

Chapitre 2

Champs de pluie et modélisation stochastique

Parmi les principaux objectifs de l'hydrologie figurent la génération des données synthétiques (par exemple pour tester la résistance d'un ouvrage face à divers scénarii) et la prévision de données futures (pluie ou débit). S'il veut exploiter le plus largement possible l'information dont il dispose et ne pas se contenter d'une analyse descriptive des données, l'hydrologue doit au préalable élaborer un modèle. En général, un modèle comporte deux volets : déterministe ou stochastique. Parfois le volet déterministe est absent et le modèle est purement stochastique, d'autres fois le volet stochastique est absent et le modèle est purement déterministe (ou mécaniste). Mais ces deux volets peuvent aussi se compléter : dans le modèle linéaire par exemple le volet stochastique est le réceptacle de l'erreur.

Les modèles purement stochastiques (ou boîtes noires) ne sont d'aucune utilité pour la modélisation des champs de pluie où l'on possède des connaissances a priori sur la causalité du phénomène (voir par exemple la hiérarchie de structures établie par Austin et Houze, 72 [3]). Les modèles purement déterministes quant à eux, sont surtout développés en météorologie et tentent de reproduire la dynamique des champs de pluie. Comme les équations exactes régissant le phénomène sont inconnues, ces modèles reposent sur des équations aux dérivées partielles de Navier-Stokes qui sont tronquées et approximées pour décrire la dynamique de l'atmosphère, puis intégrées numériquement (sous hypothèse d'homogénéité à chaque échelle). Malgré ces simplifications bien souvent éloignées de la réalité, les équations résultant de ces modèles restent complexes et les calculs numériques très lourds. De plus, les échelles sont étudiées indépendamment les unes des autres.

Les modèles à la fois stochastiques et déterministes rencontrés dans la littérature sont habituellement temporels et construits autour d'une seule échelle (de type série chronologique). Or les liens connus entre réalisation du phénomène à différentes échelles rendent inadéquats ce genre de modèles. Les modèles spatiaux et spatio-temporels intègrent plusieurs échelles en se basant généralement sur la hiérarchie des structures de champs de pluie de Austin et Houze [3]. Mais il n'est physiquement pas justifié qu'on doive privilégier certaines échelles dites caractéristiques. Dans un effort d'intégration du maximum d'information, le modélisateur se devra de mettre en jeu une gamme d'échelle aussi large que possible.

Dans cette partie, on présente brièvement l'état des connaissances que l'on possède a priori sur le phénomène pluvieux, les difficultés à le mesurer ainsi que des outils stochastiques pertinents pour la modélisation de ce dernier. Enfin, on présentera les modèles fractals et multifractals (Mandelbrot, 75 [59], Schertzer et Lovejoy, 93 [76]) qui ont connu un développement récent. Ils offrent un cadre formel prenant directement en compte une forme d'invariance d'échelle du phénomène, permettant ainsi une réduction du nombre de paramètres et une extrapolation de l'information entre les échelles.

2.1 Les champs de pluie

Avant de décrire les procédés de mesure de pluie, nous allons brièvement rappeler l'état des connaissances sur la structure des champs de pluie et les gammes d'échelle sur lesquelles il est possible de mesurer ce phénomène. Un tableau fournira ensuite les échelles pour lesquelles ces techniques de mesure sont susceptibles de conduire à des estimations fiables. Nous terminerons par quelques remarques générales sur la pertinence des hypothèses statistiques de continuité et de stationnarité des séries pluvieuses.

2.1.1 Mécanismes générateurs des champs de pluie

Les principaux types de précipitations

On distingue principalement trois types de précipitations qui peuvent survenir ensemble et se compléter :

- Précipitations liées au passage d'une perturbation frontale : Elles sont dues aux rencontres de masses d'air de températures différentes. La trace au sol de la frontière entre ces deux masses d'air est appelée front. Les fronts chauds (provoqués par un déplacement d'air chaud au dessus d'air froid dans la même direction) occasionnent des pluies de faible intensité et de longue durée. Les fronts froids (provoqués par l'avancée d'une masse d'air froid qui soulève de l'air chaud) sont marqués par de forts pics d'intensité.
- Précipitations liées à une convection locale : Un réchauffement des basses couches de l'atmosphère crée un équilibre thermique instable. L'ascendance de l'air chaud donne naissance à des cellules convectives (zone intertropicale et zone tempérée en été).
- Précipitations orographiques : Elle sont dues à la présence d'une barrière montagneuse qui, sous l'effet du vent, provoque l'élévation de masses d'air humide. Il en résulte un refroidissement de ces masses d'air. Il peut alors y avoir un équilibre thermique instable (instabilité convective) ou bien une instabilité en air saturé.

Signalons dès maintenant que selon le type de précipitation, les hypothèses d'un modèle de pluie seront différentes : l'hypothèse d'indépendance des épisodes pluvieux successifs semble plus plausible dans le cas de précipitations convectives que dans le cas de précipitations frontales.

Echelle	Spatiale	Temporelle
synoptique	$>10\ 000\ \text{km}^2$	Plusieurs jours
Grande méso-échelle	1 000 à 10 000 km^2	Plusieurs heures
Petite méso-échelle	100 à 400 km^2	Environ une heure
Cellule de pluie	Environ 10 km^2	1 min à 1/2 heure

Tableau 2.1: Organisation des champs de pluie

Hiérarchie des structures de pluie

A partir d'un suivi radar d'épisode en Nouvelle Angleterre, Austin et Houze ([3]) établissent une hiérarchie des champs précipitants en situation extra-tropicale (tableau 2.1). Ils distinguent des structures imbriquées les unes aux autres. Chaque structure est intrinsèquement liée à une échelle spatio-temporelle (comportant quatre niveaux) et possède des caractéristiques propres (durée de vie de la structure, intensité, taille).

Les deux dernières structures concernent les orages violents (par leur forte intensité qui peut atteindre jusqu'à 100 mm/h). Les cellules de pluie apparaissent isolées ou bien de façon organisée à l'intérieur d'un front.

La représentation météorologique de Austin et Houze permet donc, malgré un chaos apparent, de mettre en évidence une organisation des champs de pluie. L'inconvénient de cette analyse est qu'elle déconnecte les échelles les unes des autres, faisant apparaître artificiellement un carcan d'échelles caractéristiques. Or la pluie est un phénomène qui se développe sur une gamme d'échelles très étendue et ses grandeurs n'ont pas d'échelle caractéristique : de 1 km^2 à 10 000 km^2 en espace (averse, front pluvieux, ...), et de quelques minutes aux échelles géologiques en temps (voir tableau 2.1). Il n'est donc pas envisageable de privilégier une échelle plutôt qu'une autre, et tout effort de modélisation se doit d'établir une description unifiée des champs de pluie.

2.1.2 Mesure de la pluie

Les données de précipitations ont longtemps été produites par des mesures directes (pluviographiques ou pluviométriques). L'apparition récente des mesures indirectes (radars météorologiques et systèmes d'observation satellites) a permis de décrire la variabilité spatiale de la pluie qui jusque là n'avait été accessible que par des réseaux denses de pluviomètres ou de pluviographes.

- Mesure pluviométrique : Ce sont les mesures le plus couramment utilisées. L'eau de pluie est recueillie sur une surface de quelques cm^2 et accumulée dans un réservoir. Son volume sur un pas de temps fixe (en général 24 heures, mais parfois de l'ordre de la saison) est quantifié par divers procédés (mesure de la hauteur d'eau dans le réservoir, mais aussi pesée du réservoir). La résolution temporelle est donc égale au pas de temps de mesure et la résolution spatiale est de quelques cm^2 , mais ces mesures sont assimilées à des mesures ponctuelles en pratique. Les erreurs sont principalement

dues au vent, à des effets d'écran, à l'évaporation et parfois liées à l'obstruction ou au débordement de l'appareil.

- Mesure pluviographique : C'est une mesure un peu moins simple que la précédente. Le pluviographe classique enregistre continûment les cumuls de pluie dans le temps, mais il existe aussi des pluviographes à dépouillement manuel. Le pluviographe à basculement enregistre des dates de basculement, chaque basculement correspondant au remplissage d'un auget d'une contenance de 0.1 mm ou de 0.5 mm. La résolution spatiale est là aussi considérée comme ponctuelle, tandis que la résolution temporelle varie de 1 à 15 minutes. Les sources d'erreurs sont identiques aux précédentes mais elles peuvent aussi être dues à la saturation de l'auget qui impose une borne supérieure aux mesures.
- Mesure par spectro-pluviomètre optique : Son principe est basé sur l'atténuation produite par les gouttes de pluie lors de leur passage dans un faisceau lumineux. Le disdromètre, ou compteur de gouttes, donne aussi accès à la granulométrie des gouttes par des procédés optiques ou acoustiques
- Mesures radar : Le radar est un dispositif de mesure actif : Il émet un rayonnement incident pour mesurer la réflectivité radar Z . Celle-ci est ensuite convertie en intensité de pluie R par la relation dite de Marshall Palmer $Z = A \times R^b$, où A et b sont des coefficients calculés par une calibration utilisant des données au sol issues d'un réseau de pluviomètres. Cette relation est fondée sur des hypothèses très fortes telles qu'une répartition homogène des gouttes, une forme de goutte sphérique et une intensité de pluie constante (Andrieux 86 [2], Blanchet 93 [12], Bourel 94 [14]).

Plusieurs erreurs découlent de la relation $R-Z$. La relation de Marshall Palmer est non linéaire, ce qui implique que les coefficients de calage A et b varient par changement d'échelle. On recense dans la littérature plus d'une centaine de couples (A, b) .

Le phénomène dit de bande brillante constitue aussi une limite à la validité de la relation de Marshall Palmer. Il se produit au dessus de l'isotherme 0° , et est dû au fait que la réflectivité des particules de glace est beaucoup plus grande que celle des particules d'eau.

Ajoutons d'autres sources d'erreurs qui peuvent être dues à la mesure de réflectivité (présence d'obstacles, de masques, dépôts sur l'antenne...), ou aux écarts entre la pluie au sol et la pluie en altitude.

Malgré la multiplicité des sources d'erreurs, le radar permet néanmoins d'établir une cartographie "qualitative" de la pluie. Il tend à devenir de plus en plus opérationnel pour la mesure de la pluie.

- Mesures par satellite : Les satellites sont généralement des dispositifs passifs. Ils captent des radiations provenant de la terre essentiellement selon deux gammes d'onde :
 - L'infrarouge : L'analyse d'une image IR permet de distinguer les pixels froids des autres. Leur nombre peut être relié à une quantité d'eau précipitée en se basant sur le fait que les cellules convectives intenses sont plus froides que les

	Résolution temporelle	Résolution spatiale	Phénomènes observés
SPOT P	entre 2.5 et 26 jours	10 m	Statiques et localisés
IRS-1C	entre 5 et 24 jours	5.8 m	Statiques et localisés
Landsat TM	16 jours	30 m	Statiques et localisés
NOAA	4 fois par jour	1 km	Inondations, sécheresses

Tableau 2.2: Résolution temporelle et spatiale de quelques satellites.

sommets des nuages avoisinant (à haute altitude). Dans le cas de cellules convectives relativement isolées, ces relations sont bien explicitées. Par contre, dans le cas de nuages stratiformes ou de forçage orographique, ces relations sont quasi-inexistantes (Bourel 94 [14]). Des résultats intéressants sont obtenus par exemple au Sahel par le satellite Météosat¹. La résolution spatiale s'étend de 5 à 10 km de côté et la résolution temporelle est de l'ordre de la demi-heure.

