

Sensitivity of Statistical Models for Extremes Rainfall Adjustment Regarding Data Size: Case of Ivory Coast

Relwindé Abdoul-Karim Nassa¹, Amani Michel Kouassi², Makouin Louise Toure³

¹National Polytechnic Institute Félix Houphouët-Boigny (INP-HB), Polytechnic Doctoral School (EDP), Yamoussoukro, Ivory Coast

²National Polytechnic Institute Félix Houphouët-Boigny (INP-HB), Department of Earth Sciences and Mineral Resources (STeRMi), Laboratory of Civil Engineering, Geosciences and Geographical Sciences, Yamoussoukro, Ivory Coast

³National Polytechnic Institute Félix Houphouët-Boigny (INP-HB), Superior School of Mines and Geology (ESMG), Yamoussoukro, Ivory Coast

Email: relwinde.nassa@inphb.ci, michel.kouassi@inphb.ci, makouin.toure@inphb.ci

How to cite this paper: Nassa, R.A.-K., Kouassi, A.M. and Toure, M.L. (2021) Sensitivity of Statistical Models for Extremes Rainfall Adjustment Regarding Data Size: Case of Ivory Coast. *Journal of Water Resource and Protection*, 13, 654-674.
<https://doi.org/10.4236/jwarp.2021.138035>

Received: May 1, 2021

Accepted: August 2, 2021

Published: August 25, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The objective of this study is to analyze the sensitivity of the statistical models regarding the size of samples. The study carried out in Ivory Coast is based on annual maximum daily rainfall data collected from 26 stations. The methodological approach is based on the statistical modeling of maximum daily rainfall. Adjustments were made on several sample sizes and several return periods (2, 5, 10, 20, 50 and 100 years). The main results have shown that the 30 years series (1931-1960; 1961-1990; 1991-2020) are better adjusted by the Gumbel (26.92% - 53.85%) and Inverse Gamma (26.92% - 46.15%). Concerning the 60-years series (1931-1990; 1961-2020), they are better adjusted by the Inverse Gamma (30.77%), Gamma (15.38% - 46.15%) and Gumbel (15.38% - 42.31%). The full chronicle 1931-2020 (90 years) presents a notable supremacy of 50% of Gumbel model over the Gamma (34.62%) and Gamma Inverse (15.38%) model. It is noted that the Gumbel is the most dominant model overall and more particularly in wet periods. The data for periods with normal and dry trends were better fitted by Gamma and Inverse Gamma.

Keywords

Sensitivity of Models, Sample Size, Statistical Models of Extremes, Ivory Coast

1. Introduction

The extreme value theory was developed to estimate the probabilities of occur-

rence of rare events [1]. It is a branch of statistics which is interested in the asymptotic characterization of the maxima or minima of a random variable. It establishes the limiting behavior of the probability law tails of random variables. When these behaviors through parameters have been estimated, it becomes possible to calculate the probability of a large amplitude event [2]. The asymptotic nature of these results, however, calls for caution in the conclusions insofar as we do not have an infinite number of data. The application of the theory of extreme values in the estimation of the recurrence of extreme rains provides essential elements for the construction of infrastructures such as dikes and sanitation work, in order to effectively protect the population and their property [3]. Statistical modeling of extreme values with maximum daily rainfall values is generally preferred to that of daily rainfall above a threshold, both by researchers and planners, because it is easier to apply and often statistically more effective [4]. Thus, several authors have used the same variable of annual maximum daily rains to model extreme rains [3]-[10]. Frequency analysis is the most widely used statistical approach to quantify rainfall hazards [3] [4] [6] [10] [11] [12] [13]. Several authors who have worked with series of annual maximum daily rainfall of different sizes have come to different conclusions. According to the first group [6] [8] [9] [10] [11] [14] with size series ranging from 47 to 81 years, the authors reached the same conclusion according to which Gumbel's law best adjusts the maximum annual rainfall. However, in a study on the long pluviometric series of Athens (136 years), [15] found that the Gumbel law is not adapted to the annual maxima of the series of 136 years, whereas it appeared appropriate if, for example, only the last 34 years are considered. Other authors such as [4] [16], etc. reached conclusions similar to that of [15]. According to [16], a single conclusion is difficult to establish, hence the need to test several laws for each given local station on different sizes in order to appreciate the sensitivity of the laws according to the size of the samples. The question that guides this research is the following: are the statistical laws of extremes used for the statistical modeling of the annual maximum daily rainfall used for the sizing of hydraulic structures from the project flow rate sensitive to the size of the samples of the series?

The objective of this work is therefore to analyze the sensitivity of the statistical laws of extremes according to the size of the samples of the data used.

2. Materials and Methods

2.1. Presentation of the Study Area

Ivory Coast is located in West Africa, in the intertropical zone, between the equator and the tropic of cancer, precisely between latitudes 4°30' and 10°30' North and longitudes 8°30' and 2°30' West (**Figure 1**). It covers an area of 322,462 km² (about 1% of the African continent) and borders with the Gulf of Guinea to the South, Ghana to the East, Liberia and Guinea to the West, Mali and Burkina Faso in the North. **Figure 1** shows the study area which is Ivory Coast.



Figure 1. Presentation of the study area (Ivory Coast).

In Ivory Coast, there are four major climatic zones (**Figure 2**): the tropical transition regime or Sudanese climate in the north, the equatorial regime of attenuated transition or Baoulean climate in the center, the equatorial transition regime or the Attian climate in the South and the mountain regime or mountain climate in the West. Two main types of plant landscapes are present on Ivorian territory: a forest landscape and a savannah landscape. The first covers the

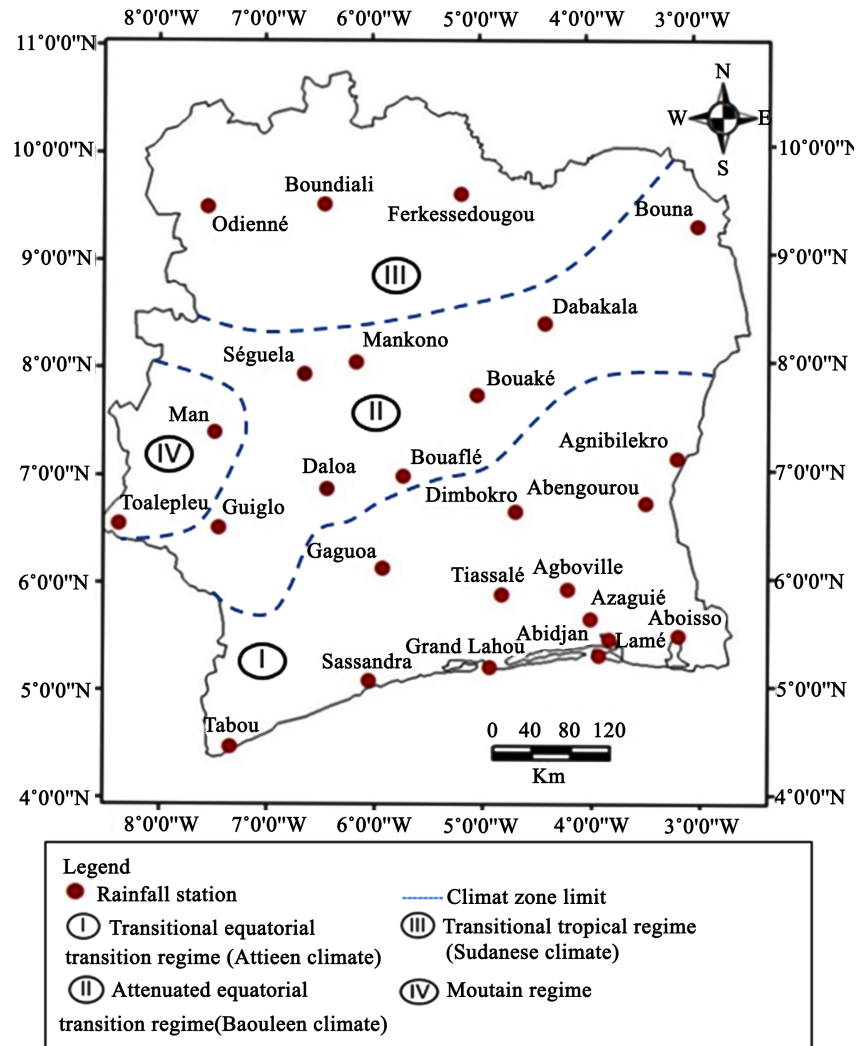


Figure 2. Main climatic zones of Ivory Coast [6].

