

Modélisation et études prévisionnelles de pluies : cas du site hydroélectrique d'Andekaleka à Madagascar

Rodolphe RAMIHARIJAFY, Jean Marc RABEHARISOA, Adolphe RATIARISON* et Jérôme VELO

*Laboratoire de Dynamique de l'Atmosphère du Climat et des Océans (DyACO),
Université d'Antananarivo, Madagascar*

* Correspondance, courriel : aratiarison@yahoo.fr

Résumé

Cette modélisation pluviométrique et les études prévisionnelles de pluie se basent sur la fonction cumulée des hauteurs de pluie. L'objectif est de prévoir la hauteur cumulée de pluie pour les années à venir, entre autre pour l'année 2020, à l'aide de la méthode des marches aléatoires. Nous avons comparé deux méthodes différentes, à savoir : la méthode de régression robuste et la méthode des marches aléatoires. Pour la modélisation par des marches aléatoires, la hauteur de pluie se distribue suivant une loi log-normale, tandis que les temps d'attente suivent une loi α -stable (avec $\alpha = 0.4$). Les résultats ont montré que, par la méthode de la régression robuste, la prévision du cumul de pluie pour l'année 2020 est de l'ordre de 1441 mm, tandis que la méthode des marches aléatoires donne une hauteur de l'ordre de 1122,8 mm. Ces deux résultats sont similaires et la hauteur est légèrement inférieure à celle de l'année 2013, qui est égale à 1519 mm. De plus, des évènements de sécheresse seront prévus pour les années 2025, 2032 et après l'année 2038.

Mots-clés : *modélisation pluviométrique, études prévisionnelles, hauteur cumulée de pluie, marches aléatoires, régression robuste, temps d'attente, loi log-normale, loi α -stable, centrale hydroélectrique d'Andekaleka.*

Abstract

Rainfall modelling and rainfall provision study : case of hydroelectric center of Andekaleka, in Madagascar

These rainfall modelling and rainfall provisional studies are based on the rainfall height accumulation function. It aims of the accumulation function rainfall provision, for example : the year 2020, using the random walk method. We have compared two different methods, such as : the robust regression method and the random walk method. For the model with the random walk method, the rainfall height distribution follows the log-normal law and the runoff has a α -establish law (with $\alpha = 0.4$). The results shows that, for the robust regression method, the annual height rainfall for the year 2020 is 1441 mm, as while the random walk gives 1122,8 mm. These two results are similar and slightly less than that of 2013 rainfall height accumulation (1519 mm). In addition, the drought events will be planned for the years 2025, 2032 and after the year 2038.

Keywords : *rainfall modelling, provisional study, rainfall height accumulation, random walk, robust regression, runoff, log-normal law, α -establish law, hydroelectric center of Andekaleka.*

1. Introduction

Cette recherche concerne l'étude prévisionnelle de pluie, en utilisant comme base de données la fonction cumulée des hauteurs de pluie. Nous savons que plusieurs nouvelles méthodes ont déjà été utilisées pour avoir un modèle sur l'étude prévisionnelle de la pluviométrie. Entre autres, le modèle basé sur l'Analyse en Composante Principale (ACP) [1], la modélisation mathématique [2], le modèle auto régressifs standard en analyse des séries temporelles saisonnières (SARIMA) [3], etc. L'objectif de notre étude est de pouvoir prédire la hauteur cumulée de pluie pour les années à venir, à l'aide de la méthode des marches aléatoires [4, 5]. Nous avons pris comme exemple le cas de la centrale hydroélectrique d'Andekaleka, et de faire la prévision de la hauteur annuelle de pluie pour l'année 2020. La méthode de la régression robuste sera utilisée afin de comparer les résultats obtenus.

2. Méthodologie

La zone d'Andekaleka fait partie du District de Moramanga, province de Toamasina. Elle se trouve à la latitude $18^{\circ}46'$ Sud et à la longitude $48^{\circ}36'$ Est, et couvre une superficie d'environ 600 km^2 entre les vallées confluentes. La zone est traversée par une petite rivière appelée Ivohitra, et cette dernière donne la chute d'eau de la centrale hydroélectrique du site.

2-1. Données

Les données pluies sont de données locales, allant du 1er janvier 1996 jusqu'au 31 décembre 2013 (durant 18 ans), comprend 6575 valeurs journalières, et provenant d'une station de mesure sur le lieu. L'unité de mesure pour la hauteur de pluie est le millimètre (mm) tandis que l'unité de temps est en jours. Les données sont complètes et à jours.

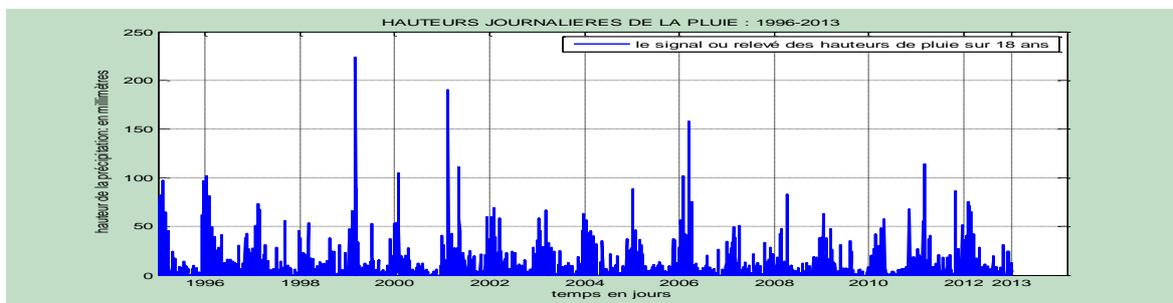


Figure 1 : Le signal : série temporelle constituée du relevé des valeurs journalières des pluies sur 18 ans (de 1996 à 2013)

Les variations observées sont très erratiques et suggèrent naturellement une évolution aléatoire. La valeur des pluies peut atteindre des valeurs supérieures à 100 mm au moment où des phénomènes extrêmes (cyclone) se présentent, tels que : cyclone, longue journée de pluie, etc. La période de 1996 à 2013 comprend 6575 jours dont 2866 jours sont avec pluie (valeur de pluie supérieure ou égale à 0,1 mm), et qui représente le 43,59 % de la période, tandis que les 3709 jours sont des jours sans pluie ou jours d'attente, et représentant le 56,41 % de la durée d'études. Il y a plus des jours sans pluie que des jours avec pluie. La valeur totale des pluies durant ces 18 années est de 26.475 mm (toutes valeurs de précipitations confondues), soit une moyenne annuelle de 1470,8 mm, soit une moyenne globale journalière de 4,026 mm avec un écart type de 10,65 mm. On peut noter que ces données renferment trois informations fondamentales. D'une part « les valeurs de pluie » qui caractérisent les précipitations, d'autre part les jours de sécheresse « temps d'attente », et enfin

les « valeurs de pluie cumulées au cours du temps » qui conjuguent les deux précédentes informations. La courbe en bleu de la **Figure 2** représente les valeurs cumulées des valeurs de pluie durant les 18 années d'affilées. Elle conjugue les deux informations concernant les jours d'attente et les jours de précipitation.

