A Bayesian analysis of the annual maximum temperature using generalized extreme value distribution

CHERAITIA HASSEN

Department of Mathematics, University of Mohamed Seddik Ben Yahia, Jijel, Algeria

(Received 6 August 2020, Accepted 15 June 2021)

e-mail: hassen.stat@hotmail.fr

ABSTRACT. The annual maximum temperature was modeled using the Generalized Extreme Value (GEV) distribution to Jijel weather station. The Mann-Kendall (MK) and Kwiatkowski Phillips, Schmidt and Shin (KPSS) tests suggest a stationary model without linear trend in the location parameter. The Kurtosis and the Skewness statistics indicated that the normality assumption was rejected. The Likelihood Ratio test was used to determine the best model and the goodness-of-fit tests showed that our data is suitable with a stationary Gumbel distribution. The Maximum Likelihood estimation method and the Bayesian approach using the Monte Carlo method by Markov Chains (MCMC) were used to find the parameters of the Gumbel distribution and the return levels were obtained for different periods.

JEL Classification: C1, C13, C46, C490.

Key words – Generalized Extreme Value (GEV) distribution, Gumbel distribution, Maximum Likelihood estimate (ML), Markov Chains Monte-Carlo method (MCMC), Maximum temperature, Return level.

1. Introduction

The Extreme Value Theory (EVT) provides a statistical framework to make inferences about the probability of very rare and extreme events. It is based on the analysis of the maximum (or minimum) value in a selected period. The modern theory of extreme value was developed between 1920 and 1940 by Fréchet (1927), Fisher and Tippett (1928); Gnedenko (1943) and Gumbel (1958). It finds application in many fields; for instance, it has been used to estimate the extreme levels of a river causing floods in hydrology: Coles and Tawn (2005); Renard et al. (2006) in France. Concerning precipitations and the risk of floods caused by this phenomenon, a considerable number of studies aiming to model such events can be found in the literature: Coles (2001); Friederichs (2010) in Germany, Kim et al. (2009) in the south of Korea, Boudriissa et al. (2017) in Algeria, Delson and Retius (2015) in Zimbabwe and Deka et al. (2011) in India. Other studies related to the risk of high temperature: Tadele et al. (2015) in South Africa and Husna Hasan et al. (2013) in Malaysia.

Different methods can be used to estimate the parameters of the extreme laws such as the Maximum Likelihood method [Coles (2001); Smith (1999)], the Weighted Moments method [Hosking (1990); Greenwood et al. (1979)] and semi-parametric methods [Pickands (1975); Hill (1975)]. The Bayesian approach offers another estimation methodology. This method is based on the processing of the unknown parameters as random variables and then obtains appropriate estimators using Markov Chains Monte Carlo methods (MCMC). Bayesian analysis is becoming increasingly popular in many fields including meteorology and has shown its practical benefits in several applied problems. As an
example, Coles and Tawn (1996) adopted a Bayesian methodology to model the oceanographic flood events on the UK east coast, Khodja et al. (1998) applied a multivariate test for a shift in the mean of rainfall data of different stations in West Africa with a Bayesian argument, Lu and Berliner (1999) considered a class of Bayesian dynamic models that involve switching among various regimes for a runoff time series, Perreault et al. (1999) presented a Bayesian approach which can be used to study a change in the mean level of a set of independent normal random variables, Thyer and Kuczera (2000) developed a hidden state Markov (HSM) model as a new approach for generating hydro-climatic time series, Wang (2001) used a Bayesian approach to estimate the parameter of a bivariate generalized extreme value distribution, Thyer et al. (2002) used a Metropolis Hasting algorithm to simulate the posterior distribution of a Box-Cox transformation, Coles et al. (2003) adopted a Bayesian approach to model a set of daily rainfall extremes data on the central coast of Venezuela, Parent and Bernier (2003) developed an inference procedure for the peak over threshold (POT) model using semi-conjugate informative priors, Renard et al. (2006) presented a Bayesian application to the regional frequency analysis of extremes in a nonstationary context. Coles and Powell (1996) reviewed the literature linking the themes of Bayesian and extreme value analysis, Smith (1999) compared the Bayesian and the frequentist approaches to study parametric predictive inference. The zyp (Bronaugh and Katz, 2011), evd (Stephenson, 2002), extRemes (Gilleland and Katz, 2011) and ismev (Stephenson, 2014) packages of R (R core Team, 2015) were used for the data analysis.

