Statistical modeling of annual maximum precipitation in Oued El Gourzi Watershed, Algeria

Nassim Bella¹ · Hadda Dridi² · Mahdi Kalla²

Received: 9 July 2019 / Accepted: 3 March 2020 / Published online: 16 March 2020
© The Author(s) 2020

Abstract

This study aims to model annual maximum precipitation based on extreme value theory for the Oued El Gourzi Watershed, Algeria. A generalized extreme value (GEV) distribution was used to determine the probability distribution of extreme values and their dependency on time for the five stations distributed across the watershed. The non-stationary models are used to represent the GEV parameters assumed an invariant shape parameter and linear functions as location and scale parameters. The best model was selected using Akaike's information criterion and Bayesian information criterion. Stationary and non-stationary return levels for different return periods have been proposed for the study area.

Keywords GEV distribution · Stationary and non-stationary models · Annual maximum precipitation · Oued El Gourzi Watershed

Introduction

The analysis of the extreme values in climatological time series is an area of intense scientific activity. The annual or monthly maximum precipitation or temperature series are the examples of this type of data. The generalized extreme value (GEV) distribution is widely employed for modeling the extreme precipitation in the environmental sciences and many other fields (Reiss and Thomas 2007; Li-Ge et al. 2013; Ngailo et al. 2016; Boudrissa et al. 2017). The assumption of independent and identically distributed data in the series with constant properties through time (stationarity) may need to be modified to reflect climate change (non-stationarity). For example, the maximum temperature and precipitation series could show trends over time (Panagoulia et al. 2014). Furthermore, due to natural climate variability or anthropogenic climate change, there is evidence that the hydroclimatic extreme series are not stationary (Shaleen and Lall 2001; Milly et al. 2008). Many

studies have analyzed extreme precipitation using either generalized GEV distribution which provides evidence of the importance of modeling precipitation from different regions of the world (Buishand et al. 2008; Carreau et al. 2013; Ender and Ma 2004). From example, Koutsoyiannis and Baloutsos (2000) applied to Greece's rainfall data set. Crisci et al. (2002) applied extreme value distributions to rainfall data set from Italy. Koutsoyiannis (2004) applied extreme value theory to rainfall data set from Europe and USA. O'Gorman (2015) analyzed precipitation extremes under climate change. The observational and statistical modeling results of the different studies have shown that there are remarkable increases in intensity of precipitation extremes. However, there has been little or no published research that has attempted to detect extreme precipitation by using GEV in regions of Algeria. Therefore, this paper would seem to be the first application of the GEV distributions for extreme precipitation in Algeria.

In this study, we analyze the annual maximum precipitation data from 1969 to 2011 in five stations in Oued El Gourzi Watershed, Algeria. Three GEV models are proposed to fit the annual maximum precipitation for each station. Then, we have compared the different fitted models and selected the best model based on the Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

✉ Nassim Bella
nassimbella@yahoo.fr

¹ Institute of the Sciences of Earth and Universe, Geography and Territory Planning Department, University of Mustapha Ben Boulaid Batna 2, Batna, Algeria

² Laboratory of Natural Hazards and Regional Planning, University of Mustapha Ben Boulaid Batna 2, Batna, Algeria

The paper is organized as follows. The annual maximum precipitation data are described in “[Study area and precipitation data set](#)” section. The models and the fitting procedure are described in “[Statistical modeling](#)” section. In “[Conclusion](#)” section, the results of the fitted models and their implications are discussed. The yearly return level estimates for the 20 and 100 years are reported.

Materials and methods

Study area and precipitation data set

Oued El Gourzi Watershed is part of the great watershed of the Constantine's Highlands in northeastern Algeria (Fig. 1). It is 400 km east of Algiers, between 6° 01' 44" and 6° 21' 15" east longitude and between 35° 25' 57" and 35° 36' 33" north latitude. It covers almost an area of 315 km². Oued El Gourzi is of great importance in the Hydrographic system, fed by five main tributaries which are Oued Tazoult

Southwest, Oued Azzeb and Oued Bouadane northwest, Oued Segouene northeast and Oued Hamla southeast; the studied basin takes these sources from the subsequent mountains. In the north, Dj Boumerzoug with an altitude of 1692 m, Dj Kassrou (1641 m), the northeast is occupied by Dj Azzab (1365 m) and Dj Bouarif (1584 m), the western part is dominated by Dj Tugurth (2091 m) and Dj Boukez-zaz (1783 m), whether the south is occupied by Dj Ich Ali, whose altitude is about 1800 m, flows into the plain by Oued El Gourzi. The town of Batna located at the outlet of the Oued El Gourzi with a large population. Batna city recorded for the year 1995, 236,669 inhabitants, the largest rate of urbanization of the wilaya (89.4%) (Sefouhi et al. 2010). In addition, the population density is about 2050 Hab/km² (The basic population data were collected from the last census of April 16, 2008).