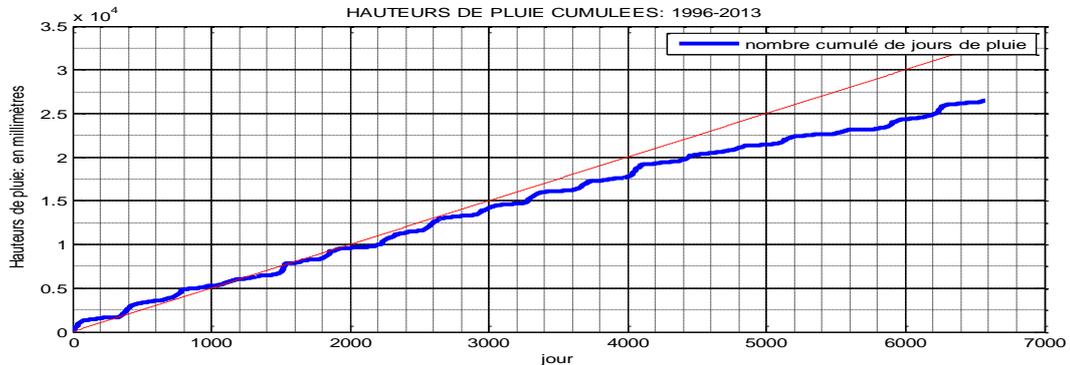


Figure 2 : *Hauteurs cumulées de pluies (1996 - 2013)*

Pour mieux savoir le taux de variation des valeurs cumulées de pluies, une droite linéaire (en rouge) représentant la tangente à la courbe, pour les jours du début des études, a été tracée. Nous avons constaté que la tendance à s'écarter de la droite tangente est très marquée à partir de l'année 2005 (année où avait débuté la tendance à la baisse). On va représenter graphiquement la fonction des hauteurs cumulées de pluie sous deux formes différentes, à savoir : Par la représentation linéaire et par la représentation normée (valeur finale est égale à 1).

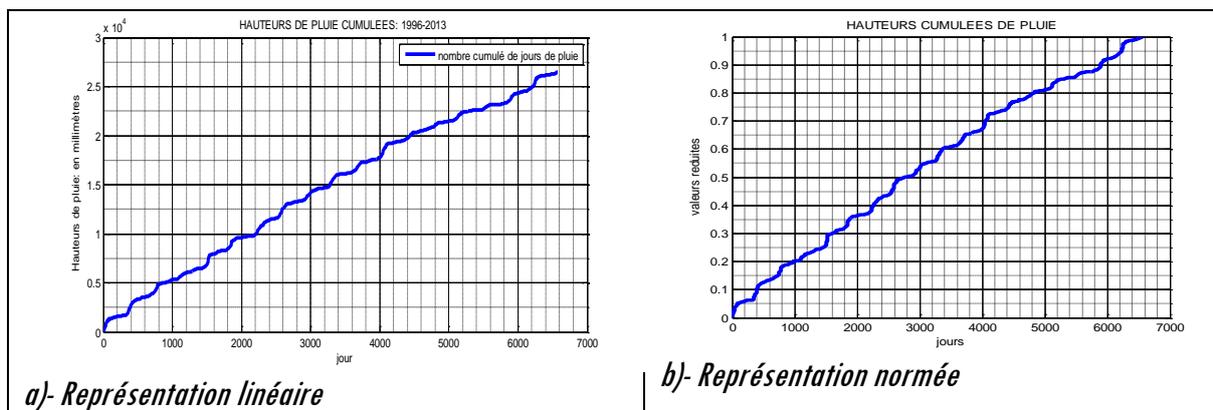


Figure 3 : *Hauteurs cumulées de pluies de 1996 - 2013*

Pour la représentation normée **Figure 3b**, la valeur finale du cumul est égale à 1. Cette présentation nous permet d'étudier la fonction cumule en matière de la probabilité.

2-2. Méthodes

Notre étude prévisionnelle de pluie se base sur la fonction cumulée des hauteurs de pluie. Pour la modélisation, nous allons utiliser de processus stochastique. Deux méthodes différentes seront à étudier, à savoir :

- i) - La méthode de régression robuste ;
- ii) - La méthode des marches aléatoires.

La méthode de régression robuste [6] est une amélioration de la régression simple. Elle minimise les écarts en éliminant les points « aberrants ».

Sur la modélisation par des marches aléatoires [5], le marcheur, ici dont la position représente la valeur de la hauteur de pluie, effectue une marche constituée d'un saut distribué suivant une loi log-normale puis s'immobilise durant un certain temps, dit temps d'attente, distribué selon une loi α -stable de type $\tau^{-1-\alpha}$ dans son comportement asymptotique avec $0 < \alpha < 1$ avant d'effectuer le saut suivant. Notre démarche consiste à déduire du modèle de marche aléatoire un modèle macroscopique de la dynamique des cumuls de pluie. Ce dernier a été obtenu par un passage à la limite de la marche aléatoire. En passant à la limite lorsque le pas des sauts et le pas de temps tendent simultanément vers zéro en gardant un rapport constant, on montre que la probabilité d'atteindre une hauteur cumulée de pluie fixée pour un temps donné est régie par les solutions de l'Equation Fractionnaire de Fokker-Planck (F.F.P.E.) [5]. L'obtention de cette équation, qui n'admet pas de solution analytique connue permet cependant d'avoir directement par calcul numérique cette probabilité et par conséquent de définir la hauteur de pluie la plus probable en localisant au mieux le maximum de la loi de probabilité par exemple pour un temps donné. Pour ce faire, on s'attache à rechercher la limite spatio-temporelle de la marche aléatoire ci-dessus. Cette limite sera une F.F.P.E. (Equation Fractionnaire de Fokker-Planck), exploitée avec des temps d'attente "subordonnés". Les simulations sont réalisées à partir de l'E.D.P.S. (Equation aux Dérivées Partielles Stochastiques) d'Îto [7] relativement simples d'utilisation. Pour un temps donné, la Fonction de la Densité de Probabilité (P.D.F) sera établie par la méthode de Monte-Carlo. L'estimation de la proximité des simulations à la courbe expérimentale sera évaluée, d'une part par les écarts de celle-ci à la trajectoire moyenne et d'autre part par la situation de la courbe expérimentale parmi les déciles des trajectoires et enfin par diverses mesures de leur dispersion.

2-2-1. Régression robuste

En théorie, la régression robuste [6] est une amélioration de la régression simple. La méthode des moindres carrés est sensible à la présence de données aberrantes situées loin de la droite de régression. On peut remédier à ce problème de deux façons :

- en éliminant les points « aberrants » ;
- en utilisant un autre critère que les moindres carrés.

La première solution peut être risquée et n'a de valeur que s'il s'agit effectivement de données erronées ou appartenant à une autre population, ce qui n'est pas toujours simple à déterminer. La deuxième approche a l'avantage d'être automatique et de fournir un modèle robuste convenant à la majorité des données, en perdant toutefois les propriétés d'optimalité des estimateurs des moindres carrés. Dans ce cas, on minimise la somme des valeurs absolues :

$$\sum_{i=1}^n |y_i^* - a - bx_i| \quad (1)$$

a. Ajustement sur des données

On dispose de « n » couples (x_i, y_i) ; $i = 1, \dots, n$ constituant un n-échantillon d'observations indépendantes de (X, Y) . On suppose vraie l'hypothèse :

$$E(Y/X) = \alpha + \beta X$$

$$\text{avec, } Y = \alpha + \beta X + \varepsilon \quad (2)$$

Le problème est donc d'estimer α, β ainsi que la variance σ^2 du résidu ε . Il suffit alors de supposer que pour chaque observation, on a $y_i = \alpha + \beta x_i + \varepsilon_i$ où les ε_i sont des réalisations indépendantes d'une variable ε d'espérance nulle et de variance constante σ^2 , quel que soit x_i .

