, however, was reduced to at-site (independent) calculation of the extreme quantiles. The existence of a spatial structure in the distribution of the model parameters was not addressed. In

addition, if a map was wanted corresponding to a different return period or hazard level, new at-site estimations and interpolation were needed.

In this paper we explore the construction of a spatially continuous probability model in which the parameters of a Generalized Pareto distribution are allowed to vary spatially. The parameters are estimated locally by common at-site methods from the climatological records in the study area, and then distributed using spatial interpolation techniques. Once the spatial models of the distribution parameters are known, different maps of probabilities, quantiles and return periods can be derived according to the needs of the user, without undertaking new spatial interpolations. Moreover, compared to using a set of unrelated at-site probability models, the analysis of the spatial distribution of the model parameters results in a much more robust regional probability model. The spatial distribution of the parameters can also be interesting from a theoretical point of view.

2. Study area and elaboration of the data base

We have tested our methodology in a Mediterranean area in which the extreme precipitation are frequent and cause important social, economic and environmental damages (White et al. 1997; García-Ruiz et al. 2000; Lasanta 2003). The study has been carried out in the middle Ebro valley (NE of the Iberian Peninsula) (Figure 1).

Figure 2 shows the spatial distribution of the mean annual precipitation (1), the average annual daily maxima (2) and the ratio of the annual maximum event to the total annual rainfall in percentage (3), for the period 1951-2000. The study area covers 22973 km². Relief isolates the valley, largely impeding the

maritime influence, resulting in a continental-like climate. These features similarly determine the complexity of its climate, its principal feature being its aridity (Cuadrat, 1991; Creus and Ferraz, 1995; Creus 2001).

The annual precipitation oscillates between 300 and 450 mm in the valley bottom, and more than 800 mm in the northern and southern mountain areas. A high interannual variability is observed as a consequence of the alternation of dominant atmospheric patterns. There are years in which the precipitation greatly exceeds the mean value; in other years, the values are less than a third of the mean, and long drought periods are particularly frequent (Vicente-Serrano and Beguería-Portugués, 2003).

Extreme precipitation events are frequent in Mediterranean regions. In the Mediterranean coastland areas are recorded the most extreme events in the Iberian Peninsula (Romero et al. 1998), as a consequence of the higher influence of Mediterranean convective cellules that affect importantly these areas (Camarasa, 1993; Llasat and Puigcerver, 1994; Millán et al. 1995; Llasat, 2001). This cause that precipitation is concentrated in a little number days and annual maxima events represent a high percentage of the total annual precipitation (De Luis *et al.*, 1996; Martín-Vide, 2004). This influence can be seen in the study area, where the highest irregularity (map 3) is found on the South-east, close to the Mediterranean Sea.

The original data based consisted of 380 series of daily precipitation with different lengths. Due to the frequent changes of position of the observatories within the same locality we created new series merging the data of the observatories located in the same town (<10 km of distance). We then selected the series with less than 15% of daily data lost (Karl et al. 1995), and with a

common record period covering the range from 1951 to 2000. We adopted this strict criterion to ensure that all the series were sufficiently long to provide reliable estimates of the probability of extreme events (Jones 1997), and that they covered the same record period to avoid variability in parameter estimation due to interannual climatic cycles. This led to a final database of 43 observatories (See location in figure 1).

The reduction of the spatial coverage of the weather stations can introduce some limitations to the analysis of climatic variables. The use of a high density of weather stations facilitates the mapping process independently of the interpolation method selected (Weisse and Bois, 2002; Vicente-Serrano *et al.*, 2003). Nevertheless, long climatologic time series are not frequent in the Iberian Peninsula due to frequent changes in the location of observatories. Also, the series can greatly differ in the period of record.

The length of the dataset is an important aspect on the analysis of climatological variables, but in the case of extreme value analysis it becomes critical, since the samples are reduced to only the highest values in the range of the variable (Jones, 1997). The problem is even higher under Mediterranean climate, characterised by high interannual variability of precipitation. In Spain significant differences in the annual averages have been found between decades (Esteban-Parra *et al.* 1998; Rodríguez-Puebla *et al.* 1998; Rodríguez *et al.* 1999), which implies the need of long series to provide robust estimations. Significant temporal variability and trends in extreme events have also been found in different areas of the Iberian Peninsula (García *et al.* 1995; González-Hidalgo *et al.* 2003).

The adequate length of series needed to obtain reliable predictions about the frequency of extreme events is subject to debate. Benson (1952) indicated that the 50 year return interval estimations, by means of annual maxima following a Gumbel distribution with a 25 % error requires 39 years of data. Porth et al. (2001) analysed the adequate sample size for return interval estimation and found that 20 years of data provide estimates with a 20 % rate of error, and in order to obtain the return periods with less than a 20 % of error 25 or more years of data were necessary.

Considering the previous discussion, we decided to maintain a large temporal extent of the database (50 years), despite the reduction of spatial coverage. Although the methods used to analyse the probability of the extreme events do not require to have complete series, we decided to fill the lost data to assure that all the observatories recorded the same precipitation events, to avoid that the lack of a high event in only one observatory could affect the frequency distribution of the variable in that location.

The process of complete missing daily data is highly problematic. It is common practice to fill the gaps by means of statistical techniques. Karl et al (1995), Karl and Knight (1998) and Brunetti et al (2001 and 2002) filled missing values by generating random rainfall amounts based on the probability distributions of the variables studied. The goal of this procedure was not giving a realistic estimate of the unknown daily values, but to obtain data series of equal length without changing the probability distributions of rainfall amounts. To fill the gaps, we followed the procedure described by Romero et al (1998) to create a complete daily precipitation dataset in the Mediterranean coastland of the Iberian Peninsula. They used the weighted average of the records at the nearest

stations to fill the gaps in the weather station under study. The spatial criterion, used by them to select the auxiliary weather stations, was a limit distance of 60 km. In our case we selected only the weather stations around 15 km of the station to be completed. The high density of weather stations allowed completing records with a high reliability. Only in the rare occasions (< 5 % of total missing data) when no data was available at the circle of 15 km we considered a wider radius of 30. A total of 47317 daily data were filled, with an average of 1100 data in each weather station (5.7% of the data) and a range from 31 records (0.16%) to 2773 records (14.88%).

The quality control and homogeneity of daily climatological records is very complex. There are no standard tools in the scientific literature to test the homogeneity of daily data sets, as it is the case for monthly data sets. For this reason, different authors have addressed the problem by first aggregating the data from daily to monthly records, and applying the homogeneity tests on the aggregated series (Manton et al. 2001; Brunetti et al. 2002). This has been the method followed in this study.

Thus, we tested the monthly aggregated series against inhomogeneities using the standard normal homogeneity test developed by Alexandersson (1986) and Alexandersson and Moberg (1997). For this purpose the ANCLIM software was used (Štípanek 2004). Since we did not have the meta-data of the weather stations we decided to apply a test of relative homogeneity. The relative homogeneity method compares the temporal evolution of the series under study with the temporal evolution of a reference series, which records the evolution of the neighbour stations. The procedure developed by Peterson and Easterling (1994) was followed to create each reference series, consisting of using the

average of the five highest correlated series with the series to check (using the first difference series).

We only found significant shifts in five monthly series, which do not coincide with changes in the reconstruction of the series. The inhomogeneities were not important neither detected in other months on the same observatory. Moreover, we did not have the meta-data of the stations to be sure that the shifts coincide with changes in the location of the observatories. The shifts were recorded in small population centres or depopulated areas in which urban effect does not exist. For this reason we considered that there were no significant inhomogeneities in the dataset used. However, some inhomogeneities may remain in the data at a daily scale, that can affect the distribution of extremes in the data set and probably have some impact on the accuracy of our results.

3. Methods

3.1. Hazard estimation by the extreme value theory

The extreme value theory deals with the analysis and estimation of the tails of random variables, adjusting different distribution functions to the highest or lowest observations in a given data set (see Hershfield 1973; Smith 1990 and 2003; Reiss and Thomas 2001). This allows estimation of the probability of events greater than those observed during the period of record. The mostly used approaches for this are the series of maxima and the partial duration series.

The exceedances or partial duration (PD) series approach has been used since long (see, i.e., Todorovic and Zelenhasic 1970; Davison and Smith 1990). It is based on censoring the original sample at a certain threshold value x_0 , taking only the exceedances over that threshold: $y_i = x_i, \forall x_i > x_0$. In contrast with the

series of series of maxima, like annual maxima, the PD approach allows to include more cases in the sample, resulting in much accurate estimation of parameters (Madsen et al. 1997). As demonstrated by Pickands (1975), the series generated by exceedances over a threshold tend to converge to a Generalized Pareto (GP) distribution, provided that the threshold value \tilde{x}_0 is high enough. In the form commonly applied to PD series, the GP distribution is described by a shape parameter κ and a scale parameter α , and has the following cumulative distribution function:

$$P(Y \leq y | \alpha, \kappa) = \begin{cases} 1 - \left(1 - \kappa \cdot \frac{y}{\alpha}\right)^{\frac{1}{\kappa}}, & \kappa \neq 0 \\ 1 - \exp\left(-\frac{y}{\alpha}\right), & \kappa = 0 \end{cases}, \quad (y_i = x_i - x_0) \quad (\text{eq. 1})$$

It contains the Exponential distribution as a special case when $\kappa = 0$ (second expression), a form that was used extensively in the first applications of the PD scheme. For $\kappa < 0$ the distribution is long-tailed, and for $\kappa > 0$ it becomes upper-bounded with endpoint at $-\alpha / \kappa$. These two forms of the GP distribution are sometimes referred in the literature as the Fréchet and Beta distributions, respectively.