- Les micro-ondes : C'est un système actif puisqu'il consiste à embarquer un radar sur un satellite à orbite polaire. Ce système sert à sonder les nuages. La résolution spatiale est de l'ordre de 20 km et la résolution temporelle est de 12 heures.
- Citons aussi le visible qui sert à la classification des nuages.

Les résolutions en espace et en temps des satellites sont cependant trop grandes pour les besoins de l'hydrologie : la fréquence minimale d'acquisition d'images reste supérieure à 30 minutes tandis que la résolution spatiale avoisine au mieux les 5 km de côté (voir tableau 2.2). Mais cette technique de mesure est en constante amélioration, puisque les satellites de la nouvelle génération atteindront une taille de pixel de l'ordre de quelques mètres (le satellite SPOT 5 en 2002 prévoit une taille de pixel de 3 mètres). D'autre part, les passages des satellites sont peu nombreux et s'effectuent toujours à la même heure.

2.1.3 Echelles spatio-temporelles

Toutes les échelles spatio-temporelles interviennent dans le processus des précipitations. Par exemple, la constitution d'un modèle météorologique (pour la prévision d'un débit de pointe par exemple) associe plusieurs niveaux d'échelle : de 10^4 à 10^2 km² pour la représentation de la cyclogénèse, et les mesures à des échelles inférieures pour la désagrégation à l'échelle du bassin versant. Mais les résultats obtenus à une certaine échelle pourraient être faussés par des mécanismes d'échelle inférieure, ou s'extrapoleraient mal pour expliquer des résultats à une échelle inférieure. Selon quel pas de temps et quel pas d'échelle spatiale est-il alors raisonnable de calculer des estimations de lames d'eau ? Ces gammes d'échelles concernent-elles les besoins pratiques de l'hydrologie ?

¹Au Sahel, l'essentiel des précipitations provient des cumulo-nimbus dont le sommet est froid (environ -40°). Cependant, il existe des nuages à sommet froid non précipitant et tous les cumulo-nimbus ne précipitent pas. Il est donc nécessaire d'effectuer un callage par des mesures au sol.

Bassin versant	Résolution spatiale	Echelle globale	Résolution en temps
Urbain	ha à 1 km ²	km ²	5 à 30 min
Petite	km ²	10 à 100 km ²	30 min à 1 h
Moyen	du km ² à la dizaine de km ²	1000 km ²	24 heures

Tableau 2.3: Ordre de grandeur des tailles et de la résolution des estimations de lames d'eau de chaque type de bassin versant.

La réponse dépend du moyen de mesure (dans le cas de mesures issues d'un réseau de pluviomètres ou de pluviographes, elle dépend de la densité) mais surtout de la taille et de la nature du bassin versant.

Faure, 93 [33] considère trois types de bassins versants :

- Les bassins versants urbains ;
- Les petits bassins ruraux (par exemple montagneux) ;
- Les grands bassins versants.

On a synthétisé dans le tableau 2.3 l'échelle globale d'espace caractérisant chacun des bassins versants et les deux résolutions (en espace et en temps) nécessaires pour l'utilisation en pratique des sorties du modèle. La résolution temporelle correspond à la durée de la réponse du bassin versant.

2.1.4 Conclusion

Dans ce chapitre, nous avons dressé une synthèse de l'état actuel des connaissances sur la dynamique des champs de pluie, puis défini les contraintes liées à la mesure de la pluie et à son utilisation et réalisé une synthèse des sources d'erreurs de mesures ainsi que des ordres de grandeur d'échelles d'espace et de temps concernées.

L'état actuel des connaissances permet d'avoir une représentation globale de la dynamique des champs de pluie et des estimations de lame d'eau plus ou moins fiables selon les dispositifs de mesure. Toutefois, cette représentation météorologique et ces dispositifs de mesure fractionne le processus en un nombre fini d'échelles qui ne correspondent pas forcément aux besoins hydrologiques, et rendent difficile l'extrapolation des estimations à d'autres échelles. L'une des ambitions de ce travail est de les traiter simultanément en se fondant sur une continuité du processus de pluie entre échelles.

2.2 La modélisation stochastique de la pluie

Dans la partie précédente, nous avons rapidement présenté les difficultés soulevées par la modélisation physique de la pluie. Bien que les phénomènes entrant en jeu ne soient pas forcément régis par le hasard, une manière de décrire les processus hydrologiques consiste à construire un modèle de pluie stochastique. Selon E. Halphen : "Les probabilités mesurent l'ignorance humaine et la statistique est ce qui tient lieu de science aux ignorants que nous sommes."

2.2.1 Pourquoi une modélisation stochastique ?

L'approche stochastique de modélisation est susceptible a priori de présenter divers avantages en hydrologie :

- Tout d'abord, elle permettrait la réalisation de transferts d'information en des sites géographiques différents, ou bien à des échelles plus petites (spatiales, temporelles ou spatio-temporelles), atout non négligeable en raison du prix et de la difficulté d'acquisition des mesures.
- Face à l'extension des zones urbaines et à l'augmentation de la vulnérabilité, l'hydrologue a besoin d'estimations à une résolution spatiale et temporelle de plus en plus fine et de prévisions de plus en plus fiables. Les modèles stochastiques réalisent des estimations dont la précision dépend de la variabilité de la série, tout en quantifiant l'erreur, qui peut être retranscrite en terme de risque (Berger, 85 [11]).
- Un autre avantage de l'approche stochastique est le caractère opérationnel du modèle pour des applications hydrologiques. En effet, on modélise toujours par rapport à un objectif, et le modèle doit être bien distingué du phénomène à modéliser (de Marsily, 94 [26]). L'hydrologue s'intéresse à la prévision des crues et des étiages, à la gestion des réservoirs ou à l'estimation des extrêmes pour le dimensionnement d'ouvrages. Une prévision de la pluie pour l'annonce des crues en milieu urbain requiert une modélisation à petit pas de temps et d'espace, tandis que la détermination de la hauteur d'une digue sera réalisée à partir d'un modèle à l'échelle du bassin versant.

Mais modéliser des grandeurs hydrologiques en utilisant des modèles stochastiques n'est généralement pas une tâche facile. En fait, les variables hydrologiques possèdent une structure de dépendance et une variabilité des moins classiques sur le plan statistique. Les séries de débits ou de pluie sont fortement asymétriques (notamment à petit pas de temps, voir figure 2.2) et possèdent généralement une fonction d'autocorrélation convergeant lentement vers zéro.

Parmi les principaux modèles stochastiques proposés en hydrologie, on distingue les modèles de type gaussien et de type non-gaussien, au sein desquels apparaissent diverses structures de dépendance statistique (ARMA, Markov) qui possèdent chacune une interprétation physique simple.

Dans cette section, nous rappelons les principales méthodes d'estimation spatiale, avant de dresser une revue des différents modèles stochastiques de précipitation rencontrés dans la littérature : nous présenterons les modèles spatiaux, temporels, puis les modèles qui appréhendent la pluie en tant que champ spatio-temporel.

2.2.2 Méthodes statistiques d'estimation spatiale

Dans ce paragraphe, nous passerons en revue les méthodes de pondération, du krigage et du variogramme qui visent le plus souvent à obtenir des informations sur la variabilité spatiale de la pluie, mais peuvent aussi être employées pour la variabilité temporelle ou spatio-temporelle.

Figure 2.1: Asymétrie des cumuls en fonction du pas d'intégration (série de Athènes).

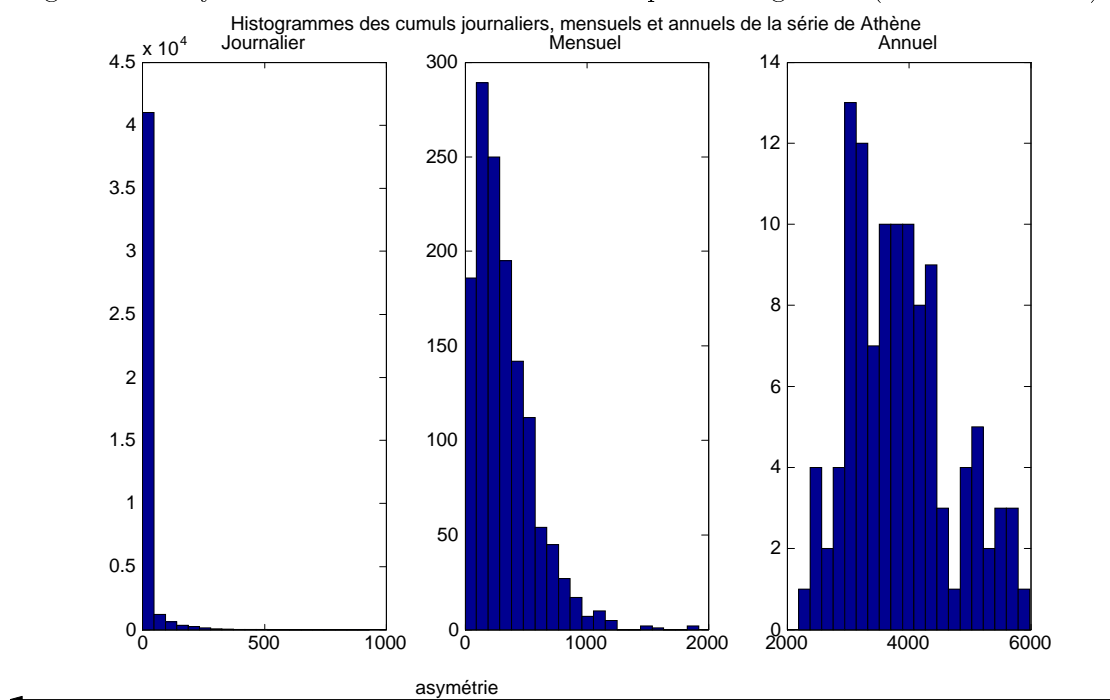


Figure 2.2: Asymétrie de la distribution des cumuls pluviométriques.

La démarche générale de ces méthode est la suivante : On aimerait déterminer, à partir d'un réseau dense de pluviomètres ou pluviographes, la hauteur de pluie tombée en tout point d'une région, ou bien en moyenne sur un domaine de cette région. On dispose de n mesures $(z_i)_{i=1,\dots,n}$ prises aux pluviomètres situés aux points $(x_i)_{i=1,\dots,n}$ et on veut estimer :

- ponctuellement la mesure z en un point x

$$z(x) = f(z_1, z_2, \dots, z_n) = f(z_1(x_1), z_2(x_2), \dots, z_n(x_n))$$

- ou en moyenne la lame d'eau Z sur un domaine X

$$Z(X) = F(z_1, z_2, \dots, z_n) = F(z_1(x_1), z_2(x_2), \dots, z_n(x_n))$$

2.2.2.a Méthodes de pondération : Les méthodes de la moyenne pondérée, des polygones de Thiessen et des isohyètes, sont des méthodes déterministes de pondération. Elles sont fondées sur l'hypothèse de continuité du champ de pluie (en temps ou en espace) et peuvent servir à une estimation spatiale ou temporelle d'une hauteur, d'un cumul ou d'une intensité de pluie. Pour un pas de temps donné, un champ pluviométrique est *continu* si la mesure en un point est représentative de la lame d'eau précipitée sur une surface assez grande.

Remarquons que ces méthodes peuvent cependant aussi être considérées comme aléatoires, l'aléa portant sur les hauteurs z_i ou sur la répartition des postes x_i .