- *Estimation de α, β, σ^2 par la méthode des moindres carrés*

La méthode des moindres carrés due à Gauss reprend sur l'échantillon la propriété que $E(Y / X) = \alpha + \beta X$ est la meilleure approximation de Y par X en moyenne quadratique. On cherche donc à ajuster au nuage des points (x_i, y_i) une droite **d'Équation** $y^* = a + bx$ de telle sorte que :

$$\sum_{i=1}^n (y_i - y_i^*)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 = F(a,b) \tag{3}$$

soit minimale. On étudie ensuite les propriétés de a et b en tant qu'estimations de α et β ainsi que l'estimation de $\hat{\sigma}^2$ de σ^2 que l'on en déduit. La méthode élémentaire de la détermination de a et b est la suivante : Ce minimum est atteint pour $\frac{\partial F}{\partial a} = \frac{\partial F}{\partial b} = 0$, ce qui donne les deux **Équations**,

$$\sum_{i=1}^n (y_i - a - bx_i) = 0 \text{ d'où } \bar{Y} = a + b\bar{x} \tag{4}$$

$$\sum_{i=1}^n x_i (y_i - a - bx_i) = 0$$

dont la solution est :

$$b = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = r \cdot \frac{S_x}{S_y} = \frac{cov(x,y)}{S_x^2} \tag{5}$$

$$a = \bar{y} - b\bar{x} \tag{6}$$

$$y^* = \bar{y} + r \cdot \frac{S_x}{S_y} (x - \bar{x}) \tag{7}$$

La droite des moindres carrés passe donc par le centre de gravité du nuage et sa pente est l'analogue empirique de la pente de la droite de régression $\rho \cdot \frac{\sigma_x}{\sigma_y}$

b. Prédiction d'une valeur ultérieure par la méthode de régression

Supposons maintenant que l'on désire prévoir à l'aide du modèle la valeur de Y pour une valeur non observée x_0 de X . La prédiction naturelle est :

$$y_0^* = a + bx_0 \tag{8}$$

Afin d'encadrer cette valeur, cherchons ce que l'on appelle un intervalle de prédiction. On a vu que y_0^* est distribué selon une loi gaussienne définie par :

$$LG(\alpha + \beta x_0; \sigma \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}) \tag{9}$$

Ainsi, $Y_0 - Y_0^*$ suit alors une loi LG $(0; \sigma \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}})$ et donc

$$\frac{Y_0 - Y_0^*}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}} \text{ suit la loi de Student } T_{n-2} \quad (10)$$

$$\text{D'après la table de la loi de Student : } P(|T_{n-2}| < 2,074) = 0,95 \quad (11)$$

D'où, pour encadrer les valeurs probables de Y_0 à 95 % de confiance, on calcule :

$$\left| \frac{Y_0 - Y_0^*}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}} \right| < 2,074 \quad (12)$$

D'où on trouve une expression de la forme : $\left| \frac{Y_0 - c_1}{c_2} \right| < 2,074$ avec c_1 et c_2 sont de constantes.

Ainsi, pour notre cas, l'Équation de la droite de régression est définie par :

$$f(x) = Y = 3,95x + 1578 \quad (13)$$

D'après l'Équation (13), pour l'année 2020, le nombre de jours du 01 janvier 1996 jusqu'au 31 décembre 2020 est égal à 9125 jours (à raison d'une année comprend 365 jours). Pour $x = 9125$, on a $Y = 37622$.

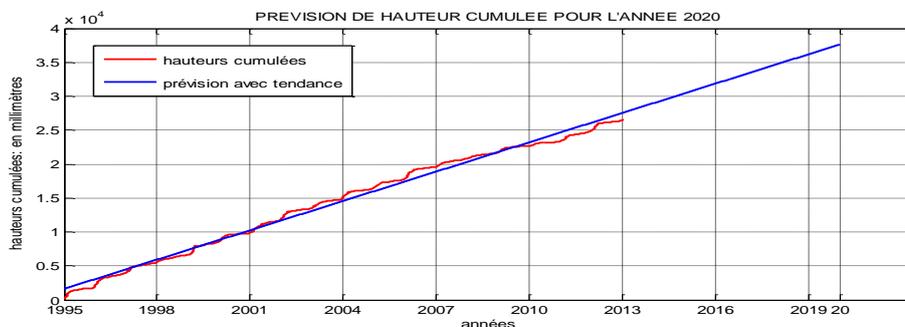


Figure 4 : Prédiction de la hauteur cumulée pour l'année 2020

La prédiction de la hauteur cumulée de 1996 à 2020 est environ 37.622 mm (ou bien 37,622 m). La prédiction de la hauteur cumulée seulement pour l'année 2020 est égale à 1441,8 mm.

Par rapport à l'année 2013 dont la valeur annuelle du cumul est égale à 1519 mm, il y aura une diminution de l'ordre de 77 mm entre les deux années (2013 et 2020). L'inconvénient de cette méthode de prédiction réside sur le fait qu'elle ne considère pas l'existence d'effet de mémoire dans les données. Cela augmente l'incertitude sur les résultats obtenus. L'erreur est de l'ordre de 5 à 10 %. Ces valeurs de prédiction seront-elles confirmées ou infirmées par la méthode de marches aléatoires ?

2-2-2. Marches aléatoires

a. Présentation de la loi alpha stable

La loi stable ou distribution de Lévy tronquée [8, 9], nommée d'après le mathématicien Paul Lévy, est une loi de probabilité utilisée en mathématiques, physique et analyse quantitative. L'hypothèse qui est faite dans cette étude est que le phénomène pluvieux local suit un processus stochastique qui reste à déterminer et pour cela, l'intensité des précipitations comme les temps d'attente entre deux pluies sont des variables aléatoires dont on s'attachera à déterminer les caractéristiques. Les variables aléatoires étant définies, il importe de les caractériser afin de les introduire dans le processus.

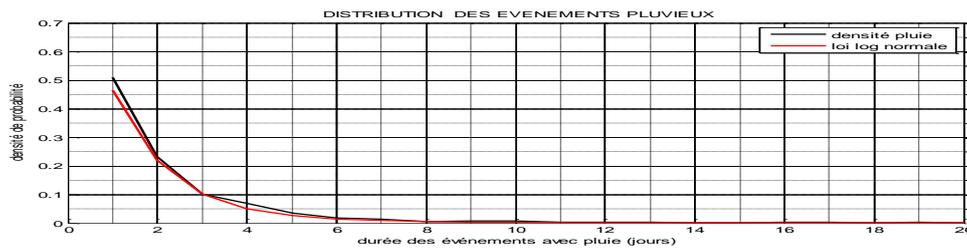


Figure 5 : Distribution des évènements pluvieux

La distribution des évènements pluvieux est en accord avec une loi de distribution de probabilité loi log-normale dont les paramètres sont : moyenne $\mu = 0,3$ et d'écart type $\sigma = 0,8$. Les résultats de nos analyses montrent que les intensités des précipitations suivent une loi de distribution de probabilité loi log-normale alors que la série des temps d'attente entre deux pluies, semble avoir une distribution de type α -stable. En effet, Gnedenko et Kolmogorov [10] ont établi une généralisation du théorème de la limite centrale selon laquelle la somme de variables aléatoires ayant des queues de distribution décroissantes selon $\frac{1}{|x|^\alpha}$ avec $0 < \alpha < 2$ (ayant donc une variance infinie) tend vers une loi stable de paramètre α .

b. Etudes des temps d'attente

La série des jours sans pluie (hauteurs de pluies inférieures à 2 mm) depuis 1996 à 2011 est donnée par la **Figure** ci-dessous :

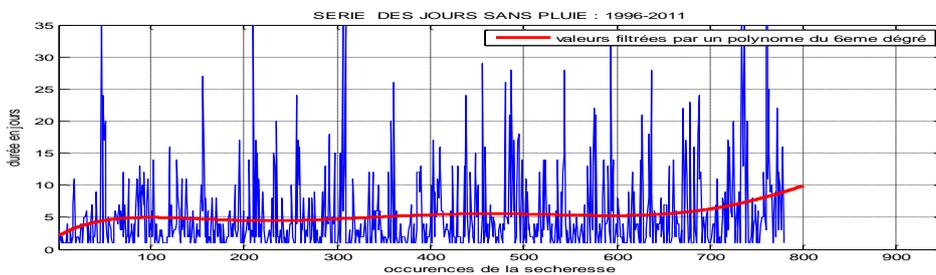


Figure 6 : Série des jours sans pluie

La tendance à la hausse des évènements de sècheresses des années 2005 - 2011 apparaissent sur le graphique des temps d'attente de la **Figure 6**. Nous allons chercher un modèle de loi de distribution de probabilité à ailes épaisses pour la série des temps d'attente.

