Regional frequency analysis of extreme precipitation in northeastern Algeria

Soufiane DAD[1) ABDEF], Tamara BENABDESSELAM[2) ACDE] ✉

[1)] Badji Mokhtar University, Soil and Hydraulics Laboratory, Annaba, Algeria; e-mail: dad.soufiane@yahoo.fr
[2)] Badji Mokhtar University, Department of Hydraulics, BP 12, 23000 Annaba, Algeria; e-mail: kondratievat@yahoo.fr

Abstract

The aim of the study is to improve the quality of estimating of the annual maximum daily precipitations of the northeastern area of Algeria. The regional frequency analysis based on L-moments was used. The investigated area is represented by 58 measuring stations. The main stages of the study were the definition of homogeneous regions and the identification of the regional distribution. It has been defined that the study region is homogeneous in terms of L-moments ratios despite the climatic differences within the region. Among the different tested distributions; the generalised extreme value (GEV) distribution has been identified as the most appropriate regional distribution for modelling precipitation in the region. The growth curve, derived from the regional distribution, was established. Therefore, to estimate the different return period's precipitation quantiles in a given site of the region, the mean precipitation of the site has to be multiplied by the corresponding regional quantile (growth factor). Comparison of the quantiles estimated from the regional and at-site frequency analysis showed that in the majority of stations (82.8%) at-site model underestimates the quantiles having high return periods.

Key words: *extreme daily precipitation, GEV distribution, homogeneous regions, L-moments, regional frequency analysis*

INTRODUCTION

Estimation of precipitation associated with of extreme events is a subject to interest in all areas related to water. The knowledge of precipitation quantiles of rare frequencies is necessary for the design of hydraulic structures such as flood protections, storm sewer networks and in many applications engineering. The estimation of these frequencies is difficult because extreme events are by definition rare and available at site data often come from short record lengths are insufficient to reliable estimate extreme quantiles. In order to solve this problem, we have recourse to regional analysis. The regionalization concept, introduced by DALRYMPLE [1960],"trading space for time" by using data from nearby or similar sites to estimate quantiles of the underlying variable at each site in the homogenous region of consideration. The quantiles estimated from the regional sample are considered to be more accurate [HOSKING, WALLIS 1997].

The frequency analysis methods were initially developed for flood estimation by DALRYMPLE [1960]. Since then these methods were continuously developed. GREHYS [1996] and OUARDA *et al*. [1999; 2008] in their studies on the regionalization of flood presented and compared different methods. The flood frequency analysis was applied to regionalization of precipitation which was then the basis for much research work. ALILA [1999] developed a hierarchical regional frequency model for precipitation of short duration in Canada. DJERBOUA [2001] and MORA *et al.* [2005] focused on the regional estimation of daily precipitation in France. NGUYEN *et al*. [2002] proposed two alternative methods for estimating extreme precipitation of various durations. KYSELY and PICEK [2007] used a method based on L-moments to estimate regional precipitation. Regional frequency analysis based on the index variable method and L-moments was utilized by NORBIATO *et al*. [2007] to analyse short duration annual maximum precipitation in Italia. GELLENS [2002] combined the regional

approach and data extension procedure for estimation of extreme precipitation in Belgium. GAAL *et al.* [2008] applied region-of-influence method to a frequency analysis of heavy precipitation in Slovakia.

In their literature review on regionalization of precipitation, ST-HILAIRE *et al.* [2003] pointed out that most regional analysis methods follow the steps of the determination of homogeneous hydrological regions, the identification of regional distribution and the estimation of parameters and quantiles of this distribution.

In this study, the regional frequency analysis based on the L-moments proposed by HOSKING and WALLIS [1997] is applied to estimate the quantiles of annual maximum daily precipitation at any available site in the study area. After describing the method used and the study area and data, the results of applying the steps of the regional frequency analysis will be presented and discussed in part entitled: Results and discussion; these steps are the screening of data, the formation of homogeneous regions, the identification of the regional frequency distribution and the estimation of the parameters and quantiles of the fitted distribution. A conclusion is finally made.

METHOD OF REGIONALIZATION

REGIONAL FREQUENCY ANALYSIS

The regional frequency analysis of extreme precipitation applied is based on L-moments and associated with the procedure of the "flood index" method [DALRYMPLE 1960] applied to hydrological data. The procedure used is a scale invariance procedure; the frequency distributions of the sites within a homogeneous region are identical except for a site-specific scale factor. Generally, the scale factor is the population average at the site [HOSKING, WALLIS 1997]. Therefore, quantiles of frequency F at site i of a homogeneous region of N sites can be determined as follows: $Q_i(F) = \mu_i q(F)$ where μ_i is the scale factor or the mean at the site i. The regional quantities $q(F)$ form the "regional growth curve" defined by a regional distribution of the reduced variable $y_{ij} = x_{ij}/\overline{x_i}$ where x_{ij} represent the annual maximum daily rainfall, $\overline{x}$ their mean at each site and $j = 1, 2, …, n_i$, n_i is the population of site i.

The parameters of the regional distribution are estimated from all the at-site statistics of the homogeneous region. To estimate this statistics we utilized the L-moments method.

L-MOMENTS

The L-moments theory was developed by HOSKING [1990]. L-moments are unbiased and less sensitive to outliers in data samples in comparison to the conventional moments. L-moments are analogous to conventional moments. Their estimation can be made from linear ordered data combinations.

For an ordered sample $x_1, x_2, …, x_n$ where $x_{1:n} < x_{2:n}, < … < x_{n:n}$ the probability weighted moments (PWM) are estimated by HOSKING and WALLIS [1997]:

$$\beta_0 = n^{-1} \sum_{j=1}^{n} x_{j:n}$$
$$\beta_1 = n^{-1} \sum_{j=2}^{n} \frac{j-1}{n-1} x_{j:n}$$
$$\beta_2 = n^{-1} \sum_{j=3}^{n} \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n}$$
$$\beta_3 = n^{-1} \sum_{j=4}^{n} \frac{(j-1)(j-2)(j-3)}{(n-1)(n-2)(n-3)} x_{j:n} \qquad (1)$$

Thus the following L-moments can be estimated using the PWM:

$$
\begin{aligned}
l_1 &= \beta_0 \\
l_2 &= 2\beta_1 - \beta_0 \\
l_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\
l_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0
\end{aligned} \qquad (2)
$$

The first L-moment (l_1) is equal to the mean of the distribution and l_2 is a scale parameter (L-standard deviation). In addition, the following L-moments ratios were introduced:

L-coefficient of variation, L-Cv, $t = l_2/l_1$ (3)

L-skewness, L-Cs, $t_3 = l_3/l_2$ (4)

L-kurtosis, L-Ck, $t_4 = l_4/l_2$ (5)

DISCORDANCY MEASURE TEST

The first step in regional frequency analysis is the screening of data. A discordance measure is used to identify those sites from a group of given sites that are grossly discordant with the group as a whole. The discordance measure is a single statistic based on the difference between the L-moment ratios of a site and the average L-moment ratios of a group of similar sites.