The probability of an exceedance Y is frequently expressed by its return period T , which is the expected time between two consecutive occurrences of the event. Under the assumption of homogeneous Poisson process, the return period is the inverse of the probability of exceedance; expressed in years, this equals:

$$t_y = \frac{1}{\lambda} \cdot (1 - P(Y \leq y)), \quad (\text{eq. 2})$$

where λ is a frequency parameter equalling the average number of occurrences of Y per year in the original sample. Expressed in the original scale, this gives:

$$E_x[\mathbf{T}] = \begin{cases} x_0 + \frac{\alpha}{\kappa} \cdot \left[1 - \left(\frac{1}{\lambda \cdot t} \right)^\kappa \right], & \kappa \neq 0 \\ x_0 + \alpha \cdot \log\left(\frac{1}{\lambda \cdot t}\right), & \kappa = 0 \end{cases} \quad (\text{eq. 3})$$

The return period is only an expected, or most probable, value. The probability that an event of magnitude X (expressed in the original scale) will occur at least once in a period of t years is:

$$P(X \geq x | \alpha, \kappa, t, \lambda) = \begin{cases} 1 - \left[1 - \left(1 - \kappa \cdot \frac{x - x_0}{\alpha} \right)^{\frac{1}{\kappa}} \right]^{\lambda \cdot t}, & \kappa \neq 0 \\ 1 - \left[1 - \exp\left(-\frac{x - x_0}{\alpha}\right) \right]^{\lambda \cdot t}, & \kappa = 0 \end{cases} \quad (\text{eq. 4})$$

A common problem when analysing natural data series is the likely presence of serial dependence or persistence in the process, due to the rainfall events tending to occur clustered in groups. This is especially true for Mediterranean climates, where long periods without rainfall are followed by events that last several days (Martín-Vide 1989; Llasat and Puigcerver 1994; Martín-Vide and Llasat 2000). This constitutes a major problem, since persistence affects the premise of homogeneous random process. For this reason we adopted a cluster approach, considering that the series of clusters of rainfall (consecutive days with precipitation higher than 0 mm) adapt better to a random process (Beguería, 2005). This is a common approach in the analysis of hydrological extremes (see eg., Cunnane 1979), and has been used with other environmental variables as well (eg. Smith 1989). Thus, series of cluster maxima (maximum daily intensities in the clusters) were constructed for each station, and the remaining analysis was performed upon this data set.

The following procedure was followed to perform the PD-GP analysis. First, a threshold value \bar{x}_0 was needed for extracting the exceedance series at every station. This is one of the most important issues in PD series modelling as the threshold level controls the size of the data sample. An important property of the GP model is that the shape parameter remains constant independently of the threshold (threshold stability), so the final predictions would be equivalent whichever the threshold value. Generally, a low threshold is to be preferred to a high one, since it maximises the amount of data used for estimation. However, if the threshold is too low, the asymptotic property would not apply, and the GP model would not fit the data correctly.

An appropriate tool to assess the validity of the GP assumption is the mean excess plot, which is a plot of the average mean excess over a threshold against the value of the threshold. If the process follows a GP distribution at a given threshold value \bar{x}_0 , then the mean excess plot should appear approximately linear from this value on. Considering this, we selected the highest 1200 cluster maxima values at every station, corresponding approximately to the 93.5 centile of the original data series ($\lambda = 23.7$ events per year). This can be considered a sufficiently low threshold level for a starting point. Mean excess plots were constructed using increasing threshold values, to evaluate the adequacy of the model to the data. In all cases the series generated were found to fit the GP model adequately at the lower threshold values.

Figure 3 shows, for example, the mean excess plot for the station of La Sotonera. As it can be seen, the dots follow a straight line from the lowest value of the threshold, which equals 7 mm. Additionally, we tested the convenience of the threshold by means of the probability-plot weighted correlation coefficient

(PPWCC, see below and also shown in Figure 3). As it can be seen, the process exhibits the highest values of this statistic around the selected threshold.

We estimated the GP parameters using the method of probability weighted moments (Hosking and Wallis 1987):

$$\begin{aligned}\hat{\alpha} &= \beta_0 - \left(\frac{\beta_0}{2 \cdot \beta_1 - \beta_0} \right) \\ \hat{\kappa} &= \left(\frac{\beta_0}{2 \cdot \beta_1 - \beta_0} \right) - 2\end{aligned}\quad (\text{eq. 5})$$

where the circumflex $\hat{}$ designates that the parameter is estimated from the sample. β_r is the order r probability weighted moment, obtained as (unbiased estimator, Landwehr et al. 1979):

$$\beta_r = \frac{1}{n} \cdot \sum_{i=1}^n \binom{i-1}{r} \cdot \binom{n-1}{r}^{-1} \cdot y_i = \frac{1}{n} \cdot \sum_{i=1}^n \frac{(i-1) \cdot (i-2) \dots (i-r)}{(n-1) \cdot (n-2) \dots (n-r)} \cdot y_i \quad (\text{eq. 6})$$

For assessing the goodness of fit of the GP models to the data, probability-plots (PP) were constructed, which compare the observed exceedances Y and the expected (GP predicted) ones \hat{Y} . This are obtained by solving equation 1 for y_j :

$$\hat{y}_j = \frac{\hat{\alpha}}{\hat{\kappa}} \cdot \left[1 - (1 - \text{EmpCDF}(j))^{\hat{\kappa}} \right] \quad (\text{eq. 7})$$

where $\text{EmpCDF}(j)$ is the empirical cumulative distribution of y_j , obtained using the plotting position formula of Landwehr et al. (1979):

$$\text{EmpCDF}(j) \equiv P(X \leq y_j) = \frac{j + A}{n + B} \quad (\text{eq. 8})$$

j being the position of y_j in the series of exceedances sorted in ascending order.

The values of $A = -0.15$ and $B = 0$ were used, as recommended by Landwehr et

al. (1979) for similar, long-tailed, distributions. The probability plot is a graphical technique for assessing if the data set follows a given distribution function (see Chambers et al. 1983).

A probability-plot weighted correlation coefficient statistic was computed as a measure of goodness-of-fit (PPWCC, based on Filliben 1975):

$$\text{PPWR}^2 = \frac{\left[\sum_{j=1}^n (y\hat{y} - n\bar{y}\bar{\hat{y}}) \cdot \omega_j \right]^2}{\sum_{j=1}^n (y_j^2 - n\bar{y}^2) \cdot \omega_j \cdot \sum_{j=1}^n (\hat{y}_j^2 - n\bar{\hat{y}}^2) \cdot \omega_j}$$

$$\bar{y} = \frac{1}{n \cdot \sum_{j=1}^n \omega_j} \cdot \sum_{j=1}^n y \cdot \omega_j \quad (\text{eq. 12})$$

$$\bar{\hat{y}} = \frac{1}{n \cdot \sum_{j=1}^n \omega_j} \cdot \sum_{j=1}^n \hat{y} \cdot \omega_j$$

where \bar{y} and $\bar{\hat{y}}$ are the weighted averages of y and \hat{y} , respectively. As the number of small exceedances is much greater than the number of high exceedances, a weighting function ω_j has been used in the calculation of the statistic. ω_j is proportional to the probability of exceedance, thus giving more importance to the highest, less-frequent, observations in the sample:

$$\omega_j = \frac{1}{1 - \text{EmpCDF}(j)} \quad (\text{eq. 11})$$

Additionally two error statistics, the probability-plot weighted bias error (PPWMBE) and the root mean square error (PPWRMSE), were computed to compare the results of the at-site GP models with the results from the regional model (see below). The PPWMBE represents the average error, and the PPWRMSE represents the standard deviation or errors. The prefix PP stands again for ‘probability-plot’, to reflect the fact that the comparison is made using

two estimators of probability, and not directly between real and estimated values, as common. The expressions for these statistics are described in table 1.

3.2. Hazard mapping using spatial interpolation methods

The results of the previous section are a series of at-site GP models from which is possible to obtain return periods, probabilities of extreme events, etc, only at places where a climatic station is located. The purpose of this section is to extend this analysis to the rest of the study area, allowing for the elaboration of maps showing the spatial distribution of the above mentioned extreme statistics. As explained above, our approach has been to develop a spatial probability model whose parameters are dependent on the location of the point of interest. The local parameters α , κ and \bar{x}_0 , obtained by the procedure outlined above, were thus the dependent climatic variables to model. With respect to the frequency parameter λ , used in eqs. 3 and 4, we observed that it did not present significant variability between stations, so we used a fixed value corresponding to the average of the at-site values.

There are different methods to predict the distribution of spatial variables: global, local, and geostatistical techniques (Burrough and McDonnell 1998). It has been shown that in mountainous areas and in regions with complex atmospheric influences, as the middle Ebro valley, the local interpolators and geostatistical methods, which do not use external variables, do not show the real spatial variability of climatic variables at different scale, and yield higher prediction errors (Daly et al. 2002; Weisse and Bois 2002; Vicente-Serrano et al. 2003). In fact, these methods usually provide bad results if the sample network is not dense (Dirks et al. 1998).

On the other hand, regression is a global approach based on empirical relationships between the variable of interest and a other spatial variables, and tend to produce maps with a lower degree of generalisation. But also, applying smoothing filters of different sizes to the independent variables it is possible to capture the variability of the target variable at different spatial scales. Regression based techniques adapt to almost any space and usually generates adequate maps (Goodale et al. 1998; Vogt et al. 1997; Ninyerola et al. 2000). We thus selected a multiple regression scheme upon location parameters (powers of the coordinates of the point) and other spatially distributed independent variables. The value of a climatic variable at non-sampled points is predicted by the following transference function:

$$z(x) = b_0 + b_1P_1 + b_2P_2 + \dots + b_nP_n \quad (\text{eq 13})$$

where z is the predicted value at the point (x) , b_0, \dots, b_n are the regression coefficients and P_1, \dots, P_n are the values of the different independent variables at point x . A forward stepwise procedure with probability to enter set at 0.01 was used to select only the significant variables.

A list of the independent variables introduced into the model is shown in Table 2. The independent variables were generated from a digital elevation model (DEM) at a resolution of 1 km using the MiraMon GIS software. Latitude and longitude (km) were calculated in UTM-30N coordinates. They were used to fit first and second order trend surfaces to the data, that can account for broad global spatial trends. Elevation (km) could be extracted directly from the DEM and introduced to the regression model, considering the well known effect of elevation on precipitation. Slope and relief (difference in height between the point of interest and the highest point in a circle of 2.5 or 25 km) were directly

derived from the DEM. In a different way, they both give information about the topography energy in the surroundings of a point, yielding high values in mountain areas and lower values on more flat areas, independently of their elevation. Also the annual potential incoming solar radiation was estimated from the DEM and included in the model (Pons, 1997). Low-pass filters with radii of 2.5, and 25 km were applied to elevation, slope and incoming solar radiation maps in order to measure the influence of these variables at different spatial resolutions.