southern half of the country and belongs to the Guinean domain. The second occupies the northern half of Ivory Coast and is part of the Sudanese domain [17]. The Guinean domain has a predominantly dense humid forest vegetation. Ivory Coast is characterized by a relief not high. Most of the land consists of trays and plains. The west of the country, mountainous region, however, presents some reliefs beyond a thousand meters (the mount nimba culminates at 1,752 m). Apart from this region, altitudes generally vary between 100 and 500 meters, with most plateaus being around 300 to 400 meters. These have different aspects. The highest tops are rigid in their shapes as well as in their materials; those of intermediate levels quite often have blunt shapes; the lower ones have a certain rigidity, but are made of loose materials. Huge and rigorously tabular and horizontal vertical expanses are sometimes present in the savanna regions, but also under the small snags of savannas included in the dense forest. The dominant element of these plates is constituted by a ferruginous armor visible on the surface in the form of rust-colored slabs, but sometimes veiled with sand.

2.2. Data

The data used to carry out this study come from the national meteorological measurement network of Côte d'Ivoire. The annual maximum daily rainfall data used covers the period 1931-2020 and comes from twenty-six (26) rainfall stations distributed throughout the country (**Figure 3**). They were made available to us by SODEXAM (Aeronautical, Airport and Meteorological Development and Exploitation Company). These stations have been classified in the main climatic zones of Ivory Coast (**Table 1**). The choice of stations was guided by the availability and quality of chronological data (fewer gaps with a threshold of 5%).

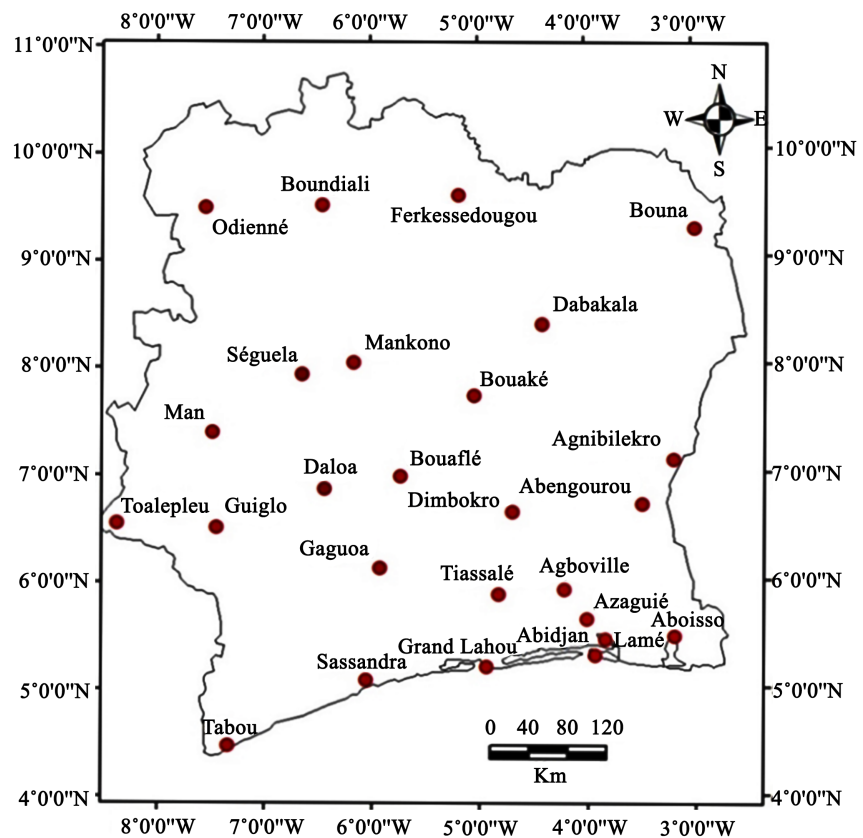


Figure 3. Location of selected rainfall stations.

Table 1. Distribution of stations according to climatic zones.

Zones Climatic	Weather	Stations
Zone I	Equatorial climate of transition	Abidjan, Aboisso, Agboville, Agnibilékrö, Azaguie, Dimbokro, Gagnoa, Grand-Lahou, Lamé, Tabou, Tiassalé, Sassandra
Zone II	Humid tropical climate	Bouaflé, Bouaké, Bouna, Dabakala, Daloa, Guiglo, Mankono, Séguéla
Zone III	Subtropical climate	Boundiali, Ferké, Odienné
Zone IV	Mountain climate	Man, Toulepleu

The various data have undergone preprocessing. Indeed, the method of double cumulations and single residuals were applied to the data of extreme rainfall to identify any erroneous values. The regional vector method and linear regression made it possible to fill in the gaps and correct the values identified as erroneous.

The characteristics of the maximum daily rainfall are presented in **Table 2**. Maximum daily rains vary between 17 (Agboville) and 480 mm (Boundiali) with an average ranging from 73.61 (Agnibilékro) to 136.59 mm (Grand-Lahou). The values of the standard deviation of the different stations oscillate between 20.32

Table 2. Descriptive characteristics of extreme rains (1931-2020).

Stations	Minimum (mm)	Average (mm)	Maximum (mm)	Standard deviation (mm)	CV (%)	Asymmetry coefficient	Kurtosis coefficient
Abengourou	29.82	83.45	234.5	37.13	44.49	1.77	3.99
Abidjan	38.6	136.44	311.6	45.95	33.67	0.8	1.58
Aboisso	42.9	132.77	327.1	55.68	41.94	1.04	1.17
Agboville	17	87.11	294.93	36.22	41.58	2.35	10.55
Agnibilékro	25.2	73.61	190.1	25.42	34.54	1.49	4.13
Azaguie	42	97.42	225.46	36.28	37.24	1.03	1.17
Bouafle	31.08	80.22	223.2	30.87	38.48	1.77	5.16
Bouake	24.43	74.4	196.68	29.23	39.29	1.35	3.19
Bouna	26.04	79.13	211.02	34.36	43.42	1.43	3.01
Dabakala	24.63	82.12	223.39	32.77	39.9	1.54	4.13
Boundiali	26.31	90.41	480	52.91	58.52	4.58	30.47
Daloa	28.2	86.12	193.5	32.22	37.41	1.22	1.92
Dimbokro	31.2	75.93	130	20.32	26.76	0.43	0.22
Ferkessédougou	26.23	75.59	185.72	30.53	40.39	0.77	0.8
Gagnoa	37.5	81.57	246.9	30.96	37.95	2.5	9.96
Grand-Lahou	34.16	136.59	296	52.74	38.61	0.82	0.61
Guiglo	39.16	99.38	264.11	43.89	44.16	1.54	2.98
Lame	39.97	125.4	293.88	44.65	35.6	1.34	2.6
Man	28.99	83.04	198.62	32.11	38.66	0.93	1.61
Mankono	30.93	83.01	157.3	32.67	39.35	0.47	-0.56
Sassandra	37	119.86	253	53.75	44.84	0.58	-0.37
Seguela	27.16	78.68	230.94	35.52	45.15	1.21	2.77
Odienné	27.92	81.25	153.5	29.65	36.49	0.21	-0.48
Tiassale	28.1	87.32	227.18	33.84	38.76	1.74	4.24
Tabou	38.25	136.48	345.7	59.29	43.44	0.43	0.57
Toulepleu	33.5	91.45	269.18	39.46	43.15	2.12	6.68

