

Université du Québec
INRS-Eau
Chaire en Hydrologie statistique

**MODÉLISATION STOCHASTIQUE DE LA PLUIE À L'AIDE DE MODÈLES
INVARIANTS D'ÉCHELLE**

Par
Hugo Gingras
Baccalauréat en statistique

Mémoire présenté pour l'obtention
du grade de Maître ès sciences (M. Sc.)
en sciences de l'eau

Jury d'évaluation

Examineur externe

Fahima Nekka, chercheure
Centre de Recherches Mathématiques et
Faculté de Pharmacie
Université de Montréal

Examineur interne

Monique Bernier, professeure
INRS-Eau

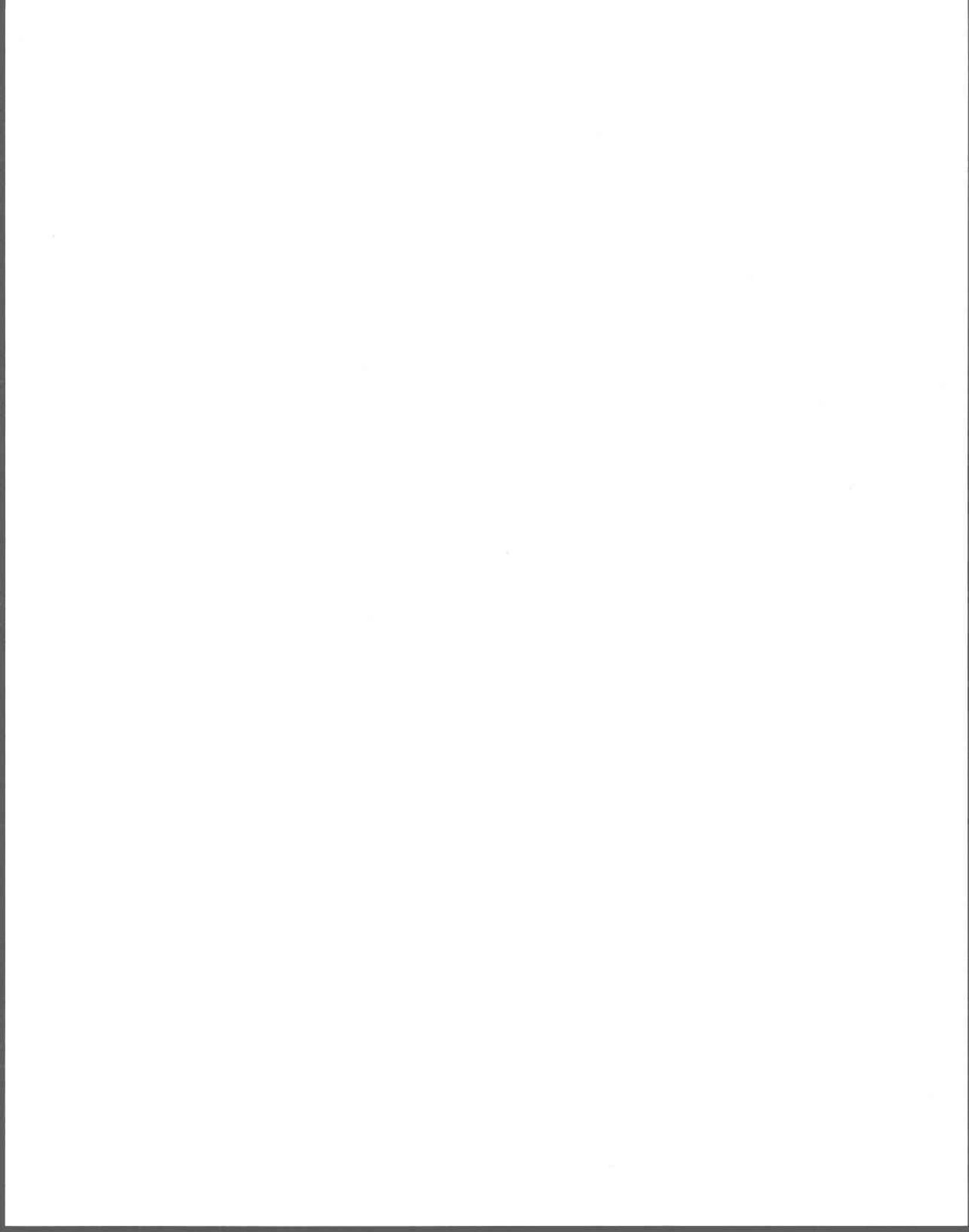
Directeur de recherche

Peter F. Rasmussen, professeur associé
Department of Civil and Geological Engineering
University of Manitoba

Codirecteur de recherche

Bernard Bobée, professeur
Titulaire de la Chaire en Hydrologie statistique
INRS-Eau

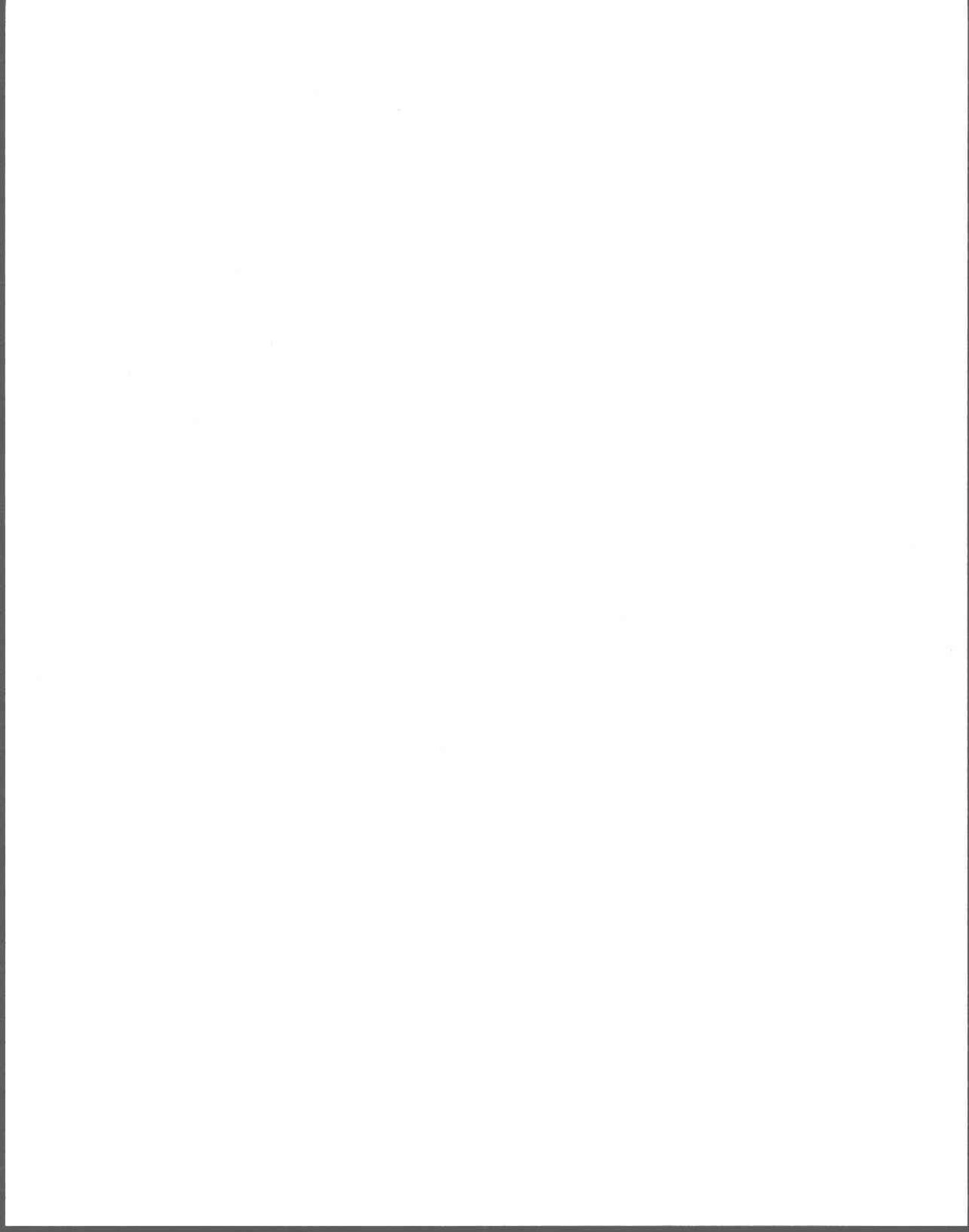
Décembre 2000



REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche, le professeur Peter F. Rasmussen, pour son dévouement et son intérêt soutenu tout au cours de ce projet. J'aimerais également exprimer ma gratitude à l'endroit de mon codirecteur de recherche, le professeur Bernard Bobée, et de toute l'équipe de la Chaire en Hydrologie statistique/Hydro-Québec/CRSNG/Alcan pour leur encadrement scientifique et leur soutien financier. Une mention toute spéciale va à Alin Cârsteanu qui a su prendre la relève après le départ de Peter pour le Manitoba ; sans sa précieuse collaboration, ce mémoire n'aurait sans doute jamais pu voir le jour.

J'adresse des remerciements sincères à l'égard de mes parents et de ma soeur pour le soutien, autant physique que moral, qu'ils ont su m'apporter tout au long de ces trois années. Je tiens finalement à envoyer un salut tout particulier à Nabil, pour son amitié qui a été bien appréciée en cette fin de parcours.



RÉSUMÉ

La modélisation des précipitations atmosphériques a toujours revêtu un intérêt particulier pour les chercheurs, afin de répondre à toutes sortes de problèmes touchant à la gestion de la ressource eau (comme la prévision des événements de précipitation extrêmes pour le dimensionnement des ouvrages hydrauliques) ou encore pour ce qui est de prévenir la population face à l'arrivée de fortes averses de pluie. Or, la grande variabilité spatio-temporelle exhibée par le phénomène a toujours su poser des défis de taille aux scientifiques cherchant à le modéliser.

L'une de ces difficultés concerne la bonne performance des modèles sur plusieurs échelles de temps. Afin de pallier à cette lacune à laquelle se sont butés la plupart de ceux-ci (notamment les modèles se basant sur les processus ponctuels), une nouvelle approche a commencé à se répandre au cours des dernières années faisant appel à la propriété d'invariance d'échelle, qui offre une solution à ce problème.

On discute donc dans ce mémoire des aspects théoriques de l'invariance d'échelle, d'une méthode permettant de détecter sa présence dans les séries de pluie, en plus de développer un modèle simple de simulation s'appuyant sur le principe des cascades multiplicatives. Une section de ce travail est aussi réservée à l'étude de l'effet de l'atome en zéro (*i.e.* la probabilité d'observer des mesures de précipitation nulles) sur le modèle en question.

Ce modèle est testé sur des données horaires en provenance de l'Alabama. Après avoir vérifié la présence de la propriété d'invariance d'échelle sur les données, des simulations ont été effectuées afin de le valider. La performance du modèle est évaluée au niveau de la reproduction des moments statistiques ainsi que de la distribution observée pour les intervalles secs. Les résultats satisfaisants obtenus permettent de conclure à l'utilité de recourir à ce type de modèle pour des applications futures. Quelques perspectives de recherche concernant la prévision des événements extrêmes sont d'ailleurs exposées.

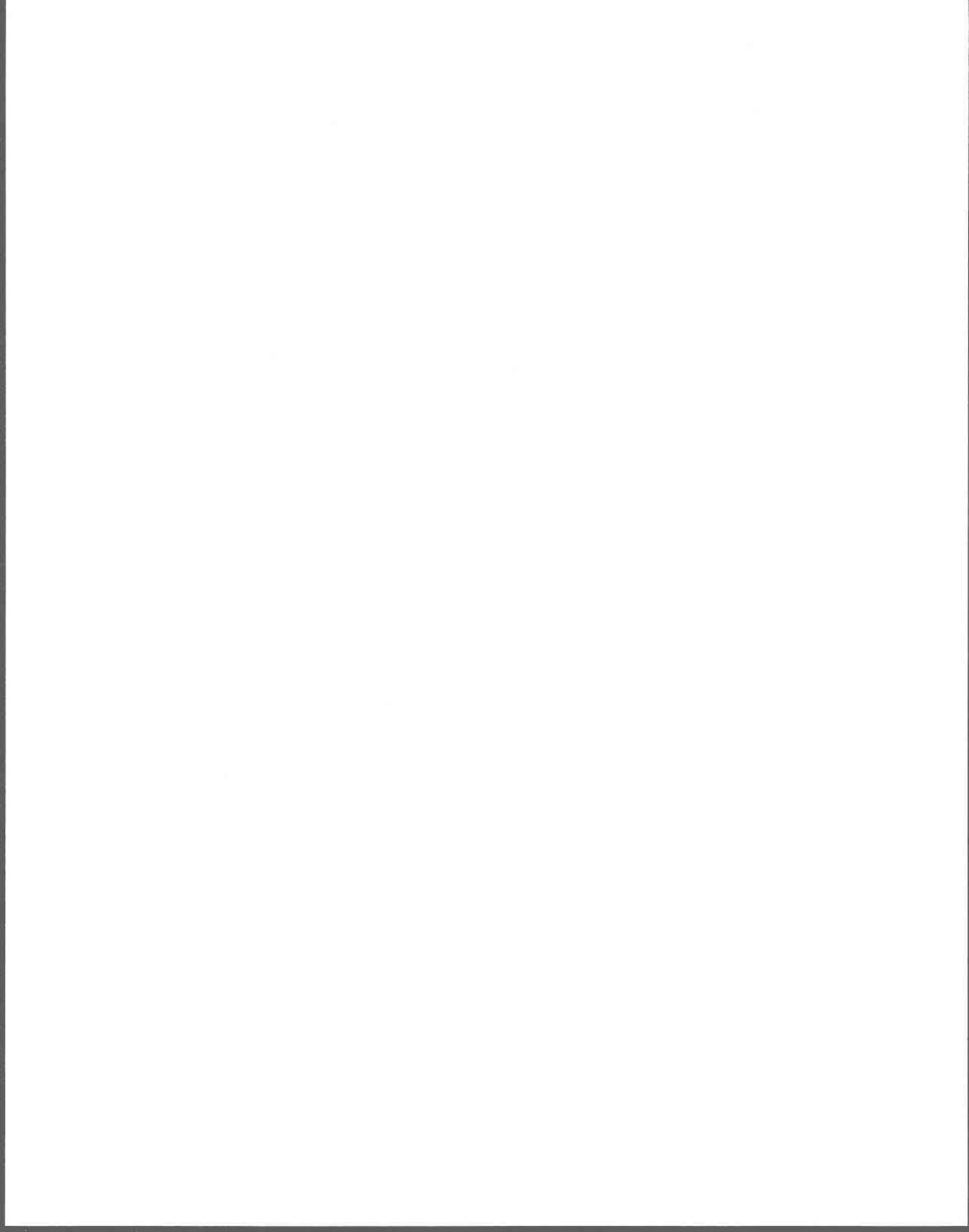


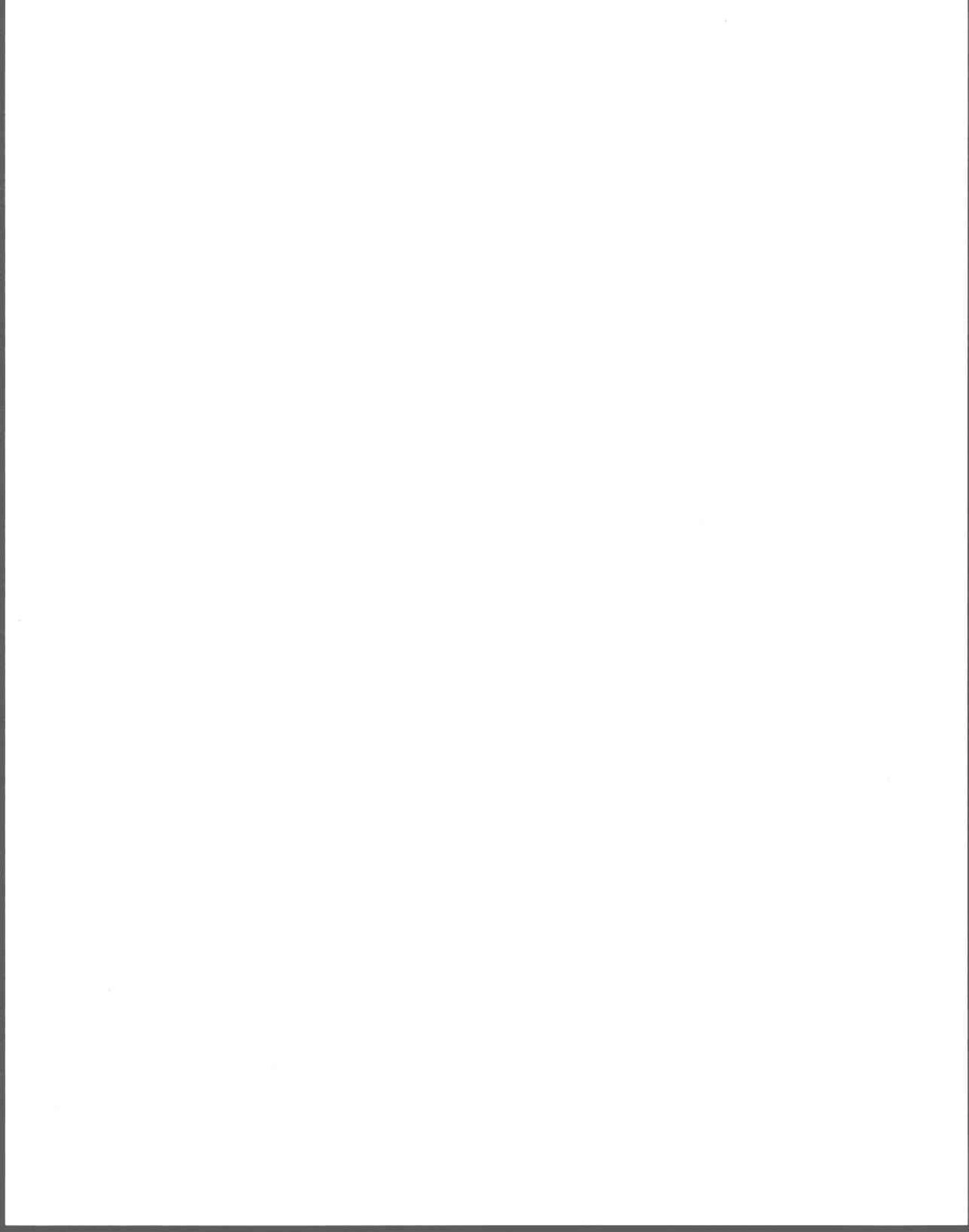
TABLE DES MATIÈRES

REMERCIEMENTS.....	iii
RÉSUMÉ.....	v
TABLE DES MATIÈRES.....	vii
LISTE DES FIGURES.....	ix
LISTE DES TABLEAUX.....	xi
1. INTRODUCTION.....	1
1.1 Caractéristiques du phénomène de précipitation.....	1
1.1.1 Les statistiques	1
1.1.2 Les aspects déterministes	3
1.1.3 L'invariance d'échelle	4
1.2 Approches de modélisation de la précipitation	5
1.2.1 Discussion sur les processus ponctuels.....	5
1.2.1.1 Description d'un processus ponctuel	5
1.2.1.2 Définition du processus ponctuel de Poisson.....	6
1.2.1.3 Représentation canonique d'un processus ponctuel.....	7
1.2.1.4 Processus ponctuels de type groupé (<i>Cluster processes</i>).....	9
1.2.1.5 Exemple d'application : modèle de Poisson.....	10
1.2.2 L'invariance d'échelle et son utilité dans la modélisation.....	14
2. MODÉLISATION DANS LE CADRE DE L'INVARIANCE	
 D'ÉCHELLE	17
2.1 Aspects théoriques.....	17
2.1.1 Invariance d'échelle simple et spectre de Rényi.....	17
2.1.2 Exemple de construction fractale et comptage de boîtes	23
2.1.3 Les cascades multiplicatives monofractales	25
2.2 Vérification de l'hypothèse d'invariance d'échelle sur les données ;	
zones d'applicabilité	29
2.2.1 Démarche méthodologique et présentation des données	29
2.2.2 Estimation de la dimension de capacité	31
2.2.3 Estimation du spectre τ	36
2.3 Description du modèle de précipitation s'appuyant sur les cascades	
multiplicatives.....	43
2.4 Compensation de l'effet de l'atome en zéro sur la distribution des	
poids reconstitués de la cascade	47

3. VALIDATION DU MODÈLE	57
3.1 Les moments statistiques de l'intensité de la précipitation à différentes échelles.....	57
3.2 Les distributions statistiques des périodes sèches à différentes échelles.....	64
4. PERSPECTIVES DE RECHERCHE : LES COURBES IDF (INTENSITÉ-DURÉE-FRÉQUENCE)	69
4.1 Le concept de courbe IDF	69
4.2 Les courbes IDF obtenues à partir des modèles invariants d'échelle	69
5. CONCLUSION.....	73
BIBLIOGRAPHIE.....	75

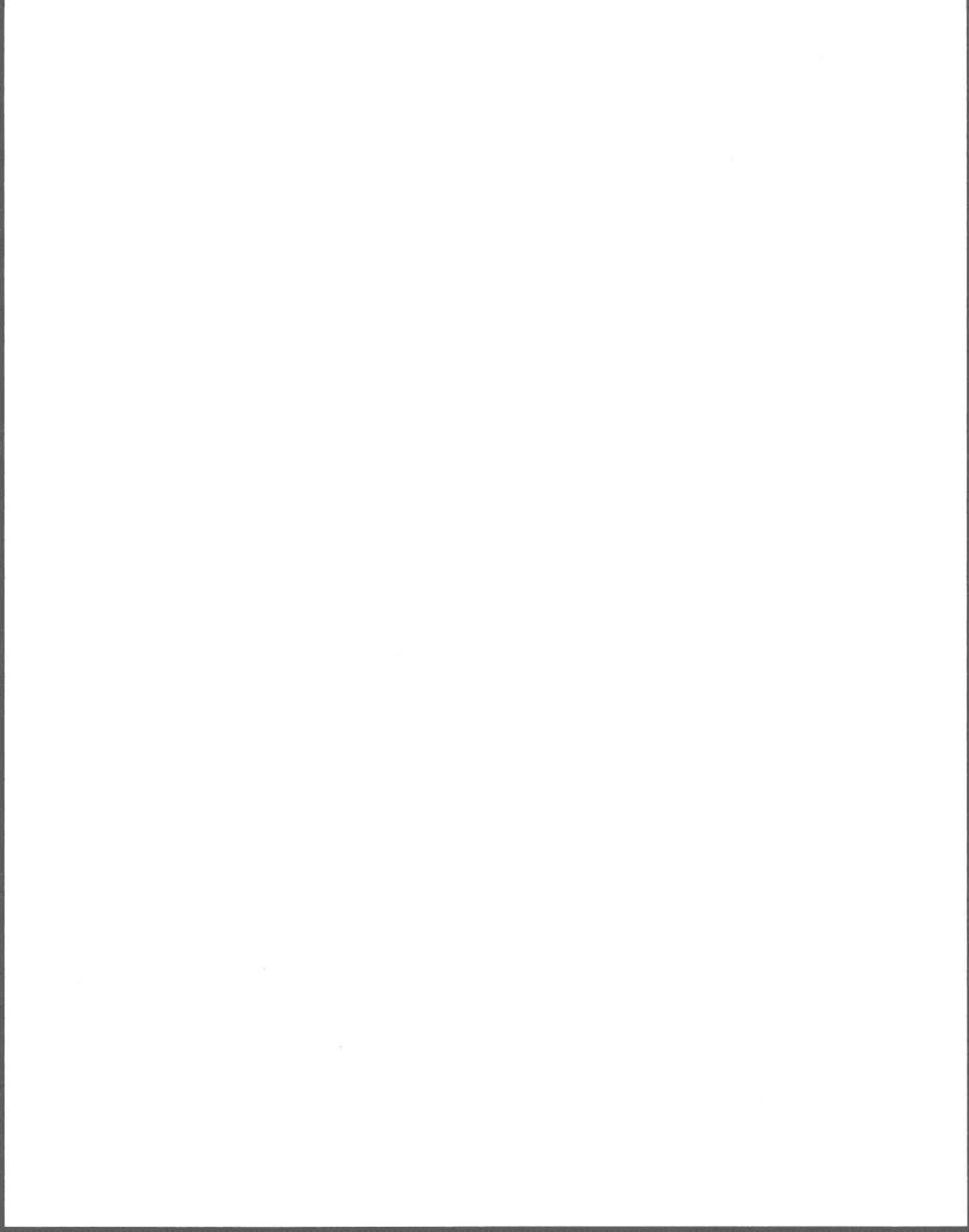
LISTE DES FIGURES

Figure 1.1:	Description schématique du modèle de Poisson (amplitudes indépendantes)	11
Figure 2.1:	Quatre étapes dans la construction d'un ensemble de Cantor.....	24
Figure 2.2:	Schéma d'une cascade à deux branches.....	26
Figure 2.3:	720 relevés pluviométriques horaires d'une station de l'Alabama	30
Figure 2.4:	Fonction d'autocorrélation estimée de la série horaire de l'Alabama	31
Figure 2.5:	Algorithme du comptage de boîtes appliqué aux données horaires de l'Alabama (estimation de la dimension de capacité).....	33
Figure 2.6:	Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 0,5$ sur les données horaires de l'Alabama.....	37
Figure 2.7:	Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 1,5$ sur les données horaires de l'Alabama.....	38
Figure 2.8:	Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 2,0$ sur les données horaires de l'Alabama.....	38
Figure 2.9:	Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 2,5$ sur les données horaires de l'Alabama.....	39
Figure 2.10:	Représentation graphique de l'estimation du spectre τ	42
Figure 2.11:	Schéma du modèle de simulation des séries de précipitation	46
Figure 2.12:	Reconstitution d'une série de données en tant que cascade à deux branches..	47
Figure 2.13:	Construction (a) et reconstruction (b) de trois niveaux d'une cascade à deux branches.....	49
Figure 3.1:	Moments des séries simulées à l'échelle d'une heure	58
Figure 3.2:	Moments des séries simulées à l'échelle de deux heures	59
Figure 3.3:	Moments des séries simulées à l'échelle de quatre heures	59
Figure 3.4:	Moments des séries simulées à l'échelle de huit heures	60
Figure 3.5:	Moments des séries simulées à l'échelle de seize heures	60
Figure 3.6:	Moments des séries simulées à l'échelle de 32 heures	61
Figure 3.7:	Moments des séries simulées à l'échelle de 64 heures	61
Figure 3.8:	Moments des séries simulées à l'échelle de 128 heures	62
Figure 3.9:	Moments des séries simulées à l'échelle de 256 heures	62
Figure 3.10:	Moments des séries simulées à l'échelle de 512 heures	63
Figure 3.11:	Étude des distributions empiriques de la longueur des périodes sèches, pour différentes échelles d'agrégation, associées aux données de l'Alabama	65
Figure 3.12:	Étude des distributions empiriques de la longueur des périodes sèches, pour différentes échelles d'agrégation, associées aux vingt séries synthétiques de précipitation.....	66
Figure 3.13:	Relevés pluviométriques de l'Alabama vs hauteurs de pluie simulées (9600 premières observations présentées)	67



LISTE DES TABLEAUX

Tableau 2.1: Estimation du spectre τ pour la série de données horaires de l'Alabama.....	41
Tableau 2.2: Étude de la distribution des poids reconstitués (en fonction de l'échelle d'agrégation) de la série de données pluviométriques de la station de l'Alabama	54
Tableau 3.1: Étude de la distribution des poids reconstitués pour les vingt séries synthétiques sensées reproduire les caractéristiques de la série de données pluviométriques de l'Alabama.....	64



1. INTRODUCTION

1.1 Caractéristiques du phénomène de précipitation

1.1.1 Les statistiques

La précipitation atmosphérique est une composante importante de tout modèle de prévision hydrologique. Afin de la modéliser, il importe de bien comprendre sa structure. Dans une perspective de gestion de la ressource eau, des éléments tels l'occurrence des événements extrêmes ainsi que la distribution des périodes sèches (*i.e.* les temps d'attente entre deux averses) peuvent donc prendre une importance considérable.