The normality of each variable was tested by the Chi-square test. \bar{x}_0 showed a normal distribution ($p = 0.70$), and this was also the case for \bar{x} ($p = 0.98$); in the case of the $\bar{\alpha}$ parameter a logarithmic transformation was needed to attain normality ($p = 0.95$).

Due to the likely presence of correlation between several of the independent variables, a conservative value of 0.01 was set for a variable to enter the stepwise procedure, as recommended by Hair et al. (1998). However, since collinearity problems can arise even with correlations as low as 0.3, a condition index test (SPSS v.12) was applied to the regression models. This test flags the models for which the coefficients uncertainty is too large due to multicollinearity, and recommends the use of a model with less number of parameters.

The spatial database was then sampled at the locations corresponding to the climatic stations. The regression analysis was performed upon this data base using SPSS v.12 software, and the final maps showing the distribution of the three parameters were obtained using Arcview v3.2 GIS.

We used a jack-knife cross-validation method to validate the maps of GP parameters. A recursive procedure removed one station at a time and calculated the regression model with the remnant points, storing the predicted values of the parameters for the removed station. This procedure was repeated until all the stations were evaluated, and the predicted values of the parameters were compared to the known values obtained before. The error statistics described in Table 3 (Willmot 1981) were used to compare both results.

Due to the global nature of the regression-based techniques, some local features can still not be represented in the model. For this reason, maps of the residuals (observed minus predicted values) of the parameters were visually inspected for spatial structure of the errors. A certain degree of spatial correlation was found in the case of the location and scale parameters, and a random pattern in the case of the shape parameter. For this reason, correction maps were constructed only for the two first parameters.

A local adaptive method, splines with tension ($\varphi = 400$), were selected for the interpolation of the errors (Mitasova and Mitas 1993). The final maps of \hat{x}_0 and $\hat{\alpha}$ were obtained by adding the regression estimates and the maps of interpolated errors.

4. Results

4.1- At-site extreme value analysis

The results of the at-site hazard estimation are shown in Figure 4. For every station, the maximum daily rainfall is related to its average return period. Note the different shapes of the curves, due to the use of local (not regionally-averaged) \hat{x} estimates.

Goodness of fit and error statistics for the at-site GP models are given in Table 4. The mean bias error shows very small skew of the predictions to the right (note that the scale is the same as the original variable, mm of exceedance above the threshold). The probability-plot error and goodness-of-fit statistics can also be considered very good. The average PPWCC is 0.972.

Two probability plots are given as examples in Figure 5, relating the observed exceedances with the theoretical values computed using eq. 6. The plots show one of the stations with best agreement (La Sotonera, PPWCC = 0.999) and one of the worst (Escatrón, PPWCC = 0.922). A certain overestimation of the highest values is seen in the second case, but in general the level of agreement can be considered very good. Very similar plots were found in the rest of stations.

4.2. Maps of the GP parameters

The standardised Beta-coefficients of the multiple regression of the GP parameters are shown in Table 5, along with the goodness of fit (R^2) and error statistics for the final models. Only the significant variables are shown. The three regressions were significant at a 95% confidence level. The R^2 values obtained can be considered good, especially for the origin (\mathfrak{x}_0) and scale ($\mathfrak{\alpha}$) parameters. The regression of the shape parameter (\mathfrak{k}) obtained a lower R^2 , as expected from the greater uncertainty involved in its estimation, but it still was significant. Error statistics for the regression models are also shown in table 6. For \mathfrak{x}_0 the mean absolute error was 0.03, a small value compared to the range of the variable (0.1 to 7.6). The agreement between observed and predicted values

is very high (0.99). The same occurs with α , which has a mean error of 0.12 for a variable range of 6.6 to 19.5. For κ the relative error is higher due to the great uncertainty of this parameter (0.06, range -0.175 to -0.025).

Figure 6 shows the maps of the different parameters. The value of \bar{x}_0 is related with the latitude, showing a gradual increase to the North (map 6.1). \bar{x}_0 also shows relation with the relief, but at a broad scale (10 km average). The spatial distribution of \bar{x}_0 mimics quite well the general distribution of the precipitation in the area.

The distribution of $\bar{\alpha}$ is slightly different, as shown in map 6.2. $\bar{\alpha}$ is negatively related with the longitude and the distance to the Cantabrian sea, capturing the gradual Mediterranean influence on climate towards the East. The consequence for the probability distribution of the extreme events is an increase in its variance, resulting in a more irregular distribution of the events. α is also related with the elevation and the relief, at mid-range scale (2.5 km).

In contrast with the high spatial detail of the previous parameters, $\bar{\kappa}$ presents a coarser spatial distribution, as it is only related to the latitude and longitude. The result is a gradient from the North towards the SW. The lowest values of $\bar{\kappa}$, resulting in greater importance of the right tail of the distribution and hence of the extreme events, are found along the SW limit of the study area.

Since the spatial regression of the parameters introduces a new source of uncertainty to the estimations, new validation statistics were calculated for the resulting models. Table 6 reports the error and goodness of fit statistics obtained comparing the empirical exceedance probability, Y , with the theoretical one obtained using the SPF method, \bar{Y} . The results from the validation were very

similar to the ones obtained with at-site estimated parameters (Figure 5), what supports that the SPF model can be reliable for estimating the degree of hazard of extreme rainfalls.

In Figure 7 are showed the probability plots for La Sotonera and Escatrón stations (same as in figure 5), using the spatially estimated parameters. Although in this case a certain overestimation is observed in La Sotonera station, the goodness of fit can be considered as being very good.

4.3. Spatial distribution of extreme precipitation hazard

With the parameter maps in figure 6 the estimation of return periods of extreme events and quantiles becomes possible at sites without climatic data records. The following figures are examples of this. The Figure 8 shows the distribution of the expected return period of a 100 mm event. In Figure 9 is shown the maximum daily-expected precipitation for different return periods. Finally, Figure 10 shows the probability that an event of 100 mm year⁻¹ occurs within different time periods. Other maps can be of course obtained for different requirements.

In general, the greatest hazard or heaviest precipitation is expected in the foothills of the Pyrenees and the Iberian Range, in the North and South-East of the study area respectively, coinciding with the greatest elevation and relief energy. But not only the relief explains the spatial distribution of the hazard of extreme precipitation, the atmospheric influences are also important. The South-east regions are highly affected by the Mediterranean perturbations that cause precipitation events of high intensity, mainly during the autumn, due to the instability caused by the contrast between the hot sea surface and the cold air

masses that arrive to this latitude (Millán et al. 1995; Serra et al. 1996). These perturbations become weaker as they ascend the Ebro valley (Creus and Ferraz 1995), and this fact can explain the reduction of the extreme precipitation hazard to the west.

Also, the higher magnitude of precipitation expected in the northern mountains, is explained by the influence of the perturbations associated to the Polar front that arrive from the West. The effect of these fronts on intense precipitation is greatly increased by the mountain ranges that oppose the West flows, creating a great contrast with the center of the valley (Creus and Ferraz, 1995; Ruiz 1982). Moreover, the mountains located to the north are affected by South-West flows associated to the negative phase of the North Atlantic Oscillation in winter (Martin-Vide and Fernandez 2001). These are regional flows that affect the whole of the Iberian Peninsula and are reactivated in the Northern Mountains due to vertical movements of the air masses (Esteban et al. 2002; Vicente-Serrano 2004).

The central areas of the valley are isolated from these air masses due to the mountains located in the north and south. The aridity is high in this area and extreme droughts are highly frequent (Vicente-Serrano and Beguería-Portugués 2003). Nevertheless the low hills located in the center of the valley (<700 m a.s.l.) modulate the spatial hazard of extreme precipitation events due to its role in the convective movements, which cause high precipitation during summer months in the center of the valley (Cuadrat 1999).

5. Discussion

The methodology presented in this paper has some points in common with the regional approaches to extreme flood estimation (see Cunnane 1988). Regional approaches were designed to maximize the information contained in different data series within a homogeneous region, with the result of providing more robust estimation of the distribution parameters. This applies specially to the shape parameter (κ), subject to a great uncertainty since high order moments are involved in its estimation. For this reason a common average value of κ is used in most regional approaches, while the other two parameters can be estimated locally or be averaged too. Regional methods have been extensively used for the estimation of extreme floods, but some examples exist of their use to model extreme rainfall (Alila 2000; Fowler and Kilsby 2002), and even some regionalisation methods have been proposed for this variable (Cong et al. 1993; DeGaetano 1998).

In this paper we have obtained statistically significant spatial models of the three parameters of the GP model, including the shape parameter, what represents an improvement over considering a simple average value. At least in our study area, we have demonstrated that spatial interpolation techniques can be used to obtain smooth continuous spatial representations of the probability model parameters, resulting in a robust regional model. Since the regional model contains the spatial distribution of the parameters of the probability distribution, this allows applying the theory of extreme value analysis also to ungaged locations. This is not the case for most of the regional approaches in the literature, where one or more parameters need to be estimated locally from a real data series.

We used a standard multiple regression model to explain the spatial distribution of the probability distribution parameters, what allowed for the inclusion of explanatory variables likely to be related with extreme rainfalls. For example, many authors have highlighted the relationship between precipitation intensity and elevation and other topographic features (Konrad II 1996; Prudhomme 1999; Prudhomme and Reed 1999; Lorente and Beguería 2002), but also the geographic location, the exposition to wind flows (Basist et al. 1994; Daly et al. 2002) or the proximity to hot seas (Llasat and Puigver 1994) can be determining factors that explain the spatial distribution of extreme precipitation.

Several comparative studies have determined the superiority of regression-based techniques in the case of variables with a complex spatial distribution and normally not very dense sampling network, as is normally the case of climatic variables. Analysing specifically extreme precipitation data, Weisse and Bois (2002) concluded that geostatistical methods (kriging) performed better than regression models only when the gauging network was dense enough. Similar results have been published for other climatic variables (Dirks et al. 1998; Vicente-Serrano et al. 2003). It must be noticed, however, that other spatial interpolation techniques like geostatistics (kriging) have also great potential for capturing the variability of a spatial variable, and its use should not be avoided a priori when analysing other data sets.

