#### Université du Québec INRS-ETE

### NOUVEAUX DÉVELOPPEMENTS DU PROCESSUS DE NEYMAN-SCOTT

Introduction d'une dépendance entre l'intensité et la durée des cellules pluvieuses et non-stationnarité

#### par GUILLAUME EVIN

Thèse présentée pour l'obtention du grade de Philosophiæ doctor (Ph.D.) en sciences de l'eau

#### Jury d'évaluation

Président du jury et

M. Alain Mailhot, professeur

examinateur interne

INRS-ETE

Examinateur externe

Université du Québec

M. Marc Fredette, professeur

Service de l'enseignement des méthodes quantitatives de gestion

HEC Montréal

Examinateur externe

M. Christian Onof, professeur

Dept. of Civil and Environmental Engineering

Imperial College London

Directeur de recherche

Mme Anne-Catherine Favre, professeure

**INRS-ETE** 

Université du Québec

thèse présentée le 31 Juillet 2008

# Remerciements

J'aimerais remercier tout particulièrement Anne-Catherine Favre pour son soutien inconditionnel tout au long de cette thèse. Sa confiance et ses conseils m'ont continuellement rassuré et aidé à franchir tous les obstacles. Un grand merci au jury qui a fait un travail remarquable et dont les remarques ont grandement aidé à l'amélioration de la thèse.

Un sincère remerciement également à Bernard Bobée qui m'a introduit dans la Chaire en hydrologie statistique et m'a ainsi permis de découvrir le Québec et le monde de la recherche.

### Résumé

Depuis une vingtaine d'années, les modèles de précipitations basés sur des processus ponctuels ont fait l'objet de nombreux efforts de recherche afin de restituer les caractéristiques de la pluie à un pas de temps court (le plus souvent horaire). La plupart des modèles utilisent un processus ponctuel pour générer des amas de cellules pluvieuses (ou cluster). Ces cellules comportent elles-mêmes une durée et une intensité aléatoire. Malgré tous les efforts fournis et les différentes variations du modèle, des problèmes demeurent, par exemple dans la reproduction des extrêmes. De plus, certaines limitations du modèle restreignent son application (par exemple la stationnarité temporelle de ce type de modèle). Cette recherche propose de nouvelles généralisations pour le modèle de pluie connu sous le nom de Neyman-Scott (ou NSRPM). Plus spécifiquement, nous nous intéressons d'une part au lien entre l'intensité et la durée des cellules pluvieuses et, d'autre part, à la possibilité d'introduire une non-stationnarité dans la fréquence des averses.

Une des hypothèses du modèle de Neyman-Scott généralement acceptée est l'indépendance des caractéristiques décrivant les cellules du processus ponctuel. Dans cette thèse, nous montrons que l'introduction d'une dépendance entre les variables constituantes du modèle permet de mieux restituer les caractéristiques des précipitations (période de sécheresse, valeurs extrêmes, etc.). Ce lien est introduit à l'aide de copules cubiques et permet d'obtenir une forme de dépendance souple. Nous étudions l'impact de l'introduction de la dépendance entre la durée et l'intensité des cellules en appliquant le modèle sur de nombreux jeux de données d'origine différente (Canada, États-Unis, Suisse, France et Belgique). Ces séries de pluies horaires nous permettent de comparer la performance des modèles que nous avons développés et de plusieurs modèles existants sur des climats variés.

La seconde généralisation proposée permet de prendre en compte la non-stationnarité des propriétés statistiques de la pluie. Dans un contexte de changement climatique, le postulat de stationnarité dans les modèles stochastiques de pluie doit être remis en question. Nous considérons alors un processus de Poisson non-homogène à la base du modèle de Neyman-Scott. La théorie de base est reconsidérée et les moments statistiques sont calculés jusqu'au troisième ordre dans le cas général ainsi qu'avec une fonction simple pour le taux d'arrivée des averses. Une méthode d'estimation des paramètres est ensuite proposée et discutée, les moments observés étant obtenus grâce aux statistiques issues d'une fenêtre mobile et d'une méthode de régression.

## Summary

Many stochastic point processes for rainfall have been developed during the last twenty years. Point rainfall models based upon Poisson-cluster processes were deeply examined. In this type of models, storms arrive according to a Poisson process and a cluster of rainfall cells is attached to each storm, the intensity and the duration of these cells being themselves randomly distributed. These models have been proved to be particularly efficient in simulating hourly rainfall, due to the presence of clusters of rainfall cells in the rainfall data. However, weaknesses remain in the reproduction of extreme events and the reproduction of dry periods. The thesis proposes new developments of this Poisson-cluster models for rainfall for the Neyman-Scott Rectangular Pulses Model (NSRPM). More specifically, we deal with the dependence between the cell intensity and the cell duration. In a second part, we study the possibility of introducing a nonstationarity in the storm arrivals.

The independence relation between cell intensity and duration can explain the limitations of the Neyman-Scott model. Although independence between cell intensity and duration turned out to be a nonrealistic assumption, only a few models link these variables. Therefore, a Neyman-Scott cluster process considering dependence between cell depth and duration is developed. We introduce this link with a cubic copula. Thanks to this flexibility, we are able to introduce a global concept of dependence between cell depth and duration. We derive the aggregated moments (first, second and third order moments) from the new model for several families of polynomial copulas and perform an application on datasets with various origins (Canada, USA, Swiss, France and Belgium). These rainfall series enable us to confront the proposed model with the existing ones on various climate types.

The second development proposed to take into account the non-stationarity in the statistical properties of rainfall. In a context of climate change, the postulate of stationarity in stochastic models for rainfall must be discussed. We consider a non-homogeneous Poisson-process within the Neyman-Scott model. The basic theory is revised and the moments are derived up to the third order in the general case and also with a simple function on storm arrivals. A calibration method is proposed and discussed. Empirical moments are computed thanks to the moving window statistics and a regression method.

# Table des matières

$\mathbf{R}_{0}$	emerciements	iii
R	ésumé	v
Sι	ummary	vii
Ta	able des matières	x
Li	iste des tableaux	xiv
Li	iste des figures	xvi
N	omenclature	XX
1	Introduction	1
2	Théorie de base des processus ponctuels  2.1 Le processus de comptage  2.2 Le processus de Poisson  2.2.1 L'intensité  2.2.2 La distribution des intervalles  2.2.3 Le processus de comptage  2.3 Espérance, variance et covariance  2.4 Le processus de cluster	. 8 . 8 . 9 . 9
3	Le modèle à cellules rectangulaires de Neyman-Scott  3.1 Modèles ponctuels de Poisson	. 17
4	La modélisation de la dépendance avec les copules  4.1 Généralités sur les copules  4.1.1 Définitions  4.1.2 Propriétés  4.2 Mesures de dépendance  4.2.1 Le rho de Spearman  4.2.2 Le tau de Kendall	. 23 . 24 . 25

	4.3	Familles de copules	27
		4.3.2 Copules archimédiennes	28 29
5	Infl	uence de la dépendance entre l'intensité et la durée des cellules	35
	5.1	La discrétisation du temps	35
	5.2	La séparation des averses	36
	5.3	L'effet cluster sur le lien entre l'intensité et la durée des averses	36
		5.3.1 Influence du paramètre $\mu_D$ sur $\tau_\lambda$	37
		5.3.2 Influence des différents paramètres sur $\tau_{\lambda}$ lorsque la durée et l'intensité des cellules sont $indépendantes$	39
		5.3.3 Influence des différents paramètres sur $\tau_{\lambda}$ lorsque la durée et l'intensité des	
		r	41
	5.4	Conclusion	43
6	Cal	cul des moments du NSRPM incluant un lien intensité/durée des cellules	45
	6.1		46
	6.2		48
	6.3	Probabilité de période sans pluie	49
7	Mét	hode d'estimation des paramètres du modèle	51
	7.1	Méthode basée sur une fonction de vraisemblance	51
	7.2		52
		7.2.1 Méthode généralisée des moments	52
		7.2.2 Méthode d'estimation des paramètres pour le cas particulier de la version de	
			54
	7.3	Méthode d'estimation adoptée	58
8	App	olication du NSRPM incluant un lien entre l'intensité et la durée des cellules	61
	8.1	Description des données disponibles	61
		8.1.1 Suisse	65
		8.1.2 France et Belgique	66
			67
			67
	8.2		70
	8.3		71
	8.4		86
	8.5		89
	8.6	Conclusion	99

9	Un	modèle de Neyman-Scott non-stationnaire	101
	9.1	Introduction	. 101
	9.2	Le processus de Poisson non-stationnaire avec amas de cellules	. 103
		9.2.1 Moment du premier ordre	. 104
		9.2.2 Variance et Autocovariance	. 104
	9.3	Un choix de $\varphi(t)$	. 106
	9.4	Calcul des moments empiriques	. 108
	9.5	Méthode d'estimation des paramètres du NSRPM-NS	. 115
	9.6	Simulation	. 117
	9.7	Application	. 117
	9.8	Conclusion	. 120
10		nclusion	123
	10.1	Travail accompli	. 123
		10.1.1 NSRPM incluant un lien entre l'intensité et la durée des cellules pluvieuses	. 123
		10.1.2 NSRPM non-stationnaire	. 125
	10.2	Discussion et pistes des travaux à accomplir	. 125
		10.2.1 Estimation des paramètres	. 125
		10.2.2 Validation du modèle	. 126
		10.2.3 Vers un modèle parfait?	. 126
Bi	bliog	graphie	127
A	Cop	oules cubiques	139
	_	Conditions nécessaires	. 139
		Minimum du tau de Kendall pour une copule avec sections cubiques en u et v $$	
В	Moi	ments avec une copule cubique	141
	B.1	La fonction d'autocovariance avec la durée et l'intensité des cellules dépendantes .	. 141
	B.2	Calcul du moment du troisième ordre	. 142
	B.3	Moment du troisième ordre avec la copule FGM	. 145
	B.4	Formules générales des moments agrégés avec une copule cubique	. 145
$\mathbf{C}$	App	olication du NSRPM avec un lien intensité/durée des cellules	147
		Fonction objectif	
	C.2	Dépendance	. 164
	C.3	Extrêmes	. 170
D		ments agrégés du processus non-stationnaire	187
		Variance	
	D.2	Covariance	. 189

# Liste des tableaux

1.1	Résumé de la structure des pluies	4
4.1 4.2	Familles de copules cubiques à un paramètre	32 32
8.1 8.2	Description des séries de pluie	62 63
8.3	Mois et stations où la méthode alternative d'estimation des paramètres a été appliquée	
8.4	Estimations des paramètres du NSRPM $\hat{\phi}$ à Zürich pour le mois de janvier	
8.5	Estimations des paramètres du NSRPM $\hat{\phi}$ à Zürich pour le mois d'août	
8.6	Estimations des paramètres du NSRPM $\hat{\phi}$ à Uccle pour le mois de janvier	
8.7	Estimations des paramètres du NSRPM $\hat{\phi}$ à Uccle pour le mois d'août	
8.8	Estimations des paramètres du NSRPM $\hat{\phi}$ à Val d'Or pour le mois de janvier	
8.9	Estimations des paramètres du NSRPM $\hat{\phi}$ à Val d'Or pour le mois d'août	
8.10	Tableau récapitulatif des meilleurs modèles	85
9.1	Biais relatifs moyens et intervalles de confiance empirique	15
9.2	Choix des poids dans la procédure d'estimation des paramètres du NSRPM-NS 1	
9.3	Estimations des paramètres du NSRPM non-stationnaire $\hat{\phi}$	.20
C.1	Estimations des paramètres du NSRPM $\hat{\phi}$ à Brest pour le mois de janvier	.49
C.2	Estimations des paramètres du NSRPM $\hat{\phi}$ à Brest pour le mois d'août	
C.3	Estimations des paramètres du NSRPM $\hat{\phi}$ à Lyon pour le mois de janvier	
C.4	Estimations des paramètres du NSRPM $\hat{\phi}$ à Lyon pour le mois d'août	.51
C.5	Estimations des paramètres du NSRPM $\hat{\phi}$ à Marseille pour le mois de janvier 1	.53
C.6	Estimations des paramètres du NSRPM $\hat{\phi}$ à Marseille pour le mois d'août	.53
C.7	Estimations des paramètres du NSRPM $\hat{\phi}$ à Trudeau pour le mois de janvier 1	.55
C.8	Estimations des paramètres du NSRPM $\hat{\phi}$ à Trudeau pour le mois d'août 1	.55
C.9	Estimations des paramètres du NSRPM $\hat{\phi}$ à Miami pour le mois de janvier 1	.57
C.10	Estimations des paramètres du NSRPM $\hat{\phi}$ à Miami pour le mois d'août	.57
C.11	Estimations des paramètres du NSRPM $\hat{\phi}$ à Seattle pour le mois de janvier 1	.59
C.12	Estimations des paramètres du NSRPM $\hat{\phi}$ à Seattle pour le mois d'août	.59
C.13	Estimations des paramètres du NSRPM $\hat{\phi}$ à Los Angeles pour le mois de janvier 1	.61
C.14	Estimations des paramètres du NSRPM $\hat{\phi}$ à Los Angeles pour le mois d'août 1	.61
C.15	Estimations des paramètres du NSRPM $\hat{\phi}$ à Minneapolis pour le mois de janvier 1	.63
C.16	Estimations des paramètres du NSRPM $\hat{\phi}$ à Minneapolis pour le mois d'août 1	.63

# Liste des figures

2.1	Construction d'un processus de <i>cluster</i>	11
3.1 3.2	Construction des modèles de Neyman-Scott et Bartlett-Lewis	14 20
4.1 4.2	Deux paires d'observations concordantes (à gauche) et discordantes (à droite) Graphe de l'espace des paramètres dans $S$ pour chaque copule cubique	27 33
5.1 5.2 5.3 5.4 5.5 5.6 5.7	$\begin{array}{l} \tau_{\lambda} \; (\text{lien dur\'e/intensit\'e des averses}) \; \text{en fonction de} \; \mu_{D} \; . \qquad . \\ \tau_{\lambda} \; (\text{lien dur\'e/intensit\'e des averses}) \; \text{en fonction de} \; D \; . \\ \tau_{\lambda} \; \text{versus} \; (\mu_{D}, \beta), \; (\mu_{D}, \eta), \; (\beta, \eta), \; (\alpha, \eta), \; (\alpha, \beta) \; \text{et} \; (\mu_{D}, \alpha) \; . \\ \tau_{\lambda} \; \text{versus} \; \mu_{D} \; \text{quand intensit\'e et dur\'ee des cellules sont li\'ees} \; . \\ \tau_{\lambda} \; \text{versus} \; \mu_{D} \; \text{quand intensit\'e} \; \text{et dur\'ee} \; \text{des cellules sont co-monotones} \; . \\ \tau_{\lambda} \; \text{versus} \; \alpha, \beta \; \text{et} \; \eta \; \text{avec un lien intensit\'e/dur\'ee} \; \text{des cellules} \; . \\ Deux \; \text{exemples d'averses dont la dur\'ee} \; \text{et l'intensit\'e} \; \text{des cellules sont co-monotones} \; . \\ (\tau_{D} = 1) \; . \; . \; . \; . \; . \; . \\ \end{array}$	38 38 40 41 42 42
7.1	Diagramme résumant la procédure d'estimation en neuf étapes	57
8.1 8.2 8.3 8.4 8.5 8.6 8.7	Carte des climats mondiaux	63 64 65 66 68 69
8.8	Trudeau	72 73
8.9	Évolution mensuelle de l'autocorrélation horaire et journalière empirique à la station de Brest	74
8.10	Évolution de la fonction objectif et des estimations des paramètres du NSRPM en fonction de l'autocovariance journalière à Brest	74
8.11	Fonction objectif pour les différents modèles testés à la station de Zürich	79

8.12	Fonction objectif pour les différents modèles testés à la station d'Uccle 81
8.13	Fonction objectif pour les différents modèles testés à la station de Val d'Or 83
8.14	Tau de Kendall entre l'intensité et la durée des cellules avec la copule Sym2 en Europe 87
8.15	Tau de Kendall entre l'intensité et la durée des cellules avec la copule Sym2 en
	Amérique du Nord
8.16	Maxima annuels journaliers à Zürich
8.17	Maxima annuels horaires à Zürich
8.18	Maxima annuels journaliers à Uccle
8.19	Maxima annuels horaires à Uccle
8.20	Maxima annuels journaliers à Trudeau $\ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ 95$
8.21	Maxima annuels horaires à Trudeau
8.22	Maxima annuels journaliers à Lyon avec le "méga-modèle"
8.23	Maxima annuels horaires à Lyon avec le "méga-modèle"
9.1	Approximations linéaires des moments et moments théoriques du NSRPM non-
	stationnaire
9.2	Biais relatif moyen et intervalle de confiance empirique à $80\%$ ( $10^{\text{ème}}$ et $90^{\text{ème}}$ cen-
	tiles) résultant de la régression robuste pour la moyenne
9.3	Biais relatif moyen et intervalle de confiance empirique à 80% (10ème et 90ème cen-
	tiles) résultant de la régression robuste pour la variance
9.4	Biais relatif moyen et intervalle de confiance empirique à 80% (10ème et 90ème cen-
	tiles) résultant de la régression robuste pour l'autocovariance de décalage un 113
9.5	Biais relatif moyen et intervalle de confiance empirique à 80% (10ème et 90ème cen-
	tiles) résultant de la régression robuste pour le moment du troisième ordre (centré). 114
9.6	Exemples de fonctions $\varphi(t)$ , $\Lambda(t)$ et $\Lambda^{-1}(s)$ pour la simulation avec $\lambda=0.01$ et
	$\epsilon = 0.01.$
9.7	Moments empiriques et régressions linéaires correspondantes à une échelle horaire
	pour le NSRPM non-stationnaire
9.8	Moments empiriques et régressions linéaires correspondantes à une échelle jour-
	nalière pour le NSRPM non-stationnaire
B.1	Illustration du calcul du moment du troisième ordre du NSRPM
C.1	Fonction objectif pour les différents modèles testés à la station de Brest
C.2	Fonction objectif pour les différents modèles testés à la station de Lyon
C.3	Fonction objectif pour les différents modèles testés à la station de Marseille 152
C.4	Fonction objectif pour les différents modèles testés à la station de Trudeau 154
C.5	Fonction objectif pour les différents modèles testés à la station de Miami 156
C.6	Fonction objectif pour les différents modèles testés à la station de Seattle
C.7	Fonction objectif pour les différents modèles testés à la station de Los Angeles 160
C.8	Fonction objectif pour les différents modèles testés à la station de Minneapolis 162

C.9	Tau de Kendall entre l'intensité et la durée des cellules avec la copule Fcub en Europe $164$
C.10	Tau de Kendall entre l'intensité et la durée des cellules avec la copule Fcub en
	Amérique du Nord
C.11	Tau de Kendall entre l'intensité et la durée des cellules avec la copule Asym3 en
	Europe
C.12	Tau de Kendall entre l'intensité et la durée des cellules avec la copule Asym3 en
	Amérique du Nord
C.13	Tau de Kendall entre l'intensité et la durée des cellules avec la copule FGM en Europe $168$
C.14	Tau de Kendall entre l'intensité et la durée des cellules avec la copule FGM en
	Amérique du Nord
C.15	Maxima annuels journaliers à Brest
C.16	Maxima annuels horaires à Brest
C.17	Maxima annuels journaliers à Lyon
C.18	Maxima annuels horaires à Lyon
C.19	Maxima annuels journaliers à Marseille
C.20	Maxima annuels horaires à Marseille
C.21	Maxima annuels journaliers à Val d'Or
C.22	Maxima annuels horaires à Val d'Or
C.23	Maxima annuels journaliers à Miami
C.24	Maxima annuels horaires à Miami
C.25	Maxima annuels journaliers à Seattle
C.26	Maxima annuels horaires à Seattle
C.27	Maxima annuels journaliers à Los Angeles
C.28	Maxima annuels horaires à Los Angeles
C.29	Maxima annuels journaliers à Minneapolis
C.30	Maxima annuels horaires à Minneapolis

### Nomenclature

### Processus ponctuel

Ξ espace du processus ponctuel

 $N(\cdot)$  processus de comptage

 $\mathbb{I}_A(\cdot)$  fonction indicatrice de l'ensemble A

 $\mathcal{H}_t$  histoire du processus au temps t

 $c(\cdot)$  fonction de densité de covariance du processus ponctuel

### Variables aléatoires du NSRPM

D nombre de cellules par *cluster* 

X intensité de la cellule

L durée de la cellule

Y(t) intensité de pluie au temps t

 $Y_i^h \quad \ \,$  pluie totale cumulée dans le ième intervalle de longueur h

### Dépendance

 $C(\cdot, \cdot)$  distribution copule

 $\Phi(\cdot)$  fonction génératrice des copules archimédiennes

 $A_i, B_i, i = 1, 2$  paramètres des copules cubiques

 $au_D$  tau de Kendall entre l'intensité et la durée des cellules  $au_\lambda$  tau de Kendall entre l'intensité et la durée des averses

### **Paramètres**

λ taux du processus de Poisson stationnaire gérant les origines d'averses  $\beta$ paramètre de la distribution exponentielle liée à la position des cellules espérance du nombre de cellules  $\mu_D$ paramètre de la distribution exponentielle liée à la durée des cellules  $\eta$ espérance de l'intensité des cellules  $\mu_X$ paramètre d'échelle lié à l'intensité des cellules  $\alpha$ paramètre de forme lié à l'intensité des cellules  $\gamma$ paramètre de la copule cubique  $f, c \text{ et } \zeta$ paramètres intensité/durée des cellules dans les modèles DD1 et DD2 vecteur des paramètres du NSRPM φ

### Moments et estimation des paramètres

distribution bivariée de X lorsque L > lkdécalage de la fonction d'autocovariance moment d'ordre r de X lorsque  $L \geq l$  $\mu_{X^r,L>l}$ ordre du moment mhniveau d'agrégation temporelle  $c_Y(\tau)$ covariance de décalage  $\tau$  du processus d'intensité pluvieuse moment du troisième ordre (centré) d'un intervalle de longueur h $\xi_h$ PD(h)probabilité d'avoir un intervalle de longueur h sans pluie 0 fonction objective dans la méthode des moments nombre de moments considérés dans la fonction objectif ppoids des moments dans la fonction objectif  $Z_i, i = 1..4, V, W$ fonctions permettant de décomposer les expressions de la variance et de la covariance dans la version de base du NSRPM échelle de fluctuation du processus  $\chi$ 

### **NSRPM** non-stationnaire

taux du processus de Poisson dépendant du temps  $\varphi(\cdot)$ paramètre de la distribution exponentielle liée à la durée des averses dans le ρ modèle de Bartlett-Lewis non-stationnaire  $\Omega$ longueur de la fenêtre mobile taux de croissance du taux du processus de Poisson coefficients de l'approximation linéaire de l'espérance agrégée  $V_0^h, V_1^h$ coefficients de l'approximation linéaire de la variance agrégée  $C_0^h, C_1^h$ coefficients de l'approximation linéaire de l'autocovariance agrégée coefficients de l'approximation linéaire du moment du troisième ordre agrégé  $\Lambda(\cdot)$ fonction de cumul du taux du processus de Poisson

# Chapitre 1

### Introduction

La répartition quantitative de l'eau à l'échelle du globe est très hétérogène. De même, la distribution spatiale de l'eau douce disponible par habitant est inégale selon les pays. Cette constatation constitue la source de conflits et de tensions politiques qui ne vont pas cesser de s'accroître dans les prochaines décennies (conflit de l'eau au Moyen-Orient, voir Bisson & Mutin, 2000). Sans parler de la consommation domestique de l'eau douce, la gestion de l'eau est devenue une priorité dans de nombreux domaines (production hydro-électrique, agriculture, loisirs, etc.). De nombreux exemples montrent les conséquences désastreuses d'aménagements hydrauliques mal planifiés (crise écologique de la mer d'Aral, stérilisation des sols, sédimentation et inondations suite à la construction du complexe hydraulique de Sanmenxia en Chine, ou encore l'assèchement des marais de Mésopotamie en Irak, etc.). Pour de plus amples informations concernant les enjeux de l'eau, le lecteur pourra consulter un ouvrage récent sur ce sujet (Anctil, 2008) ou le dossier sur l'eau douce du Centre National de la Recherche Scientifique (CNRS) disponible à l'adresse http://www.cnrs.fr/cw/dossiers/doseau/accueil.html¹.

