International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:9, No:6, 2015

Extreme Temperature Forecast in Mbonge, Cameroon through Return Level Analysis of the Generalized Extreme Value (GEV) Distribution

Nkongho Ayuketang Arreyndip, Ebobenow Joseph

Abstract—In this paper, temperature extremes are forecast by employing the block maxima method of the Generalized extreme value(GEV) distribution to analyse temperature data from the Cameroon Development corporation(C.D.C). By considering two sets of data (Raw data and simulated data) and two (stationary and non-stationary) models of the GEV distribution, return levels analysis is carried out and it was found that in the stationary model, the return values are constant over time with the raw data while in the simulated data, the return values show an increasing trend but with an upper bound. In the non-stationary model, the return levels of both the raw data and simulated data show an increasing trend but with an upper bound. This clearly shows that temperatures in the tropics even-though show a sign of increasing in the future, there is a maximum temperature at which there is no exceedence. The results of this paper are very vital in Agricultural and Environmental research.

Keywords—Return level, Generalized extreme value (GEV), Meteorology, Forecasting.

I. Introduction

EXTREME value theory or analysis is a branch of statistics dealing with extreme deviations from the median of probability distribution [1].

A. Generalized Extreme Value Distributions

The generalized extreme-value (GEV) distribution, introduced by Jenkinson [1955], has found many applications in hydrology [2]. It was recommended for at-site flood frequency analysis in the United Kingdom [2], for rainfall frequency in the United States and for sea waves [2]. For regional frequency analysis the GEV distribution has received special attention since the introduction of the index-flood procedure based on probability weighted moments (PWM) [2]. Many studies in regional frequency have used the GEV distribution [2]-[3]. In practice, it has been used to model a wide variety of natural extremes, including floods, rainfall, wind speeds, wave height, and other maxima. The physical origin of these maxima suggests that their distributions may be one of the extreme value (EV) types spanned by the GEV distribution (EV types I, II, and III) [2]. Mathematically, the GEV distribution is very attractive because its inverse has a closed form, and parameters are easily estimated by moments and L moments [3]. In probability theory and statistics, the generalized extreme value distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumbel, Frechet and Weibull families also known as type I, II and III extreme value distributions. It seeks to asses from a given ordered sample the probability of evens that are more extreme than any previously observed. Some applications of extreme value theory include predicting the probability of distribution of; Extreme floods, extreme winds, extreme rainfall, extreme temperature, the amount of large insurance losses, pipeline failures due to pitting corrosion [4]. In most countries this theory is used determine the GDP of a country, national income, birth rate, death rate, population growth rate etc.

Nkongho Ayuketang Arreyndip and Ebobenow Joseph are with the department of Physics and Applied Physics, University of Buea, Buea, P.O Box, 63, Cameroon e-mail:nkongho.anyuketang@ubuea.cm, Ebobenow@gmail.com.

B. Challenges from Global Climate Change

In the 21st century scientist have been faced with the challenging problem of global climate change or global warming. This global climate change has affected global air temperature, oceanic temperature, rainfall, wind and quality of incoming solar radiation. Global circulation model predicts 1.4 to $5.8^{o}c$ rice in global temperature by the end of the 21st century because of the increase in the concentration of green house gases. This increase in temperature has drastically increase the rate of evaporation which has resulted to the accumulation of clouds hence less radiant heat is being loss. This has lead to an increase in night temperature than day temperature. This increase in night temperature has a huge effect on agricultural production worldwide. Long and short term periods of heat stress are predicted to occur more frequently as a result of global warming, affecting many aspects of crop growth and development, reducing crop yield and decreasing crop quality. This high temperature decreases crop production by decreasing photosynthetic function, sugar and starch content, increasing respiration rate, suppressing floral bud development, causing male sterility and low pollen viability and hastening crop maturity [5].

C. Geographical Location and Climatic Conditions of Mbonge

Mbonge is a little town in Meme division, South West region of Cameroon located on the leeward site of mount Cameroon. This town like most other towns in Cameroon has an equatorial climate with temperature range between $20^{o}c$ to $35^{o}c$ and very high level of precipitation. Some tropical crops that can grow in this region are; cassava, beans, plantains, banana, cocoa, palm nuts, rubber, ground nuts ...etc. The Little town of Mbonge is known to be one of the towns where there is Cameroon's Agro industrial company called "Cameroon Development Corporation" (CDC). This company produces crops like: rubber, oil palm, bananas, coconuts, tea etc. These crops are mostly exported and also contributes greatly to the countries GDP. So following the huge impact of climatic change on crop growth and production, it is worth while necessary to study the effect of extreme temperatures in this area.