(Dimbokro) and 59.29 mm (Tabou) with an average of 38.02 mm. The values of the coefficient of variation are all greater than 25%. They vary from 26.76% (Dimbokro) to 58.52% (Boundiali). Thus, the heights of the annual maximum daily rainfall at the various stations are heterogeneous. These rainfall amounts are therefore dispersed in time and space.

As for the flattening coefficient, its values are all greater than zero with values ranging from 0.21 (Odienné) to 4.58 (Boundiali). This shows that the rainfall values of the different stations do not follow the law of normal distribution and have a less flat peak with thicker ends compared to the normal distribution. Asymmetry coefficients are generally positive. They are in proportion of 12% of negative values and 88% of positive values with values ranging from -0.56 (Mankono) to 30.47 (Boundiali); this reflects the fact that the distribution at the level of these data is spread to the right of their mean. As for the stations with negative values, the distribution of their data is spread to the left of their average.

2.3. Statistical Modeling of Annual Maximum Daily Rainfall

The aim is to analyze the influence of the size of the data on the choice of the best statistical law that best represents the extreme rainfall data. This analysis will be done by varying the size of the data in order to assess the possible impacts of the size of the data series on the statistical laws. The sizes of the series considered are 30 years (1931-1960; 1961-1990; 1991-2020) which constitute the 3 WMO reference normals, 60 years (1931-1990; 1961-2020) and 90 years (the entire chronicle 1931-2020). The minimum size of 30 years is chosen in accordance with the recommendations of the WMO, the size of 90 years is fixed according to the availability of data and the size of 60 years constitutes the intermediate size between the two periods (30 years and 90 years).

The methodological approach consisted first of verifying a certain number of statistical hypotheses in the context of the application of frequency analysis, such as independence, homogeneity, stationarity and independence of the data. Then, the different distribution laws chosen were adjusted to the annual maximum daily rainfall data over the different periods defined after the choice of the best classes of laws. Finally, an evaluation of the validity of the preselected models was carried out.

2.3.1. Assumptions of Frequency Analysis

The frequency analysis passes first of all by the appreciation of the quality of the series to have a distribution function by using the stationarity tests of Kendall (Aka *et al.*, 1996), of independence of Wald-Wolfowitz (Hache *et al.*, 1996) and of Wilcoxon homogeneity (Siegel, 1956) (Habibi *et al.*, 2013; Ague *et al.*, 2015). These tests all operate on the same principle, which consists of stating a hypothesis on the mother population and checking on the observed observations whether they are plausible within the framework of this hypothesis. The hypothesis to be tested is called H_0 (null hypothesis) and is imperatively accompanied by its alternative hypothesis called H_1 . The test will focus on validating or reject-

ing H_0 (and consequently drawing the opposite conclusion for H_0). If the result of the test leads to accept the null hypothesis H_0 , then the probability that the distribution is random is high. On the other hand, the rejection of H_0 means that this distribution of the answers conceals information which does not seem to be random and that it is advisable to deepen the analysis.

The Wilcoxon homogeneity test on an annual scale consists of dividing the series according to the different breaks and checking whether this subseries have the same average. The assumptions are therefore:

- H_0 : the means of the two sub-samples are equal;
- H_1 : the means of the two sub-samples are different.

The Kendall stationarity test (1975) cited by Manohar *et al.* (2005), which is a rank correlation test, is used to detect trends in series (Yue and Pilon, 2004). For this test, the assumptions are as follows:

- H_0 : there is no trend in the observations;
- H_1 : there is a trend in the observations.

The test of independence of Wald-Wolfowitz is useful to verify in the observations, the existence of a sequential dependence which would lead, when this one is true, to define the type and the level of this one before continuing the study of the frequency process. For this, the assumptions are as follows:

- H_0 : the observations are independent;
- H_1 : the observations are dependent.

2.3.2. Choice and Estimation of Statistical Model Parameters

The choice of the various statistical models retained for the adjustment of the annual maximum daily rains is based on theoretical considerations and the recommendations of previous work in this area [6] [10] [11] [15] [16] [18]. In general, determining the best fit law has always been tricky and the choice of model can be crucial for estimating precipitation for different extreme value return periods.

After checking the various hypotheses, the frequency analysis is carried out using several statistical tests (Jarque-Bera test, Log-Log graph, mean function of excess (FME), Hill's ratio and Jackson's statistic) [18]. Three main categories in which we can classify the ten most used distributions in hydrology, for the maximum values were distinguished by the decision support system (SAD) of the Hyfran tool (Table 3):

- Class C (regularly varying distribution): Fréchet (EV2), Halphen B Inverse (HIB), Log-Pearson type 3 (LP3), Gamma Inverse (GI);
- Class D (sub-exponential distributions): Halphen type A (HA), Halphen type B (HB), Gumbel (EV1), Pearson type 3 (P3), Gamma (G);
- Class E (exponential law).

The parameters u , α , k respectively denote the position, scale and shape parameters of the different laws. The position parameter u characterizes the order of magnitude of the series of extreme rains. The shape parameter k indicates the behavior of the extremes or the shape of the distribution. According to the sign

Table 3. Hypothesis tests of frequency analysis.

Stations	Independence test		Stationarity test		Homogeneity test	
	U	P	K	P	W	P
Abengourou	2.9	0.038	0.38	0.71	0.037	0.97
Abidjan	1.09	0.28	2.77	0.02	2.82	0.06
Aboisso	0.15	0.88	1.77	0.08	2.34	0.02
Agboville	0.01	0.1	1.81	0.07	2.2	0.03
Agnibilekro	2.88	0.01	0.66	0.51	0.42	0.68
Azaguié	0.36	0.72	2.52	0.01	2.1	0.04
Bouafle	0.71	0.48	2.64	0.01	1.99	0.05
Bouake	2.11	0.04	2.72	0.07	2.29	0.02
Bouna	2.11	0.04	0.8	0.43	0.09	0.93
Boundiali	1.96	0.05	2.18	0.03	0.87	0.39
Dabakala	0.83	0.41	1.17	0.24	0.82	0.40
Daloa	1.15	0.25	2.33	0.02	1.9	0.06
Dimbokro	0.8	0.05	0.03	0.978	0.41	0.69
Ferkessédougou	0.29	0.78	3.6	0.03	4.04	0.05
Gagnoa	2.46	0.01	0.03	0.98	0.84	0.40
Grand-Lahou	0.88	0.38	0.62	0.53	1.36	0.17
Guiglo	0.84	0.4	0.37	0.71	1	0.32
Lame	0.72	0.47	0.20	0.04	2.49	0.01
Man	3.06	0.02	2.23	0.03	0.90	0.37
Mankono	2.68	0.07	3.08	0.02	2.55	0.01
Odienné	3.93	0.01	4.35	0.01	3.31	0.09
Sassandra	2.49	0.01	4.71	0.03	4.22	0.01
Seguela	1.45	0.15	4.45	0.09	3.15	0.02
Tabou	3.96	0.01	3.86	0.01	2.46	0.01
Tiassale	0.78	0.44	1.97	0.05	1.83	0.07
Toulepleu	0.72	0.48	1.39	0.17	0.42	0.68

of this shape parameter, one defines three types of GEV laws:

- $k = 0$, light tail law (or Gumbel distribution);
- $k < 0$, heavy-tailed law (or Fréchet distribution);
- $k > 0$, bounded tail law (or Weibull distribution).

