Water Resour Manage (2017) 31:2333–2348
DOI 10.1007/s11269-017-1653-2

Obtaining Homogeneous Regions by Determining the Generalized Fractal Dimensions of Validated Daily Rainfall Data Sets

M. T. Medina-Cobo[1] · A. P. García-Marín[1] · J. Estévez[1] · F. J. Jiménez-Hornero[2] · J. L. Ayuso-Muñoz[1]

Received: 20 April 2016 / Accepted: 29 March 2017 /
Published online: 5 April 2017
© Springer Science+Business Media Dordrecht 2017

Abstract Extreme rainfall data are widely used in several hydrological models and civil engineering design. Despite high temporal resolution rainfall data are not commonly available, daily rainfall data series are easily found. When these available data series are short in length the Regional Frequency Analysis (RFA) is a good tool to enlarge them by joining stations into homogeneous regions. This is by far, the most complicated step in RFA. This work presents a new method to form homogeneous regions of extreme annual daily rainfall data series. Daily rainfall data series from 53 weather stations in the Maule Region (Chile) have been used. Their fractal dimensions spectra have been obtained by applying the box counting method. Each station has been characterized by the fractal dimensions D1 and D2. A cluster analysis has been carried out based on these at-site characteristics and three regions have been obtained. After performing a RFA of extreme daily annual rainfall data series within each region they have shown as homogeneous. Only one of the available stations has not been possible to be included into any homogeneous regions, being the local frequency analysis the only suitable method to be applied at this location.

Keywords Regional frequency analysis · Homogeneous regions · Fractal dimensions

1 Introduction

The estimation of extreme events is a crucial problem in hydrology specially when dealing with rainfall or flood, due to the impact that these events can have on society and economy

✉ A. P. García-Marín
amanda.garcia@uco.es

[1] Department of Rural Engineering, Leonardo Da Vinci Building, University of Córdoba, 14071 Cordoba, Spain

[2] Department of Graphic Engineering and Geomatics, University of Cordoba, Gregor Mendel Building (3rd floor), Campus Rabanales, 14071 Cordoba, Spain

(e.g. Shabri et al. 2011). A good solution for many hydrologic engineering problems is based on the proper knowledge of extreme rainfall. For a certain place, rainfall intensity and its duration affect the maximum discharge that can be expected. Thus, information on the magnitude and frequencies of extreme rainfall is essential. Intensity-duration-frequency (IDF) relationships allow to compute the design storm which is the expected rainfall value for a given duration and a given occurrence probability (Di Baldassarre et al. 2006). When extreme rainfall data series for different durations are available, there are many IDF models that can be fitted to the rainfall quantiles. If the series are long enough, local frequency analysis techniques can be applied to obtain the quantiles. Since reliable estimations require very long station records that are not usually available, the regional frequency analysis appear as an alternative technique to provide a framework for hazard characterization of the extreme events (Norbiato et al. 2007).The RFA increases the data at the site of interest considering data from other places that share the same probability distribution functions. The RFA leads to more accurate quantile estimations than those from local frequency analysis (Lettenmaier and Potter 1985; Wallis and Wood 1985; Hosking and Wallis 1997) when working with rain (Hosking and Wallis 1997). The improvement of the RFA over the local one depends on the regional homogeneity, always considering that in cases of extreme regional heterogeneity, local estimations could be better than those based on RFA (Lettenmaier and Potter 1985).

The regional frequency analysis method introduced by Hosking and Wallis (1993, 1997) is widely used in rainfall studies over different climatic areas (Lee and Maeng 2005; Fowler and Kilsby 2003; Di Baldassarre et al. 2006; Norbiato et al. 2007; Wallis et al. 2007; Castellarin et al. 2009; Ngongondo et al. 2011; García-Marín et al. 2011, 2015a, b; Satyanarayana and Srinivas 2011; Malekinezhad and Zare-Garizi 2014; Liu et al. 2015).This method is based on lineal moments estimations (Hosking 1990, 1992) of the data series analyzed, and their values are used in all its steps (Hosking and Wallis 1993; Hosking and Wallis 1995; Hosking and Wallis 1997; Rao and Hamed 2000). Within all the steps on RFA the determination of homogeneous regions is the most difficult task and it conditions the final results.

Two important aspects have to be considered in order to finally obtain homogeneous regions: the grouping methodology used and the at-site characteristics to be considered in the joining process. Different methodologies have been applied in rainfall regionalization including spatial correlation analysis (Gadgil et al. 1993), principal component analysis (García-Marín et al. 2011), cluster analysis (Easterling 1989; Bonell and Sumner 1992; Venkatesh and Jose 2007), combination of principal component analysis and cluster analysis (e.g. Dinpashoh et al. 2004), and clustering combined with artificial neuronal networks (Jingyi and Hall 2004; Srinivas et al. 2008; Satyanarayana and Srinivas 2011), among others.

The more common characteristics that have been used in rainfall regionalization include climatological and geographical information, statistical values and location attributes (García-Marín et al. 2011) or even atmospheric variables (e.g. Satyanarayana and Srinivas 2011). Some multifractal parameters of rainfall data series have been recently used with this aim with very good results (García-Marín et al. 2015a; b). The multifractal character of rainfall has been widely studied from a descriptive use (e.g. de Lima and Grasman 1999) to any application in engineering models (García-Marín et al. 2013). Several methodologies exist to analyze the multifractal behavior of rainfall. All of them have in common that the multifractal paremeters are independent of the available data for the different scales, and that no probability distribution function has to be assumed for the data set. The multifractal analysis based on the strange attractor formalism (e.g. Hentschel and Procaccia 1983; Grassberger 1983; Halsey et al. 1986) deals with the fractal dimensions of a data set.

Thus, the objective of this work is to compound homogeneous regions of extreme annual daily rainfall by using the fractal dimensions of the daily rainfall data sets available in the Maule Region of Chile.

2 Materials and Methods

2.1 Rainfall Data

Daily precipitation data from 53 stations located in the Maule Region of Chile and supplied by the “Dirección General de Aguas”,DGA, were used to carry out this work. The geographical distribution of the stations throughout the Region ofMaule is shown in Fig. 1. Site elevations range from 10to 1058 m abovemeansealevel, longitude, from 70° 48′ 43”to72° 25′ 17″ W and latitude, from34° 54′ 41″ to36° 21′ 29″S (Table 1).

Maule Region is located in the semiarid region of Chile (from 34°41′ to 36°33′ S latitude), withannualaveragerainfallrangingbetween 600 and 2.300 mm. As central Chile, its physiography is characterized by the Andes mountains at the Eastside (withaltitudesbordering the 4.000 m.a.s.l.), followed by a central valleyof 40 km width, the Coast mountains (withheightsof 300 and 1000 m), and the coastalplain, whichreaches a widthof 5 km and is interrupted by the riversthatflowinto the Pacific Ocean. The Maule region is located in a transition area of Chile, from the semi-arid zone and the wetzone, showing a north-south gradient in annualrainfall. Besides, the orography lets an increase of precipitation from the coastto the Andes Mountains.

