

Analyse Des Fortes Précipitations Et De Leur Evolution Future Dans La Partie Nord De Madagascar

RAMIANDRA Aina Clarc¹, RANDRIANANDRASANARIVO Raphaëlson Jacques², HARY Jean³, MAXWELL Djaffard⁴

¹Ecole doctoral génie vivant et modélisation (EDGVM)

Université de Mahajanga

Mahajanga, Madagascar

clarcaina@gmail.com

²Ecole doctoral génie vivant et modélisation (EDGVM)

Université de Mahajanga

Mahajanga, Madagascar

rrjphys4@gmail.com

³Institut Supérieur des sciences et technologies de Mahajanga (ISSTM)

Université de Mahajanga

Mahajanga, Madagascar

jjeanhary@yahoo.com

⁴Institut Supérieur des sciences et technologies de Mahajanga (ISSTM)

Université de Mahajanga

Mahajanga, Madagascar

djafmax@yahoo.fr

Corresponding Author: RAMIANDRA Aina Clarc; clarcaina@gmail.com



Résumé- L'objectif principal de cette étude consiste à modéliser et à informer les risques de fortes précipitations dans notre zone d'étude. Nous avons employé diverses techniques telles que l'analyse descriptive empirique des échantillons et la théorie statistique des valeurs extrêmes, plus précisément la distribution Pareto généralisée. Les principaux résultats indiquent que les précipitations sont très variables, avec quelques événements de pluie très élevés. La modélisation de la précipitation en excès au-delà du seuil 67.56 mm estime que les précipitations extrêmes qui se produisent plus souvent sont moins intenses. Les précipitations les plus violentes sont moins fréquentes, elles sont extrêmement violentes. Dans notre zone d'étude, où les risques d'événements extrêmes sont modérés, les conséquences sont limitées, mais il est essentiel de surveiller régulièrement afin d'éviter d'éventuelles aggravations. Les aménagements comprennent la création de systèmes de prévision et d'alerte précoce et l'amélioration des infrastructures de drainage afin de réduire au minimum les conséquences des précipitations modérées.

Mots clés- Nord de Madagascar, forte, précipitation, théorie des valeurs extrêmes, Pareto généralisée, fréquentes, intense, risque, modérées

Summary - The main objective of this study is to model and inform about the risks of heavy rainfall in our study area. We employed various techniques such as empirical descriptive analysis of samples and the statistical theory of extreme values, specifically the generalized Pareto distribution. The main results indicate that precipitation is highly variable, with some very high rainfall events. The modeling of excess precipitation beyond the threshold of 67.56 mm estimates that extreme precipitation events that occur more frequently are less intense. The most intense rainfall is less frequent; when it occurs, it is extremely violent. In our study area, where the risks of extreme events are moderate, the consequences are limited, but it is essential to monitor regularly to prevent potential worsening. The

arrangements include the creation of forecasting and early warning systems and the improvement of drainage infrastructure in order to minimize the consequences of moderate rainfall.

Keywords - North of Madagascar, heavy, rainfall, extreme value theory, generalized Pareto, frequent, intense, risk, moderate.

I. INTRODUCTION

Au cours des 30 dernières années, Madagascar a été particulièrement vulnérable à des risques de catastrophes causés par divers aléas naturels tels que les cyclones, les tempêtes tropicales, les inondations, les sécheresses et les invasions acridiennes. Ces événements extrêmes ont engendré des dommages et ont impacté la sécurité alimentaire, l'approvisionnement en eau potable et l'irrigation, les systèmes de santé publique, la gestion de l'environnement et la qualité de vie. Dans ce travail, nous allons focaliser notre recherche sur l'inondation provoquée par une forte pluie dans la région Nord de Madagascar.

Les précipitations intenses sont des pluies qui apportent sur une courte durée (d'une heure à une journée) une importante quantité d'eau. Cette quantité peut égaler celle reçue habituellement en un mois, voire en plusieurs mois. Des cumuls de l'ordre de 50 mm en 24 heures dans la plupart des régions de plaine et de l'ordre de 100 mm en 24 heures dans les régions montagneuses sont considérés comme des seuils critiques. Le dépassement de ces seuils peut provoquer, lorsque la nature du terrain s'y prête, de graves inondations [1].

Les précipitations intenses se produisent lorsque des averses entraînent une quantité importante d'eau de pluie en toute rapidité. Pour parler de fortes précipitations, il faut au moins 5 litres de pluie par mètre carré en 5 minutes ou 17 litres en une heure [2].

Cette étude vise principalement à évaluer, étudier et anticiper les risques de fortes précipitations dans la zone d'étude, à court, moyen et à long terme et ensuite informer sur les risques liés à l'inondations afin de développer des mesures appropriées. Le dernier objectif est de dresser et de renforcer un portrait des initiatives mises en œuvre par les institutions en matière de réponse ou de prévention aux événements climatiques extrêmes.

On se demande alors comment les précipitations intenses ont évolué et quelle sera l'intensité et la fréquence de ces événements à l'avenir ?

Afin de parvenir à notre but, nous avons employé diverses techniques telles que l'analyse descriptive empirique des échantillons et la théorie statistique des valeurs extrêmes (TVE). Le présent travail se subdivise en deux parties. Nous verrons dans la première partie la méthodologie qui contiendra les matériels et les méthodes. Dans la seconde partie, nous nous concentrerons sur les résultats et discussions, et nous terminerons par la conclusion.