II. Methodology

Monthly temperature over Mbonge in Cameroon for the period 1993 to 2012 was obtained from the Cameroon Development Corporation (C.D.C). The dataset contains 240 values of monthly maximum temperature for the past 20 years. Extreme value analysis was performed on this study by fitting the generalised extreme value distribution to the sample of different periods of extremes using method of maximum likelihood estimates (MLE).

A. Generalized Extreme Value Family of Distribution

We consider the generalized extreme value distribution having cumulative distribution function given by

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

Define for $\left\{x : 1+\xi\frac{(x-\mu)}{\sigma} > 0\right\}$, $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$ where μ is the location parameter, ξ is the shape parameter and σ is the scale parameter.The Location parameter specifies the center of the GEV distribution. Scale parameter, determines the size of deviations of μ. And Shape parameter which shows how rapidly the upper tail decays.
Here positive ξ implies a heavy tail while negative ξ implies a bounded tail, and the $\lim \xi \to 0$ implies an exponential tail [6].

There are three classes of the generalized extreme value family of distribution. Their difference depends only on the value of the the shape parameter ξ.

We have the Gumbel distribution with $\xi = 0$ with cumulative probability distribution given by

$$F(x) = 1 - \exp(-\exp(z)) \tag{2}$$

where

$$z \equiv \frac{x-\mu}{\sigma}$$

for Gumbel mean and

$$F(x) = \exp(-\exp(-z)) \tag{3}$$

where

$$z \equiv \frac{x-\mu}{\sigma}$$

for Gumbel max [7].

The Gumbel distribution is a particular case of the generalized extreme value distribution (also known as the Fisher Tippett distribution). It is also known as the log Weibull distribution and the double exponential distribution (a term that is alternatively sometimes used to refer to the Laplace distribution). It is often incorrectly labelled as Gompertz distribution. And it is used to model the distribution of the maximum (or the minimum) of a number of samples of various distributions it is also known to be unbounded as it takes all values in the real number line [7].

For $\xi < 0$ we have the Weibull distribution with cumulative probability distribution given by

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\sigma}\right)^{\xi}\right) \tag{4}$$

for two parameter Weibull and

$$F(x) = 1 - \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right) \tag{5}$$

for three parameter Weibull [7].

The Weibull distribution is used in cases that deal with the minimum rather than the maximum. The distribution here has an addition parameter compared to the usual form of the Weibull distribution and, in addition, is reversed so that the distribution has an upper bound rather than a lower bound. Importantly, in applications of the GEV, the upper bound is unknown and so must be estimated while when applying the Weibull distribution the lower bound is known to be zero.
For $\xi > 0$, we have the Freshet class of distribution with cumulative probability distribution given by

$$F(x) = exp\left(-\left(\frac{\sigma}{x}\right)^{\xi}\right) \tag{6}$$

for two parameter Freshet and

$$F(x) = exp\left(-\left(\frac{\sigma}{x-\mu}\right)^{\xi}\right) \tag{7}$$

for three parameter Freshet [7].
The Freshet cumulative distribution function (CDF) is the only CDF defined on the non-negative real numbers that is a well-defined limiting CDF for the maxima of random variables (RVS). Thus, the Freshet CDF is well suited to characterize RVS of large features. As such, it is important for modelling the statistical behaviour of materials properties for a variety of engineering applications [7].
Below are curves of the GEV family of Distributions

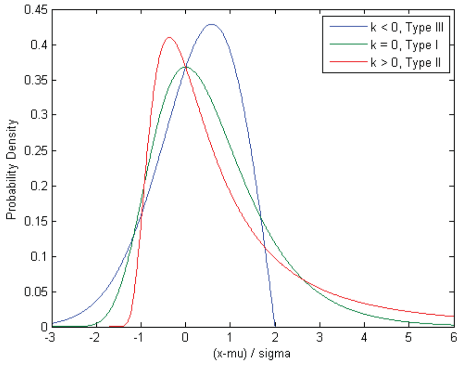

Fig. 1. GEV Family of Dsitributions

B. Return Levels

A T-year return level say x_T, is the value occurring on average ones in every T-years. For our analysis, we have a 2 years return level, 20 years return level and 100 years return level.
Extreme-value theory is often required to find return values for return periods that amply exceed the record length. This implies extrapolation of the GEV fit to a domain outside the range of the observations. In our approach, the return value determination involves little extrapolation, as series length and return periods of interest T are about equal. This considerably reduces the uncertainty in the estimate [8].
Solving for x_T in the equation