Par la méthode de régression des queues [4], le test est basé sur le fait que le comportement asymptotique de la fonction de survie $(1-F(X))$ de la variable aléatoire X est de type puissance de $(-\alpha)$ soit :

$$\lim_{t \rightarrow \infty} t^\alpha \Pr(|X| > t) = \gamma C(\alpha) \quad (14)$$

Donc à l'infini $\frac{\log(P(|X|>t))}{\log(t)}$ est équivalent à α . D'où la représentation en log-log de la fonction de survie va donner une pente finie pour de grandes valeurs de t . Cette pente est alors, en valeur absolue, une évaluation de l'exposant α . Reprendre l'ajustement ci-dessus en considérant que les écarts $\Delta Y = y_i - y_k$ restent proportionnels aux écarts $\Delta X = \log(x_i) - \log(x_k)$

$$D'où \alpha = \frac{\Delta Y}{\Delta X} = \left| \frac{Y_B - Y_A}{X_B - X_A} \right| \quad (15)$$

Une partie de cette courbe a une tendance linéaire avant de prendre la position un peu verticale. Pour la **Figure 7**, c'est la pente de la courbe qui nous intéresse. Elle est présentée par un segment de droite (en vert). Pour avoir la pente de ce segment, on sait bien que deux droites parallèles ont la même pente. Ainsi, il suffit de calculer la pente de la droite en rouge qui est parallèle au segment [11].

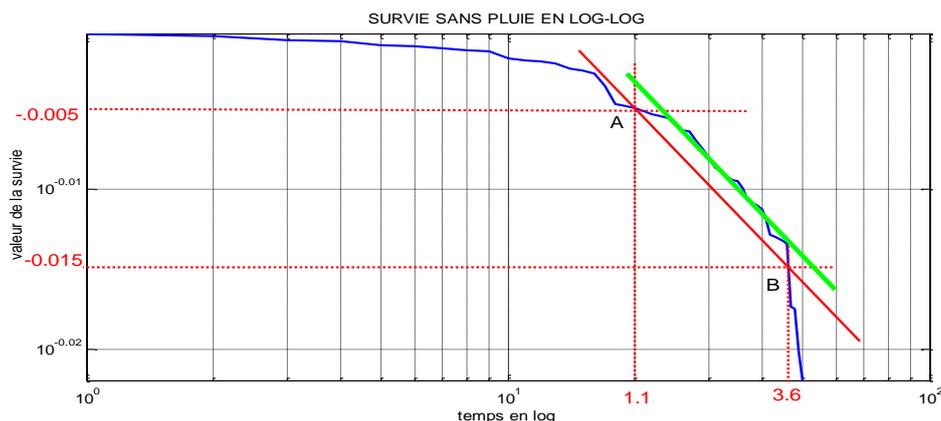


Figure 7 : Estimation graphique de l'exposant de stabilité α

Les coordonnées des deux points A et B sont marquées sur la **Figure 7** par leur exposant. Si l'on calcule la pente, on arrive à trouver :

$$On \alpha = \left| \frac{0,015 - 0,005}{3,6 - 1,1} \right| = \frac{1}{2,5} \text{ d'où } \alpha \approx 0,4 \quad (16)$$

La valeur de α est inférieure à 1. Les points A et B ont les coordonnées A (1,1 ; -0,005) et B (3,6 ; -0,015) car on ne prend que l'exposant de « 10 » comme coordonnées des points pour la représentation en log-log.

2-2-3. Construction du modèle de marches aléatoires des hauteurs cumulées de pluies

Notre approche par une marche aléatoire en espace et en temps des précipitations étant assimilées à des sauts distribués suivant la loi log-normale estimée précédemment et le temps d'attente distribué suivant une loi 0,4-stable. Chaque saut est alors la contribution à la marche aléatoire cumulée au temps t . Cette marche

aléatoire comporte donc deux dimensions, l'une spatiale et l'autre temporelle. Pour la première, la hauteur de pluie cumulée x_n du nième jour, est égale à la hauteur du jour précédent augmentée de la précipitation ou encore du saut j_n du jour.

$$x_{n+1} = x_n + j_n \tag{17}$$

où, j_n est une variable aléatoire de loi log-normale dont la loi de densité de probabilité s'écrit

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \frac{1}{x} \cdot \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \text{ avec } x > 0 \tag{18}$$

Comme il a été précisé précédemment, les caractéristiques μ, σ ; moyenne et écart-type de j_n , sont calculées à partir de la moyenne et variance (m, var) de la série expérimentale selon la **Formule** :

$$\mu = \ln(m) - 0.5 * \ln\left(1 + \frac{var}{m^2}\right) \text{ et } \sigma^2 = \ln\left(1 + \frac{var}{m^2}\right) \tag{19}$$

L'Équation (25) donne $\mu = 0,3$ et $\sigma = 0,8$.

Il est à noter ici que ces valeurs de μ et σ ne constituent qu'une estimation "à priori". Elles seront réévaluées par ajustement des simulations (trajectoires) aux données expérimentales. On peut aussi utiliser une forme centrée de la variable de saut j_n , pour réécrire **l'Équation (17)** sous une forme plus standard de marche aléatoire

$$\begin{aligned} x_{n+1} &= x_n + \mu + (j_n - \mu) \\ x_{n+1} &= x_n + V \cdot \tau + \tilde{j}_n \end{aligned} \tag{20}$$

où, τ est une variable temporelle qu'on associera à un pas de temps. Le paramètre $V = \mu / \tau$ possède la dimension d'une vitesse. Ce paramètre V représente une vitesse moyenne d'advection dans la marche aléatoire, dans le sens où elle traduit un déplacement moyen systématique (de valeur μ) à chaque pas de temps. La variable aléatoire \tilde{j}_n , obtenue à partir de j_n par translation, suit donc une loi log normale centrée $L(0, \sigma^2)$ de densité de probabilité ϕ définie par :

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \frac{1}{(x+\mu)} \cdot \exp\left(-\frac{(\ln(x+\mu)-\mu)^2}{2\sigma^2}\right) \text{ avec } x > -\mu \tag{21}$$

la marche aléatoire, dans sa dimension spatiale, peut encore s'écrire :

$$x_{n+1} = x_n + \mu + \bar{\bar{j}}_n \tag{22}$$

en fonction de la variable aléatoire $\bar{\bar{j}}_n = \frac{1}{\sigma} \tilde{j}_n$ qui, elle, suit une loi log-normale centrée réduite (notée $L(0, 1)$).