The results from the crossvalidation showed a good agreement with the estimates obtained by using standard at-site techniques at the gauged location, supporting the validity of the proposed method to estimate the hazard of extreme rainfalls also at ungauged locations. This is true, at least, for locations showing similar characteristics in terms of the predicting variables to the set of climatic

stations used for creating the spatial models. At points where one or more of this variables are far outside the range of the observations, however, it is necessary to extrapolate the results of the spatial models, and the results have to be considered as subject to higher uncertainty. This represents an important problem for any study addressing the spatial distribution of a climatic variable, since the stations rarely present a random spatial distribution, as it would be desirable in an ideal case. In this study we have considered one year as the temporal unit of study, as it is common practice in the analysis of heavy rainfalls for the purpose of hazard assessment. No seasonal effects or different synoptic situations have been considered, since they present a complete cycle within the period of one year, and thus do not introduce differences between different years. The explained methodology, however, can also be applied to datasets where the events have been classified according to the synoptic situation, or to the season of the year. This approach would allow developing different maps accounting for the spatial distribution of extreme rainfalls for every season or meteorological situation.

6. Conclusions

In this paper we address the problem of mapping the hazard of extreme precipitation, linking the theory of extreme value analysis and spatial interpolation techniques. We have showed that it is possible to obtain a probability model in which the distribution parameters vary spatially, yielding a robust regional extreme value model. As a product, maps showing the hazard of extreme precipitation at different recurrence times can be developed, or single values can be derived for non-gauged sites.

Starting from daily rainfall records, we have used partial duration series of rainfall cluster maxima, fitted to a General Pareto distribution, to obtain at-site estimates of the model parameters. The at-site magnitude-frequency curves have been validated against empirical frequency estimations using probability-plot error statistics.

We have addressed the spatial distribution of the probability model parameters using geo-regression techniques including location and other spatial independent variables as predictors, obtaining significant and well-fitted models. The residuals of two of the parameters, location and scale, have been incorporated to the maps by local interpolation using splines. A jackknife validation scheme has been used to provide goodness-of-fit and error statistics, showing very good results.

Acknowledgements

The authors want to acknowledge financial support from the following projects:

PIRIHEROS (REN2003-08678/HID); PROHISEM (REN2001-2268-C02-01/HID); BSO2002-02743; REN2002-01023-CLI and REN2003-07453, funded by Ministerio de Ciencia y Tecnología (Spain) and UE-FEDER, and "Programa de grupos de investigación consolidados" (BOA 147 of 18-12-2002), funded by Aragón Government. Research of the authors was supported by postdoctoral fellowships by the Ministerio de Educación, Cultura y Deporte (Spain).

References

- Alexandersson, H. and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for lineal trends. *International Journal of Climatology*, **17**, 25-34.
- Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *Journal of Climatology*, **6**, 661-675.
- Alila, Y, 2000: Regional rainfall depth-duration-frequency equations for Canada. *Water Resources Research*, **36**, 1767-1778.
- Basist, A., G.D. Bell and V. Meentemeyer, 1994: Statistical relationships between topography and precipitation patterns. *Journal of Climate*, **7**, 1305-1315.
- Beguiría, S. and A. Lorente, 1999: Distribución espacial del riesgo de precipitaciones extremas en el Pirineo aragonés occidental. *Geographicalia*, **37**, 17-36.
- Beguiría, S., 2005: Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value, *Journal of Hydrology*, **303**: 215-230.
- Benson, M.A. 1952: Characteristics of frequency curves based on a theoretical 1000 year record, US Geological Survey Water-Supply Paper 1043-A. US Government Printing Office, Washington DC, pp. 51-74.
- Bruce J.P., 1994: Natural disaster reduction and global change. *Bulletin of the American Meteorological Society*, **75**, 1831-1835.
- Brunetti, M., M. Mangeri, T. Nanni and A. Navarra, 2002: Droughts and extreme events in regional daily Italian precipitation series. *International Journal of Climatology*, **22**, 543-558.

- Brunetti, M., M. Mangueri and T Nanni, 2001: Changes in total precipitation, rainy days and extreme events in northeastern Italy. *International Journal of Climatology*, **21**, 861-871.
- Burrough, P.A. and R.A. McDonnell, 1998: *Principles of Geographical Information Systems*. Oxford University Press, Oxford.
- Camarasa, A., 1993: La estructura interna de las tormentas mediterráneas. *Cuadernos de Geografía*, **54**, 169-188.
- Chambers, J., W. Cleveland, B. Kleiner and P. Tukey, 1983: *Graphical methods for data analysis*, Wadsworth.
- Cong, S, Y. Li, J. Vogel and J. Schaake, 1993: Identification of the underlying distribution form of precipitation by using regional data. *Water Resources Research*, **29**, 1103-1112.
- Creus, J. and J. Ferraz, 1995: Irregularidad pluviométrica y continentalidad térmica en el valle medio del Ebro. *Lucas Mallada*, **7**, 147-164.
- Creus, J., (2001): Las sequías en el valle del Ebro. En (Gil, A. y Morales, A. Eds.) *Causas y consecuencias de las sequías en España*. CAM, Alicante: 231-259.
- Cuadrat, J.M. 1999: *El clima de Aragón*. Cai 100. Zaragoza. 109 pp.
- Cuadrat, J.M., 1991: Las sequías en el valle del Ebro. Aspectos climáticos y consecuencias económicas. *Revista de la Real Academia de Ciencias*, **85**, 537-545.
- Cunnane, C., 1979: A note on the Poisson assumption in partial duration series models. *Water Resources Research*, **15**, 489-494.

Cunnane, C., 1988: Methods and merits of regional flood frequency analysis.

Journal of Hydrology, **100**, 269-290.

Daly, C., W.P. Gibson, G.H. Taylor, G.L. Johnson and P. Pasteris, 2002: A

knowledge-based approach to the statistical mapping of climate. *Climate Research*, **22**, 99-113.

Davison, A.C. and R.L. Smith, 1990: Models for exceedances over high

thresholds (with discussion). *Journal of the Royal Statistical Society, series B*, **52**, 393-442.

De Luis, M., J.C. González-Hidalgo and J.R. Sánchez, 1996: Análisis de la

distribución espacial de la concentración diaria de precipitaciones en el territorio de la comunidad Valenciana. *Cuadernos de Geografía*, **59**, 47-62.

DeGaetano, A, 1998: A Smirnov test-based clustering algorithm with application

to extreme precipitation data. *Water Resources Research*, **34**, 169-176.

Dirks, KN., J.E. Hay, C.D. Stow and D. Harris, 1998: High-resolution studies of

rainfall on Norfolk Island. Part II: interpolation of rainfall data. *Journal of Hydrology*, **208**, 187-193.

Easterling, D.R., G.A. Meehl, C. Parmesan, S.A. Changnon, T.R. Karl and L.O.

Hearns, 2000: Climate extremes: observations, modelling and impacts. *Science*, **289**, 2068-2074.

Esteban Parra, M.J., F.S. Rodrigo and Y. Castro-Díez, 1998: Spatial and

temporal patterns of precipitation in Spain for the period 1880-1992. *International Journal of Climatology*, **18**, 1557-1574.

Esteban, P., X. Soler, M. Prohom and O. Planchon, 2002: La distribución de la

precipitación a través del índice NAO. El efecto del relieve a escala local:

el Pirineo Oriental. In *El agua y el clima*. (Guijarro J.A., Grimalt, M., Laita, M. y Alonso, S Eds.). Palma de Mallorca. pp. 25-34.

Filliben, J.J., 1975: The probability plot correlation coefficient test for normality.

Technometrics, **17**, 111-117.

Fowler, H.J. and C.G. Kilsby, 2002: A weather-type approach to analysing water resource drought in the Yorkshire region from 1881-1998. *Journal of Hydrology*, **262**, 177-192.

Gajic-Capka, M., 1991: Short-term precipitation maxima in different precipitation climate zones of Croatia, Yugoslavia. *International Journal of Climatology*, **11**, 677-687.

García, J.A., A. Marroquin, J. Garrido and V.L. Mateos, 1995: Analysis of daily rainfall processes in lower Extremadura (Spain) and homogenization of the data. *Theoretical and Applied Climatology*, **51**, 75-87.

García-Ruiz, J.M., J. Arnáez, S.M. White, A. Lorente and S. Beguería, 2000: Uncertainty assessment in the prediction of extreme rainfall events: an example from the central Spanish Pyrenees. *Hydrological Processes*, **14**, 887-898.

González-Hidalgo, J.C., M. De Luis, J. Raventos and R. Sánchez, 2003: Daily rainfall trend in the Valencia Region of Spain. *Theoretical and Applied Climatology*, **75**, 117-130.

Goodale, C.L., J.D. Aber and S.V. Ollinger, 1998: Mapping monthly precipitation, temperature and solar radiation from Ireland with polynomial regression and a digital elevation model. *Climate Research*, **10**, 35-49.

- Groisman, P.Y., R.W. Knight, T.R. Karl, D.R. Easterling, B. Sun and J.H. Lawrimore, 2004: Contemporary changes of the hydrological cycle over the contiguous United States: trends derived from in situ observations. *Journal of Hydrometeorology*, **5**, 64-85.
- Groisman, P.Y., T.R. Karl, D.R. Easterling, et al., 1999: Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climatic Change*, **42**, 243-283.
- Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black., 1998: Multivariate data analysis. Prentice Hall International, Inc. 799 pp.
- Hershfield, D.M., 1973: On the probability of extreme rainfall events. *Bulletin of the American Meteorological Society*, **54**, 1013-1018.
- Hosking, J.R.M. and J.R. Wallis, 1987: Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339-349.
- Jones, J.A.A., 1997: *Global Hydrology*. Addison Wesley Langware. Harlow. Reino Unido. 399 pp.
- Karl, T.R. and R.W. Knight, 1998: Secular trends on precipitation amount frequency and intensity in the United States. *Bulletin of the American Meteorological Society*, **79**, 231-241.
- Karl, T.R., R.W. Knight and N. Plummer, 1995: Trends in high-frequency climate variability in the twentieth century. *Nature*, **377**, 217-220.
- Katz, R.W. and B.G. Brown, 1992: Extreme events in a changing climate. Variability is more important than averages. *Climatic Change*, **21**, 289-302.