Or, les différentes manifestations du phénomène de précipitation atmosphérique présentent une grande variabilité dans le temps et l'espace. Si différentes caractéristiques statistiques peuvent être dégagées selon le type de climat, le problème consistant à déterminer des distributions de probabilité marginales et conjointes permettant de bien cerner tous les aspects du phénomène se révèle très complexe. En effet, la dépendance spatio-temporelle commune aux séries de précipitations rend très difficile l'obtention des distributions conjointes explicites (*i.e.* les fonctions exprimant les probabilités d'observer simultanément divers niveaux de précipitation en différents points de l'espace spatio-temporel).

Cela n'empêche toutefois pas de modéliser celles-ci par diverses approximations. De nombreuses pistes de solutions ont été proposées dans la littérature. Sans vouloir en faire une longue énumération, il est possible cependant de faire ressortir certains traits communs qui semblent incontournables lorsque vient le temps de choisir une distribution marginale de probabilité pour la pluie (la variable étudiée étant alors l'intensité de la précipitation).

Tout d'abord, avant de s'attaquer à l'intensité de la précipitation, on peut voir le phénomène sous l'angle d'un processus stochastique à deux états, *i.e.* soit il pleut, soit il ne pleut pas. La conséquence de cette affirmation est de dire qu'il existe une probabilité non nulle d'observer une intensité de précipitation égale à zéro, d'où l'impossibilité de considérer une distribution de type continu uniquement. On doit donc se tourner vers une distribution marginale qui se divise en deux composantes : l'une discrète (pour les valeurs nulles) et l'autre continue (pour les valeurs supérieures à zéro).

On remarque ensuite que la distribution des intensités positives est asymétrique. Cela mène au problème de la queue de la distribution, qui lui-même fait référence au problème de l'estimation des événements extrêmes. Deux points de vue sont alors régulièrement confrontés dans la littérature. L'un stipule que la décroissance de la distribution pour les grandes valeurs se fait de façon exponentielle, ou même plus rapidement encore (comme dans le cas de la loi normale). D'autres diront en revanche que le comportement exhibé est plutôt du type hyperbolique, ce qui implique une divergence des moments d'ordre élevé (*i.e.* que les estimateurs de ces moments augmentent avec la taille de l'échantillon). La décroissance de la fonction de distribution se fait alors de façon beaucoup plus lente, avec pour résultat une fréquence observée plus grande pour les événements extrêmes (Bendjoudi et Hubert [1998]).

La caractérisation de la distribution pour les événements extrêmes se révèle d'une importance cruciale pour des fins d'ingénierie hydraulique par exemple, lorsque vient le temps de planifier la construction de nouveaux ouvrages. En effet, l'allure des courbes IDF (Intensité-Durée-Fréquence), utilisées pour le design de telles constructions, sera bien différente selon le type de distribution choisie (queue de type hyperbolique *vs* exponentielle). Comme leur nom l'indique, ces courbes relient l'intensité i à la durée d de même qu'à la période moyenne de retour T , période qui se définit comme étant l'inverse de la fréquence (*i.e.* l'inverse de la probabilité au dépassement de l'événement sur une unité de temps). Étant donné que la probabilité au dépassement $P_{\text{dép}}$ pour les événements extrêmes

est plus grande dans le cas hyperbolique (Hyp) que dans le cas exponentiel (Exp), on observe donc la relation suivante concernant les périodes de retour:

$$\left[P_{\text{dép}} \right]_{\text{Hyp}} > \left[P_{\text{dép}} \right]_{\text{Exp}} \Rightarrow T_{\text{Hyp}} < T_{\text{Exp}} \quad (1.1)$$

1.1.2 Les aspects déterministes

Bien qu'issus de processus physiques déterministes, la complexité des interactions que l'on retrouve au sein des phénomènes atmosphériques fait en sorte que leur comportement peut, à la limite, être assimilé à un processus aléatoire. En termes de systèmes dynamiques, un système aléatoire est un système avec une infinité de degrés de liberté. Ainsi, un système avec un nombre relativement élevé de degrés de liberté peut, à de certaines fins, être assimilé à un système avec une infinité de degrés de liberté (*i.e.* un processus aléatoire).

La question qui se pose alors est de savoir quand il est possible de formuler un modèle mathématique déterministe permettant de reproduire le comportement apparemment désordonné du phénomène ; quand, par contre, il est recommandable d'utiliser un modèle stochastique seulement ; et quand, notamment, on peut utiliser une combinaison des deux? Le tout en notant que la fonction de distribution des valeurs d'une variable est nécessairement une des distributions invariantes à la transformation temporelle du système dynamique dont elle fait partie.

Un modèle de la pluie doit se caractériser tout d'abord par des changements brusques dans la série, ce que l'on dénomme l'intermittence. Puis, de façon opposée, on dénote également une certaine persistance dans les séries de pluie, ce qui se traduit par une décroissance lente de la fonction d'autocorrélation, possiblement liée à la notion d'invariance d'échelle qui sera expliquée dans ce qui suit. En somme, on se retrouve avec des séries qui présentent des tendances soutenues sur certaines périodes de temps, mais qui peuvent également se retrouver perturbées à tout moment.

Reproduire de telles interactions au sein d'un système dynamique requiert que ce dernier soit non linéaire et qu'il soit très sensible aux conditions initiales. Même s'il est bien connu que des systèmes chaotiques assez simples peuvent donner naissance à de telles propriétés (voir, *e.g.*, Devaney [1992]), il semble que pour les champs de précipitation, une combinaison de processus stochastiques dans une structure déterministe soit le meilleur moyen d'atteindre cet objectif, ceux-ci disposant d'un nombre élevé de degrés de liberté. Il s'agit là de l'idée qui sera développée dans la suite de ce mémoire.

1.1.3 L'invariance d'échelle

L'étude de la propriété d'invariance d'échelle dans les phénomènes hydrologiques est une façon de répondre aux problèmes concernant la compréhension, la mesure et la prédiction de ces processus sur une large gamme d'échelles. Ce concept est relié à la propriété d'auto-similarité (Mandelbrot [1989]), quoique celle-ci ait surtout une connotation géométrique. De manière abrégée, cela veut dire que le degré d'irrégularité observé dans une série chronologique sera statistiquement semblable à toutes les échelles, moyennant un facteur d'échelle. C'est donc dire que si l'on fait un zoom sur une certaine zone de la série, on peut s'attendre à retrouver le même genre de fluctuations dans cette zone que celles observées à plus grande échelle. Autrement dit, c'est comme si la structure générale de la série se répétait statistiquement à l'intérieur de celle-ci, ce qui implique qu'une composante de la variabilité de la série peut être caractérisée indépendamment de l'échelle d'observation.

En développant ainsi des modèles invariants d'échelle, on en arrive à reproduire des systèmes qui semblent évoluer naturellement dans le temps et l'espace vers un état "d'auto-organisation" critique, caractérisé par des lois de puissance.

1.2 Approches de modélisation de la précipitation

1.2.1 Discussion sur les processus ponctuels

1.2.1.1 Description d'un processus ponctuel

Alors qu'un processus stochastique représente une famille de variables aléatoires indexées par l'ensemble des nombres réels (ou entiers), un champ aléatoire se définit de façon analogue comme une collection de variables aléatoires indexées par les éléments des sous-ensembles d'un espace dans lequel le phénomène aléatoire évolue (*e.g.*, \mathbb{R}^n , $n = 1, 2, 3, \dots$). Si les variables en question ne peuvent prendre que des valeurs positives et entières, on se retrouve alors avec ce qu'il a été convenu d'appeler dans la littérature un processus ponctuel.

Pour en donner une définition plus formelle, disons simplement qu'un processus ponctuel $N(\cdot)$ représente une famille de variables aléatoires à valeurs positives et entières $N(A)$, indexées par des sous-ensembles A d'un espace approprié \mathfrak{S} , et satisfaisant les conditions suivantes:

$$(I) \quad N(\emptyset) = 0 \quad (\emptyset \text{ représentant l'ensemble vide})$$

$$(II) \quad P[N(A) < \infty] = 1 \quad (\text{si } A \text{ est un ensemble borné})$$

$$(III) \quad P\left[N\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} N(A_i)\right] = 1 \quad (\text{pour des ensembles mutuellement disjoints } A_1, A_2, \dots)$$

De façon plus concrète, on peut voir $N(A)$ comme étant la fréquence observée d'un certain phénomène aléatoire dans un intervalle ou une région $A \subset \mathfrak{S}$. Il est clair que dans un contexte d'application à des modèles de précipitation, l'espace \mathfrak{S} considéré est \mathbb{R}_+ (*i.e.* le temps) ou encore l'espace Euclidien à n dimensions \mathbb{R}^n ($n \leq 4$), pour les modèles espace-temps.

Pour un traitement plus détaillé de la théorie des processus ponctuels, le lecteur est référé, *e.g.*, à Cox et Isham [1980], qui offrent une approche concrète axée sur les applications, à Waymire et Gupta [1981a, 1981b, 1981c], qui se concentrent sur les applications aux modèles de précipitation, ainsi qu'à Daley et Vere-Jones [1988] pour une approche plus approfondie s'appuyant sur la théorie des mesures.

1.2.1.2 Définition du processus ponctuel de Poisson

De façon générale, il est possible de saisir la structure probabiliste complète d'un processus ponctuel à travers la spécification des distributions conjointes des nombres de points dans un nombre fini d'ensembles arbitraires et (mutuellement) disjoints. Dans le cas d'un processus de Poisson $N(\cdot)$ de paramètre d'intensité λ , $\lambda > 0$, cela revient à dire que pour tout $k = 1, 2, \dots$, les variables $N(A_1), \dots, N(A_k)$ sont indépendantes pour des sous-ensembles disjoints A_1, \dots, A_k de \mathbb{R} . De plus, ces variables sont distribuées selon des lois de Poisson de moyennes respectives $\lambda|A_1|, \dots, \lambda|A_k|$, où $|A_i|$ ($i = 1, \dots, k$) représente la longueur de l'intervalle A_i . On peut noter les similitudes entre la définition du processus ponctuel de Poisson et celle du processus stochastique de Poisson.

Le processus de Poisson se caractérise également à l'aide de ses propriétés fondamentales, qui s'énumèrent comme suit:

- (i) complètement aléatoire: $N(A_1), \dots, N(A_k)$ sont indépendantes pour des régions disjointes A_1, \dots, A_k , pour tout $k > 1$ (il s'agit de l'équivalent de la propriété d'incrément indépendants dans la théorie des processus stochastiques)
- (ii) sans événements simultanés: $P[N(A) \geq 2] = o(|A|)$ pour $|A| \rightarrow 0$
- (iii) homogène ou stationnaire: $P[N(A) = 1] = \lambda|A| + o(|A|)$ pour $|A| \rightarrow 0$ et $\lambda > 0$

La propriété (iii) montre que l'homogénéité du processus de Poisson découle du fait que l'on suppose le paramètre d'intensité λ comme étant constant dans le temps. Un processus

satisfaisant aux conditions (i)-(iii) mais pour lequel le paramètre λ est fonction du temps, *i.e.* $\lambda = \lambda(t)$, est appelé un processus de Poisson non homogène avec fonction d'intensité $\lambda(t)$. Pour un tour d'horizon plus général concernant les processus ponctuels poissonniens, le lecteur est référé à Abi-Zeid [1997].

Enfin, en ne considérant ce processus que dans une seule dimension (le temps) et puisque l'arrivée d'événements simultanés a une probabilité nulle, il devient également possible de le spécifier à l'aide de la distribution conjointe des temps d'attente X_j ($j = 1, 2, \dots$) entre chaque événement. En fait, il suffit de dire que les temps d'attente X_j sont indépendants et identiquement distribués selon une loi exponentielle de paramètre λ pour caractériser un processus ponctuel de Poisson de paramètre d'intensité λ , $\lambda > 0$.

1.2.1.3 Représentation canonique d'un processus ponctuel

Dans le cas d'un processus ponctuel défini sur la droite réelle, l'espace des réalisations Ω peut être vu comme l'ensemble des différentes positions que peuvent prendre les points (correspondant à l'arrivée des événements) sur la droite. Un élément w de Ω se représente donc comme une suite de nombres réels, *i.e.* que:

$$w = \{x_i\}_{i=1}^{\infty} \quad (x_1 \leq x_2 \leq \dots) \quad (1.2)$$

On suppose qu'une telle suite respecte la condition (II) de la définition d'un processus ponctuel (section 1.2.1.1), *i.e.* qu'elle ne peut contenir qu'un nombre fini de points dans tout sous-ensemble A borné de \mathbb{R} . On est alors en mesure d'affirmer que w représente une réalisation possible du processus $N(\cdot)$.

Or, on peut associer une mesure de dénombrement à toute réalisation possible $w \in \Omega$. Cette mesure se définit ainsi (Waymire et Gupta [1981b]):

$$N(w, A) = \text{card}\{i: x_i \in A\}, \quad A \subset \mathbb{R}, \quad w = \{x_i\}_{i=1}^{\infty} \quad (1.3)$$

où *card* dénote le cardinal de l'ensemble en question. Cela revient à dire que pour une réalisation w donnée du processus $N(\cdot)$, $N(w, A)$ représente le nombre de points se retrouvant dans l'ensemble A . L'utilité d'une telle définition réside dans le fait qu'elle permet d'introduire la notion d'intégrale stochastique. La notion d'intégrale d'une fonction $f(\cdot)$ définie sur \mathbb{R} est bien connue. Imaginons cependant une fonction $f(\cdot)$, qui au lieu d'être définie sur \mathbb{R} , n'est définie que sur un sous-ensemble infini (quoique dénombrable) de \mathbb{R} . Autrement dit, une fonction $f(\cdot)$ définie sur un sous-ensemble $w \in \Omega$ représentant une réalisation possible d'un processus ponctuel $N(\cdot)$. On peut comprendre intuitivement que l'intégrale d'une telle fonction sur w correspond en fait à une somme infinie de termes $f(x_i)$ pour chaque élément $x_i \in w$. L'expression d'une telle intégrale (dite stochastique) par rapport au processus ponctuel $N(\cdot)$ peut être déduite en se servant de la définition (1.3), ce qui se traduit de la façon suivante:

$$\int_{\mathbb{R}} f(x) dN(w, x) = \sum_{i=1}^{\infty} f(x_i), \quad w = \{x_i\}_{i=1}^{\infty} \in \Omega \quad (1.4)$$

où l'expression $dN(w, x)$ vaut 1 si un événement s'est produit dans l'intervalle $[x, x+dx]$ ou 0 sinon. On peut donc définir la fonction $f(\cdot)$ indépendamment sur \mathbb{R} . On utilise habituellement l'expression (1.4) sous cette notation abrégée:

$$\int_{\mathbb{R}} f(x) dN(x) = \sum_{i=1}^{\infty} f(x_i) \quad (1.5)$$

L'équivalence entre l'intégrale et la somme infinie de l'équation (1.5) peut aussi être mise en évidence en remarquant que:

$$\frac{dN(x)}{dx} = \sum_{i=1}^{\infty} \delta_{x_i}(x) \quad (1.6)$$

où δ représente la distribution de Dirac. Puisque l'on se retrouve avec une somme infinie en (1.5), des problèmes de convergence peuvent se poser. Une façon de contourner cela est de dire que $f(\bullet) \equiv 0$ en dehors d'un sous-ensemble A borné de \mathbb{R} , en plus de poser la condition que $f(\bullet)$ est elle-même bornée. Dans ce cas, la condition (II) de la section 1.2.1.1 nous assure que $f(x_i) \neq 0$ que pour un nombre fini de points, et donc que l'intégrale définie en (1.5) existe. Notons également que dans la mesure où $f(\bullet) \equiv 1$ sur \mathbb{R} , l'expression (1.5) nous permet de réexprimer la variable $N(A)$, $A \subset \mathbb{R}$, sous la forme:

$$N(A) = \int_A dN(x) = \sum_{x_i \in A} 1 \quad (1.7)$$

1.2.1.4 Processus ponctuels de type groupé (*Cluster processes*)

Les précipitations se présentent généralement sous forme groupée. Sans entrer dans les détails, c'est que l'avènement d'un front pluvieux amène avec lui son lot d'averses. Une référence fondamentale sur le sujet, souvent citée dans la littérature, concerne les travaux de Le Cam [1961]. Kavvas et Delleur [1981], par exemple, se basent sur ce dernier afin de construire leur modèle stochastique pour représenter des données journalières de pluie provenant de l'Indiana.

Ce genre de dépendance a donc amené les chercheurs à considérer un type de processus ponctuel que l'on qualifie de groupé. Soit $N_1(\cdot)$ un processus ponctuel sur \mathbb{R} que l'on dénote comme le processus des centres des groupes. Définissons également, pour tout $x \in \mathbb{R}$, un processus $N_2(\cdot | x)$ que l'on dénote comme le processus des membres du groupe dont le centre est situé au point x . Supposons que $N_2(\cdot | x_1), \dots, N_2(\cdot | x_k)$ représentent k processus indépendants pour k valeurs distinctes $x_1, \dots, x_k \in \mathbb{R}$ ($k = 1, 2, \dots$). Un processus

$N(\cdot)$ résultant de la superposition de ces divers processus est alors dit être de type groupé et peut s'exprimer ainsi:

$$N(\cdot) = \sum_{x_i \in N_1(\cdot)} N_2(\cdot | x_i) \quad (1.8)$$

Bien que définis de façon générale, les processus ponctuels groupés ne sont intéressants en pratique que dans certaines situations spéciales. Dans les cas d'application à des modèles de précipitation, on s'intéresse tout particulièrement aux processus ponctuels groupés dits de Poisson, *i.e.* dont le processus des centres des groupes est poissonnien. Un exemple simple d'un tel processus ponctuel groupé est le processus de Poisson composé au sein duquel un nombre aléatoire d'événements se produit au centre de chaque groupe. On suppose que les nombres de points pour chaque groupe sont indépendants et identiquement distribués.

Le processus de Neyman-Scott est un autre exemple de processus groupé qui présente un intérêt particulier dans la modélisation des précipitations. Comme son nom l'indique, ce processus a été introduit par Neyman et Scott [1958] afin de modéliser la distribution des galaxies dans l'espace (les galaxies étant vues comme des regroupements). En bref, il s'agit d'un processus groupé de Poisson pour lequel un nombre aléatoire de points sont dispersés de façon indépendante par rapport aux centres des groupes, et ce en accord avec une distribution de distance commune $f(\cdot)$. On suppose là aussi que les tailles (*i.e.* le nombre de points) des groupes sont indépendantes et identiquement distribuées.

1.2.1.5 Exemple d'application : modèle de Poisson

Ce modèle ne présente aucune dépendance statistique dans sa structure. On y assume que l'intensité de la précipitation consiste en une série d'événements pluvieux instantanés. La position dans le temps de ces divers événements est gouvernée par un processus de Poisson $N(t)$ de paramètre λ , d'où $E[N(t)] = \lambda t$ (section 1.2.1.2). À chaque événement est associée

une amplitude (hauteur de pluie) aléatoire U_i , de telle sorte que U_1, U_2, U_3, \dots représentent des variables aléatoires indépendantes et identiquement distribuées ayant pour fonction de densité $f_U(u)$. Ces variables sont également considérées comme indépendantes de $N(t)$. Ainsi l'intensité de la précipitation $\varepsilon(t)$ au temps t peut s'écrire sous la forme:

$$\varepsilon(t) dt = U(t) dN(t) \quad (1.9)$$

où $dN(t)$ vaut 1 si un événement survient dans l'intervalle $[t, t+dt]$ ou zéro sinon (voir section 1.2.1.3). Une illustration schématique de ce modèle est donnée à la figure 1.1.

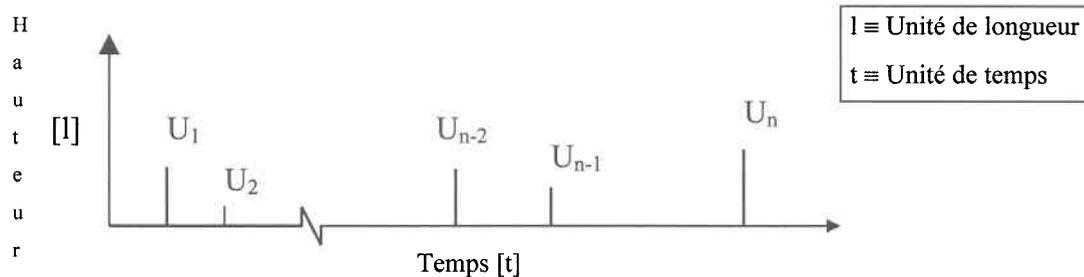


Figure 1.1: Description schématique du modèle de Poisson (amplitudes indépendantes)

On définit par la suite à partir de (1.9), de même qu'en utilisant l'équation (1.5) dérivée à la section 1.2.1.3, la variable $N_U(t)$ donnant le niveau de pluie accumulée au temps t :

$$N_U(t) = \int_0^t \varepsilon(v) dv = \sum_{n=0}^{N(t)} U_n, \quad t > 0, \quad U_0 = 0 \quad (1.10)$$

En tenant compte des hypothèses formulées précédemment, $N_U(t)$ représente donc un processus de Poisson composé (*Compound Poisson Process*). Les expressions pour la moyenne, la variance et la covariance de ce processus sont bien connues (voir, e.g., Parzen [1962]) et sont données par:

$$E[N_U(t)] = E[U]\lambda t \quad (1.11)$$

$$\text{Var}[N_U(t)] = E[U^2]\lambda t \quad (1.12)$$

$$\text{Cov}[N_U(t_1), N_U(t_2)] = E[U^2]\lambda \cdot \min(t_1, t_2) \quad (1.13)$$

où $E[U]$ représente la hauteur de pluie moyenne associée à un événement pluvieux. Il est possible d'en déduire les caractéristiques pour les hauteurs de pluie cumulées sur des périodes de temps successives (et disjointes) de longueur T . En effet, en posant:

$$Y_i = N_U(iT) - N_U((i-1)T), \quad i = 1, 2, \dots$$

les variables Y_i ainsi définies sont indépendantes et identiquement distribuées. En dénotant leur fonction de densité commune par $f_Y(y)$, il est assez simple de démontrer à partir des formules (1.11), (1.12) et (1.13) que les moments du premier et du second ordre de Y s'expriment sous la forme:

$$E[Y] = E[U]\lambda T \quad (1.14)$$

$$\text{Var}[Y] = E[U^2]\lambda T \quad (1.15)$$

$$\text{Cov}[Y_i, Y_j] = 0 \quad (i \neq j) \quad (1.16)$$

Afin de tester les propriétés de ce modèle à différentes échelles de temps T , il faut tout d'abord obtenir des estimations pour les paramètres λ , $E[U]$ et $E[U^2]$. Pour limiter le nombre de paramètres à estimer, on peut supposer que les amplitudes U_i sont distribuées selon une loi exponentielle de paramètre σ , *i.e.* que $f_U(u) = \sigma e^{-\sigma u}$. La moyenne et la variance de U sont alors données par $E[U] = 1/\sigma$ et $\text{Var}[U] = 1/\sigma^2$. Or, en se servant du fait que $\text{Var}[U] = E[U^2] - (E[U])^2$, on en déduit que $E[U^2] = 2/\sigma^2$, *i.e.* que $E[U^2] = 2(E[U])^2$, ce qui nous permet de réécrire les équations (1.14) et (1.15) en fonction des paramètres λ et $E[U]$ uniquement, d'où l'on obtient ces expressions pour λ et $E[U]$:

$$\left\langle \begin{array}{l} E[Y] = E[U]\lambda T \\ \text{Var}[Y] = 2(E[U])^2 \lambda T \end{array} \right\rangle \Leftrightarrow E[U] = \frac{\text{Var}[Y]}{2E[Y]} \quad (1.17)$$

$$\lambda = \frac{2(E[Y])^2}{\text{Var}[Y] \cdot T} \quad (1.18)$$

Rodriguez-Iturbe *et al.* [1984] soulignent l'importance de choisir des intervalles de temps considérés comme homogènes d'un point de vue hydrologique pour l'analyse des données, de façon à ce que l'estimation des paramètres garde un sens. En effet, la valeur des paramètres peut changer au fil des mois ou des saisons par exemple. Pour chaque mois ou chaque saison, on commence tout d'abord par estimer la moyenne $E[Y]$ et la variance $\text{Var}[Y]$ à partir de la séquence des hauteurs de pluie cumulées Y_i à notre disposition. Puis, en admettant une distribution exponentielle pour U , on peut ensuite estimer les paramètres λ et $E[U]$ à l'aide des équations (1.17) et (1.18), où la valeur de T correspond à la fréquence à laquelle les valeurs Y_i sont récoltées. Ainsi, avec T égal à une heure ou une journée, l'estimation de λ représenterait de façon intuitive le nombre moyen d'événements pluvieux par heure ou par jour (sachant que $E[N(T)] = \lambda T$).

