Journal of Hydrology (2008) 348, 27–39

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

A regional monthly precipitation simulation model based on an L-moment smoothed statistical regionalization approach

Javier González [a,*], **Juan B. Valdés** [b]

[a] *Department of Civil Engineering, University of Castilla-La Mancha, Spain*
[b] *Department of Civil Engineering and Engineering Mechanics, and Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), University of Arizona, USA*

Received 12 June 2007; received in revised form 5 September 2007; accepted 14 September 2007

KEYWORDS
Regional frequency analysis;
Orographic precipitation;
Smoothing regionalization;
Precipitation geostatistic;
Precipitation simulation

Summary Using rain-gauge station records for the statistical characterization and simulation modeling of spatio-temporal precipitation field involves many issues and simplistic assumptions. One major issue is related to dealing with uncertainty at-site sample statistical inference, because of the limited length of records. Regional frequency analysis uses the idea of substituting space for time in order to reduce uncertainty. It assumes equal shapes of the precipitation statistical distributions in a region. However, this assumption limits the area of the analyzed region where this assumption is valid. The extension is dependent on terrain complexity.

This work presents a new approach for the statistical regionalization of a large precipitation field, replacing the shape constancy assumption for the hypothesis of smooth spatial variation. The approach accounts for every uncertainty on site information, using an L-moment method for inference analysis. Additionally, the orographic effect is introduced in the regionalization, which substantially improves the interpolation performance and estimation of areal precipitation. The approach is used for modeling the monthly precipitation field in the Júcar River Basin Authority Demarcation (Spain), incorporating its stochastic structure, and spatial dependency coming from a geostatistical analysis. Issues related to the estimation of regional precipitation, and mean areal precipitation are discussed in the exposition.

Introduction

Regionalization is the inclusion in frequency analysis of data from sites other than the site at which statistic

* Corresponding author. Tel.: +34 926295422; fax: +34 926295391.
E-mail addresses: Javier.Gonzalez@uclm.es (J. González), jvaldes@u.arizona.edu (J.B. Valdés).

0022-1694/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2007.09.059

characterization is required. The method assumes that the frequency distributions of other sites are similar to the site of interest. Most of the statistical regionalization applications are concerned with reducing the uncertainty of extreme quantile estimation. The index-flood procedure of Dalrymple (1960) is an early example. A popular reference in applying regionalization methods for flood flow frequency analysis in the United States is the Bulletin 17B of the United States Water Resources Council (US Water Resources Council, 1976, 1977, 1981). The assumption of ''homogeneous region'' was suggested as invalid for US streamflow data (Benson, 1962). Thus there is reason to doubt whether regionalization is worthwhile. However, research has shown that even though a region may be moderately heterogeneous, regional analysis will still yield much more accurate quantile estimates than at-site analysis (Lettenmaier and Potter, 1985; Lettenmaier et al., 1987; Hosking and Wallis, 1988; Potter and Lettenmaier, 1990).

The effect of serial cross-dependence on frequency analysis has been investigated by Landwehr et al. (1979) and MacMahon and Srikanthan (1982). They found that serial cross-dependence causes a small amount of bias and a small increase in the standard error of quantile estimates. However, a small amount of serial cross-dependence in data series has little effect on the quality of quantile estimates (Hosking and Wallis, 1997). Therefore, regionalization is considered a powerful tool for improving the quality of statistical characterization and quantile estimation. The major hypothesis is the uniformity in the frequency distribution parameters. The serial cross-dependence may limit the ability to reduce uncertainty by increasing spatial information.

The uniformity and independence required in a regionalization process limit the applicability of the method. Uniformity restricts the maximum area covered by the database: distant sites are less probable to be similar in their frequency distributions. The second assumption, however, may require distant sites to be serially independent and to add information to regional statistical analysis. The space to meet these two restrictions depends much more on the nature and type of variable to be analyzed (e.g. streamflow or precipitation depth, and monthly or annual maximum), and often, on the terrain complexity (e.g. complex orography, geology, and soil uses distribution introduces spatial heterogeneity in the hydrological processes).

When large study areas are to be analyzed, the regionalization approaches fail to achieve spatial homogeneity. Uniform values of dispersion or shape frequency distribution coefficients are not to be expected, and applying regionalization for the whole area is not valid. Several approaches may be found in the literature to overcome this issue. Fiorentino et al. (1987) and Gabriele and Arnell (1991) proposed a procedure that involved a hierarchy of regions. Relatively large regions are defined over which the shape parameters are assumed to be constant, and these regions are subdivided into smaller regions over which the dispersion parameter is assumed to be constant. Hierarchical regions method may produce abrupt changes in the parameters when passing from one site to a neighboring site.

Fractional-membership procedures considers a site to provide information on several regions, rather than belonging to a particular region (Wiltshire, 1986; Acreman and Sinclair, 1986). Acreman and Wiltshire (1989) noted that the explicit construction of a region is not necessary in a fractional-membership procedure. This lead to the concept of *region of influence*, proposed by Burn (1990), where the weights must account for the influence of each site in the estimation of parameters or quantiles for any particular ''site of interest''. The largest disadvantage of the method is in the weight definition, where no universal definition exists. Another related approach is the mapping procedure, which is effective when a smooth relation to site characteristics can be found. Parameters or quantiles arising from regional frequency analysis are mapped or plotted against typical site characteristics in each region (Schaefer, 1990; Fill, 1994). A mapping approach can also be used with at-site estimates (McKerchar and Pearson, 1990). The major disadvantage when compared with a simpler regionalization procedure, such as the index-flood procedure, is the difficulty in estimating the accuracy of the final quantile estimates.

The different procedures proposed to overcome regionalization limitations seek to avoid the idea of regions defined by borders where abrupt changes occur in favor of a smooth continuity between regions. The real world produces a continuous field, where uniformity may be a sufficient approximation in some places. In this work a new regionalization approach is presented in which the uniform assumption is replaced by a more realistic assumption when large areas are analyzed: smooth spatial variation in the frequency distribution parameters. The different L-moments coefficients which define the frequency distribution function at each site are approximated by surfaces, which are fitted from at-site estimations, taking into account the sample estimation uncertainty and the orography factor. This provides a regional approximation of a random field distribution. The approach is used to model monthly precipitations in the Júcar River Basin Authority Demarcation (Spain), producing a spatio-temporal model for spatial precipitation simulation.

The precipitation database

The Júcar River Basin Authority Demarcation is located in the east of the Iberian Peninsula, along the Spanish Mediterranean coast. It comprises a vast area of 42,989 km^2, and it is composed of the aggregation of several river basins with outflow to the Mediterranean Sea: Cenia, Mijares, Palancia, Turia, Júcar, Serpis, and Vinalopó. The principal mountainous system is the Iberian System, in the northwest of the demarcation, where the maximum altitude of 2013 m above the mean sea level is achieved. The terrain provides high orographic complexity in the Iberian System and lower complexity in the Betic Mountainous System, which comprises part of the south basin borders (see Fig. 1).

The Iberian System plays an important role in the atmospheric dynamics in the region. It is exposed to the Mediterranean (east) and Atlantic (west) fronts. The mountain system provides a barrier so that few Atlantic fronts reach the Mediterranean coast, excepted in the south (Rodríguez-Puebla et al., 1998). Convection processes from the Mediterranean Sea and topographical configuration characterize the rainfall distribution over the coast (Millán et al., 1995).