To determine the measure of discordance of a site of a region of N sites, one proceeds in the following way: let $u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$ be a vector containing the samples L-moments ratios $t^{(i)}, t_3^{(i)}, t_4^{(i)}$ for site i and T denotes transposition of a vector or matrix. The average of the u_i is defined as:

$$\overline{u} = N^{-1} \sum_{i=1}^{N} u_i \qquad (6)$$

and the matrix of sums of squares and cross products:

$$A = \sum_{i=1}^{N} (u_i - \overline{u})(u_i - \overline{u})^T \qquad (7)$$

The discordance measure for site i is defined as follows:

$$D_i = \frac{1}{3} N (u_i - \overline{u})^T A^{-1} (u_i - \overline{u}) \qquad (8)$$

The site i is declared to be discordant, if D_i is greater than the critical value of the discordance statistic D_i given

in tabular form by HOSKING and WALLIS [1997]. The magnitude of the critical value depends on the number of sites in the region. For example, for a region whose number of sites $N > 15$, the critical value is 3. The critical values suggested by HOSKING and WALLIS [1997] correspond to the significance level of 10%.

HOMOGENEITY TEST

To validate the homogeneity of a region (group of stations) in terms of the L-moment ratios, the statistic homogeneity test is used in which the representative parameters of a region are the weighted average L-moment statistics.

Thus, for a region of N site having each n_i length recording, the regional L-moments and the L-moment ratios are calculated as follows:

$$\overline{t}_r = \frac{\sum_{i=1}^{N} n_i t_r^{(i)}}{\sum_{i=1}^{N} n_i}; \quad \overline{l}_r = \frac{\sum_{i=1}^{N} n_i l_r^{(i)}}{\sum_{i=1}^{N} n_i} \quad (9)$$

Where: $t_r^{(i)}$, $l_r^{(i)}$ are the values of t_r and l_r at site i.

The hypothesized homogeneous region is then tested using a Monte Carlo simulation procedure. The four-parameter Kappa distribution is used for the simulations. The Kappa distribution is then fitted using the regional weighted average L-moments and L-moment ratios. Samples drawn from this parent are arranged to replicate the number of sites and the number of observations at each. A large number of generated regions are then replicated, and the following three measures of the between-site variability of sample L-moments are calculated for each time:

- the weighted variance of t:

$$V_1 = \frac{\sum_{i=1}^{N} n_i \left(t^{(i)} - \overline{t} \right)^2}{\sum_{i=1}^{N} n_i} \quad (10)$$

- the weighted average distance from the site to the region's mean on the t versus t_3 space:

$$V_2 = \frac{\sum_{i=1}^{N} n_i \sqrt{\left(t^{(i)} - \overline{t} \right)^2 + \left(t_3^{(i)} - \overline{t}_3 \right)^2}}{\sum_{i=1}^{N} n_i} \quad (11)$$

- the weighted average distance from the site to the region's mean on the t_3 versus t_4 space:

$$V_3 = \frac{\sum_{i=1}^{N} n_i \sqrt{\left(t_3^{(i)} - \overline{t}_3 \right)^2 + \left(t_4^{(i)} - \overline{t}_4 \right)^2}}{\sum_{i=1}^{N} n_i} \quad (12)$$

Where: $t^{(i)}$, $t_3^{(i)}$, $t_4^{(i)}$ denote respectively (Equations (3), (4), (5)) the L-Cv, L-Cs and L-Ck at site i; $\overline{t}, \overline{t}_3$, and $\overline{t}_4$ denote respectively the regional L-Cv, L-Cs and L-Ck calculated

according to equation (9). If V denotes any of three values V_1, V_2 and V_3, the homogeneity criterion of a region is given by:

$$H_V = \frac{V_{obs} - \mu_v}{\sigma_v} \quad (13)$$

Where V_{obs} is the observed value of either V_1, V_2 or V_3; μ_v and σ_v are the mean and the standard deviation of V obtained by simulations. The variable H enables to measure the dispersion of observations relatively to those of the simulations. According to HOSKING and WALLIS [1997], a region is acceptably homogeneous if $H < 1$, probably heterogeneous if $1 \leq H < 2$ and definitely heterogeneous if $H \geq 2$.

IDENTIFICATION OF REGIONAL FREQUENCY DISTRIBUTION

Among the different frequency distributions, the Gumbel distribution is the most often used in Algeria in the frequency analysis of extreme precipitation events in a single site. This distribution was used by MEBARKI [2005] in the frequency analysis of annual maximum daily rainfall in eastern Algeria. This two parameters distribution is also widely used in different climatic regions. The asymptotic behaviour of the Gumbel distribution is however challenged by KOUTSOYIANNIS [2004], confirming that its effect is to underestimate the precipitation values of high frequencies compared to the distribution GEV (EV2) (generalised extreme value, type 2). ALILA [1999] raised some concern on the use of Gumbel distribution in a regional context. In his study on the regionalization of short duration precipitation in Canada, different distributions have been adjusted and the GEV distribution was identified as the most appropriate regional distribution. This latter is the most widely used for both the precipitation regional frequency analysis and flood. OVEREEM *et al.* [2007] have used it for the regionalization of short duration precipitation in whole Holland. DJERBOUA [2001], VERSIANI *et al.* [1999] and CANNAROZZO *et al.* [1995] have chosen the TCEV distribution (two-component extreme value) as the regional statistical model of annual maximum daily precipitations. To determine growth curves regional precipitations of short duration SVEINSSON *et al.* [2002] used a regional approach based on the flood frequency index method taking into account the different distributions: lognormal distribution with three parameters (LN3), GEV, lognormal (LN) and Pearson type 3 (P3).

In the present study, the hypothesis of fitting the GEV, LN3, P3 and GLO (generalised logistic) distributions with the series of annual maximum daily precipitation of the study area is made. The suitability of fitting each of these three parameters distributions is evaluated by the difference between the theoretical L-kurtosis of the fitted distribution and the regional L-kurtosis. The significance of this difference is assessed through the Z statistic [HOSKING, WALLIS 1997]:

$$Z^{DIST} = \frac{\tau_4^{DIST} - \overline{t}_4 + \beta_4}{\sigma_{\overline{t}_4}} \quad (14)$$

Where $\overline{t_4}$ is the observed regional weighted average L-kurtosis, τ_4^{DIST} is the theoretical L-kurtosis of the distribution (DIST) estimated from the observed regional L-skewness; β_4 and $\sigma_{\overline{t4}}$ are respectively the bias and the standard deviation of $\overline{t_4}$ obtained by repeated simulations of a homogeneous region with the Kappa distribution as a parent. The statistic Z is based on asymptotic normality and the fit is declared satisfactory at the 90% level if $|Z| \leq 1.64$.