Une gestion efficace de cette ressource naturelle est de première importance, notamment afin de prévenir les catastrophes naturelles (inondations, sécheresses). Le terme "inondation" désigne la submersion de zones normalement sèches par l'eau, des matériaux en suspension ou du matériau de charriage; cette submersion étant partielle ou totale et limitée dans le temps. Au Canada, les inondations sont dues à une ou plusieurs des causes suivantes<sup>2</sup>:

- fonte des neiges :

<sup>&</sup>lt;sup>1</sup>dernière consultation le 7 août 2008

<sup>&</sup>lt;sup>2</sup>site d'Environnement Canada http://www.ec.gc.ca/Water/fr/manage/floodgen/f\_cause.htm, dernière consultation le 7 août 2008

Introduction

- pluie;
- embâcles et autres obstacles;
- jökulhlaups ou inondations provoquées par l'éclatement d'un glacier;
- tempêtes sur les littoraux (tsunamis, cyclones, ouragans);
- modifications du drainage dues à l'urbanisation;
- défaillance d'un barrage (ou autre ouvrage hydraulique).

L'inondation constitue un risque majeur aux conséquences humaines et matérielles extrêmement préjudiciables. Au Canada, les inondations ont causé, de façon directe ou indirecte, la mort d'au moins 198 personnes et entraîné des dépenses de plusieurs milliards de dollars au cours du XXième siècle (voir Brooks et al., 2001). Il est donc indispensable de posséder des outils de gestion efficaces afin de prévenir ce type de catastrophe lors du dimensionnement des ouvrages hydrauliques (barrages, digues, évacuateurs de crue, égouts, etc.). Au Québec, la gestion des ressources en eau est de plus particulièrement importante en vue de la production hydroélectrique, qui représente plus de 95% de l'électricité produite.

La pluie est un des principaux acteurs du cycle de l'eau. À partir d'un modèle de pluie qui représente adéquatement ses propriétés statistiques (moyenne de l'intensité pluvieuse, variance, autocovariance, valeurs extrêmes) à plusieurs échelles de temps (mois, jour, heure, jusqu'à un pas de temps de cinq minutes, dépendamment des applications), il sera possible d'utiliser de longues séries synthétiques de pluie comme intrants dans des modèles hydrologiques (pluie-débit, infiltration, etc.). Ce sont alors les sorties de ces modèles qui permettront aux constructeurs d'ouvrages hydrauliques ou aux gestionnaires de disposer d'outils d'aide à la décision. Par exemple, la prédiction des crues peut être réalisée en couplant un modèle stochastique de pluie et un modèle hydrologique déterministe (Overney, 1997; Sinniger, 1997; Favre, 2001; Overney et al., 2005). Certains modèles hydrologiques à base quasi-physique permettent aussi de simuler l'effet de modifications des variables hydrométéorologiques (comme la pluie) dans le temps et l'espace sur les crues (voir par exemple Sinniger, 1997).

Selon Cox & Isham (1994), les modèles de pluies (dans l'espace et le temps) peuvent être regroupés en trois grandes catégories :

- Les modèles statistiques empiriques dans lesquels les relations entre les variables explicatives sont représentées via des équations empiriques (Stein, 1986);
- Les modèles de météorologie dynamiques dans lesquels les processus physiques sont décrits en utilisant les théories de la dynamique et de la thermodynamique des fluides. De larges systèmes non-linéaires d'équations différentielles doivent alors être résolus numériquement (Mason, 1986). De tels modèles sont couramment appliqués dans le domaine de la

- prévision météorologique et dans les modèles climatiques globaux (Global Climate Models, GCMs);
- Les modèles stochastiques dans lesquels un nombre modeste de paramètres, qui sont reliés à des caractéristiques physiques, sont utilisés pour représenter la variabilité temporelle ou spatiale de la pluie.

Ces trois types de modèles ont tous un rôle important à jouer et le choix du modèle dépend uniquement de l'objectif de l'analyse. Dans ces travaux, nous nous intéressons exclusivement au troisième type de modèle. Un modèle stochastique de pluie a pour but objectif de simuler des séries chronologiques reproduisant au mieux les observations en un point donné (ou une petite région) de la surface du globe. Les paramètres d'un tel modèle étant calibrés, il peut servir à extrapoler des séries chronologiques existantes pour simuler le comportement à long terme d'un système hydrologique, étudier la fréquence de ses défaillances ou d'apparition d'un phénomène particulier (débordement d'un système d'assainissement urbain par exemple) et étudier les caractéristiques des pluies qui les provoquent.

Il existe principalement trois types de modèles stochastiques de pluie. Le premier ensemble de modèles se base sur des chaînes de Markov (voir par exemple Stern & Coe, 1984). L'intensité de la pluie journalière ou horaire est alors distribuée selon une loi simple (gamma par exemple) conditionné par un état qui peut être "sec" ou "humide", ou encore "transitionnel" pour un choix d'espace à trois états. Ce type de modèle donne des résultats convaincants à une échelle journalière. Il faut cependant indiquer que ces modèles n'ont pas la prétention de nous renseigner sur la physique des processus réels. Les modèles multifractals constituent une deuxième façon d'aborder le problème de la modélisation stochastique de la pluie. Ils traduisent le plus simplement possible des propriétés d'invariance de certains paramètres (invariance spatiale, temporelle ou spatio-temporelle) à partir d'un nombre très limité de paramètres. En particulier, dans les modèles de cascades multifractales de générateur algébrique, c'est le coefficient de décroissance algébrique qui est un invariant d'échelle (voir Mandelbrot, 1975; Schertzer et al., 1997; Schertzer & Lovejoy, 2004). Le dernier grand type de modèle utilise des processus ponctuels avec amas de cellules (cluster). Ces modèles sont particulièrement intéressants car ils s'inspirent des caractéristiques physiques des champs de précipitations, principalement de leur structure hiérarchique (les averses étant représentées par des amas de cellules pluvieuses).

Les aires de pluie peuvent être classifiées selon leur extension spatiale (Northrop, 1996). Les zones plus grandes que  $10^4$  km² sont définies comme "synoptiques" et sont généralement associées à des pluies cycloniques. Les zones synoptiques ont une durée de vie pouvant aller de un à plusieurs jours. Les zones de précipitations subsynoptiques les plus grandes pouvant être contenues dans une zone synoptique sont appelées LMSA (*Large MesoScale Area*). Dans les pluies cycloniques, les LMSA apparaissent souvent comme des bandes allongées et ont ainsi été désignées quelquefois par le terme de "bande". Les LMSA, dont l'extension varie de  $10^3$  à  $10^4$  km², se construisent et se

dissipent à l'intérieur d'une zone synoptique. Leur temps de vie est de l'ordre de quelques heures, et leur nombre à l'intérieur d'une zone synoptique varie de un à six. L'intensité des précipitations à l'intérieur d'un LMSA est toujours plus élevée que celle de la région qui l'entoure.

Chaque LMSA contient des régions identifiables de pluie convective de type cumulus, communément appelées cellules "convectives". Les cellules convectives, dont la taille varie de 10 à 30 km² selon le type de pluie, se construisent et se dissipent à l'intérieur d'un LMSA et apparaissent généralement en cluster. La durée de vie d'une cellule est de l'ordre de quelques minutes à environ une demi-heure. Ces cellules suivent le mouvement relatif du LMSA et l'intensité de pluie à l'intérieur d'une cellule est toujours plus élevée que celle de la région qui la domine. Certaines observations montrent de plus la possibilité que les amas de cellules ne soient pas la seule sous-région pouvant contenir des cellules. Ces observations sur certaines pluies indiquent l'occurrence de SMSA (Small MesoScale Area), variant en taille de 10² à 10³ km², à l'intérieur d'un LMSA. Chaque SMSA contient à son tour des cellules pluvieuses. Les SMSA se construisent et se dissipent à l'intérieur d'un LMSA avec une durée de vie de quelques heures.

Cette description met en évidence une structure hiérarchique dans la formation des précipitations. Le Tableau 1.1 résume les différents éléments que l'on peut retrouver.

TAB. 1.1 – Résumé de la structure des pluies : aires de pluie avec leur extension et leur durée classique

Élément de pluie	Extension spatiale [km²]	Durée classique
Cellule pluvieuse	10 - 30	jusqu'à 30 minutes
SMSA	$10^2 - 10^3$	quelques heures
LMSA	$10^3 - 10^4$	plusieurs heures
Zone synoptique	$> 10^4$	jusqu'à plusieurs jours

Les processus stochastiques ponctuels s'avèrent être très utiles pour reproduire la variabilité de la pluie à plusieurs niveaux d'agrégation (horaire, journalier). Un modèle stochastique de pluie en un point de l'espace décrit le comportement temporel des précipitations liquides. Deux modèles à pulsations rectangulaires bien connus ont été intensivement utilisés à cette fin : les modèles de Bartlett-Lewis et de Neyman-Scott. De nombreuses applications ont déjà illustré l'efficacité de tels modèles pour l'analyse de données collectées à un pas de temps court, c'est-à-dire horaire (voir par exemple Rodríguez-Iturbe et al., 1987a; Cowpertwait et al., 1996; Onof et al., 2000). Dans cette thèse, nous nous concentrons principalement sur le modèle de Neyman-Scott (Neyman-Scott Rectangular Pulse Model ou NSRPM). Pour obtenir une revue détaillée des développements avec ce type de modèles de pluie et une discussion des performances des différentes approches, le lecteur se référera à l'article de Onof et al. (2000).

Afin de calculer facilement les moments théoriques de l'intensité de pluie (moyenne, variance,

autocovariance), la plupart des modèles supposent une relation d'indépendance entre l'intensité et la durée des cellules de pluie. Cette hypothèse est questionnable pour plusieurs raisons. D'une part, le modèle serait plus souple sans cette condition et donc pourrait s'adapter plus facilement à la représentation de la structure temporelle de la pluie. D'autre part, la dépendance entre la durée des averses et leur intensité est souvent significativement négative (De Michele & Salvadori, 2003). Kim & Kavvas (2006) remettent cette hypothèse d'indépendance en question et citent quelques développements (voir Córdova & Rodríguez-Iturbe, 1985; Singh & Singh, 1991; Bacchi et al., 1994; Kurothe et al., 1997; Goel et al., 2000) où les cellules de pluie, comportant une corrélation dans leurs caractéristiques, arrivent selon un processus de Poisson simple (c'est-à-dire sans amas). Les auteurs citent quelques distributions exponentielles bivariées [Gumbel type I (Gumbel, 1960), Downton (Downton, 1970)] qui sont utilisées pour modéliser des corrélations négatives et positives entre la durée et l'intensité pluvieuse. Cependant, ces représentations ne considèrent pas de cluster et apparaissent limitées. Sans la superposition de cellules, le modèle de Poisson est incapable de représenter plus d'une échelle de temps (Rodríguez-Iturbe et al., 1987b). Kim & Kavvas (2006) développent un NSRPM qui considère des corrélations positives et négatives avec une distribution bivariée de Gumbel type-II. Kakou (1997) inclut une dépendance dans la moyenne de l'intensité des cellules conditionnellement à leur durée. Les moments du premier et second ordres sont calculés et les proportions de périodes sèches semblent être adéquatement reproduites à toutes les échelles temporelles (voir Onof et al., 2000). Dans notre cas, une copule cubique fait le lien entre l'intensité et la durée des cellules. Ce modèle de dépendance flexible inclut comme cas particulier la distribution de Gumbel type-II. Cette dernière est en fait composée de lois marginales exponentielles et d'une structure de dépendance modélisée par la famille de copules de Farlie-Gumbel-Morgenstern (ou FGM) discutée par Morgenstern (1956), Gumbel (1958) et Farlie (1960). L'utilisation des copules cubiques est d'ailleurs novatrice considérant les rares applications avec cette classe de copules. Nous développons donc un NSRPM plus général et nous calculons les moments du premier, deuxième et troisième ordres ce qui nous permet de calibrer le modèle avec la méthode généralisée des moments.

Un deuxième apport de cette thèse aborde la question de la non-stationnarité de la pluie dans le temps. En nous questionnant sur la viabilité de l'hypothèse de stationnarité dans les modèles de Poisson avec *cluster* dans un contexte de changement climatique, nous proposons de nouveaux développements permettant de prendre en compte une augmentation de la fréquence des averses dans le futur.

La thèse est organisée de la façon suivante. Le chapitre 2 présente succinctement la théorie des processus ponctuels, nécessaire à la compréhension du modèle de Neyman-Scott. Le NSRPM est ensuite présenté en détails dans le chapitre 3, dans sa version de base. Le chapitre 4 expose les éléments essentiels de la théorie des copules et présente plus particulièrement la classe des copules cubiques. Dans le chapitre 5, nous montrons pourquoi il n'est pas possible de mesurer la dépendance entre la durée et l'intensité des cellules à partir des données de pluie. Le développement du modèle

Introduction

6

et le calcul des moments du NRSPM avec dépendance intra-cellules est ensuite présenté au chapitre 6. La problématique de l'estimation des paramètres du modèle est abordée au chapitre 7. Ce modèle est ensuite appliqué sur de nombreuses séries de pluie au Canada, aux États-Unis d'Amérique, en Suisse, en France et en Belgique et une analyse des performances des différents modèles considérés est synthétisée au chapitre 8. Le chapitre 9 présente de nouveaux développements du NSRPM afin de prendre en compte une non-stationnarité de la pluie dans le temps. Le chapitre 10 conclut et donne quelques pistes de travaux à réaliser.

# Chapitre 2

## Théorie de base des processus ponctuels

Dans ce chapitre nous exposons la théorie de base des processus ponctuels, en nous concentrant uniquement sur les notions nécessaires pour comprendre le modèle de Neyman-Scott. Les processus ponctuels sont des processus stochastiques particuliers dont les réalisations consistent en des événements ponctuels dans le temps ou dans l'espace. Ils sont utilisés dans de nombreux domaines d'applications. Les définitions, formules et notations utilisées dans ce chapitre sont tirées de Cox & Isham (1980).

### 2.1 Le processus de comptage

Un processus ponctuel est un modèle de points distribués aléatoirement dans un espace  $\Xi$ . Les points peuvent représenter l'occurrence d'un événement dans le temps, la position d'objets ou des éléments d'un espace de fonctions. En pratique, il est souvent intéressant de compter le nombre de points dans des sous-espaces de  $\Xi$  sur lequel le processus ponctuel est défini. Un **processus de comptage** est introduit dans ce but et peut être associé à tous les processus ponctuels. Soit une réalisation, notée T, d'un processus ponctuel aléatoire sur  $\Xi$ . T est un ensemble dénombrable de points sur  $\Xi$ . Cela signifie que T peut être énuméré de la façon suivante :  $T = \{t_1, t_2, \ldots\}$ , où chaque  $t_i$  dénote la coordonnée d'un point dans  $\Xi$ . Soit A un sous-espace de  $\Xi$ , et notons N(A,T) le nombre de points dans T qui appartiennent à A. Formellement :

$$N(A,T) = \sum_{i} \mathbb{I}_{A}(t_{i}), \qquad (2.1)$$

où chaque  $t_i$  est la coordonnée d'un point dans T et  $\mathbb{I}_A(t)$  dénote la fonction indicatrice de A définie mathématiquement par :

$$\mathbb{I}_A(t) = \begin{cases} 1 & \text{si} \quad t \in A \\ 0 & \text{si} \quad t \notin A. \end{cases}$$
(2.2)

Quand N(A,T) est considéré comme une fonction de A et T, il définit un processus aléatoire entier positif sur  $\Xi$ , appelé processus de comptage. Afin de simplifier la notation, nous supprimons la dépendance de N(A,T) sur T et désignons le nombre de points situés dans A simplement par N(A).

### 2.2 Le processus de Poisson

Le processus ponctuel le plus connu et le plus utilisé est sans aucun doute le processus de Poisson. De plus, il est l'élément de base du modèle de Neyman-Scott. La présente section s'attache à décrire les principales propriétés du processus de Poisson simple, c'est-à-dire où le taux  $\lambda$  est constant. Dans ce cas, le processus de Poisson est dit **stationnaire** ou **homogène**. Nous supposons ici que  $\Xi = \mathbb{R}$ . Cox & Isham (1980) définissent le processus de Poisson de trois manières équivalentes décrites dans les sous-sections suivantes.

#### 2.2.1 L'intensité

L'expression des positions de tous les points dans l'intervalle  $]-\infty,t]$  est appelée l'histoire du processus au temps t et sera notée  $\mathcal{H}_t$ . Soit N(a,b) la variable aléatoire désignant le nombre de points dans l'intervalle ]a,b]. Le processus de Poisson de taux  $\lambda$  est défini complètement par les deux propriétés suivantes qui s'appliquent pour tout t lorsque  $\delta \to 0^+$ :

$$\Pr\{N(t, t + \delta) = 1 | \mathcal{H}_t\} = \lambda \delta + o(\delta), \tag{2.3}$$

$$\Pr\{N(t, t+\delta) > 1 | \mathcal{H}_t\} = o(\delta), \tag{2.4}$$

où  $o(\delta)$  désigne une fonction petit ordre de  $\delta$ , c'est-à-dire qui tend vers 0 plus rapidement que  $\delta$ . Une fonction  $f(\delta)$  est  $o(\delta)$  si  $\lim_{\delta \to 0} \frac{f(\delta)}{\delta} = 0$ . Par exemple, la fonction  $\delta^2$  est petit ordre de  $\delta$  tandis que la fonction  $\delta^{1/2}$  ne l'est pas. Nous déduisons directement des deux propriétés précédentes la relation suivante :

$$\Pr\{N(t, t+\delta) = 0 | \mathcal{H}_t\} = 1 - \lambda \delta + o(\delta). \tag{2.5}$$

Une propriété principale résultant des équations 2.3 à 2.5 est que les probabilités définies ne dépendent pas de l'historique  $\mathcal{H}_t$ . Cela signifie que la probabilité de trouver un point dans  $]t, t+\delta[$  ne dépend pas du nombre de points qui apparaissent juste avant t. Notons aussi que l'expression (2.4) exclut le cas d'apparition de plus d'un point en même temps. L'occurrence simultanée de plusieurs points est une généralisation possible du processus de Poisson.

#### 2.2.2 La distribution des intervalles

Considérons les points  $0 < T_1 < T_2 < \dots$  qui définissent un processus de Poisson homogène de taux  $\lambda$ . Alors les variables aléatoires  $X_i$ ,  $i=2\dots n$  qui décrivent l'intervalle  $T_i-T_{i-1}$  sont indépendantes et distribuées de façon exponentielle. Leur densité de probabilité s'écrit  $f_X(x)=\lambda \exp(-\lambda x)$  pour  $x\geq 0$ . Dans ce cas également, une généralisation possible du processus de Poisson est d'associer une distribution différente que l'exponentielle aux intervalles  $T_i-T_{i-1}$ . Ces processus sont appelés **processus de renouvellement**.

### 2.2.3 Le processus de comptage

Considérons le nombre d'événements  $N(a_i,b_i)$  du processus dont l'occurrence se situe dans l'intervalle  $]a_i,b_i]$  avec  $a_i < b_i \le a_{i+1}$ , alors le processus de Poisson est complètement défini par l'équation :

$$\Pr\{N(a_i,b_i) = n_i, i = 1, \dots, k\} = \prod_{i=1}^k \frac{\{\lambda(b_i - a_i)\}^{n_i}}{n_i!} \exp\{-\lambda(b_i - a_i)\}.$$

De cette relation peuvent se déduire plusieurs propriétés importantes des processus de Poisson:

- 1. Le nombre de points dans chaque intervalle fini  $[a_i, b_i]$  possède une distribution de Poisson.
- 2. Le nombre d'événements qui surviennent dans des intervalles de temps disjoints sont indépendants. Un processus stochastique est à accroissements indépendants lorsqu'il satisfait cette condition;

Soient deux nombres réels positifs t et h. Si le nombre N(t, t+h) d'événements qui se produisent pendant l'intervalle de temps ]t, t+h] ne dépend pas de la valeur de t mais seulement de la longueur de l'intervalle h, le processus stochastique N est dit **stationnaire**.

### 2.3 Espérance, variance et covariance

Dans cette section nous présentons quelques éléments de calcul des propriétés statistiques du processus de comptage. Considérons les deux premiers moments du processus de comptage dans des ensembles arbitraires A et B:

$$\mathbb{E}\left\{N(A)\right\}, \operatorname{Var}\left\{N(A)\right\} \text{ et } \operatorname{Cov}\left\{N(A), N(B)\right\}.$$

Pour des processus stationnaires à taux finis et fixes  $\lambda$ , nous déduisons directement que :

$$E\{N(A)\} = \lambda |A|,$$

où |A| désigne la mesure de Lebesgue de l'ensemble A. Considérons la covariance pour des processus de comptage de deux ensembles disjoints A et B, nous avons :

$$2\operatorname{Cov}\left\{N(A), N(B)\right\} = \operatorname{Var}\left\{N(A \cup B)\right\} - \operatorname{Var}\left\{N(A)\right\} - \operatorname{Var}\left\{N(B)\right\}. \tag{2.6}$$

Le problème revient alors à résoudre le calcul des variances. Pour le calcul de la variance, il est utile de réécrire N(t) comme la limite de la somme de comptages à travers une partition de [0,t] en de petits intervalles :

$$N(0,t) = \int_0^t \mathrm{d}N(z),$$

où  $\mathrm{d}N(z)$  désigne plus simplement  $\lim_{\delta\to 0}N(z,z+\delta)$ . D'après la relation (2.2),  $\mathrm{d}N(z)$  vaut 1 s'il existe un point en z et 0 sinon. En appliquant la relation (2.6) à la variance d'une somme de variables aléatoires, cela conduit à :

$$\operatorname{Var}\{N(0,t)\} = \int_0^t \operatorname{Var}\{\mathrm{d}N(z)\} + 2 \iint_{\substack{0 < z < t \\ 0 < u \le t-z}} \operatorname{Cov}\{\mathrm{d}N(z), \mathrm{d}N(z+u)\},$$

où l'intégrale est ici considérée comme la limite d'une somme. Nous en déduisons alors (voir Cox & Isham, 1980, Eq. 2.25-2.26) :

$$\operatorname{Var}\left\{N(z,z+\delta)\right\} = \operatorname{E}\left\{N^{2}(z,z+\delta)\right\} - \left[\operatorname{E}\left\{N(z,z+\delta)\right\}\right]^{2}$$
$$= \operatorname{Pr}\left\{N(z,z+\delta) = 1\right\} - \left[\operatorname{Pr}\left\{N(z,z+\delta) = 1\right\}\right]^{2} + o(\delta)$$
$$= \lambda\delta + o(\delta),$$

et pour u > 0:

$$\operatorname{Cov} \left\{ N(z, z + \delta_1), N(z + u, z + u + \delta_2) \right\} \\
= \operatorname{E} \left\{ N(z, z + \delta_1) N(z + u, z + u + \delta_2) \right\} - \operatorname{E} \left\{ N(z, z + \delta_1) \right\} \operatorname{E} \left\{ N(z + u, z + u + \delta_2) \right\} \\
= \operatorname{Pr} \left\{ N(z, z + \delta_1) = N(z + u, z + u + \delta_2) = 1 \right\} - \lambda^2 \delta_1 \delta_2 + o(\delta_1 \delta_2).$$

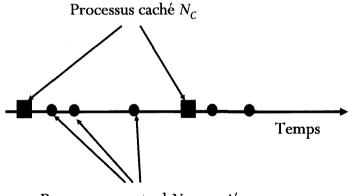
Pour un processus stationnaire, Cov  $\{N(z, z + \delta_1), N(z + u, z + u + \delta_2)\}$  dépend seulement de u lorsque  $\delta_1$  et  $\delta_2$  tendent vers 0 et est notée plus simplement c(u). La stationnarité implique que c(-u) = c(u). La fonction  $c(\cdot)$  est appelée la **densité de covariance**.

