

ROBUST EXTREME RETURN LEVEL WITH POWER NORMALIZATION FOR EXTREME EVENTS: APPLICATION OF REAL HYDROLOGY DATA

Abdellah Belhajjam^{*}, Belbachir Mohammadine, and Saad Elouardirhi

Received: July 26, 2021; Revised: June 11, 2023; Accepted: June 19, 2023

Abstract

In statistical studies of rare and catastrophic phenomena the distribution of generalized extreme values under linear normalization is always chosen as the appropriate model. It used to estimate the probabilities of events that have not yet been observed. Recently, the extreme value theory (EVT) received a lot of attention both theoretically and practically using just the classical linear model (L-Model) or linear normalization of the maximum to estimate return level. So, in this paper we propose a new multiplicative model based on the distribution of generalized extreme values under non-linear normalization, whose purpose is to raise the strong and weak points between these two models. Our main goal is to use our multiplicative model (P-model) to calculate the return level, as well as the associated confidence interval. The diagnostic fit, test, and statistical inference to compare the two models (linear and non-linear) are studied. Finally, a data analysis and discussion are applied at first on real hydrological data for Morocco and South of Australia, then on water levels of lake Erié in Canada. The results show that our multiplicative (non-linear) model is more adaptive because it takes into account the variation of the return period.

Keywords: Extreme value theory; Generalized Extreme Value (GEV); linear normalization; non-linear normalization; Peak Over Threshold (POT)

Introduction

The Extreme Value Theory (EVT), appeared between 1920 and 1940, thanks to Fréchet (1927), Fisher and Tippett (1928), Gumbel (1958), and Gnedenko (1943), is playing an increasingly important role in the treatment of modeling of rare events, since its application provides a relatively safe method for extrapolation beyond what has been observed (see, for example, book of Embrechts *et al.* (1997)). An extreme event is an event that has a very low probability of

occurring but, when it occurs, takes on very large or small values and very often generates a large impact.

The aim of extreme value theory is to study the asymptotic behavior of large or small observations of a sample of independent and identically distributed random variables (iid). The fundamental result of the Extreme Value Theory (EVT), known as the Fisher-Tippett theorem, gives the possible limit laws of the maximum

Department of Mathematics and Informatics, Mohammed V-Agdal University, Rabat B.P. 1014, Morocco.

E-mail: abdellah.belhajjam@gmail.com; mohammadine.belbachir@gmail.com; s.elouardirhi@um5s.net.ma

^{*} Corresponding Author

DOI: <https://doi.org/10.55766/sujst-2023-04-e01485>

Suranaree J. Sci. Technol. 30(4):010240(1-11)

normalized sample, and thus allows to have some knowledge on the stochastic behavior of the tail of the distribution.

In the Extreme Value Theory (EVT), there are mainly two approaches. The first approach concerns the law of generalized extreme value (GEV) and is based on the law of the maximum normalized. This approach is based on the theorem of Fisher-Tippet and is often referred to as the “maxima block approach” (BM). The second approach is based on results from Balkema and Haan (1974), and Pickands (1975). It consists of using observations that exceed a fixed threshold and is denoted POT (Peaks-Over Thresholds).

On the other hand, the first approach is developed in the analysis of extreme values, it is applied to data which consist of a set of maxima: annual daily, semi-daily daily, quarterly daily, etc., and which groups together three distributions of extreme value namely the Gumbel law, Weibull law, and Fréchet law. However, this approach has been criticized for the use of a single maximum leading to a loss of information contained in the other large values of the sample. For overcoming this problem, Pickands (1975) introduced a new approach to value analysis known as the Generalized Pareto Distribution (GPD), which is applied to data that exceeds a certain well-determined high threshold (Peaks Over Threshold).

Moreover, the literature concerning the extreme value theory is very rich, we can quote the book of Gumbel (1958) which is used as a classic reference in EVT, that of Galambos (1978) which gives a historical description of the main results of EVT. As for Leadbetter *et al.* (1983) and Resnick (1987) they provide a mathematical treatment on the notions of functions with regular variation widely used in EVT. Works such as that of Kotz and Nadargah (2000), Coles (2001), Reiss and Tomas (1997) propose practical applications. EVT is commonly used in a variety of fields, for example in Climatology studied in Feng *et al.* (2007), Morris (2020), Amar (2014), Siliverstovs *et al.* (2010), Triphonia *et al.* (2016) and Wambua (2020).

The EVT was developed and subsequently expanded in 1975 by the publication of Pancheva (1984) on the extreme value theory under power normalization. The domain of attraction of the possible limits is developed in Barakat and Nigm (2002). After a few years Barakat *et al.* (2010, 2011, 2012, 2013, 2014) compared the rate of convergence of linear and non-linear normalization. Subsequently, other studies were published in this context: Peng *et al.* (2009, 2010) on rate of convergence, and Ravi (1992) on max domains of attraction of the Fréchet law and Weibull law. While

Mohan and Ravi (1991), Subramanya (1994), Mohan and Subramanya (1991) show that the p-max stable distribution function df 's attract more than linear stable distribution function df 's. These results are unified by the approach of Christoph and Falk (1996). In the other hand, as in the linear normalization case, Nasri-Roudsari (1999) has summarized the p-max stable distribution function df 's by following Von-Mises type representation.

Until now, the distribution of generalized extreme values theory (GEV) under linear normalization has been the most used in extreme statistical studies. It makes predictions in different areas (finance, hydrology, meteorology) by using both BM or POT approaches. Currently, we can ask ourselves this question: can the non-linear model bring improvement in the processing of data in these different fields? To this end, we propose, in this paper, the use of the non-linear model to estimate the return level and to give the associated confidence intervals.

In practical analysis, especially in time series models, the choice of a multiplicative model is appropriate when the magnitude of the seasonal trend in the data depends on the magnitude of the data. Thus, the choice of additive model is appropriate when the magnitude of the seasonal trend in the data does not depend on the magnitude of the data. So, if the seasonal trend in the data is not obvious, and it's difficult to choose between additive and multiplicative models, then we try both models and choose the one that gives the smaller accuracy measures. (Extracted from <https://support.minitab.com:/statistical-modeling/time-series>).