$$F(x_T) = 1 - \frac{1}{T} \tag{8}$$

where

$$F(x_T) = \exp\left\{ - \left[1 + \xi \left(\frac{x_T - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \tag{9}$$

we have that

$$1 - \frac{1}{T} = \exp\left\{ - \left[1 + \xi \left(\frac{x_T - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \tag{10}$$

$$-\log\left(1 - \frac{1}{T}\right) = \left[1 + \xi \left(\frac{x_T - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \tag{11}$$

$$\left\{ -\log\left(1 - \frac{1}{T}\right) \right\}^{-\xi} = 1 + \xi \left(\frac{x_T - \mu}{\sigma} \right) \tag{12}$$

$$\Rightarrow x_T = \mu - \frac{\sigma}{\xi} \left[1 - \left\{ -\log\left(1 - \frac{1}{T}\right) \right\}^{-\xi} \right]. \tag{13}$$

For the case where $\xi = 0$, we have that

$$x_T = \mu - \sigma \log\left\{ -\log\left(1 - \frac{1}{T}\right) \right\} \tag{14}$$

By substituting $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ in the above equation, we obtain the Maximum Likelihood estimate for return levels. From our data, we have 20 years of monthly maximums with a maximum temperature of 36^0c. Our main aim in this work is to carry out za return level analysis for extreme temperature forecast.

C. Model Selection

Fitting the data in to the GEV family of distribution, we have the following results

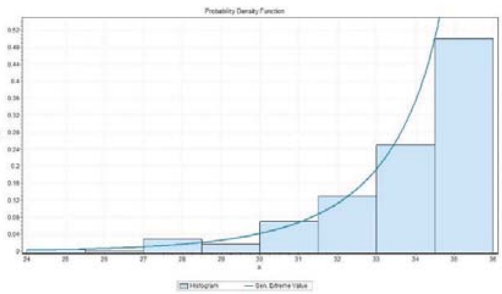

Fig. 2. GEV distribution show a good fit

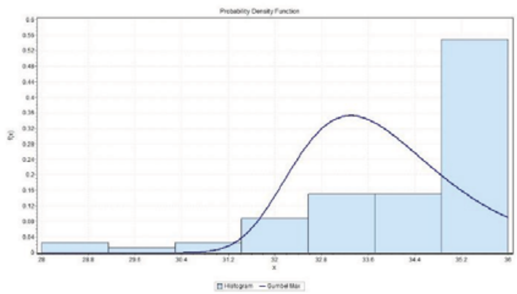

Fig. 3. Gumbel distribution shows no fit

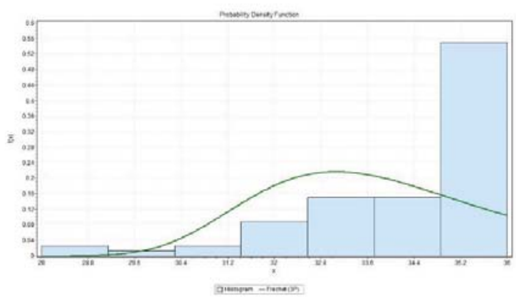

Fig. 4. Freshet distribution shows no fit

From the above graphs we see that the GEV distribution best fits our data than the Frechet and Weibull distribution. Hence it is considered as the best model.

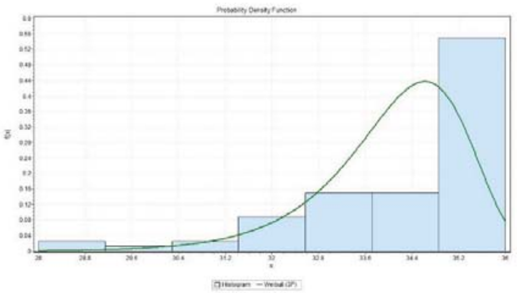

Fig. 5. Weibul distribution shows no fit

D. Selection Period

Generalized extreme value distribution uses block maxima for statistical modelling. Here, the data is partitioned into blocks of equal length, and fitting the GEV distribution to the set of block maxima. The dataset provided has 20 years of monthly maxima. So if one year block is used, then we will only have 20 points which is not good for any statistical modelling. Thus we consider and compare different selection periods which are the; Monthly, Bi-monthly, Quarterly, Half-yearly and yearly selection periods. Carrying out a test statistics, we have the following data below.