Méthode de la moyenne pondérée : On estime la variable aléatoire au point x à partir d'une moyenne pondérée sur les postes voisins $(x_i)_{i=1\dots n}$

$$\hat{z} = f(z_1, \dots, z_n) = \frac{\sum_{i=1}^n w(d_{xx_i}) \times z_i}{\sum_{i=1}^n w(d_{xx_i})}$$

avec

- \hat{z} la hauteur estimée au point x
- d_{xx_i} la distance de x à x_i .
- w une fonction de poids (par exemple $w(d) = \frac{1}{d}$) est décroissante en fonction de la distance. Le poids affecté à chaque point x_i s'écrit

$$\frac{w(d_{xx_i})}{\sum_{i=1}^n w(d_{xx_i})}$$

Méthode des polygones de Thiessen : Cette méthode, proposée par Thiessen en 1911 [82], est une généralisation de la précédente lorsque $w(d) = \frac{1}{d^k}$ et $k \rightarrow \infty$. Elle permet de réaliser une *estimation sur un domaine*. Les poids sont définis comme suit : on affecte un polygone d'influence à chaque observation z_i d'un bassin versant ; le polygone d'influence est un domaine défini par toutes les médiatrices des segments joignant les couples de points autour de chaque observation. Le poids affecté à chaque observation est le rapport de la surface du polygone d'influence sur la surface du bassin $\frac{S_i}{S}$:

$$\widehat{Z} = \sum_{i=1}^n \frac{S_i}{S} \times z_i$$

Méthode des isohyètes : Pour chaque couple d'observations, on découpe le segment les joignant en intervalles réguliers et on relie ensuite les valeurs identiques pour obtenir les courbes isohyètes $(H_i)_{i=1,\dots,n}$ correspondant à des hauteurs $(z_i)_{i=1,\dots,n}$. En notant S_i l'aire délimitée par les isohyètes H_i et H_{i+1} , on estime la lame d'eau par

$$\widehat{Z} = \sum_{i=1}^n \frac{S_i}{S} \times \frac{z_i + z_{i+1}}{2}$$

2.2.2.b Méthodes d'estimation optimales : Les méthodes du krigage et du variogramme sont toutes deux des méthodes d'estimation optimales, et ont été introduites il y a 30 ans dans des domaines aussi variés que l'industrie minière par Matheron, 65 [64], ou la météorologie par Gandin, 65 [38]. On n'en fait ici qu'une brève présentation. Pour les détails des calculs, on se référera à Delhomme, 78 [27].

On considère qu'un champ de pluie correspond à la réalisation d'une fonction aléatoire (*FA*) notée Z , à valeurs $Z(x_i)$, et on détermine l'estimateur \widehat{Z} de la hauteur Z en x qui minimise l'erreur quadratique d'estimation :

$$\min_i E \left\{ \left(\widehat{Z}(x) - Z(x_i) \right)^2 \right\}$$

Méthode du krigage (simple) : Elle repose sur l'hypothèse de stationnarité d'ordre un et deux de la *FA* Z dans l'espace :

- L'espérance est constante

$$\forall x \quad E(Z(x)) = m(x) = m$$

- La covariance entre deux points x_i et x_j de ce champ ne dépend pas séparément des deux points mais seulement de leur distance $|x_i - x_j|$:

$$\text{cov}[Z(x_i), Z(x_j)] = E[Z(x_i) \cdot Z(x_j)] - E Z(x_i) \cdot E Z(x_j) = K(|x_i - x_j|)$$

On détermine l'estimateur \widehat{Z} de la hauteur Z en x qui minimise l'erreur quadratique d'estimation $\min E \left\{ \left(\widehat{Z} - Z \right)^2 \right\}$ avec une contrainte de non biais $E(\widehat{Z} - Z) = 0$, et on recherche un estimateur de la forme

$$\widehat{Z}(x) = \sum_{i=1}^n \alpha_i Z(x_i)$$

On obtient un système sous contraintes permettant de calculer les coefficients $(\alpha_i)_{i=1, \dots, n}$ qui est résolu par la méthode des multiplicateurs de Lagrange en se ramenant à une équation matricielle :

$$\begin{bmatrix} C & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \Lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_0 \\ 1 \end{bmatrix}$$

où C et C_0 représentent respectivement la matrice de covariance des postes de mesures (dimension $n \times n$) et le vecteur des covariances entre le point à interpoler x et les postes x_i (dimension n). On estime C et C_0 à partir des réalisations de Z en ajustant un modèle $C(d)$ sur les $\frac{n(n-1)}{2}$ valeurs mesurées de la covariance.

L'inconvénient majeur de la méthode du krigage est qu'elle est fondée sur une hypothèse trop contraignante. Il existe cependant d'autres versions de cette méthode qui s'adaptent mieux aux variables hydrologiques (le krigage lagrangien de Amani et Lebel, 96 [1] par exemple).

Méthode du variogramme : Le variogramme est un outil pour quantifier la variabilité (spatiale, temporelle ou spatio-temporelle) d'un champ.

Il remplace la fonction de covariance dans le cas de processus non stationnaires car on se place sous l'hypothèse intrinsèque (Matheron, 65 [64]). Cette dernière est plus faible, car c'est une hypothèse de stationnarité d'ordre un et deux qui ne porte que sur les accroissements Z :

Pour tous points x_i et x_j distants de $|h|$:

$$\begin{aligned} E(Z(x_i) - Z(x_j)) &= m(|h|) \\ \text{Var}(Z(x_i) - Z(x_j)) &= 2\gamma(|h|) \end{aligned}$$

La variance ne dépend que de la distance $|h|$.

Définition 1 *Le variogramme $\gamma(|h|)$ de l'événement est défini par*

$$\gamma(|h|) = \frac{1}{2} \text{Var}(Z(x_i) - Z(x_j))$$

Sous l'hypothèse $m(|h|) = 0$, il s'écrit encore $\gamma(|h|) = \frac{1}{2} E \left[(Z(x_i) - Z(x_j))^2 \right]$. Le variogramme de h est l'accroissement quadratique moyen entre deux points distants de $|h|$.

On minimise l'erreur quadratique sous contrainte de non biais et sous l'hypothèse intrinsèque. On obtient par la méthode des multiplicateurs de Lagrange un système linéaire

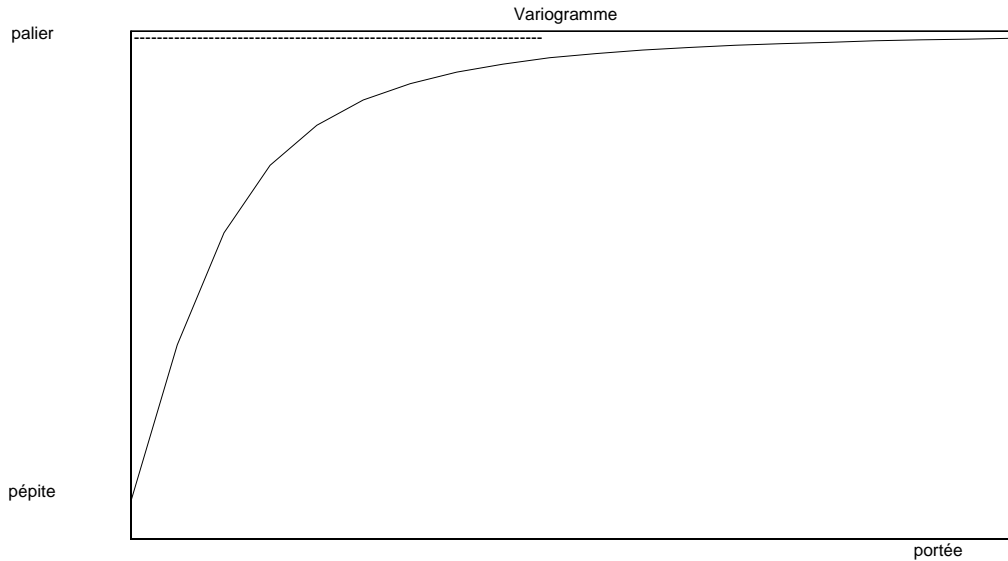


Figure 2.3: Pépite, portée et palier d'un variogramme.

permettant d'estimer les coefficients $(\alpha_i)_{i=1..n}$:

$$\begin{bmatrix} \gamma_{11} & \dots & \gamma_{1m} & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{m1} & \dots & \gamma_{mm} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \cdot \\ \alpha_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \cdot \\ \gamma_{m0} \\ 1 \end{bmatrix}$$

avec μ coefficient de Lagrange et $\gamma_{ij} = \frac{1}{2} E [Z(x_i) - Z(x_j)]$

La résolution de cette équation nécessite le choix d'un variogramme.

D'un point de vue pratique, ce variogramme est estimé en :

- Regroupant les $\frac{n(n-1)}{2}$ couples de mesures en un nombre N_C de classes d'interdistance.
- Affectant à chaque distance moyenne de classe d_m la valeur estimée et on estime $\gamma(d_m)$ par la moyenne sur les classes des carrés des écarts.

$$\hat{\gamma}(d_m) = \frac{1}{2N_k} \sum (z(x) - z(x+h))^2$$

avec N_k le nombre de couples présents dans la classe k .

- On dispose de N_C valeurs de $\hat{\gamma}(d_m)$ correspondant à l'une des N_C valeurs d_m . On ajuste un variogramme théorique (sphérique, exponentiel, gaussien, ...) au variogramme expérimental (figure 2.3). Les modèles couramment utilisés sont définis par trois paramètres :

- le palier, qui est égal à la variance du champ. Il n'apparaît pas si le phénomène est très variable.
- la portée, qui représente le seuil à partir duquel le palier est atteint.
- la pépite, qui renseigne sur la continuité de Z . Une discontinuité en zéro, qui peut être due à la discontinuité du milieu, à un problème d'échelle ou à des erreurs de mesures.

En conclusion, ces différents outils, conçus pour appréhender la variabilité spatiale d'un champ, reposent tous sur des hypothèses d'homogénéité (plus ou moins affaiblie) de ce dernier. Or on a constaté que ces hypothèses sont mises à mal lorsqu'il s'agit d'un champ pluviométrique, surtout à petit pas de temps (on pourra se reporter à l'article de Hubert et Carbonnel, 88 [44] pour une étude sur les données EPSAT-Niger).

2.2.3 Modèles temporels

On distingue les modèles d'occurrence de pluie des modèles de hauteur non nulles. Ces deux processus peuvent être soit modélisés séparément, puis recombinaison, soit modélisés simultanément. On ne présentera ici que les modèles en un seul site.

Modèles d'occurrence de pluie

Pour modéliser l'occurrence de pluie en un point fixé, on considère la variable aléatoire discrétisée à un certain pas de temps (généralement la journée) $X_t = 0$ ou 1 selon qu'il pleuve ou non sur ce pas de temps (ou bien selon que le cumul de pluie dépasse ou non un certain seuil). Deux approches ont été considérées ces dernières années : la modélisation des périodes sèches et des périodes humides et les séries à temps discret.

La première approche consiste à s'intéresser aux séquences de jours (ou heures) de pluie consécutifs. On ajuste une loi à la durée des périodes humides et des périodes sèches. Cette voie a été largement exploitée par des auteurs tels que Green en 64 [41] ou Eagleson en 78 [29] qui utilisent respectivement une loi exponentielle et une loi de Weibull pour les durées des périodes sèches et humides. Ces modèles sont en fait des processus de renouvellement dans le sens où états secs et pluvieux alternent, aucune transition dans le même état n'étant possible. Un autre modèle de renouvellement créé par Galloy, Martin et Lebreton en 82 [37] utilise la loi binomiale négative pour les durées des périodes sèches et humides.