The remaining of this paper is organized as follows. The second section provides a theoretical background and methodology. The third section is devoted to providing the return level under non-linear normalization, by the maximum likelihood method (MLM) and the calculation of the associated confidence interval by the delta method and the profile likelihood one. The fourth section presents the data used to illustrate our approach and discussion. The last section contains a brief conclusion.

Linear normalization

Let $X_1, X_2, \dots, X_n$ be n independent and identically distributed (i.i.d) random variables of distribution function F . Let $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ be the corresponding order statistics with $X_{n,n} : = \max\{X_1, X_2, \dots, X_n\}$.

Fisher and Tippet (1928) and Gnedenko (1943) showed that there are sequential normalization constants $a_n > 0$ and $b_n \in \mathbb{R}$ such as:

$$F^n(a_n x + b_n) = P\left(\frac{x_{n,n} - b_n}{a_n} \leq X \xrightarrow{w} G(x)\right) \quad (1)$$

where $G \in \{\Lambda, \Phi_\alpha, \Psi_\alpha\}$ with Λ, Φ_α and Ψ_α the standard limit laws:

Gumbel : $\Lambda(x) = \exp\{-\exp\{-x\}\}, x \in \mathbb{R}$

Frechet : $\begin{cases} 0 & \text{if } x < 0 \\ \exp\{-x^{-\alpha}\} & \text{if } x \geq 0, \text{ and } \alpha > 0 \end{cases}$

Weibull : $\begin{cases} \exp\{-(-x)^\alpha\} & \text{if } x < 0 \text{ and } \alpha > 0 \\ 1 & \text{if } x \geq 0 \end{cases}$

These families of laws are summarized by the Von-Mises type representation:

$$G_\gamma(x; \mu; \sigma) = \exp\left[-\left[1 + \left(\frac{x-\mu}{\sigma}\right)^{-\gamma}\right]^{-1}\right], \quad 1 + \left(\frac{x-\mu}{\sigma}\right)^{-\gamma} > 0 \quad (2)$$

Where $\mu, \sigma > 0$ and $\gamma \in \mathbb{R}$ are the position, scale, and shape parameters respectively.

There are two approaches to modeling extreme events. The Maxima Block (BM) method groups data into blocks of the same length to form a sample containing the maxima of these blocks. The other approach, called the Peak Over Threshold (POT) method, provides observed exceedances of a predetermined threshold.

Non-linear normalization (power normalization)

Since the 1980's, several authors have enriched the extreme value theory by introducing non-linear normalization, specifically the power normalization. Initiated by Pancheva (1984, 1994), this model was later developed by H. Barakat *et al.* (2011, 2012, 2014) and Amar (2014).

Under the assumption of existence of constants $a_n > 0$ and $b_n > 0$, it comes that:

$$F^n(a_n |x|^{b_n} S(x)) = P\left[\left|\frac{X_{n,n}}{a_n}\right|^{\frac{1}{b_n}} S(X_{n,n}) \leq x\right] \xrightarrow{w} H(x) \quad (3)$$

where $\xrightarrow{w}$ is the weak convergence and $S(x) = -1, 0, 1$ when $x < 0, x = 0$ and $x > 0$ respectively. The H distribution function, called p -max stable (power max-stable), satisfies the stability condition:

$$H^n(a_n |x|^{b_n} S(x)) = H(x), \quad x \in \mathbb{R}, n \in \mathbb{N}$$

Pancheva (1994) showed that $H \in \{\Phi_1, \Psi_1, H_{i,\alpha} : i = 1, 2, 3, 4\}, \alpha > 0$ These six limit laws are:

Type I : $H_{1,\alpha} = \begin{cases} 0 & \text{if } x \leq 1 \\ \exp\{-(\log(x))^{-\alpha}\} & \text{if } x > 1 \end{cases}$

Type II : $H_{2,\alpha} = \begin{cases} 0, & \text{if } x \leq 0 \\ \exp\{-(-\log(x))^\alpha\} & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$

Type III : $H_{3,\alpha} = \begin{cases} 0, & \text{if } x \leq -1 \\ \exp\{-(-\log(-x))^{-\alpha}\} & \text{if } -1 < x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$

Type IV : $H_{4,\alpha} = \begin{cases} \exp\{-(-\log(-x))^{-\alpha}\}, & \text{if } x < -1 \\ 1, & \text{if } x \geq -1 \end{cases}$

Type V : $\Phi_1(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \exp\{-x^{-1}\} & \text{if } x > 0 \end{cases}$

Type VI : $\Psi_1(x) = \begin{cases} \exp\{x\}, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$

When the condition (3) is satisfied, we say that F belongs to the p -domain of attraction of H and we note $F \in D_p(H)$. Necessary and sufficient conditions for $F \in D_p(H)$ were developed by Christoph and Falk (1996), Mohan and Ravi (1993), Mohan and Subramanya (1991) and Subramanya (1994). Nasri (1999) summarized these limit laws by Von-Mises representation as follows:

$$\begin{cases} H_{1,\xi}(x, a, b) = \exp\left[-(1 + \xi \log(ax^b))^{-1/\xi}\right], \\ \quad \text{if } x > 0, 1 + \xi \log(ax^b) > 0 \\ H_{2,\xi}(x, a, b) = \exp\left[-(1 + \xi \log(a(-x)^b))^{-1/\xi}\right], \\ \quad \text{if } x < 0, 1 + \xi \log(a(-x)^b) > 0 \end{cases} \quad (4)$$

Estimation of the return level

The main aim of this paper is to calculate return level, while this level is expected to be exceeded on an average of once every T time periods with probability of p .

The estimate of extreme quantiles of the annual maximum (monthly maximum) distribution is obtained by inverting Eq. (4), where $H(z_p) = 1 - p$, and $0 < p < 1$, we get:

$$z_p = \begin{cases} \frac{1}{a^{1/b}} \exp\left\{\frac{(-\log(1-p))^{-\xi} - 1}{b\xi}\right\} & \text{if } x > 0 \\ \frac{-1}{a^{1/b}} \exp\left\{\frac{1 - (-\log(1-p))^{-\xi}}{b\xi}\right\} & \text{if } x < 0. \end{cases} \quad (5)$$

As it was defined in Stuart Coles (2001), the return level z_p is associated with the return period $T = 1/p$. To estimate z_p , we will use the Maximum Likelihood Method (MLM) Smith, (1985) to estimate the parameters a, b and ξ appearing in (5).

