

Estimations Des Quantiles Extrêmes Des Précipitations Convectives Avec Les Distributions des Valeurs Extrêmes Généralisées : Cas Des Plaines De La Basse Betsiboka

[Extreme Quantile Estimation Of Convective Precipitation With Generalized Extreme Values Distributions: Case Of The Plains Of Lower Betsiboka]

RAZAFINDRALAMBO Hasina Tahiananirina¹, RAZAFIMANDIMBY Honoré², RAMAHAZOSOA Irrish Parker³, RABEHARISOA Jean Marc⁴, RATIARISON Adolphe⁵

^{1,2,3,4,5} Université d'Antananarivo, Faculté des Sciences, Laboratoire de la Dynamique de l'Atmosphère, du Climat et des Océans (DyACO), B.P. 906, Antananarivo, Madagascar,

hasinatahiananirina@gmail.com¹



Résumé – L'objectif de cette étude est d'estimer la fréquence de retour des précipitations convectives extrêmes dans les plaines de la basse Betsiboka, Madagascar. L'approche méthodologique consiste à trouver les distributions statistiques des pluies convectives journalières maximales annuelles de 1959 à 2021 en les ajustant avec des distributions des valeurs extrêmes généralisées. Les paramètres des distributions des valeurs extrêmes généralisées ont été estimés par la méthode des moments pondérés. Les résultats obtenus montrent que l'ajustement des données étudiées avec les distributions des valeurs extrêmes généralisées sont adéquats. En outre, les maxima annuels suivent des distributions à queues lourdes. Les pluies convectives extrêmes sont imprévisibles et engendrent des crues et inondations qui pourront avoir des conséquences économiques et sociales graves dans cette zone. L'estimation de la récurrence de ces événements météorologiques extrêmes est donc d'un grand intérêt.

Mots clés – Précipitation convective, maximale annuelle, distribuion GEV, méthode des moments pondérés, basse Betsiboka

Abstract – The objective of this study is to estimate the frequency of return of the annual maximum convective daily rains of the two plains of the lower Betsiboka, Madagascar. Our approach consists in finding the statistical distributions of the annual maximum daily rainfall from 1959 to 2021 by adjusting them with the generalized extreme values laws. The parameters of the generalized extreme values laws were estimated by the method of weighted moments. The results show that the generalized extreme values laws are adequate to adjust the series of annual maximum daily rainfall in this zone. Extreme rains create risks of disasters that could have serious economic and social consequences in this area. Estimating the recurrence of these meteorological events is therefore of great interest.

Keywords – Annual maximum, convective precipitation, GEV distribution, probability weighted moments, lower-Betsiboka

I. INTRODUCTION

L'estimation des pluies extrêmes faisait l'objet de nombreux travail pendant la dernière décennie [1], [2]. Elle constitue un outil indispensable au décideur pour la prévention contre les risques d'inondation. La basse vallée de la Betsiboka contient de vastes plaines, entre autres, la plaine d'Ambato-Boeny et celle de Marovoay qui est le deuxième grenier à riz du pays. Cette zone est

exposée à de fréquentes crues et inondations engendrant des conséquences économiques et sociales catastrophiques. Afin de se prémunir contre ces inondations, l'estimation de la récurrence des pluies convectives extrêmes est d'un grand intérêt. Dans cette étude, pour des raisons d'efficacité [3], les pluies extrêmes sont définies par les valeurs maximales annuelles. Théoriquement, la distribution du maximum annuelle suit une loi des Valeurs Extrêmes Généralisées (GEV) [4]. De ce fait, la loi la plus couramment utilisée en modélisation des maxima annuels est la loi GEV à deux paramètres [5], [6]. Toutefois, d'autres études [7], [8] remettent en cause cette usage systématique de la distribution de Gumbel. Nous allons dans cette étude identifier la distribution GEV la plus adéquate pour modéliser les pluies extrêmes de la basse vallée de la Betsiboka.

II. MÉTHODOLOGIE

Les données pluviométriques utilisées dans ce travail sont les hauteurs de pluies journalières du 01 janvier 1959 au 31 décembre 2021 aux points de coordonnées géographiques situés au 16,25°, respectivement 17°, de latitude Sud et 46,50°, respectivement 46,75°, de longitude Est, correspondant respectivement à la plaine de Marovoay et à celle de Maevatanana (*Figure 1*). Ce sont des données satellites, issues de l'agence européenne « European Centre for Medium range Weather Forecasts (ECMWF) ». La basse vallée de la Betsiboka est localisée dans la partie nord-ouest de Madagascar, elle est exposée à de fréquentes inondations.