Cependant ces méthodes reposent sur l'hypothèse d'indépendance des périodes pluvieuses entre elles, rarement vérifiée à petit pas de temps (à pas de temps une heure, deux séquences humides peuvent correspondre au même événement pluvieux). De plus, la durée variable des événements mériterait que l'on conditionne les cumuls à cette durée (Wilks, 89 [86]).

La seconde approche utilise les séries à temps discret, série binaire de 0 ou de 1. L'occurrence de pluie peut être modélisée par :

- Une suite de variables aléatoires de Bernoulli. C'est le modèle le plus simple mais il requiert une indépendance complète. Il est mieux adapté aux grands pas de temps (jour et plus).

- Les chaînes de Markov : Elles sont très utilisées du fait de leur simplicité. Ce sont des chaînes à deux états $X_t = 0$ ou 1 . Le précurseur de cette modélisation semble être Quetelet en 1852 (d'après Woolhiser, 91 [87]). Van-Thanh-Van Nguyen, 83 [84] détermine la loi de probabilité des cumuls journaliers appartenant à une période pluvieuse de n jours, en décrivant l'occurrence de pluie par une chaîne de Markov stationnaire d'ordre 1. Mais selon le degré de persistance désiré, les chaînes utilisées peuvent être d'ordre supérieur à un et même non homogènes. Stern et Coe [78] représentent l'alternance des jours secs et pluvieux par une chaîne de Markov non homogène en se plaçant à l'échelle d'un jour. Ces modèles servent à calculer la loi des cumuls sur n jours, la loi du maximum annuel des précipitations. Les théorèmes limites fournissent des lois asymptotiques pour les hauteurs extrêmes.

Ces modèles ne décrivent pas les persistance à long terme (par exemple les activités cycloniques persistantes durant certaines saisons) ni les effets de classement à petit pas de temps (car, par exemple, un événement pluvieux dû à un événement pluvieux précédent a une vraisemblance plus forte).

- Les processus $DARMA(p, q)$, p étant l'ordre de l'autorégression et q l'ordre de la moyenne mobile créés par Buishand en 1978 [16] tiennent mieux compte des persistance à long terme que les chaînes de Markov d'ordre élevé mais la structure du modèle est difficile à interpréter physiquement.

Modèles de hauteur non nulles

L'approche la plus courante est de décrire la distribution des hauteur $Y(t)$ indépendamment de celle de l'occurrence $X(t)$. Plusieurs types de dépendance caractérisent ces modèles :

- Les $Y(t)$ sont indépendantes : Divers lois sont utilisées telles que la loi Exponentielle, Gamma ou Weibull (Zucchini et al., 92 [89]).
- Les $Y(t)$ sont dépendantes : On utilise des modèles ARMA pour modéliser une dépendance entre les $Y(t)$ non nuls. Ils sont surtout utilisés pour modéliser les débits mais peu pour les hauteurs de pluie à cause de la quasi indépendance.
- Les $Y(t)$ sont indépendantes mais dépendent de l'occurrence du jour précédent $X(t - 1)$: Ce sont les processus dépendants en chaîne, introduits par Katz en 1977 ([48] et [49]) qui conditionnent $Y(t)$ par $X(t - 1)$. Ces modèles nécessitent un grand nombre de paramètres.

D'autres expressions de la dépendance peuvent être trouvées dans la littérature, telles que celle de Buishand, 78 [16] qui classe les événements pluvieux selon les types suivants :

- I : jour pluvieux isolé.
- II : jour pluvieux précédé ou suivi d'un jour sec.
- III : jour pluvieux précédé et suivi d'un jour sec.

Modèles Mixtes

On peut modéliser l'occurrence $X(t)$ et les hauteurs $Y(t)$ séparément ou bien simultanément $Z(t) = (X(t), Y(t))$. La plupart des modèles mixtes s'appuient sur le cadre mathématique construit par Le Cam en 1961 [17] qui adapte à l'hydrologie un processus d'agrégation de type Neyman-Scott. Ce dernier, créé en astronomie pour modéliser la formation d'étoiles, présente l'avantage de prendre en compte la dynamique des champs de pluie. De nombreux auteurs se sont attachés à analyser formellement la qualité des estimateurs (Cowpertwait, 91 [20]) ou à simuler des données à pas de temps fin, notamment en hydrologie urbaine (Thauvin et al. 97 [81]). Dans un modèle de type Neyman-Scott (figure 2.4), durées de temps sec et de temps pluvieux alternent. Les origines de périodes pluvieuses (notées '0'), arrivent selon un processus de Poisson. A l'intérieur d'un intervalle pluvieux, la pluie se décompose en une succession d'averses (représentées sur le schéma par le symbole '*'). Chaque averse est caractérisée par sa durée et son intensité. L'intensité totale en chaque point est la somme des intensités de chaque averse active.

Le processus de Neyman-Scott présente des variantes définies par la loi des intensités attribuées à chaque cellule de pluie. Citons le Neyman-Scott White Noise (NSWN) de Rodriguez-Iturbe et al. en 84 [74] ou le Neyman-Scott Rectangular Pulse (NSRP) de Rodriguez-Iturbe et al. en 87 [73]. Il fait partie d'une vaste classe de processus ponctuels présentés sur un plan théorique dans l'ouvrage de Cox et Isham [21] parmi lesquels on peut citer le processus de Poisson double (ou processus de Cox). Ce dernier est un processus de Poisson à intensité aléatoire. Kavvas et Delleur [50] appliquent le processus de Cox pour construire un modèle stochastique à deux niveaux, le RCM (Renewal Cox process with Markovian Intensity) pour l'occurrence de pluie dans lequel la vitesse d'occurrence des orages est déterminée par un autre processus appelé processus climatologique. Smith et Karr, 83 [77] exploitent ce modèle pour calculer le nombre d'orages et les durées d'inter-arrivées sur des données issues du bassin versant de la rivière Potomac.

Le processus de Neyman-Scott se généralise naturellement en un processus spatial ou spatio-temporel comme nous pourrions le constater dans la section suivante.

2.2.4 Modèles spatio-temporels

La littérature propose peu de modèles spatio-temporels. Ils ont été créés généralement en vue de simulations afin de tester des stratégies d'estimation à partir d'appareils de mesure de types différents. Les modèles présentés ici sont construits autour de structure de variabilité observées aux différentes échelles choisies selon les observations (cellules pluvieuses, bandes de cellules en espace, événement pluvieux en temps...).

Le plus simple de ces modèles est celui de Cox et Isham [22]. Il généralise le processus de Neyman-Scott en espace et en temps ; il est fondé sur des hypothèses très restrictives. Les averses arrivent selon un processus de Poisson en espace et en temps de paramètre λ . Chaque averse est modélisée par une région circulaire de pluie de rayon R qui se déplace à une vitesse aléatoire $v = (v_x, v_y)$ sur une durée aléatoire D jusqu'à disparition de la cellule. L'intensité de pluie X est supposée constante. C'est au prix de ces simplifications drastiques que le nombre de paramètres reste raisonnable (six au total : λ, R, v_x, v_y, D et X).

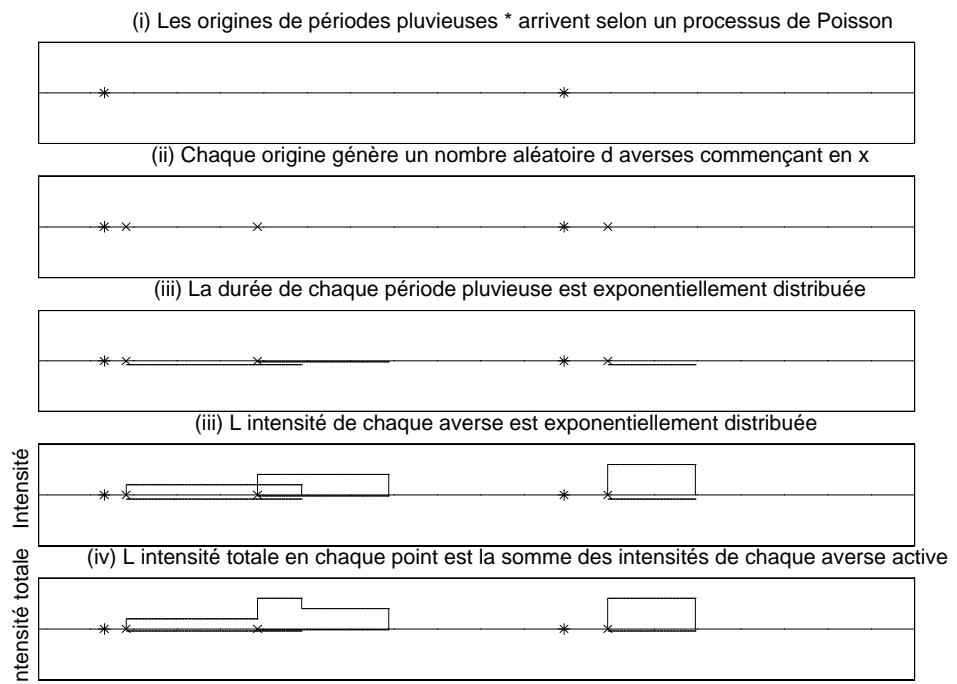


Figure 2.4: Processus de Neyman-Scott.

Dans le modèle de Waymire, Gupta, Rodriguez Iturbe, 84 [85], chaque cellule pluvieuse est caractérisée par son intensité, sa durée de vie ainsi que des paramètres de taille et de forme. Elles sont groupées en bandes pluvieuses de la petite méso-échelle (*section* 2.1.1). Leur nombre et leur emplacement suit un processus spatial de Poisson. Ces bandes pluvieuses de la petite méso-échelle sont elles-mêmes groupées dans des zones de pluie de la grande méso-échelle, leur nombre et leur emplacement suivant aussi un processus de Poisson de paramètres différents. L'intensité à l'intérieur de la cellule décroît exponentiellement depuis le centre jusqu'à atteindre zéro sur les bords. Au total, ce modèle comporte 12 paramètres. La moyenne et la variance théorique sont calculées. L'hypothèse de turbulence de Taylor (Taylor, 1938) stipule que la covariance en temps en un point fixe sur une durée Δt est égale à la covariance à une date arbitraire entre deux points séparés par une distance Δx telle que Δx soit égal à Δt à une constante près : $\Delta x = U \cdot \Delta t$. L'hypothèse de Taylor est vérifiée par les champs de pluie simulés par ce modèle.

Les chaînes de Markov peuvent elles aussi être généralisées en espace-temps. Pour des applications de ces modèles en hydrologie, on pourra consulter les travaux de Zucchini et Guttorp en 91 [88] ou la thèse de Bayomog, 94 [6] qui emploie des processus de Markov spatio-temporels pour modéliser l'occurrence de pluie.

2.2.5 Conclusion

Contrairement aux modèles temporels, on trouve peu de modèles spatiaux ou spatio-temporels dans la littérature. Les modèles temporels sont développés en vue de faire de l'estimation et de la prévision tandis que les modèles spatiaux ou spatio-temporels sont en général des modèles de simulation.