### 2.4 Le processus de cluster

Un processus de cluster est basé sur un processus ponctuel caché  $N_C$ . Dans notre modèle de pluie, ce processus ponctuel caché (donc non-observé) représente les instants de début des averses. Ces points constituent alors des origines auxquelles sont attachés des sous-ensembles de points. Ces clusters forment des familles dénombrables de processus ponctuels  $N_{(t)}$  dont la superposition est observé sur un même espace de réalisation (voir Figure 2.1). Dans notre cas, ces sous-ensembles correspondent aux cellules pluvieuses (le concept de cellule sera développé longuement par la suite). Ainsi, à chaque point primaire ou centre de cluster est associé le cluster de points secondaires (membres du cluster). Un processus de cluster se décrit d'une part par la connaissance du processus distribuant les centres des clusters et, d'autre part, par le mécanisme simulant les points à l'intérieur des clusters. Des résultats similaires à ceux précédemment présentés peuvent être obtenus pour les processus de cluster. Par exemple, pour l'espérance du processus de comptage, nous avons :

$$\mathrm{E}\left\{N(A)
ight\} = \int \mathrm{E}\left\{N_{(t)}(A)
ight\}\mathrm{E}\left\{\mathrm{d}N_C(t)
ight\},$$

où  $N_C$  désigne le comptage relatif au processus des centres de *cluster* et  $N_{(t)}(A)$  représente le nombre de points provenant d'un *cluster* ayant son centre en t.



Processus ponctuel  $N_{(t)}$  associé au premier centre de *cluster* 

FIG. 2.1 – Composition d'un processus de *cluster*. Le processus distribue d'abord des points "parents", puis un sous-ensemble de points est associé à chaque parent.

Les propriétés de comptage du second-ordre du processus de Neyman-Scott peuvent être calculées à l'aide de la fonction de densité de covariance  $c(\cdot)$ . Pour ce processus, il suffit de considérer deux cas correspondant à la position de deux points (ils peuvent appartenir au même *cluster* ou non). La fonction  $c(\cdot)$  s'écrit alors :

$$c(u) = egin{cases} \lambda \operatorname{E}(D) & ext{quand} \quad u = 0, \\ \lambda \operatorname{E}\{D(D-1)\} \int_{-\infty}^{\infty} f(x) f(x+u) \mathrm{d}x & ext{quand} \quad u > 0, \end{cases}$$

où  $\lambda$  est le taux du processus de Poisson gérant les centres de *cluster*, D dénote la variable aléatoire du nombre de cellules par *cluster* et f représente la densité de probabilité de la distance des origines des cellules au centre du *cluster*.

# Chapitre 3

# Le modèle à cellules rectangulaires de Neyman-Scott

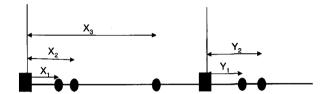
Les processus d'amas de cellules (cluster) les plus utilisés sont basés sur le processus de Poisson. Le plus simple d'entre eux est obtenu en associant une seule cellule de durée et d'intensité aléatoire à chaque point d'un processus de Poisson. Un type plus sophistiqué de mécanisme de cluster considère chaque point du processus de Poisson comme point de référence pour un sous-ensemble de cellules. Dans ce chapitre, nous présentons deux modèles de ce type souvent utilisés pour représenter la pluie : le processus de Neyman-Scott et le processus de Bartlett-Lewis. Puisque dans cette thèse nous nous concentrons principalement sur le modèle de Neyman-Scott, il est présenté plus en détail.

### 3.1 Modèles ponctuels de Poisson

Les modèles de pluie présentés dans cette section utilisent la théorie des processus ponctuels de type cluster. Nous avons indiqué dans l'introduction (chapitre 1) que la structure hiérarchique des champs de précipitations pouvait motiver l'utilisation de ce type de modèle. Comme l'indique Kakou (1997), dans ces modèles les plus grands éléments, appelés averses, sont générateurs d'éléments de plus petite échelle, appelés cellules pluvieuses. Il doit être mentionné à ce stade que le concept de cellule pluvieuse n'est pas précis. Bien que nous le considérons comme un dispositif mathématique (plutôt qu'une entité physique) utile dans le mécanisme de cluster, il est préférable que les paramètres qui le décrivent aient des interprétations physiques et prennent des valeurs

réalistes. En général, les valeurs que peuvent prendre les paramètres dépendent de la résolution des données (c'est-à-dire l'échelle temporelle ou spatiale utilisée) et du type de modèle. Il est alors plutôt arbitraire et dangereux d'avoir des idées arrêtées sur les caractéristiques de ces cellules.

### **Neyman - Scott**



Les {Xi} et {Yi} suivent classiquement une loi exponentielle, sont indépendants et identiquement distribués.

### **Bartlett - Lewis**



Les  $\{X_i\}$  et  $\{Y_i\}$  suivent classiquement une loi exponentielle et forment un processus de renouvellement.

Fig. 3.1 – Construction des modèles à cellules rectangulaires de Neyman-Scott et Bartlett-Lewis. Les deux modèles diffèrent par le type de distribution des origines des cellules par rapport aux origines des averses.

Kavvas & Delleur (1981) et Waymire & Gupta (1981a,b,c) sont les premiers à développer ce type de modèle pour la pluie. Leurs travaux ont permis le développement de deux modèles :

- 1. Le modèle le plus simple associe simplement une cellule à chaque point d'un processus de Poisson. Cette cellule possède une intensité et une durée aléatoire. Rodríguez-Iturbe et al. (1987b) montrent cependant les limites d'un tel modèle où la probabilité que les événements pluvieux se succèdent trop vite est forte.
- 2. Kavvas & Delleur (1981) considèrent un modèle de processus ponctuel de type *cluster*, à temps continu, où le premier niveau est le mécanisme de création des pluies, par exemple

les fronts pluvieux. Au second niveau sont générés les éléments du processus qui produisent réellement la pluie, que l'on désigne par le nom de cellules pluvieuses. Les centres du *cluster* arrivent selon un processus de Poisson et chacun d'eux engendre un nombre aléatoire de cellules pluvieuses. L'intervalle de temps entre l'origine de l'amas et ses membres est distribué exponentiellement. Ce modèle possède une structure de Neyman-Scott (Neyman & Scott, 1958) en temps et a été utilisé pour étudier des séquences de pluies journalières.

Rodríguez-Iturbe et al. (1987b) introduisent ensuite deux types de modèle utilisant les processus ponctuels de cluster pour étudier le comportement des précipitations d'un point fixe de l'espace. Dans les deux modèles, les origines des amas de cellules arrivent selon un processus de Poisson et génèrent un nombre aléatoire de cellules pluvieuses, qui sont représentées par des rectangles (durée et intensité). Une représentation graphique de la structure temporelle de ces deux types de modèles est donnée à la Figure 3.1. Dans le premier, appelé le processus de Neyman-Scott, le nombre de cellules par averse suit une distribution géométrique ou de Poisson. Les origines des cellules sont indépendantes et identiquement distribuées (iid). Elles sont éloignées de l'origine de l'averse selon une loi exponentielle. Le second est nommé le modèle de Bartlett-Lewis. La différence avec le précédent modèle est la distribution des origines des cellules qui forment un processus de Poisson (c'est-à-dire que les intervalles entre chaque origine successive suivent une loi exponentielle).

Le développement de ces modèles de pluie a été poursuivi par de nombreux auteurs, et diverses variations des versions de base (Rodríguez-Iturbe, 1986; Rodríguez-Iturbe et al., 1987a,b) ont été considérées (dans la suite et tout au long de cette thèse, les versions de base des modèles de Neyman-Scott (Neyman-Scott Rectangular Pulses Model, NSRPM) et de Bartlett-Lewis (Bartlett-Lewis Rectangular Pulses Model, BLRPM) correspondront à celles décrites par Rodríguez-Iturbe et al., 1987a). La surestimation systématique de la probabilité de sécheresse, c'est-à-dire la probabilité de ne pas avoir de pluie pendant une période d'une durée quelconque, incite Rodríguez-Iturbe et al. (1988) à modifier le modèle de Bartlett-Lewis en y ajoutant un paramètre. Le paramètre lié à la durée des cellules est alors tiré d'une distribution gamma. Peu après, Entekhabi et al. (1989) apportent une modification similaire au modèle de Neyman-Scott. Cowpertwait (1994) généralise le modèle en permettant aux cellules d'être de différents types : abondante (heavy) ou légère (light) selon leur nature convective ou stratiforme. La sous-estimation systématique des probabilités d'avoir des périodes sans pluie conduit Cowpertwait et al. (1996) à introduire dans la procédure d'estimation des paramètres du modèle les probabilités de transition entre les états sec et humide.

Plus récemment, il est important de souligner les nombreux développements qui ont été réalisés dans le cadre des projets DEFRA (*Department for Environment Food and Rural Affairs*). La plupart de ces développements sont formalisés dans des rapports internes (voir par exemple Chandler & Onof, 2005) et s'intéressent aussi bien à l'aspect théorie qu'à l'application. Nous pouvons citer entre autres :

- la généralisation du modèle à n types de cellules, comportant chacun 2 paramètres (intensité et durée);
- l'aléation (randomization) du paramètre associé à la durée de la cellule avec une loi gamma;
- le développement d'une procédure d'estimation des paramètres;
- l'identifiabilité des paramètres;
- le calcul d'intervalles de confiance des paramètres estimés;
- la validation du modèle sur les valeurs extrêmes.

Le modèle de Bartlett-Lewis à paramètre aléatoire (Random Parameter Bartlett-Lewis Model, RPBLM) dans lequel le paramètre associé à la durée des cellules est distribuée selon une loi gamma semble donner des résultats particulièrement intéressant (Chandler & Onof, 2005) pour un nombre de paramètres raisonnable (c'est-à-dire six paramètres dans le cas d'une distribution exponentielle pour l'intensité des cellules). Nous notons cependant que le calcul des propriétés statistiques de cette généralisation du BLRPM n'est pas direct puisqu'il n'existe pas d'expression analytique des moments (Onof, 2003). Par conséquent, nous n'avons pas adapté cette variation à la structure du Neyman-Scott. Ce modèle n'a donc pas été étudié dans cette thèse.

La différence de performance entre les modèles de Neyman-Scott et de Bartlett-Lewis n'est pas évidente et dépend des applications considérées. Kakou (1997) considère que les comparaisons entre les deux types de modèles n'ont pas démontré une différence significative, et le choix entre les deux reste plutôt arbitraire. Cette remarque est confirmée par Chandler & Onof (2005). De plus, des arguments théoriques sont fournis par Cowpertwait (1998) pour démontrer leur similarité. Il montre en effet que les versions de base des deux modèles ont des propriétés du premier et second ordres équivalentes. Chandler & Onof (2005) recommandent l'application de la structure de Bartlett-Lewis car l'expression de la probabilité de période sans pluie est disponible seulement pour le BLRPM alors que ce calcul requiert des intégrations numériques pour le NSRPM.

Dans notre cas, nous privilégions le modèle de Neyman-Scott afin d'éviter de multiplier le nombre de modèles à comparer par deux, puisque les généralisations peuvent souvent être introduites dans les deux processus. De plus, une méthode d'estimation des paramètres développée par Favre et al. (2004b) est seulement disponible pour la version de base du NSRPM. Enfin, les arguments fournis par Chandler & Onof (2005) en faveur du BLRPM, à savoir le problème du temps de calcul de la proportion de périodes sans pluie (qui doit être évalué numériquement dans le cas du NSRPM), ne semblent pas pertinents dans notre cas. Par exemple, l'évaluation du moment du troisième ordre avec les modèles développés (voir chapitre 6) est beaucoup plus exigeant en ressources informatiques. Le temps de calcul lié à l'évaluation numérique des proportions de périodes sans pluie est donc très inférieur au temps de calcul du moment du troisième ordre, en raison de la taille de son expression, et ne nous a pas limité dans l'application du modèle de Neyman-Scott.

La suite de cette partie présente le processus ponctuel de Neyman-Scott dans sa version de base.

# 3.2 Définition et propriétés du modèle de Neyman-Scott de base

Les mêmes notations que Rodríguez-Iturbe et~al.~(1987b) sont utilisées, à l'exception du nombre de cellules que nous désignons par D pour éviter toute confusion avec la copule que nous noterons C par la suite. Supposons que les temps d'arrivées des origines d'averses suivent un processus de Poisson de taux  $\lambda$ . Un nombre aléatoire D de cellules est associé à chaque origine d'averse. Les origines de cellules sont séparées indépendamment de l'origine de l'averse selon une distance qui suit une loi exponentielle de paramètre  $\beta$ . Une pulsation rectangulaire est associée avec chaque origine de cellule, représentant la cellule pluvieuse. Une représentation graphique d'une réalisation possible de ce modèle est donnée à la Figure 3.2. L'intensité de la cellule est désignée par la variable aléatoire X et la durée de la cellule par L. L'intensité totale instantanée au temps t, Y(t), est la somme de toutes les intensités de cellules actives au temps t. Soit  $X_{t-u}(u)$  une variable aléatoire qui représente l'intensité de pluie au temps t engendrée par la cellule débutant au temps t-u, alors  $Y(t) = \int_{u=0}^{\infty} X_{t-u}(u) dN(t-u)$  où dN(t-u) est un booléen indiquant la présence d'une origine de cellule à t-u (voir équation 2.1). Quand l'intensité et la durée des cellules sont deux variables indépendantes,  $X_{t-u}(u)$  est définie par :

$$X_{t-u}(u) = \begin{cases} X & \text{avec une probabilit\'e} \quad 1 - F_L(u) = e^{-\eta u}, \\ 0 & \text{avec une probabilit\'e} \quad F_L(u) = 1 - e^{-\eta u}, \end{cases}$$
(3.1)

où la durée des cellules est exponentiellement distribuée selon le paramètre  $\eta$ .

Les propriétés du second ordre de l'intensité totale Y(t) du NSRPM ont été calculées par Rodríguez-Iturbe et al. (1987b). Généralement, les données de pluie sont disponibles à des échelles de temps horaire ou journalière. Il est dans ce sens plus intéressant d'étudier certaines propriétés du modèle, comme les expressions des différents moments, à ces échelles d'agrégation (par exemple afin d'estimer les paramètres du modèle par la méthode des moments). Notons  $Y_i^h$  la quantité totale de pluie tombée dans le *i*ème intervalle de temps de longueur h, de telle façon que :

$$Y_i^h = \int_{(i-1)h}^{ih} Y(t) dt.$$

Donc, si h est mesuré en heures,  $\{Y_i^h: i=1,2,...\}$  représentent une série temporelle de pluie à un niveau d'agrégation de h heures. Les propriétés du premier et second ordre sont exprimées par les

relations suivantes:

$$\operatorname{Cov}\{Y_{i}^{h}, Y_{i+k}^{h}\} = \frac{\lambda(1 - e^{-\eta h})^{2} e^{-\eta(k-1)h}}{\eta^{3}} \left\{ \mu_{D} \operatorname{E}(X^{2}) + \frac{\operatorname{E}(D^{2} - D)\mu_{X}^{2}\beta^{2}}{2(\beta^{2} - \eta^{2})} \right\} -\lambda(1 - e^{-\beta h})^{2} e^{-\beta(k-1)h} \frac{\operatorname{E}(D^{2} - D)\mu_{X}^{2}}{2\beta(\beta^{2} - \eta^{2})},$$
(3.4)

où  $\mu_X$  désigne l'espérance de l'intensité des cellules et  $\mu_D$  l'espérance du nombre de cellules. Il est également possible d'obtenir l'expression du moment du troisième ordre (centré) qui est donné par Cowpertwait (1998) pour le modèle de Neyman-Scott :

$$\xi_{h} = \mathbb{E}[\{Y_{i}^{h} - \mathbb{E}(Y_{i}^{h})\}^{3}] = 6\lambda \,\mathbb{E}(X^{3})\mu_{D}(\eta h - 2 + \eta h \,\mathrm{e}^{-\eta h} + 2\,\mathrm{e}^{-\eta h})/\eta^{4} 
+ 3\lambda\mu_{X} \,\mathbb{E}(X^{2}) \,\mathbb{E}\{D(D-1)\}f(\eta,\beta,h)/\{2\eta^{4}\beta(\beta^{2} - \eta^{2})^{2}\} 
+ \lambda\mu_{X}^{3} \,\mathbb{E}\{D(D-1)(D-2)\}g(\eta,\beta,h)/2\eta^{4}\beta(\eta^{2} - \beta^{2})(\eta-\beta)(2\beta+\eta)(\beta+2\eta)\},$$
(3.5)

où les fonctions  $f(\eta, \beta, \theta, h)$  et  $g(\eta, \beta, \theta, h)$  s'expriment comme suit :

$$\begin{split} f(\eta,\beta,\theta,h) &= -2\eta^3\beta^2\,\mathrm{e}^{-\eta h} - 2\eta^3\beta^2\,\mathrm{e}^{-\beta h} + \eta^2\beta^3\,\mathrm{e}^{-2\eta h} + 2\eta^4\beta\,\mathrm{e}^{-\eta h} + 2\eta^4\beta\,\mathrm{e}^{-\beta h} \\ &\quad + 2\eta^3\beta^2\,\mathrm{e}^{-(\eta+\beta)h} - 2\eta^4\beta\,\mathrm{e}^{-(\eta+\beta)h} - 8\eta^3\beta^3h + 11\eta^2\beta^3 - 2\eta^4\beta \\ &\quad + 2\eta^3\beta^2 + 4\eta\beta^5h + 4\eta^5\beta h - 7\beta^5 - 4\eta^5 + 8\beta^5\,\mathrm{e}^{-\eta h} - \beta^5\,\mathrm{e}^{-2\eta h} \\ &\quad - 2h\eta^3\beta^3\,\mathrm{e}^{-\eta h} - 12\eta^2\beta^3\,\mathrm{e}^{-\eta h} + 2h\eta\beta^5\,\mathrm{e}^{-\eta h} + 4\eta^5\,\mathrm{e}^{-\beta h}, \end{split}$$

$$\begin{split} g(\eta,\beta,\theta,h) = & 12\eta^5\beta\,\mathrm{e}^{-\beta h} + 9\eta^4\beta^2 + 12\eta\beta^5\,\mathrm{e}^{-\eta h} + 9\eta^2\beta^4 + 12\eta^3\beta^3\,\mathrm{e}^{-(\eta+\beta)h} \\ & - \eta^2\beta^4\,\mathrm{e}^{-2\eta h} - 12\eta^3\beta^3\,\mathrm{e}^{-\beta h} - 9\eta^5\beta - 9\eta\beta^5 - 3\eta\beta^5\,\mathrm{e}^{-2\eta h} \\ & - \eta^4\beta^2\,\mathrm{e}^{-2\beta h} - 12\eta^3\beta^3\,\mathrm{e}^{-\eta h} + 6\eta^5\beta^2h - 10\beta^4\eta^3h + 6\beta^5\eta^2h \\ & - 10\beta^3\eta^4h + 4\beta^6\eta h - 8\beta^2\eta^4\,\mathrm{e}^{-\beta h} + 4\beta\eta^6h + 12\beta^3\eta^3 \\ & - 8\beta^4\eta^2\,\mathrm{e}^{-\eta h} - 6\eta^6 - 6\beta^6 - 2\eta^6\,\mathrm{e}^{-2\beta h} - 2\beta^6\,\mathrm{e}^{-2\eta h} \\ & + 8\eta^6\,\mathrm{e}^{-\beta h} + 8\beta^6\,\mathrm{e}^{-\eta h} - 3\beta\eta^5\,\mathrm{e}^{-2\beta h} \,. \end{split}$$

Pour l'instant, aucun choix n'a été fait pour la distribution de D et X. Les candidats classiques pour le nombre de cellules associé à chaque averse sont la loi géométrique et la loi de Poisson. Les résultats sont directs dans les deux cas. Velghe  $et\ al.\ (1994)$  indiquent cependant une meilleure performance avec la loi géométrique. Pour cette raison, nous allons privilégier cette loi dans la suite

de la recherche. Il est à noter qu'un changement de distribution est très facile à implémenter (par exemple si la distribution de Poisson est considérée), principalement à cause de l'indépendance des propriétés du processus ponctuel et des cellules. Cependant, puisque la comparaison entre la distribution de Poisson et la distribution géométrique multiplie le nombre de modèles possibles par deux et que, selon notre expérience, la différence de performance n'est pas importante, nous n'avons pas inclus cette variation dans ces travaux.

La probabilité qu'un intervalle de temps de durée quelconque ne comporte pas de pluie (*dry probability*) peut se révéler très utile dans la calibration des modèles. Quand la loi suivie pour la génération du nombre de cellules est une loi géométrique, Cowpertwait (1994) exprime cette probabilité de la manière suivante :

$$PD(h) = \exp(-\lambda h) \left\{ \exp(-\beta h)(1 - \mu_D) + \mu_D \right\}^{-\frac{\lambda}{\beta(1 - \mu_D)}} \exp\left[-\lambda \int_0^\infty \{1 - p_t(h)\} dt\right], \quad (3.6)$$

οù

$$p_t(h) = \frac{\mathrm{e}^{-\beta(t+h)}(\eta - \beta) - \eta \,\mathrm{e}^{-\beta t} + \beta \,\mathrm{e}^{-\eta t} + \eta - \beta}{(\mu_D - 1)\{\mathrm{e} - \beta(t+h)(\beta - \eta) + \eta \,\mathrm{e}^{-\beta t} - \beta \,\mathrm{e} - \eta t\} + \eta - \beta}.$$

L'intégrale  $\int_0^\infty \{1 - p_t(h)\} dt$  dans l'expression (3.6) peut ensuite être approchée en utilisant un développement de Taylor de  $\frac{1}{1+x}$ . Cette approximation paraît satisfaisante à une échelle horaire mais pour des échelles de temps supérieures, il est préférable d'utiliser une méthode numérique d'intégration.

En incluant une dépendance entre les variables constituantes des cellules de pluie, les expressions des moments du premier et second ordres ne sont plus valables (excepté la probabilité de périodes sans pluie). Le chapitre 6 décrit comment il est possible d'inclure une dépendance entre la durée et l'intensité des cellules dans le modèle de Neyman-Scott. Nous calculons les nouvelles expressions des moments, pour quelques familles de fonctions multivariées. Le chapitre suivant présente la théorie associée aux copules, les copules étant les distributions qui modélisent la dépendance.

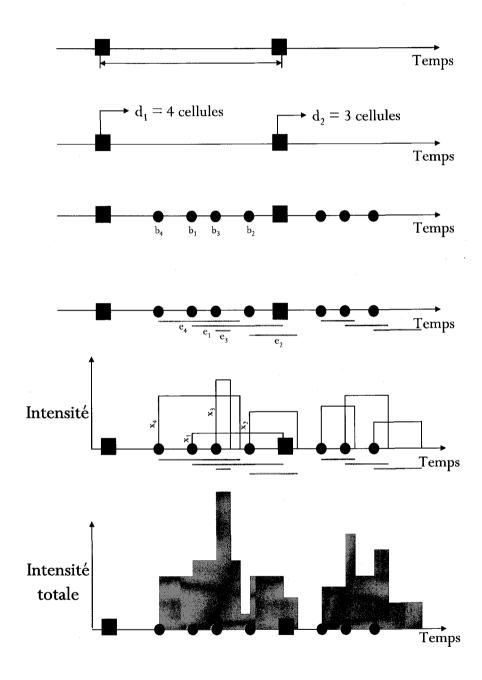


FIG. 3.2 – Réalisation possible du modèle à pulsations rectangulaires. La procédure produit en premier les origines des averses, puis les caractéristiques liées aux cellules pluvieuses, c'est-à-dire leur nombre  $(d_i)$ , leur position  $(b_i)$ , leur durée  $(e_i)$  et leur intensité  $(x_i)$ .