Concernant les traitements numériques, le logiciel utilisé a été le MATLAB 2018. Ce logiciel fournit des fonctions pour calculer les paramètres de la distribution GEV avec la méthode du maximum de vraisemblance mais pas avec la méthode des moments pondérés.

Notre démarche de l'analyse fréquentielle des maxima annuels est constituée des cinq étapes suivantes : (a) vérification de l'homogénéité, de l'indépendance et de la stationnarité de la série des maxima annuels ; (b) estimation des paramètres des lois de distribution GEV ; (c) vérification de la validité des modèles obtenus ; (d) choix du meilleur modèle ; et enfin (e) estimation des pluies maximales annuelles correspondant aux différents temps de retour.

2.1 Modélisation des valeurs extrêmes

La théorie sur la modélisation des valeurs extrêmes avec les distributions GEV est la suivante. Soit $(X_t)_{t=1,2,\dots}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d). Celle-ci est divisée en bloc de même longueur n . Soit $(Z_k)_{k=1,2,\dots}$ la suite des variables aléatoires des maxima de chaque bloc,

$$Z_k = \max(X_{(k-1)n+1}, X_{(k-1)n+2}, \dots, X_{kn}) \quad (1)$$

D'après Fisher et Tippet (1928) [4], si n est assez grand (par exemple $n = 365$), alors la distribution du maximum Z_k , $k = 1, 2, \dots$ est approximativement une loi GEV.

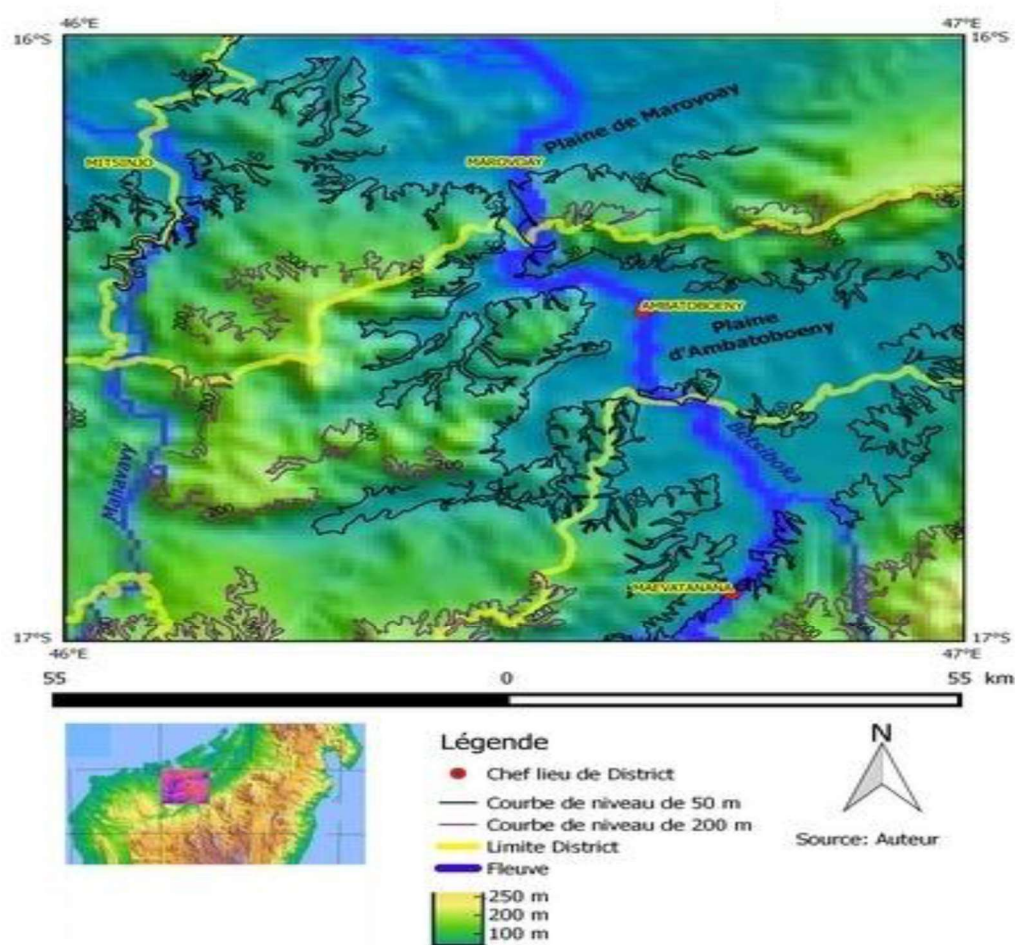


Figure 1 : Localisation de la zone d'étude.