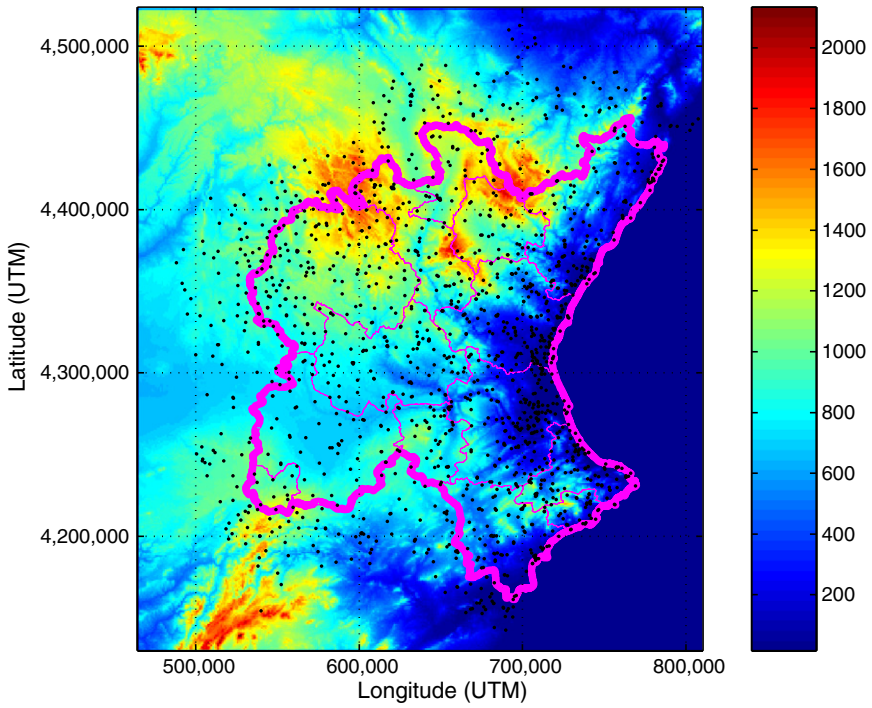

Figure 1 Digital elevation model for the Júcar River Basin Authority Demarcation (thick magenta line), and the rain-gauge stations distribution (black points). The study area has been divided into hydro-homogeneous regions (thin magenta lines). Elevation is expressed in meters above the mean sea level.

The precipitation database comprises the historical records of 1583 stations. It includes monthly precipitation depths, and extends from 1856 to 2006. During this period the number of active stations has been changing. Only a few stations were constructed before 1910, and the pace increased after that date until 1940, when almost one hundred stations were active. Then increased interest in improving the spatial precipitation information produced a progressive increase in the number of stations until 1990, when almost 750 stations were active.

No uniformity was found in the spatial distribution of rain-gauge stations. A higher station density is presented near to the coast (Fig. 1). Fig. 2 shows the $100 \times 100\ m^2$ DEM empirical cumulative probability distribution function (*c.d.f.*), which represents a bimodal distribution composed by the mountainous regions (i.e. higher than 700 m mean sea level) and the flat regions (i.e. lower than 700 m mean sea level). The empirical *c.d.f.* of the stations elevation is also plotted. Comparing both curves, a higher station spatial density around the coast is noticed. The flat region, which occupies 20% of the study area, contains more than 50% of the rain-gauge stations. However, a very low spatial density exits along the mountainous regions, where only a few stations characterize the precipitation regime in higher elevations. These regions are of great interest for water resources evaluation.

When comparing mean annual precipitation depths against elevations, two effects are observed (see Fig. 3). On one hand, orographic effects are shown in the precipitation distribution, noted specially in the mountainous regions, with an increasing precipitation trend for higher altitudes. Alternatively, annual precipitation depths along the different regions show high variability, depending on the rainfall front that reaches each region, and produced by local effects induced by the topographic configuration and sea coast proximity. The orographic effect in the flat regions is not so clearly evident in the figure because of the higher precipitation heterogeneity among flat regions. There the orographic effect is also presented, however due to the lower differences in altitude within each flat region, it has lower importance than other local factors.

Thus, the Júcar Demarcation database is characterized by irregular spatial and temporal sampling of the rainfall distribution. High variability in the mean annual precipitation depth is present, with local precipitation ranging from less than 200 mm to greater than 1000 mm. A regional statistical analysis of the monthly precipitation distributions is required to account for the whole data set, in order to obtain a better spatial description. However, because of the high variability of at-site precipitation, the sample uncertainty must be quantified and accounted for.

The L-moment smoothed statistical regionalization approach

Statistical regionalization looks to improve at-site statistical characterization by incorporating spatial data. Several methods have been described above, including *regional shape estimation* and *index-flood* to *hierarchical regions*, *fractional-membership*, *region of influence*, *mapping*, or *Bulletin 17*. However, none of these methods provide a universal approach to produce a smooth continuous spatial field of the statistical distributions. The application of these

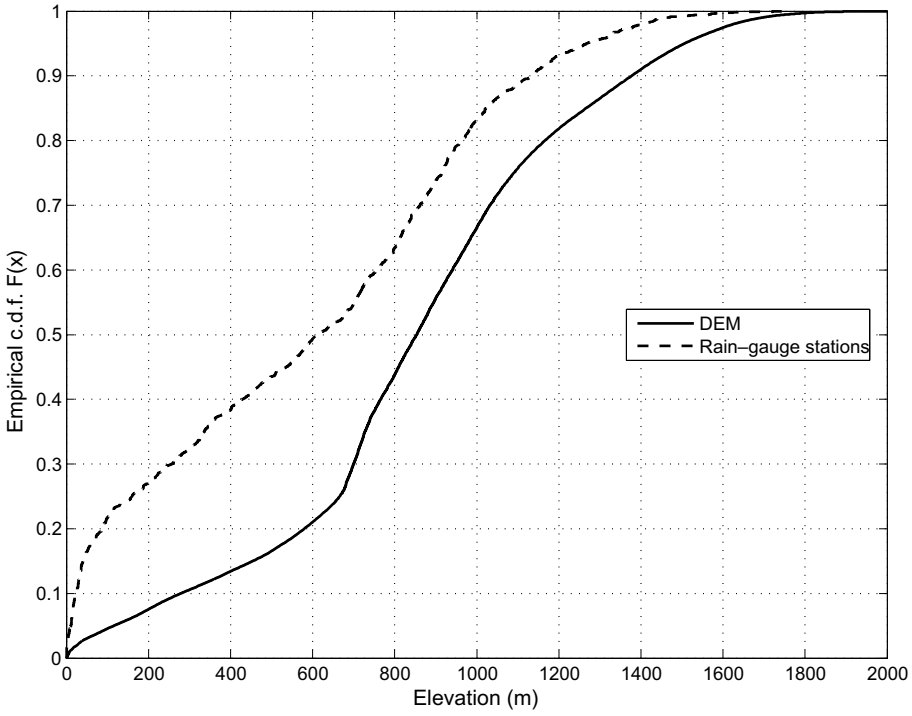

Figure 2 Empirical cumulative probability distribution function for the $100 \times 100\ m^2$ DEM and the rain-gauge stations elevation in the Júcar River Basin Authority Demarcation.

