Complete Flood Frequency Analysis in Abiod Watershed Biskra (Algeria)

S. BENAMEUR1, A. BENKHALED2, D.MERAGHNI1, F.CHEBANA*[3], A. NECIR1

[1]*Laboratory of Applied Mathematics, University of Biskra, Po Box 145, RP 07000 Biskra, Algeria*

[2]*Research Laboratory in Subterranean and Surface Hydraulics -LARHYSS, University of Biskra, Po Box 145, RP 07000Biskra, Algeria*

[3]*INRS-ETE, University of Quebec, 490 rue de la Couronne, Québec, QC, Canada*

[*]Corresponding author: **fateh.chebana@ete.inrs.ca**

November 2016

ABSTRACT

Extreme hydrological events, such as floods and droughts, are one of the natural disasters that occur in several parts of the world. They are regarded as being the most costly natural risks in terms of the disastrous consequences in human lives and in property damages. The main objective of the present study is to estimate flood events of Abiod wadiat given return periods at the gauge station of M'chouneche, located closely to the city of Biskra in a semi-arid region of Southern-East of Algeria. This is a problematic issue in several ways, because of the existence of a dam to the downstream, including the field of the sedimentation and the water leaks through the dam during floods. The considered data series is new. A complete frequency analysis is performed on a series of observed daily average discharges, including classical statistical tools as well as recent techniques. The obtained results show that the Generalized Pareto distribution (GPD), for which the parameters were estimated by the maximum likelihood (ML) method, describes the analyzed series better. This study also indicates to the decision-makers the importance to continue monitoring data at this station.

Key words : Frequency analysis; Peaks-Over-threshold ; Generalized Pareto distribution ;Threshold selection ; Flood discharges ; Extremequantiles; Biskra ; Algeria.

Introduction

The study of floods is a subject which arouses more and more interest in the field of water sciences. In spite of their low rainfall, the basins of the arid and semi-arid areas represent a hydroclimatic context where the overland flows phenomena are significant and feed a network of very active wadis. The activity of these wadis is far from being negligible from the flood in terms of their frequency and intensity. One observes on these rivers exceptional flows which sometimes, surprise by their magnitude[19]. The Abiod wadi, in the area of Biskra, is a very representative river of these basins. Moreover, the existence of Foum El Gherza dam to the downstream for the irrigation of the palm plantations makes the area more sensitive with regard to the floods. The flood events of the years 1963, 1966, 1971, 1976 and 1989 remain engraved in the memory of the inhabitants. The flood event of 11- 12th September 2009 was one of the historic floods in the Zibans area[7]. It rains 80 mm in 24 hours, while the annual total of Biskra city reaches 100 mm. The damage were 9790 palm trees, 164 flooded houses, 744 destroyed greenhouses, 200 hectares of lost cultures. The last flooding at the time of this drafting paper is that produced in October 29th 2011. All the populations living downstream of the Foum El Gherza dam were evacuated. The floods mainly occur in September and October and especially originate from exceptional storm events.

Describing and studying these situations could help in preventing or at least reducing severe human and material losses. The strategy of prevention of flood risk should be founded on various actions such as risk quantification. On this aspect, various methodological approaches can contribute to this strategy, among which flood Frequency Analysis (FA). Frequency analysis of extreme hydrological events, such as floods and droughts, is one of the privileged tools by hydrologists for the estimation of such extreme events and their return periods. The main objective of FA approach is the estimation of the probability of exceedance $P\left(X \geq x_T\right)$, called hydrological risk, of an event x_T corresponding to a return period *T*[16].This process is accomplished by fitting a probability distribution F to large observations in a data set. Two approaches were developed in the context of extreme value theory (EVT). The first one, usually based on the generalized extreme value distribution (GEV), describes the limiting distribution of a suitably normalized annual maximum (AM) and the second uses the generalized Pareto distribution (GPD) to approximate the distribution of Peaks-Over-Threshold (POT). For more details regarding this theory and its applications, the reader is referred to textbooks such as Embrechts et al.[24],Reiss and Thomas[51], Beirlant *et al.*[6] and de Hann and Ferriera [18].

Many FA models should be tested to determine the best fit probability distribution that describes the hydrologic data at hand. Specific distributions are recommended in some countries, such as the Log-normal (LN) distribution in China[10]. In the United States, the Log-Pearson type 3 distribution (LP3) has been, since 1967[44], the official model to which data from all catchments are fitted for planning and insurance purposes. By contrast, the United Kingdom endorsed the GEV distribution[45, 46]up until 1999.The official distribution in this country is now the generalized logistic (GL), as for precipitation in the United States[59]. There are several examples where a number of alternative models have been evaluated for a particular country, for example Kenya[43], Bangladesh[35], Turkey[5] and Australia [58]. Nine distributions were used with data from 45 unregulated streams in Turkey by Haktanir[26]who concluded that two parameter Log-normal (LN2) and Gumbel distributions were superior to other distributions. Recent research was conducted by Ellouze and Abida[23]in ten regions of Tunisia. They found that the GEV and GL models provided better estimates of floods than any of the conventional regression methods, generally used for Tunisian floods. Rasmussen *et al*[50]reveals that the POT procedure is more advantageous than the AM in the case of short records. Lang *et al.*[40] develop a set of comprehensive practice-oriented guidelines for the use of the POT approach. Tanaka and Takara[55] has examined several indices to investigate how to determine the number of upper extremes rainfall best for the POT approach.

In the Algerian hydrological context, during the last two decades many authors have used several approaches to study the associated risks. Recently, Hebal and Remini[29]studied flood data from 53 gauge stations in northern Algeria, between 1966 and 2008. They found that 50 %, 25 % and 22 %of the samples follow respectively the Gamma, Weibull and Halphen A distributions. Bouanani [12]performed a regional flood FA in the Tafna catchments and concluded that the AM flows fit better to asymmetric distributions such as LP3, Pearson 3 and Gamma. The FA was also used in the sediment context by Benkhaled et al. [8]where the LN2 distribution was selected in the case of the same station considered in the present study, i.e. M'chouneche gauge station on Abiod wadi.

To the best knowledge of the authors, apart from Benkhaled et al. [8], the flood FA approach has not yet been performed on data collected at this station. The primary aim of this paper is to perform a FA to the Abiod wadi flow data by the POT approach, based on GPD approximation [30].In methodological terms, all the steps constituting FA are performed from data examination to risk assessment including hypotheses testing and model selection. Due to the high importance of the latter and its impacts, more recent techniques are employed to select the appropriate distribution that fit better to the tail. A relatively large number of known distributions fit well the center of the data whereas the focus in FA is on the distribution tail. To this end, tail classification and specific graphical tools are employed, see El Adlouni et al. [22] for more technical details.

The paper is organized as follows. In Section 2, the study area and the data set are briefly described. Section 3 is devoted to the FA methodology. The results of the study are presented and discussed in Section 4. Concluding remarks are reported in Section 5.