Validation procedures are part of the quality control systems and their purpose is to identify erroneous data from meteorological sensor measurements in order to make optimal use of them (Estévez et al. 2011a). In the validation process, data of a doubtful quality must be detected and appropriately flagged. Many methods exist to validate meteorological data (Feng et al. 2004; Zahumensky 2004; Kunkel et al. 2005; Estévez et al. 2011b).The available data for Maule

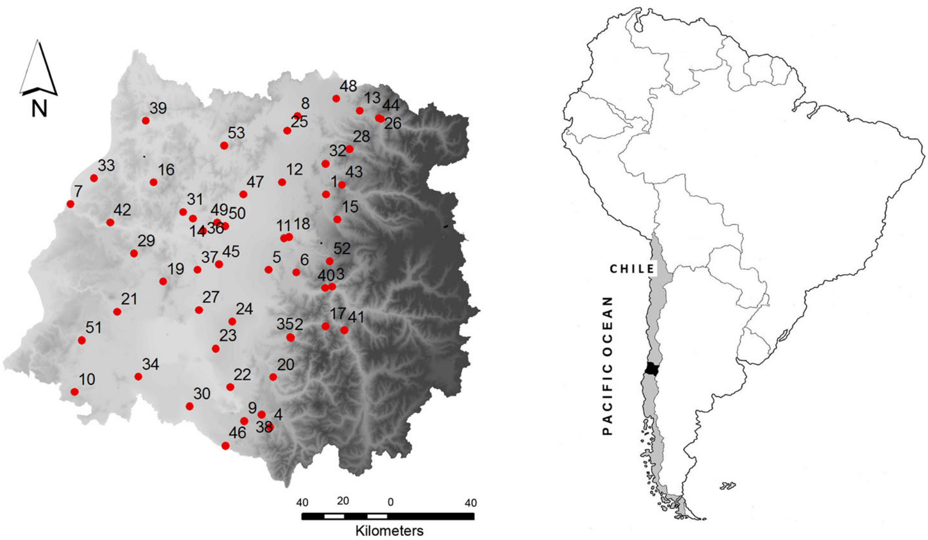

Fig. 1 Study Area: The Maule Region, Chile

Table 1 ID of the location, name, data time-period analysed, and coordinates of the weather stations used in this study (Maule Region, Chile)

ID	Name	Time period	Latitude (S) (S(N°)	Longitude (W)	Elevation (m)
1	Agua Fría	1993–2013	35° 18′ 47”	71° 05′ 54”	560
2	Ancoa Embalse	1957–2013	35° 54′ 38”	71° 17′ 45”	421
3	Armerillo	1948–2013	35° 42′ 04”	71° 04′ 38”	492
4	Bullileo Embalse	1930–2013	36° 17′ 06”	71° 24′ 51”	600
5	Colbún Maule Sur	1961–2013	35° 37′ 27”	71° 24′ 08”	280
6	Colorado	1963–2013	35° 38′ 17”	71° 15′ 38”	420
7	Constitución	1992–2013	35° 19′ 27”	72° 24′ 32”	10
8	Curicó	1971–2013	34° 58′ 52”	71° 14′ 10”	195
9	Digua Embalse	1956–2013	36° 15′ 21”	71° 32′ 53”	390
10	El Álamo	1994–2013	36° 06′ 46”	72° 25′ 17”	180
11	El Durazno	1992–2013	35° 29′ 33”	71° 19′ 06”	275
12	El Guindo	1964–2013	35° 15′ 28”	71° 19′ 26”	250
13	El Manzano	1976–2013	34° 57′ 48”	70° 55′ 04”	574
14	Fundo El Peral	1966–1986	35° 24′ 02”	71° 47′ 00”	110
15	Fundo El Radal	1992–2013	35° 25′ 08”	71° 02′ 35”	685
16	Gualleco	1961–2013	35° 14′ 38”	71° 58′ 48”	100
17	Hornillo	1962–2013	35° 52′ 02”	71° 07′ 02”	810
18	Huapi	1969–2013	35° 29′ 11”	71° 17′ 35”	250
19	Huerta Maule	1992–2013	35° 39′ 41”	71° 56′ 46”	218
20	Juan Amigo	1992–2013	36° 04′ 33”	71° 23′ 27”	460
21	La Estrella	1992–2013	35° 46′ 57”	72° 11′ 13”	200
22	La Sexta	1992–2013	36° 06′ 46”	71° 36′ 56”	229
23	Liguay	1975–2013	35° 56′ 52”	71° 41′ 03”	104
24	Linares	1979–2013	35° 50′ 17”	71° 35′ 43”	157
25	Lontue	1976–2013	35° 02′ 32”	71° 17′ 26”	199
26	Los Queñes	1931–2013	35° 00′ 03”	70° 48′ 43”	663
27	Melozal	1951–2013	35° 47′ 08”	71° 45′ 59”	96
28	Monte Oscuro	1994–2013	35° 07′ 27”	70° 58′ 29”	632
29	Nirivilo	1961–2013	35° 32′ 20”	72° 05′ 29”	200
30	Parral	1964–2013	36° 11′ 16”	71° 49′ 42”	175
31	Pencahue	1987–2013	35° 22′ 21”	71° 49′ 57”	55
32	Potrero Grande	1975–2013	35° 11′ 00”	71° 05′ 52”	445
33	Putú	1992–2013	35° 13′ 06”	72° 17′ 00”	36
34	Quella	1961–2013	36° 03′ 26”	72° 05′ 21”	130
35	Río Ancoa Morro	1999–2013	35° 54′ 31”	71° 17′ 53”	402
36	Río Claro Rauquén	1999–2013	35° 27′ 09”	71° 43′ 60”	64
37	Río Loncomilla	2001–2013	35° 37′ 01”	71° 46′ 04”	68
38	Río Longavi	2001–2013	36° 13′ 49”	71° 27′ 25”	449
39	Río Mataquito	2001–2013	34° 59′ 04”	72° 00′ 36”	20
40	Río Maule Armerillo	2001–2013	35° 42′ 22”	71° 06′ 50”	470
41	Río Maule Salto	2003–2013	35° 53′ 03”	71° 01′ 09”	730
42	Río Maule Forel	2001–2013	35° 24′ 25”	72° 12′ 30”	30
43	Río Palos	2001–2013	35° 16′ 28”	71° 00′ 56”	600
44	Río Teno	1999–2013	34° 59′ 46”	70° 49′ 14”	647
45	San Javier	1970–2013	35° 35′ 42”	71° 39′ 26”	135
46	San Manuel	1956–2013	36° 21′ 29”	71° 38′ 58”	270
47	San Rafael	1992–2013	35° 18′ 23”	71° 31′ 24”	152
48	Santa Susana	1985–2013	34° 54′ 41”	71° 02′ 07”	410
49	Talca	1964–1982	35° 25′ 10”	71° 39′ 38”	110
50	Talca UC	1982–2013	35° 26′ 09”	71° 37′ 11”	130
51	Tutuvén	1978–2013	35° 53′ 48”	72° 22′ 25”	179
52	Vilches Alto	1992–2013	35° 35′ 35”	71° 05′ 13”	1058
53	Villa Prat	1992–2013	35° 05′ 49”	71° 36′ 50”	90

Region were previously validated (García-Marín et al. 2015a). For this validation, Range (Estévez et al. 2011b) and Persistence (e.g. Hubbard et al. 2005) tests were applied.