Annual maximum precipitation (AMP) data in Oued El Gourzi Watershed for the time period 1969–2011 were obtained from the National Agency of Hydraulic Resources (NAHR). In this study, we selected five stations: Ali Ben

Fig. 1 Geographical map of the Oued El Gourzi Watershed and the five stations

Table 1 Coordinates of the selected stations

Station	Code	Longitude	Latitude	Altitude
Ali Ben Tenoun (ABT)	70304	6°19'53"	35°34'53"	1180
Batna (BAT)	70316	6°10'02"	35°325'03"	1040
Hamla (HAM)	70308	6°06'48"	35°29'13"	1174
Seguene (SEG)	70309	6°07'44"	35°35'02"	1400
Tazoult (TAZ)	70303	6°15'36"	35°25'08"	1200

Table 2 Different combinations and the regression function

Model	Parameters
Model 0 (stationary)	μ, σ and ξ are constants
Model 1	$\mu(t) = \mu_0 + \mu_1 * t$; σ and ξ are constants
Model 2	$\mu(t) = \mu_0 + \mu_1 * t$; $\sigma(t) = \sigma_0 + \sigma_1 * t$; ξ is constant

Tenoun (ABT), Batna (BAT), Hamla (HAM), Seguene (SEG) and Tazoult (TAZ) (Fig. 1). The statistical properties of annual maximum precipitation series of each station along with their geographic location are shown in Tables 1 and 2.

Statistical modeling

Generalized extreme value (GEV) distribution

The GEV distribution is widely employed for modeling extremes in the environmental sciences and elsewhere (Reiss and Thomas 2007). It depends on location (μ), scale (σ) and shape (ξ) parameters. The generalized extreme value (GEV) distribution is a flexible three-parameter model that combines the Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$) extreme value distributions. In the stationary GEV

$$F(y; \mu(t), \sigma(t), \xi) = \exp \left\{ - \left[1 + \xi \frac{y - \mu(t)}{\sigma(t)} \right]^{-1/\xi} \right\}. \quad (1)$$

In the case of non-stationary, the following regression structures could be considered for the location and scale parameters:

$$\mu(t) = \mu_0 + \mu_1 \times t \quad (2)$$

$$\sigma(t) = \sigma_0 + \sigma_1 \times t \quad (3)$$

Allowing up to linear dependence on time of both the location and scale parameters, we denote by $GEV(i, j, 0)$ the model with time dependence of order i in the location parameter and order j in the scale parameter. For example, the stationary GEV distribution is $GEV(0, 0, 0)$, obtained when the location and scale parameters are both independent of time ($\mu_1 = \sigma_1 = 0$), while the $GEV(1, 1, 0)$ non-stationary model assumes a linear trend in location and scale. Two models of varying complexity may be defined in this way (two choices for each of i and j). Table 2 shows the different GEV models and their parameters.

We followed the recommendation in the program documentation to standardize covariates. Thus, the linear term in time is actually entered into the model as

$$x_i = \frac{t_i - m}{s} \quad (4)$$

where m and s are the mean and the standard deviation of the time covariate, respectively.

Model selection

The stationary and non-stationary GEV models may be fitted to a time series $\{y(t_i): t_i \in T\}$ where $T = \{t_1, t_2, \dots, t_n\}$ by maximizing the log-likelihood function as follows:

$$\ell = \log L = \sum_{i=1}^n \left(-\log(\sigma(t_i)) - \left(1 + \frac{1}{\xi} \right) \log \left(1 + \xi \frac{y(t_i) - \mu(t_i)}{\sigma(t_i)} \right) - \left[1 + \xi \frac{y(t_i) - \mu(t_i)}{\sigma(t_i)} \right]^{-1/\xi} \right). \quad (5)$$

distribution, the three parameters are constant, while, in the non-stationary GEV distribution (El Adlouni et al. 2007; Leclerc and Ouarda 2007), these parameters are expressed as a function of time t and possibly other covariates (Coles 2001). If, as is usually done, we allow non-stationarity of the location and scale parameters but not of the shape parameter, this non-stationary $GEV(\mu(t), \sigma(t), \xi)$ distribution has distribution function:

The goodness of fit and the significance of the models were tested with the aid of log-likelihood ratio test (LRT; Sienz et al. 2010), Akaike's information criterion (AIC; Akaike 1974) and the Bayesian information criterion (BIC; Schwarz 1978). LRT, AIC and BIC methods are used to choose the best model at each station. The corrected AIC (AIC_c ; Burnham and Anderson 2002) is used to select the best model for a small sample $[(n/k) < 40]$. If $\hat{\ell}$ is the maximized value of the likelihood for a model containing p

Table 3 The characteristics of the annual maximum precipitation in the analyzed period

Station	ABT	BAT	HAM	SEG	TAZ
Min	12.0	17.2	9.0	16.4	18.0
Max	101.3	64.6	94.5	129.8	164.6
Median	34.0	36.2	29.0	43.3	37.0
Mean	38.4	37.6	35.9	50.3	44.9
SD	17.1	13.2	19.9	22.6	25.1
C_v	0.45	0.35	0.55	0.45	0.56

Min minimum, *Max* maximum, *SD* standard deviation, C_v coefficient of variation, the significance of bold: the maximum value of the max, median and mean for the five stations

Fig. 2 Boxplot of the annual maximum precipitation

parameters and n is the sample size, the criteria are defined as

(the third term is the correction) and

$$AIC_c = -2\hat{\ell} + 2p + \frac{2p(p+1)}{n-p-1} \quad (6)$$

$$BIC = -2\hat{\ell} + p \log n. \quad (7)$$

The preferred model among the GEV models is the one that minimizes the chosen criterion, although attention should also be paid to models with values close to the minimum.

All results reported in this work were obtained in the R computing environment (R Development Core Team 2019), using the *fevd* routine in the *extRemes* package (Gilleland and Katz 2016) to fit stationary and non-stationary GEV distributions by maximum likelihood.

Results and discussion

Statistical descriptive of the annual maximum precipitation data

The preliminary analysis of the annual maximum precipitation data included the calculation of descriptive statistics. Specifically, we computed minimum (Min), maximum (Max), median, mean, standard deviation (SD) and coefficient of variation (C_v). Table 3 presents the values of the descriptive statistics for the annual maximum precipitation time series for all the stations. Figure 2 shows the boxplot of the annual precipitation for each station, and the graphical presentation of the temporal variability of precipitation for each station is shown in Fig. 3. The results show that the maximum value of AMP is observed in TAZ station, while the highest mean and median values are observed in SEG station (Fig. 2). The lowest value of C_v is for BAT station (35%) and the highest for TAZ station (65%). According to Hare (2003), coefficient of variation is used to classify the degree of variability of annual maximum precipitation as less ($C_v < 20\%$), moderate ($20\% < C_v < 30\%$), high ($C_v > 30\%$), very high $C_v > 40\%$ and $C_v > 70\%$ indicate extremely high variability of annual maximum precipitation (Table 3 and Fig. 3). Based on this, from the observed data considered that all the stations had above 30% coefficient of variation highlighting the high variability of annual maximum precipitation over the Oued El Gourzi Watershed (Table 3 and Fig. 3).

In order to understand the relationship between temperature and precipitation, the ombrothermal diagram of Bagnouls and Gaussen for Batna station (BAT) is used (Fig. 4). This diagram is presented by plotting on the abscissa axis, the months of the year, and on the ordinate the precipitation on the right and the average temperatures on the left ($P=2T$) (Bagnouls and Gaussens 1953). From Fig. 4, we see clearly that the wet period ranges from November to May (7 months) for the study area.

Preliminary analysis

As a first approach to study trends in annual maximum precipitation during the study period 1969–2011, the Mann–Kendall trend test is applied. A trend is considered to be present if it has been detected by the test. The results show that at the 0.05 significance level, the annual maximum precipitation series of ABT and TAZ stations exhibited a statistically significant trend, while the other stations did not have any statistically significant trend when we considered the entire study period (Table 4 and Fig. 5).