II. METHODOLOGIE

A. MATERIELS

1) Présentation de zone d'études :

Notre zone d'étude se trouve dans la partie nord de Madagascar. Elle est délimitée par les longitudes comprises entre 47° E et 51° E et les latitudes comprises entre 12° S et 15° S. Dans cet article, nous ne considérons qu'une petite partie de cette zone qui est encadrée en couleur verte dénommée zone 1. La Zone 1 possède deux type de climat. Sa partie sud est caractérisé par le climat désertique chaud et le nord tropical humide. Ce dernier est parmi la zone la touchée par l'inondation. La zone 1 couvrit trois districts à savoir Antsohihy, Analalava et Ambanja. La Fig.1 représente la zone 1.

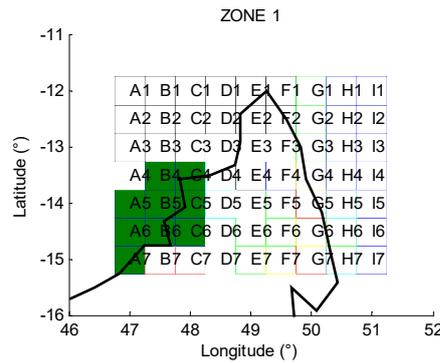


Fig. 1: Représentation de la Zone 1

2) Base de données et les matériels utilisés

Nous avons eu recours à des données pluviométriques. Ce sont des données journalières de réanalyses, pour une résolution spatiale de $1^\circ \times 1^\circ$ sur la période de 1979 jusqu'en 2018 télécharger dans ECMWF.

B. METHODE

1) Description empirique des échantillons

a) La moyenne

Notons $(x_i)_{i=1, \dots, n}$ les données observées que l'on souhaite analyser. Le premier indicateur calculé est généralement la moyenne, que nous noterons \bar{x} [3]:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

b) La médiane

La médiane est la (ou une des ...) valeurs qui partage la distribution (en probabilités) ou la série des valeurs (en statistique) en deux parties de même probabilité (0,5) ou de même effectif (50 % de l'effectif total). Elle se note le plus souvent M. [3]

c) La variance

La variance permet de mesurer la façon dont les données se dispersent autour de la moyenne. Pour le calcul de la variance, la somme des carrés des écarts est divisée par le nombre de valeurs. [3]

$$\text{var}_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

La variance décrit donc l'écart moyen au carré par rapport à la moyenne

d) L'écart-type

L'écart-type indique la dispersion d'une variable autour de sa valeur moyenne [3]. Ainsi, l'écart-type est la déviation moyenne (moyenne quadratique) de toutes les valeurs mesurées par rapport à la moyenne. On définit également l'écart-type, qui a l'avantage d'avoir la même dimension que les données :

$$s_x = \sqrt{\text{var}_x} \quad (3)$$

La différence entre la variance et l'écart-type comme indicateur de dispersion est donc que l'écart-type mesure la distance moyenne par rapport à la moyenne et que la variance mesure la distance moyenne au carré par rapport à la moyenne.

e) *Quartiles*

Les quartiles divisent vos données en quatre parties, aussi égales que possible [3]. Pour le calcul des quartiles, les données doivent être triées de la plus petite à la plus grande valeur.

- Quartile (Q1) : la valeur intermédiaire entre la plus petite valeur (minimum) et la médiane. Pour avoir Q1, on calcule $N/4$, avec N nombre d'échantillons, puis on détermine le premier entier p supérieur ou égal à $N/4$. Cet entier p est le rang de Q1.
- Quartile (Q2) : la médiane des données, c'est-à-dire que 50% des valeurs sont plus petites et 50% des valeurs sont plus grandes.
- Quartile (Q3) : la valeur intermédiaire entre la médiane et la plus grande valeur (maximum). Et Pour l'avoir, on fait de même avec $3N/4$.

f) *Coefficient d'asymétrie (Skewness)*

C'est un nombre réel sans dimension, indicateur de forme, qui mesure l'asymétrie d'une distribution probabiliste ou statistique [3]. La formule du coefficient d'asymétrie est :

$$skewness = \frac{1}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (4)$$

où : n : nombre d'observations, x_i : chaque observation, \bar{x} : la moyenne des observations, s : l'écart-type des observations.

Ce coefficient a été introduite par Fisher ; pour une distribution régulière, avec un maximum (« mode ») unique : il est positif lorsque ce maximum est à gauche de l'espérance ; donc une « queue de distribution » importante à droite, nul lorsque la distribution est symétrique, et négatif lorsque le maximum de la distribution est à droite de l'espérance ; donc une « queue de distribution » importante à gauche.

Le coefficient d'asymétrie aide à identifier si les données sont biaisées vers des valeurs plus élevées ou plus faibles. Cela peut être particulièrement utile dans l'analyse des données climatiques pour détecter des tendances inhabituelles.

g) *Coefficient d'aplatissement (Kurtosis)*

C'est un nombre réel sans dimension, indicateur de forme, qui mesure l'aplatissement d'une distribution probabiliste ou statistique. [3]

La formule du coefficient d'aplatissement est :

$$kurtoisis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (5)$$

Le coefficient d'aplatissement mesure la "hauteur" et la "largeur" des queues de la distribution. Comme sa valeur est égale à 3 pour une loi normale, Fisher a retranché 3 pour traduire l'idée que l'aplatissement standard (coefficient nul) est celui de la loi normale. Le coefficient est compris entre -2 et $+\infty$. Ce coefficient est négatif pour une distribution étalée, positif pour une distribution pointue.