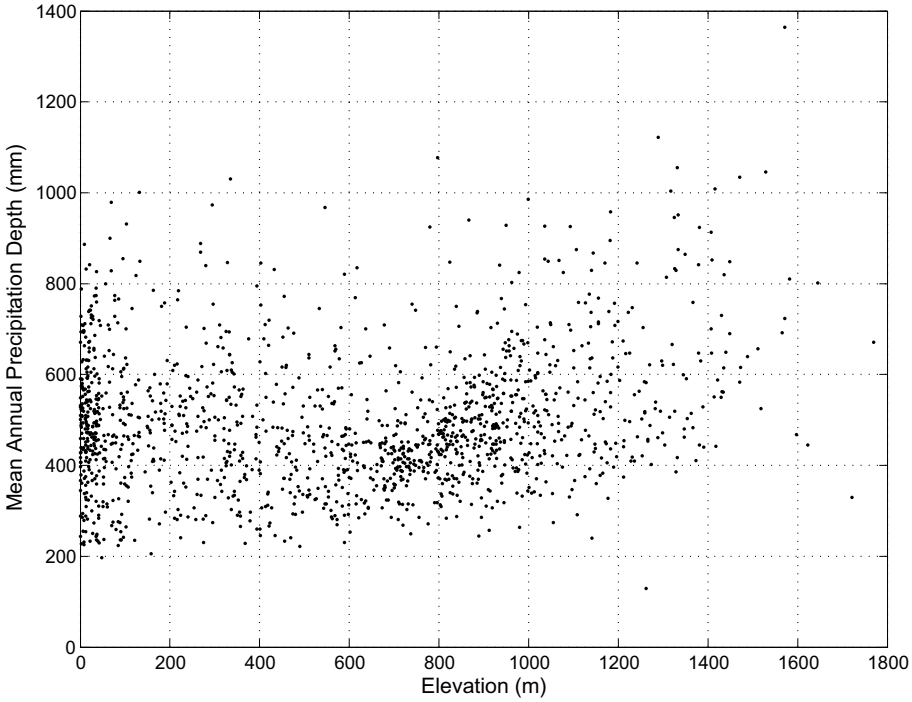

Figure 3 Mean Annual Precipitation Depth again the rain-gauge station elevation, taking from a $100 \times 100\ m^2$ DEM in the Júcar River Basin Authority Demarcation.

methods to the monthly precipitation distribution in the Júcar Demarcation is inappropriate because of its complexity. Below, a new approach is described for the statistical regionalization based on L-moments. This approach explicitly obtains the continuous spatial field of the statistical distributions of the regional precipitation. The goal is not only to characterize each site's statistical distribution, but to analyze the spatial statistical distribution of monthly pre-

cipitation over the whole Júcar Demarcation for water resources evaluation.

Fundamental steps in the proposed regionalization process are:

- selection of at-site statistical distribution models for each site in the database;
- site-by-site L-moments estimation and uncertainty quantification; and
- spatial L-moments distribution by smoothing the surface fitting, eventually accounting for orographic effects.

Selection of at-site statistical distribution model

The objective is to select one or more distribution families for statistical modeling. The distribution family may be unique for the entire study area, which is optimal because it simplifies the analysis. When complexity is large, the region might be divided in different subregions with a family distribution assigned to each one. However, this may produce abrupt changes in quantile estimation close to the subregions borders. For monthly precipitation, a flexible distribution model which produces good fit is the gamma distribution (MacKee et al., 1993).

Due to the existence of arid and semi-arid areas in the study region, some stations have months with probability of zero rainfall. Therefore, the gamma distribution must be adapted to account for this fact. The proposed model is the composite gamma distribution function

$$F(x) = \begin{cases} P_0 & x = 0 \\ \frac{1-P_0}{\beta^{\alpha}\cdot\Gamma(\alpha)} \cdot \int_0^x t^{\alpha-1}\cdot e^{-\frac{t}{\beta}}\cdot dt & x > 0 \end{cases} \quad (1)$$

The goodness of fit of the composite gamma distribution was evaluated for the data set by the Kolmogorov–Smirnov (K–S) test (Chakravarti et al., 1967). In applying the K–S test at each site, with a significance level of 5%, the null hypothesis was rejected in less than 5% of the stations. Additionally, rejected sites were dispersed over the study region, and not concentrated in only a few months. This validates the proposed distribution for monthly precipitations in the region.

Site by site L-moments estimation and uncertainty quantification

The L-moments are linear combinations of the elements of an ordered sample. L-moments have theoretical advantages over conventional moments in that they can characterize a wider range of distributions, are more robust to the presence of outliers in the data, and are less subject to bias in estimation (Hosking and Wallis, 1997). L-moments historically arose as modifications of the ''probability weighted moments'' (PWM) of Greenwood et al. (1979), defined by the quantities

$$M_{p,r,s} = E[X^p \cdot \{F(X)\}^r \cdot \{1 - F(X)\}^s] \quad (2)$$

Particularly useful special cases are the probability moments $\alpha_r = M_{1,0,r}$ and $\beta_r = M_{1,r,0}$. Measures of the scale and shape of a probability distribution are carried in certain linear combinations of the PWM: these are called the L-moments:

$$\begin{aligned} \lambda_1 &= \alpha_0 = \beta_0 \\ \lambda_2 &= \alpha_0 - 2\cdot\alpha_1 = 2\cdot\beta_1 - \beta_0 \\ \lambda_3 &= \alpha_0 - 6\cdot\alpha_1 + 6\cdot\alpha_2 = 6\cdot\beta_2 - 6\cdot\beta_1 + \beta_0 \end{aligned} \quad (3)$$

λ_1, λ_2, and λ_3 correspond with the position, the scale, and the shape moments, respectively. Landwehr et al. (1979) provides the expression for estimating the $\beta_r = M_{1,r,0}$ PWM from a finite sample

$$b_r = n^{-1}\cdot\binom{n-1}{r}^{-1}\cdot\sum_{j=r+1}^{n}\binom{j-1}{r}\cdot x_{j:n} \quad (4)$$

where $x_{j:n}$ is the jth smaller value in a n-sample ($x_{1:n} \leqslant x_{2:n} \leqslant \cdots \leqslant x_{n:n}$). It is convenient to define dimensionless versions of L-moments. The L-moment ratio $\tau = \lambda_2/\lambda_1$ is called the *coefficient of* L-*variation*, analogous to the ordinary coefficient of variation, and usually abbreviated L-CV.

For the statistical distribution fitting of the monthly precipitation depths only rain events are considered. Therefore, this corresponds to the fitting of a gamma statistical distribution. Using the L-moment estimates, the next relationships may be defined

$$\begin{aligned} \lambda_1 &= \alpha\cdot\beta \\ \tau &= \frac{\Gamma(\alpha+0.5)}{\alpha\cdot\sqrt{\pi}\cdot\Gamma(\alpha)} \end{aligned} \quad (5)$$

Therefore, the sample estimation of ℓ_1 and $t = \ell_2/\ell_1$, the $\hat{\alpha}$ and $\hat{\beta}$ parameters of each gamma distribution may be computed.

In addition to the above-mentioned advantages of using L-moments for statistical distribution fitting, the L-moments approach facilitates consideration of the sample uncertainty in the regional analysis. The sample estimation ℓ_1 follows a normal distribution, $\ell_1 \in N(\lambda_1, \sigma_\mu)$. σ_μ depends on sample size and on the distribution. The estimated L-CV is found also to follow a normal distribution, $t \in N(\tau, \sigma_t)$, when estimated from a gamma distributed population. The sample uncertainties σ_μ and σ_t can be quantified by applying Monte Carlo simulation techniques. Reiterative simulation of random n-sample generation from a population following a gamma probability distribution, in which parameters α and β are computed by assuming $\lambda_1 = \ell_1$ and $\tau = t$, and subsequent distribution fitting provide an estimation of the uncertainty coming from a n-sample estimation. The σ_μ and σ_t are estimated from the standard deviation of the simulated-fitted statistics $\ell_1(i)$ and $t(i)$, for i the simulation index $i = 1, \ldots, N$.

$$\begin{aligned} s_\mu^2 &= \frac{1}{N-1}\cdot\sum_{i=1}^{N}(\ell_1(i) - \ell_1)^2 \\ s_t^2 &= \frac{1}{N-1}\cdot\sum_{i=1}^{N}(t(i) - t)^2 \end{aligned} \quad (6)$$

Therefore, for every site not only the statistics sample estimations are computed, but also the distribution function of its estimation (Eq. (7)). These distributions are used to fit the spatial fields of both statistics.

$$\begin{aligned} \widehat{\lambda_1} &\to N(\ell_1, s_\mu) \\ \widehat{\tau} &\to N(t, s_t) \end{aligned} \quad (7)$$

Spatial L-moments distribution by smoothing surface fitting

After estimating the parameters of the distribution of the sample statistics (Eq. (7)) for the set of *Ns* rain-gauge stations (*Ns* = 1583) for every month, corresponding surfaces must be fitted to represent the spatial distributions. The process must account for a smooth spatial variation of the statistics (i.e. low surface roughness), while simultaneously the resulting field must be likely in the set of statistics distributions coming from the inference process at each site (i.e. $\ell_1 \in N(\lambda_1, \sigma_\mu)$, and $t \in N(\tau, \sigma_t)$). Thus the fitting approach must adapt to the spatial variability at the same time that it smooths sample errors. Additionally, the orographic effect must be able to be included if significant.

