# Study and Statistical Modeling of Extreme Temperatures in Northeastern Algeria

**Abderrahmane Ziari[1*], Abderrahmane Medjerab[1]**

[1]USTHB, FSTGAT, Climatology Laboratory, Bab-Ezzouar Algiers, Algeria, 16000

***Corresponding Author: Email:** ziariabderrahmane@gmail.com

Received: 12/05/2023; Accepted: 11/12/2023;

## Abstract

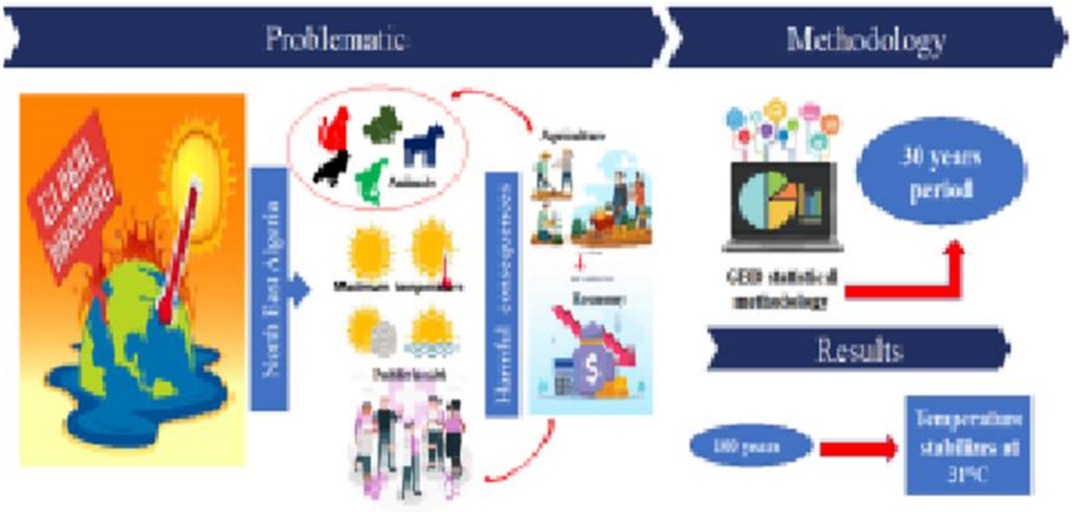
According to the (GEC2021), global warming could have a very serious impact on many sectors such as health, the environment and the economy. Scientists are highlighting the changes in intensity and frequency observed in the past, and attempting to predict the future. This is not climate change, but rather the risk of a worsening of its intensity. To this end, this article aims to identify temperature maxima in northeastern Algeria. To do this, we applied the Generalized Pareto Distribution (GPD) to data from the chosen weather station over a period of more than thirty years, in order to understand the behavior of maxima. Our study showed that whatever the return period, the monthly mean remains around 31°C.

**Keywords:** General Pareto distribution, extreme temperatures, global warming, northeastern Algeria, POT.

*Tob Regul Sci.™ 20 - -;9(2):* 2145 - 2156

*DOI: doi.org/10.18001/TRS.9.2.139*

Graphic abstract



## 1. Introduction

Algeria is a region severely affected by climate change, with temperatures rising faster than the global average and rainfall becoming increasingly erratic. These changes will have a significant impact not only on the region's ecosystems, but also on agriculture and human health. Climate projections indicate that average temperatures in the region will rise by 1.5 to 2.5°C by the end of the century (GEC 2021), increasing the frequency and intensity of extreme weather events such as droughts, heat waves, floods and storms.

These changes have had significant impacts, including limited water availability in many regions. Prolonged drought can affect agriculture, drinking water supplies and energy production. In addition, rising temperatures and changes in rainfall patterns can encourage the spread of diseases such as malaria, and the spread of certain agricultural pests. Faced with these challenges, it is imperative to develop strategies for adapting to climate change. The aim is to provide political authorities and populations with the most accurate and objective information possible, enabling them to implement mitigation strategies in the face of climatic events and adaptation strategies of populations and economies in the face of climate change.

