

**Université Pierre et Marie Curie, Université Paris-Sud,  
Ecole des Mines de Paris  
& Ecole Nationale du Génie Rural des Eaux et des Forêts**

---

**DEA Hydrologie, Hydrogéologie, Géostatistique et Géochimie  
Filière Hydrologie et Hydrogéologie Quantitatives**

**ETUDE DE LA DISTRIBUTION STATISTIQUE  
DES PLUIES ANNUELLES**

**Auteur : Harouna KARAMBIRI**

**Directeur de recherche : M. Pierre HUBERT**

**Laboratoires d'accueil: CIG Ecole des Mines de Paris  
LGA UPMC Paris 6**

**septembre 1999**



## Remerciements

Je tiens tout particulièrement à remercier mon encadreur, M. Pierre HUBERT, qui a bien voulu me proposer ce sujet et me suivre tout au long de ce travail de mémoire.

Je remercie également M. Hocine BENDJOUDI, pour sa grande disponibilité et ses conseils.

Je n'oublie pas Mlle Keltoum CHAOUICHE, pour sa collaboration et son appui aux questions mathématiques et statistiques.

Que tous ceux ou toutes celles, qui m'ont aidé et soutenu, d'une manière ou d'une autre, de loin ou de près, trouvent ici l'expression de ma profonde et sincère gratitude.

Harouna KARAMBIRI.

## SOMMAIRE

	Page
RESUME.....	1
INTRODUCTION.....	2
I- Présentation des données.....	3
I-1- Description de la base de données.....	3
I-2- Analyse à l'échelle globale.....	4
II- Multifractalité et Hydrologie.....	8
III- Distribution des séries annuelles.....	11
III-1- Limites de la loi normale.....	11
III-2- Application des multifractals à l'étude des séries annuelles.....	12
III-3- Estimation de l'exposant de décroissance algébrique $q_D$ .....	13
III-4- Essai de proposition d'une fonction densité de probabilité.....	15
III-4-1- Etude et tracé de la fonction.....	17
III-4-2- Paramètres statistiques.....	20
a) Mode.....	20
b) Médiane.....	21
c) Moments.....	22
III-4-3- Estimation des paramètres de la loi théorique à partir des données.....	23
IV- Essai d'une autre méthode d'estimation de l'exposant de décroissance algébrique $q_D$ .....	26
IV-1- Approche théorique.....	26
IV-2- Application aux données.....	30
CONCLUSION.....	31
REFERENCES BIBLIOGRAPHIQUES.....	32
ANNEXES.....	34

## Résumé

L'étude statistique des séries pluviométriques revêt une importance capitale en hydrologie, surtout en ce qui concerne le dimensionnement des ouvrages hydrauliques et l'étude des aménagements hydro-agricoles. C'est ainsi que l'on retrouve dans la littérature toute une pléthore d'ouvrages sur le sujet. Si depuis des décennies, les lois statistiques classiques ont fait le bonheur des hydrologues statisticiens, aménageurs, concepteurs et autres, aujourd'hui, un constat amer se pose : Toutes ces lois classiques, souvent utilisées abusivement, trouvent des limites dans l'étude des événements d'occurrence rare. Des études récentes (Hubert et Bendjoudi, 1996) ont montré que la loi de Laplace-Gauss (loi normale) qui est la plus utilisée pour la modélisation statistique des pluies annuelles, mais aussi toutes les autres lois à décroissance exponentielle, ne permettent pas d'ajuster d'une façon satisfaisante les queues de distribution des séries longues. Faut-il donc en finir avec la normalité ?

De nouveaux outils basés sur les démarches fractales et multifractales ont été introduits et ont permis de montrer que les séries pluviométriques suivent plutôt une loi à décroissance algébrique moins rapide qu'une décroissance exponentielle, et que l'exposant  $q_D$  pourrait être invariant d'échelle et universel.

Dans ce rapport, nous avons appliqué cette approche à 233 séries pluviométriques annuelles de durée supérieure ou égale à 100 ans. Le paramètre  $q_D$  a été estimé à **4,5±1**. Nous proposons également une nouvelle loi algébrique de distribution des probabilités ainsi qu'une autre méthode d'estimation de  $q_D$ .

# Introduction

Les précipitations constituent les principales entrées des systèmes hydrologiques tels que les bassins versants. Elles sont vitales pour les êtres humains dans la mesure où elles assurent la recharge des ressources en eau et favorisent le développement des activités humaines (agriculture, pêche, production d'énergie,...). Mais, elles peuvent devenir aussi dangereuses, voire mortelles, en causant des inondations.

Voilà plusieurs siècles que l'homme a commencé à s'intéresser à ces phénomènes naturels en essayant de les mesurer, de les analyser, de comprendre leur origine et les processus selon lesquels ils se manifestent. Mais hélas, il a été très souvent limité par les instruments de mesure et les outils d'approche. Les seules mesures fiables disponibles aujourd'hui datent seulement des siècles derniers (18<sup>ème</sup> siècle).

L'avènement des nouvelles technologies telle que l'informatique, a permis une avancée considérable en hydrologie en ce qui concerne l'acquisition, le stockage et le traitement des données pluviométriques. C'est ainsi que l'on a vu naître toute une pléiade de modèles mathématiques et statistiques.

Pendant longtemps, les hydrologues se sont contentés de développer des modèles ad hoc, spécifiques à une certaine échelle de temps, pour proposer des solutions aux problèmes pratiques urgents, avec pour seul souci la qualité de l'ajustement aux données empiriques (Hubert et Bendjoudi, 1997). Aujourd'hui, la pratique montre que ces modèles ne suffisent plus à expliquer la distribution des événements d'occurrence rare.

Les approches fractales puis multifractales, empruntées à l'étude de la turbulence hydrodynamique, ont donné une nouvelle dimension à l'analyse et à l'interprétation des champs de pluies.

Dans ce mémoire, nous nous proposons d'étudier la distribution des séries pluviométriques à partir de ces formalismes fractals et multifractals, afin de déterminer les paramètres des lois qui les régissent et de proposer de nouveaux modèles statistiques de probabilité.

Nous commencerons par une étude globale des données qui permettra d'avoir une vision spatiale de la répartition des pluies. Nous aborderons ensuite la notion de multifractalité, son introduction en hydrologie et son application sur les données. Pour finir, nous proposerons une nouvelle fonction densité de probabilité et une autre méthode d'estimation du paramètre de décroissance algébrique des queues de distributions.

# I- Présentation des données

## I-1- Description de la base de données

La base de données initiale comprend 368 séries pluviométriques annuelles. Les durées d'observation minimale et maximale sont respectivement de 39 et 299 ans, avec une moyenne de 106 ans.

Les données proviennent du monde entier. Elles ont été rassemblées par l'intermédiaire de plusieurs organismes. Les données des pays de la Méditerranée proviennent de la base AMHY du projet FRIEND-AMHY du Programme Hydrologique International de l'UNESCO (Bendjoudi et Hubert, 1997). Celles de l'Afrique de l'Ouest et du centre ont été recueillies auprès de l'IRD (ex. ORSTOM) et du CIEH. Quelques données françaises proviennent de Météo-France. Le reste des données a été collecté un peu par-ci par-là, notamment grâce à la base de données GHCN Version 2 – Precipitation Beta Release du site Internet [www.ncdc.noaa.gov](http://www.ncdc.noaa.gov) de la NOAA et de l'édition 1994 du logiciel WORLD WEATHERDISC.

Pour ce travail, nous nous limitons aux séries d'une durée de 100 ans au moins, qui sont au nombre de 233. Pour ces séries, les dates de début de mesure se situent entre 1697 et 1895 et les dates de fin de chronique, entre 1893 et 1996. La durée moyenne d'observation est de 126 ans.

Pour faciliter la lecture et le traitement de la base, les séries ont été classées par ordre décroissant de nombre d'années d'observation.

Les entrées de cette base sont (tableau 1) :

- le numéro de la station
- le nom de la station
- le code ISO du pays (AFNOR, 1994)
- année de début de la série
- année de fin de la série
- durée de la série (années)
- nature de la grandeur mesurée (P=pluie)
- pas de temps de disponibilité des séries (A=annuel)
- latitude de la station (degré, minute)
- longitude de la station (degré, minute)

A chaque station, correspond un fichier d'extension ".dat" dont la racine est le nom de la station (huit premiers caractères au maximum en partant de la gauche). Ces fichiers contiennent le nom de la station suivi du nom du pays, la date de début suivie du nombre d'années d'observation et la chronique de mesures.

Les traitements ont été effectués à l'aide de programmes en fortran 77, avec souvent des sorties sous Microsoft Excel pour le tracé des graphiques.

## I-2- Analyse à l'échelle globale

Il s'agit dans cette partie de chercher à savoir s'il existe une interaction des phénomènes se produisant aux différentes stations. Soient deux pluviomètres séparés par une distance  $d$ ; Marsan et al (1996), selon leur modèle, définissent la liaison entre les signaux  $\varepsilon_1(t)$  et  $\varepsilon_2(t)$  observés en ces deux points par l'expression:

$$\langle \varepsilon_1^q(t) \varepsilon_2^q(t + \Delta t) \rangle \approx \|d, \Delta t\|^{-K(2,q)} \quad (1)$$

où  $q$  est l'ordre des moments et  $K(2,q)$  la fonction d'échelle des moments ou fonction de structure (voir chapitre 2).

On peut généraliser cette relation pour un réseau de pluviomètres en définissant la fonction de corrélation généralisée :

$$I_n^{(q_1, q_2)}(|\Delta x|, \Delta t) = \langle \varepsilon_n^{q_1}(x, t) \varepsilon_n^{q_2}(x + \Delta x, t + \Delta t) \rangle \quad (2)$$

avec  $q_1, q_2$  : ordres des moments

En considérant les moments d'ordre 1 ( $q_1 = q_2 = 1$ ), on montre (Marsan et al, 1996) que pour deux domaines purement spatial et temporel pris séparément, la fonction de corrélation se met sous la forme:

$$\text{Domaine temporel } (0, \Delta t) : I_n^{(1,1)}(\Delta t) \approx \Delta t^{-\frac{1}{1-H}K(2)} \quad (3)$$

$$\text{Domaine spatial } (\Delta x, 0) : I_n^{(1,1)}(|\Delta x|) \approx |\Delta x|^{-K(2)} \quad (4)$$

où  $H$  est un coefficient d'anisotropie espace-temps.

A l'échelle annuelle, les séries pluviométriques peuvent être considérées comme étant indépendantes les unes par rapport aux autres. L'application de la relation (3) qui vise à caractériser et quantifier la corrélation à cette échelle ne revêt donc pas une grande importance. Elle serait plus significative à plus petite échelle (minute, heure).

Intéressons-nous plutôt à la relation (4) qui ne dépend pas de l'échelle (temporelle) et qui semble très enrichissante. On considère les pluviomètres (stations) deux à deux séparés par une distance  $\Delta x$ . On détermine pour chaque couple, le domaine de définition qui n'est rien d'autre que l'intersection des années d'observation de chacune des stations du couple. Prenons un exemple: Si on prend deux stations, l'une étant observée de 1980 à 1990 et l'autre de 1985 à 1995, on travaillera, pour ce couple, sur l'intervalle allant de 1985 à 1990.

Sur ce domaine de définition, on applique la relation (2) avec  $\Delta t=0$ ,  $q_1=q_2=1$ ; Ce qui revient à faire la moyenne des produits des deux séries préalablement normalisées,

au même pas de temps et sur leur domaine commun. La distance  $\Delta x$  séparant deux stations est calculée à partir des coordonnées géographiques des deux stations (latitude, longitude) à l'aide de la trigonométrie sphérique.

On procède ainsi pour tout le réseau et on reporte sur un graphique (graphique 1) les valeurs de la fonction de corrélation en fonction des distances considérées. Il faut signaler que nous avons retenu seulement les couples dont le domaine de définition était au moins égal à 10 ans.