Un coefficient d'aplatissement élevé (positive kurtosis) indique que les données ont plus de valeurs extrêmes que ce qui est attendu dans une distribution normale. Cela peut aider à identifier les événements climatiques extrêmes qui peuvent être cruciaux pour des études climatiques. La kurtosis donne une idée de la concentration des données autour de la moyenne. Une distribution

avec une kurtosis faible (négative kurtosis) est plus plate et a moins de valeurs extrêmes. Cela peut aider à comprendre la variabilité des données climatiques.

2) Théorie statistique des valeurs extrêmes

La Théorie des Valeurs Extrêmes (EVT, en anglais) traite du comportement stochastique la queue supérieure ou inférieure de la distribution initiale des variable aléatoire v.a. i.i.d. Elle permet ainsi d'établir les propriétés de distribution des extrêmes (maxima ou minima), des statistiques d'ordre extrême ou intermédiaire, et des excès au-dessus (au-dessous) d'un seuil élevé.

3) Loi des excès au-delà d'un seuil

La théorie des excès au-delà d'un seuil consiste à utiliser les observations qui dépassent un certain seuil, plus particulièrement les différences entre ces observations et le seuil, appelées excès. Il est clair que cette méthode nécessite la détermination d'un seuil ni trop faible pour ne pas prendre en considération des valeurs non extrêmes, ni trop élevé pour avoir suffisamment d'observations. Notons le seuil par u .

On définit un seuil $u \in \mathbb{R}$, $N_u = \text{card}\{i: i = 1, \dots, n, X_i > u\}$ et $Y_i = X_i - u > 0$ pour $1 \leq j \leq N_u$ ou N_u est le nombre des dépassements du seuil u par les $(X_i)_{1 \leq i \leq n}$ et Y_1, \dots, Y_{N_u} les excès correspondants

On cherche à partir de la loi F de X à définir une loi conditionnelle F_u par rapport au seuil u pour les variables aléatoires dépassant ce seuil. On définit alors la loi conditionnelle des excès F_u par :

$$F_u(y) = \Pr(X - u \leq y, X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, y \geq 0 \quad (6)$$

Pickands a montré que pour un seuil suffisamment élevé, c'est-à-dire lorsque $u \rightarrow \infty$, la distribution de dépassement, conditionnelle à $X > u$, est approximativement :

$$H(x) = \begin{cases} 1 - \left[1 - \frac{k}{\sigma(x-u)} \right], \xi \neq 0 \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right), \xi = 0 \end{cases} \quad (7)$$

La famille de distribution H est appelée Distribution de Pareto Généralisée (GPD, Generalized Pareto Distribution), de paramètre d'échelle σ et de forme ξ . [4]

h) Estimation des paramètres par la méthode de maximum de vraisemblance de Loi GPD

La loi GPD est indexée par deux paramètres ξ et σ que nous pouvons estimer par plusieurs méthodes. Mais dans ce travail, nous choisissons la méthode de maximum de vraisemblance.

Si Y_j désigne la $j^{\text{ème}}$ variable d'excès au-delà du seuil u , $j = 1, \dots, N_u$, dans le cas où $u = X_{n-k,n}$, la statistique d'ordre des excès est de la forme $Y_{j,k} = X_{n-k+j} - X_{n-k,n}$

avec $j = 1, \dots, N_u$. En posant $\tau = \frac{\xi}{\sigma}$, pour $1 + \xi \frac{Y_i}{\sigma} > 0$ et $i=1 \dots N_u$, nous pouvons écrire la log-vraisemblance selon [5]:

$$\log L(\tau, \xi) = -N_u \log \xi + N_u \log \tau - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \log(1 + \tau Y_i) \quad (8)$$

Les estimateurs satisfont alors

$$\frac{1}{\sigma} - \left(\frac{1}{\xi} + 1 \right) \frac{1}{N_u} \sum_{i=1}^{N_u} \frac{Y}{1 + \tau Y_i} = 0 \quad (9)$$

Et

$$\xi = \frac{1}{N_u} \sum_{i=1}^{N_u} \log(1 + \sigma Y_i) \quad (10)$$

i) Quantiles et périodes de retour

En général, l'unité de temps est l'année, et des relations simples lient les probabilités de non-dépassement et les périodes de retour. Pour simplifier, on pourra retenir qu'une valeur de période de retour T ans à une probabilité $1/T$ d'être dépassée chaque année, soit une probabilité de non-dépassement de $1-1/T$. De manière plus mathématique, la relation liant probabilité au non-dépassement et période de retour est la suivante [6]:

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

Notons qu'une correction doit être apportée à la relation ci-dessus dans le cas de méthodes d'échantillonnage ne conduisant pas à une unique valeur par an. Dans le cas de excès au-delà d'un seuil conduisant à θ valeurs par an en moyenne, cette relation devient :

$$1 - p = \frac{1}{\theta T} \quad (12)$$

4) Calcul des débits caractéristiques

Ces quantiles peuvent être calculés en inversant la fonction de répartition de la variable étudiée [6]. En effet, un quantile q_p vérifie, par définition,

$$H(q_p) = p \quad , \text{ soit } q_p = H^{-1}(p) \quad (13)$$

5) Test d'adéquation de distribution

j) QQ-plot (quantile-quantile plot)

Un QQ-plot (quantile-quantile plot) est un graphique utilisé pour comparer visuellement les quantiles d'une distribution empirique à ceux d'une distribution théorique, ou pour comparer les quantiles de deux distributions empiriques entre elles. Sur ce graphique :

- L'axe des abscisses représente les quantiles de la distribution théorique ou d'une première distribution empirique.
- L'axe des ordonnées représente les quantiles de la distribution empirique observée ou d'une seconde distribution empirique.