2.2 Vérification des hypothèses fondatrices des modèles GEV

Pour que la modélisation de la série des maxima annuels par les distributions GEV soit théoriquement valable, les observations des précipitations convectives journalières doivent être les réalisations d'une suite de variable aléatoire i.i.d. Il est donc primordial de vérifier d'abord l'indépendance, mais aussi l'homogénéité et la stationnarité des données des précipitations journalières, avant de passer à l'ajustement. Toutefois, selon Coles (2001) [9], il suffit de vérifier ces hypothèses sur la série des maxima. Pour cela, le test d'indépendance de Wald-Wolfowitz [10], le test d'homogénéité de Wilcoxon [11] et le test de stationnarité de Kwiatkowski, Phillips, Schmidt et Shin (KPSS) [12] ont été utilisées.

Le test d'indépendance de Wald-Wolfowitz permet de vérifier l'hypothèse nulle selon laquelle « les observations sont indépendantes ». Le test d'homogénéité de Wilcoxon consiste à découper la série selon différentes ruptures et à vérifier l'hypothèse nulle selon laquelle « les moyens de ces sous-échantillons sont égales ». Quant au KPSS test, l'hypothèse nulle est que « la série est stationnaire ».

2.3 Estimation des paramètres des lois de distributions GEV

La fonction de distributions des lois GEV à trois paramètres est de la forme,

$$F(x) = \exp \left[- \left(1 + \kappa \left(\frac{x - \psi}{\lambda} \right) \right)^{-\frac{1}{\kappa}} \right], \quad (2)$$

avec $\psi \in \mathbb{R}, \lambda > 0, \kappa \in \mathbb{R}$ les paramètres respectifs de position, d'échelle et de forme. Ce dernier paramètre caractérise le comportement de la queue de la distribution. Pour $\kappa < 0$, la distribution a une queue bornée, elle est appelée distribution de Weibull. Pour $\kappa > 0$, la distribution a une queue lourde, elle est appelée distribution de Fréchet.

Pour les lois GEV à deux paramètres, le paramètre de forme $\kappa = 0$. La fonction de distribution s'écrit alors :

$$F(x) = \exp \left[- \exp \left(\frac{x - \psi}{\lambda} \right) \right]. \quad (3)$$

Cette dernière distribution est à queue légère. Elle correspond à la distribution de Gumbel.

Il existe diverses méthodes pour estimer les paramètres d'une distribution statistique. Dans cette étude, nous utiliserons la méthode des moments (MOM) pour l'estimation des paramètres des distributions de Gumbel et la méthode des moments pondérés (PMW) pour celle des distributions GEV à trois paramètres [13]. La méthode d'estimation des paramètres des distributions GEV avec les moments est la suivante [14]. Soit $x = (x_j)_{j=1, \dots, n}$ la série des maximas triées par ordre croissant et X la variable aléatoire de fonction de répartition $G(x)$. Supposons que la variable aléatoire X suit une loi GEV.

Notons

$$\beta_r = E \left(X \{G(X)\}^r \right), \quad r = 0, 1, 2, \dots \quad (4)$$

Alors,

$$b_r = \frac{1}{n} \sum_{j=1}^n \frac{(j-1)(j-2) \dots (j-r)}{(n-1)(n-2) \dots (n-r)} x_j \quad (5)$$

est un estimateur sans biais de β_r .

De plus, les paramètres de la distribution GEV à trois paramètres peuvent être estimés par les formules :

$$\hat{\kappa} = 7,8590c + 2,9554c^2, \quad (6)$$

$$\hat{\lambda} = \frac{(2b_1 - b_0) \hat{\kappa}}{\Gamma \left(1 + \hat{\kappa} \right) \left(1 - 2^{-\hat{\kappa}} \right)}, \quad (7)$$

$$\hat{\psi} = b_0 + \frac{\hat{\lambda}}{\hat{\kappa}} \left\{ \Gamma \left(1 + \hat{\kappa} \right) - 1 \right\}, \quad (8)$$

avec $c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\log 2}{\log 3}$, et Γ la fonction gamma.

En particulier, les paramètres de la distribution de Gumbel sont donnés par :

$$\hat{\lambda} = \sqrt{6} \frac{\pi}{\delta} \quad (9)$$

$$\hat{\psi} = b_0 - 0,5772 \hat{\lambda}. \quad (10)$$

où δ étant l'écart-type.

2.4 Vérification de la validité des distributions obtenues

Plusieurs tests peuvent être utilisés pour vérifier si les distributions théoriques rendent bien compte des observations. Parmi ceux qui sont les plus utilisés, on trouve les tests d'adéquation non paramétriques de Chi-2 et celui de Kolmogorov-Smirnov. Dans cette étude, c'est ce dernier test qui a été employé.