2.2 Multifractal Analysis Based on the Strange Attractor Formalism

Multifractal formalisms find their origin in the theory of measures. Multifractal measures are related to the study of the distribution of a quantity over a geometric support (De Bartolo et al. 2000). The strange attractor (Hentschel and Procaccia 1983; Grassberger 1983; Halsey et al. 1986) formalism is used here to perform the multifractal analysis on daily rainfall data sets with the aim of obtaining homogeneous regions. This formalism deals with the fractal dimensions of the geometric sets associated with singularities of the measure.

The fractal dimension of a set is defined as the scaling exponent D_0

$$N(r) = \frac{A}{r^{d_0}} \quad (r \rightarrow \infty) \tag{1}$$

Where *N(r)* is the number of boxes of length or size *r*, that are necessary to cover the set, and *A* is a constant (Mandelbrot 1982; Feder 1988). Suppose the set is represented by a large number of points. If these points are uniformly distributed across the set, then the fractal dimension completely characterizes the dimension of the set. If the points are not distributed uniformly it is possible that the mass distribution of the points varies. Then, at a given box length *r*, it is possible to identify regions of the same masses μ (Feeny 2000). The mass can be estimated within a box of size *r* as$\mu_i = n_i/n$, where n_i is the number of points in the box, and n the total number of points. Then a measure can be constructed as follows,

$$M_d(q, r) = \sum_{i=1}^{N} \mu_i^q r^d \tag{2}$$

Where *N* is the number of boxes that cover the set; $d = \tau_q$ is called the mass exponent. Defining $Z(q, r) = \sum_{i=1}^{N} \mu_i^q$ as the partition function (i.e. Feder 1988), then $Z(q, r) \sim r^{-\tau_q}$and thus,

$$\tau_q = \lim_{r \rightarrow 0} \frac{\log Z(q, r)}{\log r} \tag{3}$$

τ_q can be obtained as the slope of the linear segment of a log-log plot of *Z*(*q*, *r*) versus *r*. For q> > 1, the value of *Z*(*q*, *r*) is mainly determined by the high data values, while the influence of the loB55w data values contributes most to the partition function for q < < −1 (Kravchenko et al. 1999).

The generalized fractal dimension, D_q, of moment order *q* is defined as,

$$D_q = \lim_{r \rightarrow 0} \frac{\log Z(q, r)}{(q-1)\log r} \tag{4}$$

In the limit as $q \rightarrow 1$Eq. 4 reduces to

$$D_1 = \lim_{r \rightarrow 0} \frac{\sum_{i=1}^{N} \mu_i \log \mu_i}{\log r} \tag{5}$$

Among the generalized fractal dimensions, D_0, D_1 and D_2 are frquently used to describe the measure. Thus, D_0 is the fractal dimension of the set over which the measure is carried out.D_1

is the information dimension that describes the degree of heterogeneity in the distribution of the measure. In addition, according to Davis et al. (1994), D_1 characterizes the distribution and intensity of singularities with respect to the mean. If D_1 becomes smaller, the distribution of the singularities will be sparse. On the contrary, if D_1 increases, the singularities will have lower values that exhibit a more uniform distribution. D_2 is the correlation fractal dimension, which is associated with the correlation function, and it determines the average distribution of the measure (Grassberger 1983; Grassberger and Procaccia 1983). D_q is a decreasing function with respect to q for a multifractally distributed measure (e.g., Saa et al. 2007) where $D_0 > D_1 > D_2$.

The relation between the spectrum of generalised fractal dimensions (Rényi spectrum), D_q, and multifractal spectrum, $f(\alpha)$, with α being the Lipschitz–Hölder exponent (that quantifies the strength of the measured singularities), is given through the sequence of mass exponents τ_q (Hentschel and Procaccia 1983) according to the expression:

$$\tau_q = (q-1)D_q \tag{6}$$

The multifractal or singularity spectrum $f(\alpha)$ can be obtained through (4) by means of the Legendre transform (Halsey et al. 1986) (Eq. 7). The spectrum is an inverted parabola for measures multifractally distributed. For monofractal measures, α value is identical for all the regions of the same size and $f(\alpha)$ consists of a single point (Kravchenko et al. 1999). Multifractal spectrum highest value, $f(\alpha_0)$, corresponds to the fractal dimension D_0 of the support of the measure.

$$\begin{aligned} \alpha_q &= -\frac{d\tau_q}{dq} \\ f\left(\alpha_q\right) &= q\alpha_q + \tau_q \end{aligned} \tag{7}$$

2.3 The Homogeneous Regions in Regional Frequency Analysis

The delimitation of homogeneous regions is usually the most difficult and important stage of the RFA (e.g Greis and Wood 1981; Hosking et al. 1985a, Lettenmaier and Potter 1985). If the available data cannot be joined into one homogeneous region or more, the RFA cannot be carried out. Several methodologies exist to group stations into potential homogeneous regions (e.g. Bonell and Sumner 1992; García-Marín et al. 2011; Jingyi and Hall 2004; Srinivas et al. 2008; Satyanarayana and Srinivas 2011; Yürekli and Modarres 2007) being cluster analysis of site characteristics the most practical one (Hosking and Wallis 1997). This technique has been widely used in hydrology (e.g. Burn 1989; Hall and Minns 1999; Lecce 2000; Jingyi and Hall 2004; Kyselý et al. 2007; Srinivas et al. 2008; Meshgi and Khalili 2009; Satyanarayana and Srinivas 2011; García-Marín et al. 2015a; b).

Once the potential regions have been determined, the quality of the following steps of RFA is conditioned by the degree of homogeneity found for the regions. In this work, the RFA proposed by Hosking and Wallis (1997) is followed. This methodology is based on L-Moments which are linear functions of the probability weighted moments (Greenwood et al. 1979) and were introduced by Hosking (1990, 1992). The L-Moments methodology includes from the probability distribution function characterization to the fitting of these functions to the data. For any distribution, the first four L-moments (λ_1, λ_2, λ_3, λ_4) and their ratios have to be obtained (Hosking 1990),

$$\tau = \frac{\lambda_2}{\lambda_1}\tau = \frac{\lambda_2}{\lambda_1}; \tau_3 = \frac{\lambda_3}{\lambda_2}\tau_3 = \frac{\lambda_3}{\lambda_2}; \text{ and } \tau_4 = \frac{\lambda_4}{\lambda_2}\tau_4 = \frac{\lambda_4}{\lambda_2} \tag{8}$$

Where τ, τ_3 and $\tau_4$are the L-coefficient of variation (*L-C$_v$*), L-coefficient of skewness (*L-C$_s$*) and L-coefficient of kurtosis (*L-C$_k$*), respectively. The first *L*-moment λ_1 is equal to the mean, hence it is a measure of location, and, τ, τ_3 and $\tau_4$are measures of a distribution's scale, skewness and kurtosis, respectively, which is analogous to the ordinary moments σ, γandκ, respectively (Hosking 1990).