- Konrad II, C.E., 1996: Relationships between precipitation event types and topography in the southern blue ridge mountains of the southeastern USA. *International Journal of Climatology*, **16**, 49-62.
- Kunkel, K.E., R.A. Pielke and S.A. Changnon, 1999: Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: a review. *Bulletin of the American Meteorological Society*, **80**, 1077-1098.
- Lana, X., G. Fernandez-Mills and A. Burgueño, 1995: Daily precipitation maxima in catalonia (North-east Spain): expected values and their spatial distribution. *International Journal of Climatology*, **15**, 341-354.
- Landwehr, J. M., N.C. Matalas and J.R. Wallis, 1979: Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resources Research*, **15**, 1055–1064.
- Lasanta, T., 2003: Gestión agrícola y erosión del suelo en la cuenca del Ebro: El estado de la cuestión. *Zubía*, **21**, 76-96.
- Llasat, M.C. and M. Puigcerver, 1994: Meteorological factors associated with floods in the north-eastern part of the Iberian peninsula. *Natural Hazards*, **5**, 133-151.
- Llasat, M.C., 2001: An objective classification of rainfall events on the basis of their convective features: application to rainfall intensity in the northeast of Spain. *International Journal of Climatology*, **21**, 1385-1400.
- Lorente, A. and S. Beguería, 2002: Variation saisonniere de l'intensité des précipitations maximales dans les Pyreneées Centrales: Analyse spatiale

- et cartographique. In C. Kergomard (ed.), *Publications de la Association Internationale de Climatologie, Vol. 14*, Aix-en-Provence, pp: 327-334.
- Madsen, H., P.F. Rasmussen and D. Rosbjerg, 1997: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At-site modeling. *Water Resources Research*, **33**, 747-757.
- Manton, M.J., P.M. Della-Marta, M.R. Haylock, et al., 2001: Trends in extreme daily rainfall and temperature in Southeast Asia and the South Pacific: 1961-1998. *International Journal of Climatology*, **21**, 269-284.
- Martín-Vide, J. and D. Fernández, 2001: El índice NAO y la precipitación mensual en la España peninsular. *Investigaciones Geográficas*, **26**, 41-58.
- Martín-Vide, J. and M.C. Llasat, 2000: Las precipitaciones torrenciales en Cataluña. *Serie Geográfica*, **9**, 17-26.
- Martín-Vide, J., 1989: Precipitaciones torrenciales en España. *Norba*, **6-7**, 63-79.
- Martín-Vide, J., 2004: Spatial distribution of a daily precipitation concentration index in peninsular Spain. *International Journal of Climatology*, **24**, 959-971.
- Millán, M., M.J. Estrela and V. Caselles, 1995: Torrential precipitations on the Spanish east coast: the role of the Mediterranean sea surface temperature. *Atmospheric Research*, **36**, 1-16.
- Mitasova, H. and L. Mitas, 1993: Interpolation by Regularized Spline with Tension. *Mathematical Geology*, **25**, 641-655.

- Ninyerola, M., X. Pons and J.M. Roure, 2000: A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, **20**, 1823-1841.
- Obasi, G.O.P., 1994: WMO's role in the international decade for natural disaster reduction. *Bulletin of the American Meteorological Society*, **75**, 1655-1661.
- Peterson, T.C. and D.R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *International Journal of Climatology*, **14**, 671-679.
- Pickands, J., 1975: Statistical inference using extreme order statistics, *Annals Statistic*, **3**, 119-131.
- Pons, X., 1997: Estimación de la radiación solar a partir de modelos digitales de elevaciones. Propuesta metodológica. In *VII Coloquio de Geografía Cuantitativa, Sistemas de Información Geográfica y Teledetección*. (J. Juriasti y I. Moro, Eds.). Vitoria-Gasteiz. pp: 87-97.
- Porth, L.S., D.C. Boes, R.A. Davis, C.A. Troendle and R.M. King, 2001: Development of a technique to determine adequate sample size using subsampling and return interval estimation. *Journal of Hydrology*. **251**, 110-116.
- Prudhomme C., 1999: Mapping a statistic of extreme rainfall in a mountainous region, *Physics and Chemistry of the Earth. (serie B)*, **24**, 79-84.
- Prudhomme, C. and D.W. Reed, 1999: Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland. *International Journal of Climatology*, **19**, 1337-1356.

Reiss, R.D. and M. Thomas, 2001: *Statistical Analysis of Extreme Values, from Insurance, Finance, Hydrology and Other Fields*, 2nd edition, 464 p. Birkhäuser.

Rodríguez, R., M.C. Llasat and D. Wheeler, 1999: Analysis of the Barcelona precipitation series, 1850-1991. *International Journal of Climatology*, **19**, 787-801.

Rodríguez-Puebla, C., A.H. Encinas, S. Nieto and J. Garmendia, 1998: Spatial and temporal patterns of annual precipitation variability over the Iberian Peninsula. *International Journal of Climatology*. **18**, 299-316.

Romero, R., J.A. Guijarro, C. Ramis and S. Alonso, 1998: A 30-year (1964-1993) daily rainfall data base for the Spanish mediterranean regions: first exploratory study. *International Journal of Climatology*, **18**, 541-560.

Ruiz, E., 1982: *La transición climática del Cantábrico Oriental al valle medio del Ebro*. Diputación Foral de Álava. Vitoria. 651 pp.

Serra, C., G. Fernández Mills, M.C. Periago and X. Lana, 1996: Winter and autumn daily precipitation patterns in Catalonia, Spain. *Theoretical and Applied Climatology*, **54**, 175-186.

Smith, R.L., 1989: Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, **4**, 367-393.

Smith, R.L., 1990: Extreme value theory. In W. Ledermann (ed.), *Handbook of applicable mathematics*, vol. 7, pp. 437-472, Chichester, Wiley.

Smith, R.L., 2003: *Statistics of extremes, with applications in environment, insurance and finance*, Dept. of Statistics, University of North Carolina, Chapel Hill, 62 p. Available on line at:

<http://www.stat.unc.edu/postscript/rs/semstatrls.pdf> (last visit: 27 oct 2004).

Štípanek, P., 2004: *AnClim - software for time series analysis (for Windows)*,

Dept. of Geography, Fac. of Natural Sciences, MU, Brno. 1.47 MB.

Todorovic, P. and E. Zelenhasic, 1970: A stochastic model for flood analysis.

Water Resources Research, **6**, 1641-1648.

Vicente Serrano, S.M. and S. Beguería-Portugués, 2003: Estimating extreme dry-spell risk in the middle Ebro valley (Northeastern Spain): A comparative analysis of partial duration series with a General Pareto distribution and Annual maxima series with a Gumbel distribution.

International Journal of Climatology, **23**, 1103-1118.

Vicente Serrano, S.M., 2004: *Evolución espacio-temporal de las sequías en el sector central del valle del Ebro: causas y consecuencias ambientales*.

Universidad de Zaragoza. Zaragoza. 1258 pp.

Vicente-Serrano, S.M., M.A. Saz and J.M. Cuadrat, 2003: Comparative analysis of interpolation methods in the middle Ebro valley (Spain): application to annual precipitation and temperature. *Climate Research*, **24**, 161-180.

Vogt, J.V., A.A. Viau and F. Paquet, 1997: Mapping regional air temperature fields using satellite derived surface skin temperatures. *International Journal of Climatology*, **17**, 1559-1579.

Weisse, A.K. and P. Bois, 2002: A comparison of methods for mapping statistical characteristics of heavy rainfall in the French Alps: the use of daily information. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, **47**, 739-752.

- Weisse, A.K. and P. Bois., 2001: Topographic Effects on Statistical Characteristics of Heavy Rainfall and Mapping in the French Alps. *Journal of Applied Meteorology*, **40**, 720–740.
- White, S., J.M. García-Ruiz, C. Martí, B. Valero, M.P. Errea and Gómez Villar, A., 1997: The 1996 Biescas campsite disaster in the Central Spanish Pyrenees, and its temporal and spatial context. *Hydrological Processes*, **11**, 1797-1812.
- Willmott, C.J., 1981: On the validation of models. *Physical Geography*, **2**, 184-194.

Table 1. Error statistics used to test the adequacy of the GP distributions.
 Symbols used: n , number of observations; y_j , observed exceedance value; \hat{y}_j ,
 GP predicted exceedance value; ω_j , weighting function (see text).

Probability-plot weighted mean bias error	$PPWMBE = \frac{1}{n \sum_{j=1}^n \omega_j} \sum_{j=1}^n (\hat{y}_j - y_j) \omega_j$
Probability-plot weighted root mean square error	$PPWRMSE = \sqrt{\frac{1}{n \sum_{j=1}^n \omega_j} \sum_{j=1}^n (\hat{y}_j - y_j)^2 \omega_j}$

Table 2. List of independent variables

Variable	Explanation
TREND1	First order trend surface ($y = b_1lat + b_2lon + c$)
TREND2	Second order trend surface ($y = b_1lat^2 + b_2lon^2 + b_3lat + b_4lon + b_5latlon + c$)
ELEV2.5, ELEV25	Mean elevation within a circle of 2.5 and 25 km
SLOPE2.5, SLOPE25	Mean slope within a circle of 2.5 and 25 km
RELIEF2.5, RELIEF25	Mean relief energy within a circle of 2.5 and 25 km (max. elevation - elevation at the point)
BARR_N2.5, BARR_N25, BARR_S2.5, BARR_S25, BARR_W2.5, BARR_W25, BARR_E2.5, BARR_E25	Barrier effect to the four cardinal directions (max. elevation within a wedge of radius of 2.5/25 km and mean direction N/S/W/E - elevation at the point)
RAD2.5, RAD25	Mean annual potential radiation within a circle of 2.5 and 25 km (J/m * day)