Le principe du test de Kolmogorov-Smirnov consiste à mesurer l'écart maximum qui existe entre la fonction de répartition empirique et la fonction de répartition théorique. Ainsi, l'hypothèse nulle est que les données proviennent de la distribution théorique fixée à l'avance. A cet effet, la fréquence empirique de Hazen [15] a été utilisée.

2.5 Choix du meilleur modèle

Nous utiliserons les critères d'information d'Akaike (AIC) et le critère d'information Bayésienne (BIC) [16] pour le choix du meilleur modèle.

2.6 Estimation des quantiles

La période de retour d'un événement donné représente l'intervalle de temps moyen séparant deux occurrences successives de cet événement. C'est à dire, un événement de période de retour de T années a la probabilité $p = T^{-1}$ de survenir chaque année. En termes de fréquence de non dépassement F , cela s'écrit :

$$T = \frac{1}{1 - F}. \quad (11)$$

Donc, à partir de la fonction de répartition, nous obtenons les périodes de retour d'un événement donné.

III. RÉSULTATS

Les précipitations convectives journalières du 01 janvier 1959 au 31 décembre 2021 sont divisées en bloc de chaque année. La série des maxima annuels de 1959 à 2021 est constituée par la hauteur de pluie journalière maximale de l'année d'enregistrement (Figure 2).

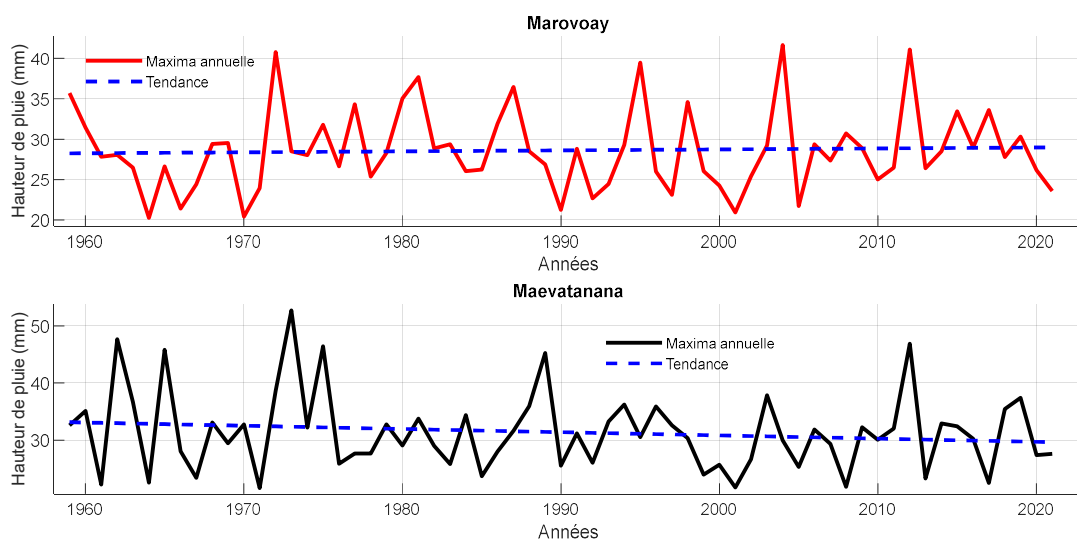


Figure 2 : Courbes des maximas annuelles

3.1 Tests d'hypothèses

Le *Tableau 1* présente les résultats des trois tests statistiques appliqués à la série des maxima annuels de chaque plaine.

Tableau 1 : *P-value des tests d'hypothèses*

	Wald-Wolfowitz	Wilcoxon	KPSS
Marovoay	0,9909	0.6418	0.1000
Maevatanana	0.4355	0.1521	0.1000

Ces résultats montrent que les conditions d'indépendance, d'homogénéité et de stationnarité ont été toutes vérifiées à un seuil de 5%. Ainsi, les hypothèses de base, selon lesquelles les séries des maxima annuelles sont les réalisations d'une suite de variables aléatoires i.i.d, sont satisfaites sur l'ensemble de la zone d'étude.

3.2 Estimations des paramètres

Les résultats des estimations des paramètres des lois GEV avec la méthode des moments pondérés pour chaque plaine respective sont donnés dans les *Tableaux 2* et *3*.