The parameters μ and σ respectively denote the mean and the standard deviation of the lognormal distribution. These parameters were determined by the method of weighted moments.

2.3.3. Calculation of Empirical Frequencies

The determination of the experimental frequencies is based on the critical and

comparative study of the different approaches for the development of empirical probability functions (FPE). Despite the recommendation to use the FPE based on the median of the order statistics as a compromise between the unbiased FPE and the FPE based on the mode of the order statistics, the Hazen formula was retained. In fact, in humid tropical zones, this has been used by most of the authors [10] [11] [14]. After a ranking in ascending order of a sample of maximum rainfall of size n , the expression of the empirical or experimental frequency of non-exceedance of Hazen for a value x of rank i is written (Equation (1)):

$$f(X_i) = \frac{i-0.5}{n} \quad (1)$$

With n the size of the sample considered.

2.3.4. Validation of the Statistical Model

Many techniques exist to compare the different methods of analysis of the laws of probability and to choose the best one. The chi-square fit test was used in this study. Visual examination of the fit charts performed, although it may seem rudimentary, is still a good way to judge the quality of a fit and should always be a preamble to any statistical test. Finally, the criteria of Akaike (AIC) proposed by Akaike (1974) and Bayesian (BIC) proposed by Schwarz (1978) are represented in [9].

The procedure of the applied chi-square test is as follows. Let be a sample of n values classified in ascending (or descending) order and for which a distribution law $F(x)$ has been determined; this sample is divided into a certain number k of classes each containing n_i experimental values. The number v_i is the theoretical number of values assigned to class i by the distribution law. This number v_i is given by (Equation (2)):

$$v_i = n \int_{x_{i-1}}^{x_i} f(x) dx = n [F(x_i) - F(x_{i-1})] \quad (2)$$

$f(x)$ being the probability density function corresponding to the theoretical law.

The expression of experimental χ^2 is presented as follows (Equation (3)):

$$\chi^2 = \sum \frac{(n_i - v_i)^2}{v_i} \quad (3)$$

The probability of overshoot corresponding to the number of degrees of freedom λ is thus determined (with $\lambda = k - 1 - n$, n being the number of parameters of the law $F(x)$). If this probability is greater than 0.05, the fit is satisfactory. Otherwise, the law is rejected.

The selection of the statistical distribution best fitted to the samples was made using two criteria, namely the Akaike criterion and the Bayesian information criterion (BIC). These two criteria make it possible to choose the best-adjusted law, considering the estimation error and parsimony (number of parameters to be adjusted). The distribution for which the values of the two criteria are lower is

the one selected.

In other words, it's about determining the best fit. Indeed, the aim of these criteria is to seek a compromise between sufficient parameterization to properly adjust a probability law to the observations, and the least complex possible parameterization. Such a compromise makes it possible to respect the principle of parsimony of the theoretical frequency distribution laws. Thus, the best law is the one in which the values of AIC and BIC are the lowest compared to the other values of the series of data analyzed. The basic foundations of these criteria are developed below.

The expression of the Akaike information criterion (AIC) is presented as follows [9] [14] (Equation (4)):

$$\text{BIC} = -2 \log(L) + 2k \log(N) \quad (4)$$

where:

- L : the likelihood;
- k : the number of parameters;
- N : the sample size.

The expression of the Bayesian information criterion (BIC) is presented as follows [9] [14] (Equation (5)):

$$\text{AIC} = -2 \log(L) + 2k \quad (5)$$

Or:

- L : the likelihood;
- k : the number of parameters.

2.3.5. Characterization of Return Periods

The best law identified which best adjusts the extreme rains was applied to the daily rainfall heights to characterize the return periods of extreme rainfall events in order to verify whether the rainy episodes, sources of flooding identified could be qualified as extreme events or no. According to [10], the return period, or return time, characterizes the statistical time between two occurrences of a natural event of a given intensity. This term is widely used to characterize natural hazards. The calculation of the frequency of occurrence of extreme rains provides interesting indications for management managers (Equation (6)).

$$T = \frac{1}{1 - F} \quad (6)$$

where:

- T : return period (year);
- F : Frequency of non-exceeding.

A rainy event is qualified as very exceptional if its return period is more than 100 years; exceptional if the return period is between 30 and 100 years; very abnormal if the return period is between 10 to 30 years; abnormal if the return period is between 6 to 10 years and normal if the return period is less than 6 years [10].

3. Results and Discussion

3.1. Analysis of the Hypotheses of Frequency Analysis

Table 3 highlights the results of the various tests applied (Wald-Wolfowitz independence test, Kendall stationarity test and Wilcoxon homogeneity test) on the extreme rainfall data series from 1931-2020.

The results obtained show that all the data from the different stations verify at least two tests at the different thresholds (1% and 5%). The towns of Odienné, Agnibilékro, Sassandra, Bouaflé and Tabou only verify two tests at the different thresholds. The independence test is validated by 92% of the stations, including seventeen (17) at the 5% threshold and seven (7) at the 1% threshold. For the stationarity test, 92% of the stations validated it, including thirteen (13) at the 5% threshold and eleven (11) at the 1% threshold. As for the homogeneity test, it is validated at 96% of stations, including sixteen (16) at the 5% threshold and nine (9) at the 1% threshold. In sum, 81% of the stations verify all the hypothesis tests. Thus, it is therefore possible to proceed with the frequency analysis.

3.2. Identification of the Best Classes of Laws

The summary of the different selected classes is presented in **Table 4**. The best classes for the 30-year series are classes C with a percentage of appearance of 57.47% and E with a percentage of appearance of 42.53%. Class C appears twenty-three (23) times alone as the best class against four (4) times for class E. For the 60-year-old heats, the best classes are classes C, D and E. Class C has an on-set percentage of 46.84%, class E has a percentage of 29.57% and class D has 23.59%. Class C appears (2) times alone as the best class against (4) times for each of the E and D classes. As for the 1931-2020 series, it presents the D class (sub-exponential distribution) in addition to the C classes and E as the best classes. 42.86% of the appearance of class C is observed against 35, 71% for class E and 21.43% for class D. Classes C, D and E appear alone as best classes, respectively, two (2), six (6) and two (2) times. The lognormal law by means of the Jarque-Bera test is inapplicable on all the data.

3.3. Identification of the Best Laws

3.3.1. Graphical Analysis of Adjustments

The identification of the best laws goes through the adjustment of the laws resulting from the different selected distribution classes. The preliminary

Table 4. Percentages of distribution of classes of laws over the different periods.