Avant d'interpréter ce graphique, il serait judicieux d'analyser d'abord pour les mêmes couples de stations, le coefficient de corrélation classique entre les séries des deux stations sur le domaine de recouvrement. Le graphique 2 montre la variation de ce coefficient en fonction de la distance. Pour des distances faibles, le coefficient de corrélation est élevé (autour de 0,8); il diminue avec l'augmentation de la distance. Au-delà d'une certaine distance (environ 4000 km) le nuage de points reste centré sur l'axe  $y=0$ . Compte tenu de l'opacité du nuage due au nombre important de points (26 982), nous avons cherché à affiner sa lisibilité en faisant des moyennes par intervalle de distance de 500 km (graphique 3). Sur ce graphique, on peut souligner une décroissance de la corrélation bien marquée pour des distances allant de 0 à environ 4000 km. Après cette distance, le coefficient oscille autour de 0 avec une moyenne pratiquement nulle. Cela nous permet de définir, pour une station donnée, un cercle de corrélation centré sur cette station, de rayon 4000 km, à l'intérieur duquel toute autre station sera corrélée avec la première.

Revenant au graphique 1, on retrouve cette portée de 4000 km où le nuage de points semble plus épais et plus compact; il s'effile et s'effrite au-delà de cette distance. On s'intéressera plus particulièrement à cette première partie de la fonction de corrélation afin de quantifier la relation (4). Pour cela, nous avons fait la moyenne comme précédemment, des valeurs de la fonction au pas de distance de 500 km (graphique 4). Un ajustement des points a été fait et on trouve:

$$I_n^{(1,1)}(|\Delta x|) \approx |\Delta x|^{-0,0091} \Rightarrow K(2) = 0,0091$$

Cette valeur est à comparer à celle de **0,22 ± 0,06** trouvée par Marsan (1998) pour des spectres d'énergie de données de réflectivité radar sur une période de 2,5 jours et un domaine carré de 500 km environ, avec une résolution de 2 km x 2 km toutes les 15 minutes.

Il faut noter que plus  $K(2)$  est élevé, plus la fonction de corrélation décroît rapidement et moins la distance de corrélation est grande. La valeur que nous avons trouvée se justifie donc par la grande échelle d'étude considérée (plusieurs milliers de km).

Nous avons repris la même étude, mais cette fois-ci, en distinguant les couples de stations ayant une orientation Est-Ouest (E-W) ou Ouest-Est (W-E) et ceux ayant une orientation Sud-Nord (S-N) ou Nord-Sud (N-S). On admettra qu'un couple a une orientation Est-Ouest, si l'angle entre la droite (plutôt l'arc) reliant les deux stations et l'Equateur géographique est inférieur à 45°. Il aura une orientation Sud-Nord dans le cas contraire.

Les graphiques 5, 6, 7, 8 montrent respectivement, en fonction de la distance, les variations des coefficients de corrélation E-W, S-N et des fonctions de corrélation E-W et S-N. Pour apprécier les différences, examinons plutôt les graphiques 9 et 10.



Sur le graphique 9, on retrouve toujours une première phase, pendant laquelle, les coefficients de corrélation restent élevés et décroissent rapidement avec la distance. Cette décroissance est plus forte dans la direction S-N que dans la direction E-W. On pourrait définir des portées de 2000 km pour la direction S-N et 3000 km pour la direction E-W. Au-delà de ces distances, les valeurs des coefficients de corrélation suivant les deux directions se rejoignent et vacillent autour de 0.

Ces mêmes portées se retrouvent sur le graphique 10, en ce qui concerne les fonctions de corrélation. Des régressions sur les premiers points, selon chaque direction, donnent les résultats suivants:

$$\text{direction } E - W : I_n^{(1,1)}(|\Delta x|) \approx |\Delta x|^{-0,0076} \Rightarrow K(2) = 0,0076$$

$$\text{direction } S - N : I_n^{(1,1)}(|\Delta x|) \approx |\Delta x|^{-0,0087} \Rightarrow K(2) = 0,0087$$

La valeur de  $K(2)$  dans la direction S-N est plus élevée que celle dans la direction E-W; ce qui se traduit par: pour une même distance, la corrélation est plus forte dans la direction E-W que S-N. La décroissance rapide de la fonction de corrélation dans cette dernière direction justifie sa faible portée.

Des explications plus physiques peuvent venir étayer ces analyses théoriques: Loin de vouloir aborder ici, en détails, les phénomènes complexes de la circulation atmosphérique et de la génération de pluie, nous tenons à rappeler simplement que la circulation dans l'atmosphère se fait différemment selon deux zones:

- La zone intertropicale qui s'étend entre les latitudes 30° S et 30° N, est marquée par l'existence de vastes cellules convectives (cellules de HADLEY) où l'air chaud s'élève près de l'Equateur et redescend à des latitudes plus élevées. Il existe également des courants de retour dans les basses couches de l'atmosphère vers l'Equateur. On appelle Zone de Convergence Intertropicale, la région où convergent ces courants venant de l'Hémisphère Nord et de l'Hémisphère Sud; laquelle zone varie en fonction des saisons. Les cellules ne montent pas toutes uniformément le long de l'Equateur, celles qui se déplacent longitudinalement sont appelées cellules de WALKER.
- La zone extratropicale s'étend au-delà des latitudes 30° N ou 30° S en direction des pôles. Cette zone est marquée par un climat beaucoup plus variable.

La vitesse du vent peut être définie en termes de système local de coordonnées cartésiennes inscrit sur la terre. A chaque latitude  $\varphi$  et longitude  $\lambda$  sur la terre (sphère) de rayon R, le vent admet deux composantes:

$$u = R \cos \varphi \frac{d\lambda}{dt} \quad \text{direction } E - W \text{ ou } W - E$$

$$v = R \frac{d\varphi}{dt} \quad \text{direction } N - S \text{ ou } S - N$$

D'une manière générale, les principaux régimes de vent à la surface de la terre s'établissent suivant une direction d'Est en Ouest aux basses latitudes et d'Ouest en Est aux plus hautes latitudes. Ces vents, appelés vents zonaux, peuvent atteindre jusqu'à plus de  $30 \text{ m.s}^{-1}$  dans les jets puissants subtropicaux centrés sur des latitudes d'environ  $30^\circ$  (Hartmann, 1994). La composante méridionale du vent est beaucoup moins élevée et correspond à la convergence des alizés vers l'Equateur. Elle peut atteindre une vitesse maximale de  $1 \text{ m.s}^{-1}$  (Hartmann, 1994). Cette circulation générale peut être perturbée par la naissance de phénomènes plus complexes telles que les moussons indiennes et africaines au mois de juillet qui ont des directions variables (Chapel et al, 1996).

On comprend donc mieux la forte corrélation des stations pluviométriques dans la direction E-W (ou W-E) à cause de la circulation générale des phénomènes atmosphériques dans cette direction, notamment les nuages qui sont à l'origine des pluies. Parlant de nuages, il faut rappeler qu'on en distingue principalement trois types selon l'altitude: les cirrus (nuages hauts), les cumulus (nuages moyens) et les stratus (nuages bas). Ils sont constitués de fines gouttelettes d'eau ou de particules de glace en suspension dans l'atmosphère. Les précipitations se produisent quand les blocs de nuages deviennent sursaturés en vapeur d'eau, provoquant ainsi condensation et chute de gouttes. Cette sursaturation est normalement causée par un refroidissement des nuages durant leur ascension qui peut être forcée par les mouvements atmosphériques. On estime à environ 3 jours le temps de séjour réel des molécules d'eau dans l'atmosphère (Hartmann, 1994). On peut donc, sans trop se tromper, tenir le raisonnement suivant: un paquet de nuages sursaturés, qui commence à précipiter et qui se déplace à la vitesse moyenne du vent de  $15 \text{ m.s}^{-1}$  soit  $54 \text{ km/h}$ , aura parcouru pendant 3 jours, une distance d'environ  $4000 \text{ km}$  avant de s'évanouir complètement. Cela pourrait expliquer la portée de corrélation des stations ( $4000 \text{ km}$ ) observée sur le graphique 2, en supposant que les signaux enregistrés à chaque couple de stations résultent de la réalisation d'un même événement pluvieux.

Ce premier traitement qui constitue une vision globale, permet de toucher du doigt toute la difficulté que représente l'étude des phénomènes atmosphériques tels que les champs de pluies. Il est bien entendu qu'une telle approche simpliste fait abstraction de beaucoup de facteurs influençant les précipitations, notamment l'orographie et l'anthropisation.

Dans la suite, nous nous intéresserons plus particulièrement à la distribution temporelle des pluies qui nécessite de nouvelles démarches plus robustes fondées sur des concepts physiques (naturels), à même de traduire fidèlement le comportement des perturbations atmosphériques.

## II- Multifractalité et Hydrologie

Le terme fractal qui vient du mot latin fractus désignant un objet de forme irrégulière et discontinue, a été utilisé pour la première fois par Mandelbrot pour décrire le chaos qui règne dans les phénomènes naturels (Mandelbrot, 1974). Cette notion qui était purement géométrique au départ (Mandelbrot, 1982), a permis de renouveler les approches de nombreux domaines de sciences de la nature en y introduisant explicitement la notion d'échelle.

L'extrême variabilité des champs géophysiques, qui résulte d'interactions fortement non linéaires entre différentes échelles et/ou champs, n'est pas bien traduite par les approches traditionnelles. Cette variabilité intervient sur de très grandes gammes d'échelles allant des échelles dites de dissipation (inférieures ou égales au millimètre) aux échelles planétaires (plusieurs milliers de kilomètres) en espace et de la milliseconde aux échelles géologiques en temps (Schertzer et Lovejoy, 1985b), (Schertzer et Lovejoy, 1994).

Pour aborder ces caractéristiques fondamentales, de nouveaux outils basés sur des démarches fractales ont été développés. La manière la plus simple d'aborder le problème (variabilité non linéaire) est de supposer qu'un processus élémentaire permet que cette variabilité se reproduise d'échelle en échelle ; d'où les termes de cascade multiplicative et d'invariance d'échelle.

Les champs géophysiques, compte tenu de leur complexité, ne peuvent se réduire à un simple cadre géométrique "blanc ou noir" ; c'est à dire l'occurrence ou la non-occurrence d'un phénomène caractérisé par sa seule dimension fractale  $D$  mesurant son degré d'irrégularité. Cela conduirait à négliger la distribution dans l'espace et/ou dans le temps de l'intensité des phénomènes étudiés. L'approche multifractale vise donc à une prise en compte simultanée des échelles et des intensités (Schertzer et Lovejoy, 1987a).

Les turbulences atmosphériques sont régies par les équations de Navier et Stokes et les modèles multifractals ont d'abord été conçus pour reproduire les propriétés de ces équations (Schertzer et Lovejoy, 1987b, 1991).

L'apparition des fractals puis des multifractals en hydrologie a surtout été motivée par le caractère intermittent tant dans l'espace que dans le temps des phénomènes pluvieux. Ils ont fortement contribué à l'appréhension et à la compréhension des processus complexes de générations pluviométrologiques. Les caractéristiques fractales ou multifractales des champs de pluie ont été observées et analysées dans l'espace (Lovejoy, 1981 ; Schertzer et Lovejoy, 1987a;...), dans le temps (Hubert et Carbonnel, 1989; Hubert et al, 1993 ; Lima et Grasman, 1999;...) et à la fois dans les deux directions (Marsan et al, 1996 ; Marsan, 1998).

Il a été montré (Kolmogorov, 1962 ; Mandelbrot, 1974) que les transferts d'énergie de gros tourbillons vers les petits s'effectuent lors de cascades d'une manière multiplicative et la fraction de flux transmise est déterminée par un facteur aléatoire.

Le cas le plus simple de telles cascades est la génération de cascades discrètes multiplicatives autosimilaires. Il faut entendre par autosimilaire, le fait qu'à chaque étape de la cascade, chaque tourbillon généré ressemble au gros tourbillon à une homothétie près.