Time	Test	C.V	P.V	S.L	Reject
1Month	K.S	0.08766	2.1413e-17	0.05	Yes
	A.D	2.5018	N/A	0.05	Yes
2months	K.S	0.12397	8.1179e-9	0.05	Yes
	A.D	2.5018	N/A	0.05	Yes
4months	K.S	0.1496	1.1662e-5	0.05	Yes
	A.D	2.5018	N/A	0.05	Yes
6months	K.S	0.21012	0.00278	0.05	Yes
	A.D	2.5018	N/A	0.05	Yes
Yearly	K.S	0.29408	0.05405	0.05	No
	A.D	2.5018	N/A	0.05	Yes

Table I: GEV test statistics (where; S.L = significant level, C.V= Critical value, P.V=P-value).

Descriptive Statistics

Statistic	Value
Sample Size	240
Range	12
Mean	33.554
Variance	3.8883
Std. Deviation	1.9719
Coef. of Variation	0.05877
Std. Error	0.12728
Skewness	-1.648
Excess Kurtosis	3.1231

Percentile	Value
Min	24
5%	29.05
10%	31
25% (Q1)	32.25
50% (Median)	34.5
75% (Q3)	35
90%	35
95%	35
Max	36

Table II: Descriptive Statistics

E. Parameter Estimation

Two common methods of estimating the GEV parameters are the method of maximum likelihood and the method of L-moments [3]. For small samples, Hosking found that L-moment estimators produced biased estimates, but were preferable to maximum likelihood estimators because they resulted in estimated quantiles with smaller variances. Also, the method of L-moments is usually computationally more tractable than the method of maximum likelihood. Research has also shown that the asymptotic standard error of the L-moment estimator when compared with those of maximum-likelihood estimators usually show that the L-moment estimator is more efficient than the Maximum-Likelihood method for parameter estimation. GEV quantiles estimated is also another method of parameter estimation from small samples using conventional method of moments estimators were more accurate than those based on either maximum likelihood or L-moments. Fitting the data into the GEV distribution, we have the following parameter estimate

Distribution	ξ	σ	μ
GEV	-1.1667	1.8967	33.688
Weibull(3P)	1.0412E+8	1.2863E+8	-1.2863E+8
Frechet(3P)	3.7853	9.9007	22.251
Gumbel(mean)	0	1.5375	34.442

Table III: Parameter estimate

III. Results and Discussion

A. Results from Raw Data

Return Period (years)	Return level (^0c)
2	34.23257
20	35.29240
100	35.34213

Table IV: Return levels estimates for monthly data

B. Results from Simulated Data

Return Period (years)	Return level (^0c)
2	32.52296
20	36.20930
100	37.35988

Table V: Return level estimate of simulated data

From the above analysis, we see that the maximum value of 36^0c can not be exceeded even in the next 100 years with the raw data. This is supported by the return level graph that gives a straight line at 35.34213^0c This was also supported by the time series graph that gave an almost constant trend over time.
This is not a good result for our situation following the concept of global warming that is alarming nowadays. So to solve this problem, we simulated the same data between the ranged given by our raw data since temperature prediction is a