Table 3- Error statistics used in the validation of spatial regression models.
 Symbols used: n , number of observations; p_i , observed parameter value; \hat{p}_i ,
 predicted parameter value; \bar{p} , mean of observed values; $\hat{p}'_i = \hat{p}_i - \bar{p}$;

$$p'_i = p_i - \bar{p}$$

Mean absolute error	$MAE = \frac{1}{n} \sum_{j=1}^n \hat{p}_j - p_j $
Root mean square error	$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{p}_j - p_j)^2}$
Willmott's D (Willmott, 1981)	$D = 1 - \frac{\sum_{j=1}^n (\hat{p}_j - p_j)^2}{\sum_{j=1}^n (\hat{p}'_j + p'_j)^2}$

Table 4. Error and goodness-of-fit statistics for the extreme value analysis

Station	PPWMBE	PPWRMSE	PPWCC	Station	PPWMBE	PPWRMSE	PPWCC
Borja	0.089	4.200	0.929	Escatrón	0.065	3.582	0.922
Luesia	0.039	1.040	0.991	Albalate del Arzobispo	0.054	1.933	0.975
Ejea	0.041	0.771	0.995	Jatiel	0.058	3.345	0.924
Luna	0.042	1.575	0.981	Santolea	0.046	2.006	0.970
Uncastillo	0.031	0.666	0.996	Alcañiz	0.074	4.266	0.905
Sádaba	0.040	1.285	0.987	Candasnos	0.048	2.059	0.970
El Bayo	0.050	1.670	0.982	Mequinenza	0.057	2.625	0.964
Ricla	0.051	1.714	0.978	Radiquero	0.039	0.973	0.995
Longares	0.055	2.363	0.961	Berbegal	0.042	1.476	0.985
La Almunia	0.056	2.538	0.949	El Tormillo	0.029	1.593	0.982
Zaragoza	0.061	3.818	0.882	Belver de Cinca	0.042	1.160	0.990
Marracos	0.065	2.087	0.981	Sariñena	0.051	1.695	0.983
Lupiñén	0.058	2.154	0.977	San Julián de Banzo	0.037	0.697	0.997
Aniés	0.050	0.908	0.996	Huesca	0.042	0.775	0.996
La Sotonera	0.033	0.308	0.999	Pallaruelo	0.047	1.624	0.981
Almudévar	0.034	0.577	0.997	San Esteban de Litera	0.044	0.940	0.994
El Temple	0.040	1.271	0.987	Binéfar	0.055	2.460	0.960
Monegrillo	0.051	1.718	0.979	Fraga	0.060	3.771	0.908
Pina de Ebro	0.056	3.169	0.924	Pena	0.150	5.167	0.964
Azaila	0.065	2.827	0.956	Maella	0.072	4.097	0.917
Sástago	0.062	3.486	0.918	Fabara	0.040	0.633	0.997

Table 5. Multiple regression results: included variables, standardised (beta) coefficients, goodness-of-fit and error statistics

Param.	Independent variables					Goodness of fit statistics			
	Lat	Long	Elev2500	Pend2500	Pend25000	R^2	MAE	RMSE	Willmott's D
x_0	0.744	0.020	0.289	---	0.327	0.858	0.031	0.688	0.997
α	---	---	---	0.676	---	0.703	0.127	0.971	0.995
κ	0.717	0.395	---	---	---	0.516	0.056	0.044	0.992

Table 6. Error and goodness-of-fit statistics using the spatially estimated parameters.

Station	PPWMBE	PPWRMSE	PPWCC	Station	PPWMBE	PPWRMSE	PPWCC
Borja	-1.062	2.501	0.986	Escatrón	-0.343	2.459	0.950
Luesia	0.184	1.244	0.989	Albalate del Arzobispo	0.557	3.627	0.951
Ejea	0.407	1.581	0.990	Jatiel	0.117	3.526	0.919
Luna	0.109	1.678	0.979	Santolea	1.803	8.204	0.863
Uncastillo	0.615	1.809	0.990	Alcañiz	-0.318	3.036	0.935
Sádaba	0.146	1.542	0.984	Candasnos	0.062	2.093	0.969
El Bayo	-0.281	1.113	0.991	Mequinenza	0.165	2.862	0.961
Ricla	0.641	3.604	0.953	Radiquero	0.895	2.713	0.990
Longares	0.344	3.307	0.944	Berbegal	0.052	1.493	0.985
La Almunia	0.078	2.609	0.948	El Tormillo	1.397	2.810	0.999
Zaragoza	-0.615	2.237	0.946	Belver de Cinca	0.099	1.252	0.989
Marracos	-0.842	1.834	0.997	Sariñena	-0.072	1.445	0.986
Lupiñén	-0.743	1.935	0.991	San Julián de Banzo	0.615	1.447	0.998
Aniés	0.074	0.952	0.996	Huesca	0.105	0.870	0.996
La Sotonera	0.441	1.164	0.998	Pallaruelo	-0.031	1.455	0.984
Almudévar	0.649	2.136	0.989	San Esteban de Litera	-0.130	0.785	0.996
El Temple	0.180	1.629	0.983	Binéfar	-0.941	2.161	0.990
Monegrillo	-0.195	1.260	0.986	Fraga	-0.779	2.339	0.962
Pina de Ebro	-0.349	2.096	0.954	Pena	-1.290	3.061	0.995
Azaila	-0.103	2.324	0.965	Maella	-0.261	3.169	0.938
Sástago	-0.369	2.277	0.951	Fabara	1.897	5.902	0.961

Figure captions

Figure 1: Location of the study area and the weather stations used in this study

Figure 2. Distribution of the mean annual precipitation (1), average annual daily maxima (2) and ratio of the annual maximum event to the total annual precipitation (3, in %)

Figure 3: La Sotonera station: mean excess plot and probability-plot weighted correlation coefficient plot

Figure 4. Quantile plots for the different studied climatic stations, expressed as maximum daily rainfall vs. return period.

Figure 5. Probability plots for the stations of La Sotonera (A) and Escatrón (B): observed vs. theoretical exceedance values, regression line (dotted) and line of perfect fit (plain).

Figure 6. Maps of the three GP-parameters. 1, origin (\bar{x}_0); 2, scale ($\bar{\alpha}$); 3, shape ($\bar{\kappa}$).

Figure 7. Probability plots for the stations of La Sotonera (A) and Escatrón (B), using the spatially estimated parameters: observed vs. theoretical exceedance values, regression line (dotted) and line of perfect fit (plain).

Figure 8. Return period maps: expected return period of a precipitation of 100 mm day⁻¹.

Figure 9. Quantile maps: daily precipitation corresponding to return periods of 5 (1), 10 (2), 25 (3) and 50 (4) years.

Figure 10. Extreme probability maps: probability that an event of 100 mm day⁻¹ occurs at least once within a time period of 5 (1), 10 (2), 25 (3) and 50 (4) years.