Si nous considérons par exemple un domaine purement spatial bidimensionnel de taille  $L \times L$ , correspondant à la structure la plus grande d'intensité ou densité du flux d'énergie  $\varepsilon_0$ , on peut développer une cascade discrète multiplicative autosimilaire par itération d'un générateur invariant d'échelle jusqu'à une échelle de résolution fine  $l=L/\lambda$ , où  $\lambda$  est le rapport d'échelle maximum. Quand  $\lambda \rightarrow +\infty$ , on observe l'apparition de singularités, extrêmes en certains points et faibles en d'autres, correspondant à une concentration de l'activité du champ sur une portion de plus en plus faible de l'espace (Schertzer et Lovejoy, 1994).

A l'échelle  $l_1=L/\lambda_1$ , on divise le domaine en  $\lambda_1 \times \lambda_1$  structures ( $\lambda_1$  entier) et on attribue à chaque structure ainsi créée une intensité  $\varepsilon_{1,i}=\varepsilon_0 * \mu\varepsilon_{1,i}$  ( $i^{\text{ème}}$  structure du 1<sup>er</sup> pas de cascade), où  $\mu\varepsilon_{1,i}$  sont les réalisations indépendantes les unes des autres d'une variable aléatoire positive  $\mu\varepsilon$ , comme nous l'avons dit plus haut.  $\mu\varepsilon$  est caractérisé par une fonction  $K(q)$  telle que :

$$\langle \mu\varepsilon^q \rangle = \lambda_1^{K(q)} \quad (5)$$

avec la condition  $\langle \mu\varepsilon \rangle = 1 \Leftrightarrow K(1)=0$  qui résulte de la conservation en moyenne d'ensemble de l'énergie totale à chaque pas de cascade.

$K(q)$  est appelée fonction de structure ou de "scaling" des moments.

Si  $\varepsilon_n$  est le champ obtenu après  $n$  itérations, on a :

$$\langle \varepsilon_n^q \rangle = \langle \mu\varepsilon^q \rangle^n = \lambda_1^{nK(q)} = \lambda^{K(q)} \quad (6)$$

En considérant  $\varepsilon_\lambda = \varepsilon_n$  on arrive à une équation fondamentale des multifractals:

$$\langle \varepsilon_\lambda^q \rangle \approx \lambda^{K(q)} \quad (7)$$

D'une façon équivalente, en termes de probabilité, cette équation peut se réécrire (Schertzer et Lovejoy, 1991) :

$$\text{Pr ob}[\varepsilon_\lambda > \lambda^\gamma] \approx \lambda^{-C(\gamma)} \quad (8)$$

où  $C(\gamma)$  est une fonction de codimension qui caractérise la probabilité d'occurrence des singularités d'ordre supérieur à  $\gamma$ .

Les auteurs font remarquer que le signe  $\approx$  signifie égalité en englobant les facteurs constants multiplicatifs et les variations lentes en échelle (ex. logarithme).

Les deux fonctions  $K(q)$  et  $C(\gamma)$  se déduisent l'une de l'autre par une transformation de Legendre (Parisi et Frisch, 1985):

$$K(q) = \max_\gamma (q\gamma - C(\gamma)) \quad C(\gamma) = \max_q (q\gamma - K(q)) \quad (9)$$

Ce qui permet d'établir des relations univoques entre ordres des moments et des singularités:

$$q = \frac{dC(\gamma)}{d\gamma} \quad \gamma = \frac{dK(q)}{dq} \quad (10)$$

En géophysique, notamment en hydrologie, les mesures faites sont des quantités habillées, puisqu'elles sont l'intégration à l'échelle d'observation de processus qui se sont poursuivis jusqu'à des échelles infiniment plus petites. (c'est l'exemple de pluies horaires, mensuelles ou encore annuelles, qui sont l'intégration temporelle de processus se développant à des échelles beaucoup plus fines). Ces propriétés habillées sont à distinguer des propriétés nues qui correspondent à des processus arrêtés à l'échelle d'observation. Les propriétés habillées qui cachent les variabilités à plus petites échelles que l'échelle d'observation, peuvent conduire à des divergences de moments d'ordre suffisamment grand (Mandelbrot, 1974; Schertzer et Lovejoy, 1987b).

$$\langle \varepsilon_\lambda^q \rangle \rightarrow \infty \text{ quand } \lambda \rightarrow \infty \text{ pour } q \text{ suffisamment grand}$$

Ce qui correspond à une transition de phase multifractale, analogue à une transition de phase thermodynamique. Cette divergence des moments est équivalente à une chute algébrique lente de la distribution des intensités au-delà d'un certain seuil. On montre (Schertzer et Lovejoy, 1992) qu'il existe un ordre de moment critique  $q_D$ , défini par  $K(q_D) = D(q_D - 1)$ ,  $q_D > 1$ ,  $D$ : dimension de l'espace support, au-delà duquel, il y a divergence. A  $q_D$ , correspond un ordre de singularité critique  $\gamma_D$ , au-delà duquel, la fonction de codimension  $C(\gamma)$  devient linéaire:

$$C(\gamma) = q_D(\gamma - \gamma_D) + C(\gamma_D) \quad \text{si } \gamma > \gamma_D \quad (11)$$

Pour  $\gamma$  suffisamment grand, on peut donc réécrire la distribution de la probabilité au dépassement du flux d'énergie  $\varepsilon_\lambda$ :

$$\text{Pr ob}[\varepsilon_\lambda > \lambda^\gamma] \approx \lambda^{-C(\gamma)} \approx \lambda^{-q_D \gamma} \quad (12)$$

Si on définit un seuil  $\varepsilon = \lambda^\gamma$ , on a :

$$\text{Pr ob}[\varepsilon_\lambda > \varepsilon] \approx \varepsilon^{-q_D} \quad (13)$$

Cette équation, très révélatrice, montre que quelle que soit l'échelle considérée, la probabilité au dépassement d'un seuil suffisamment grand, est une fonction algébrique décroissante de ce seuil et que l'exposant  $q_D$  est indépendant de l'échelle considérée. C'est ce résultat que nous utiliserons par la suite pour caractériser les queues de distributions des séries pluviométriques annuelles de la base de données.

Cette transition de phase qualifiée de transition de phase multifractale du premier ordre survient suite à des discontinuités dans les premières dérivées des fonctions  $K(q)$  et  $C(\gamma)$ ; A cause du nombre fini de données de mesures (problèmes d'échantillonnage), il apparaît une autre transition de phase multifractale du second ordre due aux discontinuités dans les dérivées secondes de  $K(q)$  et  $C(\gamma)$ . L'ordre des moments critique correspondant est  $q_S$  (Lima et Grasman, 1999; .....).

### III- Distribution des séries annuelles

#### III-1- Limites de la loi normale

Nous n'avons nullement l'intention dans cette partie de nous étaler sur les modèles statistiques employés en hydrologie. L'accent sera surtout mis sur un modèle qui a pratiquement valeur de dogme (Hubert et Bendjoudi, 1996), à savoir la loi de Laplace-Gauss (loi normale) qui est largement utilisée pour l'ajustement des séries pluviométriques annuelles. Cette loi se définit comme suit :

Si X est une variable aléatoire qui suit une loi normale, sa fonction densité de probabilité est :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad , -\infty < x < +\infty \quad , \text{ avec } \mu : \text{moyenne de la série, } \sigma : \text{écart-type de la série.}$$

L'utilisation répandue de cette loi pour l'étude et l'analyse des cumuls pluviométriques annuels est en fait basée sur le théorème central limite qui stipule que : la somme de n variables aléatoires, indépendantes, de variance finie et identiquement distribuées, est asymptotiquement normale (Kottegoda, 1980). Lequel théorème peut se généraliser à toute somme de n variables aléatoires à condition qu'aucune variable ou groupe de variables n'occupe dans la somme une place prépondérante par rapport aux autres ou encore que toutes les variables initiales soient du même ordre de grandeur (Dagnelie, 1970).

Il se dégage deux problématiques de ce théorème à savoir les notions d'indépendance et de distribution identique. Si, à des échelles de temps assez grandes (années), l'on peut considérer les événements pluvieux comme étant relativement indépendants, est-il raisonnable d'admettre cette hypothèse à petites échelles ? La pluie qui tombe à un instant t donné, ne dépend-t-elle pas de l'activité pluvio-orageuse précédente ? Il faudrait peut-être rappeler qu'en réalité, dans la nature, la pluie n'a pas d'échelle. Elle n'est ni horaire, ni journalière, ni annuelle (Hubert et Bendjoudi, 1996). C'est un phénomène qui se déroule au cours du temps selon ses propres lois. Les échelles actuelles ne traduisent en fait que notre incapacité à cerner et retrouver les paramètres invariants d'échelle qui régissent ces lois.

Si l'on peut difficilement admettre cette hypothèse d'indépendance, il est par contre inadmissible que les épisodes pluvieux soient identiquement distribués, quand on sait qu'ils varient dans un rapport de 1 à 1000 (Bendjoudi et Hubert, 1998). Les périodes pluvieuses, quelle que soit la région, ne durent que quelques heures par an. En France, cette durée varie entre 400 et 1400 heures par an, soit environ 10% du temps (Réméniéras et Hubert, 1990). Il faut noter que cette estimation de la durée des précipitations en un lieu, très liée à l'échelle d'observation à partir de laquelle elle est calculée, est bien mal définie et finalement très critiquable.

Comme nous l'avons vu dans le chapitre précédent, les instruments de mesure (pluviomètres, pluviographes, etc...) habillent les propriétés de la pluie et masquent ainsi les hétérogénéités et irrégularités qui peuvent survenir à plus petites échelles que l'échelle d'observation et entraîner des conséquences dramatiques (Schertzer et Lovejoy, 1994). C'est le cas d'une simple averse qui peut bouleverser tout un bilan

hydrologique annuel. Considérer les épisodes pluvieux du même ordre de grandeur, reviendrait donc à ignorer la nature même du phénomène pluviogène, c'est à dire l'intermittence et la variabilité spatio-temporelle.

Utilisée souvent à tort, la pratique montre que la loi normale s'écarte des valeurs de mesures en ce qui concerne les événements extrêmes. Dans la plupart des cas, ces valeurs sont tout simplement écartées sous prétexte qu'elles relèvent d'erreurs de mesure ou d'une autre logique. Faut-il continuer à développer ou à raccommoquer des modèles statistiques douteux, desquels les événements les plus riches en information sont exclus, et basés sur des hypothèses elles-mêmes douteuses ? Ou alors, faut-il rechercher des outils plus performants, rationnels et physiques, traduisant la nature intrinsèque des observations faites ? C'est là que l'approche multifractale trouve toute sa place et toute sa justification en ce sens qu'elle permet de rendre compte en une théorie unifiée des lois d'échelles et de l'intermittence associée des champs de pluie.

### III-2- Application des multifractals à l'étude des séries annuelles

Nous reprenons ici un des résultats les plus fondamentaux des multifractals à savoir que la probabilité de dépassement, quelle que soit l'échelle considérée, d'un seuil suffisamment grand, est une fonction algébrique décroissante de ce seuil, traduit par l'équation (13) :

$$\text{Pr ob}[\varepsilon_\lambda \succ \varepsilon] \approx \varepsilon^{-qD}$$

Pour un seuil assez grand h d'une précipitation H, on aura:

$$\text{Pr ob}[H \succ h] \approx h^{-qD} \quad (14)$$

Cette approche a déjà été appliquée à une banque de données pluviométriques annuelles comprenant 71 séries d'une durée supérieure ou égale à 90 ans (Bendjoudi et Hubert, 1998). Nous l'étendons ici à notre base de données de 233 séries de durée d'observation au moins égale à 100 ans.

Afin de respecter la contrainte de conservation d'ensemble de la "masse" totale, et de pouvoir comparer les séries les unes par rapport aux autres, celles-ci ont été normalisées (moyenne=1) et standardisées (écart-type=1). Si  $(Y_i)_{i=1,n}$  est une série d'observations de moyenne m et d'écart-type  $\sigma$ , la série normalisée et standardisée

$$(X_i)_{i=1,n} \text{ est telle que: } X_i = 1 + \frac{Y_i - m}{\sigma}, \quad i=1,n.$$

Les séries ont été rangées par ordre croissant et pour chaque valeur  $X_i$ , nous avons calculé la probabilité empirique au dépassement correspondante par la formule de

$$\text{WEIBULL: } \text{Pr ob}(X \succ X_i) = \frac{n+1-i}{n+1}, \quad i \text{ étant le rang et } n \text{ le nombre d'observations.}$$

Les points représentatifs de ces probabilités empiriques au dépassement ont été reportés sur un graphique log-log. A ces points, nous avons superposé la courbe correspondant à la probabilité au dépassement issue de l'ajustement des données à une loi normale (graphique 11).