## Chapitre 4

## La modélisation de la dépendance avec les copules

Les copules constituent un outil mathématique très flexible servant à modéliser la dépendance entre plusieurs variables aléatoires. En finance, elles peuvent être appliquées à la gestion des portefeuilles d'actions, l'estimation des mesures de risques (par exemple la Valeur à Risque VaR ou Value at Risk), l'agrégation du risque, les modèles de crédit (voir Embrechts et al., 2002; Cherubini et al., 2004). Nous pouvons également relever de nombreuses applications dans les sciences actuarielles (Frees & Valdez, 1998). Récemment, les copules ont fait leur apparition en hydrologie (voir par exemple De Michele & Salvadori, 2003; Favre et al., 2004a; Grimaldi & Serinaldi, 2006; Genest & Favre, 2007; Renard & Lang, 2007). La représentation par copule possède l'avantage de séparer la structure de dépendance du comportement des marges. Pour une revue détaillée de la théorie des copules, le lecteur pourra consulter les monographies de Joe (1997) et Nelsen (2006).

Il est à noter qu'il n'existe pas de consensus concernant le fait que les copules constituent la solution optimale au problème de la modélisation de la dépendance, comme l'a récemment illustré la discussion animée entre plusieurs des spécialistes des copules. L'origine de ces échanges est un papier de Mikosch (2006) qui critique l'omniprésence et l'utilisation "aveugle" des copules dans la littérature récente, comme si ces derniers fournissaient une solution à tous les problèmes impliquant une dépendance entre plusieurs variables aléatoires. Néanmoins, de nombreux auteurs (voir par exemple Genest & Rémillard, 2006; Joe, 2006; Embrechts, 2006) ont défendu l'importance des travaux sur les copules. Ils ne cautionnent pas pour autant un effet de mode certainement présent.

Dans notre cas, l'idée d'appliquer des copules polynomiales est motivée par le besoin d'obtenir

des expressions algébriques pour les moments théoriques. En effet, en raison de la difficulté à obtenir une fonction de vraisemblance appropriée (voir par exemple Chandler & Onof, 2005, p.11), la méthode généralisée des moments est la façon la plus simple d'estimer les paramètres du NSRPM. Les moments théoriques (espérance, variance, etc.) s'exprimant à l'aide des paramètres du modèle, cette méthode consiste à trouver le vecteur de paramètres pour lequel les moments échantillonnaux sont le plus proche possible des moments théoriques. Cependant, les différentes étapes à franchir afin de calculer les moments limitent le choix des distributions bivariées possibles pour décrire la structure de dépendance entre la durée et l'intensité des cellules. Elles doivent permettre d'obtenir à la fois une structure de dépendance souple (en permettant par exemple de reproduire des degrés de dépendance positifs et négatifs élevés) et des expressions algébriques explicites pour les moments (c'est-à-dire directement exprimées en fonction des paramètres du modèle). Des techniques de développements limités ou d'approximations numériques des intégrales dans l'espace multidimensionnel peuvent être difficilement envisagées dans notre cas puisqu'elles seraient préjudiciables à la calibration du modèle. Il semble en effet très délicat de supposer que les pertes liées à ces approximations soient négligeables.

Classiquement, les auteurs utilisent des familles exponentielles pour décrire les distributions des différentes variables aléatoires (voir Chapitre 3), par exemple la loi exponentielle et la distribution gamma. De nombreuses distributions bivariées avec des lois marginales appartenant à la famille exponentielle sont disponibles dans la littérature (voir Hutchinson & Lai, 1990). Nous avons essayé de calculer les moments avec un grand nombre de ces distributions bivariées. Nous citerons entre autre les essais réalisés avec les distributions bivariées exponentielles de Gumbel (1960) Type-I, de Downton (1970), de Hougaard (1986) et de Cowan (1987) ou avec la distribution bivariée Beta inversée (Lee, 1981). Les expressions à résoudre apparaissent rapidement très difficiles à manipuler (somme de lois hypergéométriques par exemple). Une alternative consiste à construire des distributions bivariées à l'aide de la théorie des copules. Cependant, la plupart des familles de copules (la classe des copules archimédiennes par exemple) sont construites par composition de fonction logarithme ou exponentielle. La manipulation de telles fonctions dans les expressions à évaluer représente alors un défi considérable. Par conséquent, nous avons étudié la possibilité d'utiliser des produits de lois exponentielles comme distribution bivariée de l'intensité et de la durée des cellules, c'est-à-dire des marges exponentielles liées par une copule cubique. Ce chapitre présente la théorie de base sur les copules et met l'emphase sur les copules polynomiales.

## 4.1 Généralités sur les copules

## 4.1.1 Définitions

**Définition 1.** Une copule C en dimension d est une fonction de distribution multivariée qui vérifie :

- $-C:[0,1]^d \to [0,1].$
- C est bornée (grounded), c'est-à-dire que

$$C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_d) = 0 \quad \forall u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_d \ et \ \forall i.$$
 (4.1)

- C est **d-croissante**, c'est-à-dire que le C-volume (différence du d<sup>ième</sup> ordre de C) est positif pour tous les pavés de  $[0,1]^d$ . Pour les copules bivariées, cette condition s'écrit :

$$C(u_2, v_2) + C(u_1, v_1) \ge C(u_1, v_2) + C(u_2, v_1)$$
 pour tous  $0 \le u_1 \le u_2 \le 1$  et  $0 \le v_1 \le v_2 \le 1$ . (4.2)

- C a des marges uniformes, c'est-à-dire que ses distributions marginales  $C_i$  sont telles que :

$$C_i(u) = C(1, \dots, 1, u, 1, \dots, 1) = u \quad \forall u \in [0, 1].$$
 (4.3)

Il se déduit de cette définition que si  $F_1, \ldots, F_d$  sont des fonctions de distribution (univariées), alors  $C\{F_1(x_1), \ldots, F_d(x_d)\}$  est une fonction de distribution multivariée avec les lois marginales  $F_1, \ldots, F_d$  (puisque  $F_i(x_i)$  est distribuée selon une loi uniforme). Il est alors facile de comprendre pourquoi les copules sont très utiles pour la construction et la simulation de distributions multivariées. Le théorème suivant est le plus important de la théorie des copules. Il est connu sous le nom de théorème de Sklar (1959) et est très utilisé en pratique.

**Théorème 1.** Théorème de Sklar : Soit F une loi de probabilité en dimension d, avec des distributions marginales  $F_1, \ldots, F_d$ . Alors F possède la représentation suivante :

$$F(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}. \tag{4.4}$$

De plus, la copule C est unique si les marges sont continues.

Le théorème de Sklar illustre le fait que, dans le cas des distributions multivariées continues, il est possible de séparer les distributions marginales et la structure de dépendance multivariée. D'ailleurs, le corollaire suivant découle du théorème 1.

Corollaire 1. Soit F une loi de probabilité en dimension d, avec des distributions marginales continues  $F_1, \ldots, F_d$  et soit C la copule correspondante (la relation 4.4 est donc vérifiée). Alors, pour tout  $u_1, \ldots, u_d$  dans  $[0, 1]^d$ :

$$C(u_1, \dots, u_d) = F\{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\},\$$

où  $F_i^{-1}$  dénote l'inverse généralisée de  $F_i$ .

## 4.1.2 Propriétés

Dans la suite de ce chapitre, nous limitons l'exposé au cas bidimensionnel (d = 2).

## Expression de la densité bidimensionnelle

Si la distribution bivariée est absolument continue, alors elle admet une densité f qui peut s'exprimer par :

$$f(x_1, x_2) = c\{F_1(x_1), F_2(x_2)\} \times f_1(x_1) \times f_2(x_2), \tag{4.5}$$

avec c(u, v) la densité de la copule C. Notons que la condition 2-croissante est alors équivalente à la positivité de la densité  $c(u, v) = \partial_{1,2}^2 C(u, v) \ge 0$  lorsque celle-ci existe. À partir de la relation (4.5), nous pouvons calculer l'expression de la densité de la copule :

$$c(u,v) = \frac{f\{F_1^{-1}(u), F_2^{-1}(v)\}}{f_1\{F_1^{-1}(u)\} \times f_2\{F_2^{-1}(v)\}}.$$

#### Ordre stochastique de concordance

La définition 2 permet d'introduire un ordre partiel sur les copules.

**Définition 2.** (relation d'ordre) Soient  $C_1$  et  $C_2$  deux copules.  $C_1$  est dite plus petite que  $C_2$   $(C_1 \prec C_2)$  si et seulement si  $C_1(u,v) \leq C_2(u,v)$  pour tout  $(u,v) \in [0,1]^2$ .

L'ordre  $\prec$  est appelé ordre de concordance. Cette relation d'ordre est dite partielle parce que les copules ne peuvent pas toutes se comparer entre elles. Considérons maintenant une famille de copules paramétriques  $C(u, v; \theta)$ , avec  $\theta$  un paramètre. Pour simplifier la notation, nous écrivons

 $C_{\theta}(u,v) = C(u,v;\theta)$ . La famille de copules  $C_{\theta}$  est dite totalement ordonnée si pour tout  $\theta_1 \leq \theta_2$ , elle vérifie :

 $C_{\theta_1} \prec C_{\theta_2}$  (la famille est positivement ordonnée),  $C_{\theta_1} \succ C_{\theta_2}$  (la famille est négativement ordonnée).

#### Symétrie

**Définition 3.** (symétrie) Une copule C est **symétrique** si C(u, v) = C(v, u) pour tout (u, v) dans  $[0, 1]^2$ .

## 4.2 Mesures de dépendance

Les paramètres des familles de copules sont étroitement liés à la notion de degré de dépendance. Un indice souvent utilisé pour mesurer la dépendance est le rho de Pearson (dénoté  $r_n$  dans ce chapitre). Cependant, ce choix d'indicateur de la dépendance n'est viable que dans le cas où la dépendance est de type gaussienne. La suite de cette section est issue de l'article de Genest & Favre (2007) et explique pourquoi il est en règle générale préférable de choisir des indicateurs non-paramétriques pour mesurer la dépendance.

## 4.2.1 Le rho de Spearman

Imitant l'approche usuelle de Pearson pour la mesure de la dépendance, une idée intuitive est de calculer la corrélation entre les paires  $(R_i, S_i)$  de rangs. Cela conduit au rho de Spearman appelé aussi corrélation des rangs puisqu'il se calcule comme :

$$\rho_n = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 (S_i - \bar{S})^2}} \in [-1; 1],$$

οù

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} R_i = \frac{1}{n} \sum_{i=1}^{n} S_i = \bar{S}.$$

Ce coefficient, qui peut être déduit de façon plus pratique par :

$$\rho_n = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^n R_i S_i - 3 \frac{n+1}{n-1} \in [-1; 1],$$

partage avec le coefficient de corrélation de Pearson classique,  $r_n$ , la propriété que son espérance devient nulle quand les variables sont indépendantes. Cependant,  $\rho_n$  est fondamentalement supérieur à  $r_n$  pour les raisons suivantes :

- $E(\rho_n) = \pm 1$  si et seulement si X et Y sont fonctionnellement dépendantes, c'est-à-dire quand la copule associée est l'une des deux bornes de Fréchet-Hoeffding,  $C^-$  ou  $C^+$  (voir sous-section 4.3.1).
- À l'inverse,  $E(r_n) = \pm 1$  si et seulement si X et Y sont des fonctions linéaires de l'autre, ce qui est beaucoup plus restrictif.
- $-\rho_n$  estime le paramètre d'une population qui est toujours bien défini, alors qu'il existe des distributions à queues lourdes (celle de Cauchy par exemple) pour lesquelles une valeur théorique de la corrélation de Pearson n'existe pas.

 $\rho_n$  est un estimateur asymptotiquement sans biais de :

$$\rho = 12 \int_{[0,1]^2} uvC(u,v) dC(u,v) - 3 = 12 \int_{[0,1]^2} C(u,v) dv du - 3.$$

#### 4.2.2 Le tau de Kendall

Une seconde mesure non paramétrique de dépendance très connue, elle aussi basée sur les rangs, est le tau de Kendall. Sa version empirique est donnée par :

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n(n-1)}P_n - 1,$$

où  $P_n$  et  $Q_n$  sont le nombre de paires concordantes et discordantes, respectivement. Deux paires  $(X_i, Y_i), (X_j, Y_j)$  sont dites concordantes quand  $(X_i - X_j)(Y_i - Y_j) > 0$ , elles sont discordantes autrement. La notion de paires concordantes et discordantes est illustré par la Figure 4.1.

Il est évident que  $\tau_n$  est une fonction des rangs des observations, puisque  $(X_i - X_j)(Y_i - Y_j) > 0$  si et seulement si  $(R_i - R_j)(S_i - S_j) > 0$ .  $\tau_n$  est un estimateur asymptotiquement sans biais de l'expression théorique du tau de Kendall, donnée par

$$au=4\int_{[0,1]^2}C(u,v)\mathrm{d}C(u,v)-1.$$

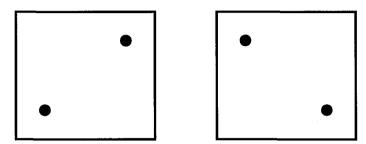


Fig. 4.1 – Deux paires d'observations concordantes (à gauche) et discordantes (à droite).

## 4.3 Familles de copules

## 4.3.1 Copules de références

Ces copules interviennent dans les définitions des structures de dépendance et représentent en quelque sorte l'intervalle de variation des copules en considérant une relation d'ordre.

#### Copule indépendance

La copule indépendance s'exprime comme :

$$C^{\perp}(u,v) = uv.$$

Cette copule est très importante puisqu'elle représente la structure de dépendance de deux variables indépendantes. Si X et Y sont deux variables aléatoires indépendantes de fonctions de répartition respectives  $F_1$  et  $F_2$ , la fonction de répartition jointe est caractérisée par la copule  $C^{\perp}$ , nous avons bien  $F(X,Y) = C^{\perp}\{F_1(X), F_2(Y)\} = F_1(X)F_2(Y)$ .

#### Les bornes de Fréchet-Hoeffding

La copule  $C^-$  désigne la borne inférieure de Fréchet et se calcule comme :

$$C^{-}(u, v) = \max(u + v - 1, 0).$$

Elle caractérise deux variables aléatoires X et Y qui sont contre-monotones, c'est-à-dire qu'il existe une variable aléatoire Z telle que X = f(Z) et Y = g(Z) avec f une fonction non-croissante et g une fonction non-décroissante.

La copule  $C^+$  dénote la borne supérieure de Fréchet et s'écrit de la manière suivante :

$$C^+(u,v) = \min(u,v).$$

Elle caractérise deux variables aléatoires X et Y qui sont co-monotones, c'est-à-dire qu'il existe une variable aléatoire Z telle que X = f(Z) et Y = g(Z) avec f et g deux fonctions non-décroissantes.

L'inégalité des bornes de Fréchet-Hoeffding dans le cas des copules s'écrit :

$$C^-(u,v) \leq C(u,v) \leq C^+(u,v) \quad \text{pour toute copule } C, \forall (u,v) \in [0,1]^2.$$

Les copules  $C^-$  et  $C^+$  constituent les bornes de l'intervalle de variation des copules puisqu'elles modélisent les structures de dépendance les plus fortes possibles (négativement et positivement). Le rho de Spearman et le tau de Kendall associés à ces deux copules sont égaux à -1 dans le cas de la copule  $C^-$  et 1 pour la copule  $C^+$ .

## 4.3.2 Copules archimédiennes

**Définition 4.** Genest & Mackay (1986) définissent une copule archimédienne de la façon suivante :

$$C(u,v) = \begin{cases} \Phi^{-1} \{ \Phi(u) + \Phi(v) \} & si \ \Phi(u) + \Phi(v) \le \Phi(0), \\ 0 & sinon, \end{cases}$$

avec  $\Phi$  une fonction de classe  $C_2$  (c'est-à-dire continue, dérivable et de dérivée continue) qui vérifie  $\Phi(1) = 0$ ,  $\Phi'(u) < 0$  et  $\Phi''(u) > 0$  pour tout  $u \in [0, 1]$ .

 $\Phi(u)$  est appelé le **générateur** (ou la **fonction génératrice**) de la copule. Il est à noter qu'il n'est pas nécessaire que  $\Phi$  soit une fonction de classe  $C_2$  partout sur [0,1] pour que C respecte les conditions 4.1 à 4.3, voir contre-exemple 4.5 dans Nelsen (2006). Cependant, la majorité des copules archimédiennes utilisées en pratique ont une fonction génératrice de classe  $C_2$ , à l'exemple de la famille de Frank appliquée dans le chapitre 5.

Les copules archimédiennes présentent un grand intérêt, car elles sont très maniables pour les calculs. Avec cette classe de copules, l'estimation des paramètres par la méthode des moments est par exemple facilité par le résultat suivant (voir Genest & Mackay, 1986):

$$au=1+4\int_0^1rac{\Phi(u)}{\Phi'(u)}\mathrm{d}u,$$

qui exprime le tau de Kendall théorique en fonction des paramètres de la copule archimédienne. L'estimation des paramètres de la copule par la méthode des moments consiste à faire correspondre une estimation empirique du tau de Kendall à cette expression théorique.

De nombreux exemples de copules archimédiennes sont exposées dans la littérature (voir Nelsen, 2006, chapitre 4). Diverses applications ont été réalisées à l'aide de telles copules et leurs propriétés

ne sont pas toutes détaillées ici (par exemple les méthodes d'estimation des paramètres des copules ou les techniques de simulation). Un exemple de copules archimédiennes est la famille de Frank (Frank, 1979) qui a le grand avantage de couvrir tout l'intervalle de dépendance. Cette famille va nous permettre dans le chapitre 5 d'introduire une dépendance entre l'intensité et la durée des cellules pour tous les degrés possibles. Le générateur de cette copule archimédienne s'écrit comme suit :

$$\Phi(u) = -\lnigg(rac{\mathrm{e}^{- heta u}-1}{\mathrm{e}^{- heta}-1}igg),$$

et son expression est donnée par :

$$C_{\theta}(u,v) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(\mathrm{e}^{-\theta u} - 1)(\mathrm{e}^{-\theta v} - 1)}{\mathrm{e}^{-\theta} - 1} \right\} \text{ avec } \theta \in \mathbb{R} - \{0\}.$$

C'est une famille ordonnée positivement. Elle est dite **complète** car elle comprend les trois copules de référence  $C^-$ ,  $C^\perp$  et  $C^+$  pour respectivement  $\theta = -\infty$ ,  $0^1$  et  $+\infty$ .

## 4.3.3 Copules polynômiales

Dans le cas de notre modèle de pluie, la copule C formalise le lien entre l'intensité et la durée des cellules pluvieuses. Les paragraphes suivants s'inspirent de la section de Nelsen (2006) qui traite des copules s'écrivant sous la forme de polynômes. Le raisonnement se base sur les questions suivantes : À quel point l'expression d'une copule peut être simple (par exemple, la copule indépendance  $C^{\perp}(u,v) = uv$  est linéaire en chacune de ses variables)? Existe-t-il d'autres copules qui peuvent s'écrire sous forme de polynômes?

#### Copules à sections quadratiques

La seule famille de copules à sections quadratiques en u et v est la famille de **Farlie-Gumbel-Morgenstern** (FGM) discutée par Morgenstern (1956), Gumbel (1958) et Farlie (1960). Pour tout  $\theta$  dans [-1,1], cette famille de copules s'exprime comme :

$$C_{\theta}(u, v) = uv + \theta u v(1 - u)(1 - v).$$

Cette copule est très utilisée en pratique car elle possède une expression analytique très simple. Pour cette famille, des solutions analytiques s'obtiennent facilement pour la plupart des calculs. Par exemple la relation entre le tau de Kendall  $\tau$  et le paramètre  $\theta$  est directe. De même, la dépendance de queue (tail dependence, c'est-à-dire la dépendance entre les valeurs extrêmes) s'écrit sous forme explicite pour cette famille de copules. Elle est aussi très simple à simuler. L'utilisation de copules quadratiques avec des marges exponentielles est a priori un choix judicieux de distributions bivariées entre l'intensité et la durée des cellules. Les moments agrégés peuvent être calculés et la

 $<sup>^{1}</sup>$ la copule de Frank n'est pas définie pour  $\theta=0$  mais elle possède cependant une continuité en 0

méthode des moments peut ensuite être appliquée. La famille de copules FGM avec des marges exponentielles est équivalente à la distribution Gumbel type-II, comme introduit dans le NSRPM par Kim & Kavvas (2006). Cependant, cette famille de copules reste limitée en raison du faible intervalle de degré de dépendance qu'elle peut modéliser ( $\tau \in \left[-\frac{2}{9}; \frac{2}{9}\right]$ ). Les copules à sections cubiques (Nelsen et al., 1997) permettent d'augmenter les degrés de dépendance en comparaison des copules à sections quadratiques. À notre connaissance, cette classe de copules n'a jamais été utilisée en pratique.

## Copules à sections cubiques

En utilisant la notation de Nelsen et al. (1997), appelons S l'union des points contenus dans le carré  $[-1,2] \times [-2,1]$  et des points à l'intérieur et sur l'ellipse  $\{(u,v) \in \mathbb{R}^2, u^2 - uv + v^2 - 3u + 3v = 0\}$ . Les copules à sections cubiques à la fois en u et v sont définies par :

$$C(u,v) = uv + uv(1-u)(1-v)\{A_1v(1-u) + A_2(1-v)(1-u) + B_1uv + B_2u(1-v)\}, \quad (4.6)$$

où  $A_1, A_2, B_1, B_2$  sont des constantes réelles telles que les points  $(A_2, A_1), (B_1, B_2), (B_1, A_1)$  et  $(A_2, B_2)$  se situent tous dans S. Les principaux théorèmes relatifs aux copules cubiques sont disponibles dans l'Annexe A.1. Plusieurs propriétés peuvent être obtenues de cette représentation d'une copule cubique. Nelsen et al. (1997) ont montré que si une copule C est définie par l'expression (4.6), les mesures de dépendance s'écrivent comme :

$$\rho = (A_1 + A_2 + B_1 + B_2)/12$$

et

$$\tau = (A_1 + A_2 + B_1 + B_2)/18 + (A_2B_1 - A_1B_2)/450. \tag{4.7}$$

Si C possède des sections cubiques à la fois en u et v, alors  $|\rho| \leq \sqrt{3}/3$ . Des calculs simples nous ont permis de trouver que nous avons aussi  $|\tau| \leq \sqrt{494\sqrt{19}-2053/25}$ . Ce résultat est démontré dans l'Annexe A.2. C est symétrique si et seulement si  $A_1 = B_2$ . De nombreuses sous-familles peuvent être construites, si l'on souhaite par exemple obtenir une famille avec un seul paramètre. Quand  $A_1 = A_2 = B_1 = B_2 = \theta$ , nous obtenons le cas particulier de la copule FGM. Lorsque  $A_1 = A_2 = B_1 = B_2 = 0$ , nous obtenons le cas particulier de la copule indépendance  $C^{\perp}$ . Le Tableau 4.1 présente plusieurs de ces familles et indique celles qui incluent l'indépendance comme cas particulier. Le Tableau 4.2 montre les degrés de dépendance qu'elles sont capables de reproduire. Une représentation de l'espace des paramètres couvert dans S pour chaque copule cubique est donnée à la Figure 4.2. Les copules cubiques déjà décrites par Nelsen et al. (1997) (c'est-à-dire les familles de Sarmanov, Frank cubique et Plackett cubique) sont introduites dans l'Annexe A.1. Des familles de copules cubiques à un paramètre sont faciles à construire. La dépendance entre l'intensité et la durée des cellules étant intuitivement négative, nous avons développé les copules suivantes dans le but d'atteindre la dépendance négative maximale.

- **AMH cubique** Nous calculons l'approximation du second ordre de la famille de copules Ali-Mikhail-Haq (AMH) (voir Ali *et al.*, 1978) donnée par

$$C(u,v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

pour  $\theta \in [-1, 1]$ . Cette famille est symétrique et inclut le cas d'indépendance pour  $\theta = 0$ .