To do this, we applied the Generalized Pareto Distribution (GPD) to data from selected weather stations to understand the behavior of maxima and make predictions over a 100-year period.

## 2-Methodology

Research on the theme of climate, disasters and risks in Algeria is initially marked by the study of water-related phenomena: precipitation and its spatio-temporal evolution Medjerab.A. (2007), Habibi.B.(2013), Other types of climatic hazards (those related to temperature, storm phenomena) have received less attention from researchers. Research on these hazards is relatively new and their number is still low, noted that the few studies conducted so far are often limited to a single type of hazard and that few studies address To this end, we study monthly maximum temperatures by applying the generalized Pareto distribution (GPD) (Cheratia 2021) to data from the Patna region. The temperature data used in this article represent the average monthly temperatures recorded over many years (1988 to 2015) at the Batna meteorological station in northern Algeria. The aim is to understand changes in maximum temperatures and build suitable predictive models that can help meteorologists and authorities understand these specific events and thus prevent climate risks. The packages zyp (Bronaugh and Werner, 2013), evd (Stephenson, 2002), extRemes (Gilleland and Katz, 2011), (Stephenson, 2014), (R core Team, 2015) were used for data analysis. A 100-year prediction was therefore made.

## 2-Presentation of the study area

### 2-1 Geographical location

Batna, the capital of the Aurès region, is located in eastern Algeria, 430 kilometers from Algiers. Geographically located in the eastern part of the country, the wilaya of Batna lies between 4° and 7° longitude and 35° and 36° north latitude. It covers an area of 12,038.76 km2. The territory of the wilaya of Batna is almost entirely formed by the confluence of the Tellian and Saharan Atlases; this is the main physical feature of the wilaya that determines its climate and human living conditions.

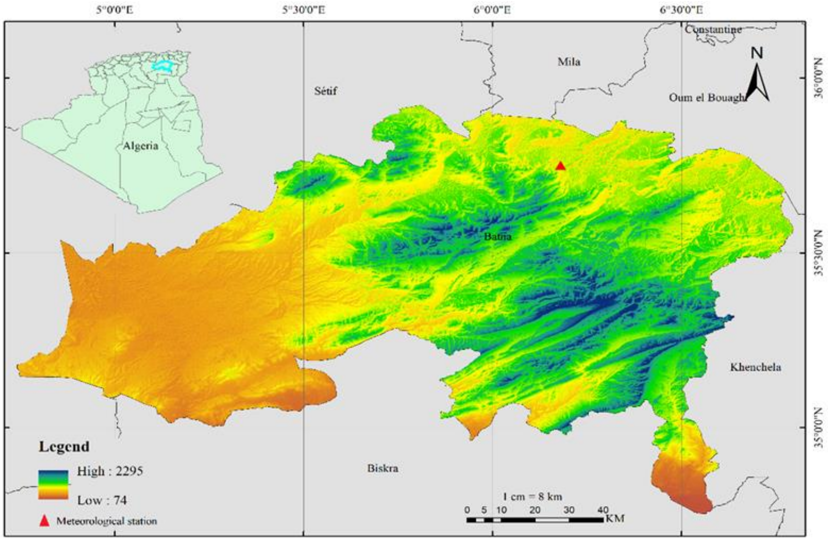



**Figure 1 : General information on Batna's climate**

Batna's bioclimatic regime is semi-arid. Based on the variance that characterizes its relief, climatic analysis reveals three distinct rainfall zones. The first is the humid zone, with annual rainfall of 900 to 1,200 mm and mountains and peaks of over 1,800 meters. Annual rainfall in the medium zone ranges from 400 to 800 mm and corresponds to the northern slopes of the massif. The arid zone, with annual rainfall of 200 to 400 mm, lies to the west and south of the massif and includes the entire Beni Imloul forest. In the shade, temperatures can reach 45°C in summer, while in winter, temperatures can reach 45°C (Castro-Camilo, 2021).