Under assumption that the $M_{1,n}, M_{2,n}, \dots, M_{n,n}$ (the maxima of n blocks) are independent random variables having the same GEV-non-linear distribution, the log-likelihood for the parameters is determined as follow:

Case 1:

When $x > 0, 1 + \xi \log(ax^b) > 0$

$$H_{1,\xi}(x, a, b) = \exp\left[-(1 + \xi \log(ax^b))^{-1/\xi}\right] \quad \text{and}$$

density $h_{1,\xi}$ is given by

$$h_{1,\xi}(x, a, b) = \frac{b}{x} (1 + \xi \log(ax^b))^{-\frac{\xi+1}{\xi}} \exp\left[-(1 + \xi \log(ax^b))^{-1/\xi}\right]$$

the log-likelihood is

$$l_{1,\xi}(x, a, b) = \log(b) - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^n \log[1 + \xi \log(ax_i^b)] - \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n (1 + \xi \log(ax_i^b))^{-1/\xi} \quad (6)$$

where $x_i > 0$ and $1 + \xi \log(ax_i^b) > 0$

Case2:

When $x < 0$ and $1 + \xi \log(a(-x)^b) > 0$

$H_{2,\xi}(x, a, b) = \exp\left[-\left(1 + \xi \log(a(-x)^b)\right)^{-1/\xi}\right]$ and density $h_{2,\xi}$ is given by

$$h_{2,\xi}(x, a, b) = -\frac{b}{x} \left(1 + \xi \log(a(-x)^b)\right)^{\frac{\xi+1}{\xi}} \exp\left[-\left(1 + \xi \log(a(-x)^b)\right)^{-1/\xi}\right]$$

the log-likelihood is:

$$l_{2,\xi}(x, a, b) = \log(b) - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^n \log[1 + \xi \log(a(-x)_i^b)] - \sum_{i=1}^n \log(-x_i) - \sum_{i=1}^n (1 + \xi \log(a(-x)_i^b))^{-1/\xi} \quad (7)$$

where $-x_i > 0$ and $1 + \xi \log(a(-x)_i^b) > 0$

Maximizing the functions in (6) and (7) gives the estimate of (a, b, ξ) .

Since no analytical solution is possible, a standard numerical optimization algorithm is developed in the non-linear case, using the Nonlinear Least Squares “**nls**” function predefined in software R.

The approximate distribution of $(\hat{a}, \hat{b}, \hat{\xi})$ is multivariate normal with mean (a, b, ξ) and variance-covariance matrix V equal to the inverse of the observed information matrix evaluated at the maximum likelihood.

The maximum likelihood estimates the return level z_p of the $1/p$ as $0 < p < 1$, is obtained by:

$$\hat{z}_p = \begin{cases} \frac{1}{\hat{a}^{1/\hat{b}}} \exp\left\{\frac{(-\log(1-p))^{-\hat{\xi}} - 1}{\hat{b}\hat{\xi}}\right\} & \text{if } x > 0 \\ \frac{-1}{\hat{a}^{1/\hat{b}}} \exp\left\{\frac{1 - (-\log(1-p))^{-\hat{\xi}}}{\hat{b}\hat{\xi}}\right\} & \text{if } x < 0. \end{cases} \quad (8)$$

Furthermore, by using the delta method, we'll get the following confidence interval for z_p is:

$$[\hat{z}_p - 1.96 * \text{Var}(\hat{z}_p); \hat{z}_p + 1.96 * \text{Var}(\hat{z}_p)]$$

where $\text{Var}(\hat{z}_p) \approx \nabla z_p^T * V * \nabla z_p$

where V is the variance-covariance matrix of $(\hat{a}, \hat{b}, \hat{\xi})$, and $\nabla z_p^T = \left[\frac{\delta z_p}{\delta a}, \frac{\delta z_p}{\delta b}, \frac{\delta z_p}{\delta \xi}\right]$ and

$$\begin{aligned} \frac{\delta z_p}{\delta a} &= \frac{-1}{b} a^{-(1+b/b)} \exp\left\{\frac{(-\log(1-p))^{-\xi} - 1}{b\xi}\right\} \\ \frac{\delta z_p}{\delta b} &= \exp\left\{\frac{(-\log(1-p))^{-\xi} - 1}{b\xi}\right\} \\ &\quad * \left[\frac{\xi \log(a) - (-\log(1-p))^{-\xi} + 1}{b^2 \xi e^{\frac{\log(a)}{b}}}\right] \end{aligned}$$

$$\begin{aligned} \frac{\delta z_p}{\delta \xi} &= -\frac{1}{a^{1/b}} \left[\frac{(\xi \log(-\log(1-p)) + 1)(-\log(1-p))^{-\xi} + 1}{b\xi^2}\right] \\ &\quad * \exp\left\{\frac{(-\log(1-p))^{-\xi} - 1}{b\xi}\right\} \end{aligned}$$

evaluated at $(\hat{a}, \hat{b}, \hat{\xi})$.

In the following, we compare the linear and non-linear models.

Results and Discussion

First application

In this section, we compute the return level and associated confidence intervals, which will allow us to choose the most efficient model for linear normalization (LM) and non-linear normalization (NLM). To get better approximation, we use the appropriate profile likelihood function which provides asymmetric confidence intervals for any specified return $\hat{z}_p$.

For checking the validity of GEV_p model and GEV_l models, we use two graphics, the probability plot and quantile plot. After, we use AIC (Akaike Information Criterion) measure to compare the two models. We start with the real relative humidity data of the capital region of Morocco, where we collected 12986 observations.

In Figure 1 and Figure 2, we observe that extreme values recur periodically, which explains the seasonal effect present in the humidity data.