Fig. 3 Graphical presentation of the annual maximum precipitation for each station

Fig. 4 Ombrothermal diagram of Bagnouls and Gaussen for Batna station (BAT)

Comparison of selection criteria

We investigated the use of the GEV distribution to model the annual maximum precipitation in Oued El Gourzi

Table 4 Mann–Kendall trend test for each station

Station	Z	p value
ABT	3.893	9.896E–05
BAT	0.659	0.510
HAM	0.576	0.565
SEG	–0.136	0.892
TAZ	2.313	0.021

Watershed. We modeled these events using both stationary and non-stationary models for the time period 1969–2011. The effect of time is taken into account. In this study, three GEV models are proposed. The parameter estimates and goodness-of-fit criteria tests (AIC_c and BIC) were calculated for different GEV models, as shown in Table 5. The best model was chosen based on the minimum values of GOF criteria. From the results, we see that the model 2, whose location parameter depends on time and whose other parameters are constant, is the best model for explaining change in the annual maximum precipitation at ABT and TZA stations. The model 0 (stationary case), where there are no significant trends, is the best model for the other stations. These

Fig. 5 Plot of annual maximum precipitation data for 1969 to 2011 with trend fitted by lowness smoother

findings are supported by Fig. 6, where we have shown the quantile and density plots of the fits for the five-time series. Figure 6 shows that the fit provided by model 0 and model 1 is good in this study area.

From Table 5, we see clearly that the location and scale parameters are high for SEG station. An analysis of the shape parameter obtained for all of the models and all of the stations except BAT station shows that this parameter is positive. A greater absolute value of this parameter corresponds to a greater annual extreme precipitation.

Return level estimates

Once the best model for the data has been determined, the interest is to derive the return levels of annual maximum precipitation. The T year return level, say x_T , is the value occurring on average once in every T years. For example,

Table 5 Parameter estimates and summary of goodness-of-fit tests for three models

Parameters	ABT	BAT	HAM	SEG	TAZ
Model 0					
μ	30.85	32.06	26.28	39.97	33.68
σ	12.82	11.48	13.19	15.60	11.69
ξ	0.01	-0.12	0.14	0.08	0.29
$-\log l$	177.69	170.27	182.32	187.89	180.74
AIC_c	361.38	346.53	370.64	381.79	367.49
BIC	366.66	351.82	375.92	387.07	372.77
Model 1					
μ_0	31.32	32.01	26.20	39.96	33.77
μ_1	8.07	1.51	-0.55	-0.01	3.25
σ	9.72	11.30	13.11	15.60	10.53
ξ	0.14	-0.11	0.15	0.08	0.36
$-\log l$	169.11	169.92	182.28	187.89	177.92
AIC_c	346.23	347.85	372.56	383.79	363.84
BIC	353.27	354.89	379.61	390.83	370.89
Model 2					
μ_0	31.14	32.09	26.24	40.06	34.13
μ_1	7.93	1.47	0.35	-1.09	4.86
σ_0	9.54	11.37	13.07	15.61	10.74
σ_1	-0.54	-0.26	1.70	-1.79	1.80
ξ	0.17	-0.12	0.15	0.07	0.32
$-\log l$	169.06	169.90	181.78	187.52	177.14
AIC_c	348.12	349.81	373.56	385.04	364.28
BIC	356.92	358.62	382.37	393.84	373.09

the 2-year return level is the median of the distribution of each station. If the model 1 is assumed, then on inverting $F(x_T) = 1 - \frac{1}{T}$ we get:

$$x_T = \mu - \frac{\sigma}{\xi} \left[1 - \left\{ -\log \left(1 - \frac{1}{T} \right) \right\}^{-\xi} \right]. \quad (8)$$

By substituting into Eq. 8, we obtain the maximum likelihood estimates of the return levels. Confidence intervals for the return level estimates are obtained by means of the delta method (Roa 1973).

The 20- and 100-year return levels for each station are shown in Table 6. It is clear that at each station, there only 1–4 observed annual maximum precipitation exceeded the 20-year return level. None of the observed annual maximum precipitation has exceeded the 100-year return level in ABT, BAT and HAM stations, while just one observed annual maximum precipitation has exceeded the 100-year return level in SEG and TAZ stations, respectively.