## 3-Modeling using the Generalized Pareto Distribution (GPD)

The first step is to determine an optimal threshold for this distribution, the Generalized Pareto Distribution GPD (Resnick 1987) presents the parameter estimation method for the statistical analysis of extreme temperature values.The temperature information used in this article is based on average monthly temperatures observed at the Batna weather station in northern Algeria over many years (from 1988 to 2015).

### 3.1 Choice of threshold

To analyze extreme maximum temperatures using the POT method, we first determine a threshold value, then adjust the GPD to higher temperature values, thanks to the work of J.

Pickands (1975), AC Davison (2003) and R.L. Smith (1989), GNEDENKO BV (1943) and others.

This technique is new and uses the Balkema-de Haan-Pickands theorem (1974), Makarov, M. (2006), to model distribution tails. It determines a threshold above which the model can be used. It is sensitive to independence assumptions and data perturbations. A threshold method is used to select a sample of independent peaks (Figure 1). It is assumed that peak values above a given threshold u are distributed according to a generalized Pareto distribution (GPD), and that the process by which peaks appear is poissonian.

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (1)$$

Where λ is a positive constant called the law parameter

The GPD distribution is written for a certain threshold u. The values obtained by the parameters provide information on the tail weights of the parent distribution. In other words, the higher the tail index, the thicker the tails of the distribution under consideration. Consequently, a tail index greater than 0 means that the probability of an extreme event occurring is greater than would be predicted by the normal distribution.

The Mean Excess ME Plot is defined as follows:

$$\{(u, e_n(u)), X_{n:n} < u < X_{1:n}\} \quad (2)$$

Where $X_{n:n}$ $et$ $X_{1:n}$ are respectively the maximum and minimum of the sample and is defined by :

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)^+}{\sum_{i=1}^{n} I_{(Z_i > u)}} = \frac{1}{N_u} \sum_{i=1}^{n}(X_i - u)^+ \quad (3)$$

Represents the sum of excesses above the threshold u divided by the number of data that exceed is the empirical estimator of the Mean Excesss ME function :

$$e(u) = \mathrm{E}[X - u\mathrm{I}\, x > u] \quad (4)$$

For GPD, the ME , is as follows:

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \quad (5)$$

Where σ ⁺ So, if the ME-plot seems to have a linear behavior above a certain value of u, it means that the excesses above this threshold follow a GPD

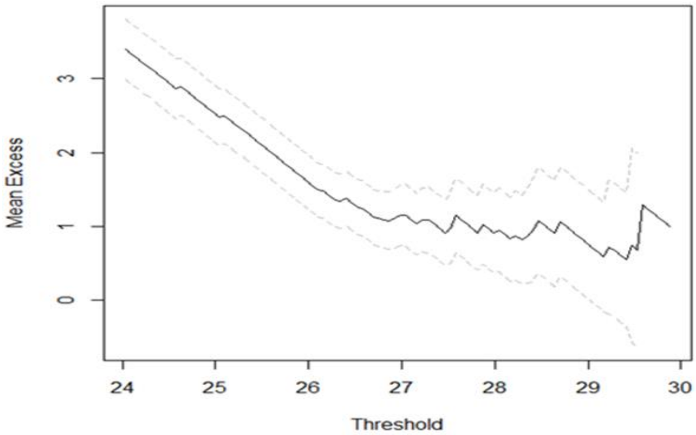
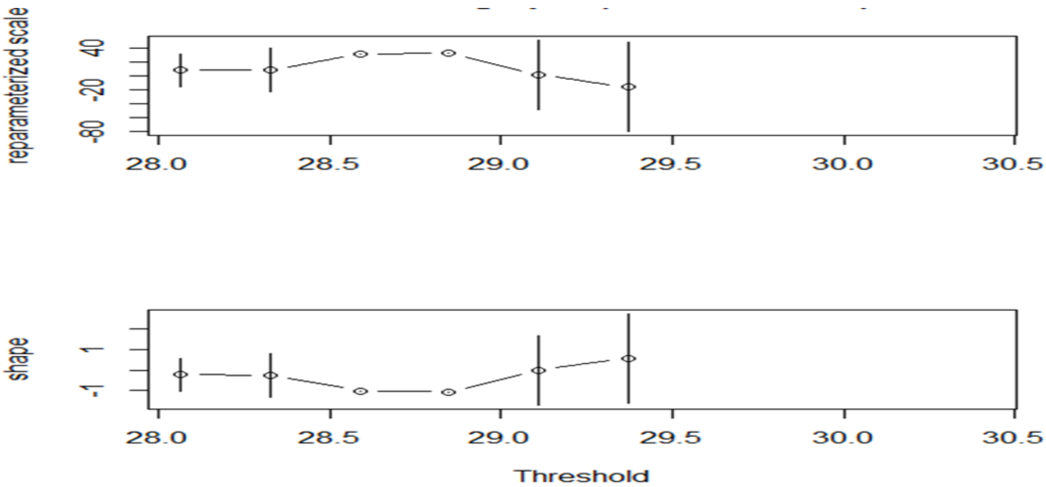