Rodriguez-Iturbe *et al.* [1984] mentionnent également que ce modèle possède quelques propriétés intéressantes permettant de vérifier son aptitude à représenter les données à différentes échelles. Par exemple, si l'autocorrélation estimée à partir des données est significativement différente de zéro (en utilisant, *e.g.*, le test d'Anderson [1941]) à l'échelle de temps considérée, cela suggère que le modèle n'est pas adéquat pour décrire le processus de précipitation à cette échelle. De plus, si les valeurs estimées des paramètres à deux échelles de temps différentes sont éloignées l'une de l'autre lorsque ramenées à une échelle de temps commune, cela peut indiquer que le modèle n'est pas valide à ces deux échelles simultanément (*i.e.* que les résultats du modèle obtenus à une certaine échelle ne peuvent pas être transposés à une autre échelle simplement par agrégation/désagrégation).

Pour d'autres exemples d'application, le lecteur est référé à Grenier [1999] qui propose une évaluation de la performance de divers modèles de précipitation basés sur des processus ponctuels de type groupé.

1.2.2 L'invariance d'échelle et son utilité dans la modélisation

L'utilisation de modèles stochastiques est une alternative pratique à l'emploi de modèles déterministes pour représenter des phénomènes complexes, comme la précipitation. On évalue habituellement la qualité d'un modèle stochastique par sa capacité à reproduire les caractéristiques statistiques présentées par le phénomène en question (*e.g.* les moments statistiques). Dans le cas de la pluie, une propriété dénommée **invariance d'échelle** retient l'attention. Pour des fins de modélisation, l'invariance d'échelle peut être vue comme une propriété permettant de caractériser un processus stochastique de façon à pouvoir en tirer aisément toute l'information probabiliste dont on a besoin. Au même titre, par exemple, que les hypothèses d'incrémentés indépendants et/ou de stationnarité, l'hypothèse d'invariance d'échelle sert donc à décrire une certaine catégorie de processus stochastiques. Dans ce cas-ci, les processus en question ont la particularité de conserver certaines propriétés (les distributions d'énergie et de masse par exemple, lorsqu'il s'agit de processus physiques) lorsqu'ils sont soumis à des opérations de changement d'échelle.

Pour en donner une définition brève, la propriété d'invariance d'échelle implique que les propriétés statistiques de la série à grande échelle (T_G) sont reliées aux propriétés statistiques à plus petite échelle (T_P) par un simple opérateur de changement d'échelle donné par $\lambda = T_G/T_P$. Cela implique généralement l'existence d'une fonction de changement d'échelle, pour tout λ , par laquelle les processus agrégés à différentes échelles peuvent être rendus équivalents en distributions de probabilité (marginales et/ou conjointes).

Une définition plus restrictive de l'invariance d'échelle peut être obtenue en considérant l'égalité au niveau des moments statistiques au lieu de l'égalité au niveau des distributions de probabilité pour les processus agrégés à différentes échelles (toujours à l'aide d'une fonction de changement d'échelle). Ce genre d'approche est plus commode pour des considérations d'ordre pratique et sera par le fait même privilégiée dans la suite de ce mémoire. Une discussion plus détaillée sur les aspects théoriques de l'invariance d'échelle fait l'objet de la section 2.1.1. Le lecteur est également référé à Gupta et Waymire [1990] à ce sujet.

L'investigation de la propriété d'invariance d'échelle dans les données de pluie est quelque chose qui reçoit une attention croissante dans la littérature concernant la modélisation des précipitations, celle-ci se révélant souvent présente (voir, *e.g.*, Schertzer et Lovejoy [1987], Olsson *et al.* [1993], Tessier *et al.* [1996]). En effet, la simulation de processus invariants d'échelle à l'aide de cascades multiplicatives (section 2.1.3) s'avère efficace pour la reproduction des séries de pluie, particulièrement au niveau de l'effet de regroupement (*clustering*) typiquement observé dans la précipitation.

Un autre avantage des modèles invariants d'échelle est qu'ils offrent une alternative simple et attrayante à l'utilisation de modèles reposant sur les processus ponctuels. En effet, bien que ces derniers puissent donner jusqu'à un certain point une description physique adéquate du phénomène de précipitation, ils ont besoin pour ce faire d'un grand nombre de paramètres, surtout pour bien reproduire l'effet de regroupement. Autrement dit, la simulation de l'invariance d'échelle au sein d'un processus ponctuel ne s'obtient qu'au prix d'un manque de parcimonie. Il en résulte alors que ces modèles ont souvent de la difficulté à représenter correctement le phénomène à différentes échelles, les valeurs estimées des paramètres ne se révélant valables que pour l'échelle de temps considérée. Dans le cas d'un modèle invariant d'échelle, les paramètres sont par construction valides à toutes les échelles et, conséquemment, peu nombreux. La présence de la propriété d'invariance d'échelle mérite donc d'être exploitée lorsque vient le temps de modéliser la précipitation. L'objectif

principal de ce mémoire est d'illustrer ce propos en donnant un bon aperçu de ce que l'on peut tirer de l'utilisation de modèles invariants d'échelle.

2. MODÉLISATION DANS LE CADRE DE L'INVARIANCE D'ÉCHELLE

2.1 Aspects théoriques

2.1.1 Invariance d'échelle simple et spectre de Rényi

Cette section vise à décrire le contexte théorique à partir duquel on peut définir la marche à suivre afin de détecter la présence de la propriété d'invariance d'échelle dans une série de données pluviométriques. Notons tout d'abord que dans le cadre de ce mémoire, on ne s'intéresse qu'aux applications sur les séries temporelles, *i.e.* à une dimension, par opposition aux séries spatio-temporelles (à dimension 2 ou 3). Cela permettra donc d'alléger la présentation qui va suivre quoiqu'une théorie générale sur l'invariance d'échelle soit déjà bien établie (voir, *e.g.*, Schertzer et Lovejoy [1987], Gupta et Waymire [1990]). Puisqu'elle sera répétée souvent, l'expression "invariance d'échelle" va désormais être remplacée par l'abréviation "IE".

Soit $\{X(t)\}$ un processus stochastique à paramètre continu (*i.e.* $t \in \mathbb{R}$). Définissons à partir de ce dernier le processus agrégé $\{Y(\Delta t)\}$, où Δt représente le pas de temps auquel on agrège (*e.g.* 1 heure pour une série horaire, 24 heures pour une série journalière, *etc...*), et que l'on peut exprimer ainsi:

$$Y(\Delta t) = \int_{\Delta t} X(t) dt \quad (2.1)$$

La propriété d'**IE simple**, habituellement associée à des processus possédant des distributions de probabilité continues, survient lorsque pour toute échelle $\lambda > 0$, il existe une constante $C(\lambda)$ telle que l'égalité suivante soit satisfaite:

$$\{[C(\lambda)]^{-1} Y(\lambda \Delta t)\} \stackrel{d}{=} \{Y(\Delta t)\} \quad (2.2)$$

où $\stackrel{d}{=}$ symbolise l'égalité au niveau des distributions. En ce sens, on parlera parfois de l'IE simple au sens strict. On peut remarquer que la fonction d'échelle $C(\lambda)$ prend la forme d'une loi de puissance car, pour deux valeurs λ_1 et λ_2 , on montre à l'aide de (2.2) que:

$$\{C(\lambda_1 \lambda_2) Y(\Delta t)\} \stackrel{d}{=} \{Y(\lambda_1 \lambda_2 \Delta t)\} \stackrel{d}{=} \{Y(\lambda_1 [\lambda_2 \Delta t])\} \stackrel{d}{=} \{C(\lambda_1) Y(\lambda_2 \Delta t)\} \stackrel{d}{=} \{C(\lambda_1) C(\lambda_2) Y(\Delta t)\}$$

et puisque la distribution de $Y(\Delta t)$ est ici supposée arbitraire, cela implique que:

$$C(\lambda_1 \lambda_2) = C(\lambda_1) C(\lambda_2) \quad (2.3)$$

Il est facile de voir que (2.3) se généralise pour n arguments ($n \geq 2$). Par conséquent, on a $C(\lambda^k) = [C(\lambda)]^k$ pour tout $k \in \mathbb{Q}$, et en autant que $C(\lambda)$ soit continue en λ , pour tout $k \in \mathbb{R}$ (pour une dérivation rigoureuse, voir Lamperti [1962]). En utilisant l'identité $e^{\ln(u)} = u$, on en déduit que:

$$C(\lambda) = C(e^{\ln(\lambda)}) = [C(e)]^{\ln(\lambda)} = e^{\ln(C(e)) \ln(\lambda)} = \lambda^{\ln(C(e))}$$

En d'autres mots, en dénotant l'expression $\ln(C(e))$ par \mathcal{H} , on a $C(\lambda) = \lambda^{\mathcal{H}}$, et l'équation (2.2) peut alors se réécrire sous la forme:

$$\{\lambda^{-\mathcal{H}} Y(\lambda \Delta t)\} \stackrel{d}{=} \{Y(\Delta t)\} \quad (2.4)$$

Il est à noter que la valeur du paramètre \mathcal{H} (dénommé l'exposant d'échelle) doit être positive dans l'équation (2.4) de façon à s'assurer que la valeur du facteur de normalisation $\lambda^{-\mathcal{H}}$ soit cohérente (*i.e.* que si $\lambda > 1$, on doit avoir $\lambda^{-\mathcal{H}} < 1$ et vice versa). Remarquons

également qu'un processus stochastique formé d'une succession de variables indépendantes et identiquement distribuées n'obéit pas de façon générale à la propriété d'IE simple. Il a été démontré par Lévy [1937] que cela est vrai si et seulement si les distributions de probabilité appartiennent à une certaine classe connue sous l'appellation de "strictement stable".

La relation décrite en (2.4) pour les densités entraîne nécessairement une relation semblable pour les moments statistiques (d'ordre q), dans la mesure où ces derniers existent, ce qui peut s'exprimer ainsi:

$$E\left[|Y(\lambda\Delta t)|^q\right] = E\left[|\lambda^H Y(\Delta t)|^q\right] = \lambda^{qH} E\left[|Y(\Delta t)|^q\right] \quad (2.5)$$

La propriété définie par (2.5) est parfois appelée l'IE simple au sens large, puisque celle-ci ne fait référence qu'aux moments, par opposition à l'IE simple au sens strict qui fait appel à l'égalité des distributions. D'un point de vue théorique, l'IE simple au sens large constitue une hypothèse moins forte que celle au sens strict. En effet, on remarque que si (2.4) implique (2.5), le contraire n'est pas vrai. Puis, si l'IE simple au sens large peut perdre son sens lorsque les moments ne sont pas définis pour certaines valeurs de q , cela n'empêche pas que l'IE simple au sens strict puisse toujours exister. D'un point de vue pratique cependant, l'IE simple au sens large peut facilement être vérifiée par l'estimation des moments en fonction de λ (pour différentes valeurs de q). En effet, il découle de l'équation (2.5) que la relation obtenue doit être linéaire dans un diagramme log-log, ce qui, le cas échéant, permet en plus d'estimer la valeur de l'exposant d'échelle \mathcal{H} .

Cela nous amène à introduire le **spectre de Rényi** [1970], qui est défini comme suit:

$$\tau(q) = \lim_{\lambda \rightarrow 0} - \frac{\ln\left(\sum_{i=1}^{T/(\lambda\Delta t)} |Y_i(\lambda\Delta t)|^q\right)}{\ln(\lambda)} \quad (2.6)$$

où T représente la période totale d'observation. Une façon de définir la **propriété d'IE** pour un processus stochastique est de dire que le rapport des log dans (2.6) peut s'exprimer indépendamment de λ pour tout q . Plus particulièrement, pour un processus obéissant à la propriété d'IE simple, l'équation (2.5) s'applique, et en la substituant dans (2.6) on obtient:

$$\begin{aligned} \tau(q) &= \lim_{\lambda \rightarrow 0} - \frac{\ln\left(\frac{T}{\lambda \Delta t} \left[\frac{\lambda \Delta t}{T} \sum_{i=1}^{T/(\lambda \Delta t)} |Y_i(\lambda \Delta t)|^q \right]\right)}{\ln(\lambda)} = \lim_{\lambda \rightarrow 0} - \frac{\ln\left(\frac{T}{\lambda \Delta t} \left[\frac{\lambda^{qH} \Delta t}{T} \sum_{i=1}^{T/\Delta t} |Y_i(\Delta t)|^q \right]\right)}{\ln(\lambda)} \\ &= \lim_{\lambda \rightarrow 0} - \frac{\ln\left(\lambda^{qH-1} \sum_{i=1}^{T/\Delta t} |Y_i(\Delta t)|^q\right)}{\ln(\lambda)} = \lim_{\lambda \rightarrow 0} - \frac{\ln(\lambda^{qH-1}) + \ln(cte)}{\ln(\lambda)} = 1 - qH \end{aligned} \quad (2.7)$$

On constate donc que l'IE simple élimine bel et bien la dépendance par rapport à λ dans l'expression (2.7), ce qui nous permet de ne pas tenir compte de la limite dans la définition du spectre τ , et ce qui implique que celui-ci est entièrement déterminé par la valeur de \mathcal{H} (puisque'il s'agit d'une fonction linéaire de pente $-\mathcal{H}$) pour des processus invariants d'échelle simple. Le résultat obtenu en (2.7) nous indique également que $\tau(0) = 1$. $\tau(0)$ se trouve en fait à représenter la dimension de l'ensemble-support, qui correspond à l'axe du temps (de dimension 1) dans le cas d'une série chronologique. Cependant, $\tau(0) = 1$ veut aussi dire que la probabilité d'observer des valeurs nulles dans la série est égale à zéro. Or, dans le cas particulier où $\{Y(\Delta t)\}$ représente un processus non négatif (tel la pluie), on s'imagine facilement qu'il y a de fortes chances pour que l'on se retrouve avec des valeurs nulles. Cela est une bonne indication que l'IE **simple** ne constitue pas un modèle approprié pour ce genre de processus (*i.e.* aux échelles où le processus est intermittent).

On peut obtenir une définition alternative du spectre τ dans le cas d'un processus non négatif par le biais de la règle de l'Hospital, qui devient alors facilement applicable à la définition (2.6), ce qui nous permet de réécrire celle-ci sous la forme:

$$\tau(q) = \lim_{\lambda \rightarrow 0} - \frac{d\left(\ln\left(\sum_{i=1}^{T/(\lambda\Delta t)} [Y_i(\lambda\Delta t)]^q\right)\right)}{d(\ln(\lambda))} \quad (2.8)$$

Ainsi, $\tau(q)$ s'interprète comme la pente de la relation du log de la sommation des puissances d'ordre q (*i.e.* la sommation des valeurs du processus élevées à la puissance q) en fonction du log de λ , lorsque λ se rapproche de 0. Pour $q = 1$, la somme des éléments reste la même pour tout λ , *i.e.* que la pente vaut 0 partout, ce qui implique que $\tau(1) = 0$ (si $\tau(1)$ existe). Sous l'hypothèse d'IE, on peut laisser tomber la limite dans la définition (2.8) car la pente de la relation ne doit pas dépendre de λ , ce qui veut dire que la relation est linéaire. Cela suggère donc que l'évaluation de la pente en question devient un estimateur naturel pour $\tau(q)$ dans le contexte d'IE, estimateur dont l'expression est donnée par:

$$\hat{\tau}(q) = - \frac{d(\ln(S(\lambda, q)))}{d(\ln(\lambda))} \quad (2.9)$$

$$\text{où } S(\lambda, q) = \sum_{i=1}^{T/(\lambda\Delta t)} [Y_i(\lambda\Delta t)]^q \quad (2.10)$$

représente l'estimation de la sommation des puissances d'ordre q au pas de temps $\lambda\Delta t$. L'estimateur défini en (2.9) se révèle donc être un outil permettant de détecter la présence de la propriété d'IE dans les données. Dans un premier temps, il s'agit de vérifier la linéarité des relations $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ pour différentes valeurs de q et d'estimer les pentes pour celles-ci. En reportant ensuite les estimations $\hat{\tau}(q)$ sur un graphique, on doit vérifier l'alignement des points pour confirmer la présence de l'IE simple (au sens large du moins).

Mentionnons cependant qu'une fonction $\tau(q)$ non linéaire nous amène à considérer un autre type d'IE qui s'avère être une généralisation de l'IE simple. En effet, la propriété d'**IE multifractale** est définie par une équation similaire à (2.5), la différence se situant au

niveau de la dérivée $d\tau(q)/dq$ qui, au lieu d'être constante, devient fonction de q . On peut démontrer que dans ce cas le spectre τ , au lieu d'être linéaire, est alors strictement convexe. On note que les étapes menant à vérifier la présence de l'IE multifractale dans les données sont les mêmes que dans le cas de l'IE simple (*i.e.*, linéarité des relations $\ln(S(\lambda, q))$ vs $\ln(\lambda)$, estimation des pentes, graphique des valeurs $\hat{\tau}(q)$ en fonction de q) car c'est l'allure de la fonction $\tau(q)$ estimée qui permet de distinguer entre l'IE simple ou multifractale.

Le terme **multifractal** est relié à la notion de **spectre des dimensions** (Hentschel et Procaccia [1983]) qui, pour des processus non négatifs, est défini comme suit:

$$D(q) \equiv \frac{\tau(q)}{(1-q)} \quad (2.11)$$

On désigne comme **monofractals** les processus dont le spectre τ s'exprime sous la forme $\tau(q) = (1-q) \cdot cte$ (par analogie avec l'I.E. simple, $\tau(q)$ est linéaire avec q dans ce cas). Or, dans le contexte de monofractalité, on remarque que $D(q)$ est constant pour tout q , *i.e.* qu'une dimension unique sert à caractériser tout le processus, soit $D(0) = \tau(0)$ ($= cte$), d'où l'utilisation du préfixe "mono". On doit noter ici que $\tau(0)$ n'est pas nécessairement égal à 1, *i.e.* que contrairement à l'IE simple, la **propriété de monofractalité** peut s'appliquer à des processus non négatifs admettant des valeurs nulles. Dans le cas multifractal, on devine bien entendu que $D(q)$ varie en fonction de q . Le spectre $D(q)$ devient alors une fonction strictement décroissante.

La quantité $\tau(0) = D(0)$ porte le nom de dimension de capacité quoique celle-ci soit souvent considérée comme étant "la" dimension fractale (par abus de langage). Dans le cas monofractal, bien qu'on ait le choix de la valeur de q pour estimer la dimension unique, on s'intéresse avant tout à $D(0)$ puisqu'il s'agit de la seule valeur du spectre $D(q)$ dont l'existence est assurée (Hentschel et Procaccia [1983]). La section 2.2.2 traite d'ailleurs

spécifiquement de l'estimation de cette quantité qui peut déjà révéler des détails intéressants sur les données dont on dispose.

Le lien existant entre la dimension de capacité et "la" dimension fractale vient du fait que celle-ci a une interprétation géométrique. En effet, tel qu'il sera explicité à la section suivante, déterminer la dimension de capacité d'une série chronologique (non négative) implique la transformation de cette dernière en une suite de 0 et de 1 (en remplaçant simplement les valeurs supérieures à zéro par des 1). En transposant la série ainsi modifiée en un ensemble de segments élémentaires dont la longueur correspond au pas de temps employé et qui seront soit vides (pour les valeurs 0) ou pleins (pour les valeurs 1), on obtient un objet géométrique qui peut être assimilé à une construction fractale. Un exemple d'une telle construction est donné à la section qui suit.

2.1.2 Exemple de construction fractale et comptage de boîtes

Les concepts introduits à la section précédente nous donnent les outils nécessaires afin de détecter la présence de la propriété d'IE dans une série de données de précipitation. Une première étape en ce sens concerne l'estimation de la dimension de capacité. Il a été fait mention à la section 2.1.1 qu'il existe un parallèle entre cette notion appliquée aux séries chronologiques et la dimension fractale servant à caractériser des objets géométriques auto-similaires issus d'une construction itérative. Un exemple classique en la matière est la poussière de Cantor (pour en savoir plus sur la géométrie fractale, le lecteur est référé à Mandelbrot [1982] ou encore à Mandelbrot [1989] pour une référence en français).

On construit cet ensemble à partir d'un segment de longueur a que l'on divise en trois pour ensuite en retirer la partie centrale (voir figure 2.1). On se retrouve alors avec un ensemble formé de deux segments de longueur $a/3$ et séparés par un vide de même longueur. La construction se poursuit en divisant à chaque itération chaque segment présent en trois et en lui ôtant sa partie centrale.

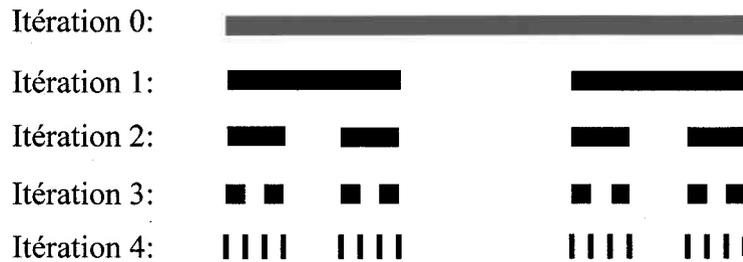


Figure 2.1: Quatre étapes dans la construction d'un ensemble de Cantor

En se référant à la figure 2.1, on constate qu'à l'itération n , on dénombre 2^n segments de longueur $a/3^n$, c'est à dire que l'on se retrouve avec un nombre grandissant de segments à chaque itération tandis que la longueur de ceux-ci diminue constamment. Ce que l'on désigne comme l'ensemble de Cantor est le résultat de cette construction prolongée jusqu'à l'infini (qui nous laisse en présence de "poussière" car les segments se transforment en points d'accumulation à la limite). Afin de caractériser cet ensemble, on utilise la "dimension d'auto-similarité" $D = \log(2^n)/\log(3^n) = \log(2)/\log(3)$ (Mandelbrot [1982] ou [1989]). De nombreuses variations de la poussière de Cantor peuvent être obtenues en faisant varier le rapport de réduction et le nombre de segments conservés. On peut même en faire des versions stochastiques en rendant aléatoires les positions des segments à chaque itération. Dans chaque cas, la dimension d'auto-similarité de l'ensemble est facilement déterminée en appliquant la formule suivante:

$$D_{\text{Cantor}} = \frac{\log(\text{nombre de segments conservés})}{\log(\text{inverse du rapport de réduction})} \quad (2.12)$$

Or, tel qu'il a été établi à la section précédente, il est possible de transformer une série chronologique en un objet géométrique, le résultat obtenu étant une suite de segments élémentaires soit vides (pour les valeurs nulles) ou pleins (pour les valeurs supérieures à zéro) dont la longueur correspond au pas de temps. En répétant cet exercice pour une même série agrégée à différentes échelles de temps, il est ainsi possible d'évoquer la construction d'un ensemble de Cantor. On remarque cependant qu'on ne peut utiliser

directement la formule (2.12) pour déterminer la dimension d'une telle série puisqu'on n'a aucune connaissance *a priori* du processus sous-jacent ayant mené à cette réalisation (*i.e.* qu'on ne connaît pas les paramètres de l'ensemble de Cantor que l'on suppose à l'origine de la série). En fait, on doit utiliser un corollaire de (2.12) qui nous dit que le nombre de segments présents P à l'itération n est fonction de la longueur l de ces derniers selon la relation:

$$P(l) = \left(\frac{a}{l}\right)^D \Rightarrow \log(P(l)) = -D[\log(l) - \log(a)] \quad (2.13)$$

où a est la longueur du segment d'origine. C'est donc dire que si l'on reporte dans un diagramme log-log le nombre de segments présents à chaque étape de la construction d'un ensemble de Cantor en fonction de la longueur de ceux-ci, la relation observée sera linéaire et de pente égale à $-D$. L'algorithme du **comptage de boîtes** s'inspire de l'équation (2.13) pour développer une méthode permettant d'estimer la dimension d'une série chronologique. Dans le cas d'une série pluviométrique, cette méthode consiste simplement à compter le nombre d'événements pluvieux (*i.e.* le nombre d'entrées supérieures à zéro) dans la série agrégée à différentes échelles et de reporter le tout dans un diagramme log-log (échelle d'agrégation vs nombre d'événements) pour vérifier tout d'abord la linéarité de la relation puis, le cas échéant, estimer la dimension (*i.e.* la pente). Tel qu'on le verra à la section 2.2.2, ceci correspond à la méthodologie décrite à la section 2.1.1 pour l'estimation de la dimension de capacité $\tau(0)$, *i.e.* que dans un cadre dépassant l'interprétation géométrique, cet algorithme contribue à la validation de l'hypothèse d'IE dans les données.

2.1.3 Les cascades multiplicatives monofractales

Les cascades multiplicatives constituent un outil régulièrement employé dans la simulation de processus invariants d'échelle. Pour en donner une description simple, disons qu'il s'agit d'une construction itérative commençant avec une "masse" initiale (m) distribuée

uniformément sur le support approprié (dans ce cas-ci, l'axe du temps). À chaque étape, le support est divisé en un certain nombre de branches (b) et autant de poids (w) sont générés de façon à ce que la masse soit redistribuée dans chaque branche par multiplication avec les poids respectifs (voir figure 2.2). Il est possible de s'assurer d'une conservation moyenne de la masse en imposant la condition que l'espérance de la somme des poids doit être égale à 1. Si la somme de ceux-ci vaut exactement 1, on dit alors que les poids sont complémentaires tandis que la cascade est considérée comme étant microcanonique. On remarque que dans le cas d'une cascade canonique, *i.e.* dont les poids sont non complémentaires, il n'est pas possible de reconstruire la cascade par agrégation des valeurs obtenues aux niveaux inférieurs afin de remonter jusqu'au niveau initial (puisque'on ne retrouvera pas exactement la valeur de la masse initiale).