Jijel City (Fig. 1) is located in the northeastern region of Algeria with an area of 2396.63 km². It is bordered from the East by Skikda city, from the West by Bejaia city, from the North by the Mediterranean Sea and from the South by Setif and Mila cities. Jijel has a Mediterranean climate characterized by mild and rainy winters with annual precipitations of 814 mm, hot and sunny summers. The annual average temperature in Jijel is around 18 °C.

This study aims to use the block maxima approach by GEV distribution to model annual maximum temperature in Northern Algeria and tries to understand the changes in temperature caused by global warming which affect Algeria’s weather specifically Jijel city. Our objective is to investigate the temperature changes to forecast its extremely high levels that can occur in Jijel city and that authorities consider it to prepare a precautionary system against the catastrophes owing to high temperatures such as forest fires.

In addition to this introduction, the paper is organized as follows: the GEV distribution, the Maximum Likelihood estimates of its parameters, the Bayesian estimate and the return level are presented in Section 2. The theoretical model is applied to data in Section 3. Finally, some conclusions are given in Section 4.

2. Methodology

2.1. Generalized Extreme Value (GEV) distribution

The extreme value theorem provides a theoretical framework to model the distribution of extreme events and the three-parameter GEV is recommended for meteorology frequency analysis. The three parameters are location, scale and shape. The GEV distribution is a family of continuous probability distributions developed within the extreme value theorem. The GEV distribution arises from the limit theorem of Fisher and Tippet (1928) and Gnedenko (1943).

Supposing that $X_1, X_2, ..., X_N$ is a sequence of independent random variables from a common distribution function $F$. The order statistic $M_n = Max (X_1, X_2, ..., X_n)$ is the maximal value of the independent identically distributed (IID) random variables. The external type theorem states that if there exist normalizing constants ($a_n > 0$) and $b_n \in R$ such that:

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P\left(M_n \leq a_n x + b_n\right) \rightarrow F^n(a_n x + b_n) \rightarrow G(x) \text{ as } n \rightarrow \infty$$

where, $G$ is a non-degenerate distribution function. The distribution of $G$ belongs to the Gumbel, Frechet, or Weibull distribution. The combination of these three distribution families into a single-family model forms the GEV distribution as proposed by Jenkinson (1955) with cumulative function:

$$G(x, \mu, \sigma, \xi) =
\begin{cases}
\exp\left[-1 + \frac{x - \mu}{\sigma}\right]^{-\frac{1}{\xi}}, & \text{for } \xi \neq 0, 1 + \xi (x - \mu)/\sigma > 0 \\
\exp\left[-\exp\left(-\frac{x - \mu}{\sigma}\right)\right], & \text{for } \xi = 0
\end{cases}
$$

$$\mu, \sigma > 0 \text{ and } \xi \text{ are location, scale and shape parameters, respectively.}$$

$$(1)$$
By derivation of the distribution function specified in expression (1), the density function is given by:

\[
g(x, \mu, \sigma, \xi) = \begin{ cases}
1 + \xi \left( \frac{x - \mu}{\sigma} \right)^{\frac{1}{\xi} - 1} \exp \left[ - \left( 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right)^{\frac{1}{\xi}} \right], & \text{if } \xi \neq 0 \\
\exp \left[ - \frac{1}{\sigma} \left( x - \mu \right) \right] \exp \left[ - \frac{x - \mu}{\sigma} \right], & \text{if } \xi = 0
\end{ cases}
\]

The classical GEV; \( G(x, \mu, \sigma, \xi) \) model assumes that the three parameters of location, scale and shape are time-independent. However, if trends are detected in the data sample, the non-stationary case where parameters are expressed as covariates \((e.g., \text{time})\) should be considered.