Kriging methods are nowadays a preferable option in the technical literature for hydro-climatic variable spatial interpolation (e.g. Hevesi et al., 1992; Phillips et al., 1992; Martínez-Cob, 1996; Holawe and Dutter, 1999; Goovaerts, 2000; Haberlandt et al., 2001). They allow for predictions at unsampled locations by capitalizing on the spatial correlation between neighboring observations. However, the kriging method does not show significantly greater predictive skill when comparing kriging and multiquadratic surface fitting, according to Borga and Vizzaccaro (1997). In fact, besides providing a measure of prediction error (kriging variance), a major advantage of kriging over simpler methods is that the sparsely sampled observations of the primary attribute can be complemented by secondary attributes that are more densely sampled. For rainfall, secondary information can take the form of weather-radar observations (Creutin and Obled, 1982; Azimi-Zonooz et al., 1989; Raspa et al., 1997), or a cheaper source of secondary information like a digital elevation model (Hevesi et al., 1992; Goovaerts, 2000).

Traditional kringing methods have the disadvantages of not being able to control fitted surface smoothness, and of not considering the sampled observations uncertainty, but exactly fitting the observed values. For these reasons, the application of kriging methods was rejected for this work. Instead spline surfaces were selected.

Smoothing spline surface fitness

Using the smoothing spline surface technique allows control of surface smoothness and accounts for the different uncertainty in the sampled observations. The fitting process is expressed as the optimization problem of minimizing Eq. (8)

$$p \cdot \sum_{i,j} w_{i,j} \cdot [z_{i,j} - ss(x_i, y_j)]^2 + (1-p) \cdot \int \left[\left(\frac{d^2 ss}{dx^2} \right)^2 + \left(\frac{d^2 ss}{dy^2} \right)^2 \right] \cdot dx \cdot dy \quad (8)$$

where $ss(x,y)$ is the smoothing spline, p is the specified *smoothing parameter*, and $w_{i,j}$ are the specified weights (de Boor, 1978). The weights must account for the differences in the uncertainty of sampled observations. Because sample estimations follow normal distributions, weights are related with the standard deviations s_μ and s_t, respectively.

Here, a cubic smoothing spline was used to fit each statistic, (λ_1 and τ), which requires a regular grid of observed points. Since rain-gauge stations do not follow a regular grid, the spatial information was translated to a designed regular grid. A quadrangular 10 km by 10 km grid was selected. The value of the statistic in each node in the grid was estimated by Eq. (9), assigning more importance to nearest stations by the inverse square distance procedure and accounting for the uncertainty. Consequently, the weight for smoothing spline fitness was evaluated by Eq. (10).The weights are inversely related to the square distance to the stations, and their standard deviation

$$v(i,j) = \frac{\sum_{k=1}^{Ns} \frac{v(k)}{\left[d_{(i,j)}^{(k)}\right]^2 \cdot s_v(k)}}{\sum_{k=1}^{Ns} \frac{1}{\left[d_{(i,j)}^{(k)}\right]^2 \cdot s_v(k)}} \quad (9)$$

$$w(i,j) = \sum_{k=1}^{Ns} \frac{1}{\left[d_{(i,j)}^{(k)}\right]^2 \cdot s_v(k)} \quad (10)$$

where v is a variable to be interpolated (i.e. ℓ_1 or t), evaluated at station (k) or at node (i,j); $d_{(i,j)}^{(k)}$ is the distance from the node (i,j) to station (k); and $s_v(k)$ is the standard deviation of the estimation of variable v at station k, with $k = 1, \ldots, Ns$, $i = 1, \ldots, I$, and $j = 1, \ldots, J$.

Correct selection of the *smoothing parameter* is one of the most critical steps in the regionalization process. The sensitivity of the results to p value may produce a surface which exactly meets each observation and has a high roughness ($p = 0$), or a flat surface that poorly matched the observations ($p = 1$). In order to select p, a maximum likelihood cross-validation criteria was used. The p that maximizes the likelihood of the interpolated value in each station was sought, when its observation was excluded from the observation set (i.e. delete one method). This is expressed in the form

$$V = \left[\prod_{k=1}^{Ns} \Phi(ss_v^{[k]}(k), v(k), s_v(k)) \right]^{\frac{1}{Ns}} \quad (11)$$

where $ss_v^{[k]}(k)$ is the interpolated v value at station k using the full set of observations, except the value at station k; and $\Phi(x, \mu, \sigma)$ is the probability density function of the normal distribution, evaluated at x, with mean μ and standard deviation σ.

This procedure provides an objective criteria for the selection of p, which is critical for the regionalization. A coupled source of sensitivity is the mesh resolution for smoothing spline fitness. Both are responsible for the level of smoothness achieved, and thus the resolution of the regionalization. A regionalization procedure looks for detracting sample randomness and filtering population randomness in the statistical inference process by accounting for spatial information. Additionally, it may try to estimate the statistical distribution at ungauged sites. However, these may be not the only goals for regionalization. For example, changing the resolution in the regionalization may be obtained by spatial averaging, thereby avoiding local effects that may come from point-measurements. Rain-gauges close to each other stations may produce different average precipitations which is unexpected from

the distance and precipitation heterogeneity. A suitable resolution in the regionalization may filter these local effects and provide a better estimation of the area average precipitation. The filter effect in the regionalization is controlled, in this case, by the grid resolution for smoothing spline fitness.

The objective in this case was to characterize the precipitation for water resources management. Therefore, a grid of 10 km by 10 km was considered suitable, given the objective and the expected spatial heterogeneity in the region. Three variables were used to characterize the statistical distribution of monthly precipitation, ℓ_1 and t for nonzero precipitation, and P_0 for zero precipitation.

Mean nonzero monthly precipitation fitting

Fig. 3 shows the orographic effect in the mean annual precipitation. Similar figures were obtained when analyzing for each month the average precipitation against elevation. However, the orographic effect in the precipitation is seasonally important, a function of the kind of fronts that produce rainfall.

To account for the orographic effect in the mean nonzero monthly precipitation the next model is proposed

$$\ell_1(x, y, z) = \ell_1^0(x, y) + \rho_{\ell_1} \cdot z \tag{12}$$

where $\ell_1^0(x, y)$ is the estimated mean precipitation in the point (x, y), considering no orographic effect (i.e. $z = 0$), which is taken into account by a linear function of the elevation z.

Eq. (12) separates the orographic effect from the local effect. To maximize the likelihood (Eq. (11)), in addition to the *smoothing parameter*, p, the slope ρ_{ℓ_1} in Eq. (12) plays a role. Thus, both variables must be adjusted simultaneously. Fig. 4 presents the resulting values of ρ_{ℓ_1}, and its seasonal behavior. The largest orographic effects are produced in May and June, when the impacts are not negligible. The mean annual slope implies an increment in the annual precipitation depth of 116 mm/km of elevation in a given area.

Fig. 5 shows the map of mean monthly precipitation in April after smoothing regionalization for the area of study. That month shows a high orographic effect in precipitation distribution, with the impact in the map.