1. Study area and data

In this section, we present the region where the site of interest is located, followed by a description of the available data.

1.1 Study Area

The Abiod wadi watershed, with an area of 1300 Km^2, is located in the Aurès massif in the southern east of Algeria in North Africa (Figure 1). It is part of the endorheic watershed Chott Melghir. The wadi length is 85 km from its origin in the Chelia (2326 m high) and Ichemoul (2100 m high) mountains. After crossing Tighanimine, the wadi gradually flows into the canyons of Ghoufi and M'chouneche gorges, and then opens a path to the plain until the Saharian gorge Foumel Gherza. The valley of the wadi is mainly composed of sedimentary rocks, comprising alternating limestone, marl, soft sediments (sandstones, conglomerates) and some evaporates (gypsum) dated of Paleogene.

The watershed is characterized by its asymmetry, a mountainous area in the north to over 2000m (Chelia) and another low area in the south (El Habel 295 m). The relief is rugged with slopes ranging between 12.5% and 25% for half of the area, and from 3% to 12.5% for another 40% of the area. Land cover is a mix of rocky outcrops, highly eroded soil, sparse vegetation, a few forests, crops, gardens and pastures [27]. In the orographic and hydrographic points of view, Abiod wadi is characterized by two distinct climatic regions: the Aurès, where rainfall averages 450 mm/year, and the Sahara plain with mean rainfall 100-150 mm/year. The climate of Abiod wadi watershed is thus semi-arid to arid. Along Abiod wadi to the Foum El Gherzadam there are six rainfall stations, and one hydrometric station located 18 km upstream of the dam, as shown in Figure 1, which was damaged during the floods of 1994-1995 and it is not operational since.

The choice of this station was made on the basis of climatic context of the study area. It is the only station on the studied basin and it is rather representative of the whole south-east region in Algeria, which is arid to semi-arid. Also, the size of the series used shows the interest of the FA application.

1.2 Data Description

The data set used in this study is provided by the National Agency of Hydraulics Resources (ANRH) of Biskra and it is the first time to be considered and studied. It consists of the daily average discharges $Q_1,\ldots,Q_N$ (with N = 8034), collected at the gauge station of M'chouneche over 22 years from 1972 to 1994.

Note that the IACWD Bulletin 17B [1] suggests that at least 10 years of record are necessary to warrant a statistical analysis. For instance, Tramblay et al. [57] used a minimum of 10 years of daily data. The short data size can affect the choice of distributions, the quantile estimations, particularly those corresponding to large return periods and the extent of confidence intervals. The size of the used data in the present study is relatively large, to perform a FA, as in a number of similar studies[15].

2. Methodology

In this section, after defining the type of series to be analyzed, namely the POT series, we briefly present the required elements to perform a hydrological FA. The latter is a statistical approach of prediction commonly used in hydrology to relate the magnitude of extreme events to a probability of their occurrence[16]. It allows, for the selected station, to estimate the flood quantiles of given return periods. In general, FA involves four main steps:

(i) characterization of the data and determination of the usual statistical indicators, such as the mean, the standard-deviation, the coefficients of skewness (Cs), kurtosis (Ck) and variation (Cv)and detection of outliers,

(ii) checking the basic hypotheses of FA, i.e. homogeneity, stationarity and independence, applicability on the studied data set,

(iii)fitting of probability distributions, estimation of the associated parameters and selection of the best model to represent the data, and

(iv) risk assessment based on quantiles or return periods, [e.g. 11, 14, 26, 49].

3.1. Peaks Over Threshold Series

The data to be extracted and then used in this approach consist in the observations that exceed a selected relatively high threshold u. Let Q represent the daily average discharge and denote by N_u the number of discharges exceeding u. Then, the sample of excesses is defined as

$$\left\{E_j = Q_{i_j} - u \quad s.t.\, Q_{i_j} > u\,;\, j = 1, \ldots N_u\right\}.$$

In this approach the selection of an appropriate threshold is crucial. This approach is useful and has some advantages compared to the AM one, even though the latter is widely used. It is of particular interest in situations where the AM could not perform well especially in situation with little extreme data or the extracted extremes by AM cannot be considered as extremes in a physical or hydrological meaning.

3.2.1.GPD Approximation

Statistically, the distribution of the POT series $E_1, \ldots, E_{N_u}$,can be determined by making use of the GPD which is a cdf $G_{\gamma,\sigma}$ defined, for $x \in S(\gamma,\sigma) := [0,\infty)$ if $\gamma \geq 0$ and $[0, -\sigma/\gamma)$ if $\gamma < 0$, by:

$$G_{\gamma,\sigma}(x)=\begin{cases}1-\left(1+\gamma\dfrac{x}{\sigma}\right)^{-1/\gamma}, & \gamma\neq 0,\\ 1-e^{-x/\sigma}, & \gamma=0,\end{cases}\quad(1)$$

where $\gamma \in R$ and $\sigma>0$ are respectively shape and scale parameters[31].

Let $F_u(x)=P(Q-u\le x|Q>u)$ denote the excess cdf of Q over a given threshold u. Then, we have the following result**:**

$$\lim_{u\to q_F}\sup_{0<x<q_F-u}\left|F_u(x)-G_{\gamma,\sigma(u)}(x)\right|=0, \quad (2)$$

where q_F is the right end point of the cdf F. This result, due to Balkema andde Haan[4] and Pickands [48],is one of the most useful concepts in statistical methods for extremes. It says that for large threshold u, the excess cdf F_u is likeley to be well approximated by a GPD.

3.2.2.Threshold Selection

In order to obtain the asymptotic result in(2), the threshold *u* should be large enough which has as a consequence a satisfactory GPD approximation. The choice of the threshold is a crucial issue in the POT procedure. Indeed, selecting a threshold that is too low results in a large bias in the estimation, whereas taking one that is too high yields a big variance[24, section 6.4 and 6.5]. Hence, a compromise between bias and variance is to be found. To this end, one can minimize the asymptotic mean squared error, which is composed by the bias and variance. Furthermore, several graphical procedures are available to select u, such as the mean residual life (MRL), threshold choice (TC) and dispersion index (DI) plots. On the other hand, the choice of *u* can be based on physical considerations, e.g. by identifying the flood level of the river of interest. For a survey of the main selection procedures, see e.g. the paper of Lang et al [40].