Pour toutes les séries, l'ajustement des données empiriques à la loi normale est satisfaisant... sauf pour les valeurs extrêmes. On note en effet pour toutes les séries et cela d'une manière systématique, que la loi normale décroche pour des seuils grands, donc des probabilités au dépassement faibles. Cette cassure se produit dans l'ensemble pour des probabilités au dépassement inférieures ou égale à 0,05 (période de retour  $\geq 20$  ans). Au-delà de cette probabilité, la loi normale n'est plus valable puisque la loi des grands nombres ne s'applique plus, d'où perte d'ergodicité, divergence des estimateurs statistiques usuels et dépendance des estimations de la taille de l'échantillon (Bendjoudi et Hubert, 1998): C'est la transition de phase multifractale du premier ordre. Elle correspond à une chute algébrique de la distribution des intensités, infiniment moins rapide qu'une décroissance exponentielle. Ce qui n'était qu'une conjecture théorique, se voit bien vérifié par les données empiriques. En effet, au-delà de la cassure, les points empiriques s'organisent autour d'une droite. C'est en fait la traduction de l'équation (14) en logarithme:

$$\text{Log}(\text{Pr ob}[H > h]) \approx -q_D \text{Log}(h) \quad -q_D \text{ étant la pente de la droite.}$$

Des études précédentes (Hubert et Bendjoudi, 1996) effectuées sur la série de Dédougou (Burkina Faso), aux échelles de temps annuelle, mensuelle et journalière, montrent une décroissance algébrique de pentes similaires des queues de distribution pour ces trois échelles. Ce qui milite en faveur d'une invariance d'échelle du paramètre  $q_D$  (graphique 12).

### III-3- Estimation de l'exposant de décroissance algébrique $q_D$

Partant donc de ce constat, nous avons ajusté une droite, pour toutes les séries, sur les points dont la probabilité empirique au dépassement est inférieure à 5%, afin d'estimer  $q_D$ . Le tableau 2 regroupe les résultats; il comprend la liste des stations, le code du pays, la durée de la série, la valeur de  $q_D$ , le coefficient de corrélation de l'ajustement et le nombre de points utilisés. Les coefficients de corrélation sont satisfaisants (moyenne = 0,96). Le nombre de points utilisés varie de 15 pour les plus longues séries à 5 pour les plus courtes avec une moyenne de 6 points. En fait ces points représentent 5% du nombre total d'observation; puisque:

$$\text{Pr ob}(X > X_i) = \frac{n+1-i}{n+1} \text{ par la formule de WEIBULL}$$

$\text{Pr ob}(X > X_i) \leq 0,05 \Leftrightarrow \frac{n+1-i}{n+1} \leq 0,05 \Leftrightarrow i \geq 0,95(n+1)$  (les séries étant classées par ordre croissant).

Les valeurs de  $q_D$  s'échelonnent de 2 à 26. La moyenne sur toutes les séries est de 5,9; la dispersion autour de cette valeur centrale est 3,7. Il est intéressant de voir comment évoluent les différentes valeurs de  $q_D$  d'une station à une autre. Le graphique 13 montre cette évolution en fonction du nombre d'années d'observation des séries. Les points ne sont pas répartis au hasard. On peut en effet remarquer une espèce de noyau délimité par les intervalles [100;140] pour les durées et [2;7]



pour les valeurs de  $q_D$ , où la densité des points est plus élevée. Au-delà, les points sont épars. La dégradation horizontale de ce noyau peut s'expliquer par le nombre de moins en moins élevé de séries de plus de 150 ans. Mais la faible densité de points au-dessus du noyau (verticalement) pourrait traduire l'existence d'une valeur de convergence de  $q_D$  au sein du noyau. Une petite statistique sur ce noyau donne une valeur moyenne de  $q_D$  de 4,4 et un écart-type de 1,3.

Entre 100 et 140 ans, on note une forte variation verticale des valeurs de  $q_D$  (2 à 26). Très rapidement, cette variation s'atténue lorsque la durée des séries augmente. Les points semblent encadrés par deux courbes enveloppes virtuelles minorante et majorante qui convergent vers une même limite lorsque le nombre d'années d'observation devient grand. Pour essayer d'approcher cette valeur "idéale" de  $q_D$  qui serait la limite d'une série infinie, nous avons utilisé les 6 plus longues séries (de plus de 200 ans). En nous fondant sur la moyenne et l'écart-type de ces séries, nous pouvons estimer la valeur limite de  $q_D$  à **4,5 ± 1**.

Cette valeur est proche de celle trouvée pour le noyau d'attraction.

Il y a lieu de relever que cette estimation de l'exposant de décroissance  $q_D$  est légèrement supérieure à celle faite dans les travaux précédents (Bendjoudi et Hubert, 1998) qui est de **3,8 ± 0,5**.

La recherche dans la littérature de nouvelles séries de plus en plus longues, s'avère indispensable pour affiner davantage la détermination de ce paramètre.

Nous avons appliqué d'autres formules empiriques de calcul de la probabilité au dépassement à savoir:

$$\text{Cunnane: } Prob(X \succ X_i) = \frac{n + 0.6 - i}{n + 0.2}$$

$$\text{Hazen: } Prob(X \succ X_i) = \frac{n + 0.5 - i}{n}$$

où  $i$  est le rang du seuil et  $n$  le nombre d'observations.

Toutes ces formules posent le même constat que celle de Weibull: le décrochage de la loi normale des données empiriques pour des probabilités inférieures ou égales à 0,05 et la chute algébrique des queues de distribution. Elles donnent cependant des estimations un peu plus élevées du paramètre  $q_D$  (**5,2 ± 1,2 pour Cunnane et 5,5 ± 1,2 pour Hazen**).

Pour mettre en évidence la divergence des moments statistiques du fait de la transition de phase multifractale, nous avons calculé les moments d'ordre 1 à 6. Le graphique 14 montre la variation de ces moments pour la station de Gibraltar (Espagne) en fonction du nombre d'années utilisé  $N$ , allant de 1 à la durée de la série (202 ans). On distingue essentiellement deux parties: Une première partie entre 1 et 70 ans, où les moments fluctuent beaucoup en raison de la taille assez faible des échantillons. Les amplitudes de ces fluctuations sont moindres pour les moments d'ordre 1 à 3 comparativement aux moments d'ordre 4 à 6.

Intéressons-nous à la deuxième partie se situant au-delà de 70 ans. Dans cette partie, les trois premiers moments se stabilisent rapidement et deviennent

constants. Alors que les moments d'ordre 4, 5 et 6 ont du mal à se stabiliser et ont tendance à diminuer quand N augmente. Cette instabilité des moments d'ordre supérieur ou égal à 4 traduit bien le fait qu'un phénomène s'est produit: c'est la transition de phase multifractale. L'ordre critique des moments se situe donc autour de 4, ce qui consolide l'estimation faite plus haut. Le moment d'ordre 4 n'existant pas, des paramètres statistiques tel que le coefficient d'aplatissement n'ont aucune signification.

Les conséquences pratiques d'une décroissance algébrique des queues de distribution des séries sont assez lourdes, surtout en ce qui concerne les risques hydrologiques et les coûts des ouvrages de génie civil. Quand on multiplie un seuil d'intensité par 10, on divise sa probabilité au dépassement par  $10^{q_D}$ , on multiplie donc sa période retour par  $10^{q_D}$ . En adoptant la valeur de  $q_D$  de 4,5, on est très loin des résultats donnés par une distribution normale qui divise par 10 la probabilité au dépassement d'un seuil 10 fois plus grand. En outre, un événement qualifié de millennal par une loi normale, ne serait en fait que centennal dans le cas de notre modèle.

Le graphique 15 permet de mieux visualiser ces différences. Nous y avons représenté la période de retour T de l'intensité maximale de chaque série, calculée par la loi normale, en fonction de la durée n de la série qui correspond à la période de retour empirique de la plus grande intensité. Pour des raisons de commodité, l'axe des ordonnées est en échelle logarithmique.

Hormis quelques rares points, la quasi-totalité des points (plus de 90%) se situe au-dessus de la première bissectrice. Le nuage de points n'est aucunement allongé le long de cette droite; il est plutôt orienté verticalement. Alors que la période retour empirique varie entre 100 et 299 ans, avec une moyenne de 126 ans, celle théorique varie de 38 ans à 4,3 millions d'années avec une moyenne de 57 000 ans. Pour des événements qui sont dans l'ensemble centennaux, 40% sont estimés plus que millennaux par la loi gaussienne.

Comme nous pouvons donc le constater, la loi normale fait une forte surestimation des périodes de retour des événements extrêmes. Ce qui a une incidence sur les coûts des ouvrages hydrauliques.

### **III-4- Essai de proposition d'une fonction densité de probabilité**

L'introduction des approches multifractales dans l'étude de la distribution des pluies annuelles ouvre de nouvelles perspectives à la modélisation spatio-temporelle des champs de précipitations. Au lieu de peaufiner des modèles statistiques basés sur des lois à décroissance exponentielle trop rapide, il est peut-être grand temps que les mathématiciens statisticiens œuvrent à la recherche de nouvelles lois présentant un comportement "fat tail" ou "heavy tail" (décroissance lente); c'est à dire des lois de probabilité à décroissance algébrique.

Nous citons pour revue:

- La loi de Pareto généralisée (Van Montfort et Witter, 1986):

$$P(X \leq x) = \begin{cases} 1 - (1 - \theta Z)^{1/\theta} & 0 < Z < \infty, \quad \theta < 0 \\ 0 < Z < 1/\theta, & \theta > 0 \\ 1 - \exp(-Z) & 0 < Z < \infty, \quad \theta = 0 \end{cases} \quad (15)$$

où  $Z = \frac{x - \nu}{\sigma}$

avec

$\nu$  : paramètre de position ( $x > \nu$ )

$\sigma$  : paramètre d'échelle ( $\sigma > 0$ )

$\theta$  : paramètre de forme.

- La loi de Halphen B-1 (Morlat, 1956):

Fonction densité de probabilité:

$$f_{B-1}(x) = \frac{2}{m^{-2\nu} ef_{\nu}(\alpha)} x^{-2\nu-1} \exp\left[-\left(\frac{m}{x}\right)^2 + \alpha\left(\frac{m}{x}\right)\right] \quad \text{pour } x > 0 \quad (16)$$

$\nu > 0$  et  $\alpha$  sont des paramètres de forme

$m > 0$  est un paramètre d'échelle

$ef_{\nu}$  est la fonction factorielle exponentielle définie comme suit:

$$ef_{\nu}(\alpha) = \int_0^{\infty} x^{2\nu-1} e^{(-x^2 + \alpha x)} dx$$

Dans cette même foulée, nous essayons ici de proposer une fonction densité de probabilité qui présente un comportement algébrique.

Soit la fonction:  $h(x) = \frac{x^a}{(m + x^b)^c} \quad \forall x \geq 0, \quad \text{avec } a > -1 ; b > 0 ; m > 0 ; c > (a+1)/b$

L'intégrale de cette fonction vaut:

$$\int_0^{+\infty} h(x) dx = \int_0^{+\infty} \frac{x^a}{(m + x^b)^c} dx = m^{\frac{a+1-bc}{b}} \frac{\Gamma\left(\frac{a+1}{b}\right) \Gamma\left(c - \frac{a+1}{b}\right)}{b \Gamma(c)} \quad (\text{Lide D.R., 1992})$$

Définissons une fonction  $f$  telle que:  $f(x) = \frac{h(x)}{\int_0^{+\infty} h(x) dx} \quad \forall x \geq 0$

On aura:

$$f(x) = \frac{b\Gamma(c)}{m^{\frac{a+1-bc}{b}} \Gamma\left(\frac{a+1}{b}\right) \Gamma\left(c - \frac{a+1}{b}\right)} * \frac{x^a}{(m+x^b)^c} \quad \forall x \geq 0 \quad (17)$$

Avec  $a > -1$  ;  $b > 0$  ;  $m > 0$  ;  $c > (a+1)/b$

$f$  est bien une fonction densité de probabilité puisqu'elle est continue, positive et dérivable sur  $]0; +\infty[$  et  $\int_0^{+\infty} f(x)dx = 1$ .