2.2. Selection Period, model choice and Stationary Test

The data of extreme value are grouped into blocks of equal length \( n \). The maximum of each block forms a series of block maxima \( M_1, M_2, M_3, \ldots, M_n \) to be fitted with the GEV distribution. The selection of block size is critical as a too-small block can lead to a bias and if too large leading to a large estimation variance (Husna et al., 2018). In this study, data of maximum temperature are divided into annual blocks. Three models are considered: a stationary model \( M_1 = \text{GEV}_1 (\mu, \sigma, \xi) \) with constant parameters \( \mu, \sigma \) and \( \xi \), a non-stationary model \( M_2 = \text{GEV}_2 [\mu(t), \sigma(t), \xi(t)] \) with time as covariates in the location parameter: \( \mu(t) = \mu_0 + \mu_1 t \), \( \sigma(t) = \sigma(t) \) and \( \xi(t) \) are constants and stationary Gumbel model \( M_0 = \text{GEV}_0 (\mu, \sigma, 0) \) with \( \xi = 0 \) and \( \mu, \sigma \) are constants. The stationarity assumption must be checked before the GEV model is fitted. The following tests are proposed:

2.2.1. Mann Kendall test

The non-parametric Mann-Kendall test is used to determine if the values of a random variable follow a monotonic trend (Ryden, 2011). The null hypothesis states that no trend exists. This test does not conform to any particular distribution and is particularly useful if datasets have missing values (Husna et al., 2018).

2.2.2. Kwiatkowski Phillips, Schmidt and Shin (KPSS) test

The KPSS test is applied to the data to check the stationarity. The null hypothesis of the KPSS test is that there is no trend, while the alternative hypothesis is that there is a trend in the data.

2.3. Model diagnostic and validation tools

The goodness-of-fit of the model is assessed by the following methods:

2.3.1. The Likelihood Ratio test

The Likelihood Ratio (LR) test is used to compare the fit of the null model \( M_0 \) and the alternative model \( M_1 \) \((M_0 \text{ and } M_1 \text{ are a special case of } M_1)\). The LR test statistic defined as:

\[
\lambda = -2 \ln \left( \frac{L_0}{L_1} \right)
\]

where, \( L_0 \) and \( L_1 \) represent the Likelihood of the model \( M_0 \) and \( M_1 \) respectively. \( \lambda \) has a Chi-Square distribution with one degree of freedom (Boudrissa et al., 2017).

The null model \( (M_0) \) is preferred if \( \lambda \leq \chi^2_{1,0.05} = 3.8415 \). Otherwise, the alternative model \( (M_1) \) is more suitable.

2.3.2. Quantile-Quantile plot (Q-Q)

The (Q-Q) plot assesses the adequacy of a fitted distribution by comparing the \((1/n + 1)\) th quantiles deriving from the theoretical and empirical distributions. If the two distributions being compared are similar, the points in the Q-Q plot will approximately lie on the diagonal line.

2.3.3. The goodness of fit tests

These tests [Kolmogorov Smirnov (KS), Anderson Darling (AD) and Cramer Von Mises (CVM)] are often used for comparing an empirical distribution determined from a sample \( F_n(x) \) to a theoretical distribution \( F(x) \). The null hypothesis is \( F_n(x) = F(x) \) and the alternative hypothesis is \( F_n(x) \neq F(x) \).

The distributions of their statistics are the subject of the statistical tables \((KS, AD \text{ and CVM tables})\), which take into account the sample size and the accepted error \( \alpha \). It is then sufficient to compare the observed value with the appropriate value in the table.