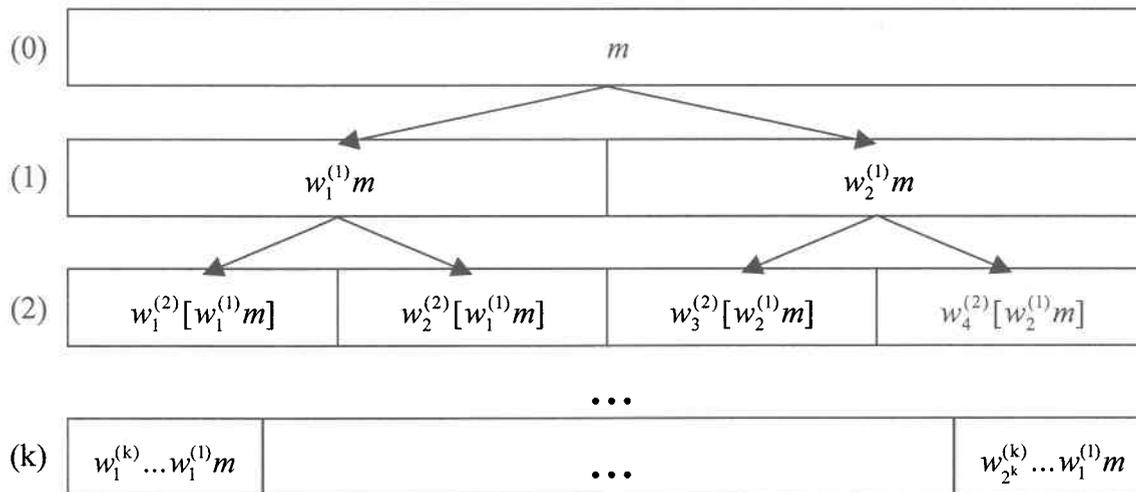


Figure 2.2: Schéma d'une cascade à deux branches

Il est possible de faire une analogie entre une cascade multiplicative et l'ensemble de Cantor (voir section précédente) dans la mesure où il existe une probabilité p_0 non nulle que la valeur des poids puisse être égale à 0. Sous cet angle, l'ensemble de Cantor représenté à la figure 2.1 peut être vu comme une cascade à 3 branches dont la probabilité p_0 vaut $1/3$. Il importe toutefois de spécifier qu'il s'agit là d'une cascade déterministe puisque la position de la branche "vide" est connue d'avance (celle-ci se retrouvant toujours au centre). On

pourrait dire également qu'il s'agit d'une cascade microcanonique puisqu'il y a exactement deux branches "pleines" à chaque sous-division (en supposant que la somme des deux poids non nuls soit égale à 1).

La dimension (fractale) de l'ensemble de Cantor telle que donnée par la formule (2.12) peut se résumer sous la forme $D = \log(c)/\log(b)$, où c représente le nombre de branches "pleines" et où b est le nombre total de branches. Ainsi, si l'on fixe au préalable la valeur de la dimension D , il est possible de déterminer des valeurs pour c et b permettant de construire une cascade dont la dimension (de capacité) sera égale à D . Par exemple, si $D = 1/3$, on doit avoir $3\log(c) = \log(b)$, ou encore $c^3 = b$, d'où l'on peut déduire la solution $c = 2$ et $b = 8$. C'est donc dire qu'en construisant une cascade à huit branches avec seulement 2 des huit poids qui sont non nuls à chaque étape, on obtiendra une série dont la dimension vaudra $1/3$. On peut y ajouter un élément stochastique en rendant aléatoires les positions des branches non nulles sans pour autant affecter la valeur de la dimension.

Dans une perspective de simulation stochastique cependant, on préfère généralement utiliser des cascades canoniques qui sont moins restrictives dans leur construction par rapport aux conditions qu'elles imposent sur le processus sous-jacent. Pour reprendre l'exemple précédent, au lieu de se retrouver avec deux poids sur huit qui sont plus grands que zéro, la valeur de chaque poids serait plutôt déterminée sur une base individuelle en accordant une probabilité $1-p_0 = 1/4$ ($= c/b = 2/8$) que ces derniers soient non nuls. Si on perd alors l'assurance que la dimension de la série résultante sera exactement égale à $1/3$, on peut en contrepartie ajuster la distribution des poids de façon à ce que l'espérance de la somme de ceux-ci soit égale à 1, avec pour résultat une conservation de la masse en moyenne. En utilisant la formule $1-c/b$ pour calculer la valeur de p_0 , puis en supposant que les poids sont générés de façon indépendante, il est facile de déterminer la distribution appropriée, qui peut s'exprimer sous la forme:

$$f_w(w) = \left(1 - \frac{c}{b}\right) \delta_0(w) + \frac{c}{b} \delta_{1/c}(w) \quad (2.14)$$

où δ représente la distribution de Dirac. La variable W ne peut donc prendre que les valeurs 0 ou $1/c$ avec des probabilités de $1-(c/b)$ et c/b respectivement. On vérifie qu'en générant b variables W_i indépendantes et identiquement distribuées selon (2.14), l'espérance de leur somme vaut bel et bien 1:

$$E\left[\sum_{i=1}^b W_i\right] = bE[W] = b\left[\left(1 - \frac{c}{b}\right)0 + \left(\frac{c}{b}\right)\frac{1}{c}\right] = b \cdot \frac{1}{b} = 1$$

On a considéré jusqu'ici les valeurs de b et c comme étant entières. Si cela est trivial pour b (pour qu'il y ait cascade, b doit être un entier plus grand ou égal à 2), on note que cette restriction n'est pas nécessaire pour c puisque la valeur de p_0 peut être obtenue uniquement à partir du nombre de branches b et de la dimension D , comme suit:

$$D = \frac{\log[c]}{\log[b]} \Rightarrow D \cdot \log[b] = \log[c] \Rightarrow b^D = c \Rightarrow b^{D-1} = \frac{c}{b} = 1 - p_0 \quad (2.15)$$

Cette constatation est utile en pratique car elle nous dit qu'il est possible de générer des cascades de dimension (sous-unitaire) quelconque en fixant au préalable le nombre de branches b . En effet, pour des valeurs données de b et D , on déduit de (2.15) que $P[W = 0] = p_0 = 1 - b^{D-1}$ et que soit $W = 0$ ou soit $W = 1/c = b^{-D}$. Remarquons que dans le cas où $D = 1$, $p_0 = 0$ et les poids valent tous automatiquement $1/b$; en d'autres mots, la masse est alors toujours répartie également le long du support à toutes les étapes de la construction de la cascade, *i.e.* qu'il n'y a pas de discontinuité du processus.

L'hypothèse de l'indépendance des poids, bien que simplificatrice, peut être relaxée afin de laisser place à des résultats plus réalistes. Dans le cas de la pluie tout particulièrement, où

l'autocorrélation à l'échelle horaire est habituellement significative (voir section suivante), l'ajout d'une certaine corrélation au niveau des poids de la cascade peut constituer une façon de mieux se conformer à la distribution observée pour les périodes sèches.

2.2 Vérification de l'hypothèse d'invariance d'échelle sur les données ; zones d'applicabilité

2.2.1 Démarche méthodologique et présentation des données

L'objectif poursuivi dans les trois prochaines sections consiste à donner un exemple concret de validation de l'hypothèse d'IE dans une série chronologique. La démarche adoptée s'inspire des grandes lignes de la section 2.1.1, *i.e.* qu'elle fait appel à l'estimation du spectre τ en différentes valeurs de q , ce qui fait l'objet de la section 2.2.3. Auparavant, à la section 2.2.2, une attention plus particulière est portée au cas $q = 0$, où l'on procède à l'estimation de la dimension de capacité $\tau(0)$ par le biais de l'algorithme du comptage de boîtes qui a été décrit à la section 2.1.2.

Toutefois, avant de procéder, il convient de dire quelques mots sur la série de données retenue pour effectuer cet exercice. Il s'agit d'un échantillon provenant de la station AL 748 (Division 3) en Alabama et contenant 19368 relevés pluviométriques horaires recueillis entre le 19 juillet 1993 et le 3 octobre 1995. Les mesures effectuées pendant cette période présentent l'avantage d'être complètes (*i.e.* pas de données manquantes) en plus de provenir d'une région géographique où la précipitation demeure relativement uniforme tout au cours de l'année (car pas d'hiver froid). Cela permettra de ne pas se soucier des problèmes relatifs aux effets saisonniers.

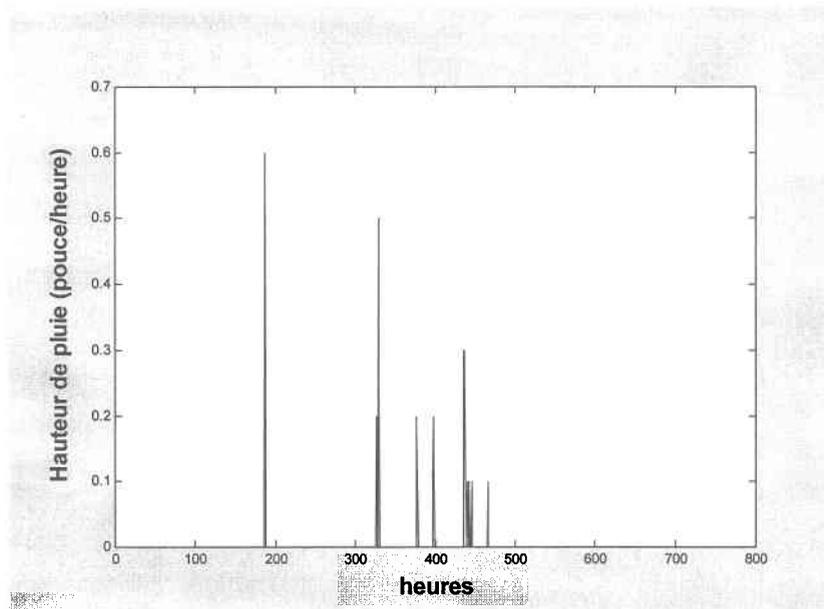


Figure 2.3: 720 relevés pluviométriques horaires d'une station de l'Alabama

Question de faire un survol rapide des statistiques de la série, on peut dire tout d'abord que la moyenne et la variance sont assez faibles à 0,0064 et 0,0023 pouce/heure respectivement. Cela n'est pas surprenant puisque bien que les mesures varient entre 0 et 1,5 pouce/heure, la proportion de valeurs nulles dans les données s'élève à 96,7%. La figure 2.3 présente les 720 premiers relevés de la série, ce qui correspond environ à un mois de données. On y constate un effet de regroupement (le *clustering*) provoqué par quelques averses se présentant de façon quasi ponctuelle. En effet, la durée moyenne d'une averse est de 1,75 heure par opposition à la durée moyenne entre celles-ci qui est de 51 heures. Cette dernière valeur peut sembler relativement faible en comparaison avec d'autres séries horaires typiques ; elle passe malgré tout à 67 heures dans la série agrégée à deux heures, puis à 80 heures dans la série agrégée à quatre heures (cela étant probablement dû à la présence de courtes périodes sèches intercalées au coeur des événements pluvieux, ce qui n'est plus visible à grande échelle).

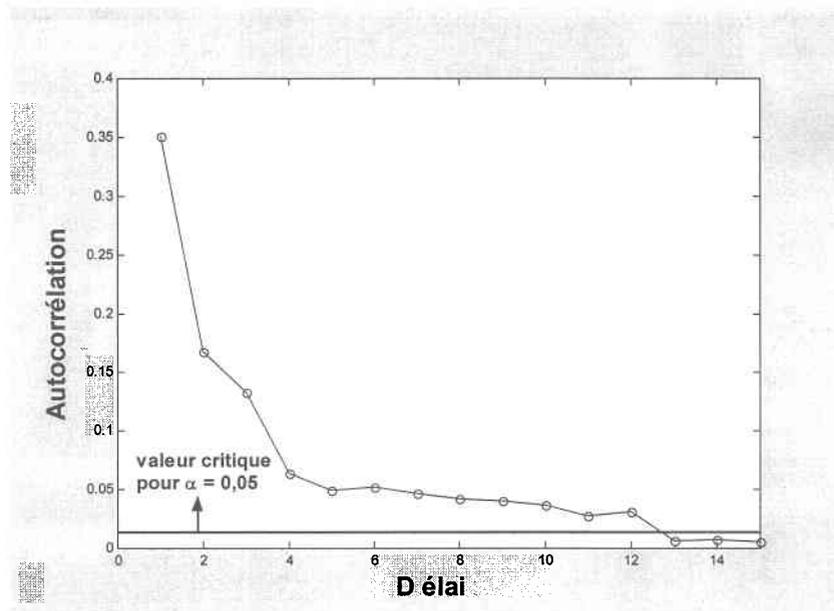


Figure 2.4: Fonction d'autocorrélation estimée de la série horaire de l'Alabama

La figure 2.4 présente la fonction d'autocorrélation estimée des données à l'échelle horaire. Le test d'Anderson [1941] nous permet d'affirmer que celle-ci est significativement différente de zéro (pour un seuil de confiance $\alpha = 0,05$ tel qu'indiqué sur la figure 2.4) jusqu'à un délai de 12 heures. En effet, la valeur critique de ce test étant donnée par $1,96 / \sqrt{N} \cong 0,014$ (où N représente le nombre total d'observations dans la série, soit 19 368), la valeur absolue de l'autocorrélation estimée ne doit donc pas excéder cette valeur critique afin d'être considérée comme non significative. Tel que mentionné à la section 1.1.2, cette décroissance lente de la fonction d'autocorrélation peut être perçue comme une manifestation du phénomène de persistance typique aux séries de précipitation.

2.2.2 Estimation de la dimension de capacité

Nous sommes intéressés ici à estimer la dimension de capacité du jeu de données à notre disposition. En se référant à la section 2.1.1, cela consiste à évaluer le spectre τ en $q = 0$. On doit alors considérer l'estimateur défini en (2.9), pour lequel il faut calculer les sommes $S(\lambda, 0)$ en fonction de différentes valeurs de λ . Or, dans le contexte de la

section 2.1.1, λ était un réel positif représentant un facteur d'échelle permettant soit de rapetisser ($\lambda < 1$) ou d'agrandir ($\lambda > 1$) l'échelle d'observation du processus original. On comprend cependant qu'en pratique nous soyons limités à la résolution des données, *i.e.* que l'on ne peut utiliser que des valeurs discrètes pour λ (1,2,3,...). Toute valeur intermédiaire impliquerait une désagrégation de la série à une échelle plus fine, ce dont on ne dispose pas. En ce qui concerne l'expression de la sommation des puissances d'ordre 0 au pas de temps $\lambda\Delta t$ ($\Delta t = 1$ heure dans ce cas-ci), elle prend la forme suivante, en accord avec (2.10):

$$S(\lambda,0) = \lim_{q \downarrow 0} \sum_{i=1}^{N/\lambda} [R_i(\lambda\Delta t)]^q \quad (2.16)$$

où $R_i(\lambda\Delta t)$ représente le $i^{\text{ème}}$ élément de la série agrégée au pas de temps $\lambda\Delta t$ et où N est toujours le nombre d'entrées dans la série. Si N/λ n'est pas entier, on se limite à la partie entière. L'expression (2.16) se résume en fait en une sommation de 0 (pour les valeurs nulles) et de 1 (pour les valeurs strictement positives). Ainsi, les valeurs nulles de la série ne contribuent pas à l'évaluation de la sommation des puissances d'ordre 0, ce qui se justifie par l'emploi de la limite dans (2.16) (en effet, bien que 0^0 soit indéfini, on remarque que $\lim_{x \downarrow 0} 0^x = 0$). $S(\lambda,0)$ correspond donc au nombre de valeurs supérieures à zéro dans la série agrégée à l'échelle $\lambda\Delta t$.

L'estimation de $\tau(0)$, qui selon (2.9) est donnée par $\hat{\tau}(0) = -\frac{d(\ln(S(\lambda,0)))}{d(\ln(\lambda))}$, s'obtient alors en

examinant les valeurs de $S(\lambda,0)$ sur un diagramme log-log en fonction de λ , *i.e.* en calculant la pente d'une éventuelle droite se dessinant sur un tel graphique. On voit que cela correspond bien à la description de l'algorithme du comptage de boîtes dont il a été question à la section 2.1.2. Nombre d'auteurs ont utilisé cette procédure pour caractériser la variabilité temporelle des précipitations (*e.g.* Hubert et Carbonnel [1989], Olsson *et al.* [1992]). Cette approche, au départ géométrique, s'est ensuite raffinée pour incorporer la

propriété d'IE dans des modèles tenant compte également de l'intensité de la précipitation, permettant dès lors d'effectuer des simulations. Un exemple d'un tel modèle, conçu pour reproduire la série de données introduite à la section précédente, et se basant sur le principe des cascades multiplicatives (section 2.1.3) est développé à la section 2.3.

Les résultats obtenus suite à l'application de la méthode du comptage de boîtes (facilement programmable d'ailleurs) sur les données horaires de l'Alabama sont illustrés à la figure 2.5. Les sommations $S(\lambda,0)$ y ont été calculées pour $\lambda\Delta t = 1$ heure, 2 heures, 3 heures, et ainsi de suite jusqu'à 8760 heures (*i.e.* un an). Ce que l'on observe tout d'abord sur la figure 2.5, c'est qu'il y a un bon alignement des points pour les petites valeurs de λ (de 1 à 6 heures en fait) et que la pente estimée (par régression) pour le segment couvrant ces périodes équivaut à $-0,37$, d'où l'estimation de la dimension est donnée par $\hat{\tau}(0) = 0,37$. La linéarité des points après six heures est moins évidente quoique ces derniers suivent assez bien une droite de pente $-0,37$ (droite qui est indiquée sur le graphique par des cercles [°°°]), en autant que l'on reste sous la barre des 6 jours environ (*i.e.* tant que $\ln(\lambda) < \ln(6 \times 24) \cong 5$).

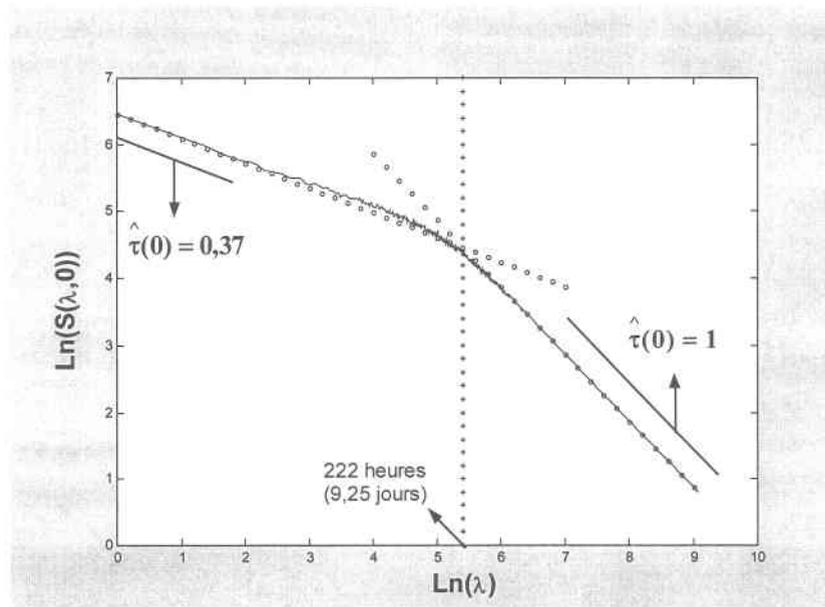


Figure 2.5: Algorithme du comptage de boîtes appliqué aux données horaires de l'Alabama (estimation de la dimension de capacité)

En effet, on s'aperçoit que pour des périodes de l'ordre de la semaine et plus, la pente de la relation $\ln(S(\lambda,0))$ vs $\ln(\lambda)$ commence à s'accroître pour tendre finalement vers -1. Cela n'est guère surprenant puisque, d'une certaine façon, cet effet est pratiquement intrinsèque au phénomène de précipitation. Ce qui se passe, c'est qu'au delà d'un intervalle de temps suffisamment grand, on peut être presque certain qu'une chute de pluie surviendra. Dans le cas de la série utilisée ici, la période sèche maximale enregistrée est de 578 heures. Cela veut donc dire que pour $\lambda > 578$, $S(\lambda,0) = \sum_{i=1}^{N/\lambda} 1 = N/\lambda$, puisqu'on est assuré que

$$R_i(\lambda\Delta t) > 0 \text{ pour tout } i. \text{ Conséquent, } \ln(S(\lambda,0)) = \ln(N) - \ln(\lambda) \text{ et } \frac{d(\ln(S(\lambda,0)))}{d(\ln(\lambda))} = -1.$$

Cela suggère que la série observée serait le résultat de la combinaison de deux processus d'IE : dans la mesure où la monofractalité est vérifiée (ce qui est le sujet de la section suivante), on aurait donc pour les petites échelles (de quelques heures à quelques jours) un processus de dimension 0,37 puis, pour les grandes échelles (au delà d'une semaine), un processus de dimension 1. Bien que, techniquement, la dimension $\tau(0)$ soit définie en (2.8) par la valeur de la pente vers $\lambda = 0$, le fait que la relation $\ln(S(\lambda,0))$ vs $\ln(\lambda)$ soit linéaire pour les grandes échelles nous permet de justifier cette extrapolation. En effet, tel qu'argumenté à la section 2.1.1, la linéarité découlant du contexte d'IE nous permet d'éliminer la limite dans la définition du spectre τ . On se retrouve alors avec deux régimes distincts d'IE, ce dont on devra tenir compte dans la modélisation.

L'identification de plusieurs régimes d'IE à l'aide de la méthode du comptage de boîtes a déjà été mise en évidence par Olsson *et al.* [1993]. Selon toute vraisemblance, un troisième segment serait apparu sur la figure 2.5 si nous avions disposé de données de résolution plus fine. D'après l'auteur, pour des échelles en deçà d'une heure, on se situe habituellement au niveau de la durée typique d'un événement pluvieux (la valeur moyenne de 1,75 heure donnée à la section précédente est quant à elle dépendante de l'échelle d'échantillonnage d'une heure, puisque 60% des averses enregistrées se sont produites à l'intérieur d'une

période d'une heure). Le processus est alors plus susceptible de présenter des caractéristiques multifractales en raison de la grande variabilité de la distribution caractérisant la durée des événements pluvieux (Olsson *et al.* [1993]). Lorsqu'on passe à des échelles allant de quelques heures à quelques jours, on se situe plutôt au niveau de la durée entre les événements. La dimension de 0,37 estimée pour ces périodes peut ainsi être vue comme une mesure de l'intermittence des événements pluvieux dans le temps.

La transition d'une pente de $-0,37$ à -1 sur la figure 2.5 ne se fait pas brutalement ; c'est donc dire qu'il existe une zone de transition (entre une et deux semaines approximativement, *i.e.* pour $5 \cong \ln(6 \times 24) < \ln(\lambda) < \ln(14 \times 24) \cong 5,8$) à l'intérieur de laquelle l'hypothèse d'IE ne s'applique pas vraiment. Comme il s'agit d'une fenêtre relativement petite, un compromis raisonnable à adopter dans la modélisation consiste à déterminer une frontière séparant les courtes périodes des longues périodes. Pour cela, on peut prendre simplement l'intersection des droites de pente $-0,37$ et -1 (indiquées par des cercles [°°°] sur la figure 2.5) représentant respectivement les petites et les grandes durées, et qui équivaut à 222 heures dans ce cas-ci, soit 9,25 jours.

Il reste toutefois à confirmer l'hypothèse de monofractalité en effectuant une estimation plus complète du spectre τ , ce qui est proposé à la prochaine section. Dans la perspective de deux régimes d'IE, il faudra de plus obtenir deux fonctions $\hat{\tau}(q)$ estimées, soit une pour les courtes périodes et une autre pour les longues périodes.

En terminant cette section, mentionnons que l'algorithme du comptage de boîtes se généralise facilement pour des dimensions du support supérieures à 1, *i.e.* qu'il est également applicable aux séries spatio-temporelles. En deux dimensions, au lieu de compter des segments, on va compter des carrés ; en trois dimensions, ce sera des cubes et ainsi de suite...

2.2.3 Estimation du spectre τ

L'estimation de la dimension de capacité $\tau(0)$ effectuée à la section précédente nous a permis d'accomplir un premier pas vers l'objectif de validation de l'hypothèse d'IE dans une série temporelle de précipitation. Il reste maintenant à estimer la fonction $\tau(q)$ pour des valeurs de q supérieures à 0. Pour confirmer la présence de la propriété de monofractalité, il faut que le spectre estimé $\hat{\tau}(q)$ obéisse à la relation $\tau(q) = (1-q) \cdot \tau(0)$ telle que donnée à la section 2.1.1 pour des processus monofractals. Tout cela sans oublier qu'on désire vérifier cette hypothèse deux fois puisque deux régimes d'IE aux propriétés différentes (*i.e.* dont la dimension $\tau(0) = D(0)$ n'est pas la même) ont été préalablement identifiés en fonction des échelles de temps considérées (soit grandes ou petites).

L'estimation de $\tau(q)$ pour une valeur de q fixée s'effectue selon la méthodologie décrite à la section précédente, excepté que des sommations $S(\lambda,0)$, on passe désormais aux sommations $S(\lambda,q)$ à évaluer en fonction de λ , de façon à obtenir $\hat{\tau}(q)$ par la pente de la relation $\ln(S(\lambda,q))$ vs $\ln(\lambda)$ (si celle-ci est linéaire), conformément à (2.9). L'expression de la sommation des puissances d'ordre q au pas de temps $\lambda\Delta t$, toujours en accord avec (2.10), est alors donnée par:

$$S(\lambda, q) = \sum_{i=1}^{N/\lambda} [R_i(\lambda\Delta t)]^q \quad (2.17)$$

Contrairement à la sommation des puissances d'ordre 0 en (2.16), où l'on ne fait qu'additionner des 0 et des 1, la sommation des puissances d'ordre q ($q > 0$) en (2.17) tient compte de l'intensité de la précipitation et donc la variabilité de l'estimateur $S(\lambda,q)$ sera d'autant plus grande que la valeur de q sera grande. C'est pour cette raison qu'en pratique on se limite à de petites valeurs de q (inférieures à 3 ou 4) afin d'effectuer l'estimation du spectre τ (pour qu'elle demeure fiable).