Series Size	Period	% C	% D	% E
30 years	1931-1960	58.54	0	41.46
	1961-1991	54.35	0	45.65
	1991-2020	59.52	0	40.48
60 years	1931-1990	45.84	14.58	39.58
	1961-2020	47.83	32.61	19.56
90 years	1931-2020	42.86	21.43	35.71

element required in order to arrive at this identification is the graphic analysis (Figures 4-6). On all the graphs of the different series, the laws which best adjust the extreme rains are the Gamma Inverse, Fréchet, Weibull, Gamma, Exponential, Log-pearson type 3, Pearson type 3 and Gumbel laws.

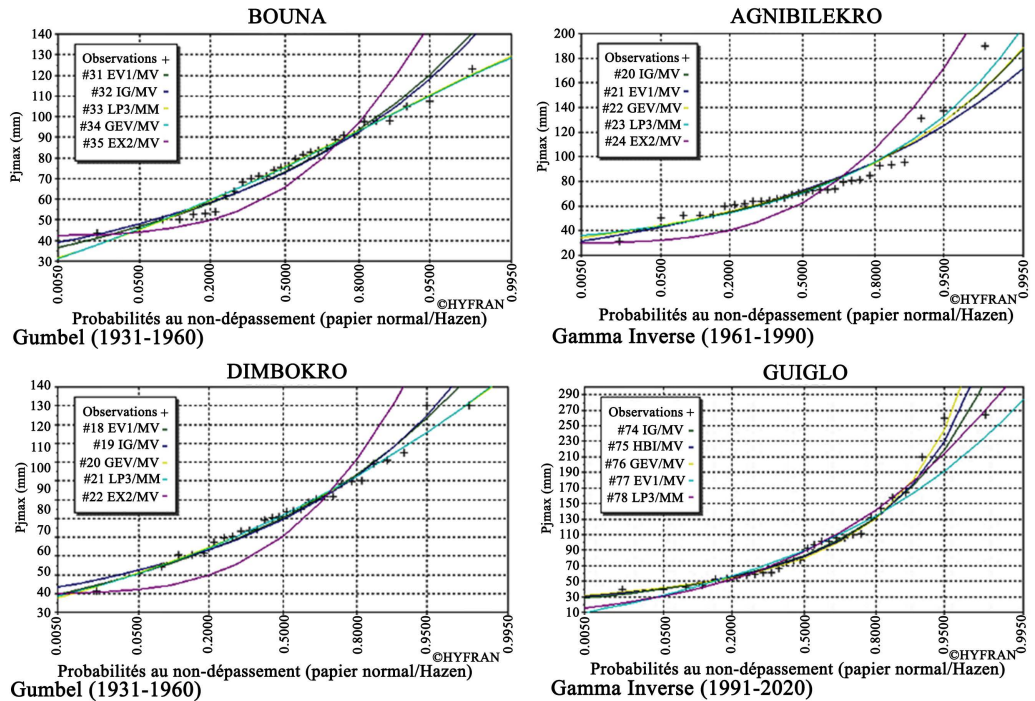


Figure 4. Comparison of adjustment laws over 30 years periods.

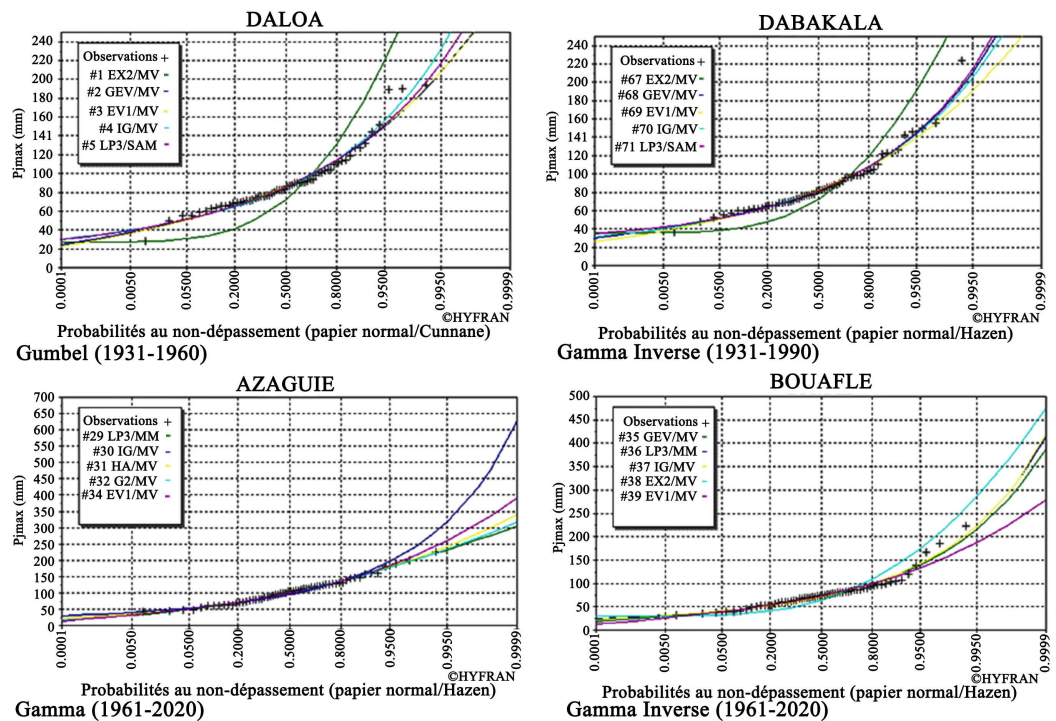


Figure 5. Comparison of adjustment laws over 60 years periods.

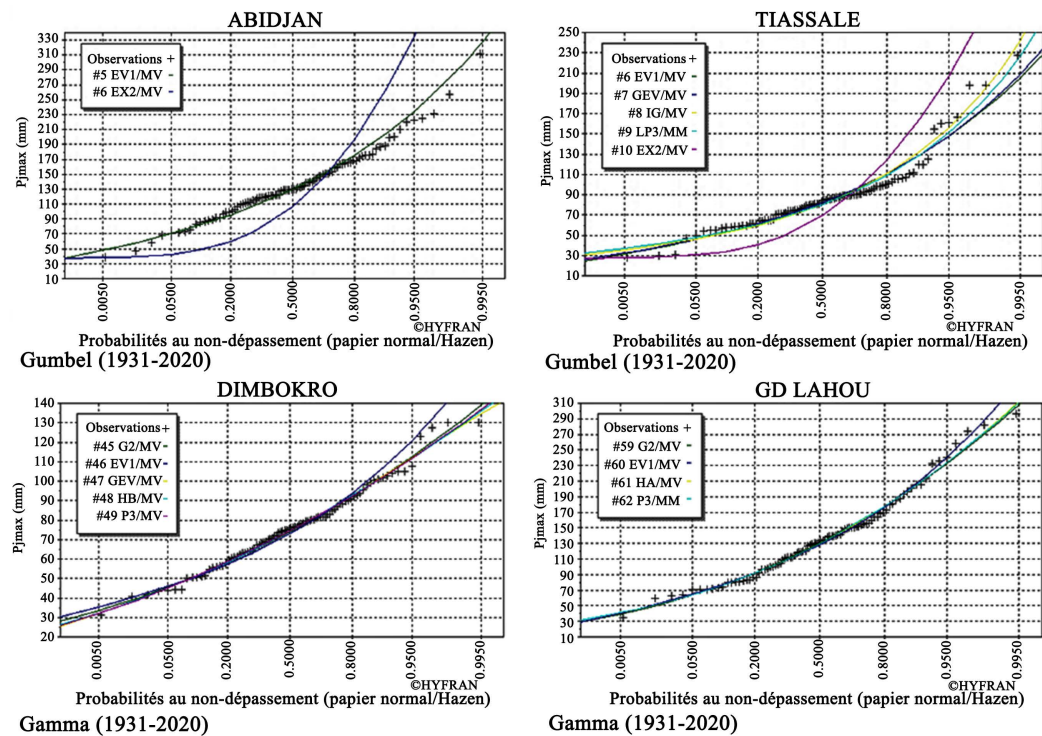


Figure 6. Comparison of the adjustment laws over the entire period (1931-2020).