Tableau 2 : *Valeurs des paramètres de la distribution de Gumbel estimées par la méthode des moments.*

	Paramètres de position	Paramètres d'échelle	Paramètres de forme
Marovoay	26.3426	3.9508	0
Maevatanana	28.2721	5.3882	0

Tableau 3 : *Valeurs des paramètres de la distribution de Fréchet par la méthode des moments pondérés.*

	Paramètres de position	Paramètres d'échelle	Paramètres de forme
Marovoay	26.3464	4.1374	0.0279
Maevatanana	28.2559	5.4576	0.0044

D'après ces résultats, les valeurs des paramètres de forme sont positives pour l'ensemble de la zone d'étude. Donc, les lois GEV à trois paramètres correspondent à la distribution de Fréchet. Ce qui correspond à des distributions à queue lourde.

3.3 Vérification de la validité des modèles obtenus

Les résultats des *p-values* des tests d'adéquations de Kolmogorov-Smirnov pour les distributions de Gumbel et de Fréchet estimés sont présentés dans le *Tableau 4*.

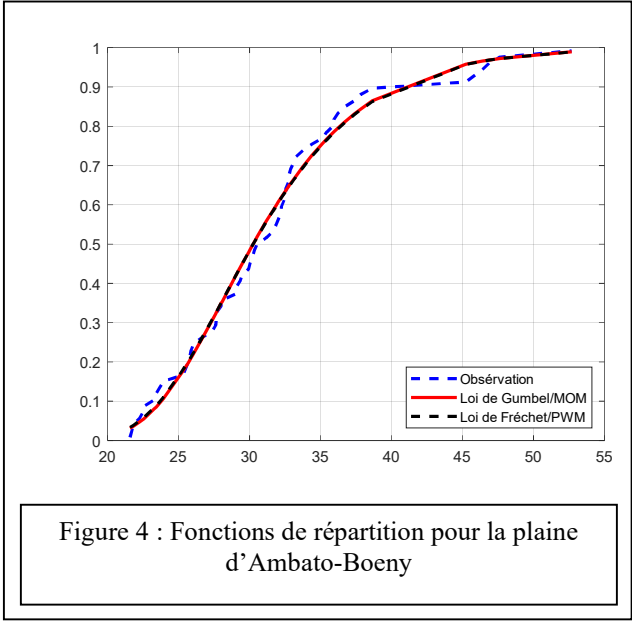
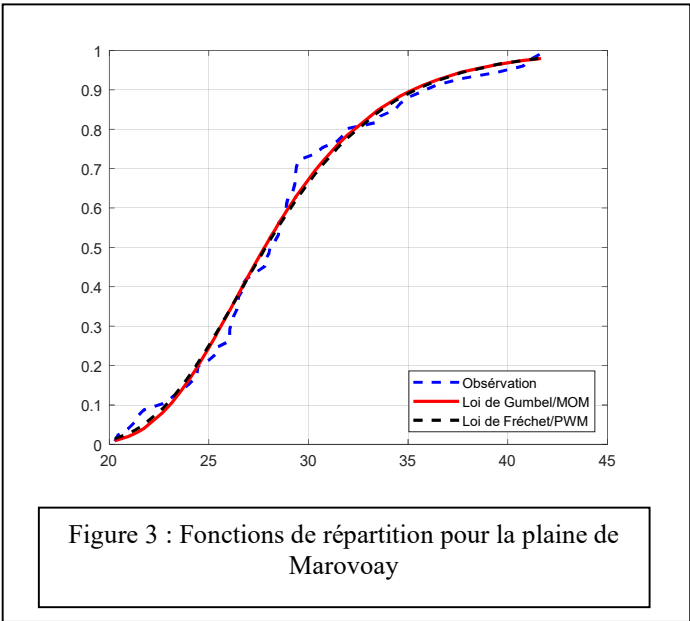
Ces résultats montrent que les distributions GEV aussi bien à deux paramètres qu'à trois paramètres sont adéquates pour ajuster les séries des maximas de l'ensemble de la zone d'étude.

Tableau 4 : *P-value des tests d'adéquations de Kolmogorov-Smirnov.*

	Distribution de Gumbel	Distribution de Fréchet
Marovoay	0.6554	0.5449
Maevatanana	0.9623	0.9687

3.4 Choix du meilleur modèle

Les Figures 3 et 4 permettent de comparer les fonctions de répartition des distributions théoriques et celles des observations.



Nous constatons que les différences sont minimales entre distributions de Fréchet et celle de Gumbel. En outre, une analyse visuelle confirme que les distributions GEV sont bien adéquates pour ajuster les précipitations convectives journalières maximales annuelles de la basse Betsiboka.

Les valeurs des critères AIC et BIC (Tableau 5 et 6) montrent sans équivoque que les meilleurs modèles restent les distributions de Gumbel.