3.2.Exploratory Data Analysis

The first step allows to check the data quality and to screen the data to avoid outlier effects. It also permits to obtain prior information, e.g. the shape, regarding the distribution to be selected. The presence of outliers in the data can have an important effect and causes difficulties when fitting a distribution[3]especially on the distribution upper part. The Grubbs and Beck[25] statistical test, based on the assumption of normality data, is designed to detect low and high outliers. In the case where the original data are not normal, they should be appropriately transformed. According to Section 1.8.3 in[49], this test is based on the following quantities:

$$x_H = \exp\left(\bar{x} + k_n s\right), \tag{3}$$

$$x_L = \exp\left(\bar{x} - k_n s\right), \tag{4}$$

where $\bar{x}$ and s are respectively the mean and standard deviation of the natural logarithms of the sample, and k_n is the Grubbs-Beck statistic tabulated for various sample sizes and significance levels by Grubbs and Beck [25]. For instance, at the 10% significance level, the following approximation is used

$$k_n = -3.62201 + 6.28446 n^{1/4} - 2.49835 n^{1/2} + 0.491436 n^{3/4} - 0.037911 n, \tag{5}$$

where n is the sample size.

The observations greater than x_H are considered to be high outliers, while those less than x_L are taken as low outliers.

3.3.Testing Independence, Stationarity and Homogeneity

Three basic assumptions are required to correctly apply FA of extreme hydrological events, namely independence, stationarity and homogeneity of the data[11]. To verify these assumptions, three tests are widely used in the literature. The Wald-Wolfowitz test is employed for the independence, the homogeneity test of Wilcoxon is applied to check whether the data come from the same distribution or not and the Mann-Kendall test allows to verify stationarity of the data, i.e. the series does not present a trend over time. These three tests have the advantage of being non-parametric and are widely used in hydrological FA. In other words, they do not require any prior knowledge on the distribution of the data.

3.4.Parameter Estimation and Model Selection

The choice of the appropriate model is one of the most important issues in FA. In practice the distribution of hydroclimatic series is not known. Using the fitted probability distribution, it is possible to predict the probability of exceedance for a specified magnitude, i.e. quantile, or the magnitude associated with a specific exceedance probability. To estimate the parameters associated to the appropriate probability distribution, popular techniques are used in hydrology, including the methods of Maximum Likelihood (ML)[e.g. 17, 46], Moments (MM) and Probability Weighted Moments (PWM) [e.g. 13, 32]. The latter is equivalent to the L-moment method which is widely used in hydrological FA[29].

The choice of the adequate distribution is determined on the basis of numerous classical and recent statistical tools, including graphical representations [34, 46] and goodness-of-fit tests such as the tests of Pearson (Chi-squared, Chi2), Kolmogorov-Smirnov (KS), Cramer-von Mises and the normality specific Shapiro-Wilk (SW) test. Due to the importance of the distribution impact in FA, these tools should be exploited. This point is widely studied in the literature [8, 20, 22, 28, 31, 38 and 47].

Nonetheless, the decision procedures mentioned above are not perfectly suited for extreme value distributions, because they are not sensitive enough to deviations in the tails. Several transformations have been proposed to overcome the limitations of the aforementioned tests [36, 39, 41]. In our application, where we focus on the upper tail of the distribution, we perform the Anderson-Darling k-sample test (k=2) implemented in the adk package of the statistical software R. This procedure is used to test the null hypothesis that k samples come from one common continuous distribution. In our case, the first sample of size 42 is the considered POT series and the second one consists in values generated from the GPD model. For more details on this test, we refer to [53].

The probability distributions that are appropriate for hydrology data are those with heavy tails. A number of them are listed, e.g. in [37, 49, 52]. In order to select the appropriate distribution among those which passed the goodness-of-fit tests, one or more criteria are required. To this end, one can consider the Akaike and Bayesian information criterion (AIC, BIC) respectively proposed by Akaike [2] and Schwartz [54]. They are given by:

$$AIC = -2\ln L + 2k, \tag{6}$$

$$BIC = -2\ln L + 2k\ln m \tag{7}$$

where L is the likelihood function, k the number of parameters and $m$the sample size. The best fit is the one associated with the smallest criterion AIC or BIC values [20, 28, 49].

3.5.Quantile Estimation

Once the appropriate distribution selected, the quantiles and return periods can be evaluated. The quantile estimation for various recurrence intervals is the main goal in hydrological practice. The notion of return period for hydrological extreme events is commonly used in FA, where the objective is to obtain reliable estimates of the quantiles corresponding to given return periods of scientific relevance or government standard requirements[49]. In the FA context the uncertainty decreases with the sample size whereas it increases with the return period when estimating quantiles.

In many environmental applications the sample size is rarely sufficient to enable good extreme quantiles estimations. Usually, a quantile of return period *T* can be reliably estimated from a data record of length *n* if $T<n$. However, in many cases, this condition is rarely satisfied –since typically $n<50$ for hydrological applications based on annual data[31].

3. Results and discussion

The application of the presented methodology in section 3 to the data described in section 2 leads to the following results, obtained by means of the packages stats, evir and POT of the statistical software R [33] and also by using the HYFRAN-PLUS software[21].

4.1.Exploratory Analysis and Outlier Detection

From Figure 2, it appears that the whole daily data series vary from a minimum value of $0\ m^3/s$ corresponding to many dry days, to a maximum value of $78.57\ m^3/s$ recorded on September 21st, 1989. The average flow of $0.39\ m^3/s$ is a relatively low in comparison with other tributary wadis of Chott Melghir like El Hai wadi and Djamorrah wadi [42]. The standard-deviation of $2.48\ m^3/s$ yields a coefficient of variation equal to 6.39 .The box-plot in Figure 3clearly shows the existence of extreme values. Indeed, the median ($0.10\ m^3/s$) is close to both 25th and 75th percentiles ($0.04\ m^3/s$ and $0.20\ m^3/s$). In addition to this graphical consideration, the values of skewness and kurtosis ($20.51\ m^3/s$ and $498.59 m^3/s$ respectively) eliminate the Gaussian model. In particular the very large value of the kurtosis indicates longer and fatter distribution tails, urging us to focus on heavy-tailed models

From **Erreur ! Source du renvoi introuvable.**, we observehigh inter-annual and the short sample size (resulting from selection AM) which leads to selecting low discharges during the driest years whereas some interesting discharges were not selected during the years where several floods have occurred. This explains the non-relevance of the AM approach for Abiod wadi data analysis and suggests that the POT approach would be more appropriate, and would lead to a more homogeneous sample of extreme discharges. This method starts with the selection of a convenient threshold, then the consideration of the observations that exceed this threshold.

In order to detect outliers, the quantities x_H and x_L are found to 508.31 and 0.08 respectively. Since there is no value greater than x_H and nor less than x_L, we conclude that, at the significant level of 10 %, no outlier exist among the excesses. Since it is difficult to use the outlier detection test with the analysis of extremes and due to the lack of regional weather data, the significance level to 10% is considered.

4.1.1.Threshold Selection

In this study, we adopt one of the available graphical tools, namely the TC-plot. From Figure 4we can choose a threshold value $u = 5.6\, m^3/s$, which results in an excess series of size 58.However, as recommended by many authors [9, 40, 56], this data set must be reduced in order to avoid the effects of dependence. We eliminated the peaks being obviously part of the same flood, and in order to keep the character of flood seasonality, we retain three peaks per year over the recorded period. Thus, the length of the data series becomes 42. Figure 5shows the distribution of these excesses and Table 1summarizes their elementary statistics.