La prise en compte de la structure d'échelle du phénomène pluvieux est importante (pour la désagrégation par exemple), et permettrait d'établir un pont entre les connaissances des hydrologues et des météorologues. Les modèles stochastiques de pluie basés sur la classification de Austin et Houze font appel à un trop grand nombre de paramètres (au minimum 12, modèle de Waymire et al., 84 [85]), et ne donnent de bons résultats que sur les échelles sur lesquelles ils ont été calibrés. Les modèles hydrologiques mettent en oeuvre des outils statistiques standards : des distributions d'intensités ou de cumuls de type exponentiel (loi exponentielle, exponentielle mélangée, normale, gamma, kappa, etc...), ainsi qu'une structure de dépendance dite à court terme telle que ARMA ou Markov. Ces modèles donnent des résultats satisfaisants pour l'estimation ou la prévision des événements courants (faibles périodes de retour), mais sont défaillants pour les événements rares. En fait, les choix de ces lois et dépendances sont arbitraires et parfois ne s'accordent même pas avec les constatations empiriques. Des études de plus en plus nombreuses tendent à montrer que les corrélations entre observations éloignées (en espace ou en temps) décroissent à une vitesse inférieure à celle obtenue avec des données indépendantes ou un modèle de type ARMA ou Markov. Un modèle stochastique adapté à cette forme de dépendance est donc nécessaire.

Dans un souci d'adéquation à la réalité physique du phénomène et dans le but d'intégrer nos connaissances a priori, un modèle de pluie devrait représenter une série à différentes échelles, avec les mêmes valeurs pour certains paramètres. Mais comment élaborer des modèles fondés sur la probable existence d'invariant(s) d'échelle ? Les fractals et mul-

tifractals, présentés dans la section suivante, constituent un outil susceptible de répondre à cette question.

2.3 Les modèles fractals

Dans cette section, on présente les principales définitions ainsi que les résultats de la théorie fractale. La géométrie fractale fournit un moyen de quantification du degré d'irrégularité (encore appelé rugosité) d'un ensemble. De plus elle permet d'extraire un paramètre indépendant de l'échelle : la dimension fractale. Des applications directes d'estimation de cette dernière sur des données sahéliennes d'occurrence de pluie (données Badinage) sont réalisées dans cette section, et leurs résultats comparés à ceux existant dans la littérature.

Nous présentons ensuite les cascades multifractales qui permettent de modéliser non seulement le support d'occurrence d'un phénomène mais aussi son intensité. Ces modèles stochastiques possèdent des caractéristiques fractales telles que des paramètres invariants d'échelle et une grande irrégularité. Enfin, une revue des principaux auteurs ayant travaillé avec les fractals et les multifractals en hydrologie (ainsi que leurs résultats) est ensuite dressée.

2.3.1 Les fractals

Les fractals, dont la racine latine *fractus* signifie irrégulier, ont été introduits par Mandelbrot en 1975 [59] pour caractériser des ensembles aux propriétés inhabituelles en géométrie classique. L'intérêt porté à cette théorie, initialement fondée sur des contre-exemples introduits par des mathématiciens entre les années 1875 et 1950, a été stimulé par ses multiples possibilités d'application à des objets naturels (tels que le contour des nuages, les côtes, les réseaux hydrographiques).

Dans cette section, on rappelle les définitions d'objet et de dimension fractals en se référant à l'ouvrage de Falconer, 90 [32]. Puis une étude des artefacts de la principale méthode d'estimation de la dimension fractale (la méthode du comptage de boîtes) est menée sur des simulations d'ensemble de Cantor.

Objet fractal

Plusieurs définitions d'objet fractal ont été proposées. Nous choisirons ici celle de Mandelbrot [59], selon laquelle un fractal est un ensemble dont la dimension fractale (qui reste à définir) est strictement supérieure à sa dimension euclidienne. Mais certains auteurs, tel que Falconer, 90 [32], pensent que cette définition est trop restrictive et préfèrent concevoir un fractal F comme un ensemble possédant quelques unes des caractéristiques suivantes :

- F a une structure fine, c'est-à-dire des détails à des échelles arbitrairement petites.
- F est trop irrégulier pour être décrit en langage géométrique traditionnel, à la fois localement et globalement.

- Souvent F présente une forme d'auto-similarité (exacte ou statistique, cf *Annexe E*). Par exemple, dans certains cas, F est défini comme la limite d'une procédure récursive.

La notion de fractal est, quoi qu'il en soit, intimement liée à celle de dimension fractale. On trouve dans la littérature plusieurs définitions de cette dernière.

Dimension fractale

Dans ce paragraphe, on se limitera à présenter la dimension de boîte et la dimension de Hausdorff mais le lecteur pourra se référer à Falconer [32] pour d'autres dimensions fractales.

Définition 2 Soit F un ensemble borné non vide de \mathbb{R} (resp. \mathbb{R}^n). Considérons la suite de segments (resp. de cubes) de maille δ :

$$[m\delta, (m+1)\delta] \text{ avec } m \in \mathbb{N}$$

$$\text{(resp. } [m_1\delta, (m_1+1)\delta] \times \cdots \times [m_n\delta, (m_n+1)\delta] \text{ avec } m_1, \dots, m_n \in \mathbb{N})$$

et soit $N_\delta(F)$ le nombre de tels segments (resp. de cubes) intersectant F . Les dimensions de boîte supérieures et inférieures sont respectivement définies par :

$$\underline{\dim}_B F = \underline{\lim}_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

$$\overline{\dim}_B F = \overline{\lim}_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

(voir²). Si elles sont égales, on définit la dimension de boîte par :

$$\dim_B F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

La dimension de boîte est la notion la plus fréquemment utilisée pour définir la dimension fractale, car elle est très facile à calculer en pratique (contrairement à la dimension de Hausdorff comme on le verra par la suite). La méthode du comptage de boîtes, qui permet parfois de calculer la dimension de boîte d'un ensemble, consiste à utiliser une suite $\delta_n = \frac{1}{b^n}$, $b > 1$ et à déterminer :

$$- \lim_{n \rightarrow \infty} \frac{\log N_{\delta_n}(F)}{n \log b}$$

²où $\underline{\lim}$ désigne la limite inférieure : soit $E_\delta = \{f(r), 0 < r < \delta\}$

$$\underline{\lim}_{\delta \rightarrow 0} f(\delta) = \lim_{\delta \rightarrow 0} (\inf E_\delta)$$

et $\overline{\lim}$ la limite supérieure :

$$\overline{\lim}_{\delta \rightarrow 0} f(\delta) = \lim_{\delta \rightarrow 0} (\sup E_\delta)$$

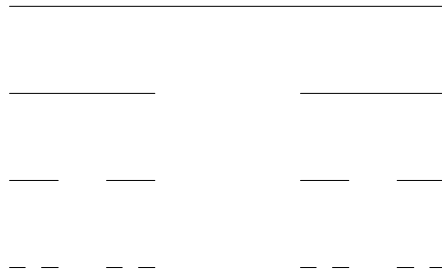


Figure 2.5: Construction de l'Ensemble de Cantor.

Exemple 3 *L'Ensemble de Cantor est un ensemble qui se construit par itérations : En partant d'un segment de longueur a que l'on divise en trois parties égales et auquel on retire le segment central, on obtient un ensemble formé de deux segments disjoints de longueur $a/3$ et séparés par un vide de même longueur (figure 2.5). On réitère indéfiniment l'opération pour obtenir un ensemble appelé Ensemble de Cantor (ou parfois poussière de Cantor).*

La dimension fractale de cet ensemble peut être déterminée par la méthode du comptage de boîtes en choisissant une suite de mailles de longueurs $\delta_n = \frac{a}{3^n}$, avec n entier, et en remarquant qu'une division par 3 de la longueur δ_n de la jauge conduit à une multiplication par 2 du nombre de boîtes $N_\delta(F)$ intersectant F , donc :

$$2 = 3^{\dim_B F} \text{ et } \dim_B F = \frac{\log 2}{\log 3} = 0.63$$

Exemple 4 *Courbe de Weierstrass : La fonction de Weierstrass est définie par :*

$$f : [0, 1] \longrightarrow \mathbb{R}$$

$$f(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t)$$

avec $\lambda > 1$ et $1 < s < 2$. On montre (Falconer [32]) que son graphe (figure 2.6) est un ensemble fractal de dimension de boîte égale à s .

D'une manière générale, la dimension fractale quantifie la puissance selon laquelle la mesure d'un ensemble diverge vers l'infini. Elle peut être interprétée comme le degré d'irrégularité de cet ensemble, conditionnellement cependant à la dimension classique (ou entière) de l'espace auquel elle appartient (tableau 2.4).

Pour un objet non fractal, la dimension de boîte est égale à sa dimension euclidienne (par exemple, un segment (resp. un carré) a pour dimension de boîte 1 (resp. 2). Pour un ensemble fractal, elle est non entière.

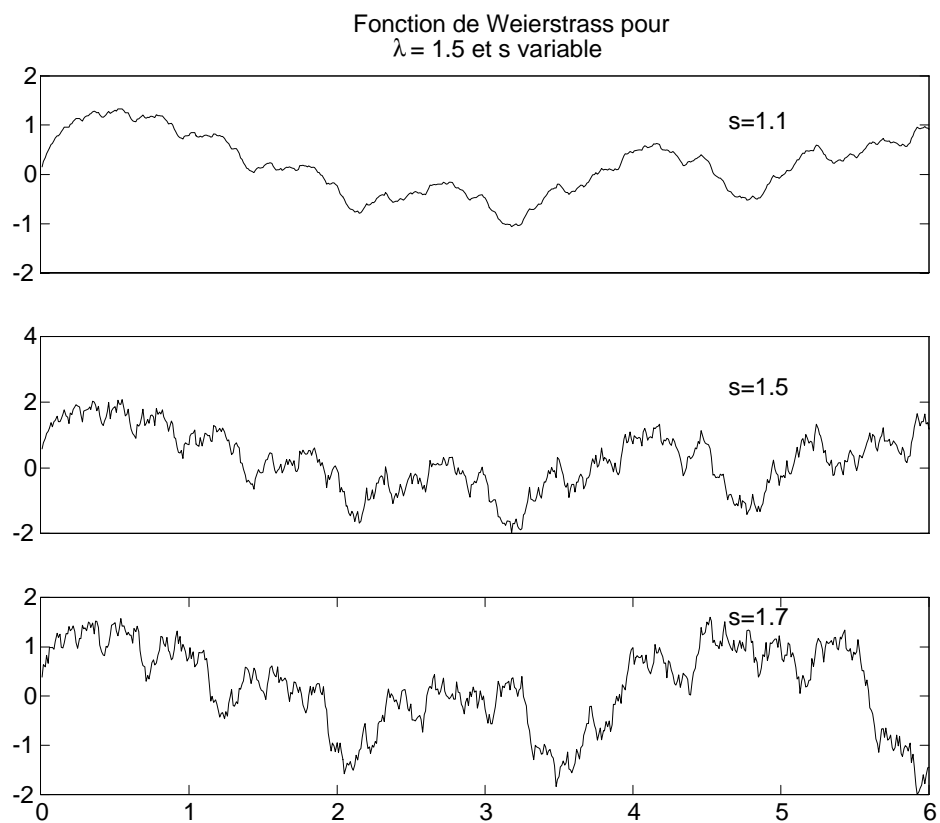


Figure 2.6: Graphes de la fonction de Weierstrass pour s variable.

Objet	Poussière sur une droite	Courbe plane très irrégulière	Surface très feuilletée
$\dim_B F$	$0 \leq \dim_B F \leq 1$	$1 \leq \dim_B F \leq 2$	$2 \leq \dim_B F \leq 3$

Tableau 2.4: Gammes de variation des dimensions fractales.

Cependant, la dimension de boîte est difficile à manipuler mathématiquement. En effet, on n'a généralement pas :

$$\dim_B U_{i=1}^{\infty} F_i = \sup_{i \geq 1} \dim_B F_i$$

Elle rend ainsi parfois très mal compte du caractère fractal de certains ensembles.

Exemple 5 L'ensemble $F = \{1, \frac{1}{2}, \dots, \frac{1}{k}, \dots\}$ a pour dimension de boîte $\frac{1}{2}$, est donc considéré comme fractal³.