ESTIMATION OF PARAMETERS AND QUANTILES OF THE REGIONAL DISTRIBUTION

The parameters of the regional distribution are estimated from the first three regional L-moments. The regional growth curve will be established on the basis of the regional distribution parameters by applying the mean as a scaling factor. In this approach, the regional L-skewness and L-coefficient of variation are assumed to be constant. Therefore, to estimate the precipitation associated with different return periods at a given site of a homogeneous region, the values of the growth factor corresponding to the same return period will be multiplied by its mean daily maximum precipitation.

STUDY REGION AND DATA

The study area is located in northeastern Algeria and covers the watersheds of two major wadis Seybouse and Medjerda (Fig. 1). Concerning the climate, the northern part of the study area is characterized by the Mediterranean climate and the southern part is subject to the semi-arid climate [KHEZAZNA *et al.* 2017; MRAD *et al.* 2018].

The annual maximum daily precipitation sets of the 58 raingauge stations were selected for this study. Most observations concern the period 1970 to 2011. The mean sample size is 22 years. The period of observations for the different stations varies from 14 to 43 years.
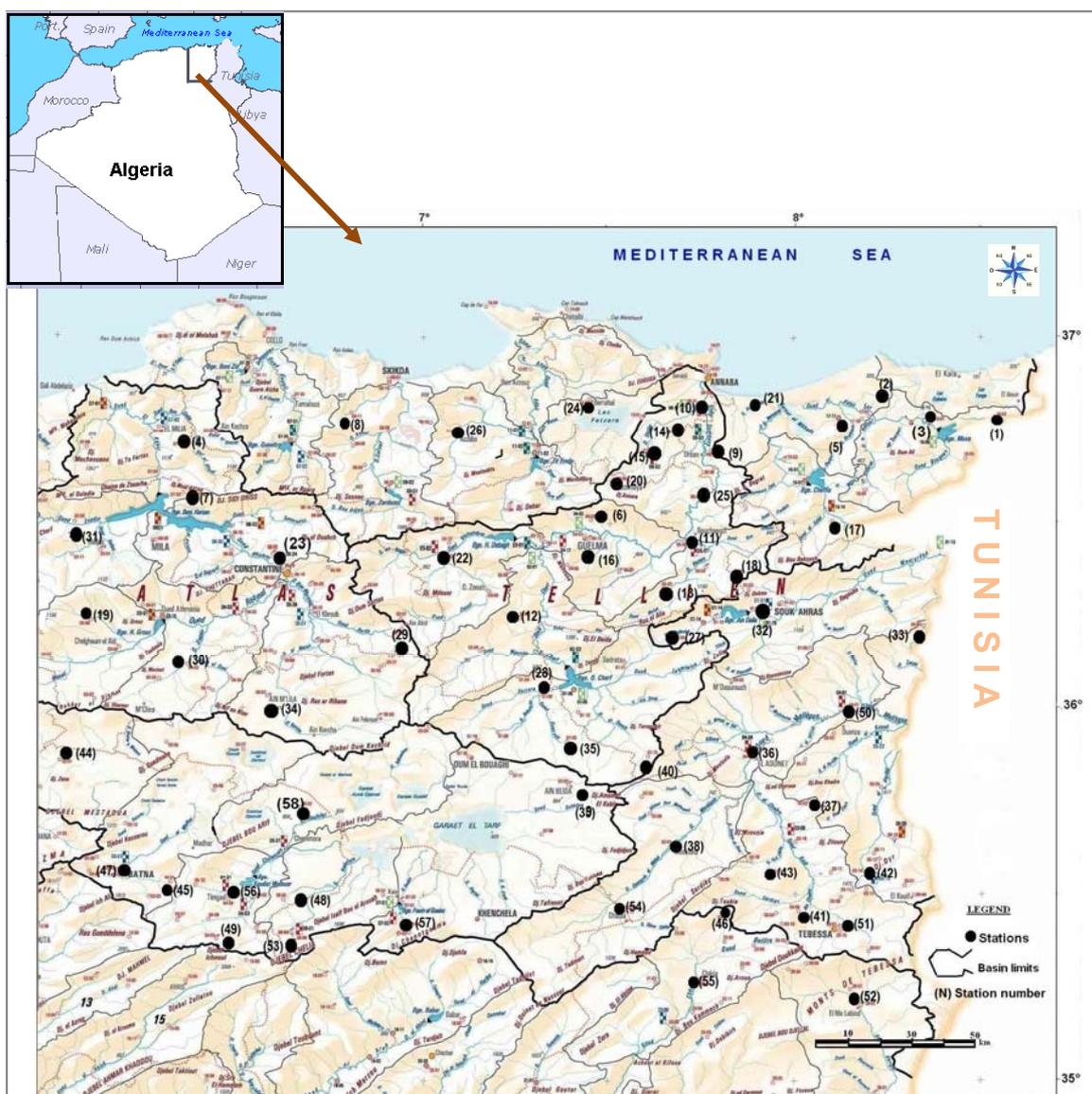
Fig. 1. Study region and location of stations; source: own elaboration

RESULTS AND DISCUSSION

In the first step of applying the regional frequency analysis method, data from 58 stations in the study region were checked in terms of discordance measure to identify stations whose statistical parameters differ markedly group. The values of discordance measure of the 58 stations range from 0.59 to 2.65. Consequently, the data of all stations can be used in regional frequency analysis.

To assess the homogeneity degree of region, 500 data regions were generated using the Kappa distribution. According to the obtained values of the heterogeneity measure H (Tab. 1), the region is homogeneous in terms of L-Cv, L-Cs and L-Ck. From the results of Table 1, negative values of H_{v2} and H_{v3} are found. This indicates that there is less dispersion among the at-site statistics than would be expected of a homogenous region with independent at-site frequency distributions. This is usually an indication of large cross-correlation between the sites' frequency distributions. Regional statistics are presented in Table 1.

The region being homogeneous, to identify the regional distribution among the GEV, P3, LN3 and GLO distributions, the Z^{DIST} statistics for these was calculated by distributions carrying out 500 simulations using the Kappa distribution. The values obtained for these latter and the theoretical L-kurtosis values of each fitted distribution are given in Table 2.