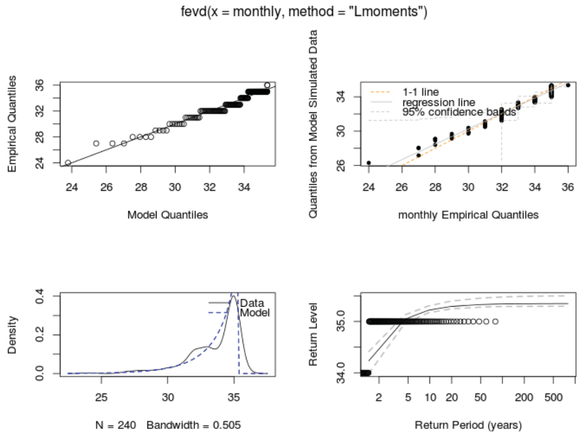

Fig. 6. Graphs from raw data

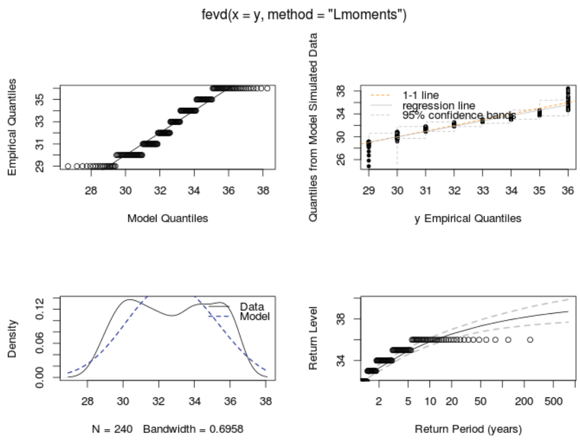

Fig. 7. Graphs from simulated data

stochastic process and we had the following results displayed in Fig: 7. Temperature values here show to be increasing but with an upper bound.

To model the GEV distribution with trend, In this work, we are only going to consider the case where the location parameter μ is time dependent and the scale and shape parameters are constants of time. The model is given by

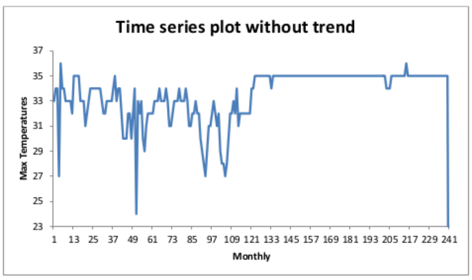

Fig. 8. Time series of raw data

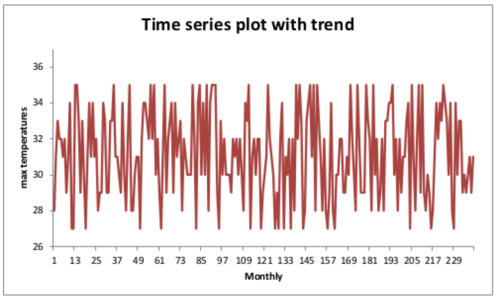

Fig. 9. Time series of simulated data

$$\begin{cases} \mu(t) = \gamma_0 + \gamma_1 t \\ \sigma(t) = \sigma \\ \xi(t) = \xi \end{cases} \tag{15}$$

The probability density function with trend in the location parameter now becomes

$$F(x) = \exp\left\{ - \left[1 + \xi \left(\frac{x - (\gamma_0 + \gamma_1 t)}{\sigma} \right) \right]^{-1/\xi} \right\} \tag{16}$$

C. Parameter Estimate and Graphs

Now fitting our data to the GEV distribution with trend, we have the following parameter estimate and the graphs are given below

Parameters	loc	scale	shape
Estimates	33.318	1.976	-0.733
Standard error	0.1301	0.1043	0.0245
Deviance: 876.9426 Optimization Information Convergence : Successful Function evaluation : 92 Gradient evaluation : 16			

Table VI: Model Parameter Estimate from Raw Data

Parameters	loc	scale	shape
Estimates	31.9807	2.4438	-0.4912
Standard error	0.19116	0.16714	0.08447
Deviance: 1059.655 Optimization Information Convergence : Successful Function evaluation : 37 Gradient evaluation : 13			

Table VII: Model parameter estimate from simulated data

From Fig: 10 and Fig: 11, we see that modelling GEV distribution taking in to consideration the existence of trend in the location parameter, the return level graph of the raw data and simulated data are similar with all of them having an upper bound. This upper bound simply means that the predicted future temperature values dough might increase, there is a maximum value that this temperature values will