Tableau 5 : Critères AIC des distributions GEV

	Distribution de Gumbel	Distribution de Fréchet
Marovoay	379.6764	381.2048
Maevatanana	414.0973	416.1714

Tableau 6 : Critères BIC des distributions GEV

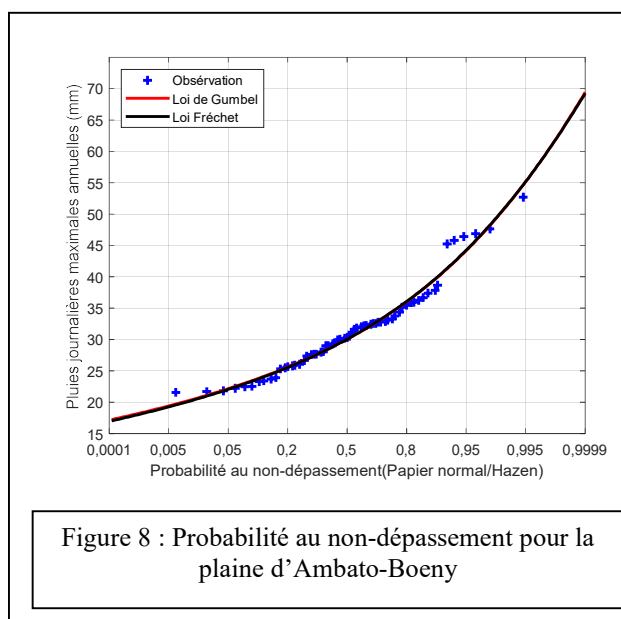
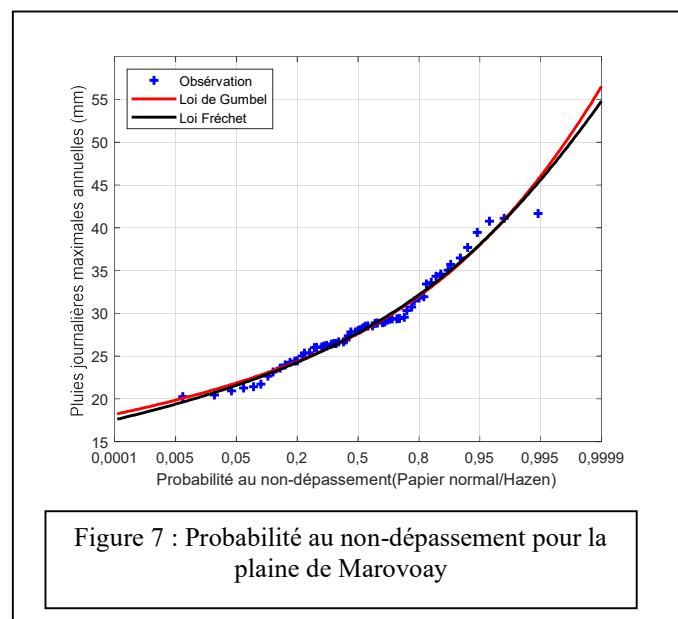
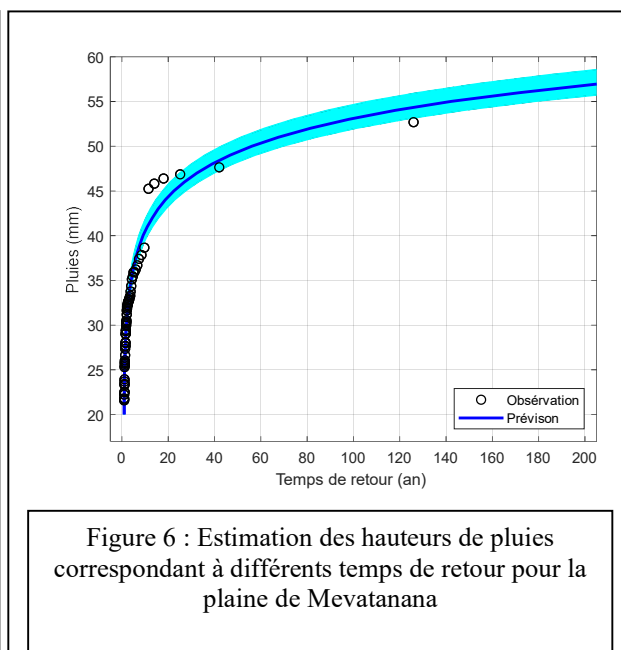
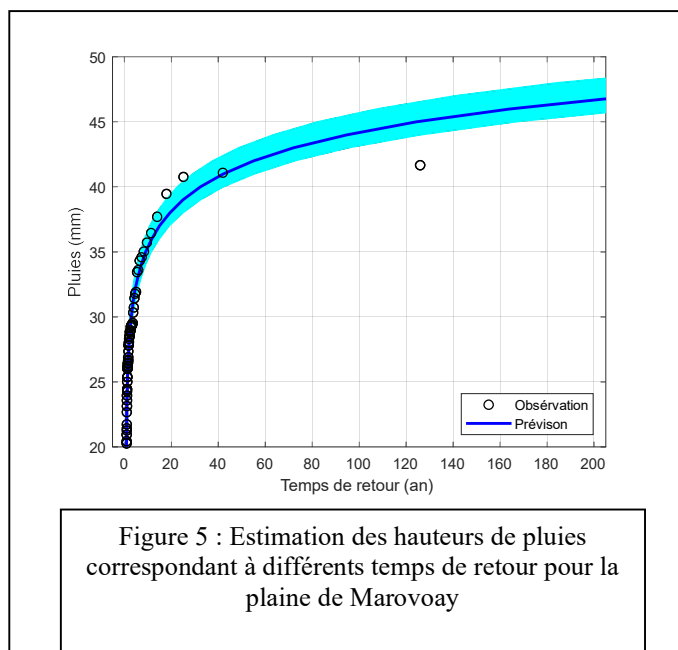
	Distribution de Gumbel	Distribution de Fréchet
Marovoay	375.6764	387.6341
Maevatanana	410.0973	422.6009