The positive skewness coefficient Cs=1.62 reveals that the data is right skewed relative to the mean excess, as shown in **Erreur ! Source du renvoi introuvable.**a. In Figure 5a, the data are arranged by classes, of length10 m^3/s each, with the associated frequencies. It can be seen that some values are more frequent than others and that the majority of excesses have a low value varying between 0 and 10 m^3/s. **Erreur ! Source du renvoi introuvable.**b, where the data are arranged according to the months of appearance, shows that the peaks generally occur in the fall season.

4.2.Testing the Basic FA Assumptions

The results of the required hypothesis testing on the considered data are given in Table 2. Applying Wilcoxon, Kendall and Walf-Wolfowitz tests respectively, we conclude that the homogeneity, stationarity and independence of the excesses are accepted at any of standard significance levels (1%, 5% and 10%).Note that for the homogeneity test, we split the data in two sub-series 1972-1981 and 1982-1994 (any other subdivision led to the same conclusion). The homogeneity can also be noted in Erreur ! Source du renvoi introuvable.a where there is only one mode (the highest frequency).

4.3.Model Fitting

To fit a statistical distribution, we consider three commonly used estimation methods of the GPD parameters (ML, MM, and PWM). Then we perform the Anderson-Darling test to check the goodness-of-fit of the model. The results are summarized in Table 3. In view of the large P-values, we deduce that the GPD can be accepted as an appropriate model for the excess at any standard significance level (for instance 5%).

To discriminate between the obtained models, we use the AIC and BIC criteria. The last two columns of Table 3as well as Figure 6bfavor the GPD of the ML fitting method. We illustrate the goodness of fit of the excesses to this model in **Erreur ! Source du renvoi introuvable.**a. Furthermore, this ML-based will be used for quantile estimation in the following section.

Note that the ML and PWM results are very similar whereas those of the MM results are slightly different but remain in the same range.

4.4.Quantile Estimation

The estimation of extreme quantiles for different return periods should take into consideration the record period and the right tail of the distribution. The formally gauged record represents a relatively small sample of a much larger population of flood events. Thus, the extrapolation for long return periods is less accurate. In the M'chouneche station only the following return periods were considered for the estimation of quantiles: 2, 5, 10, 20 and 50 years as presented in Table 4. The return period of the strongest streamflow in the 1972-1994 period, equal to 78.57 m^3/s, is estimated by means of Pareto's fitted model to be 30.62 years.

The confidence interval is a way to assess the uncertainty in the estimation of the distribution parameters and quantiles. For the GPD, the confidence bounds are obtained through asymptotic results [31]. In the present case-study, one can see from **Erreur ! Source du renvoi introuvable.**c that the GPD agrees with the observations for return levels less than 30 but not beyond even though they are all included in the confidence interval. This is probably due to the small number of peaks over the chosen threshold. Therefore, it is important to consider this distribution with care with return periods greater than 30 years. This point indicates the issue of the quantity of the required data in this station for better estimation of high return periods.

4. Conclusions

The study of the Algerian wadis floods remains a quasi-unknown field as only some very specific indications are given in the Algerian hydrological directories. Floods are one of the basic features of a stream regime. The present study, which is the first one carried out in southern east of Algeria with new data series, in the context of FA. Mean daily discharges data recorded at the gauging station of M'chouneche in Abiod wadi, near Biskra, are available and considered in the present study. Due to the high inter-annual variability of the data as well as to the relatively short record length, the AM approach is not adapted to this analysis. Hence, in this paper, we considered a more appropriate procedure, namely the POT method.

The purpose of this work is to provide a suitable model for the excesses over a chosen threshold. This allows to estimate extreme flood events of given return periods. A complete FA was applied including appropriate tools, commonly used in hydrology. The issue of threshold selection was dealt by the means of a graphical tool. Several fitting methods have led to different GPD models and according to the results, the ML-based distribution was adopted. Because of the short record length, only return periods of 2, 5, 10, 20 and 50 years were considered. It was found that most of extracted data corresponded to frequent events. In the present case study, the GPD distribution provided good estimates of return periods less than 30 years but for higher values, the estimation is not acceptable and it is associated with high uncertainty (large confidence interval).

As a conclusion, we should emphasis that, in addition to the quality of data and sample size, the right choice of a GPD model heavily depends on the threshold. To improve the flood FA at this site, future studies should focus on the importance of data monitoring. However, this conclusion even it is not new in FA, it is important for the studied area where this is the first time to be studied. It emphasis and confirms the importance of this issue, especially for local government and decision makers.

Acknowledgments

The authors are thankful to the Editor and two reviewers for their constructive comments and suggestions. The authors express their gratefulness to the financial support of Canada's International Development Research Centre (IDRC). The data used in this study where provided by the National Agency of Water Resource of Algeria (ANRH).

References

[1] (IACWD), I. C. o. W. D. (1982) Guidelines for determining flood flow frequency: Bulletin 17B (revised and corrected). *Hydrol. Subcomm., Washington, D.C*, 28.

[2] Akaike, H. (1974) A new look at the statistical model identification. *Automatic Conrol, IEEE*, 19, 6, 716 - 723.

[3] Ashkar, F. and Ouarda, T. B. M. J. (1993) Robust estimators in hydrologic frequency analysis. In: Engineering Hydrology (edited by C.Y Kuo) *Am. Soc. Civ. Eng. New York, USA*, 347-352.

[4] Balkema, A. A. and de Haan, L. (1974) Residual Lifetime at Great Age. *Annals of Probability*, 2, 792-804.

[5] Bayazıt, M., Shaban, F. and Onoz, B. (1997) Generalized Extreme Value Distribution for Flood Discharges. *Turkish Journal of Engineering and Environmental Sciences*, 21, 2, 69-73.

[6] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., de Waal, D. and Ferro, C. (2004) *Statistics of Extremes: Theory and Applications*. Wiley.

[7] Benazzouz, M. T. (2010) Flash floods in Algeria: impact and management. G-WADI meeting. Flash Flood Risk Management Expert Group. *Meeting Cairo 25-27 September 2010, Egypt*.

[8] Benkhaled, A., Higgins, H., Chebana, F. and Necir, A. (2014) Frequency analysis of annual maximum suspended sediment concentrations in Abiod wadi, Biskra (Algeria). *Hydrological Processes,28,12*, 3841-3854.

[9] Beran, M. A. and Nozdryn-Plotnicki, M. J. (1977) Estimation of low return period floods. *Hydrological Sciences Bulletin*, 22, 2 , 275-282.

[10] Bobée, B. (1999) Extreme flood events valuation using frequency analysis. *A critical review. Houille Blanche*, 54, 100-105.

[11] Bobée, B. and Ashkar, F. (1991) *The gamma family and derived distributions applied in hydrology*. Water Resources Publications.