3.3.2. Numerical Analysis of Adjustments

After the adjustments, the numerical chi-square test of the adequacy of these was applied to better assess their relative quality. The application of this test proved to be conclusive for all the adjustment laws on the annual maximum daily rainfall at the significance level of 5%. The classification of the best laws on the basis of the AIC and BIC criteria has been carried out. **Figures 7-9** present histograms of the distribution of the best laws respectively on the series of 30, 60 and 90 years. These histograms make it possible to appreciate the evolution of the laws according to the size of the period used. We notice a certain evolution of the laws on each data series.

Thus, at the level of the 30-year series, the following percentages are observed for all the stations (**Figure 7**):

- 1931-1960: Weibull's law is adjusted on one (1) station, *i.e.* 3.85%, Frechet and Exponential laws on two (2) stations each, *i.e.* 7.69%, Inverse Gamma law on seven (7) stations, *i.e.* 26.92% and Gumbel's law on fourteen (14) stations, *i.e.* 53.85%;
- 1961-1990: Weibull's law is adjusted on one (1) station, *i.e.* 3.85%, Frechet's law on three (3) stations, *i.e.* 11.54%, Gumbel's law on ten (10) stations *i.e.* 38.46% and the Inverse Gamma law on twelve (12) stations, *i.e.* 46.15%;
- 1991-2020: the laws of Frechet and Log pearson type 3 are adjusted on one (1) station each, *i.e.* 3.85%, the Exponential law on six (6) *i.e.* 23.08%, Gumbel's law on seven (7) stations, *i.e.* 26.92% and the Inverse Gamma law on eleven (11) stations, *i.e.* 42.31%.

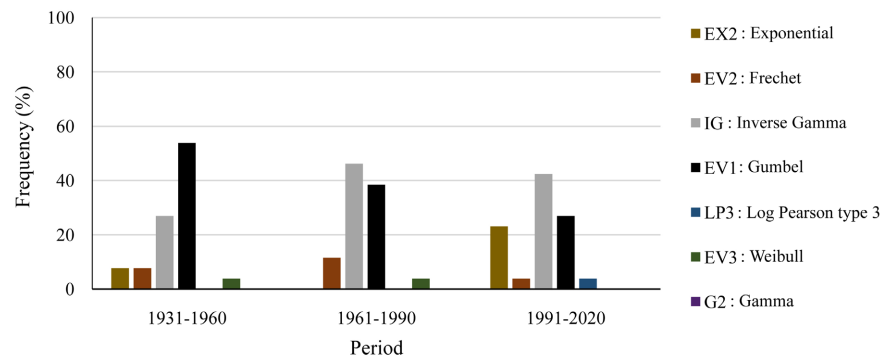


Figure 7. Best laws for the 30-year series.

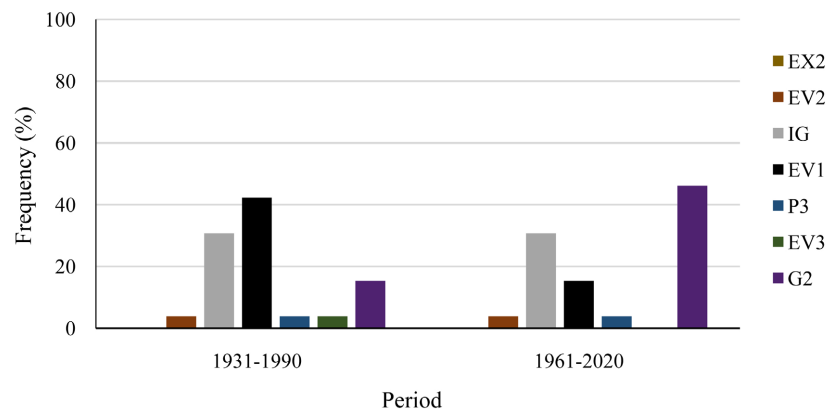


Figure 8. Best laws for the 60-year series.

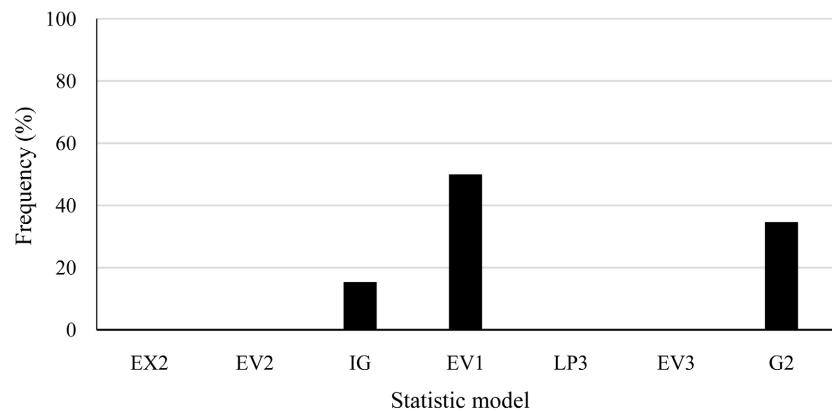


Figure 9. Best laws for the 90 years series.

In general, for the 30-year series, the supremacy of Gumbel's law deteriorates with a decrease in the occurrence rate from 53.85% for the historical normal of 1931-1960 to 26.92% for the updated normal of OMM (1991-2020) passing by 38.46% of appearance for the past normal (1961-2020). The last two thirty-year periods reverse Gumbel's trend in favor of Gamma Inverse with a respective probability of 46.15% and 42.31%. We note that the old series which integrates very old data such as 1931-1960, 1931-1990 and 1931-2020 displays Gumbel as the first law for most of the stations in the study. While the most recent

1961-1990, 1991-2020 and 1961-2020 show a dominance of the Gamma and inverse Gamma laws. This could mean that Gumbel's law fits wet extremes well and Gamma and Gamma Inverse laws lend themselves better to dry extremes. The representativeness of laws such as Fréchet, Weibull and Log Pearson 3 and Pearson is very low, especially when the size of the series is very large.

Indeed, over all the 30-year series, we notice that the laws of Gumbel and Gamma Inverse are those which fit best. Also, the closer we get to recent periods, the more the representativeness of Gumbel's law decreases and that of the Inverse Gamma law increases. We therefore observe a predominance over the last two 30-year periods (1961-1990 and 1991-2020) of the Gumbel and Gamma Inverse laws. It is noted an instability of the three best laws at the level of series of normal size ($n = 30$ years). The OMM historical normal, 1931-1960 tends to fit well by Gumbel's law, the 1961-1990 past normal used is best fitted by the Inverse Gamma law followed by Gumbel's law.