3.5 Estimation des quantiles

L'intérêt de l'ajustement d'une série annuelle des maxima avec une distribution statistique est que cela permet d'estimer la période de retour d'une hauteur de pluies donnée. Ces estimations revêtent un intérêt capital en matière de prévention des risques d'inondations.

Ainsi, à partir de la fonction de répartition de la distribution de Gumbel et de l'équation (11), nous obtenons les courbes des Figures 5 et 6 qui donnent des prévisions des hauteurs de pluies correspondant à différents temps de retour.

Les Figures 7 et 8 présentent les mêmes informations en termes de probabilité de non-retour.



3.6 Temps de retour remarquables

Les résultats remarquables de l'étude sur les quantiles et les temps de retour sont résumés dans le *Tableau 7*. Une analyse spatiale permet de conclure que les pluies extrêmes sont largement plus fréquentes dans la plaine de Maevatanana par rapport à celle de Marovoay.

Tableau 7 : Hauteur de pluie, en mm, en fonction des temps de retour.

Plaines \ Temps de retour	5 ans	10 ans	25 ans	50 ans	100 ans	200 ans
Plaine de Marovoay	32,51	35,42	38,93	42,02	44,24	48,71
Plaine de Maevatanana	36,41	40,48	45,59	49,36	53,11	56,82

IV. DISCUSSION

4.1 Estimation des paramètres

La méthode la plus utilisée pour estimer les paramètres d'une distribution statistique reste le maximum de vraisemblance. Cependant, pour les paramètres des lois GEV, Hosking et al. [14] et Hosking et Wallis [17] ont comparé les estimateurs du maximum de vraisemblance et ceux des moments pondérés et ont montré que la méthode des moments pondérés donne de meilleurs résultats en termes d'estimation des quantiles, pour des tailles d'échantillons inférieures à 100. D'autres comparaisons entre les deux méthodes d'estimation avec des échantillons de longueur entre 30 à 35 montrent que la méthode de vraisemblance est inadéquate pour l'estimation des phénomènes avec de longue période de retour [18]. Dans la présente étude, avec un échantillon de longueur 60, la différence entre les deux méthodes d'estimation est surtout observée au niveau des paramètres de forme.

4.2 Choix du meilleur modèle

La loi la plus couramment utilisée en modélisation des maxima annuels est la loi GEV à deux paramètres. Cependant, de récentes études remet en question cette usage systématique des distributions de Gumbel. En effet, la distribution de Gumbel a tendance à sous-estimer les valeurs extrêmes avec de longue période de retour [7]. Les conséquences du choix de la distribution sur les quantiles sont considérables. Pour une fréquence donnée, les quantiles d'une loi de Fréchet peuvent être deux fois plus grands que les quantiles d'une loi de Gumbel [19]. Toutefois, nos résultats montrent que la différence entre les deux distributions sont minimales pour notre zone d'étude.

V. CONCLUSION

Dans cette étude, nous nous sommes intéressés à la prévision des précipitations extrêmes des plaines de la basse Betsiboka, qui sont définie par les pluies journalières maximales annuelles. Les tests statistiques d'homogénéité, d'indépendance et de stationnarité montre que notre série de donnée peut être modélisée avec des distributions statistiques. En particulier, les tests d'adéquations certifient la validité des distributions GEV. Les distributions de Gumbel fourni ainsi une estimation des temps de retour des précipitations convectives extrêmes.