Elle présente un comportement algébrique puisque, quand  $x \rightarrow +\infty$ ,  $f(x) \approx x^{a-bc}$  ( $a-bc < 0$  et  $bc-a > 1$ ). Dans ce cas, la probabilité au dépassement vaut:

$$\text{Pr ob}[X > x] = \int_x^{+\infty} f(x)dx \approx \int_x^{+\infty} x^{a-bc} dx \approx x^{-(bc-a-1)}$$

L'exposant de décroissance algébrique est  $q_D = bc - a - 1$ .

Dans le cadre de ce travail, pour simplifier l'étude de la fonction, nous adopterons **b=1**.

On a donc:

$$f(x) = \frac{\Gamma(c)}{m^{a+1-c} \Gamma(a+1) \Gamma(c-a-1)} * \frac{x^a}{(m+x)^c}$$

En mettant  $m$  en facteur, on obtient :

$$f(x) = \frac{\Gamma(c)}{m\Gamma(a+1)\Gamma(c-a-1)} * \frac{\left(\frac{x}{m}\right)^a}{\left(1+\frac{x}{m}\right)^c} \quad (18)$$

$m$  représente le paramètre d'échelle,  $a$  et  $c$  les paramètres de forme.

### III-4-1- Etude et tracé de la fonction

Désignons par  $X$ , la variable aléatoire dont la fonction densité de probabilité (fdp) est  $f$ . Soit  $g$  la fdp d'une variable aléatoire  $U$  tel que:  $U=X/m$   
La probabilité élémentaire se définit comme suit:

$$\text{Prob}(x < X < x+dx) = f(x)dx = g(u)du$$

$$g(u) = f(x) \frac{dx}{du}$$

$$\Rightarrow \quad \Rightarrow \quad \mathbf{g(u) = mf(x)}$$

$$du = \frac{dx}{m}$$

$$g(u) = \frac{\Gamma(c)}{\Gamma(a+1)\Gamma(c-a-1)} * \frac{u^a}{(1+u)^c} \quad (19)$$

La courbe ( $C_g$ ) de la fonction est semblable à celle de  $f$  ( $C_f$ ) à un facteur d'échelle près.

La fonction  $g(u)$  peut se mettre sous la forme:

$$g(u) = \frac{1}{B(a+1, c-a-1)} * \frac{u^a}{(1+u)^c}$$

où B est la fonction bêta définie par:  $B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$

$g$  est dérivable sur  $]0 ; +\infty[$  et sa dérivée vaut :

$$g'(u) = \frac{1}{B(a+1, c-a-1)} * \frac{u^{a-1}}{(1+u)^{c+1}} * [a + (a-c)u]$$

On distingue plusieurs cas suivant les valeurs de  $a$ :

- **-1 < a < 0**

$u$  étant toujours positif, on a:  $u > \frac{a}{c-a} \Rightarrow g'(u) < 0$ :  $g$  est strictement décroissante.

$$\lim_{u \rightarrow 0} g(u) = +\infty$$

$$\lim_{u \rightarrow +\infty} g(u) = \lim_{u \rightarrow +\infty} (mu)^{a-c} = 0 \quad \text{car } a-c < 0$$

Le tableau de variation se présente comme suit:

$u$	0	$+\infty$
$g'$	-	
$g$	$+\infty$	0

- **a=0**

$$g'(u) = \frac{1}{B(c, a+1)} * \frac{-c}{(1+u)^{c+1}} < 0: g \text{ est strictement décroissante.}$$

$$g'(0) = c(1-c) \quad \text{et} \quad g(0) = (c-1)$$

Tableau de variation:

u	0		$+\infty$
g'	$c(1-c)$	-	
g	$c-1$		

• **a>0**

Le signe de  $g'(u)$  dépend du signe de  $a+(a-c)u$ .

$$u = \frac{a}{c-a} \Rightarrow g'(u) = 0 : \text{ la courbe de } g \text{ admet une tangente horizontale en}$$

$$M\left(\frac{a}{c-a}, g\left(\frac{a}{c-a}\right)\right)$$

$$u < \frac{a}{c-a} \Rightarrow g'(u) > 0 : g \text{ est strictement croissante.}$$

$$u > \frac{a}{c-a} \Rightarrow g'(u) < 0 : g \text{ est strictement décroissante.}$$

$$g\left(\frac{a}{c-a}\right) = \frac{1}{B(a+1, c-a-1)} * \frac{a^a}{c^c} (c-a)^{c-a} \quad \text{et} \quad g(0)=0$$

Tableau de variation:

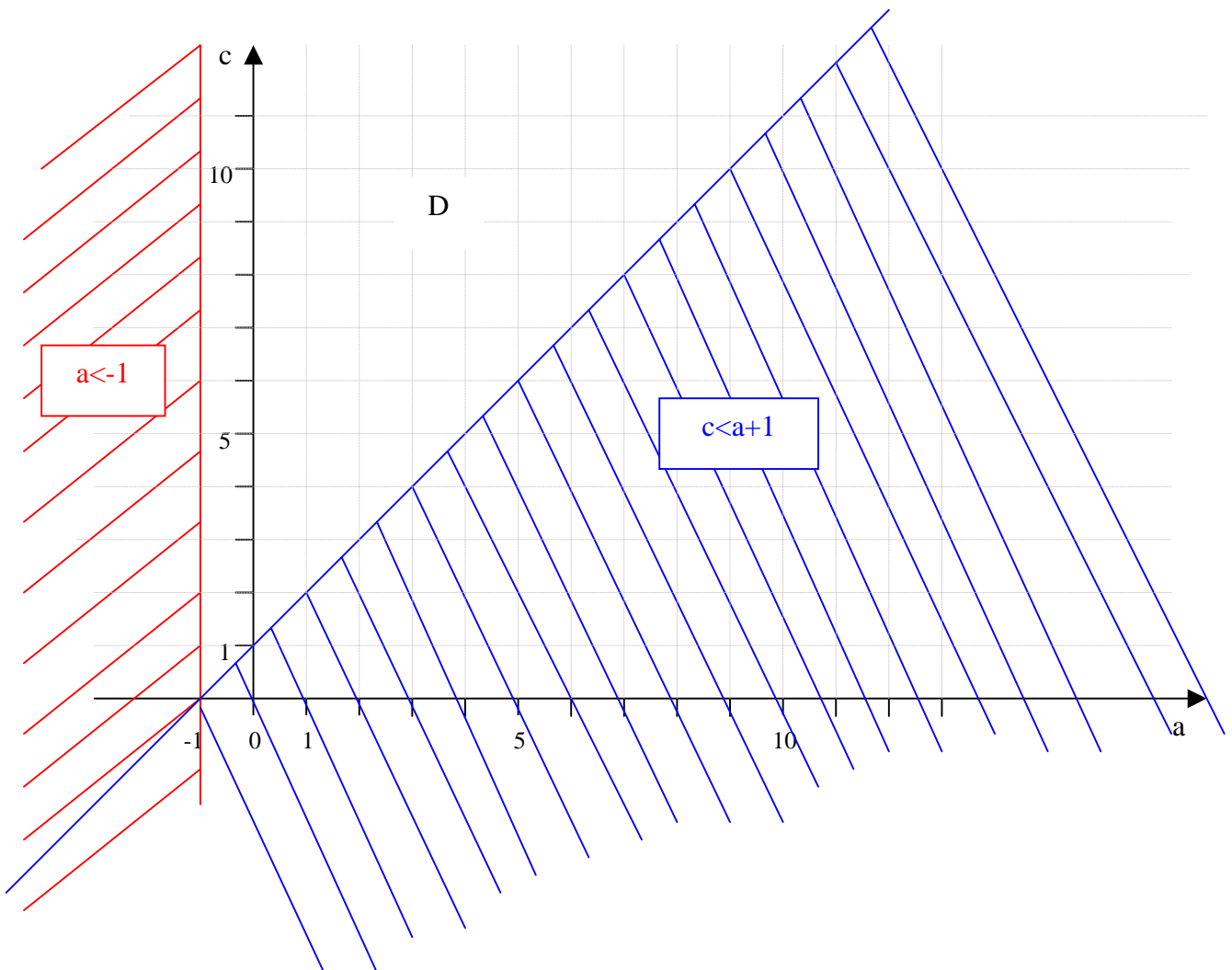
u	0	$\frac{a}{c-a}$	$+\infty$
g'		+	0
g		-	
g	0		

Il faudra remarquer que dans l'intervalle  $]0;1]$ ,  $g'(u) \rightarrow +\infty$  quand  $u \rightarrow 0$ , la courbe de  $g$  ( $C_g$ ) admet donc une tangente verticale à l'origine. Alors que dans l'intervalle  $]1; +\infty[$ ,  $g'(0)=0$ : ( $C_g$ ) admet une tangente horizontale à l'origine. Quelle que soit la valeur du paramètre  $a$ ,  $\lim_{u \rightarrow +\infty} g(u) = 0$ : ( $C_g$ ) admet l'axe des abscisses comme asymptote à l'infini.

*Choix des paramètres  $a$  et  $c$*

$a > -1$  et  $c > a+1$

Si nous représentons sur un graphe ces deux conditions (inégalités), nous obtenons un domaine  $D$  de combinaisons acceptables des paramètres  $a$  et  $c$ .



Le graphique 16 montre les allures de la fonction  $g$  pour quelques valeurs de  $a$  et  $c$ .

### III-4-2- Paramètres statistiques

#### a) Mode

Le mode est la valeur de la variable pour laquelle la fonction densité de probabilité atteint son maximum. Il s'obtient en faisant  $g'(u) = 0$ .

Pour  $-1 < a \leq 0$ , le mode n'existe pas, puisque  $g'(u) < 0$  sur l'intervalle  $[0; +\infty[$ .