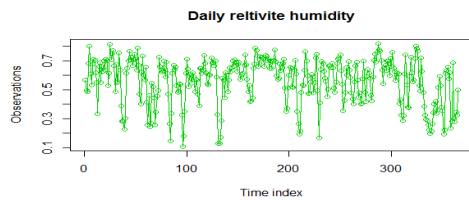


Figure 1. Distribution of observations over one year

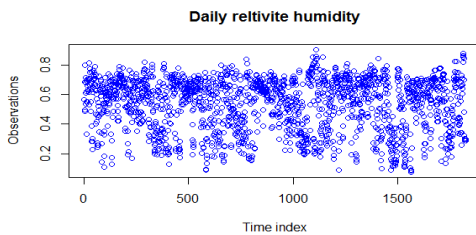


Figure 2. Distribution of observations over five years

The statistical analysis of these data can be carried out in these two steps.

Step I: Parameter Estimation for non-linear model (NLM)

For the descriptive statistical analysis, Table 1 summarizes the central trend and dispersion measures. Columns 1, 2, 3 and 4 display the mean, standard deviation, kurtosis and skewness. Column 5 displays the p-value of the normality test.

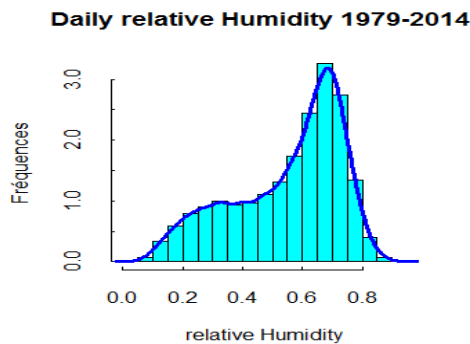


Figure 3. Histogram daily relative humidity

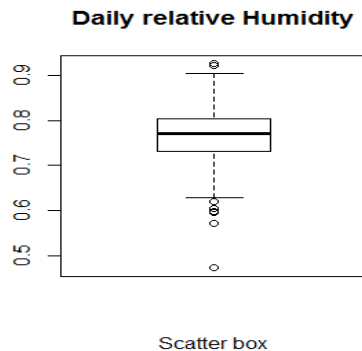


Figure 4. Box plot for daily relative humidity

Table 1. Central trend and dispersion measurements

Mean	0.7700191
Sd	0.05629804
Kurtosis	4.757318
Skewness	-0.3897856
Test Normality	A = 1.0724,
Anderson-Darling	p-value = 0.008089

The standard deviation of 0.056 indicates that there is a significant dispersion around the mean, explained by the presence of extreme values in this statistical distribution.

The negative value below 0 of the skewness ($s=-0.39$) and the value of kurtosis ($k=4.75$) indicate that firstly, the humidity distribution has a heavy left tail and secondly, the distribution curve has a bell shape

Since p-value is below 0.05 for the normality test, the null hypothesis is rejected. So, the distribution is not normal. This result appears clearly in the following two Figures.

Figure 3 gives the histogram with the density in blue line, Figure 4 gives the box plot. The latter makes it possible to visualize better the extreme values: these are the round black dots at the top and bottom of the box.

We will now apply the (MLM) to estimate the parameters of the extreme distribution of the daily humidity, where the size of each block of maxima chosen is equal to one month.

Table 2 below contains two sub-tables, the first on the left contains the estimated parameters (LM) and the second, on the right, contains the estimated parameters for (NLM).

In the linear case, we used the ismev, evir and nortet, fExtreme packages as technical support. For the non-linear case we have developed ourselves similar functions as those used in the linear case. We will therefore follow the procedure in Coles (2001), to estimate the parameters (a, b, ξ) and (μ, σ, ξ) of the two normalization models.

Columns 1, 2 and 3 contain the estimation of the three parameters of the model based on linear normalization (μ, σ, ξ) = (0.750, 0.059, -0.317). The associated standard errors are respectively 0.00294259, 0.00193603 and 0.01758717. The third row gives the 95% confidence for each parameter: [0.7451809; 0.7567159] for μ , [0.05568052; 0.06326976] for σ , and [-0.3519553; -0.2830136] for ξ .

Columns 4, 5, and 6 give the parameters estimated for the non-linear normalization model: (a, b, ξ) = (33.9087585, 12.2605486, -0.3741895). The respective standard errors are 3.85072597, 0.39842410 and 0.01612417. The third-row

Table 2: Estimation of parameters of the daily distribution of humidity.

Linear Model (LM)			Non-Linear Model (NLM)		
$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{a}$	$\hat{b}$	$\hat{\xi}$
0.75094839 (0.0029425)	0.05947514 (0.00193603)	-0.31748446 (0.01758717)	33.9087585 (3.85072597)	12.2605486 (0.39842410)	-0.3741895 (0.01612417)
[0.7451809; 0.7567159]	[0.05568052; 0.06326976]	[-0.3519553; -0.2830136]	[26.36134; 41.45618]	[11.47964; 13.04146]	[-0.4057929; -0.3425861]

confidence intervals are associated with a , b and ξ respectively:

[26.36134;41.45618], [11.47964; 13.04146] and [-0.4057929; -0.3425861].

We find that the estimated value of the ξ shape parameter is almost the same for both models. The difference in the confidence intervals is due to the difference in standard errors.

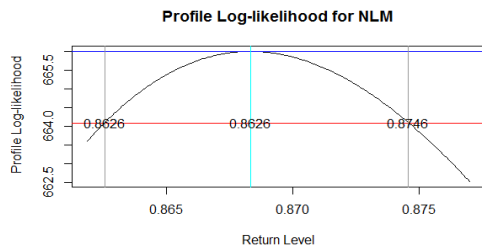


Figure 5. Profil likelihood of $\hat{z}_{0.05}$ for NLM

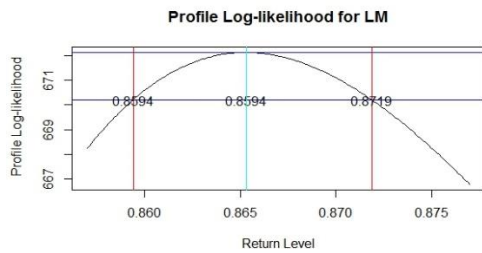


Figure 6. Profil likelihood of $\hat{z}_{0.05}$ for LM

Furthermore, the maximum likelihood estimate for ξ , is negative, corresponding to a bounded distribution.

In the next step, we use the delta-method and the results of Table 2 to compute the return levels z_p (estimated by using linear and non-linear models) and their associated confidence intervals. Table 3 contains the z_p estimates for the LM and NLM return period levels $T=20$ years.