[12] Bouanani, A. (2005) *Hydrologie, Transport solide et Modélisation. Etude de quelques sous bassins de la Tafna*. Université de Tlemcen.

[13] Chebana, F., Adlouni, S. and Bobée, B. (2010) Mixed estimation methods for Halphen distributions with applications in extreme hydrologic events. *Stochastic Environmental Research and Risk Assessment*, 24, 3, 359-376.

[14] Chebana, F. and Ouarda, T. B. M. J. (2011) Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22, 1, 63-78.

[15] Chebana, F., Ouarda, T. B. M. J., Bruneau, P., Barbet, M., El Adlouni, S. and Latraverse, M. (2009) Multivariate homogeneity testing in a northern case study in the province of Quebec Canada. *Hydrological Processes*, 23, 1690-1700.

[16] Chow, V. T., Maidment, D. R. and Mays, L. W. (1988) Applied Hydrology. *McGraw-Hill, New York, NY*.

[17] Clarke, R. T. (1994) Fitting distributions. Chapter 4 Statistical modeling in hydrology. *Wiley*, 39-85.

[18] de Haan, L. and Ferreira, A. (2006) Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering. Boston.

[19] Dubief, J. (1953) Essai sur l'hydrologie superficielle au Sahara. *GGA, Direction du Service de la Colonisation et de l'Hydraulique, Service des Etudes Scientifiques, Alger*, 457.

[20] Ehsanzadeh, E., El Adlouni, S. and Bobee, B. (2010)Frequency Analysis Incorporating a Decision Support System for Hydroclimatic Variables. *Journal of Hydrologic Engineering*, 15, 11, 869-881.

[21] El Adlouni, S. and Bobée, B. (2010)Système d'aide à la decision pour l'estimation du risqué hydrologique.*IASH Publ*, 340.

[22] El Adlouni, S., Bobée, B. and Ouarda, T. B. M. J. (2008) On the tails of extreme event distributions in hydrology. *Journal of Hydrology*, 355, 16-33.

[23] Ellouze, M. and Abida, H. (2008) Regional Flood Frequency Analysis in Tunisia: Identification of Regional Distributions. *Water resources management*, 22, 8, 943-957.

[24] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) Modelling Extremal Events for Insurance and Finance. Springer.

[25] Grubbs, F. E. and Beck, G. (1972)Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14, 847-854.

[26] Haktanir, T. (1992) Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia. *Journal of Hydrology*, 136, 1-31.

[27] Hamel, A. (2009) *Hydrogéologie des systèmes aquifères en pays montagneux à climat semi-aride. Cas de la vallée d'Oued El Abiod (Aurès)*. Université, Mentouri : Constantine.

[28] Hebal, A. and Remini, B. (2011)Choix du modèle fréquentiel le plus adéquat à l'estimation des valeurs extrêmes de crues (cas du nord de L'Algérie). *Canadian Journal of Civil Engineering*, 38, 8, 881-892.

[29] Hosking, J. R. M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*, 52, 105–124.

[30] Hosking, J. R. M. and Wallis, J. R. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29, 339-349.

[31] Hosking, J. R. M. and Wallis, J. R. (1997) Regional frequency analysis; an approach based on L-moments. *Cambridge University Press*, 224.

[32] Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985) Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics*, 257-261.

[33] Ihaka, R. and Gentleman, R. (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5, 3, 299-314.

[34] Institute of Hydrology (IH). (1999) *Flood Estimation Handbook*. Wallingford, U.K, 1999.

[35] Karim, M. A. and Chowdhury, J. U. (1995) A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrological Sciences Journal*, 40, 1, 55-66.

[36] Khamis, H. J. (1997) The delta-corrected Kolmogorov-Smirnov test for the two-parameter Weibull distribution. *Journal of Applied Statistics*, 24, 3, 301-318.

[37] Kite, G. W. (1988) Flood and Risk Analyses in Hydrology. *Water Resources Publications, Littleton, Colorado, USA*, 187.

[38] Koutsoyiannis, D. (2003) On the appropriateness of the Gumbel distribution for odeling extreme rainfall. *In: Brath, A., Montanari, A., Toth, E. (Eds.), Hydrological Risk: Recent Advances in Peak River Flow Modelling, Prediction and Realtime Forecasting. Assessment of the Impacts of Land-use and Climate Changes. Editoriale Bios, Castrolibero, Bologna; 303–319.*

[39] Laio, F. (2004) Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distribution with unknown parameters. *Water Resour. Res, W09308*, 40.

[40] Lang, M., Ouarda, T. B. M. J. and Bobée, B. (1999) Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225, 103-117.

[41] Liao, M. and Shimokawa, T. (1999) A new goodness-of-fit test for type-I extreme-value and 2-parameter weibull distributions with estimated parameters. *Journal of Statistical Computation and Simulation*, 64, 1, 23-48.

[42] Mebarki, A. (2005) *Hydrologie des basins de l'Est Algérien*. Université de Constantine.

[43] Mutua, F. M. (1994)The use of the Akaike Information Criterion in the identification of an optimum flood frequency model. *Hydrological Sciences Journal*, 39, 3, 235-244.

[44] National Research Council (NRC) (1988) *Estimating probabilities of extreme floods: methods and recommended research*. National Academy Press.

[45] Natural Environment Research Council (NERC) (1999) Flood Studies Report. *Wallingford: Institute of Hydrology*, In five volumes.

[46] Natural Environment Research Council (NERC) (1975) Flood studies report. *London*, 1.

[47] Ouarda, T. B. M. J., Ashkar, F., Bensaid, E. and Hourani, I. (1994) *Statistical distributions used in hydrology. Transformations and asymptotic properties*. University of Moncton.

[48] Pickands, J. (1975) Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 1, 119-131.

[49] Rao, A. R. and Hamed, K. H. (2000) *Flood Frequency Analysis*. City.

[50] Rasmussen, P. F., Ashkar, F., Rosbjerg, D. and Bobée, B. (1994) The POT method for

flood estimation: a review. In: Hipel, K.W. (Ed.). Stochastic and Statistical Methods in Hydrology and Environmental ngineering, Extreme Values: Floods and Droughts, Kluwer Academic, Dordrecht. *Water Science and Technology Library*, 1, 15-26.

[51] Reiss, R. D. and Thomas, M. (2007) *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser Verlag GmbH.

[52] Salas, J. D. and Smith, R. (1980) Computer Programs of Distribution Functions in Hydrology. *Colorado State. University, Fort Collins, Colorado, USA.*

[53] Scholz, F. W. and Stephens, M. A. (1987) K-sample Anderson-Darling Tests, *Journal of the American Statistical Association*, Vol 82, No. 399, 918–924.

[54] Schwartz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics,* 6, 461-464.

[55] Tanaka, S. and Takara, K. (2002) A study on threshold selection in POT analysis of extreme floods. *IAHS-AISH publication*, 271.