Pour  $a > 0$ :

$$g'(u) = 0 \Leftrightarrow \left. \begin{array}{l} u = \frac{a}{c-a} \\ u = \frac{x}{m} \end{array} \right\} \Rightarrow x = \frac{am}{c-a}$$

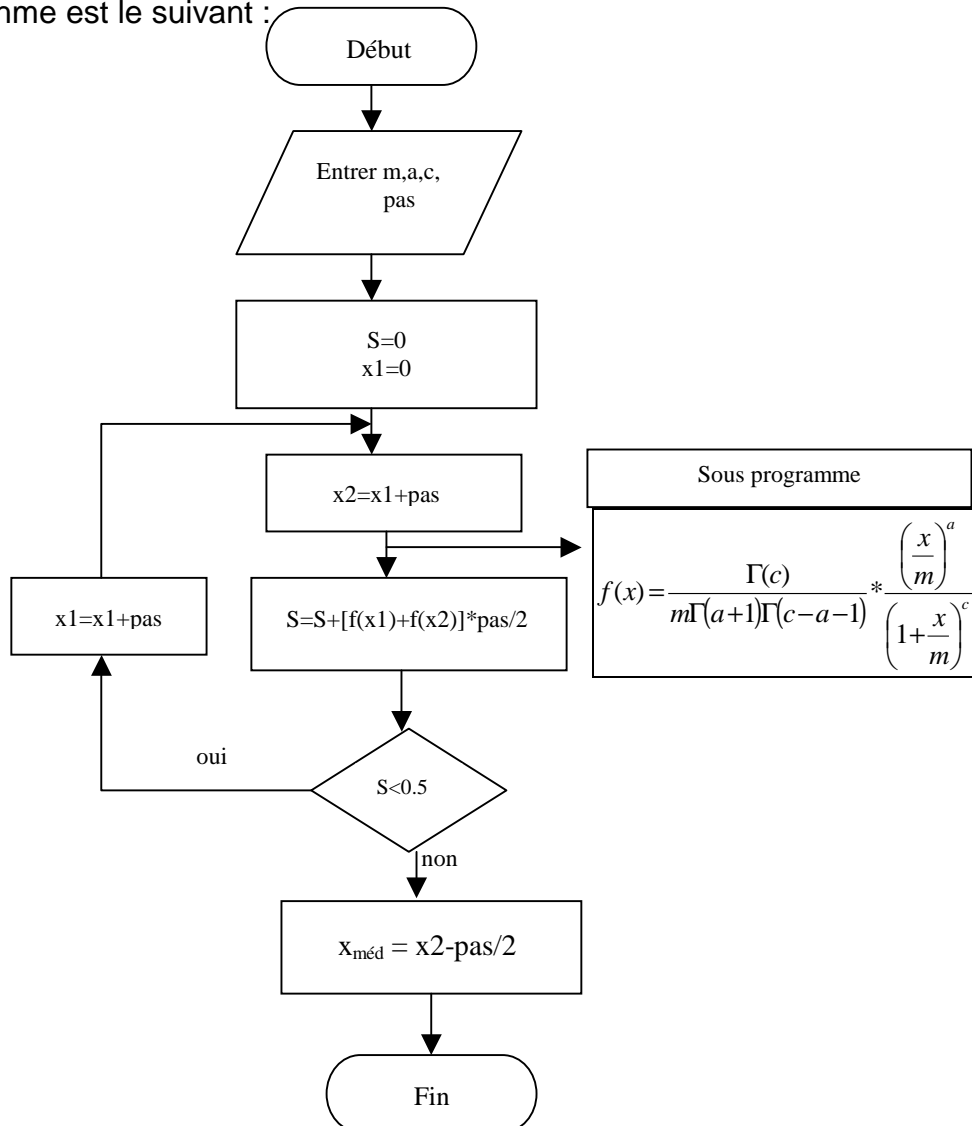
le mode vaut :  $\boxed{\text{mode} = \frac{am}{c-a}} \quad (20)$

### b) Médiane

Elle est définie comme étant la valeur de la variable de part et d'autre de laquelle, on retrouve 50% de la distribution. Si on note  $x_{\text{méd}}$  la médiane, on doit avoir :

$$\int_0^{x_{\text{méd}}} f(x)dx = \int_{x_{\text{méd}}}^{+\infty} f(x)dx = 0,5 \quad (21)$$

Cette équation étant difficile, voire impossible à résoudre littéralement compte tenu de l'intégrale de la fonction  $f$  qui est très complexe, nous proposons une résolution numérique basée sur une discrétisation trapézoïdale de l'intégrale dont l'algorithme est le suivant :





### c) Moments

Le moment d'ordre p se met sous la forme :

$$E[X^p] = \int_0^{+\infty} x^p f(x) dx = m^p \frac{\Gamma(a+p+1)\Gamma(c-a-p-1)}{\Gamma(a+1)\Gamma(c-a-1)} \quad (22)$$

On vérifie bien que, pour  $p=0$ ,  $E[X^0] = \int_0^{+\infty} f(x) dx = 1$

Les conditions d'existence de ces moments sont:  $a+p > -1$ ,  $c > a+p+1$ , et  $m > 0$ . Ce qui donne pour l'ordre des moments:  $-(a+1) < p < c-a-1 \Leftrightarrow -(a+1) < p < q_D$ . Il existe donc un ordre maximum (ordre critique  $q_D$ ) au de-là duquel les moments ne sont pas définis.

L'expression de quelques moments particuliers sont données dans la suite:

- **Moment d'ordre 1 (p=1)**

Il correspond à la moyenne théorique.

$$E[X^1] = \int_0^{+\infty} x f(x) dx = m \frac{\Gamma(a+2)\Gamma(c-a-2)}{\Gamma(a+1)\Gamma(c-a-1)}$$

En utilisant la propriété :  $\Gamma(n+1) = n\Gamma(n)$ , on obtient :

$$\boxed{E[X] = \frac{m(a+1)}{c-a-2}} \quad (23)$$

- **Moment d'ordre 2 (p=2)**

$$E[X^2] = \int_0^{+\infty} x^2 f(x) dx = m^2 \frac{\Gamma(a+3)\Gamma(c-a-3)}{\Gamma(a+1)\Gamma(c-a-1)}$$

$$\boxed{E[X^2] = \frac{m^2(a+1)(a+2)}{(c-a-2)(c-a-3)}} \quad (24)$$

On peut également calculer le moment centré d'ordre 2 qui correspond à la variance :

$$Var(X) = \int_0^{+\infty} (x - E[X])^2 f(x) dx = E[X^2] - [E[X]]^2$$

$$\boxed{Var(X) = \frac{m^2(a+1)(c-1)}{(c-a-2)^2(c-a-3)}} \quad (25)$$

### III-4-3- Estimation des paramètres de la loi théorique à partir des données

Cette estimation s'est faite à l'aide des méthodes traditionnelles:

- *Méthode du maximum de vraisemblance:*

C'est une méthode optimale, tout au moins asymptotique, car elle prend en compte tous les éléments de l'échantillon individuellement.

En considérant la fonction  $f(x,a,c,m)$  et la série d'observations indépendantes  $(X_i)_{i=1,n}$ , on définit la vraisemblance par:

$$L=f(x_1,a,c,m)*f(x_2,a,c,m)*\dots\dots\dots * f(x_i,a,c,m)*\dots\dots f(x_n,a,c,m)$$

Ou encore:

$$\text{Log}(L) = \sum_{i=1}^n \text{Log}[f(x_i, a, c, m)]$$

La méthode consiste à maximiser la fonction L ou Log(L) par rapport aux paramètres a,c,m, en annulant les dérivées partielles:

$$\frac{\partial L}{\partial a} = 0 \quad \frac{\partial L}{\partial c} = 0 \quad \frac{\partial L}{\partial m} = 0$$

On obtient ainsi autant d'équations que de paramètres qui permettent d'estimer ces derniers.

- *Méthode du maximum d'entropie:*

On définit la fonction entropie relative à un système  $X[(x_i, p_i)]_{i=1,n}$  par:

$$H(X) = -\sum_{i=1}^n p_i \text{Log}(p_i) \quad \text{où } p_i \text{ étant la probabilité associée à } x_i.$$

Pour un système continu de fonction densité de probabilité  $f$  définie sur un domaine D, l'entropie vaut:

$$H(X) = -\int_D f(x) \text{Log}[f(x)] dx$$

Il s'agit là également de maximiser la fonction H sous certaines contraintes (exemple: relations établissant la moyenne ou l'écart-type de la série).

- *Méthode des moments:*

Elle consiste à remplacer dans les relations entre moments théoriques et paramètres, les moments théoriques par les moments empiriques calculés à partir de l'échantillon. On utilise autant de moments qu'il y a de paramètres et on obtient ainsi une estimation de ces derniers.

L'application des deux premières méthodes n'est pas aisée compte tenu des complexités mathématiques engendrées par les dérivées partielles qui ne permettent pas d'expliciter les paramètres. Quant à la méthode des moments, elle s'applique facilement, mais les résultats ne sont pas satisfaisants ( $c < 0$  ou  $m < 0$ ). Cela peut être dû au fait que les moments utilisés ne soient pas assez significatifs pour un échantillon; notamment les moments d'ordre -1 et 3. Ce dernier est assez instable car étant situé dans la zone d'influence du moment critique  $q_D$ .

Ne disposant pas de ressources informatiques sophistiquées, nous avons dû, pour une première approche, simplifier davantage la fdp en adoptant  $m=1$ .

Dans ce cas, nous avons pu estimer les paramètres  $a$  et  $c$  à partir des moments d'ordre 1 (moyenne) et d'ordre 2 centré (variance).

Si on appelle  $e_1$  et  $e_2$  respectivement la moyenne et la variance déterminées à partir de l'échantillon, on a les relations suivantes:

$$\begin{cases} E[X] = \frac{(a+1)}{c-a-2} = e_1 \\ Var(X) = \frac{(a+1)(c-1)}{(c-a-2)^2(c-a-3)} = e_2 \end{cases}$$

Après résolution, on trouve :

$$a = \frac{e_1^2 \left[ \frac{(1+e_1)^2}{e_2} + 1 \right] - 1}{1+e_1} \quad (26)$$

$$c = \frac{e_1}{e_2} (1+e_1)^2 + e_1 + 2 \quad (27)$$

Ces paramètres ont été estimés pour chaque série. Afin de valider l'estimation, nous avons calculé pour toutes les séries, le mode et la médiane théoriques à partir de la loi  $f$ . Les graphiques 17 et 18 montrent les résultats des comparaisons avec le mode et la médiane empiriques des séries. Comme nous pouvons le constater, les nuages de points s'allongent très bien le long de la première bissectrice. Le coefficient de corrélation entre mode estimé et mode empirique est de 0,983; celui entre médianes estimée et empirique est 0,999. Ce qui témoigne de l'excellence des estimations.

Revenons maintenant à l'estimation d'exposant de décroissance algébrique  $q_D$ . Pour une série normalisée et standardisée, les équations (26) et (27) deviennent:  $a=2$  et  $c=7$ . On calcule  $q_D=c-a-1 \Rightarrow q_D = 4$ .

Cette valeur qualifiée de théorique est proche de celle estimée empiriquement qui est de  $4,5 \pm 1$ .

La probabilité au dépassement théorique d'un seuil  $x$  d'une variable aléatoire  $X$  se calcule par:

$$Pr ob[X > x] = \int_x^{+\infty} f(x)dx = 1 - \int_0^x f(x)dx .$$

Nous avons repris le graphique 11 en y ajoutant cette probabilité théorique (voir graphique 19). Si nous nous intéressons aux événements extrêmes qui sont les plus importants, contrairement à la loi normale, la loi  $f$  décroît lentement et colle assez

bien avec les données empiriques. Dans cette partie, l'ajustement est satisfaisant. Par contre, pour des probabilités au dépassement supérieures à 0,05, la loi  $f$  semble décrocher un peu des données empiriques par rapport à la loi normale. On distingue essentiellement pour cette partie, deux zones. L'une comprise entre les probabilités 0,05 et 0,6, où la loi  $f$  se trouve en dessous des valeurs empiriques. L'autre allant des probabilités 0,6 à 1, où la loi  $f$  se situe au-dessus des données empiriques. Toutefois, on peut remarquer que les écarts entre la loi théorique  $f$  et la loi empirique (Weibull) ne sont pas assez grands. Notons également que les valeurs de probabilité de 1, données par  $f$  pour des faibles valeurs de  $x$ , sont dues, comme nous l'avons vu plus haut, à la forme de démarrage de la courbe de  $f$  qui admet une tangente horizontale à l'origine pour  $a > 1$ .

S'il y a aujourd'hui un fait qu'on ne peut plus nier, c'est la distribution de probabilité des intensités pour des seuils élevés, suivant une loi à décroissance algébrique caractérisée par un exposant  $q_D$  qui serait invariant d'échelle et universel. Cela ne doit pas cependant nous faire oublier toute la complexité que représentent les phénomènes naturels telle que la pluie. La mise au point de modèles à base physique, reproduisant fidèlement le comportement de tels phénomènes n'est certainement pas chose facile. Cette fonction  $f$ , qui n'est qu'à sa première étude, montre toute la puissance que représentent de telles lois par rapport aux lois classiques à décroissance exponentielle beaucoup plus rapide. Il y a donc lieu d'affiner l'étude en explorant d'autres horizons ( $b \neq 1, m \neq 1$  par exemple).

## IV- Essai d'une autre méthode d'estimation de l'exposant de décroissance algébrique $q_D$

Les méthodes empiriques (Weibull, Cunnane, Hazen) permettent toutes de poser les mêmes constats, à savoir une décroissance algébrique des queues de distribution et l'existence d'un exposant  $q_D$  qui, selon toute vraisemblance, serait invariant d'échelle et universel. Mais les estimations de  $q_D$  par ces différentes méthodes restent assez disparates. En outre, ces méthodes empiriques sont limitées car, si elles calculent les probabilités au dépassement pour deux seuils  $s_1$  et  $s_2$  consécutifs, on n'a aucune information par contre sur ce qui se passe entre  $s_1$  et  $s_2$  ( $]s_1; s_2[$ ).