Table 1. Homogeneity test results and regional statistics

H_{V1}	H_{V2}	H_{V3}	$\bar{t}$	t_3	t_4
0.96	–0.66	–0.86	0.231	0.233	0.173

Source: own study.

Table 2. Theoretical L-kurtosis and Z statistic of different distributions

Distribution kind	τ_4^{DIST}	Z^{DIST}
GEV (generalised extreme value)	0.177	0.282
LN3 (three-parameter log-normal)	0.165	–0.585
P3 (Pearson type 3)	0.127	–3.197
GLO (generalised logistic)	0.211	2.630

Source: own study.

According to Z-statistics values, GEV and LN3 distributions are plausible adjustment of the regional sample. Moreover, the Z statistic value of the GEV distribution is lower than that of LN3. The quantile function of the GEV distribution is as follows:

$$x(F) = \xi + \frac{\alpha}{k}\left\{1 - \left[-\ln(F(x))\right]^k\right\} k \neq 0 \qquad (15)$$

The cumulative distribution function of the LN3 is given as follows:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^u \frac{1}{u}e^{\frac{\ln^2(u)}{2\sigma^2}}\,du \qquad (16)$$

Where: $u = (x - \xi)/\alpha$ and ξ, α and $\sigma(k)$ are the location, scale and shape parameters respectively.

For a sample, these parameters are defined from the L-moments by the following equations:
• for GEV distribution:

$$\xi = l_1 - \frac{\alpha}{k}\left[1 - \Gamma(1+k)\right] \qquad (17)$$

$$\alpha = \frac{l_2 k}{\left(1 - 2^{-k}\right)\Gamma(1+k)} \qquad (18)$$

$$k = 7.859\left(\frac{2}{3+t_3} - \frac{\ln 2}{\ln 3}\right) + 2.9554\left(\frac{2}{3+t_3} - \frac{\ln 2}{\ln 3}\right)^2 \qquad (19)$$

• for LN3 distribution:

$$\xi = l_1 - \frac{\alpha}{\sigma}\left(1 - e^{\sigma^2/2}\right) \qquad (20)$$

$$\alpha = \frac{l_2 \sigma \exp(-\sigma^2/2)}{erf(\sigma/2)} \qquad (21)$$

$$\sigma = -t_3 \frac{A_0 + A_1 t_3^2 + A_2 t_3^4 + A_3 t_3^6}{1 + B_1 t_3^2 + B_2 t_3^4 + B_3 t_3^6} \qquad (22)$$

Where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ represents the Gamma function; *erf* is the error function; A_0, A_1, A_2, A_3 and B_1, B_2, B_3 are the constants of approximation.

The regional parameters of these distributions (Tab. 3) were estimated from the regional L-moments.

Table 3. Regional the GEV and LN3 distributions parameters

Distribution kind	ξ	α	σ
GEV (generalised extreme value)	0.799	0.303	–0.096
LN3 (three-parameter log-normal)	0.905	0.371	–0.483

Source: own study.

The Z statistic test has therefore shown that there are two plausible regional distributions. To select the most appropriate distribution among these two distributions two additional tests were performed.

The first test is based on the comparison of L-Ck of the actual data with that of the distribution to be tested for each station in the region [LIN, VOGEL 2006]. For each distribution the root mean squared error (*RMSE*) is calculated:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} n_i \left(S_{i(L-Ck)} - D_{i(L-Ck)}\right)^2}{\sum_{i=1}^{N} n_i}} \qquad (23)$$

Where N is the number of stations; $S_{i(L-Ck)}$ and $D_{i(L-Ck)}$ are respectively the L-Ck of sample and distribution of the *i*th station, and n_i is the *i*th station size. The selection criterion is the least *RMSE*. According to this test, it is the GEV distribution that has the smallest *RMSE*. The *RMSE* of the latter is equal to 0.082 and that of LN3 distribution to 0.091. As a result, the L-Ck of all stations is closer to the theoretical value of GEV distribution than that of LN3.

In the second test the robustness of these two distributions is tested. The test is carried out by means of simulation and comprises the following steps:

a) selection of one of the two distributions as a population, for which the real parameters (Tab. 3) and the quantities q_T of specified return periods T are known;

b) generation, by Monte Carlo simulation, of the random variables of 58 samples related to the stations of the study region;

c) by applying the regional analysis method, quantiles $\hat{q}_T$ are estimated from the GEV distribution and the LN3 distribution for each generated sample; then, the relative biases are calculated:

$$bias(T) = \frac{q_T - \sum_{i=1}^{N} \hat{q}_{Ti} / N}{q_T} \qquad (24)$$

Where: N is the number of stations ($N = 58$).

Steps (b) and (c) are iterated 1000 times so as to obtain average values of bias for different sizes of the samples. The same procedure is applied with the second parent distribution. The results obtained are presented in Table 4.

Table 4. Bias values in quantile estimation from GEV and LN3 distributions

Return period years	Bias values for n sample size			
	$n = 42$		$n = 25$	
	GEV	LN3	GEV	LN3
Parent population GEV (generalised extreme value)				
2	–0.0019	–0.0029	–0.0039	–0.0042
5	–0.0004	–0.0013	–0.0019	–0.0017
10	0.0018	0.0010	0.0020	0.0015
20	0.0035	0.0034	0.0052	0.0053
50	0.0058	0.0060	0.0099	0.0105
100	0.0075	0.0084	0.0131	0.0147
200	0.0105	0.0116	0.0161	0.0192
500	0.0119	0.0144	0.0197	0.0250
1000	0.0138	0.0172	0.0228	0.0299
Parent population LN3 (three-parameter log-normal)				
2	–0.0024	–0.0018	–0.0041	–0.0023
5	–0.0005	–0.0008	–0.0013	–0.0006
10	0.0012	0.0019	0.0022	0.0024
20	0.0028	0.0030	0.0054	0.0059
50	0.0050	0.0057	0.0101	0.0113
100	0.0070	0.0083	0.0131	0.0155
200	0.0087	0.0109	0.0160	0.0197
500	0.0109	0.0138	0.0200	0.0256
1000	0.0123	0.0165	0.0301	0.0305

Source: own study.

From these results, it can be seen that the adjustment of the two laws on a sample of limited size leads to an underestimation of the quantiles, the values of the biases are positive. The underestimation is all the more important than the sample is small. From the point of view of L-moment estimation, the GEV distribution is more robust than LN3 (lower bias values). The difference is rather small, the same trend is observed in the results obtained by the two preceding tests.