L-CV nonzero monthly precipitation fitness

Traditional regionalization methods often hypotherize that the L-CV coefficient is constant. However in this work, this hypothesis is replaced by a continuous smooth variation of the statistic in the region of interest. After computing the standard deviation of the estimation of t at each station and month, using Eq. (6), it was checked that the hypothesis of uniformity of the L-CV coefficient in each month is unexpected. No value exceeds the 99% confidence intervals of the distribution of ts for the stations.

Therefore, a smoothed spline surface was fitted each month for the region. In the case of the t statistic, no orographic effect was found to be significant in any month, so only the ρ parameter was necessary to maximize the likelihood given by Eq. (11) for each month. Fig. 6 shows the resulting distribution of t in April for the Júcar Demarcation. The smoothness of the statistic is much more important than in the case of ℓ_1 (Fig. 5).

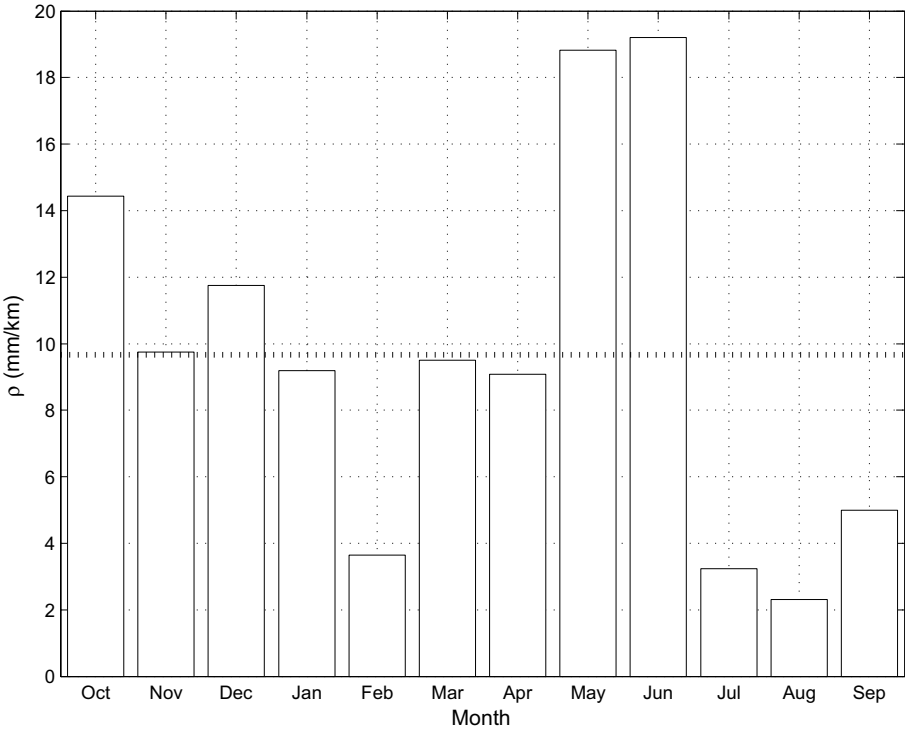

Figure 4 Seasonal orographic effect in the mean nonzero monthly precipitation, evaluated by the slope ρ_{ℓ_1} (bars), and annual mean slope (line).

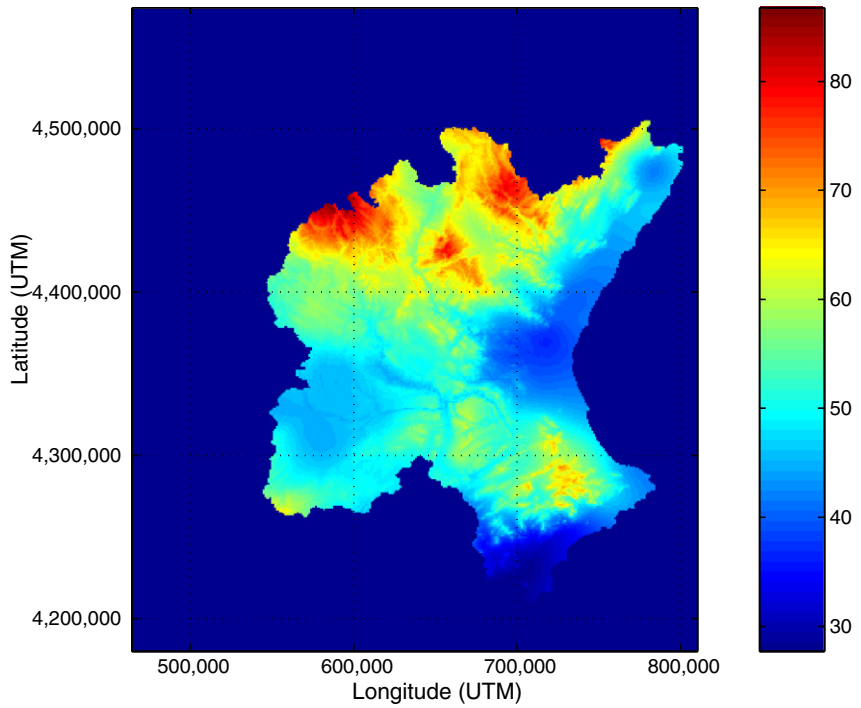

Figure 5 Spatial distribution of the mean (ℓ_1) precipitation depths (mm) in April, in the Júcar Demarcation, by smoothing regionalization, taking into account the orographic effect.

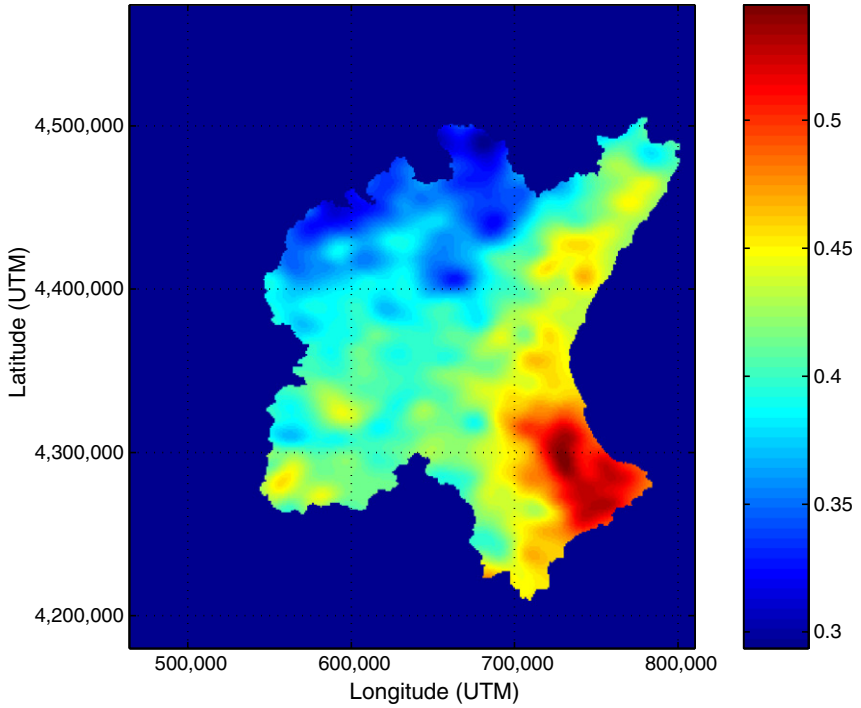

Figure 6 Spatial distribution of the coefficient of L-variation (t) of the precipitation depths in April, in the Júcar Demarcation, by smoothing regionalization.

Zero monthly precipitation probability fitness

The final statistics to define the monthly precipitation frequency distribution function along the study area is the probability of zero precipitation P_0. The event of producing zero precipitation during a month, with probability P_0, follows a binomial distribution. P_0 provides the frequency of those events, and is estimated by dividing the number of zero precipitation events (m) by the total number of observations (M) at a station. The distribution of the estimation depends on both values, m and M, but it does not follow a normal distribution. Instead, its distribution is related to the Fisher distribution (Johnson et al., 1992). Therefore,

the approach of smoothing regionalization approach presented here is not applicable. In this work, the estimation of P_0 computed at each stations was considered directly. The larger spatial variability of P_0 is produced during summer months (with lower precipitation depths), when small-extension storms introduce larger spatial variability with respect to humid months. Thus, the impact of not computing regionalization of P_0 is small for the objectives of the study.