La dimension de Hausdorff est une autre définition de la dimension fractale. Définie à partir de la mesure de Hausdorff, elle présente l'avantage d'être facile à manipuler et d'être définie pour tout ensemble. Elle est cependant difficile à calculer.

Mesure de Hausdorff : Rappelons que si U est un ensemble non vide inclus dans \mathbb{R}^n , alors son $|U|$ **diamètre** est défini par :

$$|U| = \sup \{|x - y| \mid x, y \in U\}$$

Si $\{U_i\}$ est une suite d'ensembles de diamètres inférieurs à δ recouvrant F (i.e. $F \subset U_{i=1}^{\infty} U_i$) avec $0 < |U_i| \leq \delta$, on dit qu'on a un **recouvrement** de F .

Définition 6 Soit F un ensemble inclus dans \mathbb{R}^n et soient s et δ deux nombre positifs.

Soit :

$$H_{\delta}^s(F) = \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s, \{U_i\} \text{ recouvrement de } F \right\}$$

La mesure de Hausdorff s -dimensionnelle est définie par :

$$H^s(F) = \lim_{\delta \rightarrow 0} H_{\delta}^s(F)$$

Dimension de Hausdorff : On montre qu'il existe une valeur de coupure $\dim_H(F)$ entre deux comportement relatifs à la mesure de F (figure 2.7) :

$$\text{Si } \alpha < \dim_H(F), \text{ alors } H^{\alpha}(F) = \infty$$

$$\text{Si } \alpha > \dim_H(F), \text{ alors } H^{\alpha}(F) = 0$$

³Preuve : On considère un recouvrement par les intervalles de taille $\delta > \frac{1}{2}$. Il existe un entier k tel que $\frac{1}{k(k-1)} > \delta \geq \frac{1}{(k+1)k}$. Ainsi, chaque intervalle ne peut recouvrir plus d'un seul élément de F , donc :

$$\frac{\log N_{\delta}(F)}{-\log \delta} \geq \frac{\log k}{\log [(k+1)k]} \text{ et } \underline{\dim}_B F \geq \frac{1}{2}$$

D'autre part, si $\frac{1}{2} > \delta > 0$ on choisit k tel que $\frac{1}{k(k-1)} > \delta \geq \frac{1}{(k+1)k}$. Alors $k+1$ intervalles recouvrent $[0, \frac{1}{k}]$ laissant $(k-1)$ intervalle recouvrir les autres points de F . Ainsi :

$$\frac{\log N_{\delta}(F)}{-\log \delta} \leq \frac{\log (2k)}{\log [(k-1)k]} \text{ et } \overline{\dim}_B F \leq \frac{1}{2}$$

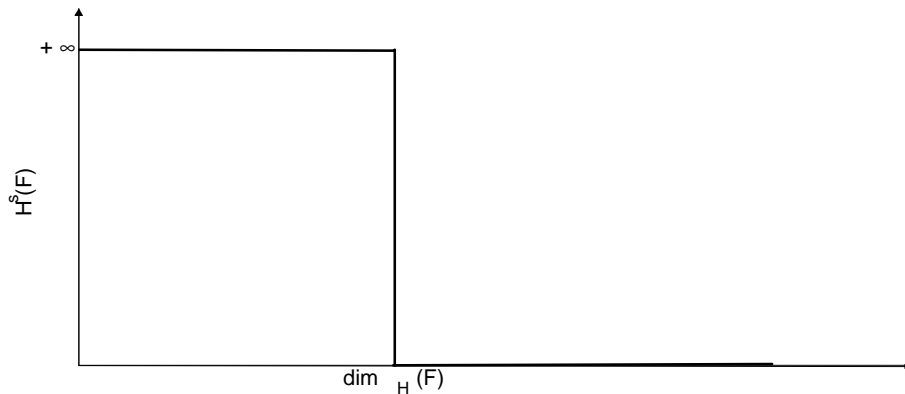


Figure 2.7: Mesure et dimension de Hausdorff.

Définition 7 La dimension de Hausdorff est définie par :

$$\dim_H(F) = \inf \{s : H^s(F) = 0\} = \sup \{s : H^s(F) = \infty\}$$

Exemple 8 Pour l'Ensemble de Cantor, on montre que, si $s = \frac{\log 2}{\log 3}$ alors $H^s(F) \in [\frac{1}{2}, 1]$. Donc $\dim_H(F) = s$.

Etude de la méthode du comptage de boîtes sur simulations

On simule un ensemble de Cantor à 3^9 éléments (soit un peu moins de 20 000 données) et on lui applique la méthode du comptage de boîtes afin de déterminer empiriquement sa dimension fractale et tester ainsi les défauts de la méthode. Plusieurs facteurs d'échelle m de taille de boîtes sont choisis (tailles de boîte variant de 2 en 2 ou de 3 en 3, etc...). On obtient (figure 2.8) des courbes à peu près linéaires pour $m = 2, 5, 7$ et un alignement parfait dans le cas où la taille des boîtes se démultiplie par 3 (qui est le facteur d'échelle d'un Ensemble de Cantor).

Pour chaque facteur d'échelle m , on calcule la pente moyenne entre points consécutifs et le nombre moyen de boîtes correspondant. Sur le tableau 2.5, on constate une surestimation (resp. sous-estimation) systématique du nombre moyen de boîtes non vides lorsque les boîtes se démultiplient par un facteur d'échelle supérieur (resp. inférieur) à 3. En *Annexe A*, on démontre ce résultat : le nombre de boîtes non vides de taille $\frac{1}{m^k}$ avec $m > 3$ (resp. $m < 3$), est supérieur (resp. inférieur) au nombre de boîtes non vides de taille $\frac{1}{3^k}$.

La méthode du comptage de boîtes présente donc en pratique quelques artefacts relatifs au choix du facteur d'échelle dont il faudra tenir compte lors des applications du paragraphe suivant.

Conclusion

La dimension fractale se révèle être parfois un moyen efficace pour extraire d'un ensemble une caractéristique invariante d'échelle. La méthode du comptage de boîtes est un

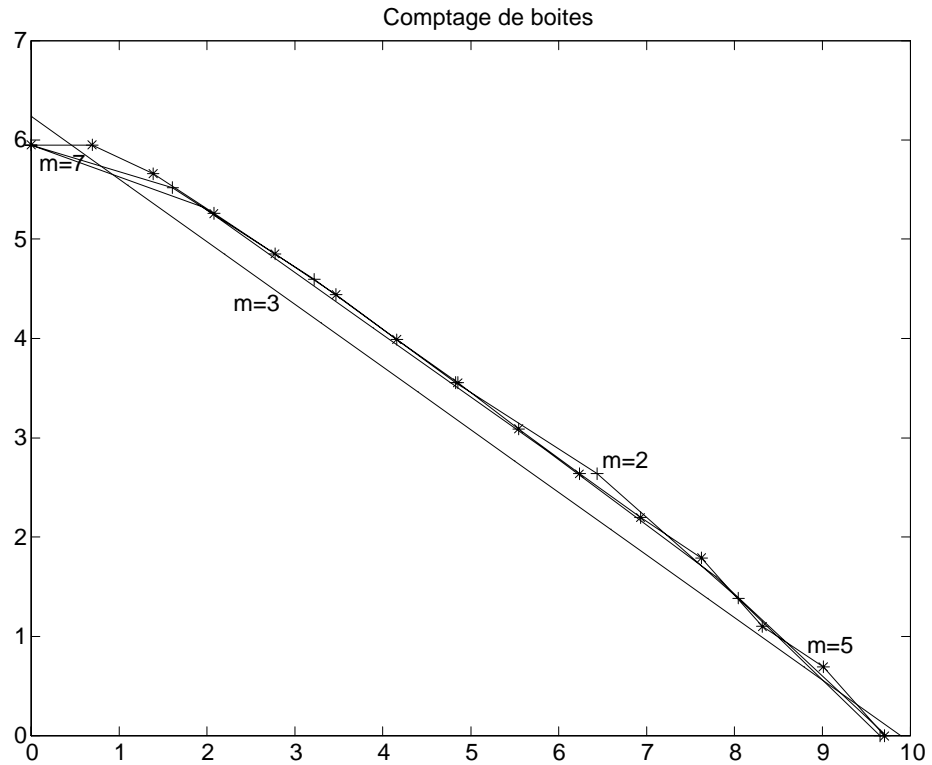


Figure 2.8: Comptage de boîtes sur un Cantor simulé à 3^9 données, avec boîtes se dé- m -uplant.

Nombre de données	2^{14}	3^9	5^6	7^5
a	2	3	5	7
Pente moyenne	-0.61	-0.63	-0.62	-0.61
N	101.1.	102.3	112.4	112.3

Tableau 2.5: Comptage de boîtes sur l'Ensemble de Cantor pour plusieurs facteurs d'échelle a (N = nombre moyen de boîtes non vides).

moyen simple d'estimation de la dimension fractale qui ne conduit cependant pas toujours à des résultats fiables. En effet, le critère de choix du facteur d'échelle n'est pas établi et l'estimateur de boîte manque de robustesse quand la série est tronquée. De plus, la forte intermittence des séries pluviométriques rend inutile l'exploration du comportement fractal en dessous d'un pas de temps minimal. Dans le but d'accéder à une quantification de la qualité d'estimation, cette méthode a aussi été étudiée dans un cadre probabiliste (Falconer, 90 [32] ou Harte, 97 [42]). Nous présentons dans le paragraphe suivant une application directe de la géométrie fractale à l'analyse de la répartition temporelle de séries pluviométriques (occurrence de pluie ou de basculement d'auget).

2.3.2 Application à quelques séries d'occurrence

On représente l'occurrence de pluie comme un nuage de segments disjoints supportés par l'axe représentatif du temps. La méthode du comptage de boîtes est appliquée à cet ensemble évoquant une poussière de Cantor, afin d'analyser la répartition temporelle d'occurrence de pluie et trouver d'éventuels invariants d'échelle.

Application 1 : Données *Badinage* Ce sont cinq mois de cumuls de pluie à pas de temps 5 minutes de l'année 1991, sur quatre mois de la saison pluvieuse (entre le 14.04.91 et le 02.09.91). On dispose des relevés en six stations au Niger. Le cumul minimum observé sur ces données est de 5 millimètres.

On applique la méthode du comptage de boîtes au nuage d'occurrence de cumuls non nuls F en reportant sur un graphe log-log la taille δ des boîtes (ici des segments de longueur $\delta = 3^k$) vs $N_\delta(F)$ le nombre de boîtes contenant au moins un segment du nuage d'occurrence de pluie (figure 2.9).

Les six graphes de comptage de boîtes se présentent sous forme de lignes brisées. Cette linéarité du graphe confirme la présence d'auto-similarité dans les séries temporelles d'occurrence de pluie. D'autre part, les deux ruptures de pente sont constatées aux mêmes abscisses sur les 6 courbes : aux pas de temps 1 heure 40 minutes ($e^3 * 5 = 100,43$ min) et 1 jour 10 heures ($e^6 * 5 = 2017$ min). La pente -1 observée sur la dernière portion du graphe est due au fait que toutes les boîtes sont pleines lorsque l'on considère des pas de temps supérieurs à un mois (il pleut au moins une fois tous les 10 jours en cette saison). Quant à la première rupture, elle indique que l'occurrence de pluie apparaît par paquets denses sur des gammes de pas de temps allant de 1 heure à 1 jour. Une explication peut être trouvée dans l'analyse des événements pluvieux et de leur durée. En effet, cette rupture peut traduire un changement de régime pluvieux car elle correspond à la durée moyenne d'un événement pluvieux dans cette région (1 heure 40 minutes). On a représenté dans le tableau 2.6 les pentes sur les deux premières portions du graphique.