Based on the results of the three tests, the GEV distribution was selected as the most robust for estimating the extreme precipitation quantiles of the study area. The regional growth curve, derived from the regional distribution GEV was plotted for specified return periods T (Fig. 2). This growth curve reflects the variation of regional quantile (growth factor) $q(F)$ versus the not exceeded probability F or versus the return period T ($T = 1/(1 - F)$).
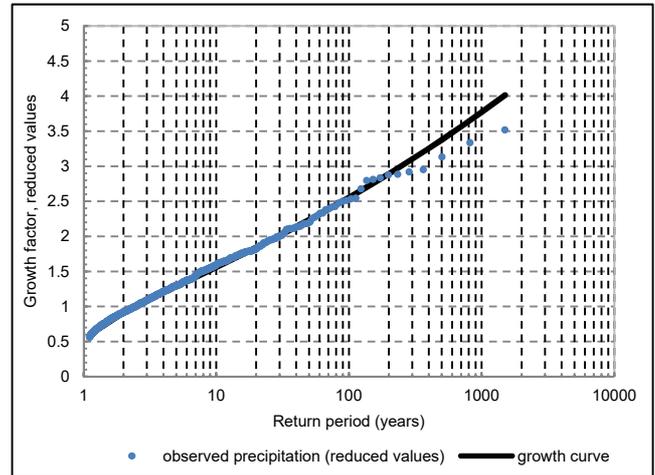
Fig. 2. Regional growth curve and empirical frequency distribution of observed precipitation; source: own study

It is observed that the asymptotic behaviour of the regional distribution GEV is quasi exponential since its parameter of form is close to zero ($k = –0.096$).

Also, in Figure 2, the empirical frequency distribution of observed precipitation (reduced values) was presented. The empirical frequencies were calculated according to the formula of Cunnane (($i – 0.4)/(N – 0.2$) where N denotes the regional sample size and i is the ranking rank. We see a good fit of the growth curve to the precipitation distribution, in particular for return periods $T < 500$ years.

To determine the accuracy and quantify the uncertainties of estimation of the regional quantiles, the Monte Carlo simulation algorithm was used. This algorithm is based on the application of the bootstrap procedure to calculate the confidence interval of the regional quantiles of specified return period (q_T). To achieve correlation between sites, a correlation matrix was used with a correlation coefficient of 0.65. By decomposing the correlation matrix (decomposition of Cholesky) we obtain the triangular matrix which is used to generate a matrix of Nn_{max} uniformly correlated random variables (N is the number of stations and n_{max} is the maximum size of a site in the region). The 10000 (R) realizations of the region have been made. For each realization, the specified regional quantities $\hat{q}_T$ have been estimated at from the adjustment of the samples generated to the GEV law. Thus we constitute a set of the simulated quantiles $q = \{\hat{q}_{T,1}, …, \hat{q}_{T,R}\}$. In the sorted series, the empirical quantities $q_{T,\alpha/2}$ and $q_{T,1-\alpha/2}$ have been determined for the not exceeding probability $\alpha/2$ and $1 – \alpha/2$ respectively. We thus obtain the confidence interval of q_T at level $1 – \alpha$: $[q_{T,\alpha/2}; \; q_{T,1-\alpha/2}]$.

The relative regional *RMSE* values and the confidence interval (90%) of the estimated growth curve for the different return periods are shown in Table 5. The low *RMSE* values demonstrate the reliability of the applied regional frequency analysis. The confidence interval (90%) obtained is narrow because of the quasi-exponential regional distribution.

Table 5. Relative root mean square error (*RMSE*) and confidence interval (90%) for estimated regional growth curve (q_T)

Return period *T* years	q_T	*RMSE*	Confidence interval (90%)	
			low	upper
2	0.912	0.028	0.752	1.121
10	1.560	0.018	1.431	1.668
50	2.233	0.029	1.952	2.455
100	2.552	0.035	2.208	2.845
200	2.891	0.042	2.489	3.221
1 000	3.771	0.061	3.151	4.296

Source: own study.

To see the effectiveness of regional frequency analysis, quantile values estimated from regional and at-site analysis were compared. In at-site analysis, the GEV distribution is used. The Chi-square test was used to test the appropriateness of the GEV distribution. This test showed that this distribution can be adopted with a significance level of 5% for all data sets, except those five stations where only 1% is accepted. In Figure 3 are presented the centennial quantile values estimated from the regional and local frequency analysis. Comparison of these values shows that in the majority of stations (82.8%) local model underestimates the quantiles having high return periods. The values of the centennial quantiles estimated from the regional model decrease from the North to the South of the study region. For stations located in the northern part characterized by a Mediterranean climate (numbered from 1 to 33 in Figure 3) these values range from 181 to 102 mm.

For stations in the southern part characterized by a semi-arid climate (numbered from 34 to 58 in Figure 3), they range from 113 to 68 mm.

To verify the coherence of the results obtained with the homogeneity hypothesis and the GEV regional law, the GEV distribution was plotted on the empirical frequency distribution of the observed data at the three stations of different sizes (Fig. 4). Note the good fit of the regional distribution GEV of precipitation recorded at the three stations in particular for the high return periods. Therefore, the results obtained are consistent with the hypothesis of homogeneity according to which the precipitation distributions follow at all points of a homogeneous region the same law. For the three stations the local model underestimates the quantiles. The difference between the quantiles estimated from regional and at-site analysis appears from the 20-year return period.

To evaluate the performance of the regional model, the relative bias and relative root mean square error (*RMSE*) were calculated for all stations as follows:

$$bias(T)(\%) = \frac{100}{N} \sum_{i=1}^{N} \left(\frac{Q_i^R - Q_i^L}{Q_i^L} \right) \quad (25)$$

$$RMSE(T)(\%) = 100 \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{Q_i^R - Q_i^L}{Q_i^L} \right)^2} \quad (26)$$