Si les points du QQ-plot se situent près d'une ligne droite diagonale (ligne $y = x$), cela indique que les deux distributions comparées sont similaires en termes de forme. En revanche, des écarts systématiques par rapport à cette ligne peuvent révéler des différences dans la forme des distributions, telles que des queues plus lourdes ou des divergences dans la symétrie ou l'étalement.

k) Test d'adéquation de Kolmogorov-Smirnov

Le test d'ajustement nous permettra de vérifier si la loi de probabilité d'un échantillon des extrêmes observés appartient à la famille de lois théorique choisie pour l'analyse des extrêmes. Soit X_1, X_2, \dots, X_n les valeurs d'un échantillon des extrêmes des dépassements d'un seuil rangé par ordre croissant, et G la fonction de répartition théorique correspondante avec les paramètres

estimés à partir de l'échantillon considéré. Le principe du test d'adéquation de Kolmogorov-Smirnov consiste simplement à trouver la différence entre les distributions de répartition empirique et théorique :

$$D = \max_x |F(x) - G(x)| \quad (15)$$

On cherche alors à tester l'hypothèse H_0 selon laquelle l'échantillon X relève de la loi G . La p-value peut-être calculée à partir de la statistique D en utilisant des tables préétablies ou des méthodes d'approximation basées sur la distribution limite de la statistique KS. Si $h = 0$: L'hypothèse nulle n'est pas rejetée, ce qui suggère que les données suivent bien la distribution théorique (ou que les deux échantillons suivent la même distribution). Sinon, $h = 1$: L'hypothèse nulle est rejetée, indiquant que les données ne suivent pas la distribution théorique (ou que les deux échantillons ne suivent pas la même distribution). La p-value est utilisée pour évaluer la signification du test. Si la p-value est inférieure à un certain seuil (généralement 0.05), l'hypothèse nulle est rejetée.

III. RESULTATS ET DISCUSSIONS

C. Statistique des précipitations

Dans cette partie, nous allons nous intéresser à la description d'une série de données. L'objectif est notamment de résumer l'information contenue dans les données grâce à quelques grandeurs caractéristiques. La Fig.2 montre l'évolution des précipitations journalières de 1979 à 2017 dans la zone 1.

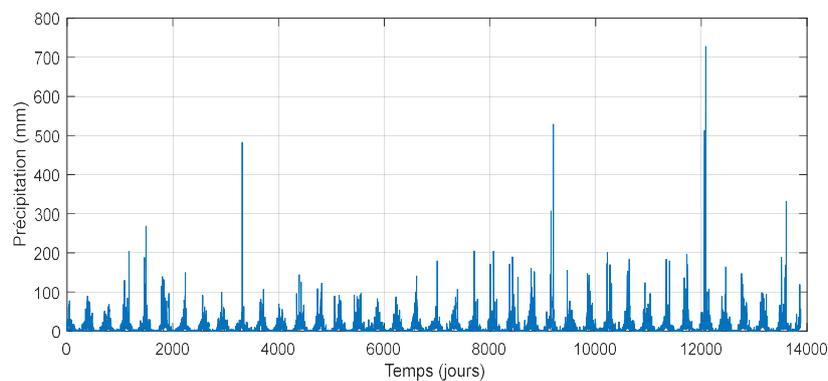


Fig. 2 : Evolution de la précipitation moyenne journalière dans la zone 1

Le tableau I récapitule le résumé statistique de hauteur des précipitations dans la Zone 1 depuis 1979 à 2017. La moyenne des précipitations est de 10,40 mm. Cela donne une indication générale de la quantité de pluie reçue sur l'ensemble de la période étudiée. La médiane est la valeur centrale des précipitations, ce qui signifie que 50% des valeurs de précipitation sont inférieures à 3,60 mm. Cela suggère que la distribution des précipitations est asymétrique, car la médiane est beaucoup plus basse que la moyenne. Le premier quartile (Q1) indique que 25% des valeurs de précipitation sont inférieures à 0,66 mm. Cela montre qu'une partie significative des précipitations est faible. Le troisième quartile (Q3) indique que 75% des valeurs de précipitation sont inférieures à 12,80 mm. Cela montre que les valeurs de précipitation s'élèvent à des niveaux plus élevés dans 25% des cas. L'écart-type des observations est 20,3824 mm. L'écart-type élevé suggère une grande dispersion des valeurs de précipitation par rapport à la moyenne. Cela signifie qu'il existe des variations significatives dans les quantités de pluie enregistrées. La variance est 415,4438 mm². La variance est le carré de l'écart-type et elle mesure également la dispersion des valeurs. Une variance élevée confirme que les précipitations varient beaucoup au cours de la période étudiée.

Le coefficient d'asymétrie est 9,6697. Ce coefficient mesure l'asymétrie de la distribution des précipitations. Une valeur positive (surtout aussi élevée) indique que la distribution est fortement asymétrique à droite, ce qui signifie qu'il y a une queue longue vers les valeurs élevées (beaucoup de faibles précipitations et quelques événements de précipitation très élevés). Le Coefficient d'aplatissement est 215,1634. Une valeur beaucoup plus grande que 3 (qui serait la valeur pour une distribution normale) suggère que la distribution est très pointue et que des valeurs extrêmes sont fréquentes.