Sur la deuxième portion (pas de temps compris entre 1 h 40 min et 1 j 10 h), la pente moyenne est de -0.27 . Ce résultat est proche de ceux de Hubert et Carbonnel en 1989 [45] : sur la série horaire de Ouagadougou (Burkina Faso), ils estiment la dimension de boîte sur ces pas gammes de temps à 0.22.

Sur la première portion du graphe (pas de temps de l'ordre de la minute), on constate une pente moyenne de -0.55 , résultat qui se rapproche de la pente 0.47 déterminée

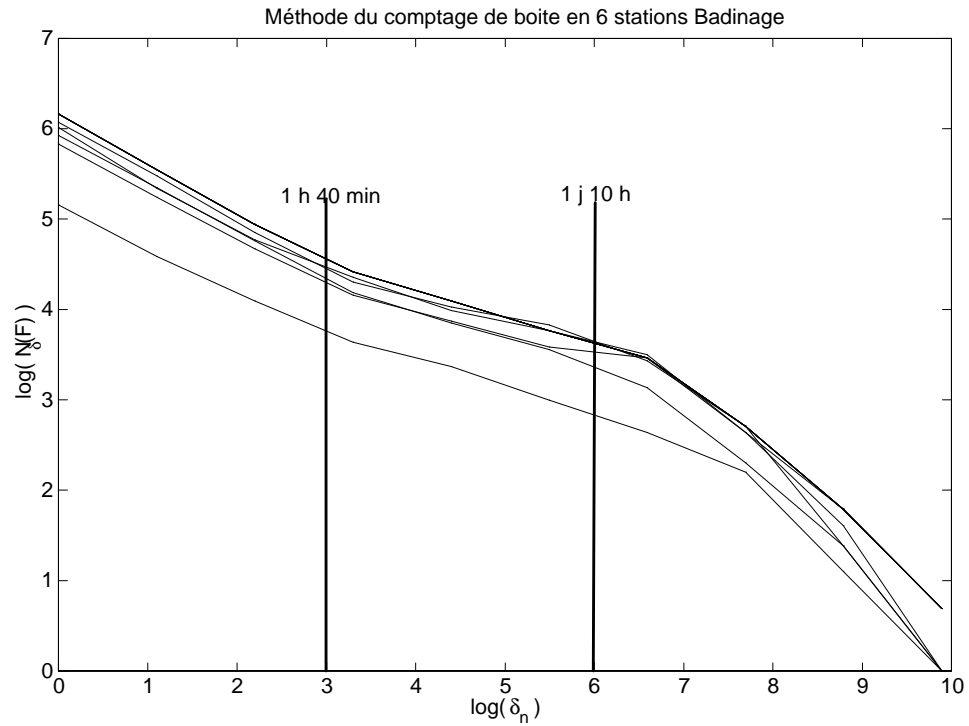


Figure 2.9: Comptage de boîtes sur les données Badinage.

Station	Longueur	Longueur tronquée	$-P_1$	$-P_2$
1	39 376	19 683	0.56	0.29
2	39 383	19 683	0.56	0.32
3	36 791	19 683	0.55	0.21
4	22 334	6 561	0.51	0.30
5	39 394	19 683	0.54	0.26
6	40 769	19 683	0.58	0.26
Moyennes	36 341	17 496	0.55	0.27

Tableau 2.6: Comptage de boîtes sur les données Badinage (P_1 = pente sur $[5min; 1h40min]$ et P_2 = pente sur $[1h40min; 1j10h]$).

m	2	3	5	7
$-P$	0.30	0.29	0.28	0.31

Tableau 2.7: Comptage de boîtes sur la station 1 ($P =$ pente sur $[1h40, 1j10h]$ et m facteur d'échelle).

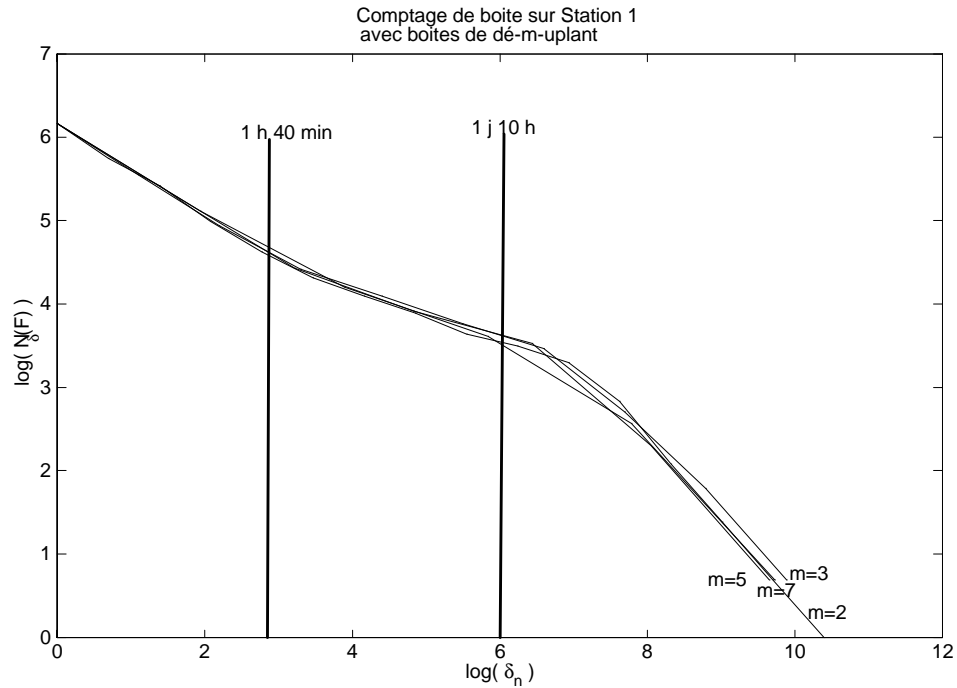


Figure 2.10: Comptage de boîtes sur la station 1.

sur les premiers points (pas de temps de l'ordre de l'heure) du graphe de comptage de boîte à Ouagadougou ([45]).

On vérifie que cette faible pente sur la portion 2 n'est pas due à un artefact de la méthode du comptage de boîtes du même type que ceux rencontrés lors des simulations sur Cantor. En effet, on a vu dans la section précédente que le choix du facteur d'échelle (s'il existe) influe sur l'estimation de la dimension de boîte d'un ensemble de Cantor. On applique donc la méthode du comptage de boîtes sur la station 1 en faisant varier le facteur d'échelle m dans l'ensemble $\{2, 3, 5, 7\}$ (figure 2.10). Les résultats (tableau 2.7) ne varient quasiment pas avec m . Il est possible que la présence d'aléas au sein de la série atténue l'artefact lié au choix du facteur d'échelle.

L'explication la plus plausible de cette sous-estimation reste celle d'un artefact lié à la troncature de la série, artefact d'autant plus présent que les pas de temps sont grands.

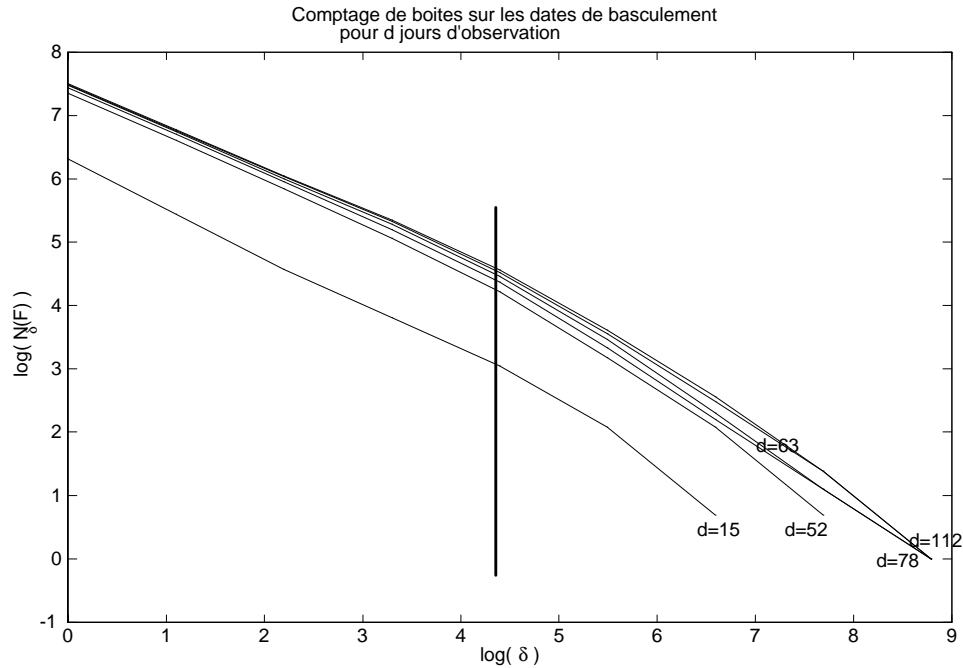


Figure 2.11: Comptage de boîtes sur la série de la Réunion.

Application 2 : Données *Réunion*

L'exploration du comportement fractal de l'occurrence de pluie aux petits pas de temps requiert des données à petits pas de temps. Les données de La Réunion sont des données à basculement d'auget. L'auget bascule dès que le cumul de pluie atteint 0.5 mm et un appareil enregistre les dates de basculement. La résolution temporelle de l'appareil est de 1 seconde, donc l'intensité maximale décelable est de 0.5 mm/s. On applique la méthode du comptage de boîtes au nuage F formé par les dates de basculement représentées sur l'axe du temps (un point pour chaque date exprimée en secondes). La densité de F reflète l'intensité de pluie ou de l'occurrence ou non de hauteurs de pluie dépassant un certain seuil.

Les tailles de boîte varient en se démultipliant par 3, en partant de 15 minutes jusqu'à 70 jours, en faisant varier la durée d d'observation entre 15 et 112 jours (figure 2.11).

La pente sur les trois derniers points (boîtes de taille supérieure à 7 jours) se stabilise à 1 puisqu'il pleut au moins une fois tous les 7 jours.

La pente de la portion de graphe correspondant à des cumuls de pluie sur des pas de temps compris entre 15 min et 1 jour (5 premiers points) décroît avec la longueur de la série mais on constate qu'elle se stabilise à 0.65 (tableau 2.8).

Conclusion : Les séries temporelles d'occurrence de pluie possèdent une structure fractale mais avec des dimensions fractales qui diffèrent avec la gamme de pas de temps considérée. Les périodes de ruptures semblent correspondre à la durée des événements pluvieux.

N	D (en jours)	$-P$
10 000	112	0.65
9 000	78	0.65
8 000	63	0.67
7 000	52	0.68
6 000	40	0.69
2 000	15	0.76

Tableau 2.8: Méthode du comptage de boîtes sur la série de la Réunion (N =Nombre de basculement, D =Durée d’observation P =Pente sur $[2.5min; 6h]$).

L’estimation de la dimension fractale est peu variable entre données sahéliennes (données Badinage et Ouagadougou). Il serait intéressant d’étudier un grand nombre de séries en ces lieux géographiques pour conclure quant à l’invariance spatiale de ce paramètre.

Mais l’interprétation de ces résultats d’estimation fractale du support de l’occurrence de pluie ne doit se faire que relativement à un seuil : celui de mesurabilité des cumuls ou des intensités des séries étudiées. En effet, la dimension fractale d’un ensemble d’occurrence de dépassement d’un seuil s varie avec ce dernier. Il serait utile d’intégrer la prise en compte de l’intensité du phénomène dans l’approche fractale. C’est l’ambition de l’approche multifractale.