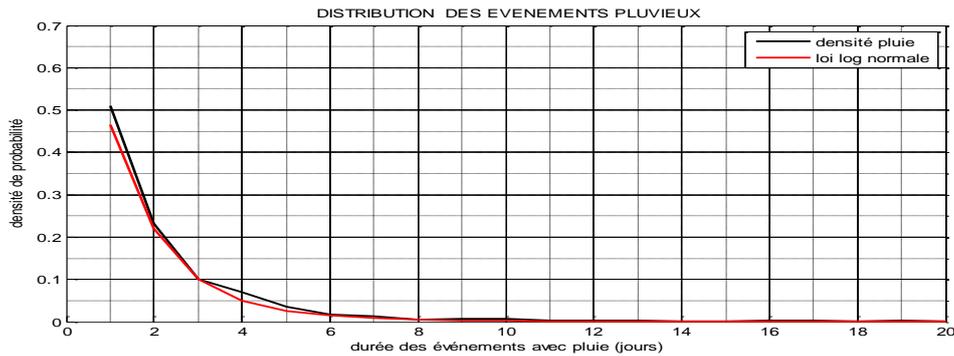


Figure 8 : Distributions log-normales de \tilde{j}_n , simulant les sauts de la marche aléatoire

En dimension temporelle, on rappelle qu'un "temps d'attente" est un temps d'attente entre deux relevés de précipitations non nulles, donc que ce temps est strictement positif. Si τ est le pas de temps de la marche aléatoire et W_n est une variable aléatoire α -stable alors la durée entre deux sauts est donnée par $\tau^{\frac{1}{\alpha}} W_n$. Avec comme condition aux limites $x_0 = 0$ et $t_0 = 0$, la marche aléatoire est alors résumée par l'Équation suivante :

$$\begin{cases} x_{n+1} = x_n + V + \ell \cdot \tilde{j}_n \\ t_{n+1} = t_n + \tau^{\frac{1}{\alpha}} W_n \end{cases} \quad (23)$$

Dans l'Équation (23) nous avons introduit un pas d'espace ℓ . Le rapport $D = \frac{\ell^2}{\tau}$ constitue le coefficient de dispersion de la marche aléatoire. Comme on le verra dans la section suivante, ce modèle admet comme limite hydrodynamique (quand τ et ℓ tendent vers zéros mais D reste constant) un processus stochastique dont la loi de probabilité marginale est gouvernée par l'équation fractionnaire de Fokker Planck.

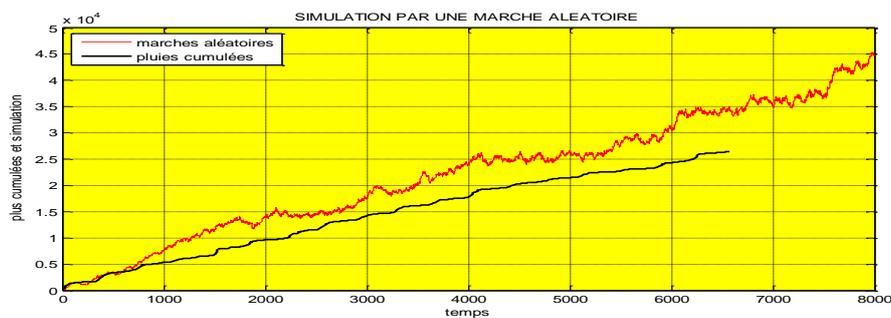


Figure 9 : Hauteurs de pluie cumulées simulées par des marches aléatoires non soumises aux caractéristiques de la série temporelle des pluies cumulées

Trois méthodes vont être utilisées pour réduire les écarts de manière à mieux cerner la réalisation dont on dispose et dont on veut établir le processus qui la régit. Rappelons que ces trois méthodes sont :

- "Tempérer" les temps d'attente en les bornant par une distribution exponentielle de moyenne λ ;
- "Trier" les temps d'attente en leur imposant une durée voisine de 6575 jours, longueur du relevé des hauteurs de pluies ;
- "Sélectionner" les temps d'attente de manière à présenter les mêmes proportions de temps d'attente entre deux pluies non consécutives que la série expérimentable.

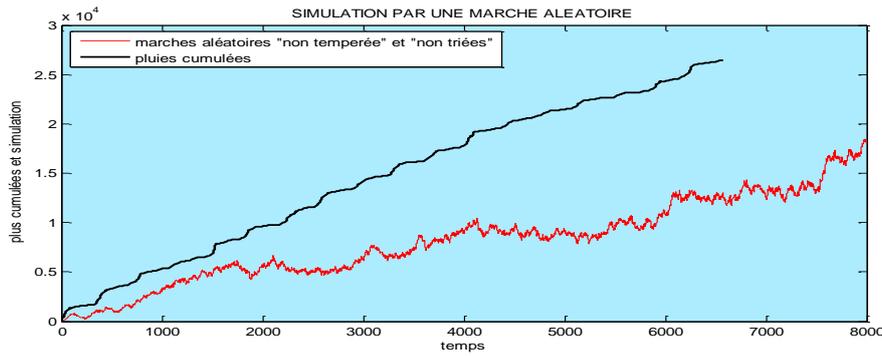


Figure 10 : Simulation par des marches aléatoires non tempérées et non triées

3. Résultats et discussion

3-1. Ajustement de la distribution des temps d'attente

a. Temps d'attente α - stables "tempérés"

Parmi les temps d'attente α -stables utilisés pour simuler les trajectoires ci-dessus certains ont une durée trop importante par rapport au processus que l'on souhaite décrire. Pour limiter la probabilité des évènements extrêmes, une méthode consiste à borner la distribution des temps $\tau \frac{1}{\alpha} W_n$ par une distribution "sans mémoire", exponentielle de moyenne λ [4, 12]. En effet, un modèle subdiffusif présentant une distribution α -stable des temps d'attente donne des simulations qui surestiment les très longues durées. De telles situations ont été observées et corrigées par l'introduction d'une loi de distribution "alpha-stable tempérées" [12]. Une telle loi est définie par la transformée de Laplace de sa fonction densité de probabilité dont les moments de tous ordres sont finis, soit $f(u) = \exp(\lambda^\alpha - (u + \lambda)^\alpha)$. Techniquement, il s'avère que la valeur de λ devra être relativement élevée pour acquérir suffisamment de valeurs ayant les caractéristiques souhaitées dans le cadre de notre phénomène pluvieux local. La **Figure 11** montre bien un encadrement très resserré des hauteurs de pluie cumulées par des simulations de marches aléatoires non tempérées, quasiment linéaires. Par ailleurs, il subsiste des temps exagérément longs qui vont au-delà de la longueur de la série expérimentale.

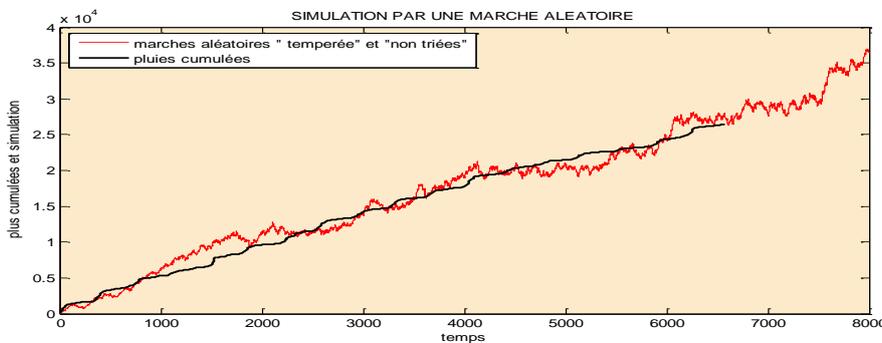


Figure 11 : Simulation par des marches aléatoires « tempérées » et « non triées »

b. Temps d'attente α -stables "triés"

Le bornage exponentiel ci-dessus produit des temps d'attente dépassant largement la dimension temporelle de la série expérimentale dont on dispose. Pour éliminer les distributions démesurées, on impose au temps de la marche aléatoire de ne pas dépasser de plus de 10 à 20 % le temps du relevé expérimental des hauteurs de pluie considérées.