For the linear model, the return level is $z_{0.05}=0.865322$ and the 95% confidence interval obtained by the delta-method is [0.859256; 0.8713882], and for the non-linear model the return level is $z_{0.05}=0.863495$ and the 95% confidence interval is [0.82762996; 0.90906984].

The fifth row displays the maximum log-likelihood (M-logl) corresponding to LM and NLM models, which gives M-logl =672.1199 for LM and M-logl =-666.0048 for NML.

We note that the value $\hat{z}_{0.05}$ is almost the same for NLM and LM. However, the confidence intervals containing $\hat{z}_{0.05}$ and computed using the delta-method, are different [0.8592560; 0.8713882] for LM and [0.82762996; 0.90906984] for NLM.

So, to get a better accuracy again we use the profile likelihood, in Figure 5 and Figure 6, we can read easily confidence interval for $\hat{z}_{0.05}$ of the two models: the first [0.8626; 0.8746] for NLM and the second is [0.8594; 0.8719] for LM while the two results are not equal to the results in table 3, explained by asymmetry in the profile log-likelihood surface.

Table 3. Estimated return level, confidence interval and log-likelihood

<i>LM</i>		<i>NLM</i>	
$z_{0.05}$	CI	$z_{0.05}$	CI
0.865322 (0.003095)	[0.8592560; 0.8713882]	0.863495 (0.0207755)	[0.82762996; 0.90906984]
M-log likelihood		M-log likelihood	
-672.1199		-666.0048	

Table 4. Estimated return levels for different years with NLM

Return period	5	10	17	20	100	200
Return level	0.82384093 (0.01727597)	0.84930732 (0.01827129)	0.86433909 (0.01992596)	0.86834990 (0.02077548)	0.88688008 0.03478746	0.90531344 (0.41094444)
CI (delta method)	[0.78998003, 0.85770183]	[0.81349559, 0.88511905]	[0.82528422, 0.90339397]	[0.82762996, 0.90906984]	[0.81869666, 0.95506349]	[0.09986233, 1.71076455]

Table 5. Estimated return levels for different years with LM

Return Period	5	10	17	20	100	200
Return Level	0.821922188 (0.002976911)	0.846589472 (0.003022708)	0.861346005 (0.003073083)	0.865322100 (0.003094967)	0.894796144 (0.003527999)	0.903413627 (0.003815199)
CI (delta method)	[0.816087442, 0.827756935]	[0.840664964, 0.85251398]	[0.855322762, 0.867369247]	[0.859255964, 0.87138823]	[0.887881266, 0.90171102]	[0.895935836, 0.910891417]

At the end of this step, in Table 4, we give the return level for different return periods for both LM and NLM models.

Table 4 above shows that the return levels obtained with the non-linear approach increase with time, which is logical. Thus, the associated confidence intervals are closer to each other. Except for the seventh column (where the return level is 200 years) where the standard error is half of the estimated value this is due to the asymmetry of the distribution.

Table 5, showing the estimated return levels in the LM case, gives the same results obtained in Table 4. The standard error remains stable around 0.03.

For the return period $T=200$ years, the confidence interval is almost the same as for the other lower return periods.

Step II: Diagnostic plots and model selection

Now we use some graphical methods to assess the accuracy of the GEV_p and GEV_l models fitted to the humidity.

Figures 7 and 8 provide diagnostic graphs, which assess the precision of the linear and non-linear models.

It can be seen from Figure 7 that the accuracy of the linear GEV model, adjusted on the humidity data of the Rabat region (Morocco), is validated. Indeed, the two curves P-Plot and Q-Plot are almost linear (very close to the diagonal).

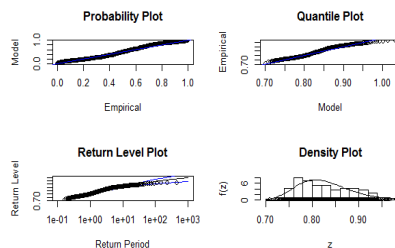


Figure 7. Adjustment with the linear GEV model

The precision of the non-linear GEV model adjusted on the same database (daily humidity) is shown in Figure 8. Both probability plot and quantile plot establish the validation of the model. Their points are almost all on the diagonal.

The third return level plot (RL Plot) shows that the curve is linear, even if few points go out of the confidence interval, which can be explained by the negative value of ξ .

b-the selection of model

We will use the AIC precision measurement $AIC = -2\hat{l} + 2k$ where $\hat{l}$ is the maximum log-likelihood and k the number of parameters. Remember also that the best model is the one with the smallest AIC.

For the non-linear model (power normalization) $AIC_p = 1338.016$ and for the linear model $AIC_L = 1350.24$.

We deduced that the non-linear model is the best. It is suitable for adjustment on humidity data.

Second numerical application

In this second numerical application, we will follow the same steps as in application 1 ie. We compute the return level and associated confidence intervals, which will allow us to choose the most efficient model.

For this study, we use the data of water levels of lake Eri  n Canadian center region, where the monthly January data measurements were taken between 1918-2019, we have 102 observations.

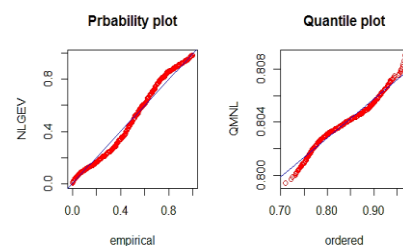


Figure 8. Adjustment with the non-linear GEV model

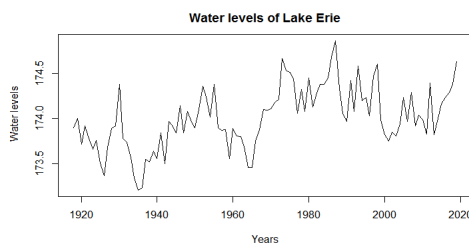


Figure 9. Monthly maximum water levels at Lake Erie (1918-2019)

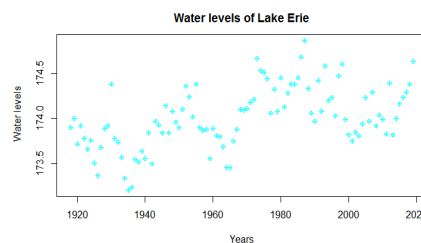


Figure 10. Adjustment with the non-linear GEV model

The data on the website: https://www.tides.gc.ca/C&A/network_means-fra.html.