- Sym1 Une copule avec des sections cubiques doit respecter les conditions définies au théorème 3 (voir Annexe A.1). Nous pouvons définir une famille de copules cubiques à un paramètre en posant  $A_2 = B_1 = \theta \in [-1, 2]$  et  $A_1 = B_2 = \theta^2 3$  de telle façon que  $(\theta, \theta^2 3)$  se situe dans  $[-1, 2] \times [-2, 1]$ . Cette famille est symétrique et couvre un large intervalle de dépendance mais n'inclut pas le cas d'indépendance comme cas particulier.
- Sym2 Soit  $A_2 = B_1 = \theta \in [-1,3]$  et soit  $A_1 = B_2 = \frac{\theta 3 \sqrt{9 + 6\theta 3\theta^2}}{2}$ . Les points  $(A_2, A_1)$ ,  $(B_1, B_2)$ ,  $(B_1, A_1)$  et  $(A_2, B_2)$  se situent tous dans S et décrivent le contour elliptique de S du point [-1,-2] au point [3,0]. Cette famille est symétrique et atteint les valeurs minimales pour  $\rho$  et  $\tau$ .
- Sym3 Le point dans S atteignant la valeur minimale pour  $\rho$  est  $[1-\sqrt{3},-1-\sqrt{3}]$ . Nous considérons alors une relation linéaire entre  $A_2=B_1$  et  $A_1=B_2$  passant par ce point et le point [0,0]. Nous obtenons une famille qui est symétrique, ayant des relations linéaires en chacun de ses paramètres, contenant l'indépendance et atteignant la valeur minimale pour  $\rho$ .
- Asym1-Asym3 Des familles asymétriques sont facilement obtenues dès que  $A_1 \neq B_2$ . La famille Asym1 a une relation quadratique en  $\theta$  et couvre un large intervalle de dépendance. Les familles Asym2 et Asym3 sont construites en posant  $A_2 = B_1 B_2$  et en s'assurant que  $A_1 \neq B_2$ . Bien qu'elles atteignent des degrés de dépendance moins élevés, la famille Asym3 donne parfois des résultats intéressants dans le cadre de notre problématique (voir chapitre 8).

Tab. 4.1 – Familles de copules à un paramètre avec des sections cubiques en u et v

Nom	$A_1$	$\mathbf{A_2}$	$B_1$	$\mathbf{B_2}$	$\theta \in$	$\subset \mathbf{C}^{\perp}$
Sarmanov	$3\theta - 5\theta^2$	$3\theta + 5\theta^2$	$3\theta + 5\theta^2$	$3\theta - 5\theta^2$	$[-\sqrt{7}/5,\sqrt{7}/5]$	OUI
Frank cub.	$3\theta - \theta^2$	$3\theta + \theta^2$	$3\theta + \theta^2$	$3\theta - \theta^2$	$[-\sqrt{3}/3, \sqrt{3}/3]$	OUI
Plackett cub.	$ heta- heta^2$	$\theta$	$\theta$	$ heta- heta^2$	[-1, 2]	OUI
AMH cub.	$\theta$	$ heta +  heta^2$	heta	$\theta$	[-1,1]	OUI
Sym1	$\theta^2 - 3$	$\theta$	$\theta$	$\theta^2 - 3$	[-1,2]	NON
Sym2	$\frac{\theta-3-\sqrt{9+6\theta-3\theta^2}}{2}$	$\theta$	heta	$\frac{\theta-3-\sqrt{9+6\theta-3\theta^2}}{2}$	[-1, 3]	NON
Sym3	$(2+\sqrt{3})\theta$	$\theta$	$\theta$	$(2+\sqrt{3})\theta$	$[1-\sqrt{3},2-\sqrt{3}]$	OUI
Asym1	$\theta - 2$	heta-1	$\theta-1$	$\theta^2/3 - 2$	[0,3]	NON
Asym2	$\theta(1-\theta)-1$	heta	heta	$\theta$	[-0.8525,1]	NON
Asym3	$\theta(1-\theta)+1$	$\theta$	$\theta$	$\theta$	[-1,0]	NON

TAB. 4.2 – Degrés de dépendance (rho de Spearman et tau de Kendall) des familles de copules à un paramètre avec des sections cubiques en u et v

Nom	$\rho \in$	$ au \in$
Sarmanov	[-0.529, 0.529]	[-0.372,0.372]
Frank cubique	[-0.577, 0.577]	[-0.400, 0.400]
Plackett cubique	[-0.500, 0.167]	[-0.340, 0.113]
AMH cubique	[-0.250, 0.417]	[-0.169, 0.280]
Sym1	[-0.542, 0.500]	[-0.378, 0.340]
Sym2	[-0.577, 0.500]	[-0.401, 0.353]
Sym3	[-0.577, 0.211]	[-0.400, 0.139]
Asym1	[-0.500, 0.500]	[-0.340, 0.340]
Asym2	[-0.428, 0.167]	[-0.289, 0.115]
Asym3	[-0.333, 0.083]	[-0.222, 0.056]

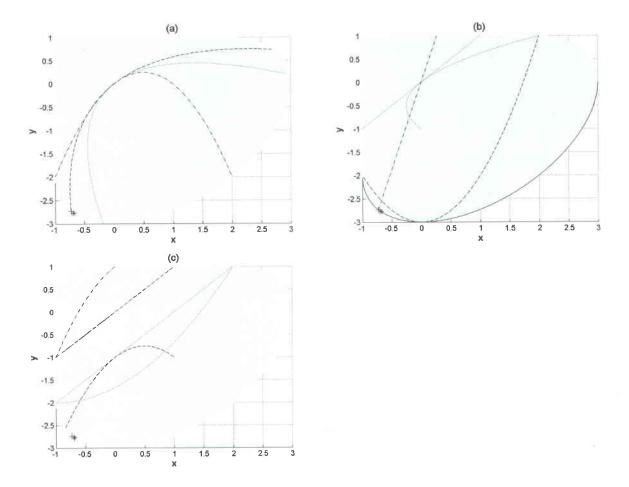


FIG. 4.2 – Graphe de  $(A_2, A_1)$ ,  $(B_1, B_2)$ ,  $(B_1, A_1)$ ,  $(A_2, B_2)$  dans S (surface grisée) pour chaque copule cubique. Les points qui indiquent les valeurs minimales pour le tau de Kendall et le rho de Spearman sont représentés par un '\*' et un '+', respectivement. (a) Sarmanov (ligne pointillée), Frank cubique (ligne segmentée), Plackett cubique (ligne mixte). (b) AMH cubique (ligne pointillée), Sym1 (ligne segmentée), Sym2 (ligne pleine), Sym3 (ligne mixte). (c) Asym1 (ligne pointillée), Asym2 (ligne segmentée), Sym3 (ligne mixte)

## Chapitre 5

## Influence de la dépendance entre l'intensité et la durée des cellules

Dans ce chapitre nous expliquons pourquoi il est impossible d'obtenir une mesure empirique de la dépendance entre l'intensité et la durée des cellules. Cette grandeur serait susceptible de nous renseigner sur la forme de la dépendance, son degré, etc. Cependant, nous n'observons pas directement ces cellules pluvieuses puisque, comme nous l'avons déjà indiqué au chapitre 3, le modèle de Neyman-Scott est purement conceptuel. Les cellules pluvieuses doivent être considérées comme un dispositif mathématique, plutôt qu'une entité physique. Il est par contre possible d'estimer la dépendance entre l'intensité et la durée des averses, en fixant une période sans pluie pour séparer ces averses et une intensité moyenne minimale qui évite de considérer du bruit (au sens statistique). Cela s'est déjà fait par exemple pour modéliser directement la distribution bivariée de l'intensité et de la durée des averses (De Michele & Salvadori, 2003). Néanmoins, il existe plusieurs raisons qui empêchent d'utiliser cette mesure empirique pour en dégager des informations pertinentes sur la dépendance intra-cellule, liées à la discrétisation du temps, à la subjectivité de la méthode de séparation des averses et à l'influence de chacun des paramètres du modèle sur la dépendance entre la durée et l'intensité des averses. Les sections ci-dessous reprennent et expliquent plus en détails chacun de ces points.

## 5.1 La discrétisation du temps

Les mesures de pluie considèrent toujours une échelle de temps qui n'est pas continue mais discrète. Lorsque la pluie est observée à une échelle horaire ou journalière, une perte d'informa-

tion importante résulte de la discrétisation du temps. Prenons le cas d'un orage très intense qui commencerait à 13h50 pour finir à 14h10, avec une intensité moyenne de 120 [mm/h]. Une échelle horaire pourrait par exemple indiquer qu'il est tombé 25 [mm] de 13h à 14h et 15 [mm] de 14h à 15h. Cela a d'abord l'effet de sous-estimer les valeurs extrêmes et dans notre cas d'atténuer la dépendance entre la durée et l'intensité des averses. Rodríguez-Iturbe et al. (1987a) nomme ce phénomène l'effet de lissage (smoothing effect). Les pluviomètres à auget permettent cependant de mesurer les intensités de pluie à des pas de temps beaucoup plus fins (de l'ordre de la minute dans certains cas). Cet effet peut donc être atténué si l'instrumentation permet d'obtenir des mesures à des échelles de temps beaucoup plus fines que l'heure.

## 5.2 La séparation des averses

La nécessité d'appliquer une méthode de séparation des averses est une autre source d'incertitude dans l'obtention d'une estimation du degré de dépendance entre l'intensité et la durée des averses. La période minimale qui doit être fixée pour considérer que deux averses sont bien distinctes (voir par exemple De Michele & Salvadori, 2003) peut, a contrario, agréger deux averses différentes en une seule et par la même atténuer le degré de dépendance entre l'intensité et la durée des averses. De plus, il est évident que développer une méthode de séparation des averses nécessite de nombreux choix très subjectifs (choix de la durée minimale sans pluie entre deux averses, intensité moyenne minimale de l'averse afin de ne pas retenir un bruit de mesure) qui ont tous une grande influence sur les valeurs d'intensité et de durée des averses obtenues.

# 5.3 L'effet *cluster* sur le lien entre l'intensité et la durée des averses

Dans cette section, nous nous intéressons à étudier l'influence des différents paramètres du modèle sur le lien (exprimé en terme de tau de Kendall) entre l'intensité et la durée des averses (noté  $\tau_{\lambda}$  dans la suite de la section). Afin d'isoler cet effet, nous procédons par simulation en temps continu, ce qui nous permet d'éviter les effets de la discrétisation du temps et de ne pas procéder à une méthode de séparation des averses. Nous simulons donc des cellules, et la durée de l'averse correspondante est simplement le temps qui sépare le début de la première cellule de la fin de la dernière (il est donc impossible que les cellules associées à une origine d'averse chevauchent les cellules associées à l'origine d'averse suivante). L'intensité moyenne de l'averse correspond au volume de pluie produit par l'ensemble des cellules divisé par la durée de l'averse. Le degré de dépendance entre la durée et l'intensité des cellules est noté  $\tau_D$ .

## 5.3.1 Influence du paramètre $\mu_D$ sur $\tau_{\lambda}$

Pour étudier l'effet des différents paramètres sur  $\tau_{\lambda}$ , nous avons procédé à un plan de simulation d'averses. Nous avons simulé 10,000 averses pour différentes valeurs de  $\mu_D$ , paramètre qui régit le nombre de cellules avec une loi géométrique ( $\mu_D \in [1; 10, 000]$ ). La densité de points est adaptée aux variations de la courbe et les valeurs de  $\eta$ ,  $\beta$  et  $\alpha$  (paramètre de la loi exponentielle utilisée pour générer l'intensité des cellules) sont fixées respectivement à 3 [h<sup>-1</sup>], 0.2 [h<sup>-1</sup>] et 0.15 [mm] (ces choix de valeurs pour  $\eta$ ,  $\beta$  et  $\alpha$ , bien qu'arbitraires, correspondent à des ordres de grandeur rencontrés en pratique). Nous avons ensuite calculé  $\tau_{\lambda}$  pour ces 10,000 averses. La Figure 5.1(a) décrit l'influence du paramètre  $\mu_D$  sur  $\tau_{\lambda}$ . L'effet de la superposition des cellules s'observe clairement. Quand il est absent ( $\mu_D = 1$ ), l'indépendance entre l'intensité et la durée des cellules ( $\tau_D = 0$ ) se mesure directement sur le lien entre l'intensité et la durée des averses et nous obtenons  $\tau_{\lambda} = 0$ . Cependant, pour un nombre de cellules faibles (c'est-à-dire  $\mu_D$  entre 2 et 10),  $\tau_{\lambda}$  est négatif. Il est important de noter qu'avec une taille d'échantillon n = 10,000, la valeur seuil qui permet de rejeter une hypothèse nulle " $\tau = 0$ " est d'environ  $\pm 0.0131$  à un risque de première espèce de 5%. En pratique, les estimations du paramètre  $\mu_D$  se situent le plus souvent entre 3 et 10. Pour de telles valeurs, nous observons sur la Figure 5.1(a) que  $\tau_{\lambda}$  est significativement différent de 0. Il est délicat d'interpréter cette dépendance en raison du caractère aléatoire de tous ces paramètres. Un début d'explication est que, lorsque les cellules ne se superposent pas et sont très éloignées les unes des autres, la durée de l'averse est d'autant plus grande que ces cellules sont éloignées. De la même manière, lorsque les cellules sont très éloignées, l'intensité moyenne est d'autant plus faible puisque le volume de pluie produit par les cellules va se répartir sur toute cette durée, d'où une dépendance négative. La Figure 5.1(b) illustre ce phénomène en montrant le type de superposition que l'obtient avec  $\eta = 3 \,[h^{-1}], \beta = 0.2 \,[h^{-1}] \text{ et } \alpha = 0.15 \,[\text{mm}] \text{ et } D = 2.$ 

La courbe de la Figure 5.1(a) croît ensuite à une vitesse logarithmique et tend vers une dépendance positive d'intensité supérieure à 0.4 pour  $\tau_{\lambda}$ . La Figure 5.1(c) nous aide à interpréter cette croissance en illustrant le type de superposition que l'obtient avec  $\eta = 3 \, [\text{h}^{-1}]$ ,  $\beta = 0.2 \, [\text{h}^{-1}]$  et  $\alpha = 0.15 \, [\text{mm}]$  et D = 10,000. La dépendance positive peut alors être expliquée par le caractère aléatoire du nombre de cellules. Un nombre de cellules important crée une averse de longue durée et avec une intensité moyenne plus élevée, d'où une dépendance positive. Pour se persuader de cet effet du caractère aléatoire, nous répétons la procédure qui a permis de dessiner la Figure 5.1(a) en fixant le nombre de cellules. La Figure 5.2 représente  $\tau_{\lambda}$  en fonction de D. La courbe ne décrit plus la croissance de la Figure 5.1(a), ce qui est cohérent avec les remarques précédentes. Elle décrit même une décroissance importante puisque  $\tau_{\lambda}$  diminue jusqu'à environ -0.9 pour D = 10,000.

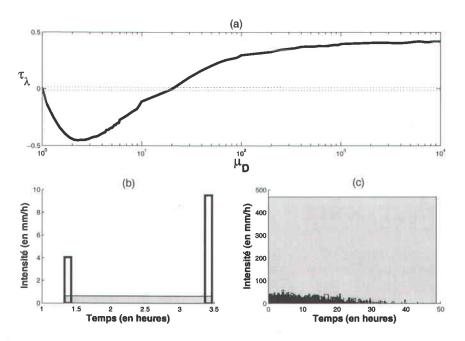


FIG. 5.1 – (a)  $\tau_{\lambda}$  mesure le lien entre la durée et l'intensité des averses en fonction de  $\mu_D$ . L'abscisse représente  $\mu_D$  sur une échelle logarithmique. Le nombre d'averses simulées pour chaque point est 10,000. Les lignes horizontales en pointillé indiquent les seuils inférieur et supérieur de significativité du tau de Kendall. (b) Exemple d'averses et les cellules correspondantes avec D=2. (c) Exemple d'averses et les cellules correspondantes avec D=10,000. La zone en gris représente la durée et l'intensité moyenne de l'averse. Nous fixons  $\eta=3\,[\mathrm{h}^{-1}]$ ,  $\beta=0.2\,[\mathrm{h}^{-1}]$  et  $\alpha=0.15\,[\mathrm{mm}]$ .

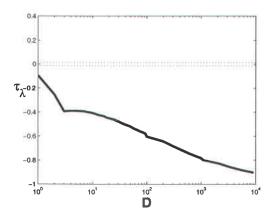


FIG.  $5.2 - \tau_{\lambda}$  mesure le lien entre la durée et l'intensité des averses en fonction de D. L'abscisse représente D sur une échelle logarithmique. Le nombre d'averses simulées pour chaque point est 10,000. Nous fixons  $\eta = 3 \, [\text{h}^{-1}]$ ,  $\beta = 0.2 \, [\text{h}^{-1}]$  et  $\alpha = 0.15 \, [\text{mm}]$ . Les lignes horizontales en pointillé indiquent les seuils inférieur et supérieur de significativité du tau de Kendall.

## 5.3.2 Influence des différents paramètres sur $\tau_{\lambda}$ lorsque la durée et l'intensité des cellules sont indépendantes

Dans cette section, nous allons réaliser le même type de simulations, mais pour différentes valeurs de couples de paramètres. Cela va nous permettre de visualiser et de comprendre plus facilement les interactions entre les paramètres et les effets de chacun de ceux-ci sur la superposition des cellules.

Sur la Figure 5.3(a), le paramètre  $\beta$  varie de 0.1 [h<sup>-1</sup>] à 10 [h<sup>-1</sup>], de telle façon que la superposition des cellules augmente avec  $\beta$  (les origines de toutes les cellules étant plus proches de l'origine de l'averse). Le paramètre  $\mu_D$  varie de 1 à 20, faisant augmenter lui aussi les chances de superposition des cellules.  $\eta$  et  $\alpha$  sont fixés respectivement à 3 [h<sup>-1</sup>] et 0.15 [mm] de façon arbitraire (cela correspond cependant à des valeurs réalistes). Il semble, dans ce cas aussi, que des valeurs de paramètres contribuant à une plus grande superposition des cellules font tendre la dépendance entre l'intensité moyenne et la durée de l'averse à une valeur d'environ 0.4 pour le tau de Kendall. La Figure 5.3(b) représente  $\tau_{\lambda}$  en fonction du paramètre  $\eta$  et de  $\mu_D$ .  $\beta$  est fixé à 0.2  $[h^{-1}]$  et  $\alpha$  à 0.15 [mm]. Les interprétations de la Figure 5.3(a) restent valables. Quand le paramètre  $\eta$  augmente, les cellules pluvieuses sont très courtes et ne se superposent que si  $\mu_D$  est très élevé. Inversement, quand  $\mu_D$  est faible, la superposition des cellules ne se fait que lorsque  $\eta$  est très faible. Les Figures 5.3(a) et 5.3(b) illustrent bien le fait que  $\eta$  et  $\beta$  ont une grand influence sur  $\tau_{\lambda}$ . Ces deux paramètres présentent une interaction importante sur la superposition des cellules. La Figure 5.3(c) représente  $\tau_{\lambda}$  pour  $\beta$  et  $\eta$  variant de 0.01 [h<sup>-1</sup>] à 10 [h<sup>-1</sup>] ( $\mu_D = 5$ ). Il est à souligner qu'en pratique, le paramètre  $\eta$  est en général très supérieur à  $\beta$ . L'allure de cette figure s'explique bien d'après les précédentes remarques. Les cellules se superposent d'autant plus que  $\beta$  est élevé (les origines de cellules sont proches) et que  $\eta$  est faible (les cellules sont plus longues). Les Figures 5.3(d) à 5.3(f) représentent l'effet du paramètre  $\alpha$  (lié à l'intensité des cellules) sur  $\tau_{\lambda}$ . Ces figures représentent bien le fait que ce paramètre a peu d'influence sur la corrélation entre l'intensité et la durée des averses puisque  $\tau_{\lambda}$  reste constant en fonction de  $\alpha$ . Cela semble parfaitement logique avec les remarques précédentes puisque ce paramètre n'influence pas la superposition des cellules.

Les paramètres  $\eta$ ,  $\beta$  et  $\mu_D$  ont donc une influence conjointe sur la superposition des cellules. Ainsi, même lorsque la durée et l'intensité des cellules sont *indépendantes*,  $\eta$ ,  $\beta$  et  $\mu_D$  ont un effet sur la valeur de  $\tau_{\lambda}$ , les cellules se superposant d'autant plus que  $\eta$  est petit et que  $\beta$  et  $\mu_D$  sont grands. Le paramètre  $\alpha$  n'a par contre aucun effet sur  $\tau_{\lambda}$ .

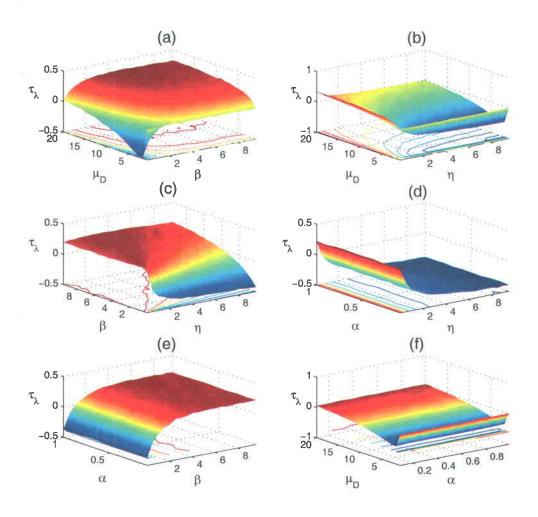


FIG.  $5.3 - \tau_{\lambda}$  en fonction de différentes valeurs des couples de paramètres (a)  $(\mu_D, \beta)$ , (b)  $(\mu_D, \eta)$ , (c)  $(\beta, \eta)$ , (d)  $(\alpha, \eta)$ , (e)  $(\alpha, \beta)$  et (f)  $(\mu_D, \alpha)$ . L'intensité et la durée des cellules sont considérées comme des variables aléatoires indépendantes. Dans chacune des figures, nous fixons  $\eta = 3 \,[h^{-1}]$ ,  $\beta = 0.2 \,[h^{-1}]$ ,  $\alpha = 0.15 \,[\text{mm}]$  et  $\mu_D = 5$  lorsque ces paramètres ne varient pas. Le nombre d'averses simulées pour chaque point est 10,000.

# 5.3.3 Influence des différents paramètres sur $\tau_{\lambda}$ lorsque la durée et l'intensité des cellules sont dépendantes

Nous nous intéressons maintenant à l'introduction d'une dépendance entre l'intensité et la durée des cellules, à l'aide de la famille de copules de Frank (voir sous-section 4.3.2). Cette famille a le grand avantage de couvrir tout l'intervalle de dépendance. Nous avons ainsi généré 10,000 averses pour plusieurs valeurs de  $\tau_D$ , le paramètre  $\mu_D$  étant également variable. Les autres paramètres sont fixés aux mêmes valeurs que précédemment ( $\eta=3\,[{\rm h}^{-1}],\,\beta=0.2\,[{\rm h}^{-1}],\,\alpha=0.15\,[{\rm mm}]$  et  $\mu_D=5$ ). La Figure 5.4 illustre les valeurs de  $\tau_\lambda$  qui résultent de chacune de ces simulations. La relation d'égalité est très nette lorsqu'il n'y a pas de superposition de cellules (c'est-à-dire que lorsque  $\mu_D=1$ , nous avons  $\tau_\lambda=\tau_D$ ). Quand le nombre de cellules augmente,  $\tau_\lambda$  est largement atténué et semble converger vers le profil de la Figure 5.1 (ceci a été vérifié en traçant la même courbe avec cette fois les variables aléatoires intensité et durée des cellules co-monotones, voir Figure 5.5).