2.4. Parameters estimates

Several methods have been introduced in the literature to estimate the parameters of the GEV distribution; for example, the method of Moments by Christopeit (1994), the L-Moments method (Hosking, 1990; Hosking and Wallis, 1997), for more details about this method see Hosking, 1990), the Bayesian method by
Smith and Naylor (1987); Lye et al. (1993); Coles and Tawn (2005) and the Maximum Likelihood method (Smith and Naylor, 1987) which is the most popular and has the advantage of allowing the addition to the fitting of co-variables such as trends, cycles, or physical variables (Katz and et al., 2002).

2.4.1. Maximum Likelihood method (ML)

Under the assumption that $X_1$, ..., $X_n$ are independent random samples having a GEV distribution, the log-Likelihood for the GEV parameters when $\xi \neq 0$ is:

$$l(x, \mu, \sigma, \xi) = -n \log(\sigma) - \sum_{i=1}^{n} \left[ 1 - \xi \left( \frac{X_i - \mu}{\sigma} \right) \right]^{1/\xi} - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^{n} \log \left[ 1 + \xi \left( \frac{X_i - \mu}{\sigma} \right) \right]$$

where, $\mu_i = \mu$ for stationary three parameters model $M_1 = \text{GEV}(\mu, \sigma, \xi)$ and $\mu_i = \mu_0 + \mu_t$ for non-stationary four parameters model $M_2 = \text{GEV}[\mu(t), \sigma, \xi]$.

We differentiated the log-likelihood of GEV to find a set of equations which we solved using numerical optimization algorithms. In the case $\xi > -0.5$, the usual properties of consistency, asymptotic efficiency and asymptotic normality hold (Delson and Retius, 2015).

In the same way, for $\xi = 0$, the Logarithm of the Likelihood function for stationary Gumbel model $M_0 = \text{GEV}[\mu, \sigma, 0]$ is given by:

$$l(x, \mu, \sigma) = -n \log(\sigma) - \sum_{i=1}^{n} \exp \left( -\frac{X_i - \mu}{\sigma} \right) - \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

Differentiating this function for the two parameters, the following system of equations is obtained

$$\begin{cases} n - \sum_{i=1}^{n} \exp \left( -\frac{X_i - \mu}{\sigma} \right) = 0 \\ n + \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \left[ \exp \left( -\frac{X_i - \mu}{\sigma} \right) - 1 \right] = 0 \end{cases}$$

2.4.2. Bayesian analysis of extreme value for GEVD

The Bayesian approach improves the estimation accuracy by assuming the parameter $\theta = (\mu, \sigma, \xi)$ as a random variable. Bayesian inference, based on the central idea of Bayes’ theorem which as follows (Choeng and Gabda, 2017):

$$\pi(\theta/x) = \frac{L(\theta/x)\pi(\theta)}{\int L(\theta/x)\pi(\theta) d\theta} \quad (2)$$

where, $x$ is the given observation. $L(\theta/x)$ denotes the likelihood function and $\pi(\theta)$ is the normal prior distribution. The denominator in equation (5) is treated as a normalizing constant so that the posterior distribution is integrated into one. Thereby, results

$$\pi(\theta/x) \propto L(\theta/x)\pi(\theta).$$

Markov chain Monte-Carlo is employed to solve the complex computational of the posterior distribution in equation (2). The Markov chain Monte Carlo techniques are applied in this paper to give Bayesian analyses of the annual maximum temperature data for Jijel weather station. Markov chain Monte-Carlo is a simulation technique that can be used to find the posterior distribution and to sample from it by constructing a Markov chain that has the target distribution as its stationary distribution. In this paper, the prior is constructed by assuming that there is no information available about the process (temperature) apart from the data. The annual temperature data have a GEV, i.e., $X_i \sim \text{GEV}[\mu, \sigma, \xi]$ and the parameters $\mu, \sigma$ and $\xi$ are treated as random variables for which we specify prior distributions. For specification of the prior, the parameterization $\phi = \log(\sigma)$ is easier to work with because $\sigma$ is restricted to be positive. The specification of priors enables us to supplement the information provided by the data. The prior density is (Delson and Retius, 2015):