**Figure2: Average monthly maximum temperature exceedance curve at Batna weather station**

Interpreting the plot of the mean excess function is not very straightforward, and selection bias may be a problem. This means that considering only values above a threshold can bias the estimates. By using different threshold ranges, we can better manage this bias by adjusting the distribution of different parts of the tail. By fitting GPD to different threshold ranges, we can explore and identify the optimal threshold that provides the most accurate estimate for our case, and better understand the distribution of extreme values in our data.

To refine the appropriateness of our choice, we will use the method of plotting scale and shape parameters against different thresholds, also known as "stable scale and shape parameters" to determine the thresholds required when fitting the data to a multi-row GPD distribution.

A different threshold is set each time. This allows us to control and localize the stability of the parameters (shape and scale). The method is implemented using the Extremes package of the R statistical analysis system software (Gilleland et al., 2004). This package contains objective tools for determining thresholds simply by examining the stability of the shape and scale parameters ε and σ, we obtained the following results (figure2).



The idea is to find the threshold interval where the graph of the stability points of the two parameters is approximately linear and stable, on these graphs it appears that the two parameters of scale and shape remain relatively constant beyond a threshold of 25C°.

### 3.2. Descriptive data analysis

The range of temperature values above the chosen threshold is shown in Figure 3, and its statistical properties are shown in Table 1. Shows that the station's mean temperature is 27.52°C. We can see that 50% of the data fall between 25.06°C

and 26.85°C.

What's more, the skewness is less than zero and the kurtosis is different from 3, so the hypothesis of normality of the monthly maximum temperature series is rejected.

| N | $N_u$ | Mean | $3^{rd}$Qu | Max | $1^{st}$Qu | Min | Kurtosis | Skewness |
|---|---|---|---|---|---|---|---|---|
| 36 | 34 | 27.52 | 28.14 | 30.88 | 26.85 | 25.06 | 0.74 | 0.70 |