C'est donc pour combler ces lacunes, que nous nous proposons de tester une nouvelle approche d'estimation de  $q_D$ , basée sur le comportement algébrique et l'irrégularité des séries pluviométriques. Cette approche fait actuellement l'objet d'un travail intensif de recherche dans le cadre de la Thèse de Keltoum CHAOUCHÉ (Laboratoire GRESE, ENGREF).

### IV-1- Approche théorique (Keltoum CHAOUCHÉ, à paraître)

Considérons une variable aléatoire  $X$  avec les réalisations  $(X_i)_{i=1,n}$  à pas de temps  $\Delta$  qui correspondent à une suite d'intensités.

Soit un seuil  $s$  variant de 0 à  $\max_{i=1,n}(X_i)$ .

Soit  $T$  une variable aléatoire de réalisations  $\tau_i$  ( $i=1,n$ ) telles que:

$$\begin{cases} \tau_i = 1 & \text{si } X_i > s \\ \tau_i = 0 & \text{si } X_i \leq s \end{cases} \quad (i=1 \dots n)$$

On définit par:

$$d(s) = \Delta \sum_{i=1}^n \tau_i : \text{la durée de dépassement du seuil } s$$

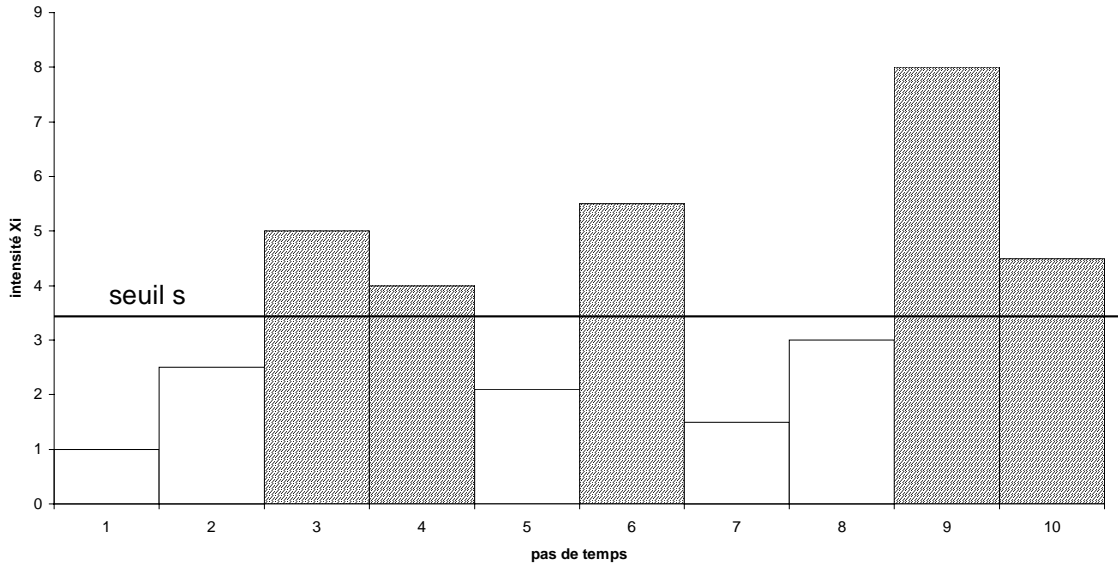
$$p(s) = \Delta \sum_{i=1}^n X_i \tau_i : \text{le cumul des intensités dépassant le seuil } s$$

$$D = n\Delta : \text{la durée totale des observations}$$

$$P = \Delta \sum_{i=1}^n X_i : \text{le cumul total des intensités}$$

Sur le graphe ci-dessous,  $d(s)$  et  $p(s)$  correspondent respectivement aux sommes des largeurs et des aires des rectangles hachurés.

définition de d(s) et p(s)



On reporte sur un graphique, pour tout seuil  $s$ , les points d'abscisse  $x(s)$  et d'ordonnée  $y(s)$  définies par:

$$x(s) = \log \left[ \frac{d(s)}{D} \right] = \log \left[ \frac{\Delta}{n\Delta} \sum_{i=1}^n \tau_i \right] = \log \left[ \frac{1}{n} \sum_{i=1}^n \tau_i \right] \quad (28)$$

$$y(s) = \log \left[ \frac{p(s)}{P} \right] = \log \left[ \frac{1}{n} \sum_{i=1}^n X_i \tau_i \right] - \log \left[ \frac{1}{n} \sum_{i=1}^n X_i \right] \quad (29)$$

soient  $(U_i)_{i=1,n}$  et  $(V_i)_{i=1,n}$  deux suites de variables aléatoires telles que:

$$\begin{aligned} U_i &= \tau_i \\ V_i &= X_i \tau_i \end{aligned} \quad i=1, \dots, n$$

On peut admettre (loi des grands nombres) que  $\frac{1}{n} \sum_{i=1}^n X_i$ ,  $\frac{1}{n} \sum_{i=1}^n U_i$ ,  $\frac{1}{n} \sum_{i=1}^n V_i$  convergent respectivement vers  $E[X]$ ,  $E[U]$  et  $E[V]$ .

On a donc:

$$x(s) = \log E[U] \quad (30)$$

$$y(s) = \log E[V] - \log E[X] \quad (31)$$

On définit pour une variable aléatoire  $X$  de densité de probabilité  $f$ , la probabilité au dépassement  $G$  par:

$$G : R \rightarrow R$$

$$G(x) = P(X > x) = \int_x^{+\infty} f(t) dt$$

En se plaçant dans le cas d'une décroissance algébrique des queues de distribution, la fonction G se met sous la forme générale (Keltoum CHAOUICHE, à paraître):

$$\forall x \geq 1 \quad G(x) = r(x).x^{-q} \quad \text{avec } q > 1 \quad (31)$$

r vérifie:

- $\lim_{x \rightarrow +\infty} r(x) = K$  et r est croissante sur  $[1, +\infty[$  (33)

- $\forall x \geq 1 \quad r'(x).x \leq Ax^{-\theta-1}$  avec  $A > 0$  et  $\theta > 0$  (34)

On appelle fonction de vitesse de dépassement de G (Keltoum CHAOUICHE, à paraître), la fonction h définie par:

$$\forall s \geq 1 \quad h(s) = \frac{sG(s)}{\int_s^{+\infty} G} \quad (35)$$

On montre (Keltoum CHAOUICHE, à paraître) que pour un seuil s positif:

$$E[U] = G(s) \quad (36)$$

$$E[V] = sG(s) + \int_s^{+\infty} G \quad (37)$$

On calcule la pente a(s) en tout point (x(s),y(s)):  $a(s) = \frac{\frac{\partial y(s)}{\partial x(s)}}{\frac{\partial x(s)}{\partial s}}$

$$\Rightarrow a(s) = \frac{G(s) + s \frac{\partial G(s)}{\partial s} + \frac{\partial}{\partial s} \int_s^{+\infty} G}{sG(s) + \int_s^{+\infty} G} * \frac{G(s)}{\frac{\partial G(s)}{\partial s}}$$

Cette pente s'exprime en fonction de la vitesse de dépassement h:

$$\boxed{a(s) = \frac{1}{1 + \frac{1}{h(s)}}} \quad (38)$$

Après une intégration par partie, on a:

$$\int_x^{+\infty} G = \frac{r(x)x^{1-q}}{q-1} + \frac{1}{q-1} \int_x^{+\infty} r'(t)t^{1-q} dt$$

$$\frac{1}{h(x)} = \frac{\int_x^{+\infty} G}{xG(x)} = \frac{1}{q-1} [1 + \varepsilon(x)] \quad \text{avec} \quad \varepsilon(x) = \frac{x^{q-1}}{r(x)} \int_x^{+\infty} r'(t)t^{1-q} dt$$

la relation (34) permet d'écrire:

$$\varepsilon(x) \leq \frac{Ax^{-(2q+\theta+1)}}{(\theta+q)r(x)}$$

Cherchons la limite de la pente  $a(s)$  lorsque le seuil  $s$  tend vers l'infini:

$$\lim_{s \rightarrow +\infty} a(s) = \frac{1}{1 + \lim_{s \rightarrow +\infty} \frac{1}{h(s)}}$$

$$\lim_{s \rightarrow +\infty} \frac{1}{h(s)} = \frac{1}{q-1} \left[ 1 + \lim_{s \rightarrow +\infty} \varepsilon(s) \right] = \frac{1}{q-1} \quad \text{car} \quad \lim_{s \rightarrow +\infty} \varepsilon(s) = 0$$

On a donc:

$$\boxed{\lim_{s \rightarrow +\infty} a(s) = \frac{q-1}{q}} \quad (39)$$

Pour un seuil assez grand, la pente devrait donc tendre vers  $\frac{q-1}{q}$

Cette pente, dans le cas d'une loi algébrique est bien différente de 1; alors qu'on montre (*Keltoum CHAOUICHE, à paraître*) que pour une loi normale ou exponentielle elle vaut 1.

Pour plus de détails sur les démonstrations, se référer à:

*Keltoum CHAOUICHE, 1999, à paraître. Estimation graphique du paramètre de décroissance algébrique. Laboratoire GRESE, ENGREF, Paris, France.*



## IV-2- Application aux données

Nous avons appliqué l'approche théorique à notre banque de données de 233 séries pluviométriques annuelles. Pour chacune d'entre elle, nous avons reporté sur un graphique en échelle log-log, pour un seuil variant de 0 à l'intensité maximale, le cumul des intensités dépassant ce seuil normalisé à l'intensité totale, en fonction de la durée de dépassement de ce seuil normalisée par la durée totale.

Sur le graphique 20, on remarque que les points sont quasiment tous alignés. En nous intéressant aux événements extrêmes, nous avons fait un ajustement linéaire sur les trois derniers points pour toutes les séries. Les coefficients de détermination sont tous égaux à 1.

Les valeurs des pentes des droites de régression sont consignées dans le tableau 3. On note que les pentes sont toutes différentes de 1, ce qui témoigne d'un comportement algébrique des séries. Nous avons cherché à savoir s'il existe une relation entre ces pentes et la durée des séries. Le graphique 21 montre cette relation. La répartition des points est comparable à celle trouvée sur le graphique 13 pour les valeurs de  $q_D$ . Dans l'intervalle de temps [100,150], les points sont assez dispersés verticalement. Plus la durée de la série augmente, moins ces dispersions sont importantes. Les points semblent enveloppés dans deux courbes inférieure et supérieure qui convergeraient vers une même limite quand la durée d'observation tend vers l'infini. Pour approcher cette limite, nous avons considéré la moyenne et l'écart-type des six plus longues séries (de plus de 200 ans). Nous l'estimons à **0,928 ± 0,025**.

A partir de la relation (39), on peut déterminer la valeur de l'exposant de décroissance algébrique  $q_D$ . En prenant la valeur moyenne de 0,928 pour la pente, on trouve  **$q_D=13,89$** .

Il est bien entendu que cette valeur de  $q_D$  reste élevée par rapport à celles trouvées précédemment au chapitre 3. On peut expliquer cet écart, d'une part, par l'application de la loi des grands nombres et le passage au logarithme qui introduisent un biais non négligeable. Il faudrait d'autre part, se pencher sur l'influence que pourrait avoir la dimension  $D$  de l'occurrence des pluies et les biais qu'elle peut introduire dans nos estimations (nous avons admis ici que la pluie se développait sur l'ensemble de la dimension temporelle, c'est à dire un ensemble de dimension  $D=1$ ). Il y a donc lieu de chercher à quantifier tous ces biais afin d'apporter une correction à la valeur de  $q_D$  trouvée.