Where: *N* is the number of stations; Q_i^R and Q_i^L are the quantiles of return period *T* estimated from the regional and at-site analysis in site *i*, respectively. The variations of the bias and *RMSE* versus the return period are shown in Figure 5. The bias and *RMSE* of quantiles are low for return periods less than 10 years (Fig. 5). Their values do not exceed 0.38 and 7.27% respectively. Beyond this level, they increase progressively to reach maximum values respectively 23.1 and 38.6 % for the 1000 years return period.
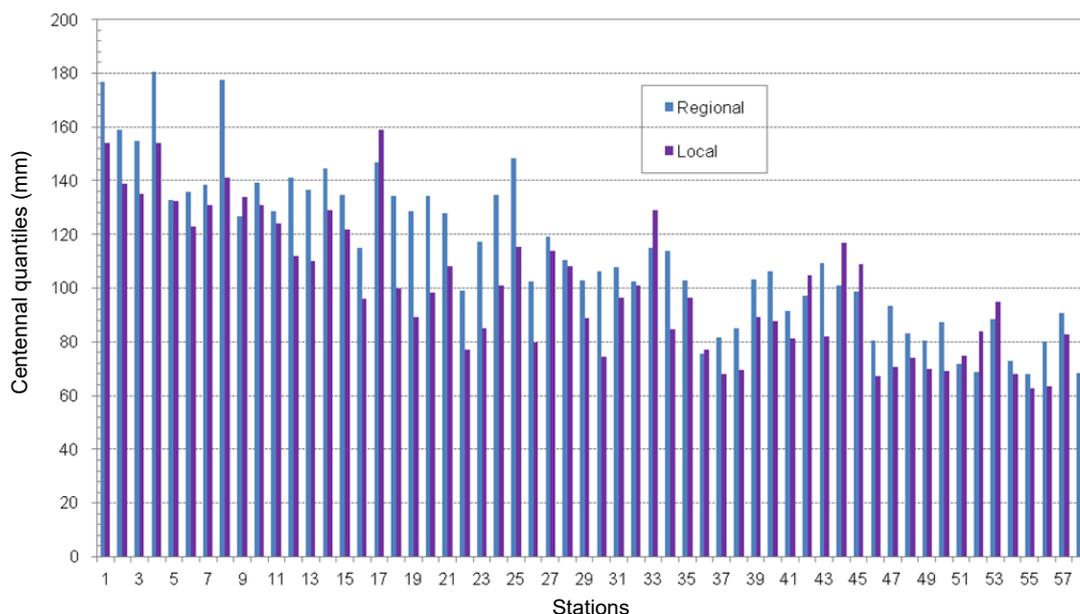
Fig. 3. Comparison of estimated centennial quantiles from regional and local frequency analysis; source: own study
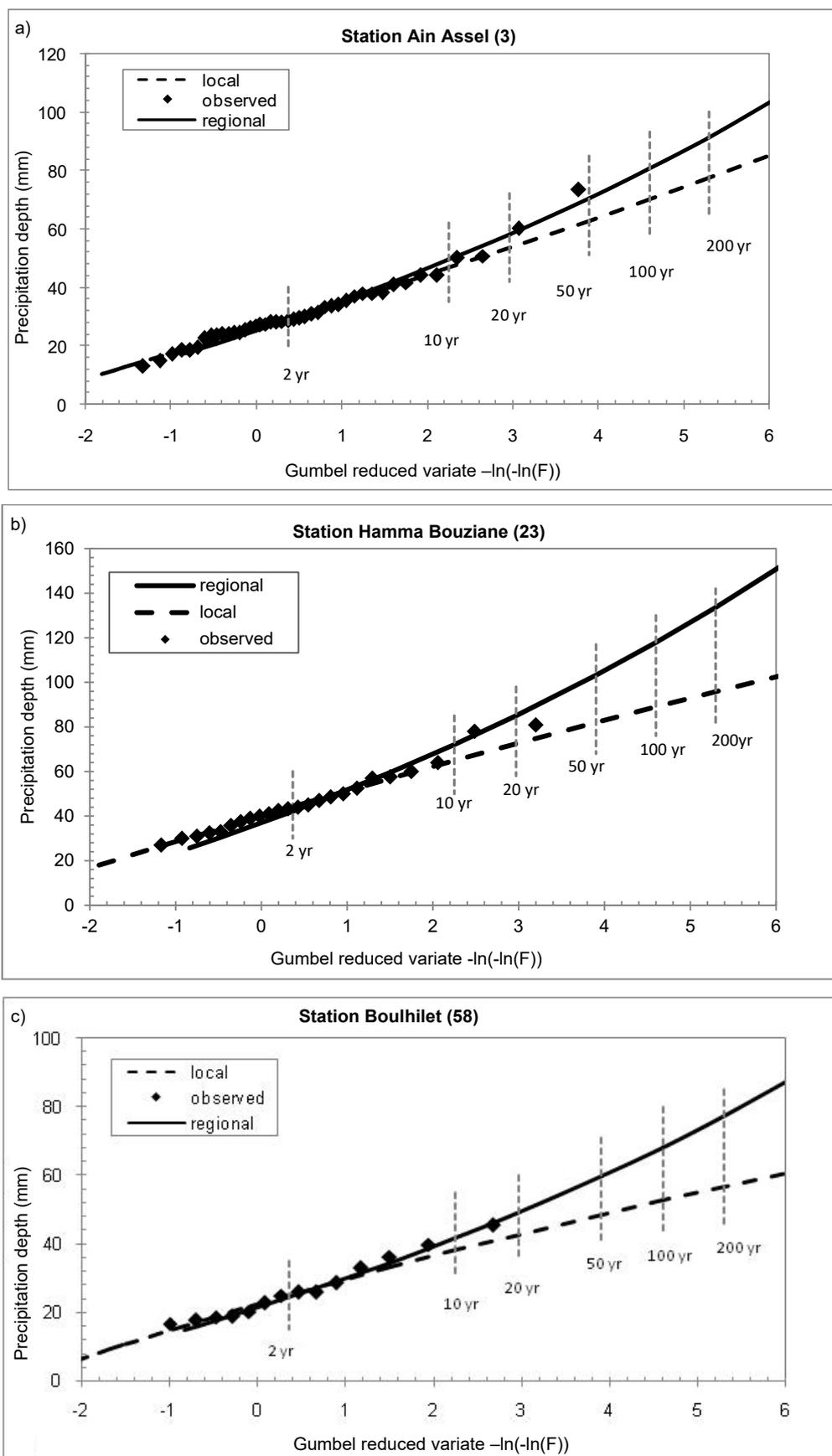
Fig. 4. Regional and local fitting of the observed precipitation at the three stations with length of record *n*:
a) *n* = 43 years; b) *n* = 24 years; c) *n* = 14 years; the numbers of these stations in Figures 1 and 3 are: a) (3); b) (23); c) (58);
source: own study
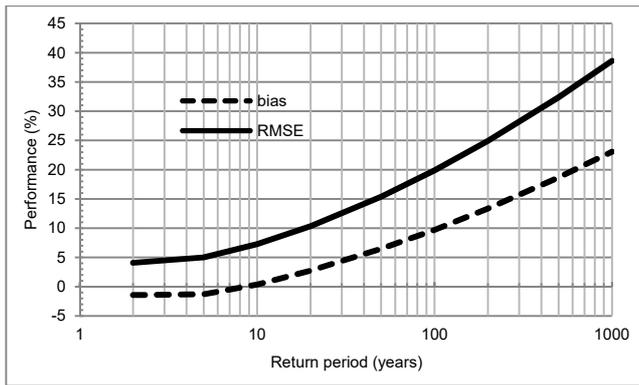
Fig. 5. Variations of root square mean error (*RMSE*) and bias related to the regional estimation of precipitation quantiles; source: own study

It is noticed that the gap between the bias and *RMSE* is relatively high for high return periods.