[56] Todorovic, P. and Zelenhasic, E. (1970) A Stochastic Model for Flood Analysis. *Water Resources Research,* 6, 6, 1641-1648.

[57] Tramblay, Y., St-Hilaire, A. and Ouarda, T. B. M. J. (2008) Frequency analysis of maximum annual suspended sediment concentrations in North America. *Hydrological Sciences Journal*, 53, 1, 236-252.

[58] Vogel, R. M., McMahon, T. A. and Chiew, F. H. S. (1993) Floodflow frequency model selection in Australia. *Journal of Hydrology*, 146, 421-449.

[59] Willeke, G. E., Hosking, J. R., Wallis, J. R. and Guttman, N. B. (1995) The national drought atlas (draft), IWR Rep *US Army, Corps of Eng*, 94-NDS-4.

List Of Tables

List Of Figures

Table 1. Statistics summary of excess data set.

Size	42 observations
Minimum	0.02 m3/s
Qu_1(25^{th} percentile)	3.36m3/s
Median	7.83m3/s
Average	15.72m3/s
Qu_2(75^{th} percentile)	19.92m3/s
Sd	19.70 m3/s
Maximum	72.97m3/s
Cs	1.62
Ck	4.48

Table 2.Stationarity, independence and homogeneity tests results.

Tests	Statistic value	p-value
Stationarity (Kendall)	0.48	0.63
Independence (Wald-Wolfowitz)	0.94	0.35
Homogeneity(Wilcoxon)	0.79	0.43

Table 3. GPD parameter estimation, Anderson-Darling goodness-of-fit test and information criterion results.

Estimation method	**Scale**	**Shape**	**Statistic value (**Anderson-Darling**)**	**p-value**	**AIC**	**BIC**
ML	10.19	0.39	-0.55	0.49	315.68	326.63
MM	12.86	0.18	-0.83	0.58	316.61	327.56
PWM	10.10	0.36	-0.86	0.59	315.72	326.68

Table 4. Estimated quantiles of excess flows from the ML-based GPD.

Return period (years)	**Estimated quantile (m3/s)**
2	8.11
5	22.80
10	37.96
20	57.82
50	93.82