Les statistiques indiquent que les précipitations dans la période 1979 à 2017 sont très variables, avec de nombreuses faibles valeurs et quelques événements de pluie très élevés. La distribution est fortement asymétrique, suggérant une tendance à avoir des périodes de faible pluie entrecoupées de rares événements de pluie intense.

Tableau I: Résumé statistique des précipitations

Indicateur	Valeur
Moyenne (mm)	10.4046
Médiane (mm)	3.6016
Quartile (Q1) (mm)	0.66226
Quantile (Q3) (mm)	12.8002
Ecart-type (mm)	20.3824
Variance (m ²)	415.4438
Coefficient d'asymétrie :	9.6697
Coefficient d'aplatissement :	215.1634

D. Ajustement de la hauteur des précipitation

La Fig.3 présente une comparaison entre les divers ajustements de la distribution de probabilité concernant la hauteur des précipitations dans la zone 1 depuis 1979 jusqu'au 2017. Il est important de noter que la loi de Pareto généralisée est la loi la plus proche de cette distribution.

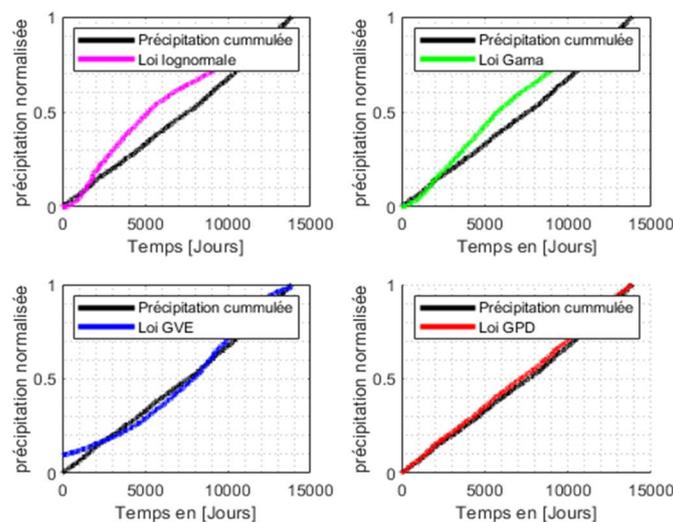


Fig. 3: Comparaison de la fonction des valeurs cumulées de pluie avec les différentes lois de probabilité

6) Estimation des paramètres de distribution adéquate

Le tableau II montre les résultats de l'ajustement d'une distribution de Pareto généralisée (GPD) à des excédents au-dessus de différents seuils (percentiles). Chaque ligne correspond à un seuil différent, et pour chaque seuil, les paramètres de la distribution GPD (le paramètre de forme ξ et le paramètre d'échelle σ) sont estimés, avec leurs intervalles de confiance.

Le seuil est défini en termes de percentiles (92.5ème, 95ème, 97.5ème, 98.23ème). Plus le percentile est élevé, plus le seuil (u) est grand. En augmentant le seuil, vous considérez des événements de plus en plus extrêmes. Cela signifie que moins de données sont utilisées pour ajuster la GPD, mais ces données représentent des événements plus rares et plus extrêmes. Les valeurs augmentent à mesure que le percentile augmente, ce qui est attendu puisque des percentiles plus élevés correspondent à des valeurs de données plus élevées. Chaque seuil correspond à un niveau différent de gravité des événements extrêmes. Un seuil à 67.5669 pour le 98.23ème percentile indique que seuls les événements au-dessus de ce niveau sont considérés comme extrêmes. Les valeurs estimées de ξ augmentent avec le seuil, passant de 0.28506 pour le 92.5ème percentile à 0.43688 pour le 98.23ème percentile. Un ξ plus élevé signifie une queue plus lourde, ce qui suggère que les événements au-dessus de seuils plus élevés ont une probabilité plus élevée de produire des valeurs extrêmes. Cependant, une augmentation de ξ avec le seuil est cohérente avec l'idée que les queues deviennent plus épaisses pour les événements les plus extrêmes. Les intervalles de confiance pour ξ sont assez larges, ce qui indique une incertitude significative dans l'estimation de ce paramètre. Les intervalles de confiance sont très larges et contiennent des valeurs proches de zéro, ce qui pourrait indiquer une certaine incertitude quant à la robustesse de l'ajustement de la GPD, surtout pour des percentiles élevés. Cette large incertitude peut également être due à la diminution du nombre de données disponibles à mesure que le seuil augmente. Les valeurs de σ augmentent également avec le seuil, indiquant une plus grande dispersion des données extrêmes pour les seuils plus élevés. Un σ plus élevé signifie que, pour les événements au-dessus de chaque seuil, il y a plus de variabilité dans les excédents. Cela est cohérent avec des événements extrêmes étant de plus en plus variables. Les intervalles de confiance pour σ sont également larges, ce qui montre une incertitude notable dans l'estimation de ce paramètre. Comme pour ξ , ces intervalles de confiance indiquent une incertitude, particulièrement pour les seuils élevés. Cela peut être dû au fait que moins de données sont disponibles pour estimer les paramètres à ces seuils.