Since $RMSE = \sqrt{\text{bias}^2 + (Var)}$, this means that the variance (*Var*) of the error in estimation of the regional quantiles is dispersed for these return periods. This is partly explained by the presence of strong quantile deviations estimated in some stations, in particularly, in stations with short length of record. In Table 6 we present the deviations of estimated quantiles from regional and at-site analysis for the three stations (Ain Assel, Hamma Bouziane and Boulhilet) presented in Figure 4. The quantile deviations for the Hamma Bouziane and Boulhilet stations are the highest compared to the other stations in the region.

Table 6. Deviations of regional and local quantiles for different return periods for the three stations Ain Assel, Hamma Bouziane and Boulhilet

Return period years	Ain Assel *n* = 43		Hamma Bouziane *n* = 24		Boulhilet *n* = 14	
	deviation					
	mm	%	mm	%	mm	%
1 000	23.8	25.0	63.9	73.0	35.5	54.0
200	15.8	17.8	38.3	39.9	20.4	35.9
100	10.4	14.8	29.4	32.9	15.2	28.8
50	7.5	11.9	21.6	26.3	10.8	22.5
20	4.4	8.2	12.9	17.8	6.1	14.1
10	2.5	5.3	7.4	11.5	3.2	8.2
5	1.0	2.4	2.9	5.1	0.9	2.7
2	–0.6	–2.1	–1.9	–4.5	–1.1	–4.5

Explanations: *n* = length of record.
Source: own study.

To test the relevance of the regional model, we reduced the initial sample size by arbitrarily removing precipitation values. Regional and local quantiles were recalculated for the new sample and then, compared to those in the initial sample. The values to be deleted were randomly drawn by the Monte Carlo simulation procedure. Finally, the initial sample was composed of 22 years of record per station and two new samples of 18 and 14 years of record per station for a total of 58 stations. In our case study, the mean of record length is 22 years. To evaluate the obtained results, we computed: i) *RMSE* between the estimated regional quantiles for initial sample and new sample;

ii) *RMSE* between the estimated local quantiles for the initial and new sample (Tab. 7). It should be noted that the estimation error increases when the local sample size decreases and that this increase is greater for the local estimation of quantiles. It can be concluded that the regional approach stabilizes the variability of quantile estimation. This is due to the fact that the regional sample is always larger than the local sample.

Table 7. Local and regional *RMSE* for different sample sizes

Return period years	Local sample size (*n*)			
	n = 18 years		*n* = 14 years	
	$RMSE_R$ (%)	$RMSE_L$ (%)	$RMSE_R$ (%)	$RMSE_L$ (%)
10	4.24	5.81	8.05	10.53
100	4.61	10.22	8.65	19.54

Explanations: $RMSE_R$= relative root mean square error between the regional quantiles; $RMSE_L$= relative root mean square error between the local quantiles.
Source: own study.

CONCLUSIONS

This study focused on the regionalization of annual maximum daily precipitation, with the aim of obtaining a reliable quality of extreme quantile estimation at the stations in the study area. The regional frequency analysis based on the L-moments was used. The homogeneity test made it possible to define that the study region represented by 58 stations is homogeneous in terms of L-moments ratios despite the climatic differences within the region. The important step in the regional analysis was the identification of the regional distribution. The three-parameter distributions GEV, LN3, GLO and P3 were tested. Using the *Z*-statistic test, it was found that the GEV and LN3 distributions had a satisfactory fit to the regional sample. Following the two other statistical tests applied, the GEV distribution was retained as being more robust than LN3. The growth curve, derived from the regional distribution (GEV), was established. Therefore, to estimate the different return period's precipitation quantiles in a given site of the region, the mean precipitation of the site has to be multiplied by the corresponding regional quantile (growth factor). To assess the precision and quantify the uncertainties of the regional frequency curve estimation, the Monte Carlo simulation test was performed. The low values of the relative *RMSE* demonstrate the reliability of the applied method. The confidence interval (90%) obtained is narrow because of the quasi-exponential regional distribution (the shape parameter is close to zero).

To assess the used of regional model relevance, the inherent bias and *RMSE* of the regional quantiles estimation were calculated. This performance investigation showed that the bias and *RMSE* are relatively heavy for the high return periods (*T* > 10 years), and that the variance of the error of quantiles estimation is quite dispersed for these return periods. The *RMSE* values vary from 7.27% to 38.6% for return periods ranging from 10 to 1000 years. The comparison of estimated quantiles from regional and at-site frequency analysis showed that in the majority of stations (82.8%) the local model underestimates, in reasonable proportions, the quantiles of high return periods. Fur-

thermore, the relevance study showed that the regional approach stabilizes the variability of quantile estimation. Thus, the obtained results allow us to conclude that the regional approach leads to more reliable estimates of precipitation quantiles that the local approach.

REFERENCES

ALILA Y. 1999. A hierarchical approach for the regionalization of precipitation annual maxima in Canada. Journal Geophysical Research: Atmospheres. Vol. 104 (D24) p. 31645–31655. DOI 10.1029/1999JD900764.

CANNAROZZO M., D'ASARO F., FERRO V. 1995. Regional and frequency analysis for Sicily using the two component extreme value distribution. Hydrological Sciences Journal. Vol. 40(1) p. 19–42. DOI 10.1080/02626669509491388.

DALRYMPLE T. 1960. Flood frequency methods [online]. US Geological Survey Water Supply Paper. 1543A p. 11–51. [Access 15.01.2018]. Available at: http://pubs.usgs.gov/wsp/1543a/report.pdf

DJERBOUA A. 2001. Prédétermination des pluies et crues extrêmes dans les Alpes franco-italiennes. Prévision quantitative des pluies journalières par la méthode des analogues [Predetermination of extreme rains and floods in the Franco-Italian Alps. Quantitative forecast of daily rainfall by the analogous method] [online]. PhD thesis. Grenoble, France. Institut polytechnique de Grenoble pp. 214. [Access 15.01.2018]. Available at: http://hydrologie.org/THE/djerboua/DJERBOUA.htm

GAAL L., KYSELY J., SZOLGAY J. 2008. Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia. Hydrology and Earth System Sciences. Vol. 12 p. 825–839. DOI 10.5194/hess-12-825-2008.

GELLENS D. 2002. Combining regional approach and data extension procedure for assessing GEV distribution of extreme precipitation in Belgium. Journal of Hydrology. Vol. 268 p. 113–126. DOI 10.1016/S0022-1694(02)00160-9.