For any step in RFA, *L*-Moments and their corresponding *L*-Moments ratios (L-C$_V$, *L-C$_s$* and *L-C$_k$*) have to be previously obtained for all the data series used in the analysis. Each data series will be considered as a site or station that could be potentially joined with other stations into a homogeneous region. The sample L-Moments ratios of a certain site is firstly considered as a point in a three-dimensional space. A group of sites will then yield a cloud of such points. Any point that is far from de centre of the cloud is considered as discordant. Mathematically, the discordance can be measured with the statistic *Di* (Hosking and Wallis 1993, 1997),

$$D_i = \frac{1}{3} N \left(u_i - \overline{u} \right)^T A^{-1} \left(u_i - \overline{u} \right) \tag{9}$$

being, $A = \sum_{i=1}^{N} (u_i - \overline{u})\ (u_i - \overline{u})$,$\overline{u} = N^{-1} \sum_{i=1}^{N} u_i, u_i = \left[LC_v^i, LC_s^i, LC_k^i \right]$ and N = the number of stations. Hosking and Wallis (1997) suggested some critical values for the discordancy test which are dependent on the number of sites in the study region. D_i is used to identify unusual sites in a potential region. If any discordant site is identified, it has to be removed from the region.

In order to asses if a proposed region is homogeneous, the heterogeneity measure *H*-statistic can be used. It is used to compare the between-site variation in sample *L*-moments for a group of sites with what would be expected for a homogeneous region (Hosking and Wallis 1997).There are three measures of the *H*-statistic, H_1, H_2, H_3, defined as

$$H_i = \frac{\left(V_{obs_i} - \mu_{v_i} \right)}{\sigma_{v_i}} \qquad i = 1, 2, 3 \tag{10}$$

Where μ_v and σ_v are the mean and standard deviation of the simulated values of *V* while V_{obs} is calculated from the regional data and is based on a corresponding *V*-statistic, defined as (Hosking and Wallis 1997)

$$\left.\begin{aligned} V_1 &= \left[\sum_{i=1}^{N} n_i \left(t^{(i)} - t^R \right)^2 / \sum_{i=1}^{N} n_i \right]^{1/2} \\ V_2 &= \sum_{i=1}^{N} n_i \left[\left(t^{(i)} - t^R \right)^2 + \left(t_3^{(i)} - t_3^R \right) \right]^{1/2} / \sum_{i=1}^{N} n_i \\ V_3 &= \sum_{i=1}^{N} n_i \left[\left(t_3^{(i)} - t_3^R \right)^2 + \left(t_4^{(i)} - t_4^R \right)^2 \right]^{1/2} / \sum_{i=1}^{N} n_i \end{aligned}\right\} \tag{11}$$

where V_1 is the standard deviation, weighted according to records length, of the at-site *L*-C$_V$s. V_2 and V$_3$ are the average distances from the site coordinates to the regional averages on a plot of *L*-C$_V$ versus *L*-skewness and a plot of *L*-skewness versus *L*-kurtosis, respectively; $t^{(i)}$, $t_3^{(i)}$and $t_4^{(i)}$ are the sample L-moment ratios at site *i;* t^R, t_3^R and t_4^R are the regional averages of the L-moment ratios; n_{i} is the record length at site *i*; and *N* is the number of sites in the region.The realization of at least 500 simulations let to obtain the mean and standard deviation values μ_{v_i} and σ_{v_i}.

The *H*-statistics (Eq. 10) indicate that the region under consideration is acceptably homogeneous when $H < 1$; possibly heterogeneous when $1 < H < 2$ and definitely heterogeneous when

$H > 2$. The statistic H_1, based on V_1 measurements, is the most decisive when discriminating between homogeneous or heterogeneous regions (Hosking and Wallis 1993; Castellarin et al. 2001), whereas H_2 has no power as a heterogeneity measurement (Viglione et al. 2007).

For any region that can be catalogued as homogeneous, the following steps of RFA can be successfully performed.

3 Results and Discussion

3.1 Aplication of Strange Attractor Formalism for Multifractal Analysis of Rainfall Data

No data from the stations were flagged by range test and very few with the persistence one (García-Marín et al. 2015a). After de validation proccess of the 53 available daily rainfall data series in the Maule region, the fractal behaviour was studied through their fractal dimensions. The strange attractor formalism was applied and the Rényi dimensions (Eqs. 4 and 5) obtained at each site. Figure 2 shows the generalized fractal dimensions Dq for q values from −10 to 10 for a selection of 4 sites: Agua Fría, Bullileo, Liguay and San Rafael. The expected decreasing behaviour of Dq function can be observed for all the stations with a strong dependence of Dq on the values of q, confirming the multifractal nature of the series analysed. The values of D_0 are 1 for all the sites which shows a full fill in the entire 1D domain. For lower and higher q values, different values of D_q are obtained, with $D_0 > D_1 > D_2$. For q values lower than 0 the highest D_q values are obtained for Agua Fria station, followed by San Rafael and Liguay, being the lowest values those from Bullileo Station. An opposite behaviour is found for q values higher than 0, being Bullileos' D_q values the highest ones, followed by Liguay and San Rafael, being Agua Fria's the lowest D_q. Table 2 shows the values of D_1 and D_2 for all the data series analysed. The information dimension D_1 provides a measure of the degree of heterogeneity (Davis et al. 1994) and characterize the distribution and intensity of singularities with respect to the mean (Ariza-Villaverde et al. 2013).The lowest value of D_1 is the one obtained for Rio Maule Salto (0.877382) showing a more sparse distribution of singularities than in the rest data series, for which greater D_1 values were obtained (Table 2) and a more homogeneous distribution of singularities can be expected. The highest D_1 value is the one of La Sexta station (0.984788). Correlation dimension values (D_2) are also shown in Table 2 with the lowest and highest values obtained for the same stations as D_1, being 0.793721 and 0.969889 for Rio Maule Salto and La Sexta, respectively. The correlation dimension describes the probability of finding data belonging to the set within a given distance when starting on a data belonging to the set (Ariza-Villaverde et al. 2013).