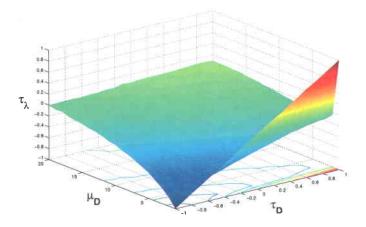


FIG.  $5.4 - \tau_{\lambda}$  en fonction de différentes valeurs du paramètre gérant le nombre de cellules  $\mu_D$ . L'intensité et la durée des cellules sont des variables aléatoires liées par la copule de Frank à un degré indiqué par  $\tau_D$ . Nous fixons  $\eta = 3 \, [\mathrm{h}^{-1}]$ ,  $\beta = 0.2 \, [\mathrm{h}^{-1}]$  et  $\alpha = 0.15 \, [\mathrm{mm}]$ . Le nombre d'averses simulées pour chaque point est 10,000.

Nous avons réitéré cette procédure en faisant varier les autres paramètres. Les résultats des simulations sont présentés à la Figure 5.6. Lorsque  $\alpha$  et  $\tau_D$  varient ensemble,  $\tau_\lambda$  évolue dans un plan et nous retrouvons le fait que  $\alpha$  n'influence pas  $\tau_\lambda$ . Pour les paramètres  $\beta$  et  $\eta$ , l'allure des graphiques est cohérente avec les remarques exposées précédemment. Lorsque les chances de superposition des cellules augmentent (c'est-à-dire pour  $\beta$  supérieur à  $\eta$ ) et que la durée et l'intensité des cellules sont directement liées ( $\tau_D = 1$ ),  $\tau_\lambda$  est très élevé. Dans ces cas, il y aura très souvent une cellule prépondérante qui contribuera pour beaucoup à l'allure de l'averse (voir Figure 5.7). Le volume de pluie produit par l'averse sera en grande partie due à une cellule de grande durée et intensité.

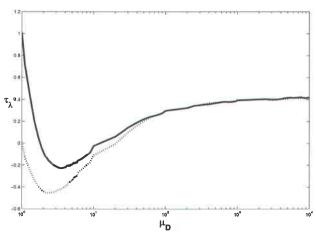


FIG.  $5.5 - \tau_{\lambda}$  en fonction de différentes valeurs du paramètre  $\mu_D$  avec cette fois l'intensité et la durée des cellules co-monotones ( $\tau_D = 1$ ) représenté en trait plein, à comparer avec la courbe de la Figure 5.1 (intensité et durée des cellules indépendantes) en trait pointillé. Nous fixons  $\eta = 3$  [h<sup>-1</sup>],  $\beta = 0.2$  [h<sup>-1</sup>] et  $\alpha = 0.15$  [mm]. Le nombre d'averses simulées pour chaque point est 10,000.

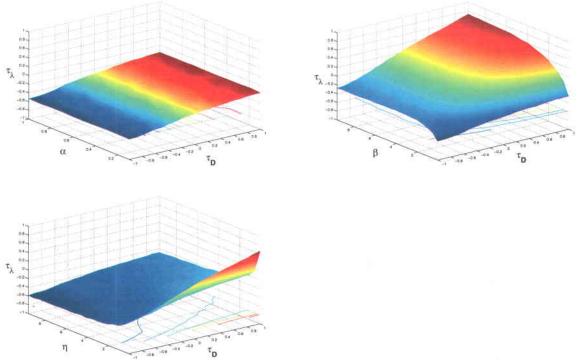


FIG.  $5.6 - \tau_{\lambda}$  en fonction de différentes valeurs des paramètres  $\alpha$ ,  $\beta$  et  $\eta$ . L'intensité et la durée des cellules sont des variables aléatoires liées par la copule de Frank à un degré indiqué par  $\tau_D$ . Dans chacune des figures, nous fixons  $\eta = 3 \, [h^{-1}]$ ,  $\beta = 0.2 \, [h^{-1}]$ ,  $\alpha = 0.15 \, [mm]$  et  $\mu_D = 5$  lorsque ces paramètres ne varient pas.

5.4 Conclusion 43

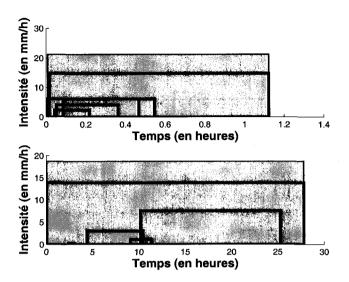


FIG. 5.7 – Deux exemples d'averses dont la durée et l'intensité des cellules sont co-monotones  $(\tau_D=1)$ . En haut,  $\beta=10\,[\mathrm{h}^{-1}], \eta=2\,[\mathrm{h}^{-1}], \mu_D=5$  et  $\alpha=0.15\,[\mathrm{mm}],$  en bas  $\beta=0.2\,[\mathrm{h}^{-1}], \eta=0.1\,[\mathrm{h}^{-1}], \mu_D=5$  et  $\alpha=0.2\,[\mathrm{mm}].$  La zone en gris représente la durée et l'intensité moyenne de l'averse.

## 5.4 Conclusion

Dans les chapitres 1 et 3, nous avons mentionné le fait que les cellules pluvieuses ne sont pas directement observées et doivent être plutôt considérées comme un concept mathématique. Les figures précédentes montrent de plus qu'il est problématique de déduire une information pertinente de façon empirique en ce qui concerne l'allure de la dépendance entre la durée et l'intensité des cellules lorsque tous les paramètres qui interviennent dans la formation de la durée et de l'intensité des averses sont considérés ( $\mu_D$ ,  $\beta$ ,  $\eta$  et  $\alpha$ ). Les paramètres  $\mu_D$ ,  $\beta$  et  $\eta$  agissent conjointement sur la forme des amas de cellules (cluster) et ont ainsi une influence sur  $\tau_{\lambda}$ . Nous notons par exemple que  $\tau_{\lambda}$  peut prendre une valeur de 0.4 sans qu'il n'y ait aucune dépendance entre l'intensité et la durée des cellules ( $\tau_D = 0$ ). De plus, pour des valeurs de  $\mu_D$  rencontrées en pratique (entre 3 et 10), la dépendance introduite entre l'intensité et la durée des cellules est considérablement atténuée par leur agrégation; avec par exemple  $\tau_D = 1$  et  $\mu_D = 5$ , nous obtenons  $\tau_{\lambda} \approx -0.1$  (voir Figure 5.5).

À la lumière des ces remarques, il apparaît extrêmement difficile de développer une méthode permettant d'avoir une mesure empirique de dépendance entre l'intensité et la durée des cellules à cause du grand nombre de paramètres du modèle de Neyman-Scott intervenant dans ce processus. Il n'est donc pas possible d'inclure un moment statistique directement lié à la dépendance entre l'intensité et la durée des cellules dans la méthode généralisée des moments.

## Chapitre 6

# Calcul des moments du NSRPM incluant un lien intensité/durée des cellules

Dans ce chapitre, nous développons le calcul des moments du Neyman-Scott quand une copule cubique modélise le lien entre l'intensité et la durée des cellules (les détails de ces calculs sont disponibles à l'Annexe B). Quand un lien existe entre X et L, la relation (3.1) n'est plus valable. Nous souhaitons obtenir la distribution de l'intensité des cellules X lorsque L > l. Nous notons cette distribution bivariée  $\Psi_{X,L}$ . La copule C désigne la distribution jointe de X et L. Nous avons que :

$$\Psi_{X,L}(x,l) = \Pr(X \le x, L > l) 
= \Pr(X \le x) - \Pr(X \le x, L \le l) 
= F_X(x) - C\{F_X(x), F_L(l)\},$$
(6.1)

où  $F_X$  et  $F_L$  désignent les distributions univariées de X et L, respectivement. Il est important de souligner qu'à ce stade, la relation (6.1) ne limite pas le choix de la distribution bivariée de l'intensité et de la durée des cellules. N'importe quelle distribution bivariée peut s'exprimer à partir d'une copule bivariée et des distributions marginales selon le théorème de Sklar. Cependant, les distributions bivariées classiques rendent le calcul des moments agrégés très complexe et même impossible. L'utilisation des copules cubiques et des marges exponentielles est un moyen élégant d'obtenir une classe flexible de distributions bivariées pour lesquelles les moments agrégés sont facilement calculables.

## 6.1 Espérance, Variance et Covariance

Le moment d'ordre m de la pluie agrégée  $Y_i^h$  peut s'exprimer comme suit :

$$\begin{aligned} \mathrm{E}\{(Y_i^h)^m\} &= \int_0^h \int_0^h \dots \int_0^h \mathrm{E}\left\{Y(t_1)Y(t_2)\dots Y(t_m)\right\} \mathrm{d}t_1 \mathrm{d}t_2 \dots \mathrm{d}t_m \\ &= m! \int_0^h \int_{t_1}^h \dots \int_{t_{m-1}}^h \mathrm{E}\left\{Y(t_1)Y(t_2)\dots Y(t_m)\right\} \mathrm{d}t_1 \mathrm{d}t_2 \dots \mathrm{d}t_m, \end{aligned}$$

avec  $t_1 \leq t_2 \leq \cdots \leq t_m$ . Une des propriétés des processus ponctuels est l'indépendance entre la génération du processus ponctuel  $dN(t_m - u_m)$  et la distribution de l'intensité  $X_{t_m - u_m}(u_m)$ . Nous pouvons alors écrire :

L'espérance de Y(t) peut alors être représentée comme le produit du taux d'apparition des cellules et de l'espérance de la quantité de pluie produite par une cellule (Rodríguez-Iturbe et al., 1987b, paragraphe 3.2), à savoir :

$$E\{Y(t)\} = \int_{u=0}^{\infty} E\{X_{t-u}(u)\} E\{dN(t-u)\}$$
$$= \lambda \mu_D \int_0^{\infty} E\{X_{t-u}(u)\} du$$
$$= \lambda \mu_D \int_0^{\infty} \mu_{X,L \ge u} du,$$

où  $\mu_{X^r,L\geq u}$  désigne  $\int_0^\infty x^r \frac{\partial \Psi_{X,L}(x,u)}{\partial x} \mathrm{d}x$ . L'autocovariance de l'intensité pluvieuse avec un décalage  $\tau$  est donnée par :

$$c_{Y}(\tau) = \text{Cov}\{Y(t), Y(t+\tau)\}\$$

$$= \int_{0}^{\infty} \int_{0}^{\infty} \text{E}\{X_{t-u}(u) X_{t+\tau-v}(v)\} \text{Cov}\{dN(t-u), dN(t+\tau-v)\},$$
(6.3)

où l'autocovariance du processus de comptage peut être exprimé selon Cox & Isham (1980, p.78) comme suit :

$$\operatorname{Cov}\{dN(t-u), dN(t+\tau-v)\} = \lambda \mu_D \Big[ \mathbb{I}_0(\tau+u-v) + \frac{1}{2}\mu_D^{-1} \operatorname{E} \{D(D-1)\} \times \beta \operatorname{e}^{-\beta(\tau+u-v)} \{1 - \mathbb{I}_0(\tau+u-v)\} \Big] dudv$$
(6.4)

quand  $\tau + u - v \ge 0$ .  $\mathbb{I}_0(.)$  est la fonction indicatrice définie par  $\mathbb{I}_0(0) = 1$  et  $\mathbb{I}_0(x) = 0$  autrement (voir la définition 2.2). En introduisant cette expression dans l'équation (6.3), nous obtenons (voir Annexe B.1):

$$c_{Y}(\tau) = \lambda \mu_{D} \int_{\tau}^{\infty} \mu_{X^{2}, L \geq u} du + \frac{\lambda}{2} \operatorname{E} \{D(D-1)\} \beta \left[ \int_{0}^{\infty} \int_{0}^{u+\tau} \mu_{X, L \geq u} \mu_{X, L \geq v} e^{-\beta(u+\tau-v)} dv du \right] + \int_{0}^{\infty} \int_{u+\tau}^{\infty} \mu_{X, L \geq u} \mu_{X, L \geq v} e^{-\beta(v-u-\tau)} dv du \right].$$

$$(6.5)$$

Notons  $\alpha$  et  $\eta$  les paramètres respectifs des distributions exponentielles de l'intensité et de la durée des cellules. Avec le choix d'une copule cubique (avec un paramètre  $\theta$ ) et de marges exponentielles pour modéliser la distribution bivariée de l'intensité et de la durée des cellules, une expression analytique est facilement calculable. Avec la copule FGM, nous obtenons :

$$c_Y(\tau) = \lambda \mu_D \{ -3\theta e^{-2\eta\tau} + 2(4+3\theta) e^{-\eta\tau} \} / (4\alpha^2\eta) + \beta\lambda E\{D(D-1)\} [2\beta e^{-\eta\tau}(\beta^2 - 4\eta^2)$$

$$(2+\theta)(6+\theta) - \beta e^{-2\eta\tau}(\beta - \eta)(\beta + \eta)\theta(8+\theta) + 6 e^{-\beta\tau} \eta \{-2\beta + \eta(4+\theta)\} \{2\beta + \eta(4+\theta)\} ] / \{48\alpha^2\eta(\beta^4 - 5\beta^2\eta^2 + 4\eta^4)\}.$$

Finalement, l'espérance, la variance et l'autocovariance du processus agrégé  $Y_i^h$  peuvent être calculées (Rodríguez-Iturbe et al., 1984) en utilisant les expressions suivantes :

$$\mathrm{E}(Y_i^h) = h \; \mathrm{E}\{Y(t)\},$$
  $\mathrm{Var}\{Y_i^h\} = 2 \int_0^h (h-u) \, c_Y(u) \mathrm{d}u,$   $\mathrm{Cov}\{Y_i^h, Y_{i+k}^h\} = \int_{-h}^h c_Y(kh+v)(h-|v|) \mathrm{d}v,$ 

où k est un entier représentant le décalage de la fonction d'autocovariance. Pour la copule FGM et des marges exponentielles, des expressions assez simples peuvent être obtenues comme suit :

$$E(Y_{i}^{h}) = \frac{h \lambda \mu_{D} (4 + \theta)}{4 \alpha \eta},$$

$$Var(Y_{i}^{h}) = \lambda \mu_{D} [-3\theta + 8 e^{\eta h} (4 + 3\theta) + e^{2\eta h} \{-32 - 21\theta + 2\eta h (16 + 9\theta)\}] / (8\alpha^{2} e^{2\eta h} \eta^{3})$$

$$+ \lambda E\{D(D - 1)\} [8\beta^{3} (\beta^{2} - 4\eta^{2}) (e^{-\eta h} + \eta h - 1) (2 + \theta) (6 + \theta)$$

$$- \beta^{3} (\beta - \eta) (\beta + \eta) (e^{-2\eta h} + 2\eta h - 1) \theta (8 + \theta) + 24\eta^{3} (e^{-\beta h} + \beta h - 1)$$

$$\{\eta^{2} (4 + \theta)^{2} - 4\beta^{2}\} ] / \{96\alpha^{2} \eta^{3} (\beta^{5} - 5\beta^{3} \eta^{2} + 4\beta \eta^{4})\},$$

$$Cov(Y_{i}^{h}, Y_{i+k}^{h}) = \lambda \mu_{D} \{-3(e^{2\eta h} - 1)^{2} \theta + 8 e^{\eta h (1+k)} (e^{\eta h} - 1)^{2} (4 + 3\theta)\} / (16\alpha^{2} e^{2\eta h (1+k)} \eta^{3})$$

$$- \lambda E\{D(D - 1)\} [-8\beta^{3} e^{(-\eta)h (1+k)} (e^{\eta h} - 1)^{2} (\beta^{2} - 4\eta^{2}) (2 + \theta) (6 + \theta)$$

$$+ \beta^{3} e^{-2\eta h (1+k)} (e^{2\eta h} - 1)^{2} (\beta - \eta) (\beta + \eta) \theta (8 + \theta) + 24 e^{-\beta h (1+k)} (e^{\beta h} - 1)^{2} \eta^{3} \{2\beta - \eta (4 + \theta)\} \{2\beta + \eta (4 + \theta)\}] / \{192\alpha^{2} \eta^{3} (\beta^{5} - 5\beta^{3} \eta^{2} + 4\beta \eta^{4})\}.$$

$$(6.8)$$

## 6.2 Moment du troisième ordre

Le moment du troisième ordre nous donne des informations sur l'asymétrie d'une distribution. Selon Cowpertwait (1998) et Onof (2003), la reproduction des extrêmes est grandement améliorée quand le moment du troisième ordre est inclus dans la procédure d'ajustement du modèle. À l'exemple de Cowpertwait (1998), nous calculons les expressions de ce moment dans le cas d'une dépendance entre l'intensité et la durée des cellules. Nous considérons à cette fin l'expression (6.2) avec m=3. Soient  $t_1 < t_2 < t_3$  les positions de trois origines de cellules. L'espérance du processus de comptage est exposé dans l'article de Cowpertwait (1998, équation 2.7) et s'écrit comme suit :

$$E\{dN(t_1)dN(t_2)dN(t_3)\} = \lambda^3 \mu_D^3 dt_1 dt_2 dt_3 
+ \frac{1}{2} \lambda^2 \beta \mu_D E(D^2 - D) \{e^{-\beta(t_3 - t_1)} + e^{-\beta(t_3 - t_2)} + e^{-\beta(t_2 - t_1)}\} dt_1 dt_2 dt_3 
+ \frac{1}{3} \lambda \beta^2 E\{(D^2 - D)(D - 2)\} e^{-\beta(t_3 + t_2 - 2t_1)} dt_1 dt_2 dt_3 + \circ (dt_1 dt_2 dt_3),$$
(6.9)

où  $t_1 < t_2 < t_3$ . Il nous reste à exprimer  $\mathrm{E}\{X_{t_1-u_1}(u_1)X_{t_2-u_2}(u_2)X_{t_3-u_3}(u_3)\}$ . Nous devons considérer différents cas selon la position des trois origines des cellules (par exemple le cas où  $t_1-u_1=t_2-u_2$  pour prendre en compte le cas où  $X_{t_1-u_1}$  et  $X_{t_2-u_2}$  concernent la même cellule). Ainsi,

Les détails de ces calculs sont donnés dans l'Annexe B.2. Avec la copule FGM et des marges exponentielles, l'expression du moment du troisième ordre (centré) se réduit à :

$$\xi_{h} = \mathbb{E}[\{Y_{i}^{h} - \mathbb{E}(Y_{i}^{h})\}^{3}] = 3\lambda\mu_{D}\{-21\,\mathrm{e}^{-2\eta h}(1+\eta h)\theta + 24\,\mathrm{e}^{-\eta h}(2+\eta h)(8+7\theta) + 3(-128+64\eta h) - 105\theta + 49\eta h\theta)\}/(16\alpha^{3}\eta^{4}) + \lambda\,\mathbb{E}\{D(D-1)\}f(\eta,\beta,\theta,h)/\{384\alpha^{3}\eta^{4}\beta(\beta+2\eta)^{2}(\beta+\eta)^{2}\} + \lambda\times \mathbb{E}\{D(D-1)(D-2)\}g(\eta,\beta,\theta,h)/\{1280\eta^{4}\alpha^{3}\beta(\beta-2\eta)^{2}(\beta-\eta)^{2}(2\beta+\eta)(\beta+\eta)(\beta+2\eta)(\beta+3\eta)(\beta+4\eta)\},$$

$$(6.10)$$

où  $f(\eta, \beta, \theta, h)$  et  $g(\eta, \beta, \theta, h)$  sont définis dans l'Annexe B.3.

Quand  $\theta = 0$  pour la copule FGM, nous avons C(u, v) = uv, ce qui correspond au cas particulier de l'indépendance entre l'intensité et la durée des cellules. Pour  $\theta = 0$ , l'équation (6.10) est en accord avec celle obtenue par Cowpertwait (1998). Les formules générales des moments agrégés (excepté le moment du troisième ordre dont l'expression est trop fastidieuse pour être présentée

dans ce document) obtenues en utilisant les copules cubiques peuvent être consultées dans l'Annexe B.4.

## 6.3 Probabilité de période sans pluie

Comme indiqué dans Kakou (1997), les autres propriétés d'intérêt, telles que la probabilité d'avoir une intensité nulle de pluie durant un intervalle de temps arbitraire ou les probabilités de transition entre les intervalles secs et humides ne sont pas des fonctions de l'intensité des cellules de pluie. Cela signifie en particulier que leurs expressions (comme l'expression 3.6) sont les mêmes que X et L soit dépendants ou non.

## Chapitre 7

# Méthode d'estimation des paramètres du modèle

La méthode d'estimation des paramètres du modèle de Neyman-Scott la plus répandue, et qui paraît la plus naturelle, se base sur les moments. Les autres approches classiques sont plus difficiles à appliquer. Par exemple, l'expression de la fonction de vraisemblance de l'intensité de pluie agrégée est impossible à obtenir pour ce type de modèle en raison de la grande complexité de la structure de dépendance temporelle (voir Smith & Karr, 1985a,b; Obeysekera et al., 1987; Chandler & Onof, 2005). Diggle (1983) propose une méthodologie qui consiste à choisir une fonction score (par exemple la densité spectrale ou la distribution du plus proche voisin) et à minimiser ensuite l'erreur quadratique moyenne intégrée (Mean Integrated Square Error, MISE) entre les valeurs théoriques et observées. Cependant, cette méthode dépend de constantes d'étalonnage arbitraires.

## 7.1 Méthode basée sur une fonction de vraisemblance

La méthode d'estimation des paramètres qui consiste à maximiser une fonction de vraisemblance comporte de nombreux avantages. En particulier, la distribution asymptotique des estimateurs du maximum de vraisemblance permet le calcul d'intervalles de confiance pour les paramètres inconnus. Chandler (1995) développe une approche basée sur le théorème de la représentation spectrale, en construisant une fonction de vraisemblance basée sur les coefficients de Fourier. Onof et al. (2000) qualifient cette méthode en ces termes :

- objective, parce qu'il n'est pas nécessaire de choisir quelles statistiques, obtenues à partir

- des données, sont utilisées pour l'ajustement des paramètres (comme dans la méthode des moments),
- exhaustive dans l'utilisation des données, puisque le théorème de Fourier assure qu'il n'y a pas de perte d'information dans le changement de domaine temps/fréquence,
- flexible et facilement réutilisable pour différents modèles.

Chandler (1997) applique cette approche sur les versions de base du BLRPM et du NSRPM. Onof et al. (2000) recommandent alors de combiner cette méthode et celle des moments. Cependant, l'utilisation de cette 'vraisemblance spectrale' approchée nécessite une simplification importante des modèles afin d'obtenir cette fonction. Pour cette raison, Chandler & Onof (2005) conseillent d'appliquer la méthode généralisée des moments. Même si ce choix représente l'inconvénient de ne pas pouvoir utiliser les propriétés asymptotiques des estimateurs du maximum de vraisemblance, nous notons que des méthodes alternatives existent dans le cas de la méthode généralisée des moments, afin par exemple de construire des intervalles de confiance sur les estimations des paramètres (voir Chandler, 2003).

## 7.2 Méthode basée sur les moments

## 7.2.1 Méthode généralisée des moments

La méthode généralisée des moments est décrite dans Rodríguez-Iturbe et al. (1987b). Les fonctions sont hautement non-linéaires en chacun des paramètres de telle sorte qu'une solution exacte n'est pas disponible. Notons  $\phi$  le vecteur des paramètres du modèle. Une solution approchée peut être obtenue en minimisant la fonction objectif

$$\mathcal{O}(\phi) = \sum_{i=1}^{p} \left(1 - \frac{M_i}{\hat{M}_i}\right)^2,\tag{7.1}$$

où p est le nombre de moments inclus dans la procédure. Les quotients  $\frac{M_i}{\hat{M}_i}$  entre les expressions théoriques des statistiques et les estimations correspondantes issues de l'échantillon ne doivent pas trop s'éloigner de 1. L'utilisation de ces rapports nous assure que de grandes valeurs numériques ne dominent pas la procédure d'ajustement des paramètres. Le choix des statistiques observées  $\hat{M}_i$  constitue alors une étape cruciale. Plusieurs possibilités s'offrent à nous, parmi les statistiques dont nous possédons les expressions théoriques, à savoir :

- la moyenne  $E(Y_i^h)$  (expressions 3.2, 6.6 et B.1),
- la variance  $Var(Y_i^h)$ , (expressions 3.3, 6.7 et B.2),
- l'autocovariance de décalage k,  $Cov(Y_i^h, Y_{i+k}^h)$ , avec  $k \ge 1$  (expressions 3.4, 6.8 et B.3),

- la probabilité d'intervalle de temps sec (3.6), la probabilité de transition de l'état sec à humide,
- le moment centré d'ordre 3,  $\mathbb{E}\left[\left\{Y_i^h \mathbb{E}(Y_i^h)\right\}^3\right]$  (expressions 3.5 et 6.10).