En ce qui concerne l'application de la méthode aux données de l'Alabama, les sommations $S(\lambda, q)$ ont été calculées pour $\lambda = 1, 2, 3, \dots, 8760$ et pour $q = 0,1, 0,2, 0,3, \dots, 2,9, 3$. Dans le cas $q = 1$, puisque la sommation des éléments de la série agrégée à différentes échelles reste la même pour tout λ , on a que $S(\lambda, 1)$ est constant, d'où $\hat{\tau}(1) = 0$ car la pente est nulle. Les figures 2.6, 2.7, 2.8 et 2.9 présentent les graphiques des relations $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ pour $q = 0,5, 1,5, 2,0$ et $2,5$ respectivement.

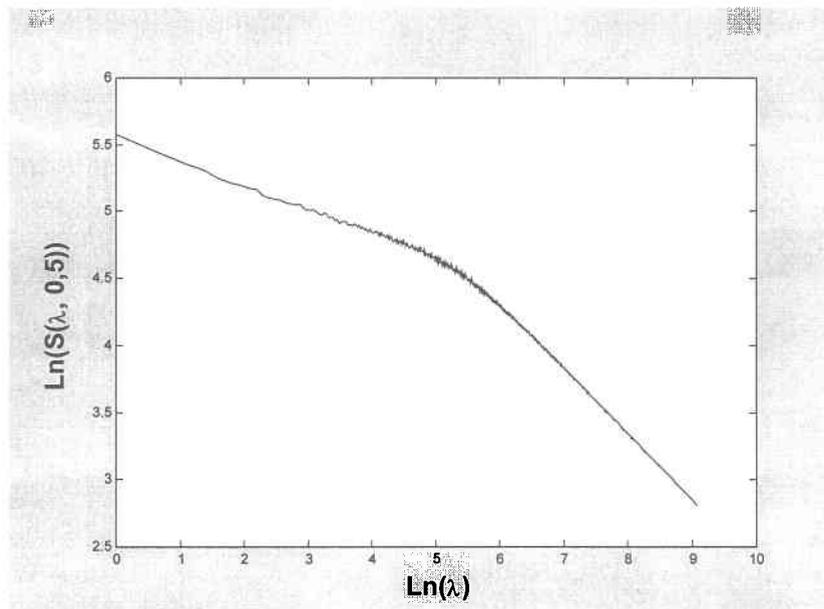


Figure 2.6: Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 0,5$ sur les données horaires de l'Alabama

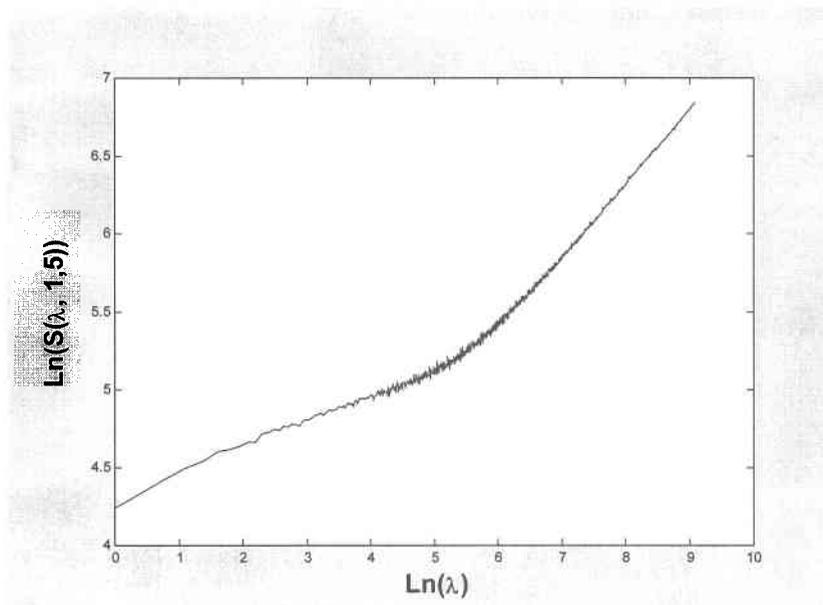


Figure 2.7: Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 1,5$ sur les données horaires de l'Alabama

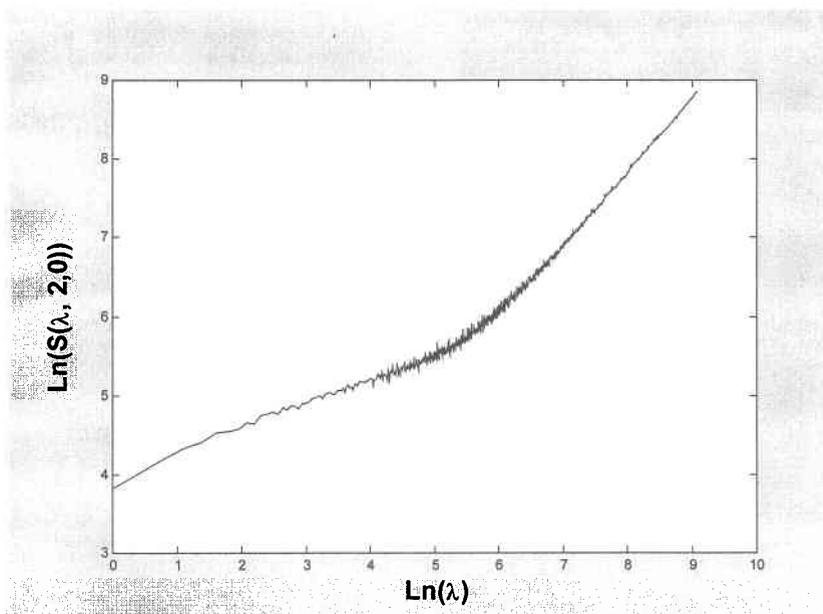


Figure 2.8: Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 2,0$ sur les données horaires de l'Alabama

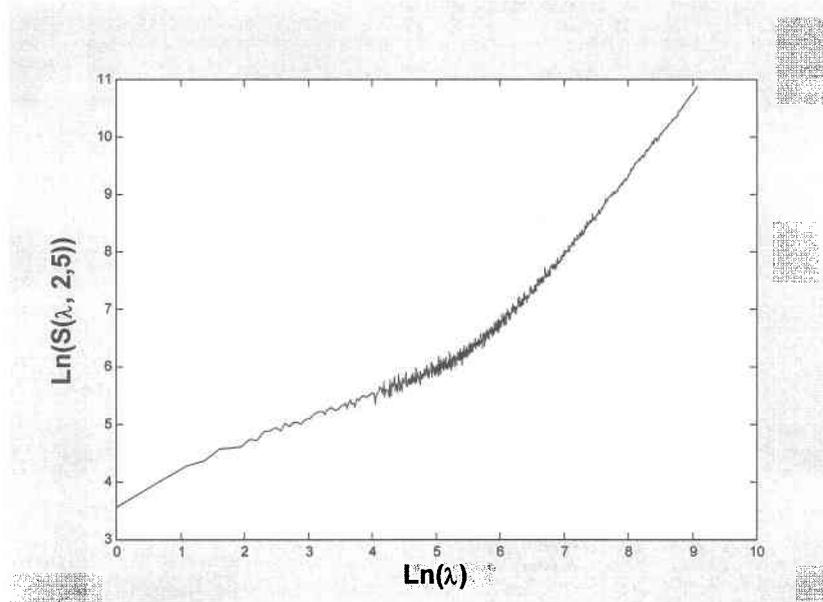


Figure 2.9: Relation $\ln(S(\lambda, q))$ vs $\ln(\lambda)$ calculée pour $q = 2,5$ sur les données horaires de l'Alabama

Concernant ces quatre figures, on peut remarquer tout d'abord que le cas $q = 0,5$ ressemble beaucoup au cas $q = 0$ (voir figure 2.5 à la section précédente). On constate par la suite que les valeurs de $\ln(S(\lambda, q))$ tendent à osciller de plus en plus à mesure que q augmente (bien que ces oscillations soient atténuées en coordonnées log-log), ce qui rejoint ce qui a été dit auparavant par rapport à la variabilité de l'estimateur $S(\lambda, q)$ qui grandit avec q . À propos de l'alignement des points, on peut dire de façon générale qu'il a la même allure que dans le cas $q = 0$, c'est-à-dire qu'il est bon pour les courtes et les longues périodes tandis qu'une zone de transition est observée entre les deux (cette zone se situant sensiblement dans les mêmes intervalles pour $\ln(\lambda)$ que lorsque $q = 0$).

Avant de passer à l'estimation des différentes valeurs $\hat{\tau}(q)$, il importe de définir les intervalles sur lesquels elle sera faite. Dans le cas de $\hat{\tau}(0)$, on a retenu l'intervalle allant de 1 à 6 heures pour les petites périodes bien qu'on aurait pu prolonger celui-ci (jusqu'à quelques jours même) tout en gardant un bon coefficient de détermination ($R^2 \cong 0,99$) dans

la régression. Cependant, la linéarité des points pour les courtes périodes devient de moins en moins bonne lorsque la valeur de q s'accroît, ce qui peut être observé sur les figures 2.6 à 2.9. Il vaut donc mieux retenir un petit intervalle sur lequel l'estimation sera très bonne quel que soit q , quitte à étendre ce résultat pour toutes les périodes allant de 1 à 222 heures (222 heures étant la frontière entre les courtes et les longues périodes telle que déterminée à la section 2.2.2), en acceptant l'idée que l'alignement des points pour ces périodes peut être assimilé à celui observé pour les six premières heures.

Le même raisonnement s'applique pour les longues périodes, *i.e.* que l'estimation se fera sur un intervalle où la linéarité est très bonne pour ensuite prolonger le résultat jusqu'à la frontière séparant les petites des grandes durées. Un choix logique ici consiste à prendre l'intervalle allant de 578 à 8760 heures puisque, comme il a été expliqué à la section 2.2.2, les points de la relation $\ln(S(\lambda,0))$ vs $\ln(\lambda)$ s'ajustent à une droite de pente -1 lorsque $\lambda > 578$. On dispose dès lors de tous les éléments nécessaires pour réaliser l'estimation du spectre τ , dont les principaux résultats sont exposés dans le tableau 2.1.

Expliquons tout d'abord la signification des symboles employés dans ce tableau : q demeure bien sûr l'ordre de puissance de la sommation, $\hat{\tau}(q)_C$ et $\hat{\tau}(q)_L$ sont les estimations du spectre τ en q pour les courtes et les longues périodes respectivement, tandis que R^2_C et R^2_L sont les coefficients de détermination associés provenant de la régression. Seule l'estimation de $\tau(1)$ a été omise puisque cette dernière est triviale (*i.e.* $\hat{\tau}(1)_C = \hat{\tau}(1)_L = 0$). On remarque que les valeurs des coefficients de détermination sont généralement bonnes bien que $R^2_C < 0,99$ lorsque q est supérieur à 2,1. Il s'agit là d'un premier signe que la linéarité commence à s'effriter pour les petites périodes avec q qui s'agrandit. Heureusement, ce problème ne se présente pas dans le cas des longues périodes puisque les valeurs des coefficients R^2_L sont toutes supérieures à 0,998. Il faut préciser toutefois que ce très bon alignement des points aux grandes échelles est un effet imputable à la

représentation en coordonnées log-log qui masque les écarts (en écrasant les grandes valeurs de $S(\lambda, q)$).

Tableau 2.1: Estimation du spectre τ pour la série de données horaires de l'Alabama

q	$\hat{\tau}(q)_C$	R^2_C	$\hat{\tau}(q)_L$	R^2_L	q	$\hat{\tau}(q)_C$	R^2_C	$\hat{\tau}(q)_L$	R^2_L
0	0,3698	0,9977	1,0000	1,0000	1,6	-0,2564	0,9947	-0,5715	0,9996
0,1	0,3407	0,9981	0,8969	1,0000	1,7	-0,2974	0,9939	-0,6648	0,9996
0,2	0,3096	0,9984	0,7946	1,0000	1,8	-0,3375	0,9931	-0,7577	0,9995
0,3	0,2764	0,9987	0,6930	1,0000	1,9	-0,3768	0,9923	-0,8500	0,9995
0,4	0,2413	0,9989	0,5921	1,0000	2,0	-0,4151	0,9914	-0,9418	0,9994
0,5	0,2044	0,9991	0,4919	0,9999	2,1	-0,4525	0,9906	-1,0350	0,9993
0,6	0,1658	0,9991	0,3923	0,9999	2,2	-0,4890	0,9897	-1,1262	0,9992
0,7	0,1259	0,9991	0,2933	0,9999	2,3	-0,5246	0,9888	-1,2170	0,9992
0,8	0,0848	0,9990	0,1950	0,9999	2,4	-0,5594	0,9879	-1,3073	0,9991
0,9	0,0427	0,9987	0,0972	0,9998	2,5	-0,5934	0,9870	-1,3971	0,9991
1,0	0	—	0	—	2,6	-0,6268	0,9860	-1,4866	0,9990
1,1	-0,0431	0,9980	-0,0966	0,9998	2,7	-0,6595	0,9850	-1,5756	0,9990
1,2	-0,0864	0,9974	-0,1927	0,9997	2,8	-0,6916	0,9840	-1,6642	0,9989
1,3	-0,1295	0,9968	-0,2882	0,9997	2,9	-0,7232	0,9829	-1,7523	0,9989
1,4	-0,1724	0,9962	-0,3832	0,9997	3,0	-0,7544	0,9817	-1,8401	0,9988
1,5	-0,2147	0,9954	-0,4776	0,9996					

Pour juger de la linéarité des fonctions $\hat{\tau}(q)_C$ et $\hat{\tau}(q)_L$, il faut les représenter graphiquement, ce qui est fait à la figure 2.10. Les droites théoriques correspondant à la monofractalité, soit $\tau(q)_C = (1-q) \cdot \hat{\tau}(0)_C$ et $\tau(q)_L = (1-q) \cdot \hat{\tau}(0)_L$ (voir section 2.1.1), y ont

été ajoutées. On peut constater au premier coup d'oeil que l'adéquation semble bonne pour les deux régimes d'IE. Les lignes verticales apparaissant sur la figure représentent les intervalles de confiance (à 95%) calculés lors de la régression pour chaque valeur estimée. On voit que ceux-ci commencent à s'accroître rapidement dans le cas des courtes périodes pour $q > 2$, reflétant ainsi l'incertitude grandissante de l'estimation dans cette région. Dans le cas des longues périodes, les intervalles sont tout simplement trop petits pour apparaître sur la figure (en raison du fort coefficient de détermination).

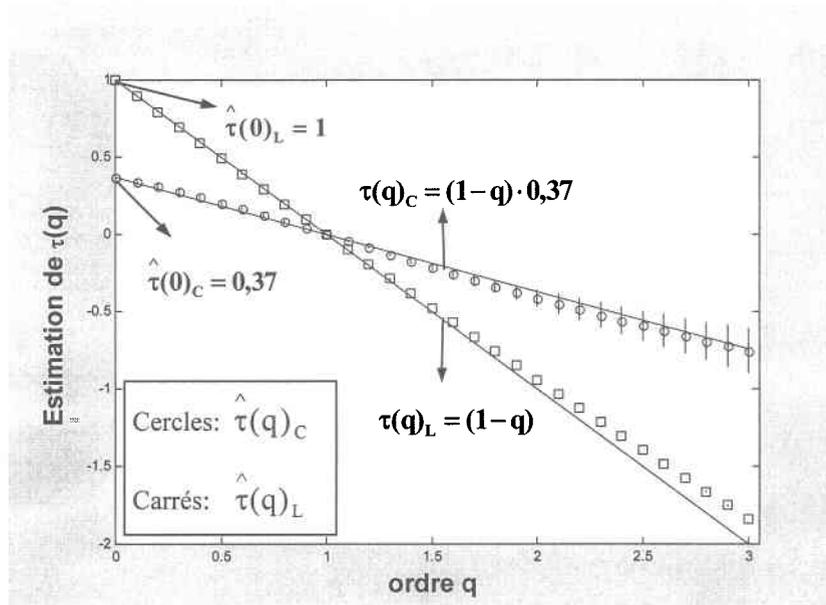


Figure 2.10: Représentation graphique de l'estimation du spectre τ

Les valeurs théoriques de $\tau(2)_C$ et $\tau(2)_L$ reliées à la monofractalité sont données par $\tau(2)_C = (1-2) \cdot \hat{\tau}(0)_C = -0,37$ et $\tau(2)_L = (1-2) \cdot \hat{\tau}(0)_L = -1$. Si on les compare avec les valeurs estimées du tableau 2.1, soit $\hat{\tau}(2)_C = -0,415$ et $\hat{\tau}(2)_L = -0,94$, on peut considérer qu'il s'agit là d'un écart raisonnable permettant d'accepter l'hypothèse de monofractalité pour les deux régimes. Bien que ce choix en faveur du cas monofractal soit discutable (surtout pour les grandes échelles), une discussion plus en profondeur sur les aspects multifractals de l'IE est réservée pour des travaux futurs.

2.3 Description du modèle de précipitation s'appuyant sur les cascades multiplicatives

L'ensemble des éléments qui ont été développés depuis le début de ce chapitre va maintenant permettre d'élaborer un modèle de simulation pouvant reproduire des séries de précipitation aux caractéristiques similaires à celle de l'Alabama avec laquelle nous travaillons dans la présente étude. Concernant cette série, il a été établi à la section 2.2.2 que celle-ci présente des caractéristiques d'IE, avec deux régimes distincts cependant, soit un pour les courtes périodes et un autre pour les longues périodes, la frontière entre ces deux régimes ayant été déterminée à 222 heures. L'étude du spectre τ à la section 2.2.3 a permis de confirmer cette hypothèse, *i.e.* que la série provient de la combinaison de deux processus invariants d'échelle monofractals (section 2.1.1) et dont les dimensions associées sont de 0,37 pour les petites échelles et de 1 pour les grandes échelles.

Le modèle repose donc sur le principe des cascades monofractales décrites à la section 2.1.3. En ce sens, l'équation (2.15) nous indique quels sont les paramètres de la cascade (canonique) à déterminer. Or, pour les petites échelles, la dimension D a été estimée à 0,37. Pour une question de simplicité, on peut poser $b = 2$ (mentionnons que le choix du nombre de branches est arbitraire puisque l'analyse du spectre τ ne nous révèle que la dimension), ce qui nous permet de calculer la valeur de p_0 , *i.e.* $p_0 = 1 - b^{D-1} = 1 - 2^{(0.37-1)} \cong 0,354$ (selon (2.15)). On peut déduire ensuite la valeur à employer pour les poids non nuls, donnée par $b^{-D} = 2^{-D} = 2^{(-0.37)} \cong 0,774$. La cascade ne pouvant être développée que sur des échelles (en heures) correspondant à des puissances de 2, il faudra se limiter à l'échelle de $2^7 = 128$ heures afin de ne pas excéder la frontière séparant les deux régimes d'IE, soit 222 heures.

Pour les longues périodes (*i.e.* 2^k heures, où $k \geq 8$), la dimension passe à 1, ce qui veut dire que la masse est en fait simplement divisée en deux à toutes les étapes de génération de la cascade, tel qu'expliqué à la section 2.1.3. Il suffit alors de simuler des valeurs que pour la plus petite des grandes échelles, soit $2^8 = 256$ heures, pour ensuite les désagréger

conformément au processus de construction de la cascade correspondant aux courtes périodes. Or, pour les données de l'Alabama agrégées à 256 heures, on observe une autocorrélation d'ordre 1 plutôt faible à $-0,067$. Cela suggère que les totaux agrégés à cette échelle peuvent être simulés de façon indépendante. En effet, la valeur critique du test d'Anderson [1941] (voir section 2.2.1) est alors donnée par $1,96 / \sqrt{75} \cong 0,226$ (au seuil de confiance $\alpha = 0,05$), où 75 correspond à la longueur de la série agrégée à 256 heures. Il devient donc naturel dans ce contexte de considérer une distribution de Lévy pour simuler ces valeurs, puisqu'il a été prouvé par Lévy [1937] qu'une suite de variables aléatoires indépendantes et identiquement distribuées obéit à la propriété d'IE simple (formule (2.4), section 2.1.1) si et seulement si les distributions appartiennent à une certaine classe dénommée "strictement stable" (ou bien "Lévy-stable"), dont la densité est donnée par:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-s|u|^\alpha} \cos(-xu - \beta s^\alpha |u|^{\alpha-1} u \tan(\pi\alpha/2)) du \quad (2.18)$$

et où les 3 paramètres sont : $0 < \alpha \leq 2$, l'exposant caractéristique (qui équivaut à l'inverse de l'exposant d'échelle \mathcal{H} , *i.e.* $\alpha = 1/\mathcal{H}$) ; $-1 \leq \beta \leq 1$, le coefficient d'asymétrie, et $s > 0$, le paramètre d'échelle. En admettant que le processus agrégé à 256 heures ne contient pas de zéro, il est alors possible de lui appliquer la propriété d'IE simple et l'on peut ainsi se servir de l'expression dérivée en (2.7) pour le spectre τ , soit $\tau(q) = 1 - q\mathcal{H}$, afin de déterminer la valeur du paramètre $\alpha = 1/\mathcal{H}$. En effet, en ayant ici identifié un régime d'IE monofractale de dimension 1 pour les grandes échelles, il en découle que le spectre des dimensions $D(q)$, tel que défini en (2.11), est constant pour tout q (*i.e.* que $D(q) \cong 1$), d'où $\tau(q) \cong 1 - q$. Il en ressort que $\tau(q) = 1 - q\mathcal{H} \cong 1 - q \Rightarrow \mathcal{H} \cong 1$, et par conséquent on a $\alpha \cong 1$. En ce qui concerne le coefficient d'asymétrie β , son calcul est simple également puisque le cas $\beta = 1$ est le seul qui permette de ne générer uniquement que des valeurs positives, ce qui ne nous laisse d'autre choix puisque l'on travaille sur un processus de précipitation. Quant au paramètre d'échelle s , celui-ci peut être ajusté de façon empirique étant donné qu'il n'agit qu'au niveau de l'amplitude des résultats.

L'intégrale contenue dans l'expression de la densité d'une variable aléatoire de Lévy en (2.18) ne présente de solution explicite que dans peu de cas, notamment pour $[\alpha = 1 ; \beta = 0]$ et $[\alpha = 2 ; \beta = 0]$ qui correspondent respectivement aux distributions de Cauchy et normale. Cela implique que la simulation de variables "Lévy-stable" dans le cas général doit être effectuée à l'aide d'approximations. L'algorithme privilégié ici se base sur le théorème suivant tiré du livre de Fristedt et Gray [1997], et qui dit en substance qu'une fonction de répartition $F(\bullet)$ est attirée par une distribution Lévy de paramètre α si et seulement si:

$$\lim_{y \rightarrow \infty} \frac{\int_0^y (1 - F(u)) du}{\int_0^z (1 - F(u)) du} = z^{1-\alpha} \quad (2.19)$$

Or, si l'on définit une variable aléatoire X telle que $X = \frac{U}{1-U}$, où la distribution de U est uniforme sur l'intervalle $(0,1)$, on a $F_X(x) = \frac{x}{1+x}$ et l'on vérifie aisément que la distribution de X satisfait l'équation (2.19) pour $\alpha = 1$. De plus, comme X est définie sur l'intervalle $(0,\infty)$, il est donc possible de s'en servir pour générer des variables Lévy dont le paramètre β vaut 1. En fait, le théorème donné en (2.19) est une version généralisée du théorème central limite qui nous permet d'affirmer que la distribution de la variable $Y = s \frac{\sum_{i=1}^N X_i}{N}$, où les variables X_i sont indépendantes et identiquement distribuées selon $F_X(x)$, tend vers une Lévy de paramètres $\alpha = 1, \beta = 1$ et s lorsque N tend vers l'infini.

En somme, la simulation des séries pluviométriques s'effectue en deux étapes. Dans un premier temps, il s'agit de simuler les totaux agrégés Y à 256 heures selon l'algorithme qui vient d'être décrit (*i.e.* en simulant un certain nombre N de variables X_i distribuées selon $F_X(x)$, puis en les additionnant). Afin de s'ajuster à la longueur de la série originale qui comporte 19 368 observations, on peut simuler 75 valeurs de départ, ce qui permettra d'obtenir des séries synthétiques comportant $75 \times 256 = 19\,200$ valeurs à l'échelle horaire.

Ensuite, pour chaque valeur Y , on amorce le processus de cascade multiplicative avec les paramètres déterminés au début de la section, soit $b = 2$ et $p_0 \cong 0,354$. La série résultante provient donc de la juxtaposition de ces 75 cascades. Si théoriquement, chaque cascade doit être développée jusqu'à l'infini, on peut se contenter en pratique de simuler jusqu'à quelques niveaux sous l'échelle horaire (les résultats finaux étant obtenus par agrégation à partir du dernier niveau généré). La figure 2.11 reprend de façon illustrée le modèle développé au cours de cette section.