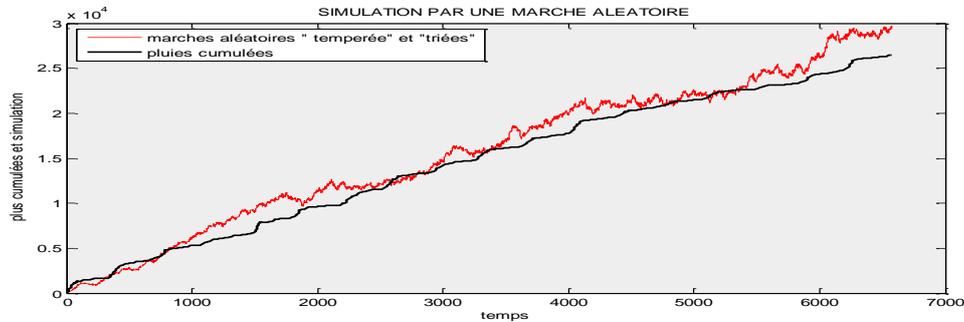


Figure 12 : Simulation par des marches aléatoires « tempérées, Triées et non sélectionnées »

b. Temps d'attente α -stables "sélectionnés"

La seule réalisation du processus dont on dispose, les hauteurs de pluie cumulées, présente 43,6 % "de sauts" qui se succèdent à un jour d'intervalle et par conséquent 56,4 % de temps d'attente supérieurs à un jour entre deux sauts. Graphiquement, ceci se traduit par des "plateaux" au cours desquels la hauteur de pluie cumulée n'évolue pas. L'idée retenue ici est de sélectionner, parmi les simulations de trajectoires tempérées et triées, celles qui présentent un taux de temps d'attente comparable; entre 70 et 90 %. Il apparaît alors un "resserrement" des trajectoires autour des hauteurs de pluie cumulées. Pour une lecture plus aisée des graphiques qui permettra de mieux apprécier la corrélation entre la série expérimentale et les trajectoires simulées, on peut réduire l'ensemble des trajectoires à la trajectoire moyenne d'une part et situer la série expérimentale par rapport aux déciles des trajectoires. Situer la courbe expérimentale par rapport aux déciles des simulations. Pour chaque trajectoire de la marche aléatoire, apparaît en abscisse son temps opérationnel propre, c'est-à-dire définie aléatoirement. Afin de représenter la trajectoire moyenne et les "déciles" il est nécessaire de prendre les points des trajectoires qui ont une même abscisse. Un pas de temps de un jour étant choisi et les trajectoires étant "en escalier", il suffit de prendre en ordonnée, celle du point d'abscisse la plus voisine "à gauche" du jour considéré.

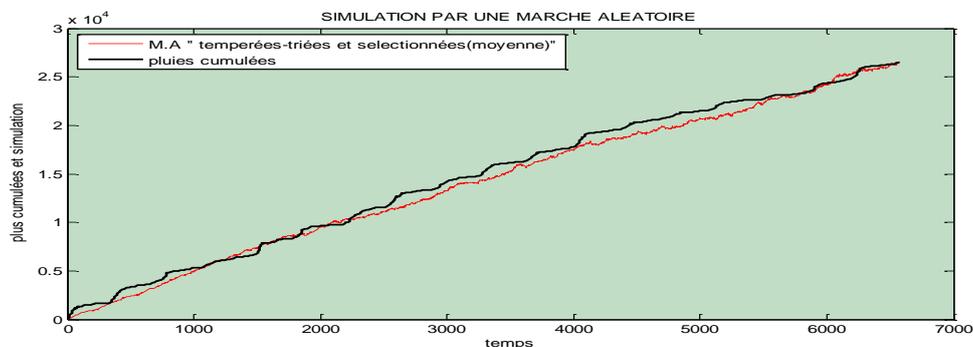


Figure 13 : Simulation par des marches aléatoires tempérées, triées et sélectionnées

Il faut noter le "fléchissement" de la trajectoire moyenne qui marque tout l'intérêt des distributions tempérées des temps d'attente. Ayant cerné le processus qui approche notre phénomène pluvieux local, on va évaluer les hauteurs de pluie les plus probables pour un temps donné et les comparer à la série initiale. Cette marche aléatoire est très proche de la marche aléatoire géométrique tempérée dont les paramètres sont définis par :

$$\alpha = 0,4 \quad \lambda = 10^{-5} \quad \sigma = 0,5 \tag{24}$$

En connaissant ces paramètres, nous pouvons les utiliser pour les prévisions à court ou à long terme. Mais l'erreur augmente au fur et à mesure que la date de prévision éloigne la dernière date des données.

3-2. Limites des marches aléatoires

L'objet de ce passage à la limite pour la marche aléatoire ci-dessus est de tenter de gérer le phénomène de manière continue en temps et en espace pour déterminer la probabilité pour qu'une trajectoire suivie par une hauteur de pluie cumulée prenne une valeur x à un instant t .

a. Équation Fractionnaire de Fokker-Planck (F.F.P.E)

[5] ont montré que dans le cas de distributions α - stables avec $\alpha < 1$ donc de moyenne infinie, en limite d'une C.T.R.W. on obtenait une F.F.P.E.(Fractionary Fokker-Planck Equation) et semblablement, Marie-Christine Néel et al montraient la même chose dans le cas $\alpha > 1$. Ainsi, si $P(x, t)$ est la probabilité pour qu'une trajectoire prenne la valeur x à l'instant t , alors $P(x, t)$ est une solution de **l'Équation** différentielle partielle :

$$\partial_t P(x, t) = - \partial_x (V(P(x, t))) + D (\sigma^2 - 2K) \partial_{x^2}^2 (P(x, t)) \tag{25}$$

avec, $2K = 2 \int_{-\mu}^0 Y^2 \varphi(Y) dY$ et $D = \frac{\ell^2}{2\tau}$

On remarque que cette **Équation** est celle de Fokker-Planck adaptée à une variable aléatoire distribuée suivant une loi log-normale c'est-à-dire que c'est le logarithme de la variable qui suit une loi normale et non pas la variable elle-même. Un tel mouvement est dit "brownien géométrique". La différence entre les deux expressions porte sur le terme multiplicatif $(\sigma^2 - 2K)$.

b. Distribution des temps d'attente suivant une loi α – Stables

L'étude de la série des temps d'attente entre deux pluies faite précédemment a montré que celui-ci pouvait être approché par une distribution qui suivrait une loi de type α -stables avec ici $\alpha = 0.4$, α étant inférieur à 1, non seulement la variance est infinie mais la moyenne aussi. Ce cas a fait l'objet d'études spécifiques [5] qui conduisent à l'Equation fractionnaire de Fokker-Planck notée F.F.P.E. (Fractionary Fokker-Planck Equation). Celle-ci, en tenant compte du paragraphe précédent où le mouvement brownien géométrique se substitue au mouvement brownien prend la forme :

$$\partial_t P(x, t) = {}_0D_t^{1-\alpha} \left[- \partial_x (V(P(x, t))) + D (\sigma^2 - 2K) \partial_{x^2}^2 (P(x, t)) \right] \tag{26}$$

ou, ${}_0D_t^{1-\alpha}$ est la dérivée fractionnaire de Riemann-Liouville pour α compris entre 0 et 1, ce qui est le cas ici.