Fig. 6 Quantile and density plots for each station

Table 6 Return levels for each station

Station	$T=20$	$T=100$
ABT	74.95	101.47
BAT	60.67	72.47
HAM	74.89	111.54
SEG	92.02	126.06
TAZ	93.14	161.59

Conclusion

In the current study, the annual maximum precipitation of the five stations in the Oued El Gourzi Watershed is fitted by generalized extreme value (GEV) distribution. The effect of linear trend in time has been analyzed in this research. Stationary model (model 0) and two non-stationary models, model 1 (linear trend in location), model 2 (linear trend in both location and scale), were proposed. The different proposed models are compared using AIC_c and BIC criteria. The results show that the model 1 and model 2 are the most adequate models for explaining the variance in annual maximum precipitation data over the Oued El Gourzi Watershed. This case study shows that it is necessary to incorporate non-stationarity into annual maximum precipitation by linking time with the distribution parameters to improve estimations.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Bagnouls F, Gaussen G (1953) Période de sécheresse et végétation. *Les Comptes rendus de l'Académie des sciences* 236:1076–1077
- Boudrissa N, Cheraitia H, Halimi L (2017) Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009. *Meteorol Appl* 24:114–119
- Buishand TA, de Haan L, Zhou C (2008) On spatial extremes: with application to a rainfall problem. *Ann Appl Stat* 2(2):624–642
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
- Carreau J, Neppel L, Arnaud P, Cantet P (2013) Extreme rainfall analysis at ungauged sites in the South of France: comparison of three approaches. *J Soc Fr Stat* 154(2):119–138
- Coles S (2001) An introduction to statistical modeling of extreme values. Springer, New York
- Crisci A, Gozzini B, Meneguzzo F, Pagliara S, Maracchi G (2002) Extreme rainfall in a changing climate: regional analysis and hydrological implications in Tuscany. *Hydrol Process* 16:1261–1279
- El Adlouni S, Ouara TBMJ, Zhang X, Roy R, Bobee B (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value distribution. *Water Resour Res* 43:W03410
- Ender M, Ma T (2004) Extreme value modeling of precipitation in case studies for China. *Int J Sci Innov Math Res* 2(1):23–36
- Gilleland E, Katz R (2016) extRemes 2.0: an extreme value analysis package in R. *J Stat Soft* 72:8
- Hare W (2003) Assessment of knowledge on impacts of climate change—contribution to the specification of art. 2 of the UNFCCC. Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen
- Koutsoyiannis D (2004) Statistics of extreme and estimation of extreme rainfall II: empirical investigation of long rainfall records. *Hydrol Sci J* 4:591–610
- Koutsoyiannis D, Baloutsos G (2000) Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences. *Nat Hazards* 22:29–48
- Leclerc M, Ouara T (2007) Non-stationary regional flood frequency analysis at ungauged sites. *J Hydrol* 343:254–265
- Li-Ge C, Jun Z, Bu-Da S, Jian-Qing Z, Gemmer M (2013) Probability distribution and projected trends of daily precipitation in China. *Adv Climate Change Res* 4(3):153–159
- Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: Whither water management? *Science* 319:573–574
- Ngailo TJ, Reuder J, Rutalebwa E, Nyimvua S, Mesquita MDS (2016) Modelling of extreme maximum rainfall using extreme value theory for Tanzania. *Int J Sci Innov Math Res* 4(3):34–45
- O'Gorman PA (2015) Precipitation extremes under climate change. *Curr Clim Change Rep* 1:49–59
- Panagoulia D, Economou P, Caroni C (2014) Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. *Environmetrics* 25:29–43
- R Core Team (2019) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna. <http://www.rproject.org/index.html>
- Reiss RD, Thomas M (2007) Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields, 3rd edn. Birkhauser, Basel
- Roa CR (1973) Linear statistical inference and its applications, 2nd edn. Wiley, New York
- Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Sefouhi L, Kalla M, Aouragh L (2010) Étude pour une gestion durable des déchets ménagers de la ville de Batna (Algérie), Déchets, Sciences et Techniques (DST), vol 58
- Shaleen J, Lall U (2001) Floods in a changing climate: does the past represent the future? *Water Resour Res* 37:3193–3205
- Sienz F, Schneidereit A, Blender R, Fraedrich K, Lunkeit F (2010) Extreme value statistics for North Atlantic cyclones. *Tellus A* 62: 347–360. <https://doi.org/10.1111/j.1600-0870.2010.00449.x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.