Exceptée la moyenne, puisque  $\mathrm{E}(Y_i^{kh})=k\,\mathrm{E}(Y_i^h)$ , les moments calculés à différents niveaux d'agrégation (de l'heure à deux jours) peuvent être introduits dans le système d'équations. Calenda & Napolitano (1999) font une présentation exhaustive des méthodes d'estimation des paramètres et s'intéressent particulièrement à l'étude de l'influence des niveaux d'agrégation sur l'estimation des moments. Calenda & Napolitano (1999) montrent aussi l'importance de bien choisir les échelles d'agrégation pour deux raisons principales :

- les caractéristiques de la fonction objectif (7.1) changent substantiellement avec l'échelle d'agrégation;
- les résultats de l'algorithme d'optimisation varient quand les points de départ de la recherche changent, et ceci tout spécialement quand les échelles sont trop similaires.

Cowpertwait et al. (1996) suggèrent d'utiliser un ensemble de moments qui dépassent en nombre les paramètres du modèle en accordant des poids à chaque statistique, selon l'importance souhaitée dans la procédure d'estimation. Un dernier calcul permet enfin de réduire l'espace de minimisation d'une dimension. En effet, poursuivant le travail de Vanmarcke (1983), Calenda & Napolitano (1999) définissent l'échelle de fluctuation du processus observé par :

$$\chi = \lim_{T \to \infty} \left\lceil \frac{\operatorname{Var} \left\{ \int_{t - \frac{T}{2}}^{t + \frac{T}{2}} Y(t) dt \right\}}{T \operatorname{Var} \{ Y(t) \}} \right\rceil.$$

Pour la version de base du processus de Neyman-Scott, avec une distribution géométrique pour le nombre de cellules, Rodríguez-Iturbe (1986) démontre que :

$$\chi = \frac{2}{\eta} \frac{1 + \mu_D}{\left\{2 + \frac{\beta(\mu_D - 1)}{\eta + \beta}\right\}}.$$
 (7.2)

L'échelle de fluctuation peut alors être utilisée pour exprimer un des trois paramètres présents dans l'équation (7.2) en fonction des deux autres. Pour réduire encore l'espace de minimisation d'une dimension, une technique consiste à réécrire l'équation (3.2) afin d'exprimer un des paramètres en fonction de la moyenne  $E(Y_i^h)$  et des autres paramètres restants (voir Cowpertwait, 1998), utilisant le fait que cette équation permet facilement d'isoler un paramètre, même lorsque la durée et l'intensité des cellules sont des variables aléatoires dépendantes.

#### 7.2.2 Méthode d'estimation des paramètres pour le cas particulier de la version de base du NSRPM

Pour le cas précis du NSRPM de base, Favre et al. (2004b) ont développé une approche basée sur les moments qui permet de réduire l'espace de minimisation à un espace à deux dimensions. Cette algorithme développé par Favre et al. (2004b) est très performant mais est cependant limité à un choix bien précis de moments. Les auteurs proposent l'utilisation des données de pluie horaires et journalières dans la procédure d'estimation, ce choix étant appuyé par Calenda & Napolitano (1999). Comme indiqué précédemment, il est important que les échelles de temps soient assez distinctes afin d'assurer la stabilité de l'algorithme d'optimisation de la fonction objectif. En utilisant les échelles horaires et journalières, nous nous assurons également que les propriétés statistiques de la pluie seront bien reproduites pour ces pas de temps, ce qui est de première importance lors des applications hydrologiques. Cependant, il va de soit que certaines applications nécessitent l'utilisation d'échelles de temps beaucoup plus fines (par exemple jusqu'à un pas de temps de cinq minutes en milieu urbain). Dans cette thèse, nous nous sommes limités à des applications sur des mesures de pluie à un pas de temps horaire, principalement afin d'obtenir plusieurs longues séries de données provenant de sites aux régimes pluviométriques variés (voir chapitre 8). Favre et al. (2004b) proposent d'inclure l'espérance horaire, la variance horaire et journalière et l'autocovariance horaire et journalière de décalage un (1) dans la fonction objectif (7.1).

Concernant la minimisation de la fonction objectif (7.1), des méthodes telles que la convergence quadratique de Powell (Press et al., 1986) ont été prônées par plusieurs auteurs (Velghe et al., 1994; Calenda & Napolitano, 1999) afin de résoudre ce problème non-linéaire d'optimisation. La principale difficulté réside dans le choix des paramètres initiaux. En effet, la convergence de l'algorithme est fortement dépendante de son point de départ. En outre, la minimisation est opérée dans un espace à cinq dimensions et les minima locaux sont ainsi difficiles à éviter. Ces limitations incitent, par conséquent, à choisir une approche alternative. Au lien de considérer l'optimisation des cinq paramètres du NSRPM, le système formé par les équations (3.2) à (3.4) peut être séparé. Considérons les variables V, W et  $Z_i, (i = 1, ..., 4)$  apparaissant dans les formules de la variance et de l'autocovariance :

$$\operatorname{Var}\left(Y_{i}^{h}\right) = VZ_{1}(\eta, \beta, h) + WZ_{2}(\eta, \beta, h), \tag{7.3}$$

$$\operatorname{Var}\left(Y_{i}^{h}\right) = VZ_{1}(\eta, \beta, h) + WZ_{2}(\eta, \beta, h), \tag{7.3}$$

$$\operatorname{Cov}\left(Y_{i}^{h}, Y_{i+k}^{h}\right) = VZ_{3}(\eta, \beta, h) + WZ_{4}(\eta, \beta, h). \tag{7.4}$$

Nous pouvons maintenant identifier V, W et  $Z_i$  en fonctions des paramètres du NSRPM, à partir des relations (3.3) et (3.4) :

$$V(\lambda, \mu_D, \mu_X) = 2\lambda \mu_D \,\mathcal{E}(X^2),\tag{7.5}$$

$$W(\lambda, \mu_D, \mu_X) = \lambda E(D^2 - D)\mu_X^2, \tag{7.6}$$

$$Z_1(\eta, \beta, h) = \frac{1}{\eta^3} (\eta h - 1 + e^{-\eta h}), \tag{7.7}$$

$$Z_2(\eta, \beta, h) = Z_1(\eta, \beta, h) \frac{\beta^2}{\beta^2 - \eta^2} - \frac{\beta h - 1 + e^{-\beta h}}{\beta(\beta^2 - \eta^2)},$$
(7.8)

$$Z_3(\eta, \beta, h) = \frac{1}{2\eta^3} (1 - e^{-\eta h})^2 e^{-\eta(k-1)h}, \tag{7.9}$$

$$Z_4(\eta, \beta, h) = \frac{Z_3(\eta, \beta, h)\beta^2}{(\beta^2 - \eta^2)} - \frac{(1 - e^{-\beta h})^2 e^{-\beta(k-1)h}}{2\beta(\beta^2 - \eta^2)}.$$
 (7.10)

Cela suggère de séparer l'ensemble des paramètres en deux sous-espaces  $\{\eta,\beta\}$  et  $\{\lambda,\mu_D,\mu_X\}$ . La procédure commence avec un choix arbitraire de  $(\eta,\beta)$  à partir duquel les valeurs des  $Z_i$ ,  $(i=1,\ldots,4)$  définies par les relations (7.5) à (7.10) peuvent être calculées. En utilisant les variances observées horaires (h=1) et journalières (h=24), la relation (7.3) conduit à deux équations à deux inconnues, qui peuvent être résolues pour V et W. La relation (7.4) permet ensuite d'obtenir des estimations des autocovariances horaires et journalières de décalage un. Les deux moments calculés sont ensuite comparés avec les moments empiriques par le biais d'une fonction objectif, similaire à la fonction  $\mathcal{O}$  de la relation (7.1). La fonction objectif est évaluée et utilisée afin de choisir de nouvelles valeurs de  $(\eta,\beta)$ . La procédure réduit ainsi le nombre de paramètres obtenus par minimisation. La méthode d'optimisation utilisée est basée sur le simplex de Nelder-Mead (Nelder & Mead, 1965). En combinant les valeurs optimales pour V et W avec l'expression de la moyenne horaire (équation 3.2), nous obtenons les trois paramètres restants :

$$\hat{\mu}_D = 2\frac{\widehat{W}}{\widehat{V}} + 1,$$

$$\hat{\mu}_X = \frac{\widehat{V}h}{4\mu_1 \hat{\eta}},$$

$$\hat{\lambda} = \frac{\mu_1 \hat{\eta}}{\hat{\mu}_C \hat{\mu}_X h},$$

où  $\mu_1$  dénote la moyenne issue de l'échantillon des pluies horaires. Le diagramme de la Figure 7.1 résume cet algorithme. Cette méthode a été appliquée sur des données de pluie d'origine variée et est beaucoup plus stable que la procédure classique de minimisation directe dans l'espace des paramètres. Ainsi, la solution obtenue avec cette méthode est contrainte (puisqu'elle implique une égalité stricte de le la relation (7.3) à une échelle horaire et journalière) et n'est pas théoriquement équivalente à la minimisation de la fonction objectif dans l'espace à cinq dimensions. Des approximations d'intervalles de confiance sur ces estimations ont de plus été données par Favre et al.

(2004b) en utilisant la méthode delta (Halperin & Mantel, 1963). Cependant, nous pouvons noter que cette approche (que ce soit l'estimation des paramètres ou l'approximation des intervalles de confiance) se base sur une séparation algébrique qui n'est possible que grâce à la forme particulière du système d'équations avec cet ensemble de moments. Avec l'ajout d'un lien entre la durée et l'intensité des cellules, cette séparation ne peut plus avoir lieu et cette méthode ne peut donc plus s'appliquer.

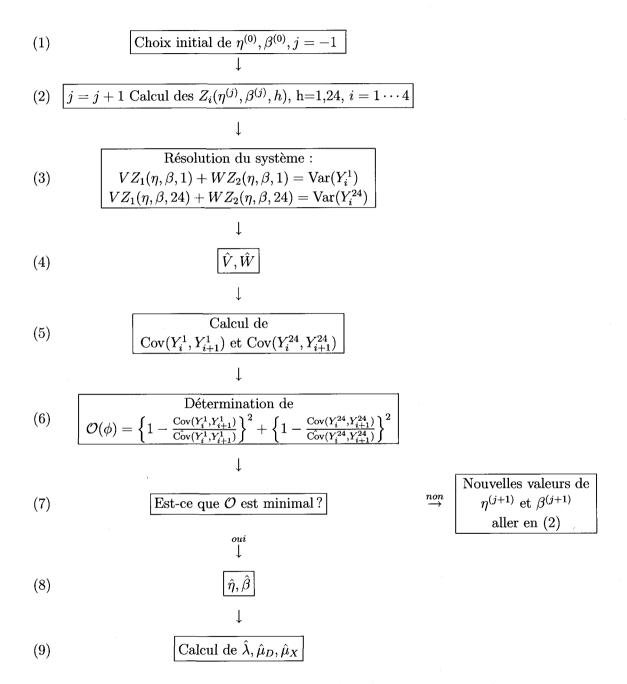


Fig. 7.1 – Diagramme résumant la procédure d'estimation en neuf étapes.

# 7.3 Méthode d'estimation adoptée

Les modèles sont généralement ajustés en utilisant une combinaison de moyennes, variances, autocorrélations et proportions d'intervalle sans pluie à des échelles variant de 1h à 24h. Pour un modèle à six paramètres, six expressions théoriques de moments sont suffisantes pour contraindre le problème d'optimisation. Cependant, tel que suggéré par Cowpertwait  $et\ al.\ (1996)$ , un ensemble plus grand de moments permet de forcer le modèle à s'adapter au minimum à ces propriétés. De plus, Chandler & Onof (2005) indiquent que les problèmes d'identifiabilité des paramètres diminuent lorsque le nombre de propriétés de la fonction objectif augmente. Nous choisissons d'inclure la moyenne horaire, la variance horaire, l'autocovariance horaire de décalage un (1), la variance journalière, l'autocovariance journalière de décalage un, le moment centré d'ordre trois à une échelle horaire et la probabilité de période sans pluie à une échelle horaire et journalière. Nous utilisons alors huit moments (p=8). Ce nombre de propriétés conduit généralement à des résultats satisfaisants (Chandler & Onof, 2005). Cowpertwait (1998) indique que l'introduction du moment d'ordre trois dans la méthode d'estimation permet une meilleure adéquation du modèle aux valeurs extrêmes.

La minimisation de  $\mathcal{O}$  n'est pas aisée et doit être réalisée numériquement. De nombreuses études ont montré que les estimations obtenues sont très sensibles au point de départ de la recherche (Calenda & Napolitano, 1999).

Plusieurs méthodes, telles que le Shuffled Complex Evolution, SCE-UA (voir Duan et al., 1992, 1993) ou les approches basées sur les algorithmes génétiques (Mitchell, 1996), ont été envisagées et brièvement testées lors de cette étape de minimisation. Ces méthodes ont l'avantage de parcourir un espace important en multipliant les directions de la recherche ou les points de départ de façon aléatoire. Elles sont très performantes pour des espaces de très grandes dimensions (c'est-à-dire supérieur à vingt). Cependant, nous nous sommes rapidement aperçus que, dans notre cas, le problème n'est pas tant de parcourir un espace très grand que d'explorer très finement un domaine de taille réduite en dehors duquel la fonction objectif est très aplatie. Ces méthodes ne donnaient finalement pas de meilleurs résultats et nous avons préféré développer un algorithme qui ciblait adéquatement l'espace de variation des points de départ de l'optimisation. La procédure d'optimisation que nous avons adoptée s'inspire de l'algorithme développé par Chandler & Onof (2005) et peut être résumée par les points suivants :

- 1. Une solution dans l'espace à cinq dimensions est obtenue grâce à la méthode généralisée des moments décrite à la section 7.2.2.
- 2. La solution obtenue à l'étape 1 est utilisée comme valeur de départ pour les paramètres  $\lambda, \beta, \eta, \mu_D, \alpha$ . La valeur de départ du sixième paramètre (par exemple  $\theta$  dans le cas des modèles utilisant une copule cubique) a été fixée de telle façon qu'il corresponde au NSRPM

dans sa version de base quand c'était possible (par exemple pour les copules cubiques qui comprennent le cas d'indépendance). Dans les autres cas, le choix est plus arbitraire. La procédure d'optimisation contraint l'espace des paramètres dans une région large mais réaliste :  $0.00001 < \lambda < 0.1 \ [h^{-1}]$ ;  $0.0001 < \beta < 0.99 \ [h^{-1}]$ ;  $0.01 < \eta < 200 \ [h^{-1}]$ ;  $1 < \mu_D < 500$ ;  $0.01 < \alpha < 1000 \ [mm]$ . La restriction des paramètres entre ces bornes nous garantit que les estimations obtenues ne sont pas physiquement invraisemblables. Cependant les bornes sont larges et l'expérience montre que les paramètres sont toujours contenus dans un petit sous-espace de cet espace des valeurs possibles (Cowpertwait, 1998). Un premier ensemble de ces paramètres est estimé en minimisant l'expression (7.1) à l'aide d'une méthode de programmation séquentielle quadratique (sequential quadratic programming, SQP), voir Fletcher (1987).

- 3. Les étapes 4 à 7 sont répétées plusieurs fois (par exemple cinq fois).
- 4. Nous effectuons 50 optimisations contraintes. Les points de départ sont générés par le biais de perturbations aléatoires tirées d'une distribution beta, l'espérance de cette loi étant constituée du meilleur vecteur de paramètres obtenu à l'itération précédente. Le support de cette distribution est défini par les bornes courantes (les bornes sont définies à l'étape 2 et évoluent à l'étape 6). Cette étape nous permet donc de tirer des points de départ aléatoires autour de la meilleure solution obtenue à ce stade.
- 5. Nous rejetons tous les vecteurs de paramètres pour lesquels la fonction objectif dépasse  $10O_{[50]}$ , où  $O_{[50]}$  est la valeur de fonction objectif, trouvées à l'étape 4, de rang 50.
- 6. Les bornes évoluent selon les meilleurs ensembles de paramètres trouvés jusque-là (à savoir ceux qui minimisent la fonction objectif). Si les meilleurs ensembles sont proches les uns des autres, les bornes sont restreintes. Si elles sont proches d'une borne existante, les bornes sont espacées. Ces choix sont arbitraires mais tiennent compte de l'échelle de valeurs des paramètres et sont motivés par le souhait d'aider l'algorithme à sortir du domaine d'attraction de minima locaux, en tenant compte du fait que l'adaptation de ces bornes ne s'effectue qu'un faible nombre de fois (cinq fois dans notre cas).
- 7. Seuls les 50 meilleurs ensembles de paramètres sont conservés. Le meilleur ensemble de paramètres parmi ces 50 ensembles devient le point de départ des futures optimisations.

Cet algorithme possède comme premier avantage de parcourir un important sous-espace de l'espace des paramètres, notamment grâce à l'évolution des bornes et des points de départ aléatoires. Sa seconde force consiste à éviter l'échec de l'algorithme de minimisation. Cela peut arriver lorsque de mauvais points de départ sont fournis, la fonction objectif devenant alors trop aplatie.

# Chapitre 8

# Application du NSRPM incluant un lien entre l'intensité et la durée des cellules

Dans ce chapitre, nous comparons les performances de différentes variantes du NSRPM (voir section 8.2) afin de savoir si le modèle est amélioré de façon significative quand la corrélation entre l'intensité et la durée des cellules pluvieuses est prise en compte. Nous disposons pour cela de plusieurs ensembles de données de pluies horaires provenant de régimes pluviométriques différents.

Dans cette thèse, nous nous intéressons uniquement à la modélisation stochastique des précipitations liquides, le type de modèle utilisé n'étant pas adéquat pour caractériser les précipitations solides. Par exemple, la modélisation de la neige implique nécessairement un couplage avec un modèle de températures. Cette limitation revêt une grande importance dans les régions nordiques, notamment au Québec, où les précipitations hivernales sont principalement constituées de neige. De plus, la succession de précipitations solides et liquides durant le même événement y est très courante selon l'évolution de la température. Dès lors, les interprétations des modèles de pluie pour les stations du Québec sont limitées pour les mois de novembre à avril.

# 8.1 Description des données disponibles

À partir des seules données de pluie (principalement à échelles mensuelle et annuelle), certains auteurs ont dressé un portrait global des régimes pluviométriques dans le monde. Pour identifier et classer les diverses régions pluviométriques du globe, on recourt habituellement aux précipitations (liquides et solides) mensuelles ou annuelles évaluées sur une longue période et leurs variations.

Une classification pluviométrique générale basée sur les données annuelles est fournie au Tableau 8.1. Elle distingue sept régimes différents, le plus humide étant le régime équatorial humide avec plus de 200 [cm] de précipitation annuelle moyenne, et le plus sec étant le régime polaire et arctique avec moins de 30 [cm] de précipitation annuelle moyenne.

Tab. 8.1 – Régimes pluviométriques du monde (Tiré de Champoux & Toutant, 1988)

(précipitations liquides et solides).

Régime	P. annuelles moyennes	Caractéristiques	Région typique
	> 900	à l'intérieur des conti-	bassin de
humide	> 200  cm	nents et sur les côtes	l'Amazone
subtropical	100-150 cm	à l'intérieur des conti-	pointe sud-est de
humide	100-150 cm	nents et sur les côtes	l'Amérique du Nord
subtropical	< 2° am	à l'intérieur des conti-	le sud du
sec	< 25  cm	nents et sur les côtes ouest	Maghreb
:tt	> 150 am	sur des zones côtières	côtes est de
intertropical	> 150 cm	étroites; humidité	l'Amérique centrale
continental	10 50 000	à l'intérieur des conti-	Plaines de l'ouest
$temp\'er\'e$	10 - 50  cm	nents; déserts ou steppes	de l'Amérique du Nord
$oc\'eanique$	> 100 am	sur les côtes ouest	la Colombie Britan-
$temp\'er\'e$	> 100 cm	des continents	nique, l'Europe
polaire	< 20 cm	nord du 60e parallèle;	le Grand Nord
et arctique	< 30 cm	grands déserts froids	canadien

La carte de la Figure 8.1 donne un aperçu global des climats de ce monde. En nous intéressant à l'Amérique du Nord et à l'Europe, nous possédons déjà un bon éventail des différents climats. En utilisant les séries de pluie provenant de la Suisse, de la Belgique, de la France, du Canada (province de Québec) et des États-Unis, presque tous les climats sont représentés, exceptés les climats équatorial et de mousson.

Le Tableau 8.2 donne les détails de ces séries de données qui ont été choisies pour leur longueur et la situation géographique de leur station. La Figure 8.2 présente les moyennes mensuelles des précipitations liquides pour les différents jeux de données sur lesquels nous appliquons les modèles de pluie. Nous nous sommes limités à 11 jeux de données pour des raisons de temps de calcul. Nous disposons cependant d'une grande variation de type de climats avec ces seules 11 stations, certaines étant très arrosées (par exemple Miami), d'autres recevant de la pluie assez régulièrement tout au long de l'année (à Zürich ou à Uccle) et enfin certaines avec des saisons très marquées (par exemple à Los Angeles ou à Val d'Or).

La suite de la section détaille les climats des différents pays considérés.

TAB. 8.2 – Description des séries de pluie horaire sur lesquelles sont appliqués les différents modèles

de Neyman-Scott.

Station	Pays	Coordonnées	Période couverte	Durée
Zürich	Suisse	47.22N\8.32E	01/01/1979 - 31/12/1999	21 ans
Uccle	Belgique	50.48N\4.20E	01/01/1968 - 31/12/1995	28 ans
Brest	France	48.23N\4.29O	01/01/1993 - 31/12/2005	13 ans
Lyon	France	45.45N\4.50E	29/07/1990 - 31/12/2005	15.4 ans
Marseille	France	43.17N\5.22E	01/01/1993 - 31/12/2005	13 ans
Trudeau	Canada	45.31N\73.39O	27/03/1943 - 01/11/1994	51.6 ans
Val d'Or	Canada	48.06N\77.47O	01/05/1961 - 01/11/1995	34.5 ans
Miami	USA	26.00N\80.00O	01/01/1949 - 31/12/1998	50 ans
Seattle	USA	48.00N\122.5O	01/01/1949 - 31/12/1998	50 ans
Los Angeles	USA	34.00N\117.5O	01/01/1949 - 31/12/1998	50 ans
Minneapolis	USA	46.00N\92.5O	01/01/1949 - 31/12/1998	50 ans

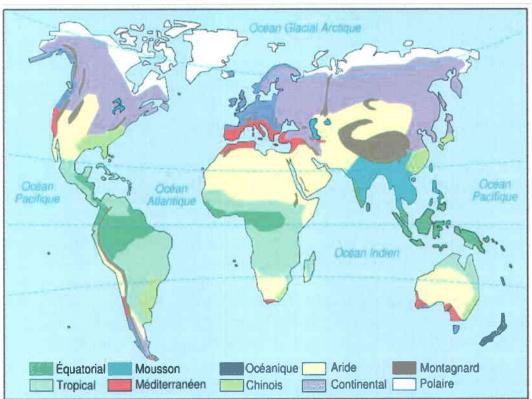


FIG. 8.1 - Carte des climats mondiaux (Source *Wikimedia Commons*, disponible à http://fr.wikipedia.org/wiki/Image:Climats\_dans\_le\_Monde.svg, dernière consultation le 7 août 2008).