Cette **Équation** régit la dynamique d'un système, ici l'évolution des hauteurs de pluie cumulées, de type anormal, sous diffusif, qui suit une décroissance assez lente, en "puissance" et non "exponentielle". Dans le cas qui nous occupe, $\alpha = 0.4$, l'**Équation** ci-dessus qui régit le processus stochastique suivi par les hauteurs de pluie cumulées, n'a pas aujourd'hui de solution analytique. Un moyen d'approcher une solution est de simuler des trajectoires répondant au processus. Une telle approche sera faite au moyen de deux processus qui substituent un "temps opérationnel τ " au "temps physique t ". C'est le principe de subordination.

3-3. Études prévisionnelles

Ces études prévisionnelles reposent sur deux méthodes différentes, dont :

- L'extrapolation du processus tronqué ;
- La résolution numérique de l'FFPE :
 - i) simulée par l'équation différentielle d'Itô ;
 - ii) par discrétisation directe de l'FFPE.

3-3-1. Méthode de processus tronqué [12]

Le processus est dit « tronqué » si les trajectoires utilisées sont « tempérées, triées et sélectionnées ». Les résultats obtenus ci-dessous sont assimilés par ordinateur à l'aide du langage de programmation Matlab. On a déjà vu que la prévision de la valeur totale du cumul depuis l'année 1996 jusqu'au 31 décembre 2020 se calcul à partir du rang du dernier jour de l'année 2020, qui est égal au 9125^e jour. On prend comme période $T = 9125$ (triage). Reprenons les paramètres tout à l'heure avec des plusieurs simulations et prenons la moyenne (sélection).

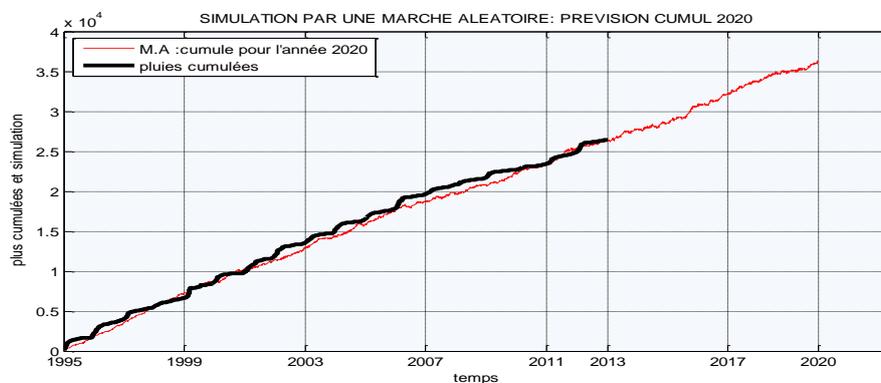


Figure 14 : Prévion par une marche aléatoire du cumul pour l'année 2020

Cette courbe nous donne la prévision des valeurs des hauteurs cumulées suivantes : Hauteur totale cumulée de 1996 à 2020 est égale à 36 249 mm (ou 36,25 m), tandis que la hauteur cumulée annuelle de l'année 2020 est égale à 1263 mm. Par rapport à la méthode de régression robuste, on a trouvé une diminution de valeurs pour la hauteur cumulée totale de l'ordre de 1373 m (soit 1,37 m) et celle du cumul annuel de l'année 2020 dont le décalage est de 179 mm. L'ordre d'erreur de confiance est de 5 %.

Pour la Prévion à long terme, notre principal objectif sur cette étude relative à la prévion à long terme, vue l'existence d'effet de mémoire dans les données, est de savoir le comportement dans son ensemble des phénomènes pluviométriques plutôt que de connaître les hauteurs cumulées. Car, pour cette dernière, l'ordre d'erreur s'augmente suivant l'éloignement de la date de prévion.

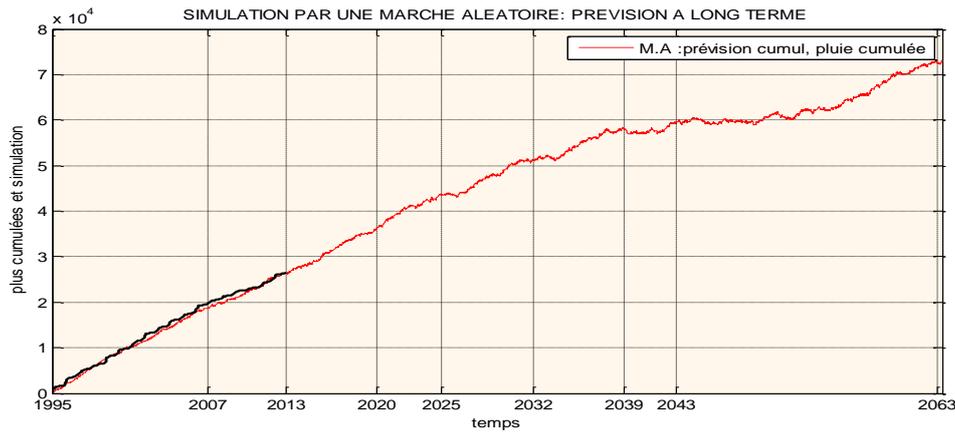


Figure 15 : Prévision à long terme du cumul par des marches aléatoires

La **Figure 15** nous montre qu'il y a de moment où la courbe se fléchisse et se stabilise sous la forme d'un plateau. Ce phénomène nous signale qu'il y aura des évènements de sécheresse très marqué durant les années correspondantes, car le cumul s'immobilise vue l'insuffisance de pluie. Ces années sont les suivantes : Années 2025, 2032 et après l'année 2038. La marge d'erreur est de l'ordre d'une année.

3-3-2. Méthode régie par l'équation fractionnaire de Fokker-Planck (F.F.P.E)

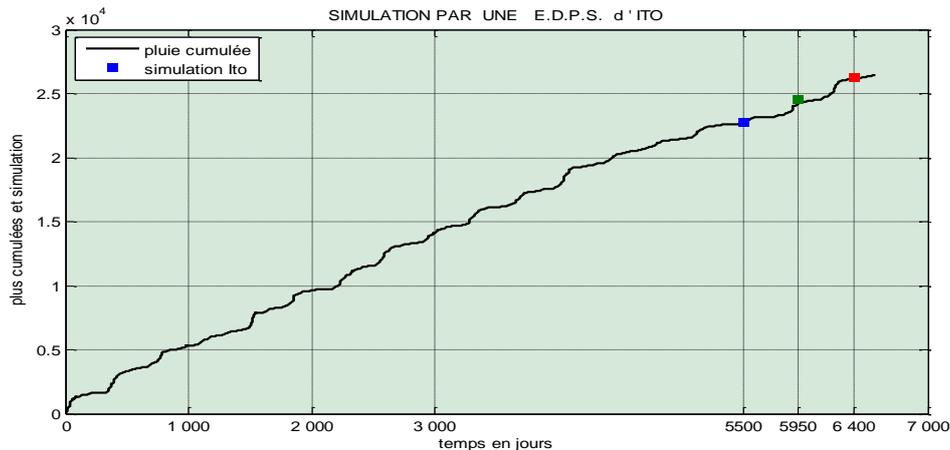


Figure 16 : Estimation du cumul par la simulation de l'EDPS d'Îto

Cette résolution arrive à suivre le fléchissement de la courbe. Du moins pour les trois moments ci-après : 5500^e, 5950^e et 6400^e jour. Pour la prévision du cumul de l'année 2020 par la simulation de l'E.D.P.S d'Îto, on trouve les valeurs probables suivantes : la hauteur cumulée de pluie du 1^{er} janvier 1996 jusqu'au 31 décembre 2020 est égale à 36371mm (soit 36,37 m) et la hauteur cumulée annuelle de l'année 2020 vaut 1234,6 mm.