Regarding the 60 years series, they present the following observations (**Figure 8**):

- 1931-1990: Pearson's type 3, Frechet and Weibull laws are adjusted on one station each, *i.e.* (3.35%), the Gamma law on four (4) stations *i.e.* 15.38%, the Inverse Gamma law on eight (8) or 30.77% and Gumbel's law on eleven (11) stations or 42.31%;
- 1961-2020: Pearson type 3 and Frechet laws are adjusted on one station each, *i.e.* (3.35%), Gumbel's law on four (4) stations *i.e.* 15.38%, Inverse Gamma law on eight (8) stations, *i.e.* 30.77% and the Gamma law on twelve (12) stations, *i.e.* 46.15%.

Over the periods of the 60-year series, the laws of Gumbel, Gamma Inverse and Gamma are those which fit best (**Figure 8**). A certain stability of the Inverse Gamma law is observed. It is also noted that the closer we get to recent data, the more the representativeness of Gumbel's law decreases and that of the Gamma law increases. Here, we notice an inverse evolution of the laws of Gumbel and Gamma. From the age of 60, the first best law remains either Gumbel's law (EV1) or the Gamma law (G2) and the Inverse Gamma law (IG) remains the second-best law of the 60-year series. This marks a certain stability of the laws for the series of average size ($n = 60$ years).

As for the entire series, the Inverse Gamma law is adjusted on four (4) stations, *i.e.* 15.38%, the Gamma law on nine (9) stations, *i.e.* 34.62% and Gumbel's law on thirteen (13) stations, *i.e.* 50% (**Figure 9**). Indeed, the complete chronicle of 1931-2020 presents a notable supremacy of the law of extreme values of type 1 (Gumbel) over the Gamma and Gamma Inverse laws. The laws of Weibull, Frechet, Exponential, Log pearson type 3 and Pearson type 3 have a low representativeness.

3.4. Discussion

The main results showed that the 30 years series are better adjusted by the Gumbel (26.92% - 53.85%) and Inverse Gamma (26.92% - 46.15%) laws. The

supremacy of Gumbel's law deteriorates from the wet period (1931-1960) with an occurrence rate of 53.85% in the dry period (1991-2020) (26.92%) in favor of the Gamma law Inverse with respective probabilities of 46.15% (1961-1990) and 42.31% (1991-2020). Regarding the 60-year series, they are better adjusted by the inverse Gamma (30.77%), Gamma (15.38% - 46.15%) and Gumbel (15.38% - 42.31%) laws. The Gumbel law is predominant over the first relatively wet period (1931-1960) with an occurrence rate of 42.31%. As for the second relatively less humid period (1991-2020), it is dominated by the Gamma (46.15%) and Inverse Gamma (30.77%) laws. The complete chronicle of 1931-2020 presents a notable supremacy of 50% of the Gumbel law over the Gamma (34.62%) and Gamma Inverse (15.38%) laws. It is noted that the Gumbel law is the most dominant law overall and more particularly in wet periods. The data for periods with normal and dry trends were better fitted by the Gamma and Inverse Gamma laws. The laws used to adjust the annual maximum daily rainfall data are therefore sensitive to the size of the samples and to the climatic context of the series. Indeed, the laws are more stable when the size of the data series becomes large (at least 60 years) and when the series considers a wet component (before 1970) and a dry component (after 1970).

The sensitivity of the statistical laws applied to extreme rains has shown that all the probability distributions of the annual maximum daily rains in Ivory Coast could not only be assimilated to a single law regardless of the size and climatic context of the data series. The use of the DSS tool (decision support system) for frequency analysis has revealed certain laws such as the Gamma and Gamma Inverse laws which are rarely used for the adjustment of rainfall maxima. These results reflect the sensitivity of statistical models for adjusting extreme values to the size of the data samples and to the climatic context.

Several authors who have worked with series of annual maximum daily rainfall of different sizes have reached conflicting conclusions. According to the first group, for large series (47 - 81 years), the authors concluded that Gumbel's law is predominant over other laws (lognormal, Fréchet, Weibull, GEV, etc.). However, according to the second group, the Gumbel law is not predominant over the other laws for large series.

In the first case, several studies have been carried out. Thus, the work of [6] [14] based on the frequency analysis of maximum daily annual rains on the one hand of 34 rainfall stations in Ivory Coast covering the period 1947-1995 (49 years) and on the other hand on 47 Ivorian rainfall stations with annual maximum rainfall data covering the period 1947-1993 (47 years), came to the same conclusion that Gumbel's law and lognormal law best adjust the annual maximum rainfall. According to the work of [11], carried out in Ivory Coast on the maximum daily annual rains during the period 1942-2002 (61 years), the best laws retained are respectively Gumbel's law (34.1%), Fréchet's law (29.5%), Lognormale law (22.7%) and the Weibull law (13.6%). According to the study by [10] from annual maximum daily rainfall data cover the period from 1961 to 2014 (54 years) on the Abidjan station (Port-Bouët), the law which best fits these

data is Gumbel's law. The results of the work of [9] carried out using data from 35 stations covering the period from 1921 to 2001 (81 years) overall, showed a predominance of the Gumbel (51.43%) and Lognormale (28, 57%). The work of [10] carried out in Benin (Sota basin) from daily rainfall data ranging from 1965 to 2008 (44 years) from eight (8) stations showed that Gumbel's law and the log-Pearson Type law III are the predominant laws.

In the second case, several studies have also been carried out. According to [4], a frequency analysis of annual series of maximum daily rainfall was carried out on data from 27 rainfall stations from the period 1970 and 2005 (35 years) of the Chott Chergui basin (Algeria). The GEV law has shown a good adequacy to the series of maximum daily rains in the Chott Chergui basin (Algeria). [16] came to the same conclusion as Habibi *et al.* (2013) regarding the supremacy of the GEV law. In a study on the long pluviometric series of Athens (136 years), [15] found that the Gumbel law is not adapted to the annual maxima of the series of 136 years, whereas it seemed appropriate if for example, only the last 34 years are considered. The results of the work of [3] in the Cheliff region have shown from a comparative analysis between the GEV and Gumbel methods based on four data samples spread over a period of between 21 and 30 years, the GEV methods and Gumbel provided similar results.

The results obtained during this work are more in agreement with the results of the first group. Beyond the size, the difference in results could be due to the climatic context. These results also raise the debate on the skepticism of the predominance of the model of Gumbel at the level of the estimate of the annual maximum daily rains. It can be said that the Gumbel law is therefore not always the best model in the adjustment of extreme rains.

4. Conclusions

The objective of this study is to analyze the sensitivity of the statistical laws of extremes as a function of the size of the data samples. The methodological approach is based on the statistical modeling of annual maximum daily rainfall. The adjustments were made on several sample sizes, namely samples of 30 years (1931-1960; 1961-1990; 1991-2020), 60 years (1931-1990; 1961-2020) and 90 years (1931-2020). Several return periods (2, 5, 10, 20, 50 and 100 years) were retained.

The results of the tests prior to the frequency analysis indicated that the hypotheses of application to the frequency analysis were verified on almost all the series. Then, the series of annual maximum daily rains are constituted by independent, homogeneous and stationary values. Indeed, the independence test is validated by 92% of the stations, including seventeen (17) at the 5% threshold and seven (7) at the 1% threshold. For the stationarity test, 92% of the stations validated it, including thirteen (13) at the 5% threshold and eleven (11) at the 1% threshold. As for the homogeneity test, it is validated at 96% of the stations, including sixteen (16) at the 5% threshold and nine (9) at the 1% threshold. In sum, 81% of the stations verify all the hypothesis tests. Thus, it is therefore

possible to proceed with the frequency analysis.