Afin de comparer notre résultat, il serait indispensable d'établir d'autre modèles, entre autres, des modèles obtenus en estimant les bornes supérieurs et inférieurs des périodes de retour en utilisant l'inégalité de Markov, ou en caractérisant les valeurs extrêmes par des observations dépassant un seuil.

REFERENCES

- [1] AGUE A.I. et A. AFOUDA. " Analyse fréquentielle et nouvelle cartographie des maxima annuels de pluies journalières au Bénin". Int. J. Biol. Chem. Sci., 9 (2015) 121-133.
- [2] BENABDESSELAM T. et H. AMARCHI. " Approche régionale pour l'estimation des précipitations journalières extrêmes du Nord-Est algérien". Courrier du Savoir, 17 (2013) 175-184.
- [3] C. Cunnane, " Review of Statistical Models for Flood Frequency Estimation", Paper pres. In : Inst. Symp. On Flood Frequency and Risk Analysis. Baton Rouge. La., Publ. In Singh, V.P. (Ed.), Hydrologic Frequency Modelling, Reidel Publ. CO., Dordrecht : (1987) 49-95.
- [4] R.A. Fisher, L.H.C. Tippett, " Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample", Proceedings of the Cambridge Philosophical Society, 24 (1928) 180-190.

- [5] FALLOT J.M. et J.A. HERTIG. “ Détermination des précipitations extrêmes en Suisse à l’aide d’analyses statistiques et augmentation des valeurs extrêmes durant le 20ème siècle”. Mémoires de la Société vaudoise des Sciences naturelles, 2 (2013) 23-34.
- [6] GOULA B.T.A., B. KONAN, Y. BROU, I. SAVANE, F. VAMORYBA et B. SROHOUROU. “ Estimation des pluies exceptionnelles journalières en zone tropicale : cas de la Côte d’Ivoire par comparaison des lois Lognormale et de Gumbel”. Hydrol. Sci. J., 5 (2007) 49-67.
- [7] S. Coles, L. R. Pericchi, S. Sisson, “ A fully probabilistic approach to extreme rainfall modeling”, J. of Hydrology, 273(1–4) (2003) 35–50.
- [8] D. Koutsoyiannis, “ On the appropriateness of the Gumbel distribution for modelling extreme rainfall”, Preceedings of the ESF LESC Exploratory Workshop, Hydrological Risk, University of Bologna, Bologna (2003).
- [9] S. Coles, “ An Introduction to Statistical Modeling of Extreme Values”, Springer Series in Statistics (2001).
- [10] A. WALD et J. WOLFOWITZ, “ An exact test for randomness in the non-parametric case based on serial correlation”, Annals of Mathematic and Statistical, vol. 14, n° 4 (1943) 378-388.
- [11] F. WILCOXON, “ Individual comparisons by ranking methods”, Biometrics, vol. 1, n° 6 (1945) 80-83.
- [12] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin. “ Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root”, Journal of Econometrics. Vol. 54 (1992) 159-178.
- [13] J. Greenwood, J. Landwehr, N. Matalas et J. Wallis, “ Probability weighted moments : definition and relation to parameters of several distributions expressable in inverse form ”, Water Resources Research, 15 (1979) 1049–1054.
- [14] J. R. M. Hosking, J. R. Wallis et E. F. Wood, “ Estimation of the generalized extreme value distribution by the method of probability weighted moments”, Technometrics 27(3) (1985) 251-261.
- [15] D. ROSBJERG, J. CORREA et P.F. RASMUSSEN, “ Justification des formules de probabilité empirique basées sur la médiane de la statistique d'ordre”, Revue des Sciences de l’Eau, vol. 5, n° 4 (1992) 529-540.
- [16] H. Akaike, “ A new-look at the statistical model identification”, IEEE Transactions on Automatic Control, 19, 6 716-723.
- [17] J. Hosking and J. Wallis, “ Parameter and quantile estimation for the generalized Pareto distribution”, Technometrics, 29 (1987) 339-349.
- [18] BENKHALED A. “ *Distributions statistiques des pluies maximales annuelles dans la région du Cheliff. Comparaison des techniques et des résultats*”. Courrier du Savoir, 8 (2007) 83-91.
- [19] MULLER A. (2006). “ Comportement asymptotique de la distribution des pluies extrêmes en France”. Thèse de doctorat, Univ. Montpellier 2, France, 214 p.