$$\pi(\theta) = \pi(\mu, \phi, \xi) = \pi_\mu(\mu)\pi_\phi(\phi)\pi_\xi(\xi)$$

where, each marginal prior is normally distributed with large variances. The posterior distribution is given as:

$$\pi(\theta|x) = \pi(\mu, \phi, \xi/x) \propto L(\mu, \phi, \xi/x) \pi(\mu, \phi, \xi)$$

where,

$$L(\mu, \phi, \xi/x) = \frac{1}{\sigma^n} \exp \left[ -1 + \xi \left( \frac{X_i - \mu}{\sigma} \right) \right]^{1/\xi} \prod_{i=1}^{n} \left[ 1 + \xi \left( \frac{X_i - \mu}{\sigma} \right) \right]^{1/\xi}$$

is the likelihood with $\sigma$ replaced by $e^\phi$. 
2.5. Metropolis-Hastings algorithm

The Metropolis Hasting algorithm is a form of generalized rejection sampling. The proposal value $\theta^*$ for $\theta_{i+1}$ is generated from arbitrary rule $q(. / \theta_i)$. The Markov chain moves to $\theta^*$ with a specified acceptance probability. Specifically, let

$$
\alpha_i = \min\left\{1, \frac{\pi(\theta^*) q(\theta_i | \theta^*)}{\pi(\theta_i) q(\theta^* | \theta_i)}\right\}
$$

where, $q(\theta^* | \theta_i)$ is denoted as proposal distribution. The candidate is accepted if the probability is equal to $\alpha_i$. Otherwise, the Markov chain remains in the current state $\theta_i$. The steps involved can be illustrated into the following algorithm (Choeng and Gabda, 2017).

(i) Initialize $\theta_0$

(ii) In $i$ iteration

- Draw a candidate $\theta^*$ from proposal distribution $q(\theta^* | \theta_i)$
- Calculate the acceptance probability
- Draw $u \sim \text{Uniforme}(0, 1)$
  
  If $\alpha_i < u$ then Set $\theta_{i+1} = \theta^*$ Else $\theta_{i+1} = \theta_i$

(iii) Increment 1 and return to step 2

2.6. Return level estimate

The return level is defined as a level that is expected to be equaled or exceeded on average once every interval of time (T) with a probability of $P$. The return level $x_P$ for the return period $\frac{1}{P}$ is the quantile of order $1 - P$ of the GEV distribution for $0 < P < 1$. 

![Fig. 1. Location of Jijel city](image-url)
TABLE 1
Statistical properties of the annual maximum temperature at Jijel weather station

<table>
<thead>
<tr>
<th>N</th>
<th>Min</th>
<th>1st Qu</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>24.9</td>
<td>25.70</td>
<td>26.10</td>
<td>26.35</td>
<td>26.75</td>
<td>28.60</td>
<td>0.75</td>
<td>2.75</td>
</tr>
</tbody>
</table>

The return level \( x_P \) is given by:

\[
    x_P = \begin{cases} 
    \mu - \frac{\sigma}{\xi} \left(1 - \log(1 - P)\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\
    \mu - \sigma \log(1 - P), & \xi = 0 
\end{cases}
\]  

(10)

The Maximum Likelihood estimates of the return level \( x_P \) can be obtained using the ML estimates of \( \mu \), \( \sigma \) and \( \xi \).

3. Results and discussion

3.1. Data description

The temperature data used in our study correspond to 31 value of the annual maximum temperature from 1988 to 2018, measured at Jijel weather station. To analyze the extreme temperature statistically in Jijel city (Fig. 1) with the GEV method, the monthly temperatures are grouped into blocks of size one year and the maximum of each block forms a series of block-maxima.