Once the multiscaling of the rainfall data series were detected and analysed from the D_q function, the multifractal spectrums $f(\alpha)$ (Eq. 7) were obtained for all of them. For the same stations of Fig. 2, the multifractal spectrums are shown in Fig. 3. For all of them, the spectrum show as an inverted parabola. The singularity spectrum quantifies in details the long- range correlation properties of a series. It gives information about the relative importance of various fractal exponents present in the series. It is a measure of how wide the range of fractal exponents found in the signal is and, thus, it measures the multifractality degree (MD) of the series (Telesca and Lovallo 2011). Greater the value of the width, greater will be the multifractality of the spectrum. For a monofractal set, the width will be zero (Maity et al. 2015). According to Telesca et al. (2004) the width of the spectrum can be obtained from the D_q function, being the difference between D-5 and D5 values. The higher the *MD* value the larger the heterogeneity. For all the stations, Table 2 shows the values of the *MD* for all the stations, with the minimum and maximum values of 0.137865 for La Sexta station and 1.481674 for Rio Ancoa station, respectively. If we focus on the sites shown in Fig. 2, the highest multifractal degree is found for Agua Fria (0.920359), followed by San Rafael (0.822637) and Liguay (0.632138). The lowest multifractal degree corresponds to Bullileo (0.468812).Some information can also be obtained from the shape of the multifractal spectra (Fig. 3) (Serrano et al. 2013). Rounder and wider spectra correspond to higher variability in the distribution of the values. Agua Fria and Bullileo stations' spectra are different in shape, being Agua Fria's rounder and wider than Bullileo's. The different behaviour between rainfall data series for both stations were previously detected by García-Marín et al. (2015a) being related to the percentage of no-rain days and with the presence of rare and extreme events in the time series.

3.2 RFA: Looking for Homogeneous Regions

As the objective is to test if the available stations in the Maule region can be grouped into regions according to the extreme annual daily rainfall, 53 extreme daily annual rainfall data series were obtained from the validated daily rainfall data series. Each site was characterized by its *L*-moments values and ratios ($L\text{-}C_v$, $L\text{-}C_s$ and $L\text{-}C_k$) (Table 3). With all the *L*-moments data from Table 3 a region called Maule was firstly tested and a RFA of extreme annual daily rainfall was performed. Considering Hosking and Wallis' (1997) criteria for regions with more than 15 sites, all the stations showing D_i values (Eq. 9) higher than 3.00 had to be removed

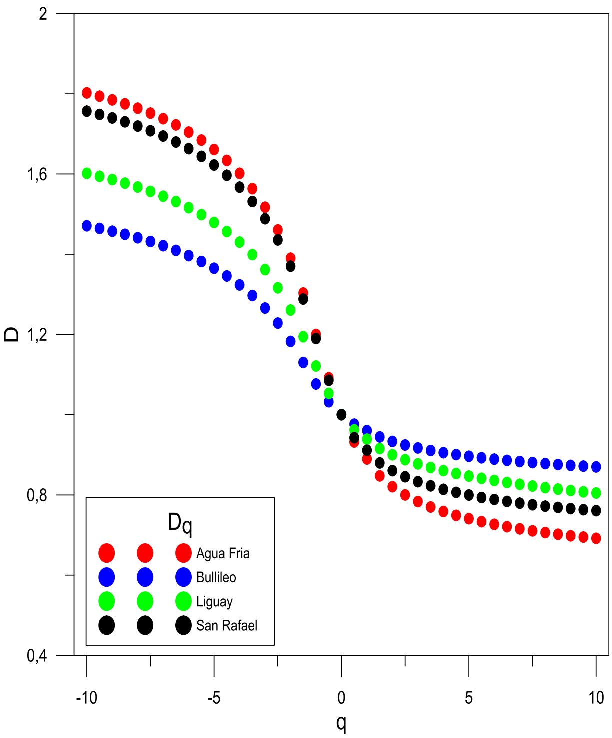

Fig. 2 The generalized fractal dimension function D_q for the stations Agua Fria, Bullileo, Liguay and San Rafael

Table 2 Values of fractal dimensions D_1 and D_2, and the multifractal degree (MD) for the available sites (ID) in the Maule Region

ID	D_1	D_2	MD	ID	D_1	D_2	MD
1	0.889615	0.820564	0.920359	28	0.937466	0.893687	0.584380
2	0.955456	0.925041	0.507552	29	0.974909	0.955247	0.299732
3	0.942059	0.905053	0.725544	30	0.955529	0.925813	0.477947
4	0.960375	0.933376	0.468767	31	0.956797	0.923613	0.413641
5	0.974106	0.954068	0.272129	32	0.953460	0.914992	0.377158
6	0.980991	0.965033	0.204177	33	0.951121	0.911541	0.416012
7	0.879698	0.802271	1.224827	34	0.978045	0.959457	0.236557
8	0.959815	0.926981	0.336645	35	0.883600	0.822819	1.481674
9	0.958850	0.930099	0.392157	36	0.919515	0.872760	0.994822
10	0.919259	0.864137	0.574581	37	0.913781	0.857483	0.899522
11	0.951433	0.913196	0.451888	38	0.918115	0.860271	0.833337
12	0.973490	0.953580	0.321003	39	0.892489	0.823538	1.143640
13	0.920847	0.873019	0.834535	40	0.909142	0.859638	0.951834
14	0.950020	0.901733	0.390810	41	0.945099	0.901751	0.500260
15	0.954170	0.919747	0.398820	42	0.877382	0.793721	1.108003
16	0.982265	0.966985	0.197977	43	0.918266	0.867041	0.823724
17	0.972172	0.948569	0.263800	44	0.955153	0.918287	0.437054
18	0.970621	0.945680	0.359593	45	0.961036	0.931249	0.376100
19	0.966911	0.942063	0.345346	46	0.980835	0.964870	0.199753
20	0.978528	0.962984	0.290351	47	0.911176	0.860429	0.822637
21	0.916204	0.864462	0.862946	48	0.977476	0.960844	0.279331
22	0.984788	0.969899	0.137865	49	0.958848	0.925689	0.430586
23	0.938829	0.900063	0.632138	50	0.929421	0.883229	0.554809
24	0.938470	0.901135	0.608994	51	0.930459	0.890896	0.767668
25	0.921801	0.874771	0.864936	52	0.899884	0.833084	0.759191
26	0.966235	0.941503	0.396674	53	0.899037	0.840771	0.958326
27	0.972364	0.949774	0.317954				

from the region. Thus, five stations showed discordance and were eliminated: Agua Fria ($D_i = 3.35$), Quella ($D_i = 3.01$), Río Ancoa ($D_i = 3.96$), Río Loncomilla ($D_i = 5.04$) and San Rafael ($D_i = 4.66$). The values of the H-statistic (Eq. 10) for the region (now composed by 48 sites) were 2.35, 1.17 and 1.38, for H_1, H_2 and H_3, respectively (Table 4). H_1, the most restrictive heterogeneity measurement, shows the heterogeneity of the Maule Region and new groups of stations (Sub-regions) had to be composed.

Since different values of the Rénji spectrum were obtained at each site, these differences were used as the basis of the joining criteria. Thus, with the D_1 and D_2 values as the at-site vector characterization (Table 2) a cluster analysis was performed and two sub regions were obtained, composed by 19 and 34 sites, respectively. For the first sub regions, only one station showed as discordant and was removed (San Rafael station, $D_i = 3.45$). The H-statistic values for the group were 0.33, −0.34, and 0.44, for H_1, H_2 and H_3, respectively (Table 4). The group with 34 sites showed as possibly heterogeneous, with a final value of H_1 of 1.89 after removing the discordant stations (Liguay, Los Queñes, Quella and Talca) from the analysis.