Tableau II: Choix du seuil et paramètre de distribution Pareto généralisée

Choix du seuil	Paramètre de position u	Paramètre de forme ξ		Paramètre d'échelle σ	
		Valeur estimée	Intervalle de Confiance	Valeur estimée	Intervalle de Confiance
92,5 ^{ème} percentile	31.9663 mm	0.28506	[0.20965, 18.3592]	20.1943	[0.36046, 22.2128]
95 ^{ème} percentile	40.8857 mm	0.28741	[0.19826, 20.0274]	22.4487	[0.37656, 25.1627]
97,5 ^{ème} percentile	58.0941 mm	0.38132	[0.23624, 20.2686]	24.1368	[0.52639, 28.7432]
98,23 ^{ème} percentile	67.5669 mm	0.43688	[0.24583, 20.2643]	25.2931	[0.62794, 31.5698]

En augmentant le seuil, les valeurs de ξ et σ augmentent, ce qui est cohérent avec l'idée que des seuils plus élevés correspondent à des événements plus extrêmes, plus rares, et avec une plus grande variabilité. Les intervalles de confiance pour ξ et σ sont larges, indiquant une incertitude dans les estimations. Cela pourrait être une indication que le modèle GPD est plus incertain ou moins stable pour les seuils plus élevés. Une large incertitude peut également indiquer que les données ne sont pas suffisamment nombreuses pour estimer avec précision ces paramètres pour les seuils les plus élevés. Étant donné l'incertitude croissante dans les estimations des paramètres pour les seuils plus élevés, il est important de faire preuve de prudence lorsque vous interprétez les résultats pour ces seuils. Il pourrait être utile d'explorer d'autres méthodes de sélection de seuil ou d'envisager l'utilisation de plusieurs seuils pour obtenir une vue plus robuste.

7) Adéquation par courbe de QQ-plot

La Fig.4 illustre le traçage des quantiles théoriques de la GPD contre les quantiles empiriques des excédents au-dessus du seuil choisi. Si les points suivent une ligne droite, cela indique un bon ajustement. Les points forment une courbe concave vers le haut (c'est-à-dire, au-dessus de la ligne droite), cela indique que la queue de la distribution empirique est plus légère que celle de la GPD ajustée. De telles déviations indiquent que le modèle GPD pourrait ne pas être bien adapté pour capturer l'extrémité de la distribution des données. Alors nous n'avons pas retenu les trois premiers seuils sur le tableau 2. Les points sont répartis autour de la ligne droite, sans tendance apparente à se regrouper ou à se disperser excessivement. Une distribution homogène des points autour de la ligne droite suggère un bon ajustement global de la distribution GPD. Et On peut rajouter aussi que les extrémités du QQ-plot (les valeurs les plus élevées des excédents) sont proches de la ligne droite. Le modèle capture correctement les événements les plus rares.

Le QQ-plot montre une bonne correspondance entre les quantiles empiriques et théoriques (alignement avec la ligne droite, absence de déviations systématiques, homogénéité des points), nous pouvons considérer que la GPD est bien ajustée pour les excédents au-delà du seuil 67.5669.

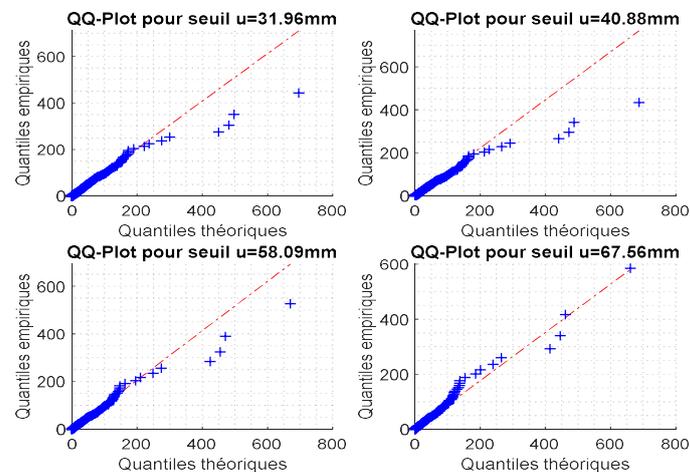


Fig. 4: comparaison des quantiles théoriques de la GPD contre les quantiles empiriques des excédents au-dessus du seuil choisi

8) Adéquation par Test d'adéquation de Kolmogorov-Smirnov

Le tableau III fournit les résultats du test d'adéquation de Kolmogorov-Smirnov (KS) pour différents seuils (paramètres de position u). Pour un seuil de 31.9663, le test d'adéquation de Kolmogorov-Smirnov a rejeté l'hypothèse nulle ($h = 1$), avec une p-value de 0.00027901, indiquant que les données ne s'ajustent pas bien à la distribution de Pareto généralisée pour ce seuil. La très faible p-value indique une forte divergence entre les données et la distribution théorique. Pour un seuil de 40.8857, le test d'adéquation de Kolmogorov-Smirnov n'a pas rejeté l'hypothèse nulle ($h = 0$), avec une p-value de 0.12954, suggérant que les données peuvent être considérées comme bien ajustées à la distribution de Pareto généralisée pour ce seuil. Pour un seuil de 58.0941, le test d'adéquation de Kolmogorov-Smirnov n'a pas rejeté l'hypothèse nulle ($h = 0$), avec une p-value de 0.40348, ce qui indique que les données s'ajustent bien à la distribution de Pareto généralisée pour ce seuil. Pour un seuil de 67.5669, le test d'adéquation de Kolmogorov-Smirnov n'a pas rejeté l'hypothèse nulle ($h = 0$), avec une p-value de 0.21237, indiquant que les données peuvent être considérées comme bien ajustées à la distribution de Pareto généralisée pour ce seuil.