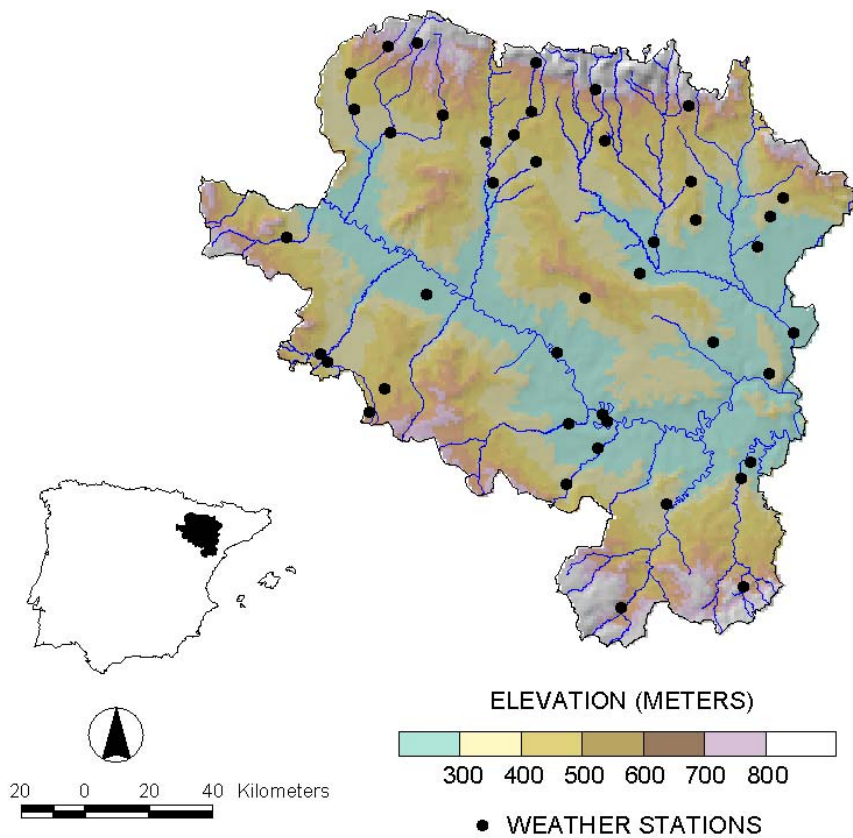


Figure 1

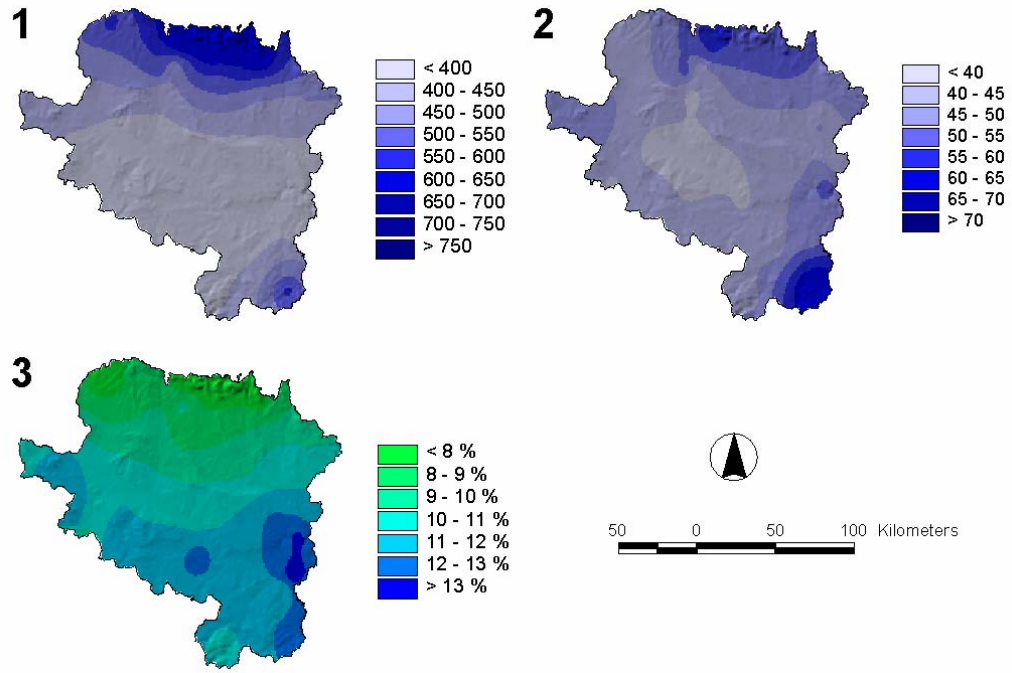


Figure 2

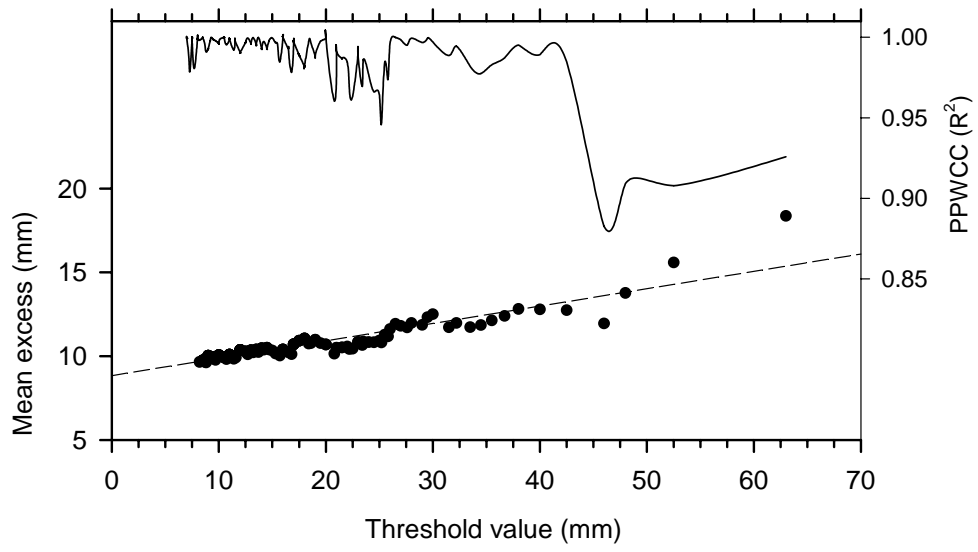


Figure 3

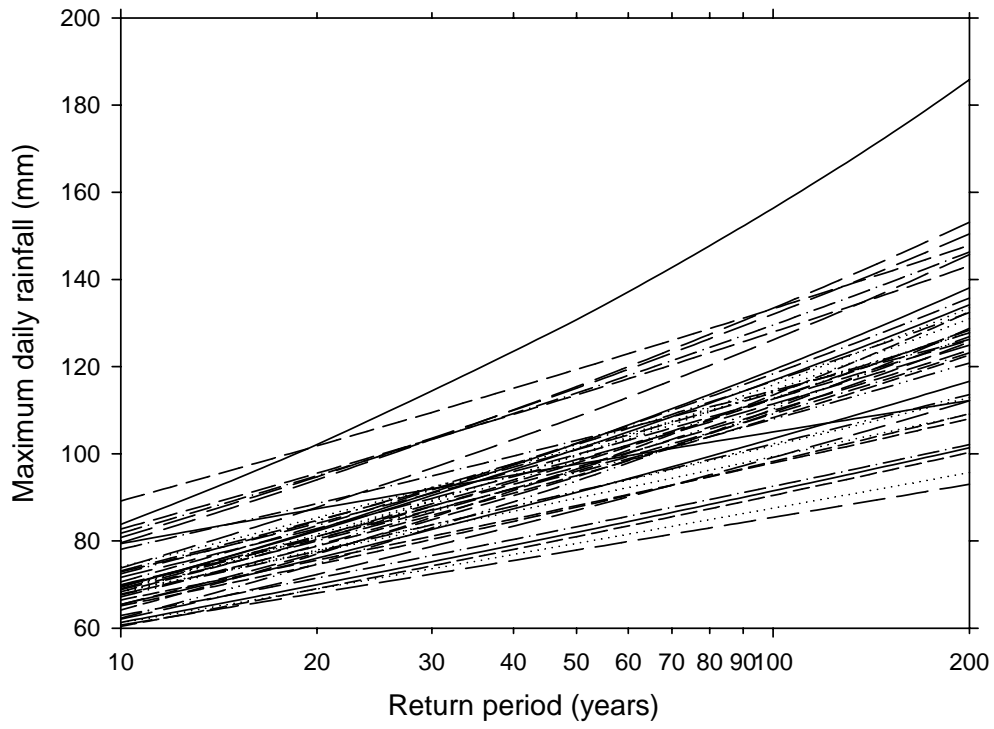


Figure 4

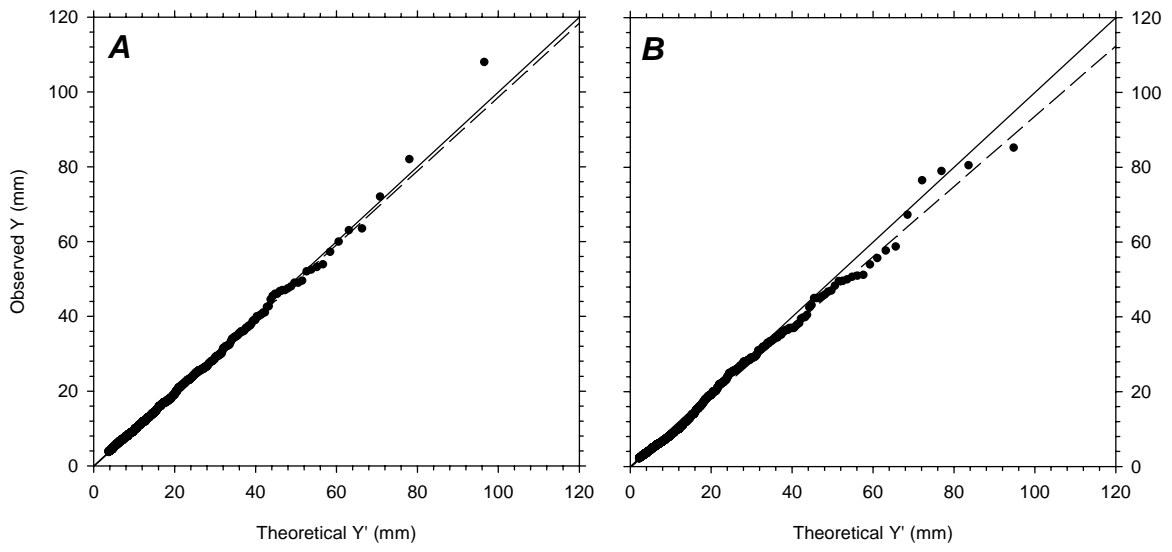


Figure 5