A new cluster analysis was then performed with all the stations that were not included in the first homogeneous region (Region 1 in Table 4). Two groups were obtained, with 20 stations and 15 stations respectively. The stations Liguay ($D_i = 3.23$), San Rafael ($D_i = 3.55$) and Talca ($D_i = 3.03$), showed as discordant in the first new group. The rest of stations (17) behaved as a homogeneous region, with H-statistic values of 0.76, −0.33, and 0.60, for H_1, H_2 and H_3, respectively (Region 2 in Table 4).The sub-region with the 15 stations had no discordant sites and the value of H_1 was 1.00. This last value is the lowest value to classify a region as a

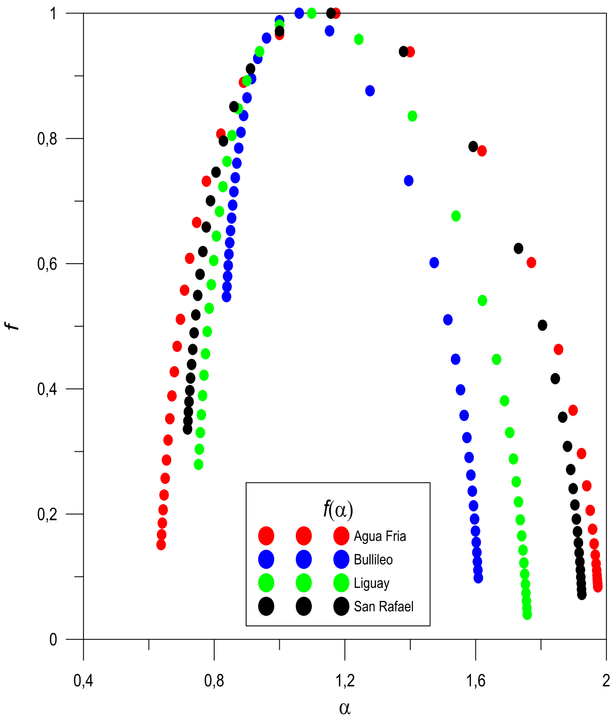

Fig. 3 The multifractal spectrum *f(α)* for the stations Agua Fria, Bullileo, Liguay and San Rafael

possibly heterogeneous. Since some spare stations were available (those removed from Region 2 for being discordant), a new group or region was formed by adding them to the sub-region with the 15 stations. Thus, an 18-site region was available and its homogeneity was tested. Only one station was discordant (San Rafael station, D_i = 3.18), but the 17 stations left, behaved as an homogeneous region (Region 3 in Table 4) with *H*-statistic values of 0.80, 0.46, and 1.17, for H_1, H_2 and H_3, respectively.

Figure 4 shows the three homogeneous regions obtained by colouring each station with the reference colour of the region: red for Region 1, blue for Region 2, and green for Region 3. Only one station (San Rafael) stays in black in Fig. 4 because it was not possible to include it in any of the three homogeneous regions detected (Table 4). For this station, only the local frequency analysis of extreme annual daily rainfall is then possible. Moreover, if the results that the authors present in this work are compared to those in García-Marín et al. (2015a), the latter let five spare sites that could not be included into any homogeneous region. This fact shows a clear improvement in the process of forming homogeneous regions. Thus the methodology presented in this work is the easiest and most direct when looking for homogeneous regions.

4 Summary and Conclusions

This paper presents a new methodology for grouping stations into regions when performing a RFA of extreme daily annual rainfall data. According to the results, grouping daily rainfall data series into homogeneous regions using the generalized fractal dimensions (also known as Rénji spectrum) of daily rainfall data is a useful method. The novelty of this work is that only

Table 3 L-Moment Ratios for the 53 stations

ID	L-Cv	L-Cs	L-Ck
1	0.17021	0.02470	0.02980
2	0.19991	0.16800	0.18040
3	0.17742	0.03650	0.07030
4	0.16041	0.06240	0.10440
5	0.20386	0.24130	0.22860
6	0.15524	0.09070	0.20070
7	0.16360	0.02740	0.11080
8	0.18376	0.10180	0.14270
9	0.16922	0.16470	0.15940
10	0.23922	0.24720	0.11410
11	0.13473	0.09070	0.05050
12	0.17035	0.21500	0.15470
13	0.18315	0.15410	0.20060
14	0.22194	0.17490	0.06870
15	0.19496	0.12300	−0.00780
16	0.21462	0.21770	0.16540
17	0.22309	0.16160	0.12500
18	0.17780	0.24470	0.14250
19	0.14048	0.03500	0.23520
20	0.16596	0.15820	0.18680
21	0.12567	0.02960	0.11540
22	0.22131	0.27010	0.10090
23	0.16202	0.38420	0.33930
24	0.15056	0.11300	0.08560
25	0.19731	0.21080	0.21990
26	0.23472	0.25290	0.28920
27	0.21118	0.23650	0.17940
28	0.17386	0.06220	0.09350
29	0.22222	0.24950	0.12030
30	0.19221	0.18550	0.11790
31	0.14389	0.14170	0.16590
32	0.15280	0.05750	0.11810
33	0.19044	0.23040	0.11650
34	0.18842	0.36920	0.42320
35	0.18862	−0.00410	0.23530
36	0.13467	0.09730	0.28580
37	0.12764	−0.08870	0.31450
38	0.20819	0.16120	0.00810
39	0.15165	0.01650	−0.03490
40	0.21089	0.29450	0.20090
41	0.19334	−0.01480	0.01010
42	0.19528	0.02530	−0.00380
43	0.18205	0.01940	−0.07320
44	0.20574	0.07030	0.05610
45	0.16841	0.11990	0.13870
46	0.16179	0.08170	0.06450
47	0.07668	0.07870	0.08400
48	0.18966	−0.01380	0.06250
49	0.14436	0.20050	0.32820
50	0.15721	0.07360	0.09140
51	0.16742	0.10020	0.12600
52	0.18563	0.15510	0.28900
53	0.23550	0.27310	0.28130

with two fractal dimensions, D_1 and D_2, from the Rénji spectrum, homogeneous regions in RFA are easily obtained.

Table 4 Results of the homogeneity tests for the regions formed by using cluster analysis

Region	Initial Sites	Discordant sites	H1	H2	H3
Maule	All	23,34,35,37,47	2.35	1.17	1.38
Region 1	1, 7, 10, 13, 21, 25, 35, 36, 37, 38, 39, 40, 42, 43, 47, 50, 51, 52, 53.	47	0.33	−0.34	0.44
Region 2	2, 3, 4, 8, 9, 11, 14, 15, 23, 24, 28, 30, 31, 32, 33, 41, 44, 45, 47, 49	23, 47, 49	0.76	−0.33	−0.60
Region3	5, 6, 12, 16, 17, 18, 19, 20, 22, 23, 26, 27, 29, 34, 46, 47, 48, 49	47	0.80	0.46	1.17

The Regional Frequency Analysis methodology used in this work was the one proposed by Hosking and Wallis (1997). This method is widely used in hydrology and is based on the L-moments of the data series analysed. Thus, the main *L*-moments and *L*-moments ratios of extreme annual daily rainfall data series from 53 sites in the Maule region (Chile) were obtained. Considering each station characterized by it *L*-moments, a first RFA was made considering a region composed by the whole sites. This region showed heterogeneous and had to be divided into new sub-regions potentially homogeneous.