2.3.3 Modèle en cascade multifractale

Les modèles en cascades multifractales fournissent les plus simples structures de dépendance retranscrivant des propriétés d’invariance de la distribution des observations. Ce sont des modèles possédant des propriétés multifractales telles qu’elles sont rappelées en *Annexe D* (pour toute précision supplémentaire sur les modèles multifractals généraux, on pourra se référer à l’ouvrage de Schertzer et Lovejoy 93 [76]). Dans cette section, on commence par définir les cascades multiplicatives multifractales avant de constater leur propriété distinctive : l’invariance d’échelle. On dresse ensuite une revue de leurs principales utilisations en de l’hydrologie.

Définition

Une cascade multiplicative multifractale de générateur η est une suite de variables aléatoires $(\mu_{j,k})$ définies par la relation de récurrence (figure 2.12) :

$$\begin{aligned} \mu_0 &= \mu_{0,0} > 0 \\ \text{et } \forall j \geq 1 \quad \forall k = 0 \cdots 2^j - 1 \quad \mu_{j,k} &= \mu_{j-1, \lfloor k/2 \rfloor} \cdot \eta_{j,k} \end{aligned}$$

L’interprétation d’une telle suite est immédiate en hydrologie :

- Supposons que l’on dispose d’un cumul μ_0 sur $[0, T[$ (à l’échelle T ou au niveau 0).
- Au niveau 1, il est désagrégé en $\mu_{1,0}$ sur $[0, \frac{T}{2}[$ et de $\mu_{1,1}$ sur $[\frac{T}{2}, T[$.

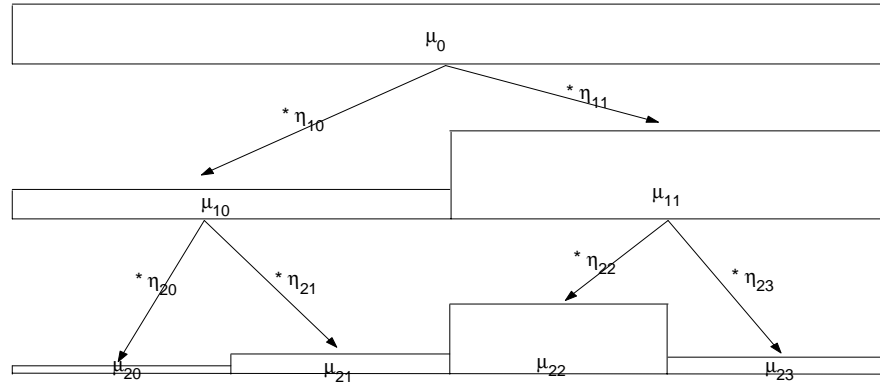


Figure 2.12: Cascade multifractale.

- etc ...
- En désagrégeant jusqu'au niveau d'homogénéité n , on obtient une suite de variables aléatoires $(\mu_{n,k})_{k=0 \dots 2^n - 1}$, les cumulés sont de $\mu_{n,0}$ sur $[0, \frac{T}{2^n}[$, ..., de $\mu_{n,2^n - 1}$ sur $[(2^n - 1) \frac{T}{2^n}, T[$.

Il est important de remarquer qu'en hydrologie, on ne s'intéresse pas à une désagrégation "infinie" (n tend vers l'infini) mais plutôt à un niveau de désagrégation limite N nommé niveau d'homogénéité de la cascade. Pour une étude de l'existence et de la définition d'un processus en cascade (désagrégation infinie), on pourra se reporter aux travaux de Kahane, 85 [47].

Générateurs

Les générateurs $(\eta_{j,k})_{j=1 \dots N, k=0 \dots 2^j - 1}$ sont choisis indépendants deux à deux et identiquement distribués :

$$(\eta_{j,k})_{j=1 \dots N, k=0 \dots 2^j - 1} \text{ iid}$$

Leur distribution f est choisie log – infiniment divisible c'est-à-dire invariante par multiplication (par exemple, la loi log – normale puisque la loi normale est infiniment divisible). Ce choix de distribution, associé à la convention $\mu_0 \equiv 1$, permet d'aboutir à un résultat d'invariance d'échelle. En effet, les variables désagrégées $\mu_{N,i}$ à tout niveau N

suivront toutes la même loi f puisque par itérations :

$$\forall N > 0 \quad \mu_{N,*} = \prod_{j=1}^N \eta_{j,*}$$

Cascades universelles

Ces cascades ont été construites dans le but de détecter et d'estimer d'éventuels paramètres communs à toutes les séries hydrologiques (à différents pas de temps et en différents site, sur des séries de pluie ou de débit) qui seraient donc des invariants directement liés à la phénoménologie des champs de pluie.

Dans ces cascades, la loi du générateur est une loi stable (parfois appelée loi de Lévy, Feller 71 [35]). Ces lois ne sont définies que par leur fonction caractéristique :

$$\Phi_{\alpha,\beta,c,\gamma}(x) = \exp \left\{ i\gamma x - c|x|^\alpha \left[1 + i\beta \frac{t}{|t|} \tan \left(\frac{\pi\alpha}{2} \right) \right] \right\}$$

avec $0 \leq \alpha \leq 2$, $-1 \leq \beta \leq 1$, $c \geq 0$ et $\gamma \in \mathbb{R}$.

Ces distributions (dont la loi normale fait partie puisque $\Phi_{norm}(x) = \exp \{-c|x|^2\}$) peuvent être vues comme des généralisations de la loi normale dans la mesure où elles sont stables⁴ et attractives⁵.

Définition 9 La fonction d'échelle des moment est définie par :

$$K(q) = \log_{2^N} E[\mu_N]$$

On montre que, par le théorème de Lévy Khinchine :

$$\begin{aligned} K(q) &= \frac{C_1}{\alpha - 1} (q^\alpha - q) \text{ si } \alpha \neq 1 \\ K(q) &= C_1 q \log q \text{ si } \alpha = 1 \end{aligned}$$

⁴Stabilité : La loi f_α est dite stable lorsque pour tout échantillon $(X_i)_{1 \leq i \leq n} \stackrel{iid}{\sim} f_\alpha$, on a :

$$\frac{\sum_{i=1}^n X_i}{n^{1/\alpha}} \sim f_\alpha$$

⁵Attractivité : La loi f_α est dite attractive lorsque pour tout échantillon $(X_i)_{1 \leq i \leq n}$ vérifiant :

$$\begin{aligned} E[X_1] &= 0 \text{ et } E[X_1^2] = \infty \\ \exists \alpha < 2, \quad E[X_1^\alpha] = 1 \text{ et } \forall \beta > \alpha \quad E[X_1^\beta] = \infty, \end{aligned}$$

on a : $\frac{\sum_{i=1}^n X_i}{n^{1/\alpha}} \xrightarrow{n \rightarrow \infty} f_\alpha$

Le cas $\alpha = 0$ correspond au modèle mono-fractal β (voir⁶) et le cas $\alpha = 2$ correspond aux cascades log-normales (voir⁷).

Les paramètres α et C_1 , qui caractérisent à eux seuls le processus, sont dits paramètres universels.

Décroissance algébrique des queues de distribution

Une distribution est dite de type algébrique si sa fonction de survie décroît algébriquement :

$$G(x) \stackrel{x \rightarrow \infty}{\propto} x^{-q} \text{ (ou } \lim_{x \rightarrow \infty} \frac{G(x)}{x^{-q}} = c^{te})$$

On a donc en particulier le résultat suivant : Si le générateur suit une loi de type algébrique invariante par multiplication, alors la cascade ainsi construite est une suite de variables aléatoires de type algébrique à tout niveau d'agrégation. Le paramètre de décroissance q est en particulier invariant d'échelle⁸.

2.3.4 Revue bibliographique de modèles hydrologiques multifractals

L'emploi des fractals et des multifractals pour la modélisation n'a commencé que récemment. Ils attirent à la fois les physiciens pour qui la notion d'invariance d'échelle est familière et les statisticiens qui y trouvent le moyen d'intégrer une très forte variabilité dans une structure de dépendance relativement simple.

Les estimations du paramètre de décroissance algébrique q_D se recoupent entre les divers auteurs et les différents pas de temps (tableau 2.9) : Elles sont comprises entre 2.1 et 3.6 pour des cumuls de pluie à pas de temps supérieur à 15 minutes mais diminuent nettement pour les pas de temps très petits (dans les travaux de Rodriguez-Iturbe, 91 [72], $q_D = 1.5$ pour un pas de temps de 15 secondes).

⁶C'est le plus simple des modèles en cascade multifractale. Le générateur η suit une loi de Bernoulli :

$$\begin{aligned} P(\eta = \lambda^c) &= \lambda^{-c} \\ P(\eta = 0) &= 1 - \lambda^{-c} \end{aligned}$$

⁷Dans ce modèle en cascade multifractale, le générateur η suit une loi log-normale :

$$\begin{aligned} \eta &= e^\theta \\ \theta &\sim \text{Normale} \end{aligned}$$

⁸Par exemple, si le générateur est log-Gamma,

$$\eta \sim \log \Gamma(n, q) \text{ (ou bien } \log \eta \sim \Gamma(n, q))$$

alors, à tout niveau N de désagrégation :

$$\mu_{N,*} = \prod_{j=1}^N \eta_{j,*} \sim \log \Gamma(Nn, q)$$

Les cumuls désagrégés suivent tous une loi log-Gamma de même paramètre de décroissance algébrique q .

Des valeurs du même ordre de grandeur sont trouvée sur les séries de débit. L'existence d'un paramètre invariant d'échelle et spatial, lié à la nature même du phénomène pluvieux n'est donc pas inacceptable.

	$\widehat{q_D}$	débit/pluie	Pas de temps	Site
Bolgov et al. [13],1999		débit	Max annuels	Russie
Tessier et al. [80],1996	2.7	débit	1 jour	France
Pandey et al. [69],1998	3.1	débit	8 jours	EU
Hubert et Carbonnel [44],1988	3.4	pluie	1 jour	Burkina Faso
Van Monfort and Witter [65],1986	2.9	pluie	1 heure	Pays-Bas
Ladoy et al.[53],1993	3.0	pluie	12 heures	Nîmes
Rodriguez-Iturbe [72],1991	1.5	pluie	15 s	Boston, EU
Georgakakos et al.[39],1994	[1.7 ; 2.9]	pluie	haute résolution	Iowa City, EU
de Lima [25],1998	3.1	pluie	15 min	V. Formoso
	3.6	pluie	1 jour	V. Formoso
	3.6	pluie	1 mois	V. Formoso
	2.6	pluie	1 heure	Assink
	2.1	pluie	15 min	Nancy

Tableau 2.9: Revue bibliographique des modèles multifractals

Mais l'approche de modélisation multifractale n'a pas encore été totalement explorée, notamment au niveau de la critique des résultats d'estimation. C'est l'une des ambitions de ce travail où, dans le chapitre suivant, on se penche sur l'étude de la qualité des estimateurs standards du paramètre de décroissance algébrique q_D , avant de proposer un outil original de détection et d'estimation de comportement algébrique. On verra dans la section suivante que les distributions à décroissance algébrique constituent un cadre théorique bien exploré en probabilités : elles appartiennent au domaine d'attraction de la loi de Fréchet. Bien que leurs propriétés s'apparentent à celles des lois de la famille exponentielle, elles attribuent un poids radicalement différent aux grandes observations, ce qui n'est pas sans conséquences pratiques en hydrologie.