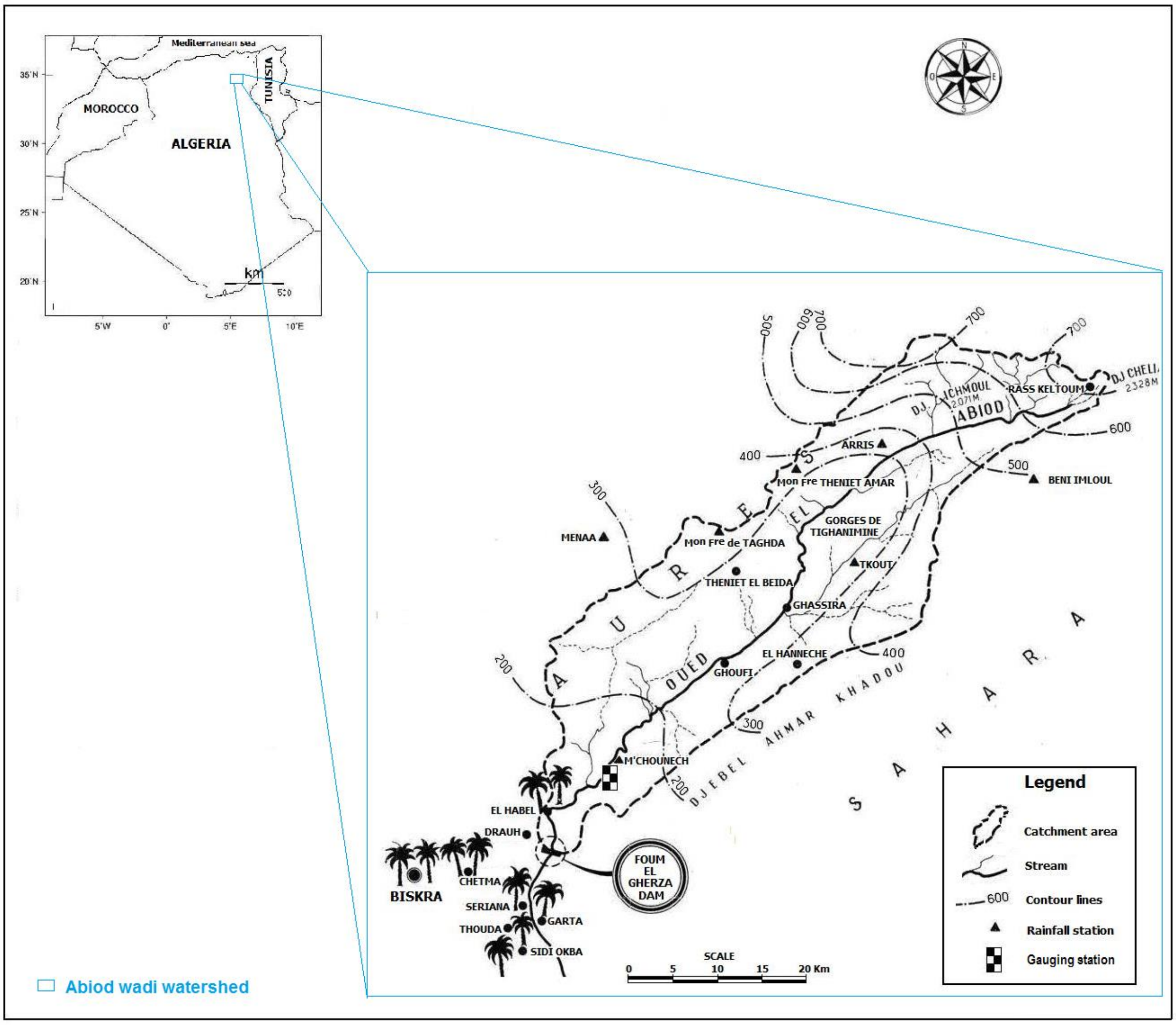


Figure 1.Geographical location of the Abiod wadi watershed

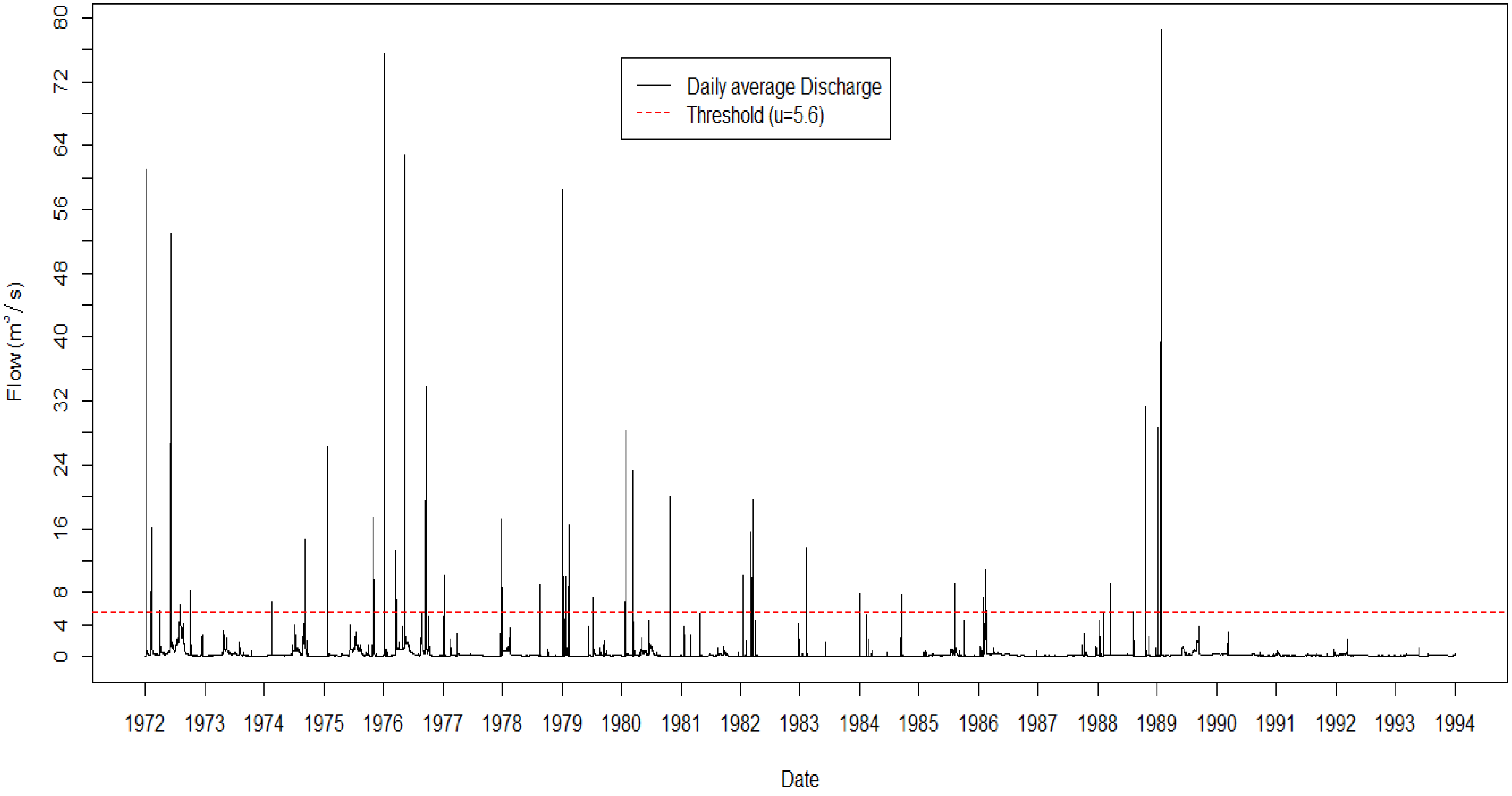


Figure 2. Time series plot of the daily average discharge at M'chouneche station covering the period 01/09/1972-31/08/1994.

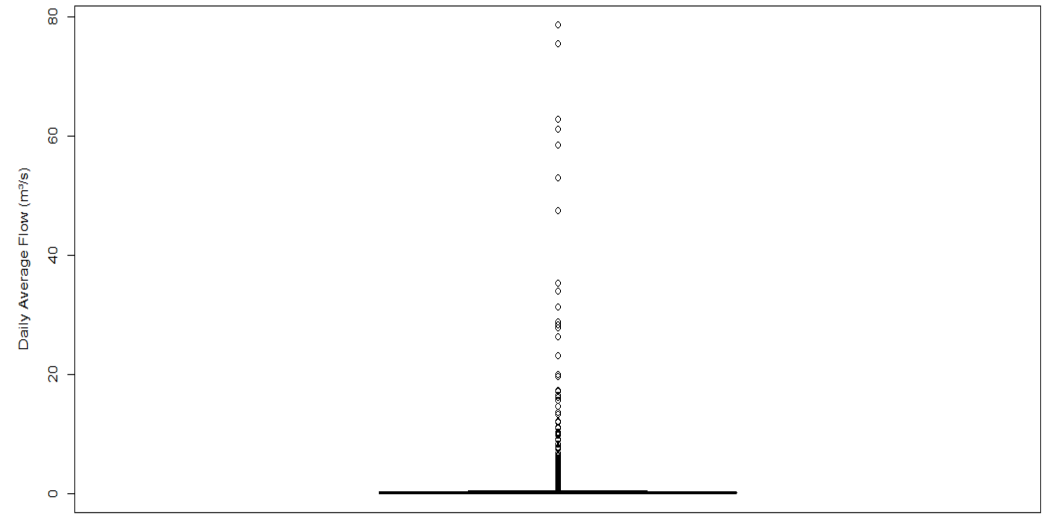


Figure 3.Box plot of daily average discharge at M'chouneche station.