Cluster analysis was performed in order to divide the whole region into new sub-regions. For this purpose, each station was characterized by two fractal dimensions from the Rénji Spectrum. The spectrum was obtained by applying the box counting method to each daily rainfall data (previously validated). The differences between the Rénji spectrums indicated that some of their fractal dimensions could be used as site characteristics in the cluster analysis.

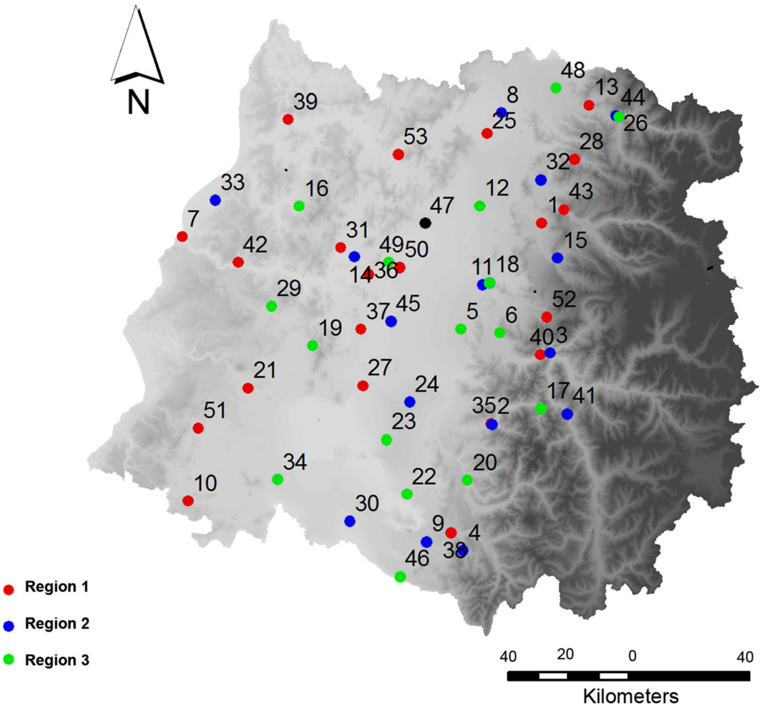

Fig. 4 The final regions obtained (*red*, *blue* and *green sites*) and the sparse station (*black*)

Two representative dimensions in fractal analysis, D_1 and D_2, were then used. D_1 characterizes the distribution of the rainfall data series and D_2 is related to the correlation function.

The fractal dimension-based cluster analysis led to form three fully homogeneous regions of 17, 18 and 17 stations respectively. Only one site stayed out of these homogeneous regions being the local frequency analysis the only option when dealing with its extreme annual daily rainfall data.

Acknowledgements We applied the 'sequence-determines-credit' (SDC) approach for the sequence of authors. The authors also want to thank the CTHA (Chile) and the DGA (Chile) for providing meteorological data to carry out this work. F.J. Jimenez-Hornero gratefully acknowledges the support from ERDF Project CGL2014-54615-C2-1-R (Spanish Ministry of Economy and Competitiveness).

References

Ariza-Villaverde AB, Jiménez-Hornero FJ, Gutiérrez de Ravé E (2013) Multifractal analysis applied to the study of the accuracy of DEM-based stream derivation. Geomorphology 197:85–95

Bonell M, Sumner G (1992) Autumn and winter daily precipitation areas in Wales, 1982–1983 to 1986–1987. Int J Climatol 12:77–102

Burn DH (1989) Cluster analysis as applied to regional flood frequency. J Water Resour Plan Manag 115:567–582

Castellarin A, Burn DH, BrathA (2001) Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. J Hydrol 241(3–4):270–285

Castellarin A, Merz R, Blöschl G (2009) Probabilistic envelope curves for extreme rainfall events. J Hydrol 378: 263–271

Davis A, Marshak A, Wiscombe W, Cahalan R (1994) Multifractal characterization of non stationarity and intermittency in geophysical fields: observed, retrieved or simulated. Journal of Geophysical Resources 99:8055–8072

De Bartolo SG, Gabriele S, Gaudio R (2000) Multifractal behavior of river networks. Hydrology and Earth Sciences 4(1):105–112

de Lima MIP, Grasman J (1999) Multifractal analysis of 15-min and daily rainfall from a semi-arid region in Portugal. J Hydrol 220:1–11

Di Baldassarre G, Brath A, Montanari A (2006) Reliability of different depth-duration-frequency equations for estimating short-duration design storms. Water Resour Res 42:W12501. doi:10.1029/2006WR004911

Dinpashoh Y, Fakheri-Fard A, Moghaddam M, Jahanbakhsh S, Mirnia M (2004) Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. J Hydrol 297:109–123

Easterling DR (1989) Regionalization of thunderstorm rainfall in the contiguous United States. Int J Climatol 9: 567–579

Estévez J, Gavilán P, García-Marín AP (2011a) Data validation procedures inagricultural meteorology. A prerequisite for their use. Adv Sci Res 6:141–146

Estévez J, Gavilán P, Giráldez JV (2011b) Guidelines on validation procedures for meteorological data from automatic weather stations. J Hydrol 402:144–154

Feder J (1988) Fractals. Plenum, New York

Feeny BF (2000) Fast multifractal analysis by recursive box covering. International Journal of Bifurcation and Chaos in Applied Sciences and Engineering 10(9):2277–2287

Feng S, Hu Q, Qian Q (2004) Quality control of daily meteorological data in China, 1951–2000: a new dataset. Int J Climatol 24:853–870

Fowler HJ, Kilsby CG (2003) A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. Int J Climatol 23:1313–1334

Gadgil S, Joshi Y, Joshi NV (1993) Coherent rainfall zones of the Indian region. Int J Climatol 13:547–566

García Marín AP, Estévez J, Medina Cobo MT, Ayuso Muñoz JL (2015b) Delimiting homogeneous regions using the multifractal properties of validated rainfall data series. J Hydrol 529:106–119

García-Marín AP, Ayuso-Muñoz JL, Taguas Ruiz EV, Estévez J (2011) Regional analysis of the annual maximum daily rainfall in the province of Málaga (southern Spain) using the principal component analysis. Water and Environmental Journal 25(4):522–531

García-Marín AP, Ayuso-Muñoz JL, Jiménez-Hornero FJ, Estévez J (2013) Selecting the best IDF model by using the multifractal approach. Hydrol Process 27:433–443