A spatio-temporal simulation model

The spatial L-moments distribution statistically characterizes the monthly precipitation in the Júcar Demarcation. This characterization is used here to fit a simulation model, which allows the generation of synthetic monthly precipitation traces that are consistent with fitted frequency distributions. In addition, the model produces likely spatio-temporal distributions. The temporal and spatial structure of the data is analyzed and modeled next.

Before starting the analysis, and in order to compute correlation coefficients to analyze temporal and spatial relationships, the data sets for every month and site have been transformed to a normal distribution. The normalization was performed using the statistical distribution in each month and site. Each normalized value corresponds to the $N(0,1)$ variable which cumulative probability distribution function (c.d.f.) coincides with the c.d.f. of the monthly precipitation. These rainfall c.d.f. distributions are given in Eq. (1). In order to apply the transformation in the lower tail of the distribution, the c.d.f of zero precipitation was considered $P_0/2$ instead of P_0. This reduces the truncation effect in the normalized distribution. With the normalized data set, the stochastic structure was analyzed first, followed by the spatial dependency.

The stochastic structure

The stochastic analysis of the normalized data set sought to define the temporal dependency of the monthly precipitation. First, the Pearson correlation coefficient was computed for every site, and between each two consecutive months: $(\text{Month}, \text{Month}+1) \rightarrow r(\text{Month}, \text{St})$; where St represent the station index. Also, to analyze the seasonal behavior in r, the confidence interval of the estimation was computed. The confidence interval of r is obtained by the Fisher Z-transformation (Fisher, 1915)

$$Z = \frac{1}{2} \cdot \ln\left(\frac{1+r}{1-r}\right) \tag{13}$$

For the transformed Z, the approximate variance $\sigma^2(z) = 1/(n-3)$ is independent of the correlation, and only depends on the sample size n. Furthermore, even the distribution of Z is not strictly normal; it converges rapidly to normal as the sample size increases for any values of ρ.

Knowing not only the estimated rs, but also their distributions, it is possible to check the hypothesis of stationary ρ (null hypothesis, H0), against seasonal ρ (alternative hypothesis, H1). Therefore, with a test significance $\alpha = 0.05$ it was accepted the null hypothesis in a percentage of month-sites lower than 10%. The temporal distribution of the month-sites that reject the hypothesis was accepted uniform, and it does not concentrate in a group of months. Additionally, the spatial distribution of the stations is disperse. Thus the hypothesis of stationarity in the monthly autocorrelation coefficient with $lag = 1$ was considered, and the stationary r(St)s were computed.

Smooth spatial variation in the ρs is also expected, much like it occurs with L-moment statistics. The hypothesis of uniform ρ over the region of study was not accepted, comparing the 95% confidence intervals of the computed ρs at every site. Therefore, a smooth spline surface was fitted for a smooth regionalization of r. The fitting was performed over the transformed Z, taking into account the corresponding variance at each site, following the above procedure (Section ''Smoothing spline surface fitness''). An orographic effect was observed for ρ and Z, and the next model will take this into account

$$Z(x,y,z) = Z^0(x,y) + \rho_Z \cdot z \tag{14}$$

where $Z^0(x,y)$ is the estimated Z-transformed monthly autocorrelation coefficient in the point (x,y), considering no orographic effect (i.e. $z = 0$), which is taken into account by a linear function of the elevation z. The slope ρ_Z takes a value of 0.0073 1/km. Fig. 7 shows the spatial distribution of r after smooth regionalization.

The largest monthly autocorrelation coefficient is produced in the north, with $r \approx 0.2$, along the Iberian Mountainous System. No significant autocorrelation is produced over the southern mountains, because of their smaller size and proximity to the coast.

The monthly autocorrelation coefficients regional distribution may be used to model the stochastic structure of monthly precipitation in the area, considering its temporal dependence. An autoregressive stochastic model AR(1) was considered to define the model (Box and Pierce, 1976):

$$X_{t+1} = \phi \cdot X_t + a_{t+1} \tag{15}$$

where X_t would correspond to the normalized monthly precipitation series in each station, ϕ is considered equal to r, and a_t is an independent random variable, which follows a normal distribution, $a \in N(0, \sigma_a^2)$, with $\sigma_a^2 = (1-r^2) \cdot \sigma_X^2$. The normal assumption of a_t is checked for every station, in addition to its independence. The series a_t is called the shock series, because it represents the fluctuation effect in the model. Thus, the model defined with Eq. (15) represents the stochastic structure in the monthly precipitations.

Spatial dependency model

In order to produce a simulation model of monthly precipitation in the Júcar Demarcation, in addition to the stochastic structure it is necessary to account for the spatial dependence of the at-site simulations. This spatial dependence is evaluated with the cross-correlation between two stations. The variable used for analyzing this spatial dependency is the residual of the stochastic model at each station, or shock time series a_t. Using this variable the temporal dependence model is disaggregated from the spatial dependence model, thereby simplifying the simulator model. This assumes that the shocks follow a correlated spatial structure that are cross-correlated in the at-site monthly precipitation series.

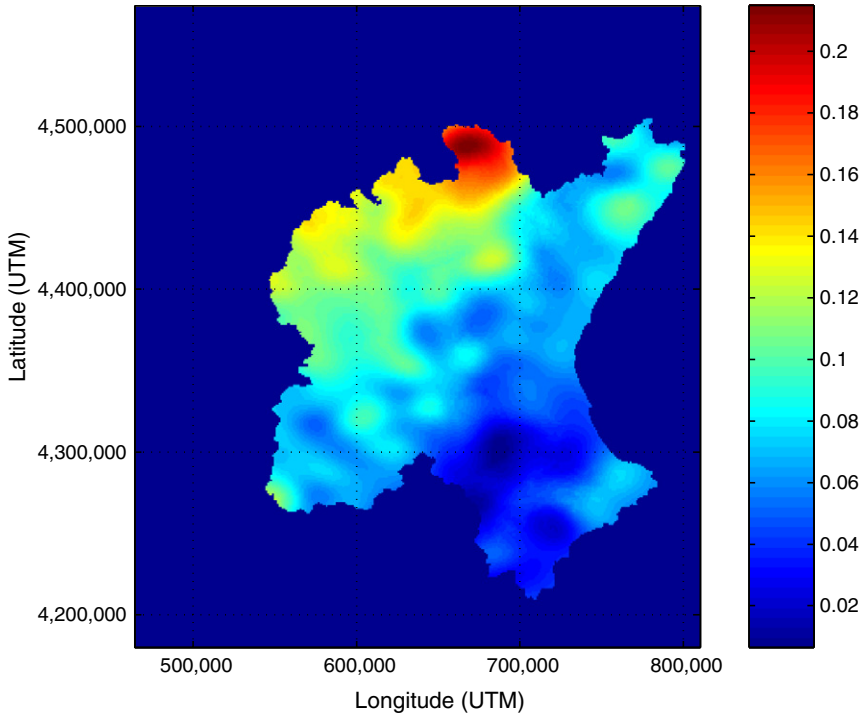

Figure 7 Spatial distribution of the stationary monthly coefficient of autocorrelation (r) of the precipitation depths in the Júcar Demarcation, by smoothing regionalization.