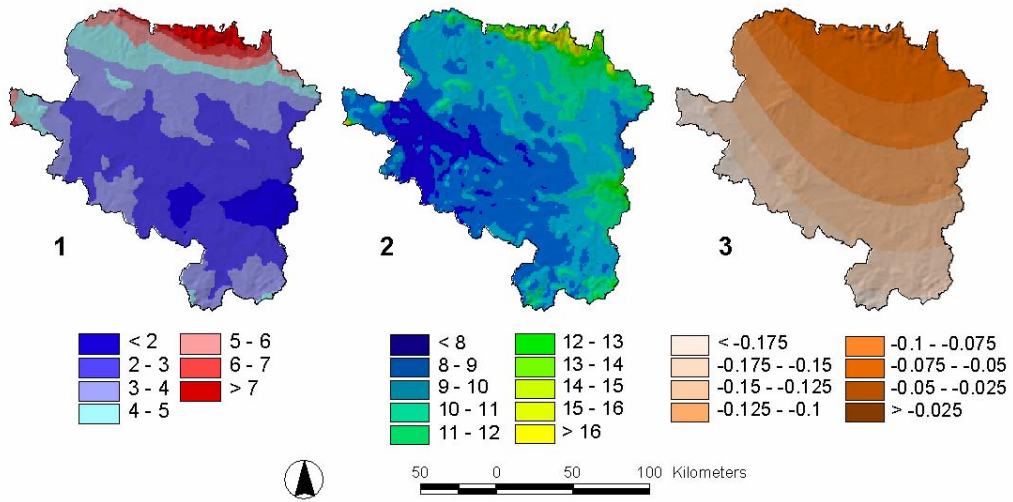


Figure 6

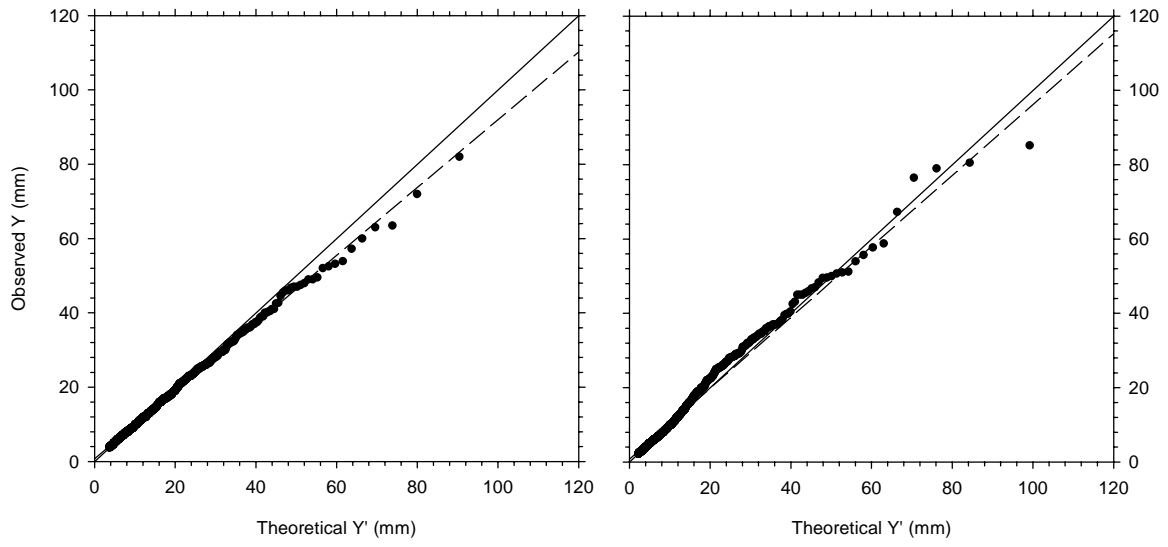


Figure 7

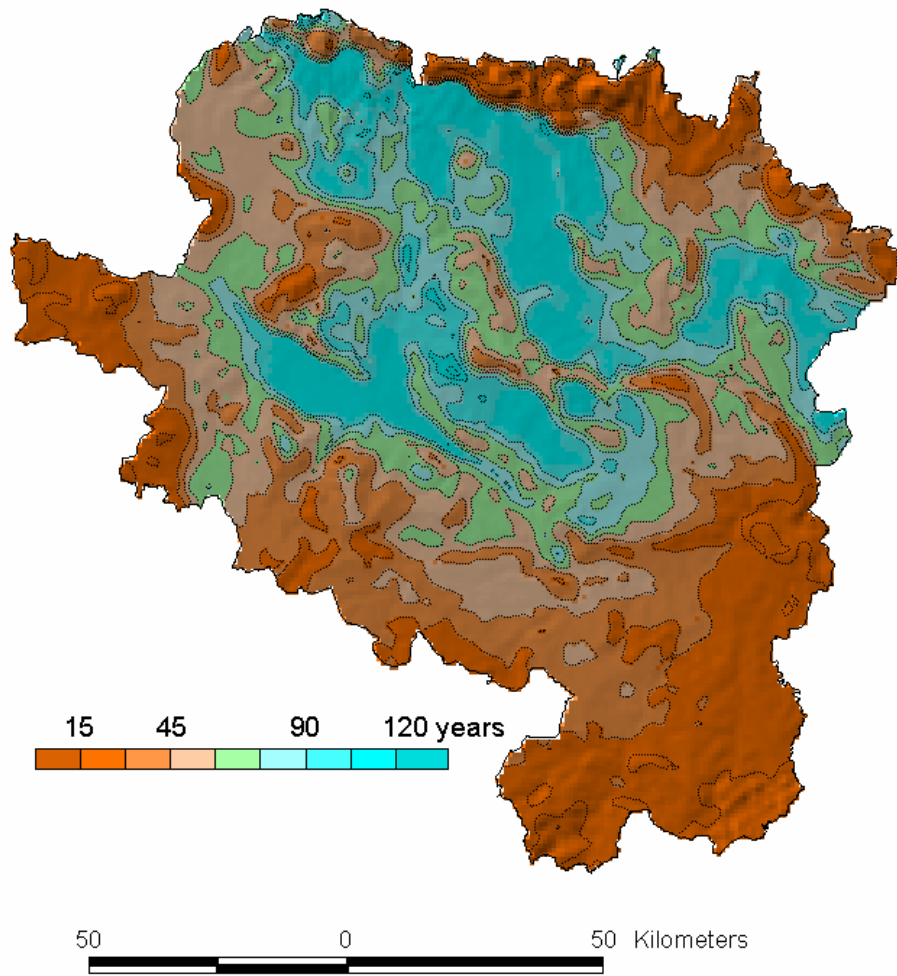


Figure 8

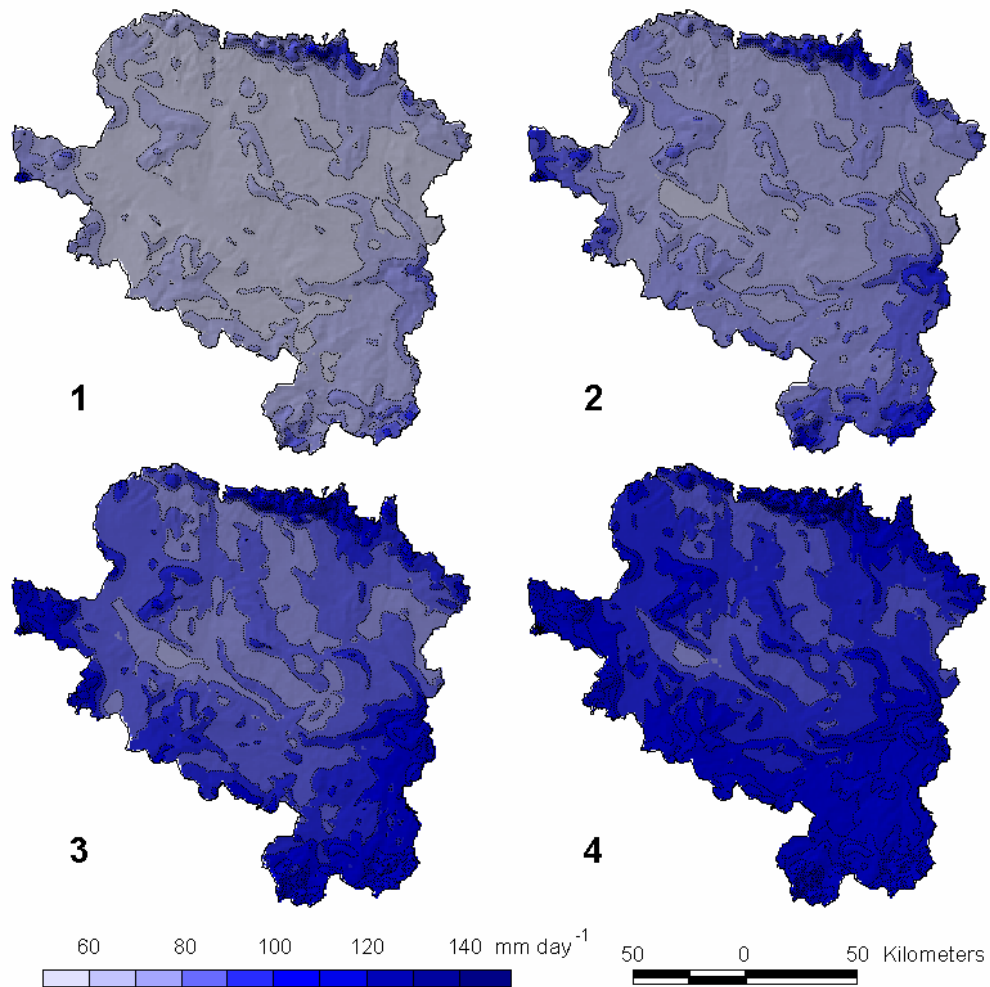


Figure 9

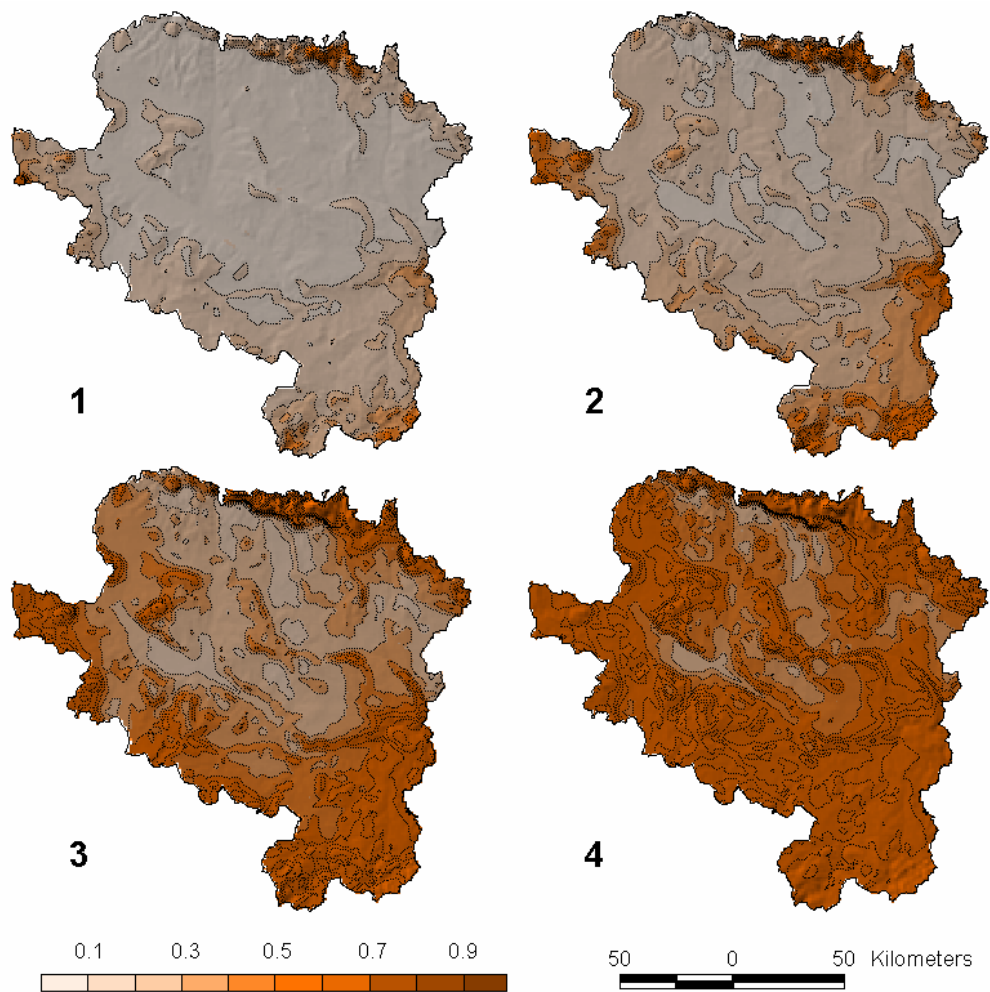


Figure 10