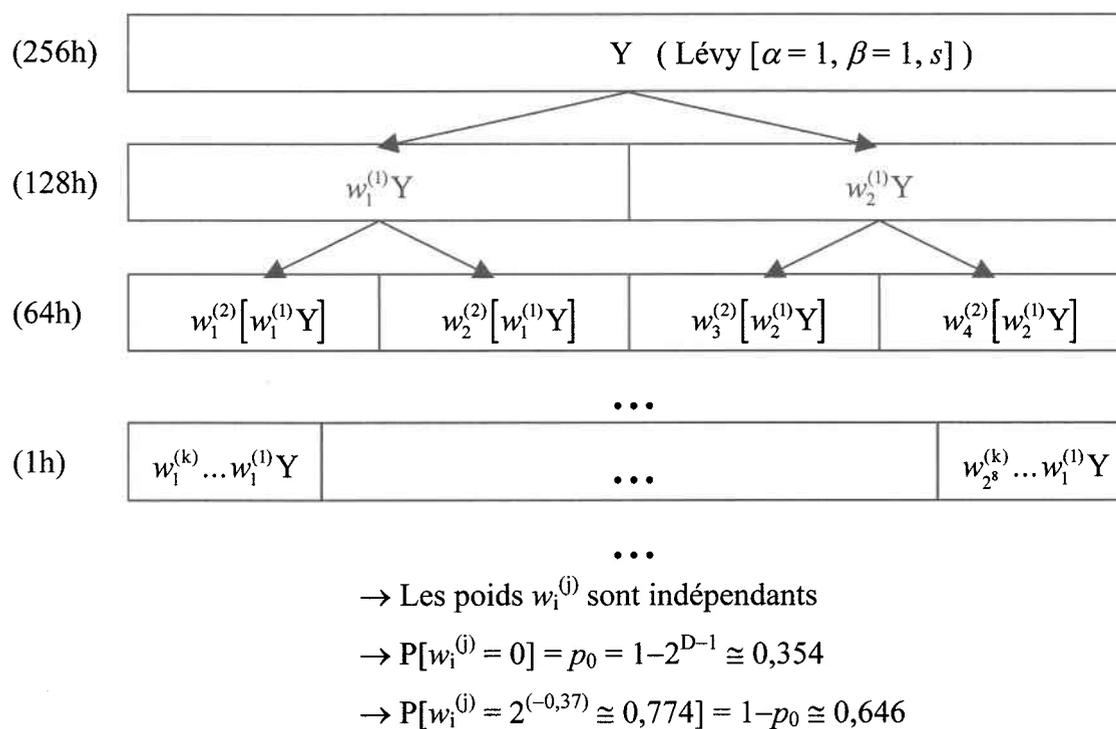


Figure 2.11: Schéma du modèle de simulation des séries de précipitation

En terminant cette section, il est important de mentionner que l'hypothèse de l'indépendance des poids n'a pas véritablement été testée. Cependant, une vérification sera effectuée de façon empirique au chapitre 3 (où l'on traite de la validation du modèle) en comparant la distribution des intervalles secs (à différentes échelles) pour les séries simulées avec la distribution des intervalles secs de la série originale.

2.4 Compensation de l'effet de l'atome en zéro sur la distribution des poids reconstitués de la cascade

Il a déjà été fait mention à la section 2.1.3 qu'il n'est pas possible de reconstruire une cascade multiplicative en partant des niveaux inférieurs, à moins que celle-ci ne soit microcanonique. En se basant sur l'exemple de la cascade à deux branches de la figure 2.2, l'illustration de ce problème est donnée à la figure 2.12. On y voit bien qu'il faut que chaque paire de poids soit complémentaire afin de retomber sur les valeurs originales.

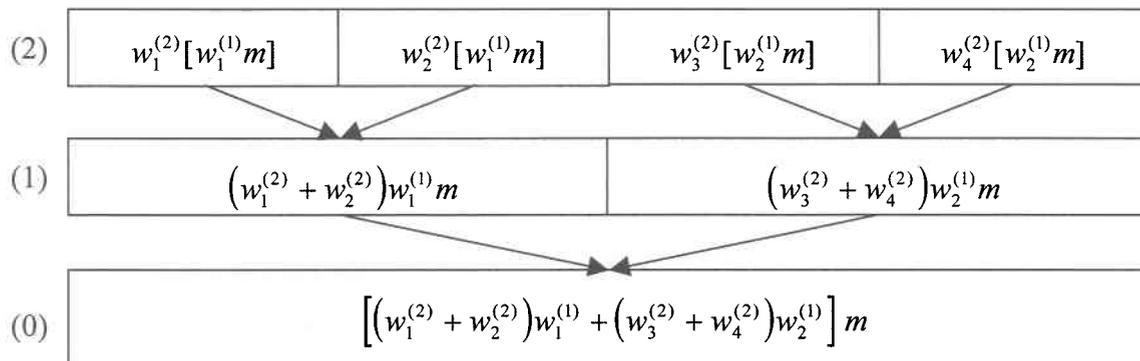


Figure 2.12: Reconstitution d'une série de données en tant que cascade à deux branches

Or, il s'agit là d'une difficulté qui a un effet au niveau de la simulation puisqu'elle vient modifier la valeur finale de la probabilité p_0 que l'on détermine à l'aide de la dimension estimée de la série (formule (2.15), section 2.1.3). En effet, si l'on suppose que les poids de la cascade de la figure 2.2 sont générés de façon indépendante et non complémentaire, les poids du niveau (1) possèdent chacun une certaine probabilité p_0 d'être nuls. On comprend cependant qu'à la figure 2.12 il faille également tenir compte de la probabilité que la somme des poids du niveau (2) soit nulle. C'est donc dire que la probabilité d'observer des valeurs nulles au niveau (1) est supérieure à p_0 lorsque l'on reconstruit la cascade par agrégation des valeurs du niveau (2). Ainsi, en agrégeant les résultats de la simulation, on observera une valeur de p_0 différente de celle estimée à partir de la série. La question qui se pose alors est

de savoir s'il est possible de corriger la situation en modifiant, par exemple, la valeur initiale de p_0 ?

Pour y répondre, il convient tout d'abord d'examiner la distribution des poids reconstitués en fonction du nombre de niveaux d'agrégation. Plus spécifiquement, on va s'intéresser à l'expression de la fonction de masse de la variable $W_+^{[k]}$ qui représente le nombre de poids supérieurs à zéro après k niveaux d'agrégation. Au cours de la construction de la cascade, la distribution de cette variable (W_+) est la même à chaque niveau ; de plus, si les poids sont générés de façon indépendante, il s'agit alors d'une binomiale de paramètres b et p , où b représente le nombre de branches et p est la probabilité que les poids soient plus grands que zéro (*i.e.* que $p = 1 - p_0$). Cette première observation permet donc d'affirmer que la distribution des poids au dernier niveau de la cascade est binomiale, *i.e.* que $W_+ \sim \text{Bin}(b, p)$.

Afin de vérifier ce résultat de façon empirique, il n'est possible cependant que de compter le nombre de valeurs positives pour chaque bloc de b valeurs. En effet, en se plaçant dans la perspective où l'on dispose comme seule information du résultat final de la simulation, on ne peut savoir si la présence de b valeurs nulles adjacentes résulte de la génération de b poids nuls au niveau précédent ou bien si cela s'est produit auparavant à un niveau supérieur (empêchant ainsi de connaître la véritable valeur de W_+).

Pour mieux fixer les idées à ce propos, imaginons la construction et la reconstruction d'une cascade à deux branches, au bout de trois niveaux de génération, telles que représentées à la figure 2.13. La partie (a) de la figure illustre la construction de la cascade. La masse initiale est donnée par $A^{(0)} (> 0)$, tandis que les valeurs $A_j^{(i)}$ ($i = 1, 2, 3$ et $j = 1, 2, \dots, 2^i$) sont les valeurs résultantes, au niveau i , de la multiplication avec les poids respectifs $w_j^{(i)}$. À la partie (b), on reprend les valeurs du dernier niveau de la cascade développée en (a) (*i.e.* $B_j^{(3)} = A_j^{(3)}$, $j = 1, 2, \dots, 8$) pour les additionner deux par deux, de façon à tenter de reconstituer la cascade de la partie (a).

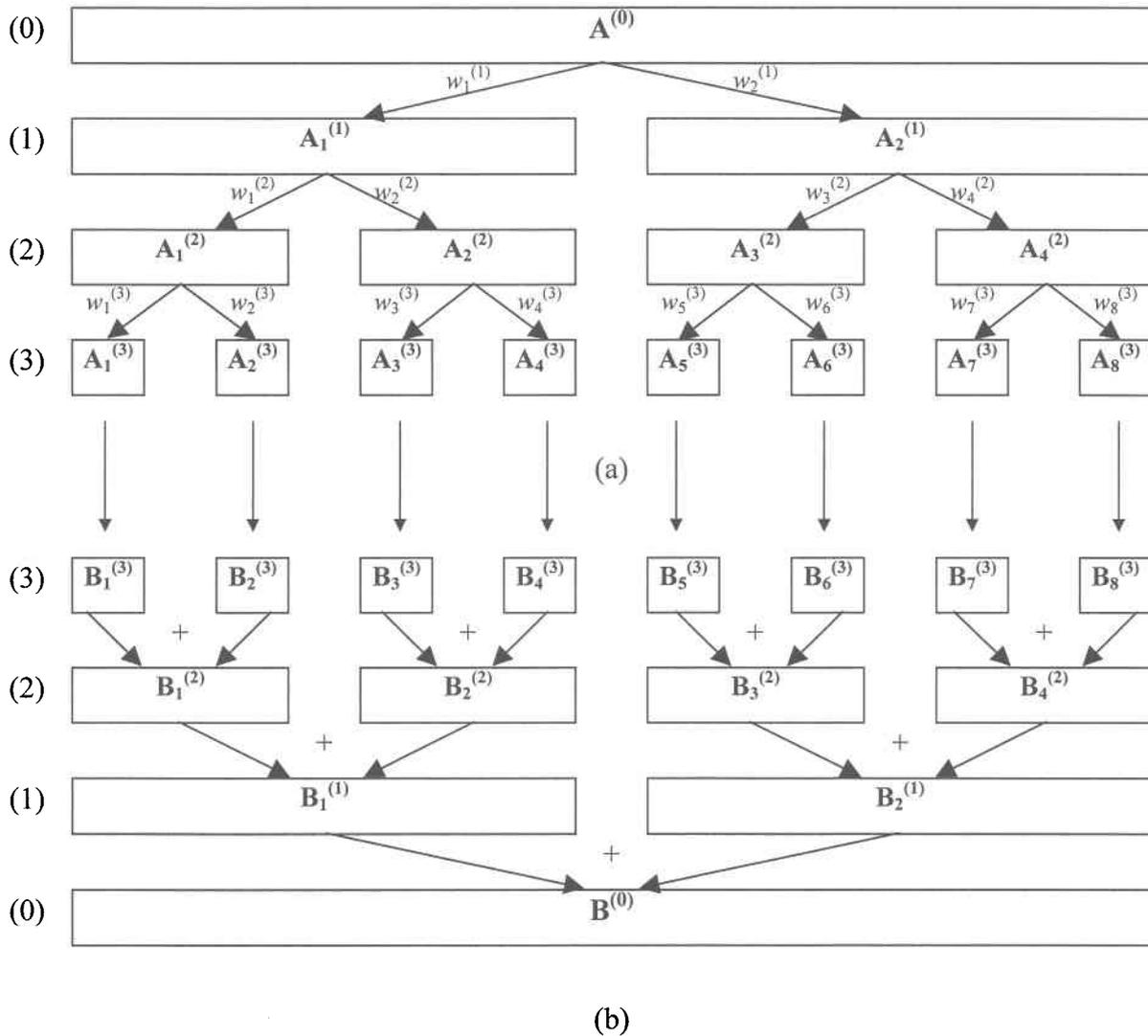


Figure 2.13: Construction (a) et reconstruction (b) de trois niveaux d'une cascade à deux branches

Pour retourner maintenant à la discussion sur la vérification empirique du fait que la distribution de la variable W_+ est binomiale, en se basant désormais sur le schéma 2.13, il faut commencer par repérer les valeurs $B_j^{(3)}$ ($j = 1, 2, \dots, 8$) qui sont supérieures à zéro. En ayant $B_j^{(3)} > 0$ pour une valeur de j donnée, on est alors assuré que $A_{j/2}^{(2)} > 0$ (si j est pair) ou encore que $A_{(j+1)/2}^{(2)} > 0$ (si j est impair), et par conséquent on sait que $w_j^{(3)} > 0$. On ne peut affirmer par contre que $w_j^{(3)} = 0$ si $B_j^{(3)} = 0$, car si $A_{j/2}^{(2)} = 0$ (ou bien $A_{(j+1)/2}^{(2)} = 0$ pour j impair), on se retrouve avec $B_j^{(3)} = 0$ peu importe la valeur de $w_j^{(3)}$. Tout cela pour dire enfin qu'on ne peut observer directement le cas $W_+ = 0$ sans faire référence au processus

générateur de la cascade. On peut toutefois établir des probabilités conditionnelles pour les cas $W_+ = 1$ et $W_+ = 2$ sachant que $W_+ > 0$. Effectivement, en admettant que $W_+ \sim \text{Bin}(2, p)$, on a :

$$\begin{aligned} \rightarrow P[W_+ = 1 \mid W_+ > 0] &= \frac{P[W_+ = 1]}{1 - P[W_+ = 0]} = \frac{2p(1-p)}{1 - (1-p)^2} = \boxed{\frac{2-2p}{2-p}} \\ \rightarrow P[W_+ = 2 \mid W_+ > 0] &= \frac{P[W_+ = 2]}{1 - P[W_+ = 0]} = \frac{p^2}{1 - (1-p)^2} = \boxed{\frac{p}{2-p}} \end{aligned} \quad (2.20)$$

On peut donc s'attendre à ce que les nombres de cas $W_+ = 1$ et $W_+ = 2$ au dernier niveau d'une cascade à deux branches (avec des poids indépendants) respectent ces proportions par rapport au nombre total de cas $W_+ > 0$. Que se passe-t-il cependant après un niveau d'agrégation, *i.e.* comment s'exprime la distribution de la variable $W_+^{[1]}$? Encore là, il n'est possible que d'observer les cas $W_+^{[1]} = 1$ et $W_+^{[1]} = 2$. On cherche alors de nouveau à établir les probabilités conditionnelles à l'événement $W_+^{[1]} > 0$. Prise isolément, cette condition permet d'affirmer dans un premier temps que $A_j^{(1)} > 0$ (pour $j = 1$ ou 2 , *cf.* figure 2.13) ; puis, en supposant toujours que les poids sont générés indépendamment les uns des autres, le chemin menant à la réalisation de l'événement $B_j^{(2)} > 0$ ($j = 1, 2, 3$ ou 4) s'exprime par la combinaison des probabilités $P[A_j^{(2)} > 0]$ et $P[W_+ > 0]$, *i.e.* que :

$$\begin{aligned} P[B_j^{(2)} > 0 \mid A_i^{(1)} > 0] &= P[A_j^{(2)} > 0] \cdot P[W_+ > 0] = p[1 - (1-p)^2] \\ (j = 1, 2, 3 \text{ ou } 4 ; i = 1 \text{ si } j = 1 \text{ ou } 2 ; i = 2 \text{ si } j = 3 \text{ ou } 4) \end{aligned} \quad (2.21)$$

Ainsi, en posant $p^{[1]} = p[1 - (1-p)^2]$, on en déduit que $W_+^{[1]} \sim \text{Bin}(2, p^{[1]})$ et il suffit de remplacer p par $p^{[1]}$ dans les formules de (2.20) afin d'obtenir les probabilités conditionnelles recherchées, soit :

$$\begin{aligned} \rightarrow P[W_+^{[1]} = 1 \mid W_+^{[1]} > 0] &= \frac{2 - 2p^{[1]}}{2 - p^{[1]}} \\ \rightarrow P[W_+^{[1]} = 2 \mid W_+^{[1]} > 0] &= \frac{p^{[1]}}{2 - p^{[1]}} \end{aligned} \quad (2.22)$$

En poussant l'exercice à un second niveau d'agrégation, on détermine de façon analogue à la démarche qui a mené à (2.21) que $p^{[2]} = P[A_j^{(1)} > 0] \cdot P[W_+^{[1]} > 0] = p[1 - (1 - p^{[1]})^2]$, où $j = 1$ ou 2 . Les probabilités conditionnelles se déduisent dès lors selon le calcul établi en (2.20), sachant que $W_+^{[2]} \sim \text{Bin}(2, p^{[2]})$.

On généralise facilement ces concepts pour une cascade à n niveaux et à b branches, et dont les poids sont indépendants. En effet, en répétant l'essence de la démarche établie jusqu'ici, on en arrive rapidement à la conclusion que $W_+^{[k]} \sim \text{Bin}(b, p^{[k]})$ pour $k = 0, 1, 2, \dots, n-1$, où $p^{[k]} = p[1 - (1 - p^{[k-1]})^b]$ pour $k = 1, 2, \dots, n-1$, en prenant $p^{[0]} = p$ (bien qu'on ne puisse définir la variable $W_+^{[n]}$, on comprend que l'expression $[1 - (1 - p^{[n-1]})^b]$ représente la probabilité de retrouver une masse initiale non nulle). Il devient donc intéressant d'étudier la suite de nombres réels définie par $(p^{[k]})$ puisqu'elle permet potentiellement de répondre à la question posée au début de cette section, à savoir comment corriger la valeur initiale de p_0 de façon à ce que les résultats de la simulation respectent la valeur de p_0 estimée à partir des données originales.

Commençons par démontrer que $(p^{[k]})$ forme une suite décroissante. En effet, en partant du fait que $0 < p < 1$ (car il s'agit d'une probabilité), on peut procéder par récurrence:

$$1) \quad p^{[1]} < p^{[0]} \equiv p$$

$$\triangleright 0 < p < 1 \Rightarrow (1 - p)^b < 1 \Rightarrow 1 - (1 - p)^b < 1 \Rightarrow p^{[1]} = p[1 - (1 - p)^b] < p <$$

$$2) p^{[k-1]} < p^{[k-2]} \Rightarrow p^{[k]} < p^{[k-1]} \quad (k > 1)$$

$$\begin{aligned} \triangleright p^{[k-1]} < p^{[k-2]} &\Rightarrow (1 - p^{[k-1]})^b > (1 - p^{[k-2]})^b \Rightarrow 1 - (1 - p^{[k-1]})^b < 1 - (1 - p^{[k-2]})^b \\ &\Rightarrow p^{[k]} = p \left[1 - (1 - p^{[k-1]})^b \right] < p \left[1 - (1 - p^{[k-2]})^b \right] = p^{[k-1]} \triangleleft \end{aligned}$$

On démontre ensuite aussi facilement que $p^{[k]} > 0$ pour tout k puisque:

$$1) p^{[1]} > 0$$

$$\triangleright 0 < p < 1 \Rightarrow (1 - p)^b > 0 \Rightarrow p^{[1]} = p \left[1 - (1 - p)^b \right] > 0 \triangleleft$$

$$2) p^{[k-1]} > 0 \Rightarrow p^{[k]} > 0 \quad (k > 1 ; \text{ on sait aussi que } p^{[k-1]} < 1 \text{ car } p^{[k-1]} < p < 1)$$

$$\triangleright 0 < p^{[k-1]} < 1 \Rightarrow (1 - p^{[k-1]})^b > 0 \Rightarrow p^{[k]} = p \left[1 - (1 - p^{[k-1]})^b \right] > 0 \triangleleft$$

En résumé, la suite $(p^{[k]})$ est décroissante et minorée (par 0), ce qui prouve qu'elle est convergente, *i.e.* que $\lim_{k \rightarrow \infty} p^{[k]}$ existe. Si l'on dénote cette limite par $p^{[\infty]}$, elle peut se calculer de la façon suivante:

$$p^{[\infty]} = p \left[1 - (1 - p^{[\infty]})^b \right] \quad (2.23)$$

Si $b = 2$ par exemple, on obtient l'équation quadratique $p^{[\infty]} \left(\left[2 - \frac{1}{p} \right] - p^{[\infty]} \right) = 0$ en remplaçant dans (2.23), équation dont les racines sont données par 0 et $2 - (1/p)$. On en conclut que $p^{[\infty]} = 0$ si $p \leq 0,5$ et que $p^{[\infty]} = 2 - (1/p)$ lorsque $p > 0,5$. C'est donc dire que la cascade finit par disparaître lorsqu'on simule avec un p inférieur ou égal à 0,5 (*i.e.* que l'on se retrouve éventuellement avec des valeurs nulles partout au fur et à mesure que l'on avance dans la construction de la cascade). On peut démontrer qu'en général, pour b branches, $p^{[\infty]} = 0$ lorsque $p \leq 1/b$. Le calcul de $p^{[\infty]}$ (lorsque $p > 1/b$) devient plus complexe

par contre pour $b > 2$, c'est pourquoi on préférera en pratique en revenir au cas $b = 2$, dont les formules reliant la valeur de $p^{[\infty]}$ à celle de p et vice versa sont données par:

$$\begin{cases} p^{[\infty]} = 2 - \frac{1}{p} & (\text{où } p > 0,5) \\ \Rightarrow p = \frac{1}{2 - p^{[\infty]}} \end{cases} \quad (2.24)$$

L'expression de la valeur de p en fonction de $p^{[\infty]}$ donnée en (2.24) est intéressante puisqu'elle met en évidence un autre moyen permettant de tester l'hypothèse d'IE sur une série de données, puis par le fait même d'estimer sa dimension. En effet, si l'on s'imagine que la série considérée provient de la génération d'une cascade à deux branches (avec des poids indépendants) développée jusqu'à l'infini, l'étude de la distribution des poids reconstitués de la série devrait faire ressortir la valeur de $p^{[\infty]}$. Afin de valider la présence de l'IE, il faut toutefois s'assurer que la valeur de $p^{[\infty]}$ ainsi estimée reste à peu près semblable sur plusieurs échelles. Puis, en utilisant (2.24) pour découvrir le p_0 associé (car $p_0 = 1-p$), il ne reste plus qu'à déduire la dimension à l'aide de (2.15), d'où découle cette formule:

$$D = 1 + \frac{\log(1-p_0)}{\log(2)} = 1 + \frac{\log(p)}{\log(2)} = 1 - \frac{\log(2-p^{[\infty]})}{\log(2)} \quad (2.25)$$

Les résultats de cet exercice effectué sur la série de données horaires de l'Alabama sont reportés dans le tableau 2.2. Pour reprendre la notation utilisée au cours de cette section, il s'agit d'évaluer les proportions $P[W_+ = 1 \mid W_+ > 0]$ et $P[W_+ = 2 \mid W_+ > 0]$ pour différents pas de temps (1 heure, 2 heures, 4 heures, etc...). Les formules de (2.20) permettant de calculer la valeur de $p^{[\infty]}$ reliée à ces probabilités, on peut ensuite estimer la dimension à partir de l'expression donnée en (2.25). Le tableau 2.2 nous indique donc que l'IE semble respectée pour des échelles n'excédant pas les 64 heures et que la dimension de la série est alors

estimée autour de 0,345. Cette valeur de la dimension est assez voisine de celle obtenue à la section 2.2.2, soit 0,37, bien que la zone d'IE soit plus restreinte qu'à cette section, où l'effet d'IE (*scaling*) semblait s'étendre facilement jusqu'à une centaine d'heures. Et puis l'on constate de nouveau que la dimension passe à 1 pour les grandes échelles.

Tableau 2.2: Étude de la distribution des poids reconstitués (en fonction de l'échelle d'agrégation) de la série de données pluviométriques de la station de l'Alabama

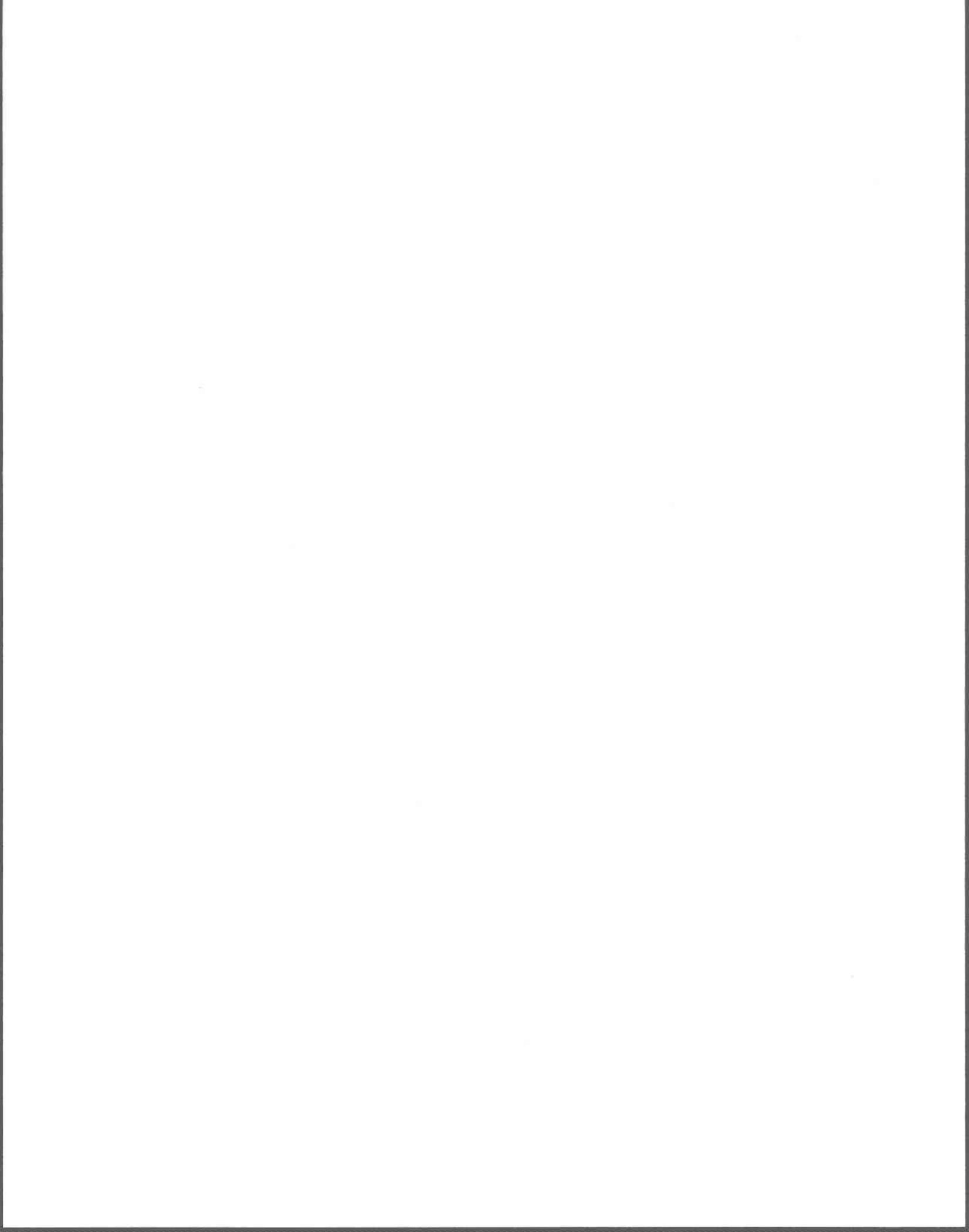
Pas de temps (heures)	$P[W_+ = 1 W_+ > 0]$	$p^{[\infty]}$	D
	$P[W_+ = 2 W_+ > 0]$		
1	2,6522	0,4299	0,3491
2	2,6239	0,4325	0,3516
4	2,6471	0,4304	0,3496
8	2,6970	0,4258	0,3454
16	2,9796	0,4016	0,3234
32	2,7317	0,4227	0,3425
64	1,7317	0,5359	0,4500
128	0,5455	0,7857	0,7199
256	0,2333	0,8955	0,8566
512	0,0	1,0	1,0

Si la dimension estimée D détermine la valeur de p à utiliser durant la simulation des séries synthétiques, il n'en demeure pas moins que la transition de la valeur de p à $p^{[\infty]}$ au cours du processus d'habillage de la cascade (*i.e.* lors de la reconstruction) résulte en une perte d'information sur la construction de la cascade. Cela se répercute tout particulièrement sur le moment d'ordre 0, qui représente la proportion espérée de valeurs non nulles dans la série. Au cours de la construction d'une cascade dont les poids sont indépendants, celui-ci s'exprime par p^k lorsqu'on se situe au $k^{\text{ième}}$ niveau de génération. En revanche, une fois revenu au même niveau au cours de l'agrégation (en admettant que la construction s'est

poursuivie jusqu'à l'infini), le moment d'ordre 0 s'exprime alors par $p^{k-1}p^{[\infty]}$, *i.e.* que celui-ci se trouve réduit par un facteur de $p^{[\infty]}/p$.