**Table1. Descriptive statistics.**

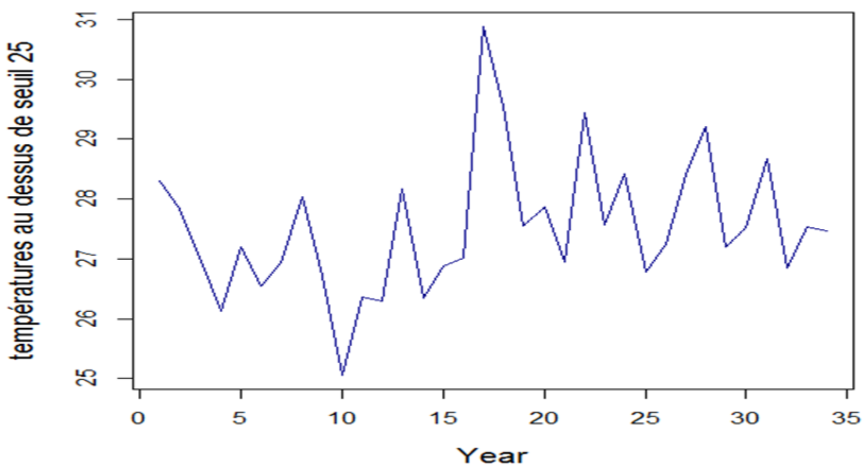



**Figure 3: Maximum monthly temperature above 25°C at Batna station.**

From the graph, it appears that the series of maxima is not affected by a trend; the series appears stationary.

### 4-Parameter estimation and model validation

Once the excess values have been identified, the GPD parameters can be estimated from these data. Parameters include a shape parameter, which characterizes the shape of the distribution of extreme values, and a scale parameter, which measures the deviation of the extreme values from the selected threshold. Parameters were estimated using ML maximum likelihood. Let's assume that our sample of excesses is identically distributed with the GPD given by :

$$GPD_u(x) = 1 - \left[1 - \frac{\xi(x-u)}{\tilde{\sigma}}\right]^{\frac{-1}{\xi}} \quad \xi \neq 0 \qquad (6)$$

The density function g of is then for

$$\frac{d}{dx} GPD_u(x) = g(x) \qquad (7)$$

$$g(x) = \frac{1}{\sigma}\left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}-1} \qquad (8)$$

The likelihood is given by $\mathcal{L}(\xi, \mu, \sigma, X) = \prod_{i=1}^{n} g(X_i)$ (9)

The log likelihood is given by :

$$l(\xi, \sigma, X) = \ln \mathcal{L}(\xi, \sigma, X) \qquad (10)$$

$$\iota(\xi, \delta; X) = -N_u \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{1}^{N_u} \ln\left(1 + \frac{\xi}{\sigma} X_i\right) \qquad (11)$$

Deriving this function in

$$\frac{\partial \iota(\xi, \sigma; X)}{\partial \xi} = 0 \qquad (12)$$

$$\frac{\partial \iota(\xi, \sigma; X)}{\partial \sigma} = 0 \qquad (13)$$

We obtain a system with two equations and two unknowns, the solution being the Maximum Likelihood estimators ʒ

And for ʒ=0, we have

$$g(x) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) \qquad (14)$$

$$l(0, \sigma; X) = -N_u \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^{N_u} X_i \qquad (15)$$

We then obtain the estimator $\hat{\sigma}_{N_u} = \sum_{i=1}^{N_u} X_i / N_u$ (16)

which is simply the empirical average of the excesses

Scale and shape parameter estimates, associated 95% confidence intervals and covariance matrices are presented in Table2.

| | Scale | Shape |
|---|---|---|
| Estimates | 3.88 | -0.64 |
| Std.err | 0.74 | 0.14 |
| CI | (2.42 , 5.34) | (-0.91 ,-0.36) |
| Estimated parameters covariance matrix | | |
| Scale | 0.55 | -0.10 |
| Shape | | 0.019 |

Table 2. ML estimates, confidence intervals (CI) and covariance matrices for shape and scale parameters of the GPD model fitted for maximum monthly temperature

Table2 shows that the shape parameter is negative (meaning that the GP distribution is of the Pareto II Resnick (1987) type). Its value is close to zero, meaning that the exponential distribution is excluded. Confidence intervals to support this conclusion. The QQ plot technique was used to validate the selected models. It can be seen that the QQ plot (figure 5) for the station is almost linear. Shows that the GPD type II model with a threshold of25° C is suitable for the maximum monthly temperature at the Batna weather station.