GREHYS 1996. Presentation and review of some methods for regional flood frequency analysis. Journal of Hydrology. Vol. 186 p. 63–84. DOI 10.1016/S0022-1694(96)03042-9. Groupe de recherche en hydrologie statistique.

HOSKING J.R.M. 1990. L-moments: Analysis and estimation of distribution using linear combination of order statistics. Journal of the Royal Statistical Society. Ser. B. Vol. 52 (1) p. 105–124.

HOSKING J.R.M., WALLIS J.R. 1997. Regional frequency analysis: An approach based on L-moments. Cambridge, UK. Cambridge Univ. Press. ISBN 978-1-108-00491-6 pp. 244.

KHEZAZNA A., AMARCHI H., DERDOUS O., BOUSAKHRIA F. 2017. Drought monitoring in the Seybouse basin (Algeria) over the last decades. Journal of Water and Land Development. Vol. 33 p. 79–88. DOI 10.1515/jwld-2017-0022.

KOUTSOYIANNIS D. 2004. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrological Sciences Journal. Vol. 49(4) p. 591–610. DOI 10.1623/hysj.49.4.591.54424.

KYSELY J., PICEK J. 2007. Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. Climate Research. Vol. 33 p. 243–255. DOI 10.3354/cr033243.

LIN B., VOGEL J.L. 2006. A comparison of L-moments with method of moments. Proceedings of the Symposium on Engineering Hydrology. New York. ASCE p. 443–448.

MEBARKI A. 2005. Hydrologie des bassins de l'est Algérien: Ressources en eau, aménagement et environnement [Hydrology of the basins of eastern Algeria: Water resources, development and environment] [online]. PhD thesis. Mentouri University of Constantine, Algeria pp. 320. [Access 15.01.2018]. Available at: http://hydrologie.org/THE/mebarki/MEBARKI.htm

MORA R.D., BOUVIER C., NEPPEL L., NEEL H. 2005. Regional approach for the estimation of low-frequency distribution of daily rainfall in the Languedoc-Roussillon region, France. Hydrological Sciences Journal. Vol. 30 (1) p. 85–109. DOI 10.1623/hysj.50.1.17.56332.

MRAD D., DJEBBAR Y., HAMMAR Y. 2018. Analysis of trend rainfall: Case of north-eastern Algeria. Journal of Water and Land Development. Vol. 36 p. 105–115. DOI 10.2478/jwld-2018-0011.

NGUYEN V.T.V., NGUYEN T.D., ASHKAR F. 2002. Regional frequency analysis of extreme rainfall. Water Science and Technology. Vol. 45 (2) p. 75–81.

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. Journal of Hydrology. Vol. 345 p. 149–166. DOI 10.1016/j.jhydrol.2007.07.009.

OUARDA T.B.M., LONG M., BOBEE B., BERNIER J., BOIS P. 1999. Synthèse de modèles régionaux d'estimation de crues utilises en France et au Québec [Analysis of regional flood models utilized in France and Québec]. Revue des sciences de l'eau. Vol. 12(1) p. 155–182. DOI 10.7202/705347ar.

OUARDA T.B.M., ST-HILAIRE A., BOBEE B. 2008. Synthèse des développements récents en analyse régionale des extrêmes hydrologiques [A review of recent developments in regional frequency analysis of hydrological extremes]. Revue des sciences de l'eau. Vol. 21(2) p. 219–232. DOI 10.7202/018467ar.

OVEREEM A., BUISHAND A., HOLLEMAN I. 2007. Rainfall depth-duration-frequency curves and their uncertainties. Journal of Hydrology. Vol. 348 p. 124–134. DOI 10.1016/j.jhydrol.2007.09.044.

ST-HILLAIRE A., OUARDA T.B.M., LACHANGE M., BOBEE B., BARBET M., BRUNEAU P. 2003. La régionalisation des précipitations : une revue bibliographique des développements récents [The regionalization of precipitation: A bibliographic review of recent developments]. Revue des sciences de l'eau. Vol. 16(1) p. 27–54. DOI 10.7202/705497ar.

SVEINSSON O.G.B., SALAS J., DUANE G.B. 2002. Regional frequency analysis of extreme precipitation in northern Colorado and the Fort Collins flood of 1997. Journal of Hydrologic Engineering. Vol. 7(1) p. 49–63. DOI 10.1061/ (ASCE)1084-0699(2002)7:1(49).

VERSIANI B.R., DE ANDRADE PINTO E.J., BOIS P. 1999. Analyse des pluies extrêmes annuelles sur la région de Minas Gerais (Brésil) : modèle de régionalisation TCEV [Analysis of extreme annual rainfall over the Ninas Gerais region (Brazil): TCEV regional model]. In: Hydrological extremes: Understanding, predicting, mitigating [online]. Ed. L. Gottschalk, J.C. Olivry, D. Reed, D. Rosbjerg. Proceeding of IUGG 99 Symposium HSJ. Birmingham July 1999. IAHS Publ. Vol. 255 p. 201–207. [Access 15.01.2018]. Available at: http://iahs.info/uploads/dms/iahs_255_0201.pdf

Soufiane DAD, Tamara BENABDESSELAM

Analiza częstotliwości ekstremalnych opadów w regionie północno-wschodniej Algierii

STRESZCZENIE

Celem pracy była poprawa jakości ocen maksymalnego w ciągu roku dobowego opadu na obszarze północno-wschodniej Algierii. Wykorzystano statystykę L-momentów do analizy częstotliwości regionalnej. Badany obszar jest reprezentowany przez 58 stacji pomiarowych. Głównymi etapami badania było zdefiniowanie regionów homogenicznych i identyfikacja rozkładu regionalnego. Stwierdzono, że badany region jest homogeniczny w znaczeniu proporcji momentów L, mimo różnic klimatycznych w obrębie regionu. Spośród testowanych rozkładów najbardziej odpowiedni do modelowania opadu w regionie okazał się uogólniony rozkład wartości ekstremalnych (GEV). Ustalono krzywą wzrostu wyprowadzoną z rozkładu regionalnego. Aby oszacować kwantyle okresów powtarzalności opadów dla danego stanowiska w regionie, średni opad w tym stanowisku należy pomnożyć przez odpowiedni kwantyl regionalny (czynnik wzrostu). Porównanie analizy częstotliwości w odniesieniu do regionu i stanowiska wykazało, że w większości stacji (82,8%) model stanowiskowy obniża wyniki oszacowania kwantyli o długim okresie powtarzalności.

Słowa kluczowe: *analiza regionalnej częstotliwości, ekstremalne opady dobowe, momenty L, regiony homogeniczne, rozkład GEV*