Figure 9 and 10: Shows monthly water-levels recorded at Lake Erie, over the period 1918-2019.

The statistical analysis of these data can be carried out in these two steps as already done above.

Step I: Parameter Estimation (NLM)

For the descriptive statistical section, Table 6 summarizes the central trend and dispersion measures. Columns 1, 2, 3 and 4 display the mean, standard deviation, kurtosis and skewness. Column 5 displays the p-value of the normality test.

Table 6: Central trend and dispersion measurements

Mean	174
Sd	0.339
Kurtosis	2.684
Skewness	0.03318
Test Normality	A = 0.2671,
Anderson-Darling	p-value = 0.6808

The negative value below 0 of the skewness ($s=0.03318$) and the value of kurtosis ($k=2.684$) indicate that firstly, the water levels distribution has a heavy left tail and that; secondly, the distribution curve has a bell shape.

The p-value above 0.05 of the normality tests (Anderson-Darling test) leads to no rejecting the null hypothesis. So, the distribution is normal. This result appears clearly in the following two Figures.

Figure 11 gives the histogram with the normal density in blue line, Figure 12 gives the density of water levels data, and the polygon shows that the distribution is almost normal.

We now apply the Maximum Likelihood Method (MLM) to estimate the parameters of the extreme distribution of the water level. Table 7

below contains two sub-tables, the first on the left contains the estimated parameters (LM) and the second on the right contains the estimated parameters (NLM).

Column 1, 2 and 3 contains the estimation of the three parameters of the model based on linear normalization: $(\mu, \sigma, \xi) = (0.8888, 0.3357, -0.2766)$; the associated standard errors are respectively 0.03646, 0.02565 and 0.05984. The third row indicates that the confidence interval associated with each parameter estimated: $[0.8173; 0.9603]$ for μ , $[0.2854; 0.3860]$ for σ , and $[-0.394; -0.1593]$ for ξ .

Columns 4, 5 and 6 gives the parameters estimated for the non-linear normalization model:

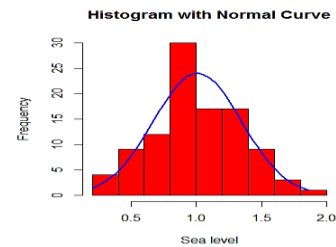


Figure 11. Histogram monthly water-levels

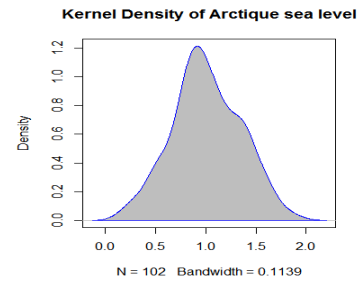


Figure 12. Density of monthly water-levels

Table 7. Estimation of parameters of the daily distribution of minimal temperature

Linear Model			Non-Linear Model		
$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{a}$	$\hat{b}$	$\hat{\xi}$
0.8888 (0.03646)	0.3357 (0.02565)	-0.2766 (0.05984)	1.4409 (0.1405)	2.3194 (0.1829)	-0.5390 (0.04753)
[0.8173; 0.9603]	[0.2854; 0.3860]	[-0.394; -0.1593]	[1.166; 1.716]	[1.961; 2.678]	[-0.6322; -0.4459]

Table 8. Estimated return level, confidence interval and log-likelihood

LM		NLM	
$z_{0.01}$	CI	$z_{0.01}$	CI
174.762455 (0.000236846)	[174.62111; 174.90380]	174.777743 (0.09128)	[174.599; 174.9567]
M-log likelihood		M-log likelihood	
33.45		34.97192	

(a, b, ξ) = (1.4409, 2.3194, -0.5390); the respective standard errors 0.1405, 0.1829 and 0.04753. The third line confidence intervals are associated with a, b and ξ respectively: [1.166 ;1.716], [1.961 ;2.678] and [-0.6322; -0.4459].

In the next step, we use the "delta-method" and the results of Table 2 to compute z_p return levels (estimated using linear and non-linear models) and their associated confidence intervals.

Table 3 contains the z_p estimates for the LM and NLM return period levels T=100 years.

For the linear model, the return level is $z_{0.01} = 174.762455$ and the 95% confidence interval obtained by the delta-method is [0.9830363; 0.9839647] where the standard error is 0.000236846. and for the non-linear model the return level is $z_{0.01} = 174.777743$ and the 95% confidence interval is [174.599; 174.9567], where the standard error is (0.09128).

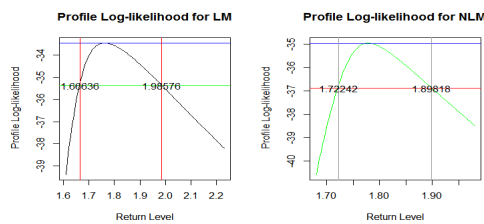


Figure 13. The profile likelihood for LM and NLM

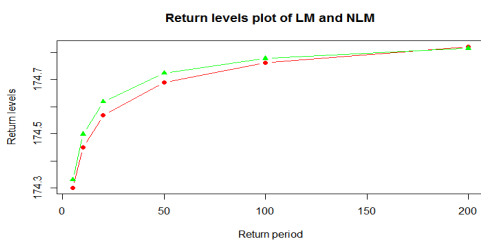


Figure 14. Return level for different years of LM and NLM

The fifth row displays the maximum log-likelihood (M-log) corresponding to LM and NLM, which gives M-log =33.45 for LM and M-log = 34.97192 for NML.

We can see that the value $z_{0.01}$ is almost the same for NLM and LM. However, the confidence intervals, containing $z_{0.01}$, and computed using the delta-method, are different.

But to obtain more accuracy for these intervals, containing the return level, we use the method of the log-likelihood profile (profile log-likelihood) Figure 13, we find that in the NLM the confidence interval for $\hat{z}_{0.01}$ obtained by the log-likelihood profile we find that CI: [1.722; 1.898], and for LM the confidence interval for $\hat{z}_{0.01}$ is [1.666;1.986].