Parmi les seuils testés, les valeurs de $u = 40.8857$, $u = 58.0941$, et $u = 67.5669$ semblent bien ajuster les données à la GPD, avec des p-values supérieures à 0.05. Le seuil u égale à 58.0941 offre le meilleur ajustement avec la p-value la plus élevée (0.40348). Ces résultats suggèrent que l'utilisation d'un seuil plus élevé (comme 58.0941) permet un meilleur ajustement des données à la distribution GPD, selon le test KS.

Tableau III : Résultat de test Kolmogorov-Smirnov

Seuil u	Hypothèse (H ₀)	p_value
31.9663 mm	1	0.00027901
40.8857 mm	0	0.12954
58.0941 mm	0	0.40348
67.5669 mm	0	0.21237

En regroupant les résultats venant de ces deux tests d'adéquation, notamment Test KS et QQ-plot, nous décidons de retenir que le seuil 67.5669 mm est la plus adéquate des données à la distribution GPD.

Le Fig.5 représente la courbe des fonctions de densité de probabilité (à gauche) et fonctions de distribution cumulative (à droite) des lois GPD ($\xi = 0.43$, $\sigma = 25.29$) de la précipitation au excès au-delà du seuil $u = 67.56$ mm. La courbe fonctions de distribution cumulative est croissante. Elle commence à 0 (probabilité que la variable soit inférieure au seuil u) et tend vers 1 à mesure que x augmente. La forme exacte de la courbe dépend du paramètre de forme ξ , comme $\xi = 0.43$ de signe positif (queue lourde) alors la courbe croît lentement au début puis accélère, indiquant qu'il y a une probabilité non négligeable d'observer des valeurs extrêmement grandes. La partie initiale de la courbe (près de u), la courbe est relativement plate, cela signifie que peu de probabilités sont accumulées pour ces valeurs. Dans le milieu de la courbe, la fonction de distribution cumulative commence à croître plus rapidement. Cela indique la transition où une part plus significative de la probabilité est accumulée. Et sur la partie finale de la courbe, elle continue de croître mais ne tend jamais totalement vers 1 (sauf à l'infini).

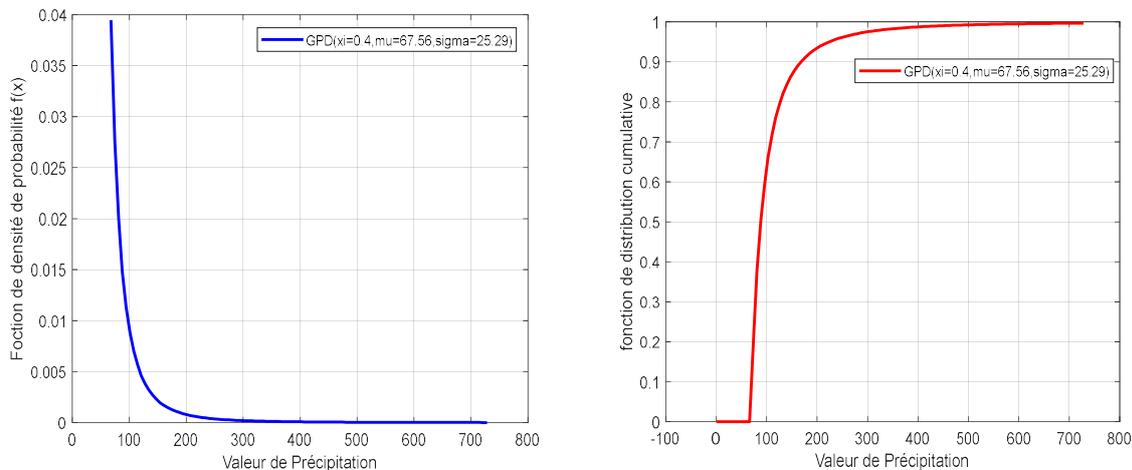


Fig. 5: Allure des fonctions de densité de probabilité (à gauche) et fonctions de distribution cumulative (à droite) des lois GPD pour paramètre de forme $\xi = 0.43$, du paramètre de position $u = 67.56$, et le paramètre d'échelle $\sigma = 25.29$

Le tableau IV présente les périodes de retour, les probabilités de non-dépassement associées, et les précipitations caractéristiques pour la zone 1 située au nord de Madagascar. Comme par définitions, une période de retour est le temps moyen entre deux événements extrêmes (ici, les précipitations) d'une certaine ampleur. Par exemple sur tableau 4, une période de retour de 10 ans signifie qu'un événement de cette ampleur ou plus se produit en moyenne une fois tous les 10 ans. Ainsi que la probabilité de non-dépassement est la probabilité qu'un événement ne dépasse pas une certaine valeur au cours d'une année donnée. Par exemple sur le Tableau 4, une probabilité de non-dépassement de 0.98374 pour une période de retour de 10 ans signifie qu'il y a 98.374 % de chances que les précipitations ne dépassent pas 359.7524 mm dans une année donnée. Et la précipitation caractéristique (en mm) est la quantité de précipitations associée à la période de retour spécifiée. Pour une période de retour de 10 ans, Il y a une chance de 2.899 % ($1 - 0.97101$) que les précipitations dépassent 407.3204 mm dans une année donnée. Cet événement est censé se produire

en moyenne une fois tous les 10 ans. Pour une période de retour de 20 ans, Il y a une chance de 98.551 % que les précipitations ne dépassent pas 466.2526 mm dans une année donnée. Cet événement est censé se produire en moyenne une fois tous les 20 ans. Pour une période de retour de 50 ans, Il y a une chance de 99.42 % que les précipitations ne dépassent pas 543.9394 mm dans une année donnée. Cet événement est censé se produire en moyenne une fois tous les 50 ans. Pour une période de retour de 80 ans, Il y a une chance de 99.638 % que les précipitations ne dépassent pas 583.6924 mm dans une année donnée. Cet événement est censé se produire en moyenne une fois tous les 80 ans. Pour une période de retour de 100 ans, Il y a une chance de 99.71 % que les précipitations ne dépassent pas 602.5432 mm dans une année donnée. Cet événement est censé se produire en moyenne une fois tous les 100 ans.