Il faut donc apporter la correction suivante afin de retrouver un moment d'ordre 0 correspondant à la valeur initiale de p au cours de la reconstruction : au premier niveau, il suffit de modifier la valeur de p en la multipliant par le facteur $p/p^{[\infty]}$, puis on utilise à nouveau le p initial pour la génération des niveaux subséquents. En conséquence, le moment d'ordre 0 pendant la construction devient $p^{k+1}/p^{[\infty]}$ au $k^{\text{ième}}$ niveau ; tandis qu'après avoir agrégé les valeurs de la cascade à partir de l'infini pour revenir au même niveau, sa valeur est diminuée d'un facteur $p^{[\infty]}/p$, *i.e.* que l'on retrouve p^k . En bref, si l'on se réfère au schéma 2.11 du modèle de simulation, on doit y inclure cette modification:

$$\begin{cases} \mathbb{P}[w_i^{(j)} > 0] = p \frac{p}{p^{[\infty]}} = \frac{p^3}{2p-1} & \text{si } j = 1 \quad (i = 1 \text{ ou } 2) \\ \mathbb{P}[w_i^{(j)} > 0] = p = 1 - p_0 & \text{si } j \geq 2 \quad (i = 1, 2, \dots, 2^j) \end{cases} \quad (2.26)$$



3. VALIDATION DU MODÈLE

3.1 Les moments statistiques de l'intensité de la précipitation à différentes échelles

Une façon de valider le modèle présenté à la section 2.3 consiste à calculer les moments des séries simulées et de les comparer avec ceux de la série originale, et ce pour différentes échelles d'agrégation. Il ne faut pas oublier cependant que la distribution de Lévy postulée pour les totaux agrégés à 256 heures (motivée par l'apparente indépendance de la série à cette échelle), de paramètre $\alpha = 1/\mathcal{H} \cong 1$, nous impose la contrainte suivante : on doit se contenter d'évaluer les moments d'ordre compris entre 0 et 1 puisque les moments de la distribution divergent pour des ordres supérieurs à 1. Cela pourrait expliquer pourquoi le spectre $\tau(q)$ estimé pour les grandes échelles commence à diverger de la monofractalité lorsque $q > 1$ (voir la figure 2.10 à la section 2.2.3).

Malgré tout, en raison du problème soulevé à la section 2.3 concernant la modification de la valeur de p_0 tout au long du processus d'habillage (*i.e.* au cours de l'agrégation au-dessous de la plus petite échelle modélisée) de la cascade, on constate que la distribution initiale est modifiée par des facteurs multiplicatifs (voir section 2.4 et Gupta et Waymire [1993]). On peut toutefois émettre l'hypothèse que ce facteur ne modifie pas la valeur du paramètre \mathcal{H} .

Afin de vérifier cette conjecture, vingt simulations ont été effectuées pour lesquelles les moments d'ordre q ($q = 0, 1/15, 2/15, \dots, 1$) ont été calculés aux échelles allant de une à 512 heures, *i.e.* pour tous les pas de temps 2^k , où $k = 0, 1, 2, \dots, 9$ (figures 3.1 à 3.10 respectivement). Le tout s'est réalisé selon le modèle de la section 2.3, sans oublier d'appliquer la correction pour le moment d'ordre 0 définie en (2.26). Ainsi, avec une valeur de $p = 1 - p_0 \cong 1 - 0,354 = 0,646$, celle-ci doit donc passer à $p^2/p^{[\infty]} = 0,923$ au premier niveau

de la cascade, où $p^{[\infty]} = 2 - (1/p) \cong 0,452$. Étant donné qu'on ne génère que deux valeurs au premier niveau, il a été jugé opportun d'arrondir la probabilité p de 0,923 à 1,0. Par la suite, la cascade a été développée jusqu'à une échelle de 1/32 heure, en ramenant cette fois-ci la valeur de p à 0,646. Or, après cinq niveaux d'agrégation pour revenir à l'échelle horaire, on a que la valeur théorique pour $p^{[5]}$ équivaut 0,474 (section 2.4), ce qui est déjà assez voisin de la valeur de $p^{[\infty]}$ (0,452) pour justifier que l'on interrompe la construction de la cascade à cette échelle. Il est à noter toutefois que certaines simulations ont dû être rejetées en cours de route, en raison de l'instabilité de l'algorithme de génération des variables aléatoires de Lévy (pour les totaux de départ), qui peut parfois produire des valeurs dont l'ordre de grandeur semble aberrant par rapport aux valeurs observées (et ce même en utilisant $N = 1000$). Enfin, mentionnons que le paramètre d'échelle s pour les variables Lévy a été ajusté (par essais et erreurs) à 0,11.

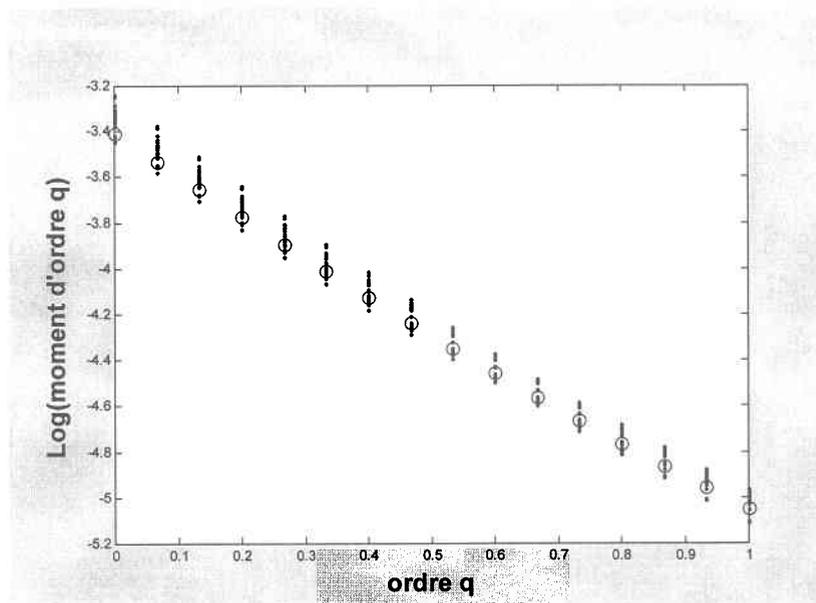


Figure 3.1: Moments des séries simulées à l'échelle d'une heure

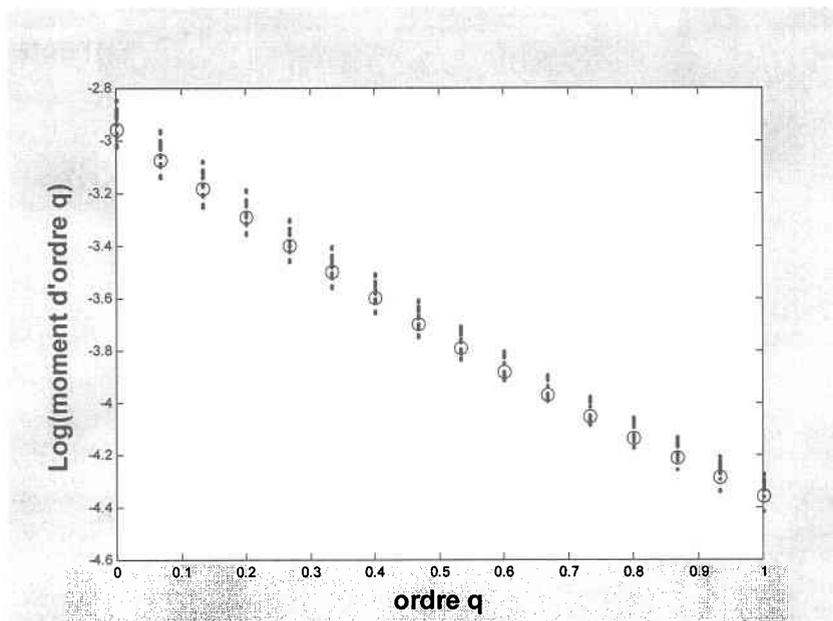


Figure 3.2: Moments des séries simulées à l'échelle de deux heures

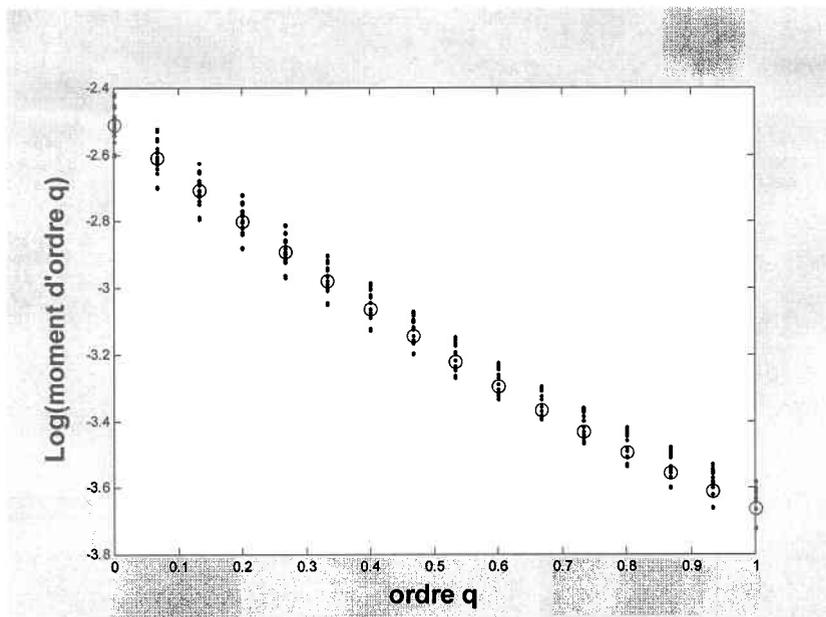


Figure 3.3: Moments des séries simulées à l'échelle de quatre heures

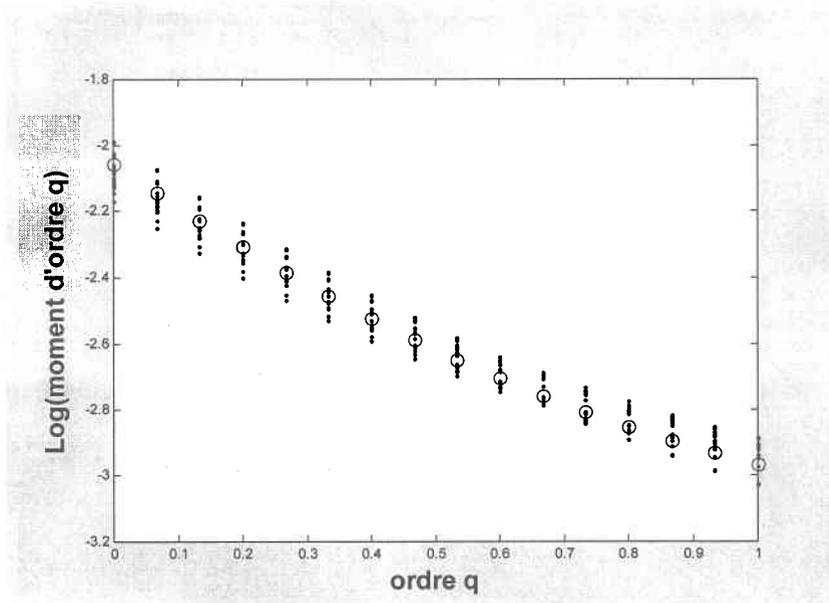


Figure 3.4: Moments des séries simulées à l'échelle de huit heures

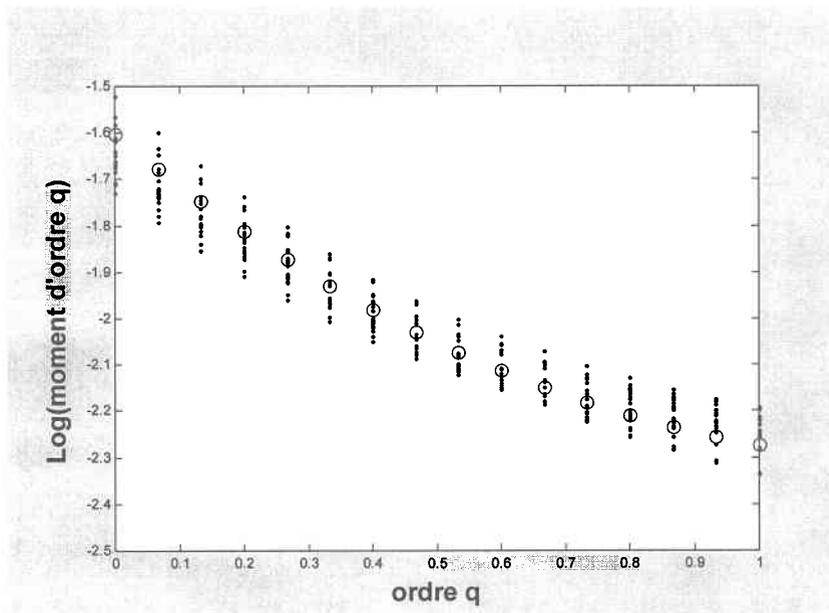


Figure 3.5: Moments des séries simulées à l'échelle de seize heures

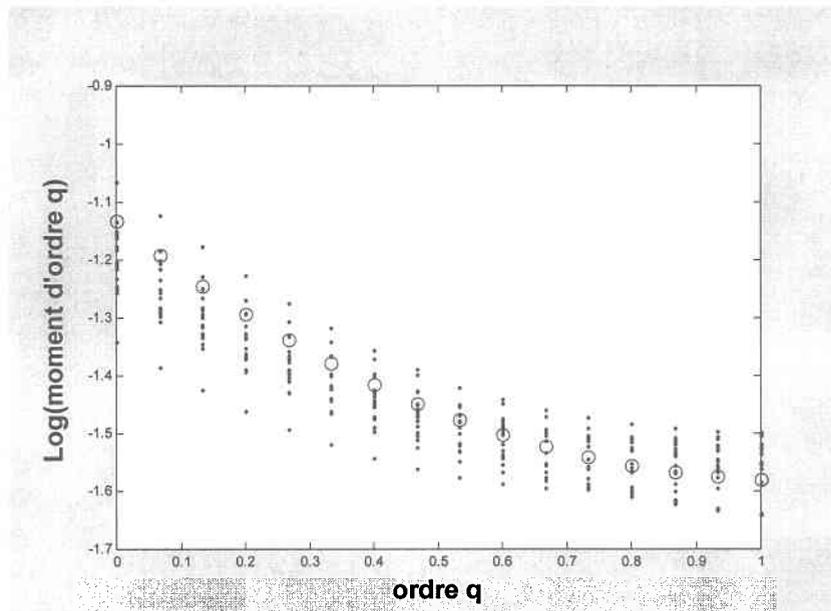


Figure 3.6: Moments des séries simulées à l'échelle de 32 heures

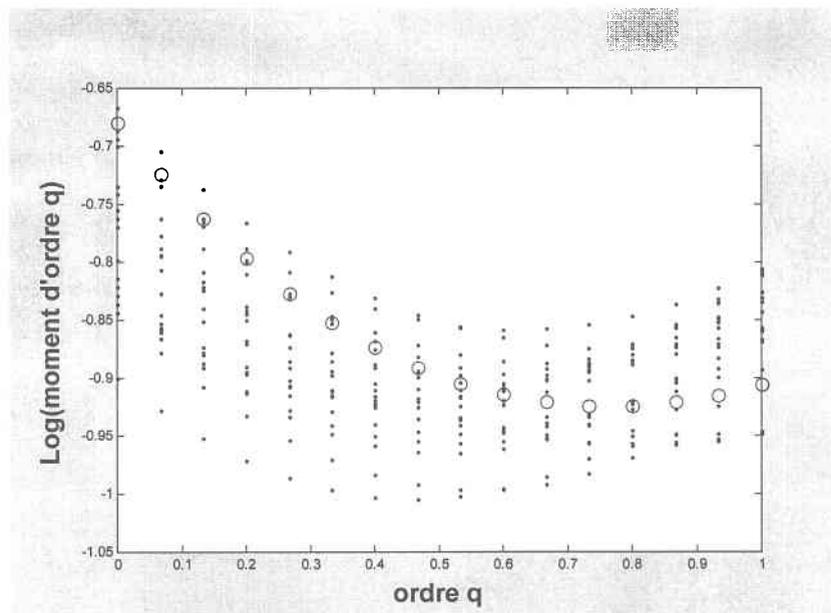


Figure 3.7: Moments des séries simulées à l'échelle de 64 heures

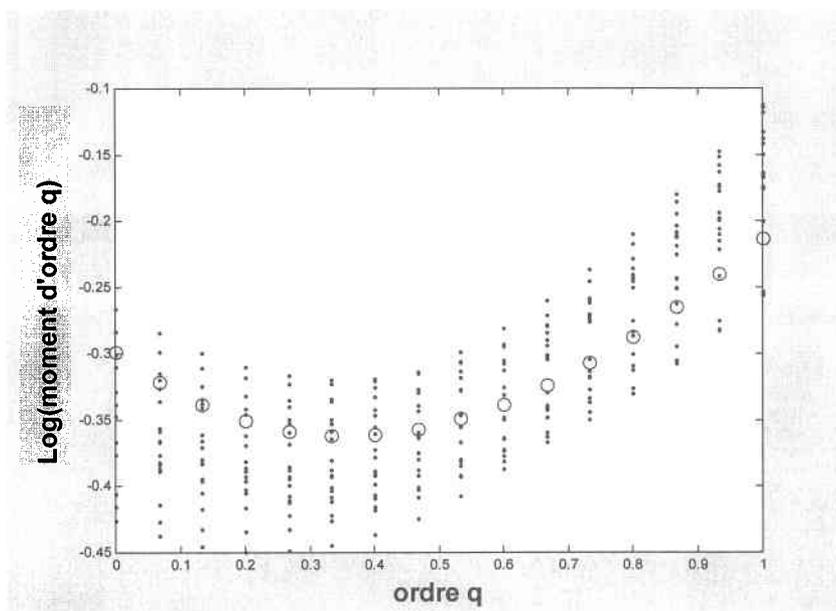


Figure 3.8: Moments des séries simulées à l'échelle de 128 heures

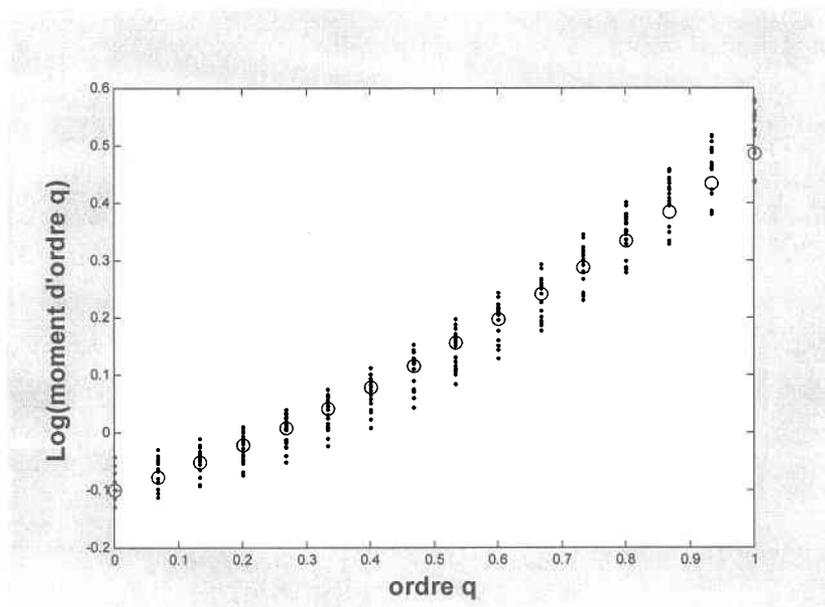


Figure 3.9: Moments des séries simulées à l'échelle de 256 heures

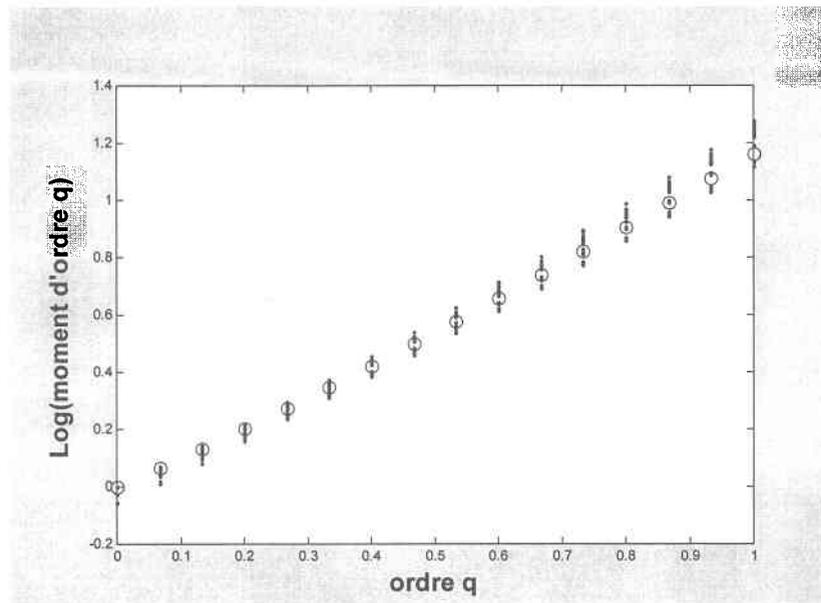


Figure 3.10: Moments des séries simulées à l'échelle de 512 heures

Les moments sur les figures 3.1 à 3.10 (chaque point correspondant à l'une des vingt simulations) ont été reportés sur une échelle logarithmique. Les cercles vides représentent les moments de la série originale. On peut donc voir que l'adéquation des moments des séries simulées est plutôt bonne pour les échelles en deçà de 64 heures, après quoi on observe des fluctuations plus accrues, quoique l'adéquation redevient bonne à partir de 256 heures, *i.e.* lorsqu'on se situe au niveau du régime d'I.E. de dimension 1 pour les grandes périodes. Ces résultats semblent donc en accord avec l'analyse du tableau 2.2 montrant que l'effet d'IE était moins bon pour les périodes comprises entre 64 et 256 heures précisément.

Le tableau 3.1 reprend d'ailleurs la même analyse des poids reconstitués qu'au tableau 2.2 pour estimer la dimension des séries simulées cette fois. Chaque entrée dans le tableau représente en fait une moyenne calculée à partir des vingt simulations. Comme prévu, on constate que la dimension des séries synthétiques est voisine de 0,37 pour les échelles inférieures à 128 heures tandis qu'elle augmente ensuite pour tendre vers 1. La valeur de la dimension estimée à l'échelle de 128 heures, n'étant voisine ni de 0,37 ni de 1, témoigne

ainsi de la transition qui s'opère dans le modèle entre les deux régimes d'IE (et qui, incidemment, se trouve à reproduire la transition observée dans la série originale).

Tableau 3.1: Étude de la distribution des poids reconstitués pour les vingt séries synthétiques sensées reproduire les caractéristiques de la série de données pluviométriques de l'Alabama

Pas de temps (heures)	$\frac{P[W_+ = 1 W_+ > 0]}{P[W_+ = 2 W_+ > 0]}$	$p^{[\infty]}$	D
1	2,1450	0,4825	0,3983
2	2,2182	0,4741	0,3904
4	2,1911	0,4772	0,3933
8	2,3431	0,4605	0,3775
16	2,4517	0,4493	0,3670
32	2,3138	0,4636	0,3805
64	2,3470	0,4601	0,3772
128	0,8767	0,6952	0,6162
256	0,1619	0,9251	0,8958
512	0,0198	0,9902	0,9859

3.2 Les distributions statistiques des périodes sèches à différentes échelles

Définissons la variable aléatoire $D(\Delta t)$ comme étant la longueur des périodes sèches (*i.e.* la longueur des séquences de zéros consécutifs) que l'on retrouve à l'intérieur de la série de données agrégée à l'échelle de temps Δt . Les fonctions de répartition empiriques des variables $D(\Delta t)$ associées aux données de l'Alabama pour les pas de temps $\Delta t = 1$ heure, 2 heures, 4 heures et 8 heures (le nombre d'intervalles secs diminuant significativement lorsque l'échelle d'agrégation est supérieure à 8 heures) sont illustrées à la figure 3.11. Les

graphiques y sont présentés en coordonnées logarithmiques pour les valeurs de D en abscisse tandis que les probabilités empiriques au non-dépassement correspondantes $P[D \leq d]$ sont reportées en ordonnée. Ces probabilités ont été estimées par la formule de Weibull, donnée par $P[D \leq d] = \text{rang}(d)/(N+1)$, où N est la taille de l'échantillon (*i.e.* le nombre total d'intervalles secs observés).