The hypothesis of stationary, (no seasonal) cross-correlation between each pair of stations was evaluated using the 5% confidence intervals computed by the Fisher Z-transformation (Eq. (13)). Assuming stationarity, the spatial dependence was analyzed by geostatistical procedures. The objective was to model this dependency by fitting an expression for the correlogram. The distance was found to be a significant variable in the correlogram definition. As distance increases, the cross-correlation between shock series decreases, with exponential and gaussian models being the best options for representing the behavior (Deutsch and Journel, 1997). Additionally, anisotropy effect was found in the correlogram, so the orientation between the two sites is relevant in the cross-correlation coefficient.

The best model considered was a composite model, which combined an exponential and a gaussian decay of the correlation coefficient C

$$C(h,\theta) = e^{-\frac{h}{A(\theta)}} \cdot e^{-\frac{h^2}{B(\theta)^2}} \quad (16)$$

where h is the distance, $\theta \in [0,\pi)$ is the azimut direction, and $A(\theta)$ and $B(\theta)^2$ are the decay coefficients functions of the exponential model and the gaussian model, respectively. These decay coefficients functions were expressed in the form of Fourier series

$$A(\theta) = a_0 + \sum_{i=1}^{n} [a_{i\cdot 2-1} \cdot \sin(2 \cdot i \cdot \theta) + a_{i\cdot 2} \cdot \cos(2 \cdot i \cdot \theta)] \quad (17)$$

$$B(\theta) = b_0 + \sum_{i=1}^{n} [b_{i\cdot 2-1} \cdot \sin(2 \cdot i \cdot \theta) + b_{i\cdot 2} \cdot \cos(2 \cdot i \cdot \theta)] \quad (18)$$

The coefficient a_i and b_i, $i = 1,\ldots,n$, were fitted to meet the maximum likelihood criteria, using the statistical distribution of the sampled cross-correlation coefficients, using the Fisher Z-transformation. The number of terms in the Fourier series, n, was chosen following the F-stopping-criteria, i.e. sequentially increasing the number of terms until the decrease in the mean square error is not significant with respect the decrease in the degrees of freedom. Resulting coefficients are presented in Table 1.

The obtained correlogram produces a uniform spatial dependence model over the region, which represents the cross-correlation between stations. This is used in the simulation model, which is validated below.

Model validation

The monthly precipitation smooth regionalization of the frequency distributions and the *lag* = 1 autocorrelation coefficients, in addition to the spatial dependency model, provide the components to simulate monthly precipitation time series in the Júcar Demarcation. The steps for N-years simulation at a set of stations are:

- Compute $12 \times N$ random set of values at each station, following a normal standard distribution, in accordance with the cross-correlation matrix produced by Eq. (16).
- Transform the time series at each station to $N(0,\sigma_a^2)$, with the σ_a^2 corresponding to the station, by σ_a multiplication.
- Incorporate the $AR1$ stochastic structure (Eq. (15)), and compute the normalized monthly precipitation time series at each station.
- Produce the monthly precipitation time series at each station by inverting the normalizing transformation, accounting for the frequency distribution at each station, given by the smooth regionalization (Eq. (8)).

Table 1 Fourier series coefficients (Eq. (17) and (18)) for the parameters of the correlogram model at Eq. (16)

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
a_i	6.4E−6	−1.2E−7	−2.8E−7	−1.5E−8	−3.7E−8	−1.3E−8	−1.2E−8	−1.7E−8	−8.2E−9	4.8E−9	7.0E−9	1.2E−8	−1.1E−9	−5.3E−9	−1.0E−8	1.7E−9	−4.0E−9
b_i	−4.1E−12	−2.6E−12	−9.1E−13	−2.6E−13	2.8E−13	3.1E−14	−1.8E−13	−2.2E−14	8.6E−15	−1.0E−14	−1.1E−14	−4.1E−15	−6.6E−15	9.7E−15	−4.9E−15	5.5E−15	5.1E−15

The model produces likelihood at-site simulations. It retains dependency properties for simultaneous simulations in a set of stations, and can simulate at sites with no previous observations. In order to validate the model it is appropriate to check the performance over areal averaging precipitations because the model has been fitted from at-site data. The regionalization used in the fitting procedure is expected to produce good spatial representation of the precipitation realizations. However, samples are only derived from at-site precipitation. The estimation of areal averaging precipitation from at-site data may be computed by the Thiessen polygon method (Thiessen, 1911), or an inverse square distance (ID2) procedure. Nevertheless, these areal averaging methods may produce significant errors because they do not consider orographic effects, for example. Therefore, comparing the statistical distributions of areal average precipitation computed from the at-site data and then provided by the simulation model (i.e. generating precipitation in a denser net) is not a valid test.

To adapt the validation test to the available data, the frequency distribution of areal average monthly precipitation coming from the sample data, and the simulated data were compared, using the realizations produced at the same time sequences and sites. This validation test followed the next steps:

- A representative time period and station set was selected from the sample. Stations with more than 20 year of data were selected, which avoid noise in the areal averaging coming from eventual further data. This produces a set of 1085 rain-gauge stations. Then a time period with enough number of simultaneous stations was sought with the period 1945/46 to 2005/06 selected.
- The selected data subset was used to compute the areal average monthly precipitation at subregions within the Júcar Demarcation. The subregions correspond to subcatchments, where hydrological processes are considered homogeneous, which conform the hydro-homogeneous regions in Fig. 1 (source: Júcar Demarcation). Over each areal average precipitation time series statistics are computed: the average and standard deviation of the monthly precipitation, and cross-correlation coefficient between pairs of hydro-homogeneous regions average precipitations.
- Monthly precipitation was simulated at the selected sites, with a duration equal to the selected period (i.e. 61 years). Because not all stations provided a sample for the full period, and this would impact in the areal average, only the simulated data corresponding with times and stations with observations in the selected period was considered in the areal averaging computation. The same statistics from the simulations were computed.
- The empirical cumulative probability distribution function (e.c.d.f.) was computed for every averaging area and statistic. Then, the e.c.d.f. for the statistics computed from the samples was calculated.
- The model was validated when the quantiles corresponding to the observed data fall into the expected intervals.

The validation test checks the areal behavior of the model, accounting for the areal averaging estimation error coming from the sample size and spatial structure, and the

Table 2 Validation test results: hydro-homogeneous regions and months percentages where the sampled data statistics fall into the 95% probability central interval of the simulated e.c.d.f.

Areal average method	Mean (%)	Standard deviation (%)	Correlation coefficient (%)
ID2	98.2	97.6	90
Thiessen	96.4	96.4	90

averaging method. It evaluates the sample likelihood based on the fitted model. The test has been performed using the Thiessen or ID2 averaging methods.

Table 2 compares the sampled data statistics to the empirical cumulative probability distribution function of the statistics obtained from simulation. In less than 5% of cases, mean and standard deviations for each month and hydro-homogeneous region fall into the 95% probability central interval of the probability distribution of the simulated statistics. For the cross-correlation between pairs of hydro-homogeneous regions and each month, the level of rejection is 10%, larger than the test significance. This may imply that the spatial correlation between regions is not fully represented. However, the level of rejection is still low. The correlation coefficient is quite sensitive to outliers or has a lack of robustness for skew distribution variables. Taking this into account, the rejection level may be due to high skewness of some monthly precipitation distributions. Nevertheless, the spatial correlation model, coming from the geostatistical analysis, assumes stationary correlogram in the study area. However, this hypothesis, usual in geostatistics, may be only an approximation in this problem. For large areas, the relationship between precipitation in two sites may depend not only on the distance and orientation, but also on the location of both points.

Conclusions

Smooth regionalization is thought to incorporate spatial information with smooth spatial continuity of the statistics used in frequency distribution analysis. This idea cannot be simplified with the hypothesis of uniformity when large areas are considered for study. Using spatial information may improve the statistical characterization of a variable, in comparison with using only at-site information, but requires an awareness of the uncertainty of every at-site statistic. The L-moment statistics provide a means of incorporating this spatial information while accounting for the uncertainty in the sample statistic estimations. The smoothing degree in the fitted surface is selected by a cross-validation maximum likelihood criteria. Therefore, three statistics, zero precipitation probability, and mean and standard deviation of nonzero monthly precipitation were fitted by the smooth regionalization approach. For the nonzero monthly precipitation, an orographic effect was observed and taken into account by introducing the elevation as an external drift in the regionalization. This improved the capacity of the regionalization for considering terrain complexity and its performance in interpolation problems.