Tableau IV : Analyse statistique des précipitations caractéristiques pour différentes périodes de retour, ainsi que les probabilités de non-dépassement

Période de retour T (ans)	Probabilité de non dépassement	Précipitation caractéristique (mm)
10	0.98374	359.7524
20	0.99187	483.5682
50	0.99675	716.8637
80	0.99797	878.0607
100	0.99837	966.9822

Pour des périodes de retour plus courtes (comme 10 ou 20 ans), les précipitations caractéristiques sont modérées (359.7524 mm et 483.5682 mm respectivement). Pour des périodes de retour plus longues (comme 50, 80, ou 100 ans), les précipitations caractéristiques sont beaucoup plus élevées (jusqu'à 966.9822 mm). Cela montre que des événements de précipitation très intenses sont rares, mais peuvent causer des inondations ou d'autres impacts importants lorsqu'ils se produisent.

Ces résultats montrent que plus la période de retour est longue, plus la précipitation caractéristique augmente. Cela indique que des événements de précipitations extrêmes deviennent de moins en moins fréquents, mais lorsqu'ils se produisent, ils peuvent être très intenses. Ces informations sont essentielles pour la planification et la gestion des risques liés aux inondations et aux événements climatiques extrêmes.

IV. CONCLUSION

Trois étapes importantes ont structuré ce travail. En premier lieu, nous avons fait les analyses statistiques préliminaires pour déterminer le comportement de nos données. Les résultats de ces analyses relèvent que les précipitations dans notre zone d'étude ont des variabilités, avec de nombreuses faibles valeurs et quelques événements de pluie très élevés. Dans le deuxième temps, le test d'ajustement des lois nous montre que la distribution de Pareto généralisé est la distribution la plus adéquate à la précipitation de 1979 à 2017 dans la zone 1. Et enfin, après avoir estimé les trois paramètres de lois GPD, à savoir le paramètre de forme ξ et le paramètre d'échelle σ ($\xi = 0.43$, $\sigma = 25.29$) de la précipitation en excès au-delà du seuil 67.56 mm, nous avons les résultats suivants. Les précipitations extrêmes plus fréquentes sont moins intenses. Les probabilités de non-dépassement sont élevées, mais les précipitations caractéristiques restent modérées. Les événements de précipitations plus rares sont beaucoup plus violents. Les probabilités de non-dépassement sont encore plus élevées, signifiant que ces événements sont moins susceptibles de se produire chaque année, mais lorsqu'ils se produisent, ils sont associés à des niveaux de précipitation très élevés.

Notre zone d'étude présentant des risques modérés d'événements extrêmes, voient des impacts limités mais nécessitent une surveillance continue pour prévenir d'éventuelles aggravations. Les mesures d'adaptation incluent la mise en place de systèmes de prévision et d'alerte précoce ainsi que l'amélioration des infrastructures de drainage pour minimiser les impacts des précipitations modérées. Cela permet aux décideurs et aux gestionnaires de risques de comprendre la fréquence et l'intensité probable des événements de précipitation extrême, ce qui est crucial pour la planification et la préparation aux catastrophes naturelles.

REMERCIEMENTS

Le moment est venu pour moi d'exprimer ma profonde gratitude à tous ceux qui m'ont aidé et encouragé dans les moments difficiles, dont l'objectif sera d'obtenir le diplôme de Docteur de l'Université de Mahajanga. Je voudrais d'abord remercier chaleureusement mon directeur de thèse, le professeur MAXWELL Djaffard, qui a accepté de superviser mes premiers pas en recherche. Merci à tous mes amis ou camarades de l'école doctorale à l'université qui ont aidé avec cette tâche. Enfin, je voudrais remercier tous les membres de ma famille, en particulier mon père RAMIANDRA Jean Augustin, ma mère MANAVAVY Olga et ma femme RAMIARIJAONA Albertinah Natacha qui m'ont soutenu dans tous les domaines. Je tiens à exprimer ma gratitude à mon fils RAMIANDRA Hatea Nathan.

REFERENCES

- [1] <https://alarmemeteo.ch/blog>
- [2] RAHOBIVONJY Alexandre Christian Sergio, analyse topographique de l'inondations de la plaine, mahavazina à Vontovorona, université d'Antananarivo école supérieure polytechnique, 2023, pp :3-4
- [3] François Dress, les probabilités et les statistiques d'A à Z, DUNOD, pp : 107-130
- [4] Goubanova Katerina, Une étude des événements climatiques extrêmes sur l'Europe et le bassin Méditerranéen et de leur évolution future, UNIVERSITÉ de PARIS 6, 21 novembre 2007, pp :15-22.
- [5] Bechir RAGGAD, fondements de la théorie des valeurs extrêmes, Ses principales applications et son apport à la gestion des risques du marché pétrolier, pp :38-40
- [6] Benjamin Renard. Probabilités et Statistiques appliquées à l'Hydrologie. France. 2014, pp.51-52. hal- 02597049