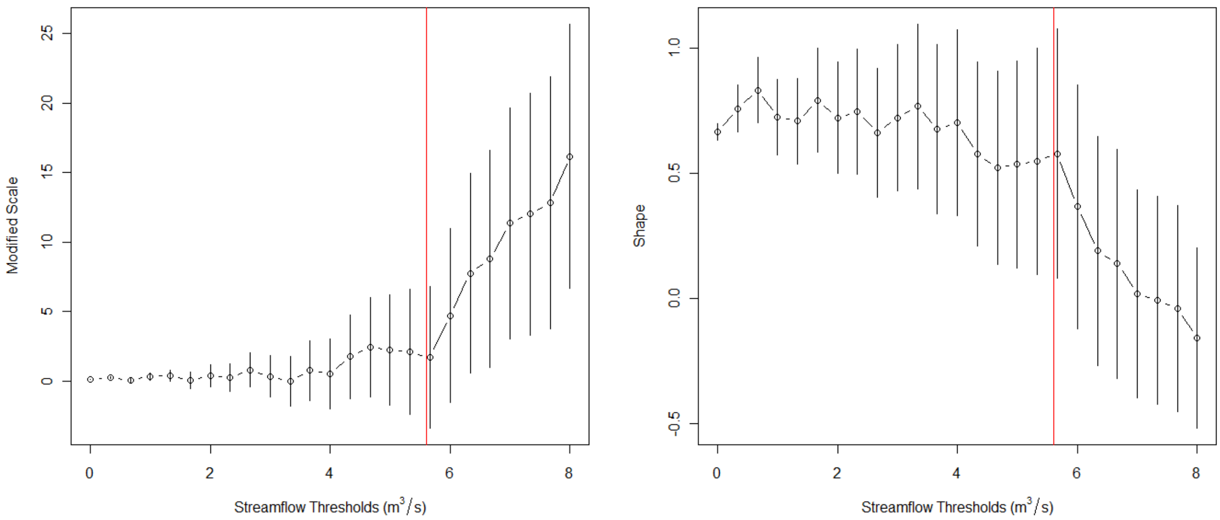
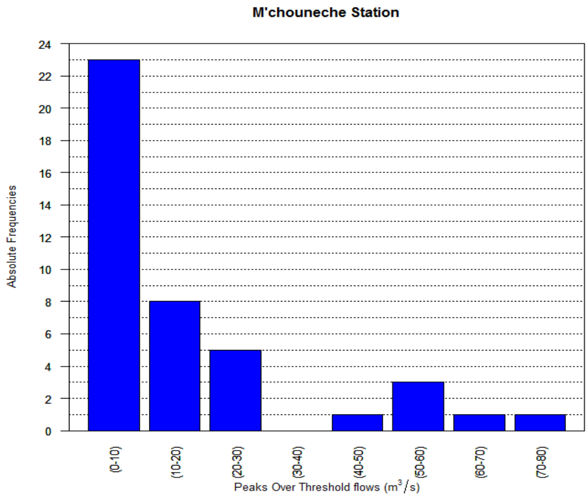
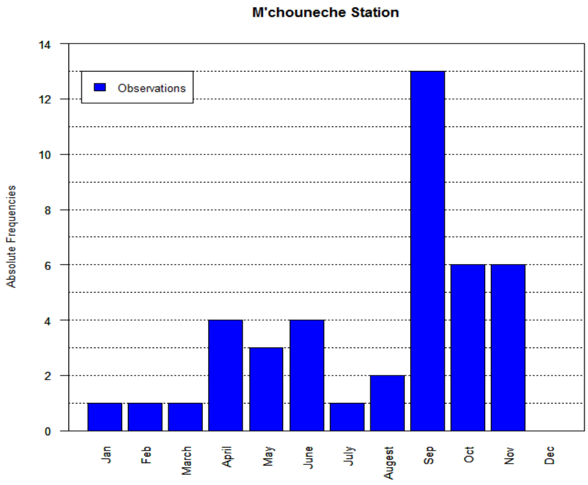


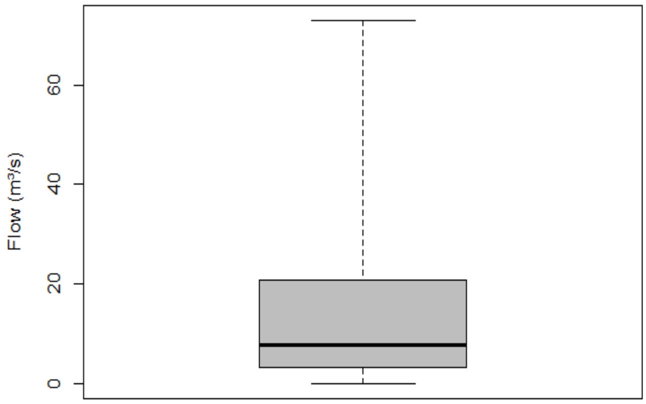
Figure 4.Graphical results of threshold selection applied for daily average discharge of Abiod wadi at M'chouneche station (tc-plot), vertical line corresponding to the threshold.



a)

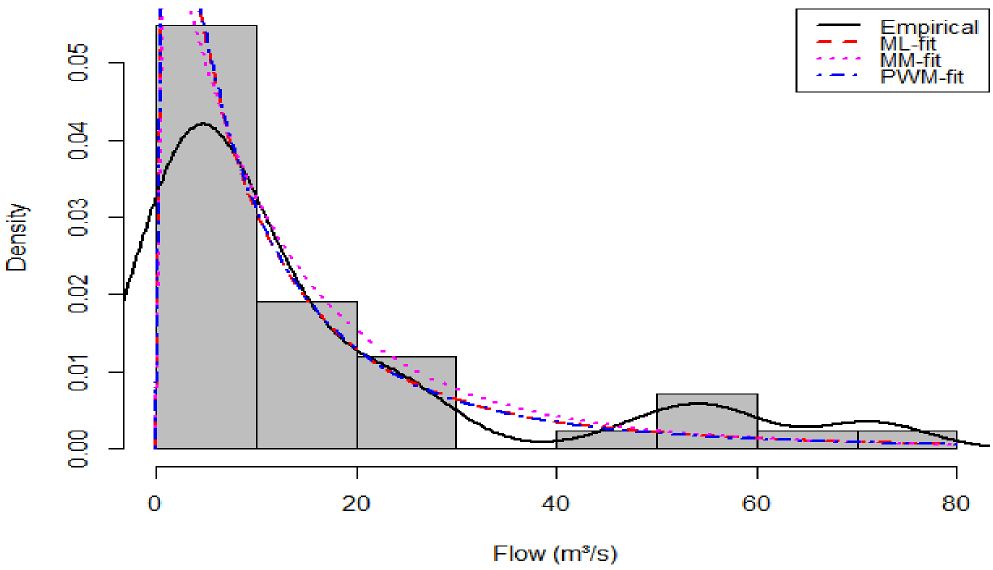


b)

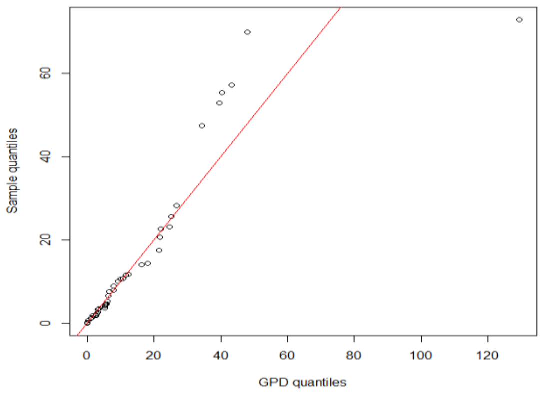


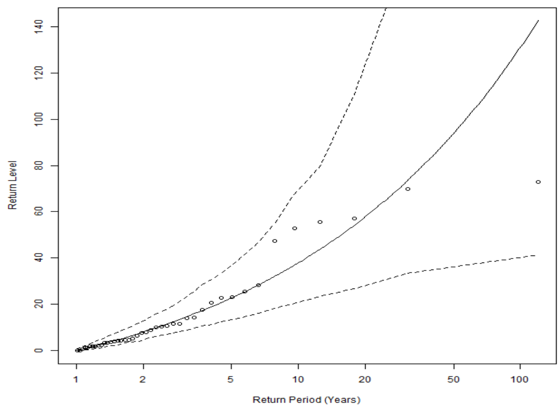
c)

Figure 5. Distribution of excess series at M'chouneche station a) Histogram by flow classes b) Histogram by month c) boxplot.



a)





b) c)

Figure 6. Best fitted distributions of excess flows at M'chouneche station a) distributions b) qq plot of ML-based GPD c) Return level plot (95% confidence interval)