## Conclusion

Les phénomènes naturels telles que les précipitations, n'ont pas fini de nous enseigner, à cause de leur complexité due à leur extrême variabilité. Il reste certainement encore beaucoup à dire et à faire sur l'étude de la distribution des pluies annuelles. Dans ce rapport, nous n'avons peut-être analysé qu'une petite partie de ce qui pourrait être demain tout un "iceberg" risquant de bouleverser les connaissances hydrologiques actuelles. Durant ce travail, nous nous sommes rendus compte de toute la difficulté qu'il y a à modéliser les queues de distribution. Cette étude devrait donc être poursuivie sur des séries de plus en plus longues qui sont celles qui permettent de mieux mettre à l'épreuve les modèles statistiques, car on y retrouve en principe davantage d'événements rares dont la probabilité d'apparition est faible (Hubert et Bendjoudi, 1996). Les conséquences attachées à ces événements rares sont très lourdes (crues, inondations, dimensionnement et coûts des ouvrages) d'où tout l'intérêt à les maîtriser afin de mieux les estimer.

Les approches fractales et multifractales ouvrent une nouvelle ère à la modélisation espace-temps des champs de pluies. Elles permettent de mieux rendre compte de la nature intermittente spatio-temporelle des précipitations. Contrairement aux lois classiques à décroissance exponentielle beaucoup trop rapide, les démarches multifractales mettent en évidence une décroissance algébrique des queues de distribution avec un exposant  $q_D$  qui serait invariant d'échelle et universel. Les implications d'un tel résultat sont nombreuses et considérables tant sur le plan scientifique, économique que social.

Pour terminer, nous osons espérer que ces nouveaux outils (fractals et multifractals) ont de beaux jours devant eux et que les chercheurs s'investiront de plus en plus dans cette voie en vue d'une meilleure connaissance et compréhension des phénomènes naturels.

## Références bibliographiques

- ABRAMOWITZ M., STEGUN A. Irène, 1970. Handbook of mathematical functions, with formulas, graphs and mathematical tables, Dover Publications, Inc., New York, 1045 p.
- AFNOR, 1994. Codes pour la représentation des noms de pays; norme NF EN 23166, ISO 3166.
- BENDJOUDI H., 1983. Application du concept d'entropie dans les sciences de l'eau, Thèse de Doctorat, ENSMP, UPMC Paris 6, 175 p. + annexes.
- BENDJOUDI H. et HUBERT P., 1997. Longues séries disponibles dans la base AMHY-FRIEND Flow Regime from International Experimental and Network Data, Projets H-5-5 (IHP IV) and 1.1 (IHP V) Third report: 1994-1997, CEMAGREF éditions, pp. 362-366.
- BENDJOUDI H., HUBERT P., SCHERTZER D., LOVEJOY S., 1997. Interprétation multifractale des courbes intensité-durée-fréquence des précipitations, CR Acad. Sci. Paris, Science de la vie, 1997, 325, pp.323-326.
- BENDJOUDI H., HUBERT P., 1998. A propos de la distribution statistique des cumuls pluviométriques annuels: Faut-il en finir avec la normalité? Revue des sciences de l'eau, 4, pp. 627-640.
- BERNHEIM M., 1991. FORTRAN: Mode d'emploi, InterEditions, Paris, 220 p.
- BOBEE B., 1996. Eléments de Statistiques, Polycopié de cours DEA HHGG, ENSMP, UPMC Paris VI, 56 p.
- CHAPEL A., FIEUX M., JACQUES G., JACQUES J.-M., LAVAL K., LEGRAND M., LE TREUT H., 1996. Océans et Atmosphère - Science de l'Univers, collection Synapses, Hachette Education, Paris, 160 p.
- DANIEL-JEAN D., 1984. Le Langage FORTRAN, Editests-Paris (France), 141 p.
- DENIS-PAPIN M., 1959. Mathématiques générales, Tome I, Dunod, Paris, 213 p.+annexes.
- DE LIMA M. I. P., GRASMAN J., 1999. Multifractal analysis of 15-min and daily rainfall from a semi-arid region in Portugal, Journal of Hydrology, 220, pp.1-11.
- FOUERE Pascale, 1990. Essais d'estimation et de caractérisation fractale des précipitations sahéliennes, Mémoire DEA national d'Hydrologie, ENSMP, UPMC Paris VI, 36 p.
- HAAN C. T., 1977. Statistical methods in hydrology, The Iowa State University Press, Ames, 378 p.
- HARTMANN L. D., 1994. Global Physical Climatology, Academic Press, New York, 409 p.
- HUBERT P. et CARBONNEL J. P., 1988. Caractérisation fractale de la variabilité et de l'anisotropie des précipitations tropicales. CR Acad. Sci. Paris, vol. 2, 307, pp. 909-914.
- HUBERT P. et CARBONNEL J. P., 1989. Dimension fractale de l'occurrence de pluie en climat Soudano-Sahélien. Hydro. Continentale, 4, pp. 3-10.
- HUBERT P., TESSIER Y., LOVEJOY S., SCHERTZER D., SCHMITT F., LADOY P., CARBONNEL J. P., VIOLETTE S., 1993. Multifractals and extreme rainfall events, Geophysical Research Letters, Vol. 20, n° 10, pp. 931-934.
- HUBERT P., TOMA C. A., SCHERTZER D., LOVEJOY S., SCHMITT F., LADOY P., 1994. Introduction aux Fractals et Multifractals, Cahier GANG-France publié par URA-CNRS 1367, 59 p.
- HUBERT P., FRIGGIT F., CARBONNEL J. P., 1995. Multifractal structure of Rainfall Occurrence in West Africa, Z.W. Kundzewicz [Ed.], New Uncertainty Concepts in Hydrology and Water Ressources, Cambridge University Press, pp.109-113.
- HUBERT P. et BENDJOUDI H., 1996. Introduction à l'étude des longues séries pluviométriques, XIIème journées hydrologiques de l'ORSTOM, Montpellier.
- HUBERT P. et BENDJOUDI H., 1997. Caractéristiques fractales des séries pluviométriques, FRIEND Flow Regime from International Experimental and Network Data, Projets H-5-5 (IHP IV) and 1.1 (IHP V) Third report: 1994-1997, CEMAGREF éditions, pp. 352-356.

- HUBERT P., BENDJOUDI H., SCHERTZER D., 1998. Indépendance d'échelle de la distribution statistique des précipitations extrêmes, International Hydrological Program (IHP-V), Technical Document in Hydrology, n° 20, Unesco, Paris.
- KOTTEGODA N. T., 1980. Stochastic Water Ressources Technology, The MacMillan Press Ltd, 384 p.
- LECOINTE P. et TRIEU L., 1987. Le formulaire: Théorèmes-Théories et Formules, Ed. Pierre Bordas et Fils, Paris, 370 p.
- LIDE D. R., 1992. Handbook of Chemistry and Physics, 73 rd edition, CRC Press, pp. A58-A60.
- MAIDMENT D. R., 1992. Handbook of Hydrology, McGRAW-HILL, Inc, USA, pp.18.1-18.66.
- MANDELBROT B.B., 1974. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, J.Fluid Mech., 62, pp. 331-350.
- MANDELBROT B.B., 1975. Les objets fractals, forme, hasard et dimension, Paris, 192 p.
- MANDELBROT B.B., 1982. The fractal geometry of nature, freeman, San Francisco, 460 p.
- MARSAN D., SCHERTZER D., LOVEJOY S., 1996. Causal Space-Time Multifractal Processes: Predictability and forecasting of rain fields, Journal of Geophys. Research, 101, D21, pp. 26,333-26,346.
- MARSAN D., 1998. Multifractals espace-temps, dynamique et prédictibilité; Application aux précipitations. Thèse de Doctorat Physique Théorique, UPMC Paris VI, 181 p.
- METCALF M. et REID J., 1993. FORTRAN 90: Les concepts fondamentaux, AFNOR Paris (France), 376 p.
- MORLAT G., 1956. Les lois de probabilité de Halphen, Revue de statistique appliquée, 4, pp. 21-43.
- NDAM NGOUPAYOU J. R., 1989. Distribution des pluies sur dix postes pluviométriques d'Afrique de l'Ouest (Burkina Faso), Mémoire DEA National d'Hydrologie, UPMC Paris VI, 113 p.
- PRESS W. H., FLANNERY B. P., TENKOSLSKY S. A., VETTERLING W. T., 1990. Numerical Recipes, the art of scientific computing (FORTRAN Version), Cambridge University Press, 874 p.
- REMENIERAS G. et HUBERT P., 1990. Article Hydrologique de l'Encyclopædia Universalis, Paris, volume XI de l'édition 1990, pp. 796-806.
- SCHERTZER D. et LOVEJOY S., 1985. Generalized scale invariance in turbulent phenomena, Phs. Chem. Hydrodyn. J., 6, pp. 623-635.
- SCHERTZER D. et LOVEJOY S., 1987a. Physically based rain and cloud modeling by anisotropic scaling of multiplicative processes, Journal of Geophysical Research D, 92, pp. 9693-9714.
- SCHERTZER D. et LOVEJOY S., 1987b. Singularités anisotropes et divergences des moments en turbulence, Ann. Sc. Math. du Québec, 11, pp. 139-181.
- SCHERTZER D. et LOVEJOY S., 1991. Scaling nonlinear variability in geodynamics: Multiple singularities, observables, universality classes; Non-linear variability in Geophysics, Scaling and fractals, D. Schertzer and S. Lovejoy eds, Kluwer, pp. 41-82.
- SCHERTZER D. et LOVEJOY S., 1992. Hard and soft multifractal processes, Physical Ann., 195, pp. 187-194.
- SCHERTZER D. et LOVEJOY S., 1993. Nonlinear Variability in Geophysics: Scaling and Multifractal Processes, Lecture Notes. Institut d'études scientifiques de Cargèse.
- SIDNEY I. RESNICK, 1997. Heavy tail modeling and teletraffic data, The Annals of Statistics, vol. 25, n°5, pp. 1805-1869.
- STROHMEIER A., 1982. FORTRAN 77: Approche systématique illustrée d'exemples, Edition Eyrolles, Paris, 140 p.
- VAN MONFORT M. A. J. et WITTER J. V., 1986. The Generalized Pareto distribution applied to rainfall depths, Hydrological Journal, Journal des Sciences Hydrologiques, 31, pp. 151-162.

# Annexes

[Tableau 1: Présentation de la base de données](#)

[Tableau 2: liste des stations avec les valeurs de  \$q\_D\$ , le coefficient de corrélation et le nombre de points utilisés](#)

[Tableau 3: Liste des stations avec les valeurs des pentes](#)

**Graphique 1:** Variation de la fonction de corrélation

**Graphique 2:** Variation du coefficient de corrélation

**Graphique 3:** Variation du coefficient de corrélation moyen

**Graphique 4:** Variation de la fonction de corrélation moyenne

**Graphique 5:** Variation du coefficient de corrélation E-W

**Graphique 6:** Variation du coefficient de corrélation N-S

**Graphique 7:** Variation de la fonction de corrélation E-W

**Graphique 8:** Variation de la fonction de corrélation N-S

**Graphique 9:** Comparaison des coefficients de corrélation

**Graphique 10:** Comparaison des fonctions de corrélation

**Graphique 11:** Probabilité empirique et ajustement à une loi normale

**Graphique 12:** Série pluviométrique de Dédougou: Invariance d'échelle

[Graphique 13: Variation de  \$q\_D\$  en fonction de la durée des séries](#)

**Graphique 14:** Evolution des moments d'ordre 1 à 6 en fonction du nombre N d'années utilisé ( Gibraltar )

[Graphique 15: Comparaison des périodes de retour empirique n et théorique T \(loi normale\)](#)

**Graphique 16:** Tracé de la fonction densité de probabilité g

[Graphique 17: Comparaison des modes empirique et estimé](#)

[Graphique 18: Comparaison des médianes empirique et estimée](#)

**Graphique 19:** Probabilité empirique et ajustements loi normale et loi f proposée

[Graphique 20: Autre méthode d'estimation de  \$q\_D\$ : Ajustement sur les 3 derniers points](#)

[Graphique 21: Variation de la pente en fonction de la durée des séries.](#)