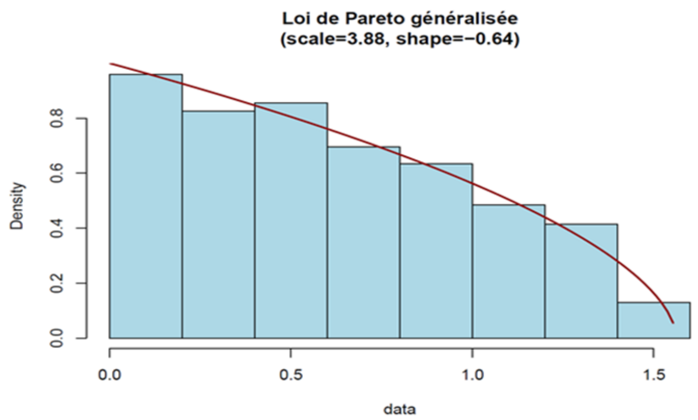



**Figure 4: Estimated distribution.**

## 4-1 Validation of the selected model

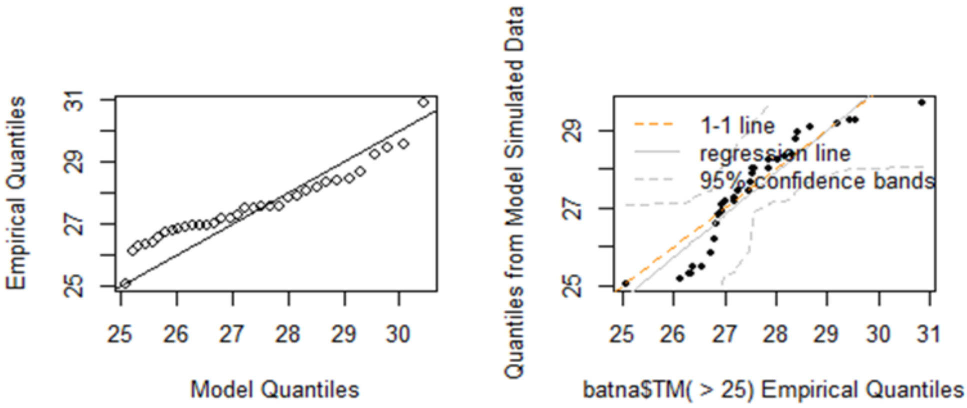



**Figure 5: QQ plot for Batna station**

The purpose of the two graphs is to examine Henry's straight line and the deviations of the observations. We note that all the points represented are aligned, forming a linear-shaped cloud, so the annual temperature maxima for the Batna station are well fitted by GPD.

### 4.2 Estimated GPD return

Table 3 shows the return rates estimated using the ML method for different monthly maximum temperature return periods with 95% profile probability confidence intervals (CI). The return level of is the upper quantile of the distribution. These correspond to values that are expected to be exceeded every "n" years on average, where "n" is the inverse of the probability of exceedance.

For example, a yield level of 100 years means that you can expect to exceed this value on average once every 100 years. This will help assess and manage the risks associated with rising temperatures.

So we can expect to exceed 30.96°C every 2 years.

Similarly :

31.03°C every 20 years

31.045°C every 50 years

31.049°C every 100 years

### 4-3 Return levels

| Return period | Estimated return level (in °C) |
|---|---|
| 2-years | 30.96 |
| 20-years | 31.03 |
| 50-years | 31.045 |

| | |
|---|---|
| 100-years | 31.049 |

**Table 3. Monthly maximum temperature return levels and 95% CI (°C) using GPD.**

## Conclusion

In this study, monthly maximum temperatures at the Batna weather station from 1988 to 2018 were modeled using the generalized Pareto distribution (GPD) to control and predict maximum temperature behavior.

Using maximum likelihood (ML) to estimate the parameters, we find that the Pareto II type (confined tail) with a threshold of 25°C is more suitable for the Batna weather station.

Return rates are estimated for several return periods, with temperatures stabilizing at around 31°C after a 2-year period.

For example, the model suggests that it will take around 100 years for average monthly temperatures to stabilize at a value of 3.049 °C.