The statistical regionalization was used to design a monthly precipitation simulation model in the Júcar Demarcation. The gamma distribution function was selected for modeling the probability distribution of nonzero precipitation in the area to normalize the data. Then the stochastic structure was analyzed and modeled by a stationary autoregressive model AR(1), which was again fitted by smooth regionalization. Orographic effect was also observed in the fitting and taken into account. Spatial dependence was modeled over the shocks of the stochastic model. After geostatistical analysis, a composed exponential and gaussian model was used for the correlogram, for which the parameters depend on the orientation.

The simulation model was validated, while checking if spatial averaging of the simulated precipitation results are likely. The proposed model was validated to represent the spatial mean and variance. Only the cross-correlations between areal average precipitation in separated areas produced lower levels of adequacy, and this may be an area for future improvements. However, the results are sufficient to consider the simulation model a suitable stochastic approximation of the complex spatio-temporal precipitation regime in the Júcar Demarcation.

Acknowledgements

The Júcar River Basin Authority Demarcation provided financial support and valuable assistance. The Spanish National Institute of Meteorology is acknowledge for providing the data set. All contributions are gratefully acknowledged.

References

Acreman, M.C., Sinclair, C.D., 1986. Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland. J. Hydrol. 84, 365—380.

Acreman, M.C., Wiltshire, S., 1989. The regions are dead: long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. In: Roald, L.K.N., Hassel, K.A. (Eds.), FRIENDS (Flow Regimes from Expimental and Networks Data) in Hydrology. International Association of Hydrological Sciences, Wallingford, Oxon, pp. 175—188, IAHS Publication 187.

Azimi-Zonooz, A., Krajewski, W.F., Bowles, D.S., Seo, D.J., 1989. Spatial rainfall estimation by linear and non-linear cokriging of radar-rainfall and raingage data. Stoch. Hydrol. Hydraul. 3, 51—67.

Benson, M.A., 1962. Evaluation of methods for evaluating the ocurrence of floods. Water Supply Paper, 1550-A, US Geological Survey, Reston, VA.

Borga, M., Vizzaccaro, A., 1997. On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. J. Hydrol. 195, 160—171.

Box, G.E.P., Pierce, D.A., 1976. Time Series Analysis Forecasting and Control. Holden-Day, San Francisco.

Burn, D.H., 1990. Evaluation of regional flood frequency analysis with a region of influence approach. Water Resour. Res. 26, 2257—2265.

Chakravarti, I.M., Laha, R.C., Roy, J., 1967Handbook of Methods of Applied Statistics, vol. I. John Wiley and Sons, pp. 392—394.

Creutin, J.D., Obled, C., 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison. Water Resour. Res. 18, 413—431.

Dalrymple, T., 1960. Flood frequency analyses. Water Supply Paper, 1543—A, US Geological Survey, Reston, VA.

de Boor, C., 1978. A Practical Guide to Splines. Springer-Verlag.

Deutsch, C., Journel, A., 1997. Geostatistical Software Library and Userś Guide, second ed. Oxford University Press, New York.

Fill, H.D., 1994. Improving Flood Quantile Estimates Using Regional Information, Ph.D. Thesis, Cornell University, Ithaca, NY.

Fiorentino, M., Gabriele, S., Rossi, F., Versace, P., 1987. Hierarchical approach for regional flood frequency analysis. In: Singh, V.P. (Ed.), Regional Flood Frequency Analysis. D. Reidel, Norwell, Mass, pp. 35—49.

Fisher, R.A., 1915. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. Biometrika, 507—521.

Gabriele, S., Arnell, N., 1991. A hierarchical approach to regional flood frequency analysis. Water Resour. Res. 27, 1281—1289.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hdrol. 228, 113—129.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resour. Res. 15, 1049—1054.

Haberlandt, U., Klöcking, B., Krysanova, V., Becker, A., 2001. Regionalisation of the base flow index from dynamically simulated flow components — a case study in the Elbe River basin. J. Hydrol. 248, 35—53.

Hevesi, J.A., Flint, A.L., Istok, J.D., 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part i: structural analysis. J. Appl. Meteorol. 31, 661—676.

Holawe, F., Dutter, R., 1999. Geostatistical study of precipitation series in Austria: time and space. J. Hydrol. 219, 70—82.

Hosking, J.R.M., Wallis, J.R., 1988. The effect of intersite dependence on regional flood frequency analysis. Water Resour. Res. 24, 588—600.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, NY, USA.

Johnson, N.L., Kotz, S., Kemp, A.W., 1992. Univariate Discrete Distributions, second ed. Wiley.

Landwehr, J.M., Malatas, N.C., Wallis, J.R., 1979. Probability-weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. Water Resour. Res. 15, 1055—1064.

Lettenmaier, D.P., Potter, K.W., 1985. Testing flood frequency estimation methods using a regional flood generation model. Water Resour. Res. 21, 1903—1914.

Lettenmaier, D.P., Wallis, J.R., Wood, E.F., 1987. Effect of regional heterogeneity on flood frequency estimation. Water Resour. Res. 23, 313—323.

MacKee, T.B., Doesken, N.J., Kleist, J., 1993. Drought monitoring with multiple timescales. In: Eighth Conference on Applied Climatology, Anaheim, Calif. Am. Meteorol. Soc.

MacMahon, T.A., Srikanthan, R., 1982. Log pearson type 3 distribution effect of dependence, distribution parameters and sample size on peak annual flood estimates. J. Hydrol. 52, 815—826.

Martínez-Cob, A., 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. J. Hydrol. 174, 19—35.

McKerchar, A.I., Pearson, C.P., 1990. Maps of flood statistics for regional flood frequency analysis in New Zealand. Hydrol. Sci. J. 35, 609—621.

Millán, M., Estrela, M.J., Caselles, V., 1995. Torrential precipitations on the Spanish east coast: the role of the mediterranean sea surface temperature. Atmos. Res. 36, 1—16.

Phillips, D.L., Dolph, J., Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitations in mountainous terrain. Agr. Forest. Meteorol. 58, 119—141.

Potter, K.W., Lettenmaier, D.P., 1990. A comparison of regional flood frequency estimation methods using a resampling method. Water Resour. Res. 26, 415—424.

Raspa, G., Tucci, M., Bruno, R., 1997. Reconstruction of rainfall fields by combining ground raingauges data with radar maps using external drift method. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong'96. Kluwer Academic, Dordrecht, pp. 941—950.

Rodríguez-Puebla, C., Encinas, A.H., Nieto, S., Garmendia, J., 1998. Spatial and temporal patterns of annual precipitation variability over the Iberian Peninsula. Int. J. Climatol. 18, 299—316.

Schaefer, M.G., 1990. Regional analyses of precipitation annual maxima in Washington State. Water Resour. Res. 26, 119—131.

Thiessen, A.H., 1911. Precipitation averages for large areas. Mon. Weather Rev. 39 (7), 1082—1084.

US Water Resources Council, 1976. Guidelines for determining flood flow frequency. Bulletin 17, Hydrology Committee, Washington, DC.

US Water Resources Council, 1977. Guidelines for determining flood flow frequency. Bulletin 17A, Hydrology Committee, Washington, DC.

US Water Resources Council, 1981. Guidelines for determining flood flow frequency. Bulletin 17B, Hydrology Committee, Washington, DC.

Wiltshire, S.E., 1986. Regional flood frequency analysis I: homogeneity statistics. Hydrol. Sci. J. 31, 321—333.