3.2. Descriptive analysis

The analysis was based on the annual maximum temperature at Jijel weather station from 1988 to 2018. The series is given in Fig. 2 and its statistical characteristics are given in Table 1. According to Fig. 2, it seems that the series is not affected by a trend, it appears stationary. From Table 1, it is found that our data is characterized by a mean of 26.35 °C. We note that 50% of the data are between 25.70 °C and 26.75 °C. The Normality assumption is rejected for the series of the annual maximum temperature because the Skewness is greater than zero and the Kurtosis is different from 3.

3.3. Stationarity checking

It is necessary to test whether the trend exists in our data. We have two types of stationarity tests: The Mann-Kendall (MK) and the KPSS test. Those tests aim to check if data (annual maximum temperature) is stationary. Estimation of the trend by the least-squares method may not be appropriate, therefore Kendall’s tau test and the
Sen’s slope estimator are proposed to compute trend (Sen, 1968). The results of this analysis are presented in Table 2. The results show that the trend is not significant for the annual maximum temperature (KPSS p-value = 0.13 and Mann-Kendall p-value = 0.068).

### 3.4. Maximum Likelihood (ML) estimates of the parameters

The Maximum Likelihood method was used to estimate the three parameters of the GEV distribution. The ML estimates of the location, scale and shape parameters and the associated 95% confidence intervals are given in Table 3.

From Table 3, it is noted that the shape parameter $\xi$ is positive in both models $M_1$ and $M_2$ ($\xi_{M_1} = 0.024$, $\xi_{M_2} = 0.04$), this is the case of Frechet distribution. Its value is near to zero implying that Gumbel distribution is candidate, the confidence interval of $\xi$ confirms this conclusion $0 \in CI_{\xi}$.

### 3.5. Model selection : the Likelihood Ratio (LR) test

The LR test aims to compare the stationary GEV model $M_1$ with the non-stationary one $M_2$ at the first time and with stationary Gumbel model $M_0$ at the second time. The results are presented in Table 4.

As we compare the result of model $M_1$ versus model $M_2$ and Model $M_0$ versus Model $M_1$, it can be seen that the Gumbel Model $M_0$ is preferred for Jijel weather station because of lowest observed Chi-Squared $LR_{(M_1,M_2)} = 3.74$ and $LR_{(M_0,M_1)} = 0.02$.

### 3.6. Model validation

This section aims to validate the selected Gumbel model $M_0$. In this step, the goodness of fit criteria is taken into account by using a graphical method (the Quantile-Quantile plot) and analytical methods (the non-parametric tests).

---

**TABLE 3**

ML estimation of the parameters and theirs confidence intervals

<table>
<thead>
<tr>
<th>Models</th>
<th>ML estimate</th>
<th>$\mu_0$</th>
<th>$\mu_1$</th>
<th>$\sigma$</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary GEV model</td>
<td>Estimates</td>
<td>-</td>
<td>25.90</td>
<td>0.73</td>
<td>0.024</td>
</tr>
<tr>
<td>$M_1$: GEV((\mu, \sigma, \xi))</td>
<td>Std.err</td>
<td>-</td>
<td>0.15</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>-</td>
<td>(25.6, 26.2)</td>
<td>(0.5, 0.95)</td>
<td>(-0.29, 0.33)</td>
</tr>
<tr>
<td>Non stationary GEV model</td>
<td>Estimates</td>
<td>25.49</td>
<td>0.03</td>
<td>0.68</td>
<td>0.04</td>
</tr>
<tr>
<td>$M_2$: GEV t((\mu(t), \sigma, \xi))</td>
<td>Std.err</td>
<td>0.25</td>
<td>0.01</td>
<td>0.1</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>(24.99, 25.98)</td>
<td>(0.0, 0.05)</td>
<td>(0.48, 0.89)</td>
<td>(-0.25, 0.34)</td>
</tr>
<tr>
<td>Gumbel model</td>
<td>Estimates</td>
<td>-</td>
<td>25.91</td>
<td>0.73</td>
<td>-</td>
</tr>
<tr>
<td>$M_0$: GEV((\mu, \sigma, 0))</td>
<td>Std.err</td>
<td>-</td>
<td>0.14</td>
<td>0.10</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>-</td>
<td>(25.64, 26.18)</td>
<td>(0.53, 0.94)</td>
<td>-</td>
</tr>
</tbody>
</table>