At the end of this step, in Table 9 and 10, we give the return level for different return periods for both LM and NLM models.

Table 9 above shows that the return levels obtained with the non-linear approach increase with time, which is logical. Thus, the associated confidence intervals are closer to each other.

Table 10, showing the estimated return levels in the LM, gives the same results obtained in Table 9. The standard error remains stable around 0.01. for the return period T=200 years, the confidence interval is almost the same as for the other higher return periods.

The Figure 14, gives the return levels of LN and NLM for different returns period; we can clearly see that the return levels with non-linear model (green line) overcame the return levels with linear model (red line).

Step II: Fit and Model Selection Adjustment

We are now using some graphical methods to validate the model.

Figures 15 and 16 provide diagnostic graphs, which assess the precision of the linear and non-linear model.

It can be seen from Figure 13 that the precision of the linear GEV model, adjusted on the water levels data of the Lake Erie, is validated. Because

Table 9. Estimated return levels for different years with NLM

Return period	5	10	17	20	100	200
Return level	174.331 (0.11)	174.499 (0.1035)	174.618 (0.09841)	174.724 (0.09365)	174.778 (0.09128)	174.815 (0.08966)
CI (delta method)	[174.1; 174.6]	[174.3; 174.7]	[174.4; 174.8]	[174.5; 174.9]	[174.6; 175.0]	[174.6; 175.0]

Table 10. Estimated return levels for different years with LM

Return Period	5	10	17	20	100	200
Return Level	174.3009 (0.03969)	174.45115 (0.04258)	174.5687 (0.0484)	174.69 (0.06065)	174.76245 (0.07211)	174.822 (0.08437)
CI (delta method)	[174.22311; 174.37870]	[174.36769; 174.53461]	[174.4739; 174.6636]	[174.57114; 174.80888]	[174.621; 174.904]	[174.65661; 174.98735]

the two curves P-Plot and Q-Plot are almost linear (very close to the diagonal). The third level of return graph (RL Plot) shows that the curve is linear, even if few points go out of the confidence interval, this is explained by the negative value of ξ^* .

The precision of the non-linear GEV model adjusted on the same database (water levels) is shown in Figure 16. The two curves P-Plot and Q-Plot are almost linear (very close to the diagonal). The third level of return graph (RL Plot) shows that the curve is linear. And for all the return levels are between the blue lines of confidence interval.

The choice of model

We will use the AIC precision measurement where $AIC = -2\hat{l} + 2k$ where $\hat{l}$ the maximum log-likelihood and k the number of parameters. For the non-linear model (power normalization) $AIC_P = -63.9434$ and for the linear model $AIC_L = -60.90$. We deduced that in this case the non-linear model is the best. So, this model is suitable for adjustment water levels data.

Third numerical application

The last application will be applied on the annual maxima of Port Pirie of South Australia which are already treated in Coles (2001) in the context of the linear model. In this work we will treat the same data using the non-linear model and compare the return level results for the two models.

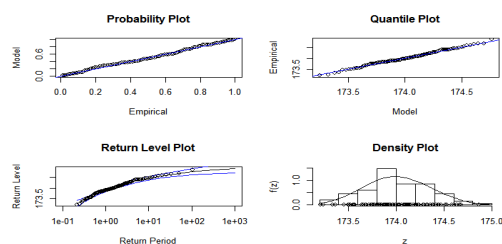


Figure 15. Adjustment with the linear GEV model

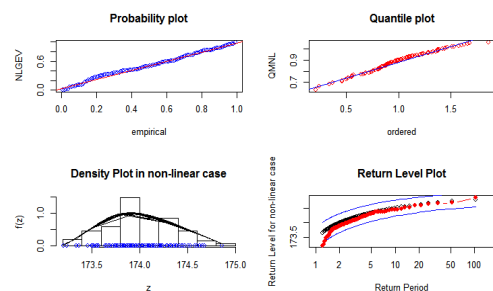


Figure 16. Adjustment with the non-linear GEV model

We follow the procedure used in the two applications above to get the following table, which contains the return level, the variance associated with delta-method and the confidence interval.

We see from the table that the return level increases proportionally with the return period. For the variance calculated with delta-method for NLM the variation is small, leading the shorter intervals. On the contrary in the case LM, the variation is greater and the intervals wider.

Conclusions

We opted for the study of the two linear and non-linear models to see which one is the best for our climate data. Having compared these two models, using the AIC accuracy measure, we concluded that the non-linear model best adjusts our humidity Moroccan data and water levels Canadian data. And for the annual maxima of Port Pirie of South Australia we can see for largest return period the two models give the same return level, but the difference in confidence interval, while in the non-linear model these intervals vary with the return period. So, we recommend using our model as a comparison model.

References

- Amar, A. (2014). Contribution of the theory of extreme values to the modeling and management of stock market, financial and hydro-meteorological risks. Available from: <https://toubkal.imist.ma/handle/123456789/13059>
- Balkema, A.A., and de Haan, L. (1974). Residual Lifetime at Great Age. *The Annals of Probability*, 2(5):792-804. <https://doi.org/10.1214/aop/1176996548>
- Barakat, H.K., and A.R. Omar. (2011). Limit theorems for intermediate and central order statistics under nonlinear normalization. *Journal of Statistical Planning and Inference*, 14(1):524-535. <https://doi.org/10.1016/j.jspi.2010.06.031>
- Barakat, H.M., Nigm, E.M. and El-Adll, M.E. (2010). Comparison between the rates of convergence of extremes under linear and under power normalization. *Statistical Papers*, 51(1):149-164. <https://doi.org/10.1007/s00362-008-0128-1>
- Barakat, H.M. (2013). The use of power normalization as a new trend in the order statistics limit theory. *Journal of Statistics Applications and Probability*, 2(3):251. <https://doi.org/10.12785/jsap/020308>
- Barakat, H., and Nigm, E.S. (2002). Extreme order statistics under power normalization and random sample size. *Kuwait Journal of Science and Engineering*, 29(1):27-41.
- Barakat, H.M. Nigm, E.M., and Khaled, O.M. (2014). Statistical modeling of extreme under linear and power normalizations with applications to air pollution. *Kuwait J. Sci.*, 41(1):1-19
- Barakat, H., Omar, A., and El-Arishy, S. (2012). Asymptotic theory of order statistics under power normalization. (1st ed.). LAP LAMBERT Academic Publishing. Retrieved from: <https://www.perlego.com/book/3357548/asymptotic-theory-of-order-statistics-under-power-normalization-order-statistics-under-power-normalization-pdf>.