According to the Decision Support System (SAD), the best classes for the 30 years series are classes C (57.47%) and E (42.53%). Class C appears twenty-three (23) times alone as the best class against four (4) times for class E. For the 60-years, the best classes are classes C (46.84%), D (29.57%) and E (23.59%). As for the 1931-2020 series, it presents respectively as the best classes the classes C (42.86%), E (35.71%) and D (21.43%). The lognormal law by means of the Jarque-Bera test is inapplicable on all the data.

The main results showed that the 30-years series were better adjusted by the Gumbel (26.92% - 53.85%) and Inverse Gamma (26.92% - 46.15%) laws. The supremacy of Gumbel's law deteriorates from the wet period (1931-1960) with an occurrence rate of 53.85% in the dry period (1991-2020) (26.92%) in favor of the Gamma law Inverse with respective probabilities of 46.15% (1961-1990) and 42.31% (1991-2020). Regarding the 60-years series, they are better adjusted by the inverse Gamma (30.77%), Gamma (15.38% - 46.15%) and Gumbel (15.38% - 42.31%) laws. The Gumbel law is predominant over the first relatively wet period (1931-1960) with an occurrence rate of 42.31%. As for the second relatively less humid period (1991-2020), it is dominated by the Gamma (46.15%) and Inverse Gamma (30.77%) laws. The complete chronicle of 1931-2020 presents a notable supremacy of 50% of the Gumbel law over the Gamma (34.62%) and Gamma Inverse (15.38%) laws. It is noted that the Gumbel law is the most dominant law overall and more particularly in wet periods. The data for periods with normal and dry trends were better fitted by the Gamma and Inverse Gamma laws. The laws used to adjust the annual maximum daily rainfall data are therefore sensitive to the size of the samples and to the climatic context of the series. Indeed, the laws are more stable when the size of the data series becomes large (at least 60 years) and when the series considers a wet component (before 1970) and a dry component (after 1970).

Acknowledgements

The authors of this article thank the instructors whose reviews and suggestions have improved this article. They also thank SODEXAM for providing them with the rainfall data used in this study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Embrechts, P., Klüppelberg, C. and Mikosch, T. (2003) *Modelling Extremal Events for Insurance and Finance Applications of Mathematic*. Springer-Verlag, Berlin, Germany, 648 p.
- [2] Garrido, M. (2002) *Modélisation des événements rares et estimations des quantiles extrêmes, méthodes de sélections des modèles pour les queues de distributions*.

- Thèse de Doctorat de Sciences, Université Joseph Fourier, Saint-Martin-d'Hères, 244 p.
- [3] Benkhaled, A. (2007) Distributions statistiques des pluies maximales annuelles dans la région du Cheliff: comparaison des techniques et des résultats. *Courrier du Savoir*, **8**, 83-91.
 - [4] Habibi, B., Meddia, M. and Boucefiane, A. (2013) Analyse fréquentielle des pluies journalières maximales: Cas du Bassin-Chergui. *Revue Nature & Technologie, Sciences de l'Environnement*, **8**, 41-48.
 - [5] Onibon, H., Ouarda, T., Bobee, B., Barbet, M., Saint-Hilaire, A. and Bruneau, P. (2004) Regional Frequency Analysis of Annual Maximum Daily Precipitation in Quebec. *Hydrological Sciences Journal*, **4**, 717-735.
 - [6] Goula, B.T.A., Konan, B., Brou, Y., Savane, I., Vamoryba, F. and Srohourou, B. (2007) Estimation des pluies exceptionnelles journalières en zone tropicale: Cas de la Ivory Coast par comparaison des lois Lognormale et de Gumbel. *Hydrological Sciences Journal*, **1**, 49-67. <https://doi.org/10.1623/hysj.52.1.49>
 - [7] Zahar, Y. and Laborde, J.P. (2007) Modélisation statistique et synthèse cartographique des pluies journalières extrêmes de Tunisie. *Revue des Sciences de l'Eau*, **4**, 409-424. <https://doi.org/10.7202/016914ar>
 - [8] Koumassi, D., Tchibozo, A.E., Vissin, E. and Houssou, E. (2014) Analyse fréquentielle des événements hydro-pluviométriques extrêmes dans le bassin de la Sota au Bénin. *Afrique Science*, **2**, 137-148.
 - [9] Ague, A.I. and Afouda, A. (2015) Analyse fréquentielle et nouvelle cartographie des maxima annuels de pluies journalières au Bénin. *International Journal of Biological and Chemical Sciences*, **1**, 121-133. <https://doi.org/10.4314/ijbcs.v9i1.12>
 - [10] Kouassi, A.M., Nassa, R.A.K., Koffi, B.Y. and Biemi, J. (2018) Modélisation statistique des pluies maximales annuelles dans le district d'Abidjan (Sud de la Côte d'Ivoire). *Revue des Sciences de l'eau*, **2**, 147-160. <https://doi.org/10.7202/1051697ar>
 - [11] Soro, G. (2011) Modélisation statistique des pluies extrêmes en Côte d'Ivoire. Thèse de Doctorat, l'Université Nangui-Abrégoua, Sciences et Gestion de l'Environnement, Abidjan, 193 p.
 - [12] Benabdesselam, T. and Amarchi, H. (2013) Approche régionale pour l'estimation des précipitations journalières extrêmes du Nord-Est algérien. *Courrier du Savoir*, **17**, 175-184.
 - [13] Neppel, L., Arnaud, P., Borch, F., Carreau, J., Garavaglia, F., Lang, M., Paquet, E., Renard, B., Soubeyroux, J.M. and Veyssière, J.M. (2014) Résultats du projet Extraflo sur la comparaison des méthodes d'estimation des pluies extrêmes en France. *La Houille Blanche*, **2**, 14-19. <https://doi.org/10.1051/lhb/2014011>
 - [14] Goula, B.T.A., Soro, G.E., Dao, V.T., Kouassi, F.W. and Srohourou, B. (2010) Frequency Analysis and New Cartography of Extremes Daily Rainfall Events in Côte d'Ivoire. *Journal of Applied Sciences*, **10**, 1684-1694. <https://doi.org/10.3923/jas.2010.1684.1694>
 - [15] Koutsoyiannis, D. and Baloutsos, G. (2000) Analysis of a Long Record of Annual Maximum Rainfall in Athens, Greece, and Design Rainfall Inferences. *Environmental Science, Natural Hazards*, **1**, 29-48. <https://doi.org/10.1023/A:1008001312219>
 - [16] Muller, A. (2006) Comportement asymptotique de la distribution des pluies extrêmes en France. Thèse de Doctorat ès Science, Université Montpellier II (France), Montpellier, 246 p.
 - [17] Brou Y.T. (2005) Climat, mutations socio-économiques et paysages en Côte d'Ivoire.

Mémoire de Synthèse des Activités Scientifiques Présenté en vue de l'obtention de l'Habilitation à Diriger des Recherches, Université des Sciences et Techniques de Lille (France), Villeneuve-d'Ascq, 212 p.

- [18] El Adlouni, S., Bobee, B. and Ouarda, T.B.M.J. (2008) On the Tails of Extreme Event Distributions. *Journal of hydrology*, **1-4**, 16-33.
<https://doi.org/10.1016/j.jhydrol.2008.02.011>