**TABLE 4**

The likelihood ratio test

<table>
<thead>
<tr>
<th>Models</th>
<th>Observed Chi-Squared</th>
<th>Table Chi-Squared</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1 - M_2$</td>
<td>3.74</td>
<td>3.84</td>
<td>0.053</td>
</tr>
<tr>
<td>$M_0 - M_1$</td>
<td>0.02</td>
<td>3.84</td>
<td>0.87</td>
</tr>
</tbody>
</table>
3.6.1. The Quantile-Quantile (QQ) plot

To validate the chosen model, the Quantile-Quantile (QQ) plot technique (Fig. 3) was used. It can be seen that for the annual maximum temperature the QQ plot is approximately linear; showing that the stationary Gumbel model $M_0$ is adequate for the annual maximum temperature at Jijel weather station.

3.6.2. The Goodness of fit tests

The results of the non-parametric tests (Kolmogorov Smirnov (KS), Anderson Darling (AD) and Cramer Von Mises (CVM)) are resumed in Table 5.

For the 5% significant level, the calculated values of the 03 tests are less than all the critical ones. This leads to the decision of non-rejection of the null hypothesis. We conclude that the annual maximum temperature follow the specified stationary Gumbel model $M_0$.

3.7. Bayesian estimates of the parameters for the stationary Gumbel model $M_0$

The MCMC method was applied to the annual maximum temperature data. The GEV scale parameter was re-parameterised as $\sigma = \log \phi$ to retain the positivity of this parameter. The prior density was chosen to be:

$$\pi_\mu(\mu) \sim N(0,10000)$$
$$\pi_\phi(\phi) \sim N(0,10000)$$
$$\pi(\mu,\phi) \sim \pi_\mu(\mu)\pi_\phi(\phi)$$

The proposal distribution function is:

$$q(\theta_i | \theta_{i-1}) \sim N(0,0.1)$$

A MCMC Metropolis Hastings random walk algorithm was applied such as:

$$\mu' = \mu_i + \epsilon_\mu, \quad \epsilon_\mu \sim N(0,0.02)$$
$$\phi' = \phi_i + \epsilon_\phi, \quad \epsilon_\phi \sim N(0,0.1)$$

To check that the chain will converge to the correct place, the ML estimate of the three parameters was used as starting points $(\mu_0 = 25.91, \phi_0 = 0.73)$. 10000 iterations

### TABLE 5

<table>
<thead>
<tr>
<th>Tests</th>
<th>Kolmogorov Smirnov test (KS)</th>
<th>Anderson Darling test (AD)</th>
<th>Cramer Von Mises test (CVM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs. Value</td>
<td>0.078</td>
<td>0.27</td>
<td>0.037</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.99</td>
<td>0.67</td>
<td>0.72</td>
</tr>
<tr>
<td>Crit. Value</td>
<td>0.23</td>
<td>0.75</td>
<td>0.21</td>
</tr>
</tbody>
</table>

### TABLE 6

<table>
<thead>
<tr>
<th>Model</th>
<th>Bayesian estimate</th>
<th>$\mu_1$</th>
<th>$\sigma$</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel model</td>
<td>Estimates</td>
<td>25.90</td>
<td>0.83</td>
<td>-</td>
</tr>
<tr>
<td>$M_0$: GEV$_0$(μ, σ, 0)</td>
<td>Std.err</td>
<td>0.203</td>
<td>0.18</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>(25.53, 26.33)</td>
<td>(0.55, 1.25)</td>
<td>-</td>
</tr>
</tbody>
</table>
Fig. 4. Trace plots of the Gumbel distribution parameters using non-informative priors for annual maximum temperature at Jijel weather station (1988-2018)