- Barakat, H.M., E.M. Nigm., and O.M. Khaled. (2013). Extreme value modeling under power normalization, *Applied Mathematical Modelling*, 37(24):10,162-10,169, ISSN 0307-904X, <https://doi.org/10.1016/j.apm.2013.05.045>
- Christoph, G., and Falk, M. (1996). A note on domains of attraction of p-max stable laws. *Statistics & Probability Letters*, 28(3):279-284. [https://doi.org/10.1016/0167-7152\(95\)00136-0](https://doi.org/10.1016/0167-7152(95)00136-0).
- Coles, S. (2001). Introduction. In: *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer, London. p. 1-17. https://doi.org/10.1007/978-1-4471-3675-0_1
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer. <https://doi.org/10.1007/978-3-642-33483-2>
- Feng, S., Nadarajah, S., and Hu, Q. (2007). Modeling Annual Extreme Precipitation in China Using the Generalized Extreme Value Distribution. *Journal of the Meteorological Society of Japan. Ser. II*, 85(5):599-613. <https://doi.org/10.2151/jmsj.85.599>.
- Fisher, R., and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180-190. doi:10.1017/S0305004100015681
- Fréchet, M. (1927) Sur la loi de Probabilité de l'écart Maximum. *Annales de la Société Polonaise de Mathématique*, 6:93-116.
- Galambos, j. (1978). The asymptotic theory of extreme order statistics, editor(s): chris p. Tsokos, i.n. Shimi, the theory and applications of reliability with emphasis on bayesian and nonparametric methods, p. 151-164 <https://doi.org/10.1016/b978-0-12-702101-0.50014-7>
- Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44(3):423-453. <https://doi.org/10.2307/1968974>
- Gumbel, E. (1958). *Statistics of Extremes*. New York Chichester. West Sussex: Columbia University Press. <https://doi.org/10.7312/gumb92958>
- Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London. P. 196. <https://doi.org/10.1142/p191>
- Leadbetter, M.R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, 65(2):291-306. <https://doi.org/10.1007/BF00532484>
- Mohan, N.R., and Ravi, S. (1993). Max Domains of Attraction of Univariate and Multivariate p -Max Stable Laws. *Theory of Probability & Its Applications*, 37(4):632-643. <https://doi.org/10.1137/1137119>
- Mohan, N., and Subramanya, U.R. (1991). Characterization of max domains of attraction of univariate p-max stable laws. Available from: <https://www.semanticscholar.org/paper/Characterization-of-max-domains-of-attraction-of-Mohan-Subramanya/834fb59a72209348124f4c41763130826c700265>
- Morris, T.H. (2020). Experiential learning - a systematic review and revision of Kolb's model. *Interactive Learning Environments*, 28(8):1,064-1,077. doi:10.1080/10494820.2019.1570279
- Nasri-Roudsari, D. (1999). Limit distributions of generalized order statistics under power normalization, *Communications in Statistics - Theory and Methods*, 28(6):1,379-1,389. <https://doi.org/10.1080/03610929908832362>
- Pancheva, E. (1994). Extreme Value Limit Theory with Nonlinear Normalization. In: Galambos, J., Lechner, J., Simiu, E. (eds) *Extreme Value Theory and Applications*. Springer, Boston, MA. p. 305-318. https://doi.org/10.1007/978-1-4613-3638-9_18
- Pancheva E. (1984). Limit theorems for extreme order statistics under nonlinear normalisation, *Lect. Notes Math.*, 1155:284-309.
- Peng, Z., Nadarajah, S., and Lin, F. (2010). Convergence Rate of Extremes for the General Error Distribution. *Journal of Applied Probability*, 47(3):668-679. <https://doi.org/10.1239/jap/1285335402>
- Peng, Z., Bin, T., and Nadarajah, S. (2009). Tail Behavior of the General Error Distribution. *Communications in Statistics-theory and Methods - COMMUN STATIST-THEOR METHOD*, 38:1,884-1,892. <https://doi.org/10.1080/03610920802478367>
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119-131. <https://doi.org/10.1214/aos/1176343003>
- Ravi, S. (1992). Tail equivalence of distribution functions belonging to the max domains of attraction of multivariate p-max stable laws. Available from: <https://www.semanticscholar.org/paper/Tail-equivalence-of-distribution-functions-to-the-Ravi/fbaeff3fa55ee629c6e4183ac5aa87ff557f8f4f>
- Reiss, R.-D., and Thomas, M. (1997). *Statistical Analysis of Extreme Values*. Birkhauser Verlag AG.s
- Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer. <https://doi.org/10.1007/978-0-387-75953-1>
- Smith, R.L. (1985). Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika*, 72(1):67-90. <https://doi.org/10.2307/2336336>
- Siliverstovs, B., Ötsch, R., Kemfert, C., Jaeger, C.C., Haas, A., and Kremers, H. (2010). Climate change and modelling of extreme temperatures in Switzerland. *Stochastic Environmental Research and Risk Assessment*, 24(2):311-326. <https://doi.org/10.1007/s00477-009-0321-3>
- Subramanya, U.R. (1994). On max domains of attraction of univariate p-max stable laws. *Statistics & Probability Letters*, 19(4):271-279. [https://doi.org/10.1016/0167-7152\(94\)90176-7](https://doi.org/10.1016/0167-7152(94)90176-7)
- Triphonia, J.N., Joachim, R., Edwin, R., Shaban, N., and Michel, d.s. M. (2016). Modelling of Extreme maximum Rainfall using Extreme Value Theory for Tanzania. *International Journal of Scientific and Innovative Mathematical Research*, 4(3):34-45. <https://doi.org/10.20431/2347-3142.0403005>
- Wambua, M.M., Mung'atu, J.K., and Aduda, J.A. (2020). Modelling Extreme Temperature Using Extreme Value Theory: A Case Study Northern Kenya. *International Journal of Data Science and Analysis*, 6(5):130-136. <https://doi.org/10.11648/j.ijdsa.20200605.12>