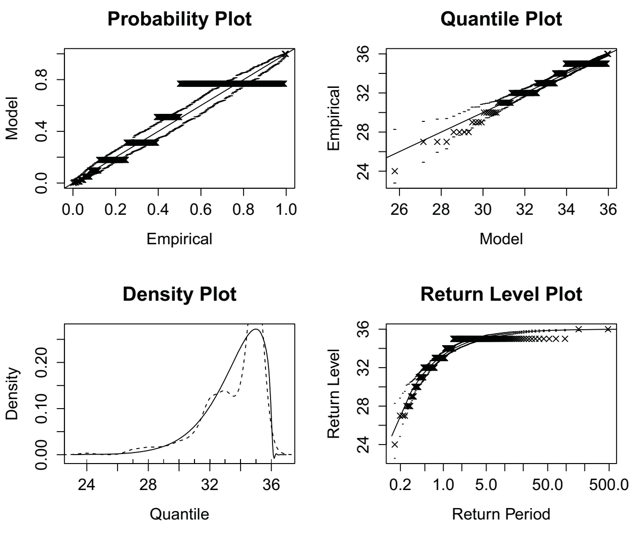

Fig. 10. Graphs with trend from raw data

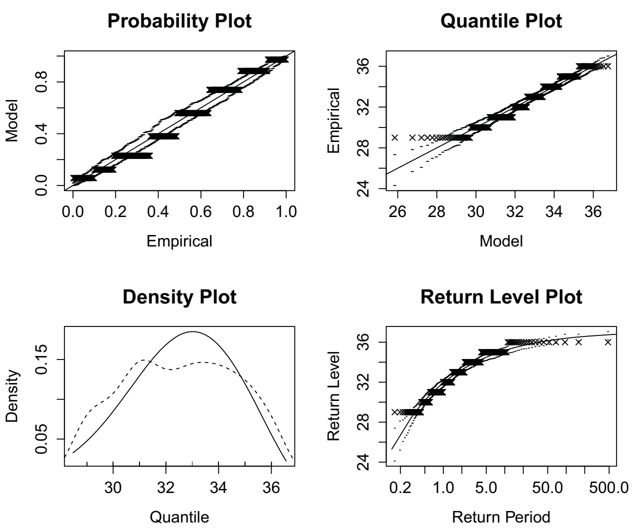

Fig. 11. Graphs with trend from simulated data

not exceed even in the next 1000 years. This simply means that if we consider the earth as a black body, it will absorb cosmic rays from the sun and also emit radiations and there will come a point where the amount of radiation absorbed is equal to the radiation emitted. This will give a kind of thermal equilibrium environment.

IV. Conclusion

We have been provided with 20 years of monthly maximum temperature from the Cameroon Development Corporation (C.D.C) Mbonge. Fitting this data into the Generalized extreme value family of distributions, it was found that the three parameter Generalized extreme value model best fit our data as compared to the Weibull, Freshet and Gumbel models. We went further to look for the best selection periods that can be model with this generalized extreme value distribution where it was found that the monthly, bi-monthly and quarterly selection periods are the best selection periods to be modelled with the GEV distribution. The time series analysis of the maximum temperature data does not show any strong evidence of trend hence we decided to simulate the same data where the time series of the simulated data does not still show a strong evidence that there exist trend over time in the temperature values this made us to conclude that temperature values in the tropics are fairly constant and does not show any strong evidence of increasing in the future. While the return level analysis of our raw data does not give us any good evidence that temperature values will exceed the maximum at 2, 20, and 100 years return periods, the return level analysis of our simulated data clearly show that the temperatures show an increasing trend over time and will exceed the maximum in 20 and 100 years return periods. From this return level analysis, we realize that even though temperature values in the tropics might show signs of increasing in the future, there is an upper bound where temperature values will not exceed even in the next 1000 years.

In the future, we will expect to have temperature data together with data from the yield or growth of these crops from C.D.C Mbonge for proper analysis.

Acknowledgment

The authors would like to thank the AIMS-NEI network most especially AIMS-Cameroon for their material and financial support.

We will like to thank the Cameroon Development Corporation (C.D.C) research department for providing us with the data.

References

[1] Myriam Charras-Garrido, Pascal Lezaud. *Extreme Value Analysis : an Introduction*. Journal de la Societe Francaise de Statistique et Societe Mathematique de France, 2013, 154 (2), pp 66-97. (hal-00917995).
[2] S. Eduardo. Martins et al, *Generalized maximum-likelihood generalised extreme-value quantile estimators for hydrologic data*, 2000.
[3] Hosking et al *Estimation of the Generalized Extreme value distribution by the method of the probability-weighted moments*, 1985 .
[4] Husna Hasan, N. Salam *Modelling extreme temperature in Malaysia using generalised extreme value distribution*, 2012.
[5] Meehl, George A, Tebaldi, Claudia, *More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century*. Science 305 (5686): 9947. Bibcode:2004Sci...305..994M. doi:10.1126/science.1098704. PMID 15310900, (13 August 2004).
[6] Sheng Ngong *Estimation of hot and cold spells with extreme value theory*, 2012.
[7] *Notes on probability distributions in Easy-fit pro*.
[8] H. W. Van Den Brink, *Estimating return periods of extreme evens from ECMWF seasonal forecast ensembles*, 2004.