García-Marín AP, Estévez J, Sangüesa-Pool C, Pizarro-Tapia R, Ayuso-Muñoz JL, Jiménez-Hornero FJ (2015a) The use of the exponent *K(q)* function to delimit homogeneous regions in regional frequency analysis of extreme annual daily rainfall. Hydrol Process 29:139–151

Grassberger P (1983) Generalized dimensions of strange attractors. PhysicsLetter A 97:227–230
Grassberger P, Procaccia I (1983) Measuringthestrangeness of strangeattractors. Physica D 9:189–208
Greenwood J, Landwehr J, Matalas N, Wallis J (1979) Probability weighted moments: definition and relation to parameters of several distributions expressed in inverse form. Water Resour Res 15(6):1049–1054
Greis NP, Wood EF (1981) Regional flood frequency estimation and network design. Water Resour Res 17(4). doi:10.1029/WR017i004p01167
Hall MJ, Minns AW (1999) The classification of hydrological homogeneous regions. Journal of Hydrological Science 44(5):693–704
Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI (1986) Fractal measures and their singularities: the characterization of strange sets. Phys Rev A 33:1141–1151
Hentschel HGE, Procaccia I (1983) The infinite number of generalized dimensionsof fractals and strange attractors. Physica D: Nonlinear Phenomena 8:435–444
Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of Royal Statistical Society B 52:105–124
Hosking JRM (1992) Moments or L-moments? An example comparing two measures of distributional shape. The American Statiscian 46(3):186–189
Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. Water Resour Res 29(1): 271–281
Hosking JRM, Wallis JR (1995) Correction to “some statistics useful in regional frequency analysis”. Water Resour Res 31(1):251
Hosking JRM, Wallis JR (1997) Regional frequency analysis-an approach based on L-moments. Cambridge University Press, Cambridge
Hosking JRM, Wallis JR, Wood EF (1985a) An appraisal of the regional flood frequency procedure in the UK flood studies report. Hydrol Sci J 30:85–109
Hubbard KG, Goddard S, Sorensen WD, Wells N, Osugi TT (2005) Performance of qualityassuranceproceduresforanappliedclimateinformationsystem. J Atmos Ocean Technol 22:105–112
Jingyi Z, Hall MJ (2004) Regional flood frequency analysis for Gan-Ming river basin in China. J Hydrol 296:98–117
Kravchenko AN, Boast CW, Bullock DG (1999) Multifractal analysis of soil spatialvariability. Agron J 91:1033–1041
Kunkel KE, Easterling DR, Hubbard K, Redmond K, Andsager K, Kruc MC, Spinar ML (2005) Quality control of pre-1948 cooperative network observer data. Journal Atmospherie and Oceanic Technology 22:1691–1705
Kyselý J, Picek J, Huth R (2007) Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic. Stud Geophys Geod 51:327–344
Lecce SA (2000) Spatial variations in timing of annual floods in the southeastern United States. J Hydrol 235:151–169
Lee SH, Maeng SJ (2005) Estimation of drought rainfall using L-moment. Irrig Drain 54:279–294
Lettenmaier DP, Potter KW (1985) Testing flood frequency estimation methods using a regional flood generating model. Water Resour Res 21(12):190–1914
Liu J, Doan CD, Shie-Yui L, Sanders R, Dao AT, Fewtrell T (2015) Regional frequency analysis of extreme rainfall events in Jakarta. Nat Hazards 75:1075–1104
Maity AK, Pratihar R, Mitra A, Dey S, Agrawal V, Sanyal S, Banerjee A, Sengupta R, Ghosh D (2015) MultifractalDetrended fluctuation analysis of alpha and theta EEG rhythms with musical stimuli. Chaos, Solitons and Fractals 81:52–67
Malekinezhad H, Zare-Garizi A (2014) Regional frequency analysis of daily rainfall extremes using L-moments approach. Atmósfera 27(4):411–427
Mandelbrot BB (1982) The fractal geometry of nature. Freeman, New York
Meshgi A, Khalili D (2009) Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. I A re-visit to regional homogeneity Stochastic Environmental Research and Risk Assesment 23(1):119–135
Ngongondo CS, Xu CY, Tallaksen LM, Alemaw B, Chirwa T (2011) Regional frequency analysis of rainfall extremes in southern Malawi using the index rainfall and L-moments approaches. Stoch Env Res Risk A 25:939–955
Norbiato D, Borga M, Sangati M, Zanon F (2007) Regional frequency analysis of extreme precipitation in the eastern Italian alps and the august 29, 2003 flash flood. J Hydrol 345:149–166
Rao AR, Hamed KH (2000) Flood frequency analysis. CRC Press, Boca Raton
Saa A, Gasco G, Grau JB, Anton JM, Tarquis AM (2007) Comparison of gliding box and box counting methods in river network analysis. Nonlinear Process Geophys 14:603–613
Satyanarayana P, Srinivas VV (2011) Regionalization of precipitation in data sparse areas using large scale atmospheric variables - a fuzzy clustering approach. J Hydrol 405:462–473
Serrano S, Perán F, Jiménez-Hornero FJ, Gutiérrez de Ravé E (2013) Multifractal analysis application to the characterization of fatty infiltration in Iberianand white pork sirloins. Meat Sci 93:723–732
Shabri AB, Daud ZM, Ariff NM (2011) Regional analysis of annual maximum rainfall using TL-moments method. Theor Appl Climatol 104:561–570

Srinivas VV, Tripathi S, Rao AR, Govindaraju RS (2008) Regional flood frequency analysis by combining self-organizing feature maps and fuzzy clustering. J Hydrol 348:148–166

Telesca L, Lovallo M. (2011). Analysis of time dynamics in Wind records by means of multifractal detrended fluctuation analysis and Fisher-Shannon information plane. J Stat Mech: Theory Exp P07001. doi:10.1088/1742-5468/2011/07/P07001

Telesca L, Colangelo G, Lapenna V, Macchiato M (2004) Fluctuation dynamics in geoelectrical data: aninvestigationbyusingmultifractaldetrendedfluctuationanalysis. PhysicsLetters A 332:398–404

Venkatesh B, Jose M (2007) Identification of homogeneous rainfall regimes in parts of western Ghats region of Karnataka. Journal of Earth System Science 116:321–329

Viglione A, Laio F, Claps P (2007) A comparison of homogeneity tests for regional frequencyanalysis. Water Resour Res 43:W03428. doi:10.1029/2006WR005095

Wallis JR, Wood EF (1985) Grouping basins for regional flood frequency analysis. Hydrol Sci J 30:151–159

Wallis JR, Schaefer MG, Barker BL, Taylor GH (2007) Regional precipitation-frequency analysis and special mapping for 24-hour and 2-hour durations for Washington state. Hydrol Earth Syst Sci 11(1):415–442

Yürekli K, Modarres R (2007) Regionalization of maximum daily rainfall data over Tokat province, Turkey. Int J Nat Eng Sci 1(2):1–7

Zahumensky I. (2004). Guidelines on Quality Control Procedures for Data from Automatic Weather Stations.WMO-No. 955. Geneva