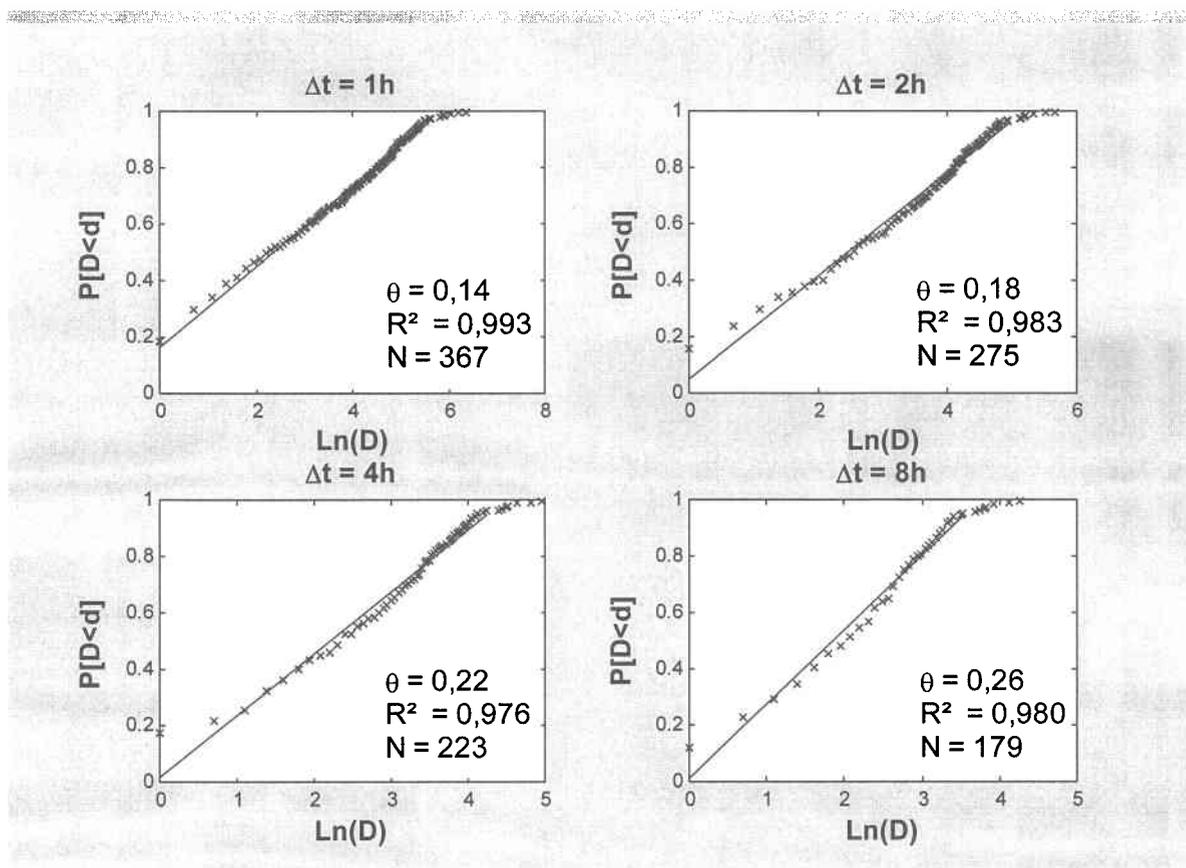


Figure 3.11: Étude des distributions empiriques de la longueur des périodes sèches, pour différentes échelles d'agrégation, associées aux données de l'Alabama

On constate tout d'abord que les graphiques de la figure 3.11 ont à peu près la même allure pour toutes les échelles considérées, *i.e.* que l'on observe un alignement des points assez bon, excepté pour les grandes valeurs de D puisqu'à partir d'un certain seuil une courbure apparaît. Pour les zones où l'alignement est bon, les pentes θ ont été calculées par

régression et sont inscrites sur les différents graphiques (les coefficients de détermination R^2 et les tailles d'échantillon N sont aussi donnés à titre indicatif).

Voyons maintenant ce qui se passe si l'on refait la même étude sur les séries simulées cette fois. Tout comme à la section précédente, un total de vingt séries synthétiques de précipitation ont été utilisées pour accomplir cette tâche dont les résultats sont présentés à la figure 3.12. Disposant de vingt séries, les tailles des échantillons N s'en trouvent donc substantiellement augmentées pour l'estimation des probabilités empiriques, ce qui permet d'expliquer l'allure plus régulière des graphiques de la figure 3.12.

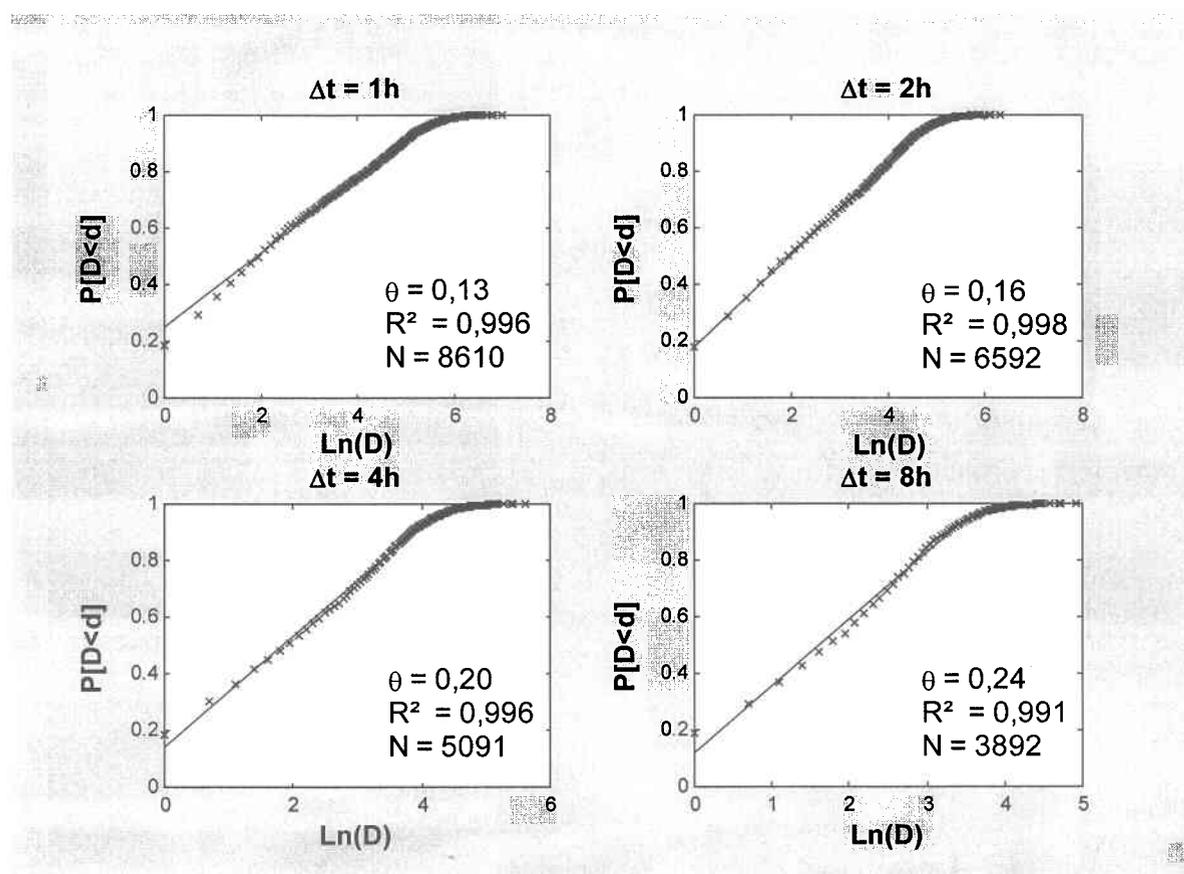


Figure 3.12: Étude des distributions empiriques de la longueur des périodes sèches, pour différentes échelles d'agrégation, associées aux vingt séries synthétiques de précipitation

La première constatation qui ressort lorsqu'on compare les résultats de la figure 3.12 avec ceux de la figure 3.11, c'est que les graphiques exhibent la même allure générale, soit un bon alignement des points jusqu'à un certain seuil où une courbure survient par la suite. Les différents seuils se situent d'ailleurs à peu près aux mêmes endroits dans les deux cas. Une fois de plus, les pentes θ à la figure 3.12 ont été calculées par régression et l'on remarque que ces dernières sont comparables avec celles calculées précédemment, bien qu'elles soient toutes inférieures. Les coefficients de détermination R^2 sont tous très bon, ce qui est dû au plus grand nombre de valeurs disponibles, ainsi qu'à l'effet de la représentation en coordonnées logarithmiques qui réduit les écarts entre les points.

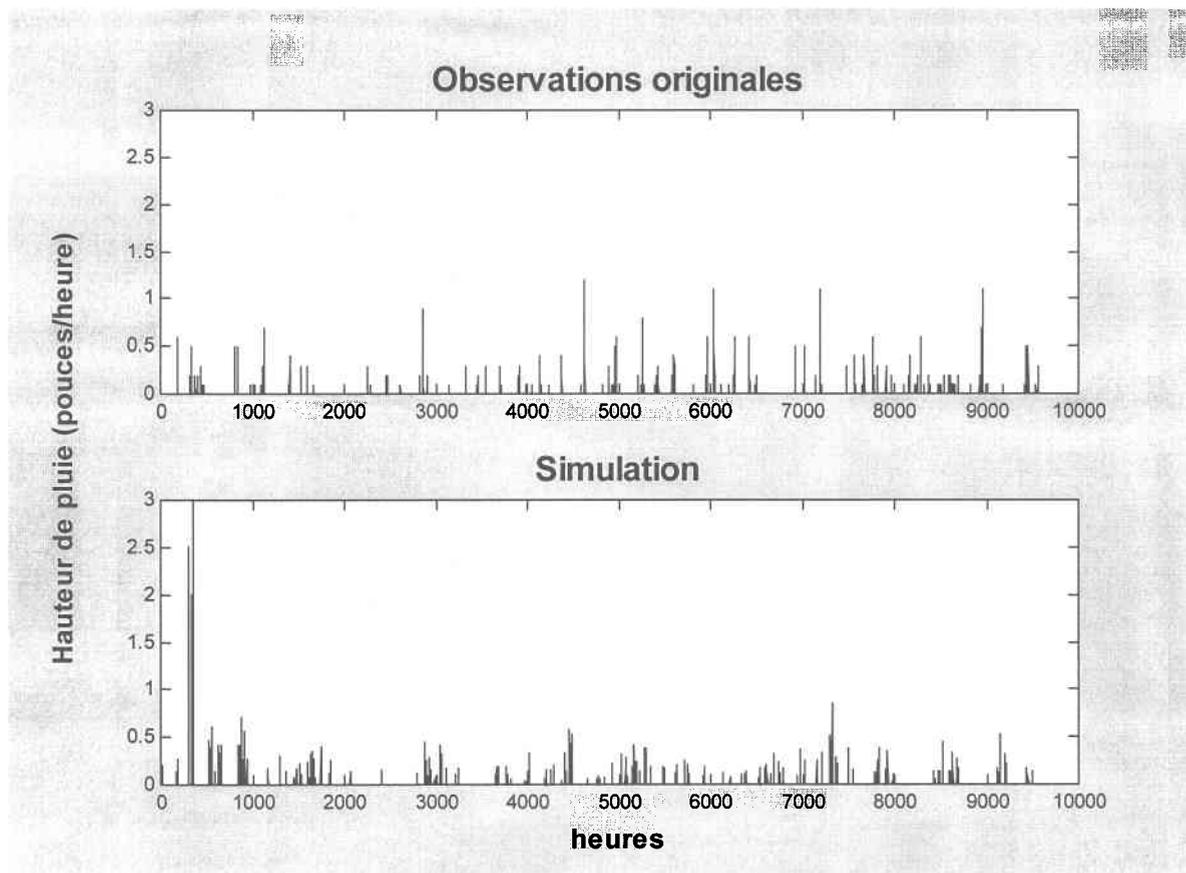


Figure 3.13: Relevés pluviométriques de l'Alabama vs hauteurs de pluie simulées (9600 premières observations présentées)

La concentration des événements pluvieux semble plus importante dans les séries simulées. Il s'agit là d'un phénomène que l'on peut détecter visuellement en observant le graphique d'une série simulée et en le comparant à celui de la série originale (figure 3.13). À l'appui de cette affirmation, il est intéressant de noter que le nombre de périodes sèches plus petites ou égales à huit heures est de 225 en moyenne pour l'ensemble des vingt simulations effectuées ici, tandis que l'on en compte seulement que 176 dans les données originales.

Tel que suggéré à la section 2.1.3, l'ajout d'une certaine dépendance au niveau des poids de la cascade constituerait une façon de s'attaquer à ce problème (Cârsteanu et Fofoula-Georgiou [1996]). Néanmoins, il faudrait étudier auparavant l'effet de l'atome en zéro (*i.e.* l'existence d'une probabilité non nulle d'observer des valeurs nulles de précipitation) sur l'évaluation des moments conjoints, en particulier sur l'autocorrélation des poids, puisque l'analyse réalisée à la section 2.4 n'a permis pour l'instant que d'explicitier les moments marginaux de la distribution des poids reconstitués.

4. PERSPECTIVES DE RECHERCHE :

LES COURBES IDF

(INTENSITÉ-DURÉE-FRÉQUENCE)

4.1 Le concept de courbe IDF

Les courbes IDF (Intensité-Durée-Fréquence) constituent un outil permettant d'évaluer les risques inhérents à l'émergence d'événements extrêmes. Comme leur nom l'indique, elles relient l'intensité i (du phénomène à l'étude) à la durée d ainsi qu'à la période moyenne de retour T , période qui se définit comme étant l'inverse de la fréquence (*i.e.* l'inverse de la probabilité au dépassement de l'événement sur une unité de temps).

Même si trois variables interviennent dans leur construction, on les représente habituellement en deux dimensions en fixant au préalable certaines valeurs pour la période moyenne de retour T , puis en traçant pour chaque valeur de T la courbe reliant l'intensité i à la durée d . Néanmoins, on peut s'attendre à voir de plus en plus de représentations en trois dimensions des courbes IDF grâce aux outils informatiques qui ne cessent de s'améliorer.

4.2 Les courbes IDF obtenues à partir des modèles invariants d'échelle

Les courbes IDF que l'on observe empiriquement pour la précipitation s'apparentent à des droites lorsqu'on les représente en coordonnées log-log. Cette observation ressort de la plupart des études qui ont été effectuées en ce sens. C'est pourquoi l'on modélise souvent ces courbes à l'aide de lois de puissance, *i.e.* que les équations de ces courbes sont alors données sous la forme de relations empiriques telles que:

$$T \propto i^m d^n \quad (4.1)$$

où m et n représentent des paramètres d'ajustement. Une justification théorique pour ces relations en lois de puissance peut être obtenue dans le cadre de l'invariance d'échelle (Bendjoudi *et al.* [1997]), qui comme on le sait maintenant, est fréquemment observée dans les processus de précipitation.

En effet, en se référant aux définitions (2.4) et (2.6) données à la section 2.1.1 (concernant l'invariance d'échelle et le spectre τ respectivement), puis en définissant $R(\Delta t)$ comme l'intégrale du processus d'intensité de la précipitation $\{I(t)\}$ sur l'intervalle de temps Δt , il est possible de dériver l'expression suivante pour la période moyenne de retour T :

$$\begin{aligned} T &= \frac{d}{P[I(d) > i]} = \frac{d}{P\left[\frac{1}{d}R(d) > i\right]} = \frac{d}{P\left[\left(\frac{d}{d_0}\right)^{\tau(0)} \frac{1}{d}R(d_0) > i\right]} = \frac{d}{P\left[d_0^{\tau(0)}R(d_0) > d^{1-\tau(0)}i\right]} \\ &= \frac{d}{1 - F_{d_0^{\tau(0)}R(d_0)}\left(d^{1-\tau(0)}i\right)} \end{aligned} \quad (4.2)$$

où $F_R(r)$ représente la fonction de répartition de la variable R et d_0 est une durée de référence. La fonction de survie pour R donnée par $1 - F_R(r)$ est fonction de $r^{-\alpha}$, où α représente encore le paramètre de la distribution de Lévy employée dans le modèle de simulation (section 2.3). On en déduit, en remplaçant dans (4.2), la relation qui suit pour T :

$$T \propto \frac{d}{\left[d^{1-\tau(0)}i\right]^{-\alpha}} = i^\alpha d^{1+\alpha(1-\tau(0))} \quad (4.3)$$

En substituant dans (4.3) les résultats $\hat{\tau}(0) = 0,37$ (section 2.2.2) et $\alpha \cong 1$ (section 2.3), on obtient donc $T \propto id^{1,63}$.

Puisque la plupart des travaux récents tendent à démontrer que la précipitation est un processus anisotrope invariant d'échelle dans le champ spatio-temporel, cela conduit naturellement vers l'introduction du concept d'invariance spatiale dans les courbes IDF, qui deviendraient alors des courbes Intensité-Durée-Surface-Fréquence (IDAF), et dont l'expression est donnée simplement par:

$$T \propto i^m d^n A^p \quad (4.4)$$

Les courbes de conversion "Hauteur-Surface" est l'outil qui a été privilégié jusqu'ici afin d'inclure la dimension spatiale dans le contexte des courbes IDF, la principale raison étant que les hypersurfaces ne sont pas très commodes à représenter graphiquement. Toutefois, sous l'hypothèse d'invariance d'échelle appliquée aux échelles spatiales, une nouvelle variable peut être introduite, soit $G = A^z d$, où z représente un exposant d'échelle dynamique (tel que calculé par, *e.g.*, Marsan *et al.* [1996] ou encore Venugopal *et al.* [1999]). L'expression de la période moyenne de retour T devient alors:

$$T \propto i^a G^{1+a(3-\tau(0))} \quad (4.5)$$



5. CONCLUSION

Le travail accompli dans le cadre de ce mémoire a permis de démontrer l'applicabilité de modèles invariants d'échelles pour reproduire le phénomène de précipitation atmosphérique. En effet, la propriété d'IE a été détectée sur une série de données de précipitations horaires en provenance de l'Alabama. La méthode employée fait appel à l'estimation du spectre de Rényi, qui inclut comme cas particulier l'algorithme du comptage de boîtes permettant de déterminer la dimension d'une série chronologique. L'utilisation du spectre de Rényi a aussi permis de définir deux régimes distincts d'IE pour les données de l'Alabama, soit un pour les petites échelles (1 à 222 heures), dont la dimension a été estimée à 0,37, et un autre pour les grandes échelles (supérieures à 222 heures) de dimension 1.

À partir d'un modèle de simulation simple s'appuyant sur le principe des cascades multiplicatives, on a pu mettre en évidence la capacité de reproduction des moments statistiques de la série originale à différentes échelles, de même qu'une adéquation plutôt satisfaisante à la distribution observée pour les périodes sèches (ce qui constitue une caractéristique importante pour une éventuelle intégration à des modèles de gestion de la ressource eau). L'étude de l'effet de l'atome en zéro sur la distribution des poids reconstitués de la cascade a permis quant à elle de bien comprendre comment la proportion de zéros dans la cascade est affectée lors du processus d'habillage, de façon à y remédier.

Le potentiel d'exploitation de ce type de modèles semble donc bien établi, particulièrement au niveau de la prévision des événements extrêmes par le biais des courbes IDF. En effet, le cadre théorique exposé au chapitre 4 illustre bien la possibilité d'adapter la construction de ces courbes dans le contexte d'un modèle invariant d'échelle. De plus, en permettant la généralisation du concept de courbe IDF aux champs spatio-temporels de précipitation, voilà qui constitue une motivation supplémentaire à l'emploi de tels modèles.

Néanmoins, considérant que les aspects théoriques de l'invariance d'échelle traités dans ce mémoire sont demeurés plutôt élémentaires, il y a encore beaucoup de place pour des développements futurs. On peut penser ici notamment aux aspects multifractals de l'IE, dont il a été fait brièvement mention à la section 2.1.1. Ou encore à une étude plus en profondeur de l'impact de l'atome en zéro sur les poids reconstitués de la cascade (de façon à bien comprendre son influence sur l'autocorrélation des poids par exemple, dans le but de mieux s'ajuster à la distribution des intervalles secs).

Bref, les développements récents s'appuyant sur la propriété d'IE pour la modélisation des précipitations laissent entrevoir un nouveau champ de recherche intéressant et stimulant, puisqu'il offre une alternative prometteuse aux modèles de précipitation se basant sur des méthodes plus classiques (tels les processus ponctuels) et dont la performance laisse plus souvent qu'autrement à désirer.

BIBLIOGRAPHIE

Abi-Zeid, I. (1997). *La modélisation stochastique des étiages et de leurs durées en vue de l'analyse du risque* [Thèse de Doctorat es Sciences (Eau)]. Québec, Université du Québec, 278 p.

Anderson, R. L. (1941). Distribution of the serial correlation coefficients. *Annals of Mathematical Statistics*, 8(1): 1-13.

Bendjoudi, H., P. Hubert, D. Schertzer et S. Lovejoy (1997). Interprétation multifractale des courbes intensité-durée-fréquence des précipitations. *Comptes rendus de l'Académie des Sciences de Paris (Sciences de la terre et des planètes)*, 325: 323-326.

Bendjoudi, H. et P. Hubert (1998). À propos de la distribution statistique des cumuls pluviométriques annuels. Faut-il en finir avec la normalité? *Revue des Sciences de l'Eau*, 4: 617-630.

Cârsteanu, A. et E. Foufoula-Georgiou (1996). Assessing dependence among weights in a multiplicative cascade model of temporal rainfall. *Journal of Geophysical Research*, 101D(21): 26 363-26 370.

Cox, D. R. et V. Isham (1980). *Point processes*. New York, Chapman and Hall, 188 p.

Daley, D. J. et D. Vere-Jones (1988). *An introduction to the theory of point processes*. New York, Springer-Verlag, 702 p.

Devaney, R. L. (1992). *A first course in chaotic dynamical systems*. Don Mills (Ontario), Addison-Wesley, 302 p.

Fristedt, B. et L. Gray (1997). *A modern approach to probability theory*. Boston, Birkhäuser, 756 p.

Grenier, D. (1999). *Évaluation de la performance de modèles ponctuels de précipitation à l'aide des courbes IDF* [Mémoire de Maîtrise es Sciences (Eau)]. Québec, Université du Québec, 97 p.

Gupta, V. K. et E. C. Waymire (1990). Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research*, 95D(3): 1999-2009.

Gupta, V. K. et E. C. Waymire (1993). A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology*, 32(2): 251-267.

Hentschel, H. G. E. et I. Procaccia (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica*, 8D: 435-444.

Hubert, P. et J. P. Carbonnel (1989). Dimensions fractales de l'occurrence de pluie en climat soudano-sahélien. *Hydrologie Continentale*, 4: 3-10.

Kavvas, M. L. et J. W. Delleur (1981). A stochastic cluster model of daily rainfall sequences. *Water Resources Research*, 17(4): 1151-1960.

Lamperti, J. (1962). Semi-stable stochastic processes. *Trans American Mathematical Society*, 104: 62-78.

Le Cam, L. (1961). *A stochastic theory of precipitation* [paper presented at the Fourth Berkeley Symposium on Mathematics, Statistics, and Probability]. Berkeley (Calif.), University of California.

Lévy, P. (1937). *Théorie de l'addition des variables aléatoires*. Paris, Gauthier-Villars, 385 p.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*. San Francisco, Freeman, 461 p.

Mandelbrot, B. B. (1989). *Les objets fractals : forme, hasard et dimension* (3^{ème} éd., revue et augmentée d'un "Survole du langage fractal"). Paris, Flammarion, 268 p.

Marsan, D., D. Schertzer et S. Lovejoy (1996). Causal space-time multifractal processes : predictability and forecasting of rain fields. *Journal of Geophysical Research*, 101D(21): 26 333-26 346.

Neyman, J. et E. L. Scott (1958). A statistical approach to problems of cosmology. *Journal of the Royal Statistical Society*, 20B: 1-43.

Olsson, J., J. Niemczynomicz, R. Berndtsson et M. Larson (1992). An analysis of the rainfall time structure by box-counting – some practical implications. *Journal of Hydrology*, 137: 261-277.

Olsson, J., J. Niemczynomicz et R. Berndtsson (1993). Fractal analysis of high-resolution rainfall time series. *Journal of Geophysical Research*, 98D(12): 23 265-23 274.

Parzen, E. (1962). *Stochastic processes*. San Francisco, Holden-Day, 324 p.

Rényi, A. (1970). *Probability theory*. Londres, North-Holland Pub., 666 p.

Rodriguez-Iturbe, I., V. K. Gupta et E. C. Waymire (1984). Scale considerations in the modeling of temporal rainfall. *Water Resources Research*, 20(11): 1611-1619.

Schertzer, D. et S. Lovejoy (1987). Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research*, 92D(8): 9693-9714.

Tessier, Y., S. Lovejoy, P. Hubert, D. Schertzer et S. Pecknold (1996). Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. *Journal of Geophysical Research*, 101D(21): 26 427-26 440.

Venugopal, V., E. Foufoula-Georgiou et V. Sapozhnikov (1999). Evidence of dynamic scaling in space-time rainfall. *Journal of Geophysical Research*, 104D(24): 31 599-31 610.

Waymire, E. C. et V. K. Gupta (1981a). The mathematical structure of rainfall representations 1 : a review of the stochastic rainfall models. *Water Resources Research*, 17(5): 1261-1272.

Waymire, E. C. et V. K. Gupta (1981b). The mathematical structure of rainfall representations 2 : a review of the theory of point processes. *Water Resources Research*, 17(5): 1272-1285.

Waymire, E. C. et V. K. Gupta (1981c). The mathematical structure of rainfall representations 3 : some applications of the point process theory to rainfall processes. *Water Resources Research*, 17(5): 1287-1294.