Fig. 5. Posterior densities of the Gumbel distribution parameters using non-informative priors for annual maximum temperature at the Jijel weather station

Figs. 6(a&b). Return levels of the annual maximum temperature: (a) ML return levels estimates and (b) Bayesian return levels estimates
TABLE 7  
Return levels estimates

<table>
<thead>
<tr>
<th>Method / Period</th>
<th>ML return levels</th>
<th>Bayesian Return levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( x_p ) (in °C)</td>
<td>CI</td>
</tr>
<tr>
<td>2-year</td>
<td>26.18 (25.87, 26.48)</td>
<td>26.21 (25.80, 26.72)</td>
</tr>
<tr>
<td>20-year</td>
<td>28.10 (27.35, 28.84)</td>
<td>28.70 (27.50, 32.23)</td>
</tr>
<tr>
<td>50-year</td>
<td>28.78 (27.86, 29.71)</td>
<td>29.87 (27.94, 36.87)</td>
</tr>
<tr>
<td>100-year</td>
<td>29.30 (28.23, 30.36)</td>
<td>30.93 (28.20, 42.17)</td>
</tr>
<tr>
<td>200-year</td>
<td>29.81 (28.60, 31.02)</td>
<td>32.20 (28.20, 49.27)</td>
</tr>
</tbody>
</table>

of the algorithm were carried out with a time of burn-in = 500. The results are resumed in Table 6. Fig. 4 shows the Markov chain Monte Carlo trace plots and the estimated posterior densities for the Gumbel parameters are given in Fig. 5.

3.8. Return level estimates

The estimation of the T-year return levels for \( T = 2, 20, 50, 100 \) and 200 are estimated using both ML and Bayesian methods with 95% confidence intervals (CI) as shown in Table 7 and Figs. 6(a&b). It can be seen from this Table that the return levels for maximum annual temperature and theirs confidence intervals increase slowly for higher return periods in both estimation methods.

Note also that estimated return levels for Jijel weather station are slightly differ for the two estimation methods; the Bayesian return level estimates are consistently greater than ML ones for all return periods.

4. Conclusions

In this study, the annual maximum temperature from 1988 to 2018 was modeled using stationary Gumbel distribution. The stationarity of our data was approved by the Mann-Kendall and KPSS tests which showed a non-significant trend in the data. This result was confirmed by applying the likelihood ratio test to compare the stationary GEV versus the non-stationary one. Model diagnostics that included QQ plot and goodness fitting tests (KS, AD and CVM) showed that the annual maximum temperature follows a stationary Gumbel distribution. We rely on an alternative method of Bayesian MCMC based on the Metropolis-Hastings algorithm to estimate the parameters in order to predict different levels return and theirs coming periods. The Gumbel parameter estimates using the Bayesian approach were close to the maximum likelihood estimates, with larger standard deviations. Bayesian return level estimates show that it will be a high annual temperature in the next years, it is approximately 30.93°C for 100 years-return level. The Algerian government should take the measure of prevention of the risk that can be caused by this expected pick of temperature by setting up an alert and vigilance plan. The use of an informative prior or using a pick over high threshold approach POT based on the generalized Pareto distribution GPD may improve the work.

Acknowledgement

We would like to thank everyone who helped us. We thank the editor and the anonymous reviewers for their valuable comments which improved the earlier version of the manuscript. We would also like to thank the Algerian Office of Meteorology for providing the data.

Disclaimer

The contents and views expressed in this research paper/article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References


HAHSEN: A BAYESIAN ANALYSIS OF THE ANNUAL MAXIMUM TEMPERATURE