This indicates that whatever the time of return, the average monthly temperature is likely to exceed 31°C, confirming the dryness and persistence of drought conditions in the region, in relation to global warming, which will increase water stress and therefore the vulnerability of a region where the issue of water is problematic.

## Acknowledgments

The authors would like to thank the National Office of Meteorology and the National Water Resources Agency for providing the data.

## Declaration of interest statement

The authors declare no conflict of interest and they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Bibliography

[1] 1-Balkema, A. A., and de Haan, L. (1974). Residual life time at great age. The Annals of Probability 2(5), 792–804.

[2] 2-Bronaugh, D., and A. Werner. (2013) zyp: Zhang + Yue-Pilon trends package. R package

[3] 3-Castro-Camilo, D., Huser, R., Rue, H. (2021).: Practical strategies for GEV-based regression models for extremes. arXiv preprint arXiv: 2106.13110 https://doi.org/10.48550/arXiv.2106. 13110

[4] 4-Cheraitia Hassen (2021) « Modélisation de la température mensuelle maximale 1988 à 2018 » Revue d'Economie et de Statistique Appliquée, Volume 18, Numéro 1, Pages 45-55. https://www.asjp.cerist.dz/en/article/156834

[5] 5-Desvina, A.P., Safitri, E., Rahma, A.N. (2019): Statistical modelling of extreme data of air pollution in Pekanbaru City. InPrime: Indon. J. Pure Appl. Math. 1(1), 48–56

[6] 6-Davison AC, Smith RL.(1990),« Models for exceedances over high thresholds (withdiscussion) », Journal of the Royal Statistical Society, Series B, PP. 393–442

[7] 7-Gilleland E. Katz R W (2011)., « New software to analyze how extremes change over time», Transactions of the American Geophysical Union, 92 (2), 2011,PP. 13–14.

[8] 8-Gnedenko BV (1943)., « Sur la distribution limite du terme maximum d'une série aléatoire», Ann. Math. 44,PP. 423–453.

[9] 9-Gumbel EJ. « Statistics of Extremes », Columbia University Press, New York, 1958

[10] 10-Habibi, B., Meddi, M., & Boucefiane, A. (2013). Analyse fréquentielle des pluies journalières maximales Cas du Bassin Chott-Chergui. Nature & Technology, (8), 41B.

[11] 11-Khan, M., Hussain, Z., Ahmad, I. (2021).: Effects of L-moments, maximum likelihood and maximum product of spacing estimation methods in using Pearson type-3 distribution for modeling extreme values. Water Res. Manag. 35(5), 1415–1431 https://doi.org/10.1007/s11 269-021-02767-w

[12] 12-Kotz, S., Nadarajah, S. (2000). "Extreme Value Distributions: Theory and Applications." London: Imperial College Press.

[13] 13-Makarov, M. A. (2006). On permutations generated by infinite binary words. Sibirskie Èlektronnye Matematicheskie Izvestiya [Siberian Electronic Mathematical Reports], 3, 304-311.

[14] 14-Medjerab A. (2007) - Les situations pluviométriques extrêmes dans le nord-ouest de l'Algérie. In : Climat, tourisme, environnement, Actes du XXème colloque de l'Association Internationale de Climatologie (Carthage, Tunisie), H. BEN BOUBAKER édit., p. 381-386.

[15] 15-Castro-Camilo, D., Huser, R., Rue, H. (2021- Évaluation des changements climatiques au Maghreb. Étude du cas des régions du quart nord-est algérien. In : Risques et changements climatiques, Actes du XXIIIème colloque de l'Association Internationale de Climatologie (Rennes), p. 463-468.

[16] 16-Pickands, J. (1975). "Statistical inference using extreme order statistics", Annals of Statistics, 3, 119–131.

[17] 17-R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[18] http://www.R-project.org/

[19] 19-STEPHENSON AG.(2002) « evd: Extreme value distributions. R News», URL: http://CRAN.R-project.org/doc/Rnews/, PP.641–644.