De plus, une résolution numérique du modèle macroscopique est possible. Les sections précédentes ont permis d'estimer au mieux les paramètres du processus. Ils peuvent être alors utilisés dans une discrétisation de F.F.P.E. réalisée sur un P.C. qui accepte un nombre suffisamment grand de simulations donc en affinant les résultats. La discrétisation de la dérivée fractionnaire de Riemann-Liouville a été réalisée de manière implicite en temps et en utilisant la méthode de Grünwald-Letnikov en x.

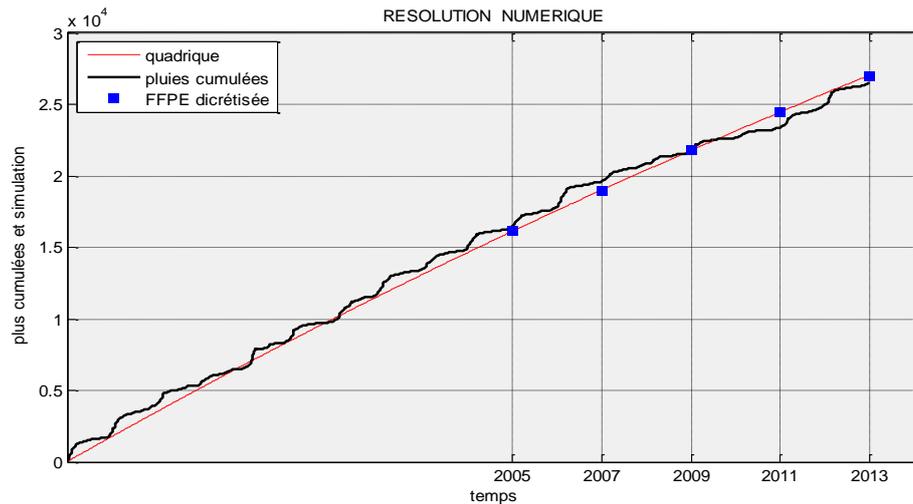


Figure 17 : Estimation du cumul par la discrétisation de la FFPE.

La **Figure 17** illustre les hauteurs de pluie cumulées les plus probables obtenues avec la solution numérique de FFPE pour les années : 2005, 2007, 2009, 2011 et 2013. On remarque que la modélisation ne reproduit pas le fléchissement.

3-3-3. Prédiction de la valeur du cumul de pluie en 2020

A titre de curiosité, nous avons réalisé une extrapolation de ces processus à 2020, nous avons trouvé un cumul total de 35341 mm (soit 35,341 m). Et, pour le cumul annuel de l'année 2020 est égal à 1122,8 mm. Le **Tableau 1** montre la comparaison des résultats de la prédiction du cumul pour l'année 2020, par les trois méthodes prévisionnelles.

Tableau 1 : Comparaison des résultats de la prédiction du cumul pour l'année 2020

Pour l'année 2020	Régression Robuste	Processus Tronqués	Simulation par L'EDPS d'Îto	Résolution Numérique de la FFPE
Cumul total	37622 mm	32649 mm	36371 mm	35341 mm
Cumul annuel	1441,8 mm	1263 mm	1234,6 mm	1122,8 mm

La méthode de la régression robuste prévoit une grande valeur pour le cumul de l'année 2020, aussi bien pour le cumul total que pour le cumul annuel, tandis que le processus tronqués et la résolution numérique de la FFPE trouvent respectivement de valeur minimale pour le cumul total et le cumul annuel.

4. Conclusion

L'étude de prévision par la régression robuste nous a servi comme référence, en tant qu'une méthode classique, à l'étude des marches aléatoires. Le processus élaboré à partir de la marche aléatoire spatio-temporelle suit une loi log normale en espace et une loi 0,4-stable en temps. On a montré qu'il approchait de façon relativement satisfaisante le phénomène pluvieux local représenté ici par la série des hauteurs de pluie cumulées. En passant à la limite lorsque saut et pas de temps tendent simultanément vers zéro en gardant un rapport constant, les solutions de l'Equation Fractionnaire de Fokker-Planck adaptée aux mêmes conditions

que la marche aléatoire montre, à son tour et dans un premier temps, une approche convenable de ce même phénomène. Pour tenir compte du fléchissement de la courbe, le processus a été « tempéré ». On peut alors admettre que les divers modèles proposés ; la marche aléatoire et les trajectoires F.F.P.E. en Temps « subordonné, tempéré et sélectionné » réalisent une bonne approche du phénomène pluvieux étudié. La prévision du cumul pour l'année 2020 par la méthode de régression robuste nous a donné une valeur de 37.622 mm, tandis que le cumul annuel de l'année 2020 est de l'ordre de 1441 mm. Cette valeur prévisionnelle de la hauteur cumulée de pluie pour l'année 2020 est inférieure à celle de l'année 2013, qui est égal à 1519 mm. En ce qui concerne la prévision à long terme, quelques événements de sécheresse seront prévues pour les années 2025, 2032 et après l'année 2038.

Références

- [1] - P. NATHALIE, Une nouvelle approche pour la prévision statistique des précipitations saisonnières en Afrique de l'Ouest et de l'Est. Thèse de Doctorat, Université de Bourgogne. UFR des Sciences humaines. Présentée le 20/12/2002
- [2] - J. DE REFFYE, Etude phénoménologique des précipitations pluvieuses. Modélisation mathématique des intensités de pluie en un point du sol. *Revue de Statistique Appliquée*. Tome 30, N°3 (1982) 39 - 63
- [3] - F. KADRI, S. CHAABANE, F. HARROU, et C. TAHON, Time series modeling and forecasting of emergency department overcrowding, *Journal of Medical Systems*, 38 (9) (2014) 107. DOI : 10.1007/s10916-014-0107-0. Epub 2014 Jul 23
- [4] - A. CHAKRABARTY and M. M. MEERSCHAERT, Tempered stable laws as random walk limits. *Statistique and Probability letters*. Vol. 81, N°8 (2011) 989 - 997. DOI : 10.1016/j.spl.2011.01.019. 2011
- [5] - E. BARKAI, R. METZLER and J. KLAFTER, From continuous time random walks to the fractional Fokker-Planck equation. *Physical Review E*, Vol. 61, N° 1 (2000) 132 - 138. DOI : 10.1103/PhysRevE.61.132. 2000
- [6] - JOHN Fax. Robust Regression. Appendix to R and S. Plus Companion to Applied Regression, (2002)
- [7] - P. KUNDU and J. TRAVIS, A Stochastic Fractional Dynamics Model of Rainfall Statistics. Geophysical Research Abstracts, Vol. 15, EGU General Assembly 2013. Held 7-12 April, 2013 in Vienna, Austria, id. EGU2013-1491. 2013EGUGA.15.1491K
- [8] - C. L. NIKIAS and M. SHAO, Signal Processing with Alpha-Stable Distributions and Applications. Ed. Prentice Hall, (1995)
- [9] - A. ALVAREZ et P. OLIVARES, Méthode d'estimation pour des lois alpha-stable avec des applications en finance. *Journal de la Société Française de Statistique*, Tome 146, N° 4 (2005)
- [10] - D. KOUTSOYIANNIS, A. PASCHALIS, N. Theodoratos Two-dimensional Hurst-Kolmogorv process and its application to rainfall fields, (2011)
- [11] - J. H. MCCULLOCH, Measuring tail thickness in order to estimate the stable index α : a critique. *Business and Economic Statistics*, Vol. 15 (1997) 74 - 81
- [12] - J. GAJDA and M. MAGDZIARZ, Fractional Fokker-Plank equation with tempered alpha-stable waiting times : Langevin picture and computer simulation. *Physical Review E*, Vol. 82 (2010)