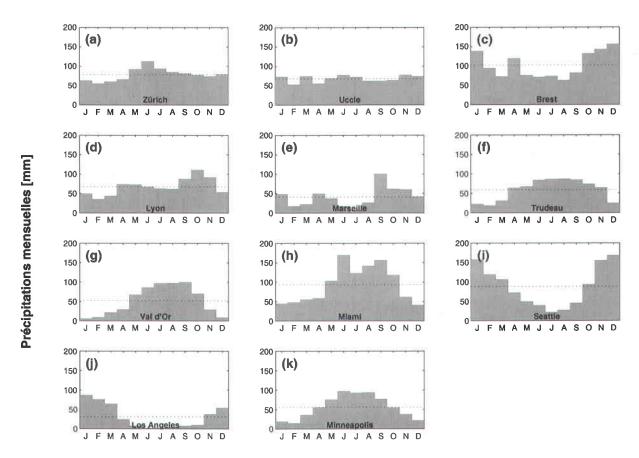


FIG. 8.2 – Moyennes mensuelles des précipitations liquides pour (a) Zürich : climat semi-continental (b) Uccle : climat océanique (c) Brest : climat hyper-océanique (d) Lyon : climat océanique de transition (e) Marseille : climat méditerranéen (f) Trudeau : climat continental humide (g) Val d'Or : climat continental (h) Miami : climat tropical (i) Seattle : climat tempéré de type océanique (j) Los Angeles : climat méditerranéen (k) Minneapolis : climat continental. La ligne horizontale en pointillé indique pour chaque station la moyenne annuelle divisée par 12 à titre de comparaison. Remarque : ces moyennes n'ont pas été calculées sur les mêmes périodes selon les mois considérés à cause des données manquantes.

#### **8.1.1** Suisse

Le réseau ANETZ (Automatisches meteorologisches Mess-und Beobachtungsnetz) comporte 72 stations de pluie automatisées. Les premières stations ont été mises en service à la fin des années 70 (20<sup>ième</sup> siècle). Une grande partie des stations ANETZ ont remplacé les stations conventionnelles. Le réseau ANETZ couvre toutes les régions et altitudes de la Suisse. Les mesures sont effectuées automatiquement toutes les 10 minutes et transmises à l'ordinateur central à Zürich. Pour certaines stations, des mesures sont effectuées par les observateurs trois à huit fois par jour. À l'aide de ce réseau, MétéoSuisse (l'institut national de Météorologie) a en particulier collecté des pluies horaires pour une vingtaine de stations du plateau Suisse, pour les années 1979 à 1999. Nous disposons pour notre part de dix séries de données de pluies horaires.

La Suisse est un pays montagneux et les stations disponibles sont toutes situées sur le plateau dans la partie la plus plate du pays (voir carte de la Figure 8.3). Le climat est de type semi-continental avec des influences océaniques et méditerranéennes dans une moindre mesure. La pluie est assez régulièrement répartie tout au long de l'année. Les variations du climat sont cependant importantes en raison des grandes différences d'altitude et d'exposition.



Fig. 8.3 – Carte de la Suisse (Source Hachette Multimédia).

#### 8.1.2 France et Belgique

Le service national de météorologie de France, MétéoFrance, nous a fourni dix séries de données de pluies horaires pour des périodes de 13 à 15 ans. Cette région de l'Europe est complétée par la série des données de pluie à Uccle en Belgique. La station de collecte de ces données est gérée par l'Institut Royal Météorologique (IRM). Les intensités de pluies horaires sont disponibles sur une période de 28 ans, couvrant 1968 à 1995.

La France, par sa latitude, est globalement située dans une zone tempérée, cependant différents types de climat peuvent être distingués (voir Figure 8.4 à gauche). Le climat est océanique à l'ouest, avec des pluies fréquentes toute l'année, continental à l'est avec une légère influence océanique. Au sud-est, le climat méditerranéen est caractérisé par des pluies irrégulières et peu nombreuses, ainsi qu'une sécheresse durant l'été. Les régions montagneuses (Alpes, Pyrénées, Massif central, Massifs des Vosges et Massif du Jura) possèdent aussi un climat particulier, les pluies étant plus nombreuses et les températures variant beaucoup en fonction de l'altitude. Le climat du centre de la France correspond à un climat océanique dégradé où l'influence océanique est encore perceptible, mais dû à l'éloignement de la côte, il se rapproche d'un climat continental avec des pluies plus faibles.

En raison de l'influence maritime, la Belgique présente un climat océanique dans la grande région de Bruxelles (voir carte de la Figure 8.4 à droite). Dans le sud du pays, le relief fait obstacle aux masses nuageuses qui se déversent sur les plateaux ardennais, les cumuls moyens de pluie y sont donc plus faibles.



FIG. 8.4 – Carte de la France et de son climat (à gauche, Source *Encarta*) et de la Belgique (à droite, Source *Hachette Multimédia*).

#### 8.1.3 États-Unis

Aux États-Unis, nous disposons des données de pluies horaires du Climate Prediction Center, qui fait partie du National Oceanic and Atmospheric Administration (NOAA)/Office of Oceanic and Atmospheric Research (OAR)/Earth System Research Laboratory (ESRL)- Physical Sciences Division (PSD), Boulder, Colorado, USA¹. Les stations sont disponibles sur une grille qui quadrille la région 140W-60W, 20N-60N avec des rectangles de 2.0° × 2.5° en utilisant un schéma de Cressman modifié. Les séries couvrent une période de 50 ans, de 1949 à 1998 et sont obtenues à partir de techniques d'interpolation et de réanalyses (une description détaillée est disponible à http://www.cdc.noaa.gov/cdc/data.cpc\_hour.html, dernière consultation le 7 août 2008).

Il serait très long de décrire complètement le climat de ce territoire immense. Dans notre cas, il est cependant important de souligner que l'étalement latitudinal et la disposition du relief nous procurent une grande diversité de climats. Du froid polaire du nord de l'Alaska au climat tropical de Floride, presque tous les climats sont représentés aux États-Unis. La carte de la Figure 8.5 représente les pluies totales annuelles moyennes aux États-Unis à partir des données reconstituées dont nous disposons. Nous distinguons très nettement les différents régimes de pluie, qui vont de régions très arrosées en Floride (>150 [cm]) à des régions très sèches dans le désert de l'Arizona par exemple (<20 [cm]).

#### 8.1.4 Canada

Au Canada, nous nous intéressons à la province de Québec pour des raisons de disponibilité de données. Les données de pluies horaires nous ont été fournies par Environnement Canada qui produit ce service via le Service météorologique du Canada. Nous disposons de 13 stations qui couvrent des périodes s'étendant de 26 à 52 ans.

Quatre types de climat principaux se distinguent dans la totalité du territoire québecois (une carte des précipitations totales moyennes annuelles au Québec est fournie à la Figure 8.6): continental humide, au sud du 50<sup>ième</sup> degré; subarctique, entre les 50<sup>ième</sup> et 58<sup>ième</sup> degrés; arctique, à l'extrême nord; maritime de l'Est aux îles de la Madeleine. Les conditions climatiques québécoises sont influencées par la position septentrionale de la province et par une double exposition aux eaux froides de la baie d'Hudson et aux courants océaniques froids, le long de la côte du Labrador. Nous nous intéressons ici à la région du Québec la plus au Sud et la plus habitée qui correspond donc au premier type de climat. Les saisons les plus marquées restent l'été et l'hiver. Les étés sont chauds et souvent très humides. Les hivers sont froids, plutôt longs et neigeux. Nous notons d'importantes

<sup>&</sup>lt;sup>1</sup>disponible à http://www.cdc.noaa.gov/cdc/data.cpc\_hour.html, dernière consultation le 7 août 2008

précipitations (solides et liquides) annuelles, dont un tiers environ sous forme de neige.

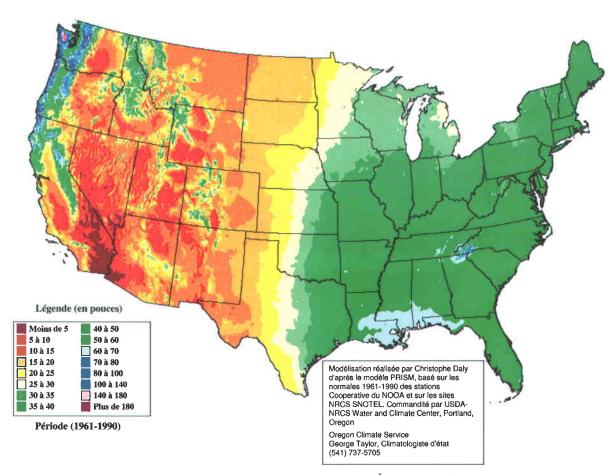


FIG. 8.5 – Carte des pluies totales moyennes annuelles aux États-Unis (Source *NOAA*, disponible sur leur site à http://www.wrcc.dri.edu/pcpn/us\_precip.gif, dernière consultation le 7 août 2008).

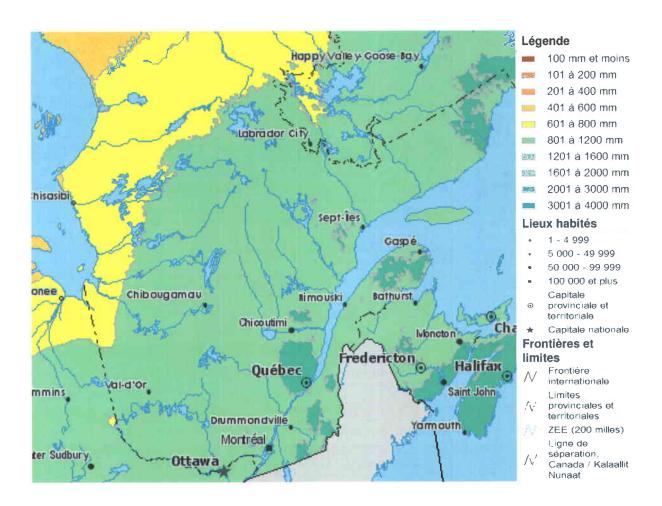


FIG. 8.6 – Carte des précipitations (solides et liquides) totales moyennes annuelles [mm] au Québec (Source Ressources naturelles Canada), disponible à http://atlas.nrcan.gc.ca/site/francais/maps/environment/climate/precipitation, dernière consultation le 7 août 2008.

#### 8.2 Modèles considérés

Nous allons appliquer le NSRPM incluant une dépendance intra-cellule sur différentes séries de pluie en utilisant les familles de copules décrites au Tableau 4.1. Pour éviter de surcharger les figures et les tableaux, nous montrons les résultats pour seulement quatre familles de copules, à savoir FGM, Frank cubique (Fcub), Sym2 et Asym3. Il doit être noté que plusieurs familles peuvent donner des résultats similaires en raison de leur proximité dans l'espace des fonctions pour certaines valeurs de  $\theta$ . Par exemple, les copules Frank cubique, Sym2 et Sym3 peuvent être équivalentes puisqu'elles atteignent toutes les trois la valeur minimale du rho de Spearman pour une copule cubique (voir Figure 4.2). Les familles Frank cubique (Fcub), Sym2 et Asym3 sont sélectionnées en raison de résultats intéressants sur des essais préliminaires. L'application de la copule FGM nous permet d'analyser la différence de performance entre cette famille de copule quadratique et les familles de copules cubiques. Les distributions marginales pour la durée et l'intensité des cellules sont supposées exponentielles. Nous obtenons ainsi une distribution bivariée à trois paramètres pour caractériser le comportement de l'intensité et de la durée des cellules. Nous rappelons que la copule FGM avec des marges exponentielles correspond à la distribution bivariée de Gumbel Type-II. Ce modèle particulier nous permet donc de comparer les performances des différents modèles avec l'approche développée par Kim & Kavvas (2006). Nous notons cependant plusieurs erreurs de calcul dans l'article de Kim & Kavvas (2006) et des résultats décevants lors de l'application (sans doute liés aux expressions erronées des moments). Nous ne comparerons donc pas nos résultats à ceux illustrés dans cet article. Afin de comparer équitablement les modèles avec et sans dépendance, il est nécessaire de considérer des modèles ayant le même nombre de paramètres. Nous appliquons dans ce but une distribution gamma à deux paramètres pour l'intensité des cellules dans le cas de l'indépendance (densité de probabilité  $f(x) = \alpha^{\gamma} x^{\gamma-1} e^{-\alpha x} / \Gamma(\gamma), x \ge 0, \gamma > 0$ ). Nous examinons également le cas d'une distribution de Pareto de densité de probabilité  $f(x) = \gamma \alpha^{\gamma}/(\alpha + x)^{(\gamma + 1)}, x \ge 1$  $0, \gamma > 0$  pour l'intensité des cellules, où  $\alpha$  est le paramètre d'échelle et  $\gamma$  le paramètre de forme. Ces deux distributions possèdent l'avantage de pouvoir posséder une queue inférieure lourde. De plus, la distribution de Pareto comporte une queue supérieure lourde, ce qui intervient directement sur les propriétés des valeurs extrêmes du modèle. Même si nous souhaitons comparer des modèles comportant le même nombre de paramètres, il nous a semblé intéressant d'étudier également le cas classique d'une distribution exponentielle simple pour l'intensité des cellules. Ce modèle correspond donc à la version de base présentée à la section 3.2 et comporte cinq paramètres.

Le dernier type de modèle considéré est le modèle 'Intensité-Durée Dépendantes' (Dependent Depth-Duration Model) développé par Kakou (1997). Ce modèle conserve la même forme d'occurrence pour les averses et les cellules. L'intensité des cellules X possède une forme exponentielle. La structure de dépendance entre l'intensité et la durée des cellules est introduite par le biais d'une intensité moyenne de cellule dépendante de L, c'est-à-dire que  $\mathrm{E}(X|L)=g(L)$ . La distribution exponentielle de paramètre  $\eta$  pour la durée des cellules L est conservée. Deux versions pour la

fonction g(L) sont considérées :  $g(L) = f e^{-cL}$  et  $g(L) = f L e^{-cL}$ , où f et c sont des scalaires positifs. Par la suite, les modèles DD1 et DD2 désigneront respectivement ces deux versions du modèle 'Intensité-Durée Dépendantes'. Dans le modèle DD1, les variables X et L sont toujours négativement corrélées, avec comme fonction de corrélation (rho de Pearson) :

$$\frac{-\zeta}{\zeta+1}\sqrt{\frac{2\zeta+1}{2\zeta^2+2\zeta+1}},$$

où  $\zeta = c/\eta$  (Kakou, 1997). Avec le modèle DD2, la fonction de corrélation (rho de Pearson) entre l'intensité et la durée des cellules pluvieuses s'exprime comme suit :

$$\frac{1-\zeta}{\zeta+1}\sqrt{\frac{(2\zeta+1)^3}{4(\zeta+1)^4-(2\zeta+1)^3}}.$$

Pour ce dernier type de modèle, nous calculons donc l'ensemble des moments impliqués dans la fonction objectif (voir chapitre 7). En particulier, le moment centré d'ordre trois n'était pas disponible dans la littérature pour ces modèles et a donc fait l'objet d'un travail important.

Pour l'ensemble de ces modèles, nous rappelons que la durée des cellules suit une distribution exponentielle et que le nombre de cellules suit une loi géométrique.

# 8.3 Estimation des paramètres et fonction objectif

Nous nous sommes heurtés à plusieurs difficultés lors de l'estimation des paramètres pour certaines stations de la France, du Canada et des États-Unis. Nous avons mentionné au chapitre 7 que nous utilisons la moyenne horaire, les variance horaire et journalière et les autocovariance horaire et journalière de décalage un (1) comme premier ensemble de moments dans la procédure d'estimation des paramètres. Cependant, il apparaît que l'autocorrélation journalière échantillonnée prend parfois des valeurs très faibles (par exemple pour les mois d'été dans certaines régions du Québec, voir Figure 8.7).

La procédure d'estimation décrite dans la sous-section 7.2.2 devient instable lorsqu'elle est appliquée avec une autocorrélation journalière très faible. Si l'autocovariance journalière de décalage un varie alors que les autres moments sont fixes, les estimations obtenues ne sont satisfaisantes que si l'autocovariance journalière dépasse un certain seuil. Nous avons appliqué cette procédure aux données de pluie de la station Trudeau pour le mois de juillet. La Figure 8.8 représente les valeurs de la fonction objectif  $\mathcal{O}(\phi)$  et des estimations des paramètres du NSRPM  $(\hat{\lambda}, \hat{\eta}, \hat{\beta}, \hat{\mu}_D, \hat{\alpha})$  en fonction de l'autocovariance journalière variant de 0 à 10 [mm/h] (l'espérance horaire, les variance horaire et journalière et les autocovariances horaire et journalière de décalage un conservant

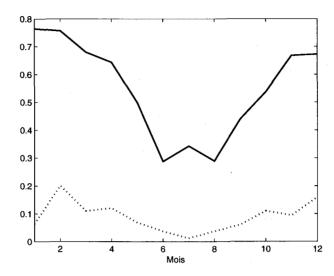


Fig. 8.7 – Évolution intra-annuelle (mensuelle) de l'autocorrélation horaire (trait plein) et de l'autocorrélation journalière (trait pointillé) empiriques de décalage un à la station Trudeau.

leurs valeurs calculées sur l'échantillon). Les estimations deviennent cohérentes pour des autocovariances journalières dépassant 2 [mm/h]<sup>2</sup>. Le même phénomène s'est reproduit plusieurs fois dans nos applications et se retrouve par exemple à Brest (voir Figures 8.9 et 8.10). Certaines limites du modèle de Neyman-Scott apparaissent ici puisqu'il n'est pas en mesure de reproduire des structures de pluie où l'autocorrélation journalière décalage un de l'intensité pluvieuse est trop faible. Il est par exemple impossible au NSRPM de reproduire des autocovariances négatives (l'expression théorique de l'autocovariance est le résultat de la sommation de quantités positives, voir chapitre 6, équation 6.5), ce qui pourrait être observé en réalité. Il est également possible que les difficultés rencontrées soient liées à la forme de la fonction objectif 7.1 qui évalue les quotients entre les moments théoriques et échantillonnaux. Le rapport entre deux quantités très faibles pourrait alors exploser et rendre la procédure d'optimisation instable. Ce problème particulier peut être résolu en remplaçant la fonction 7.1 par une somme quadratique des différences entre les moments théoriques et échantillonnaux. Il s'agit alors d'associer un poids à chaque terme de la somme afin de tenir compte des différences d'échelle des divers moments statistiques considérés.

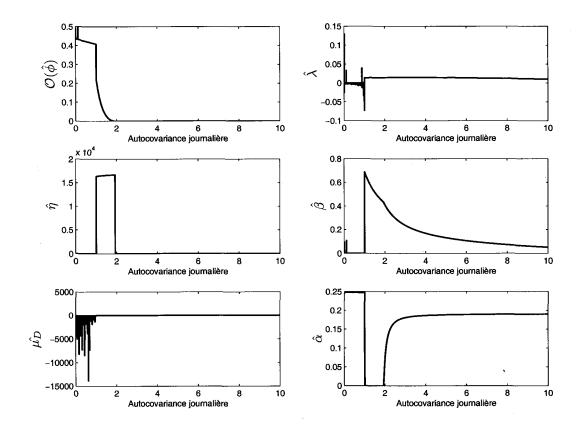


Fig. 8.8 – Évolution de la fonction objectif et des estimations des paramètres du NSRPM dans sa version de base pour le mois de juillet à la station Trudeau en fonction de l'autocovariance journalière de décalage un.

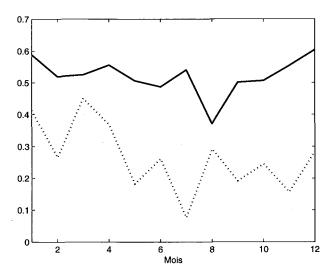


FIG. 8.9 – Évolution intra-annuelle (mensuelle) de l'autocorrélation horaire (trait plein) et journalière (trait pointillé) empiriques de décalage un à la station de Brest.

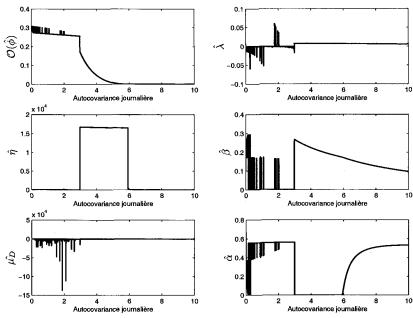


Fig. 8.10 – Évolution de la fonction objectif et des estimations des paramètres du NSRPM dans sa version de base pour le mois de juillet à la station de Brest en fonction de l'autocovariance journalière de décalage un.

Lorsque la méthode d'estimation décrite dans la sous-section 7.2.2 ne fonctionnait pas, nous avons adapté la procédure explicitée au chapitre 7 en remplaçant l'autocovariance journalière de décalage un par l'autocovariance semi-journalière (12h) de décalage un (de façon à augmenter la valeur de l'autocorrélation). Cette légère différence dans le choix des moments semble avoir des conséquences positives sur les estimations des paramètres. En effet, les résultats sont beaucoup plus cohérents pour les mois et les stations concernés par ce changement dans la méthode d'estimation. Ces mois et ces stations sont décrits dans le Tableau 8.3 et représentent une proportion de 21.21%. Seules les stations de Zürich, Uccle et de Miami n'ont pas nécessité de modifications dans la procédure d'estimation des paramètres. Nous devons noter sur ce point que, même s'il aurait été peut-être plus logique de considérer pour toutes les applications l'autocovariance semi-journalière de décalage un (afin par exemple d'utiliser le même ensemble de moments pour toute l'année), nous avons préféré conserver, autant que possible, le choix des échelles horaires et journalières motivé au chapitre 7. Il semble de plus que ce choix ait peu d'implications sur les résultats pour les mois où l'autocovariance journalière de décalage un est suffisamment élevée.

Tab. 8.3 – Mois et stations où la méthode alternative d'estimation des paramètres a été appliquée.

G1 - 1:						M	ois					
Station	J	$\mathbf{F}$	M	A	M	J	J	A	$\mathbf{S}$	Ο	N	D
Zürich												
Uccle												
Brest							X					
Lyon	×						×	×				,
Marseille	×	X			×				×			
Trudeau					×	×	X	X	×			
$Val\ d'Or$		×	×					×			×	
Miami												
Seattle							X					
$Los\ Angeles$			×					X		X		
Minneapolis	×	×	×	×				×			×	×

Les valeurs de la fonction objectif donnent une mesure globale de l'adéquation du modèle. Cela indique donc que les modèles qui minimisent cette fonction surpassent les autres modèles pour ce critère (c'est-à-dire l'adéquation du modèle aux propriétés statistiques considérées dans la fonction objective) et pour cette application. Il est important de souligner que plusieurs modèles appliqués ici incluent comme cas particulier le modèle de base où l'intensité des cellules est distribuée exponentiellement. Il s'agit des modèles gamma, FGM, Fcub et DD1 (les copules cubiques Sym2 et Asym2 n'incluant pas l'indépendance, voir Tableau 4.1). Cela signifie donc que, avec l'ajout d'un paramètre supplémentaire, ces modèles devraient obtenir une valeur plus faible de fonction objectif, ou au pire, égale. Dans le cas contraire, cela indique que l'algorithme d'optimisation a rencontré des problèmes numériques importants (voir par exemple le modèle Fcub à Los Angeles pour le mois d'août au Tableau C.14).

Pour chaque station, nous représentons les valeurs minimales de fonctions objectif atteintes par les différents modèles dans les Figures 8.11-8.13 et en annexe dans les Figures C.1-C.8. Afin de faciliter la lecture de ces figures, nous ne dessinons pas les courbes associées aux modèles FGM, DD1 et DD2, la performance de ces modèles étant généralement inférieure aux autres modèles. Plus précisément, la copule FGM semble être limitée par le degré de dépendance qu'elle peut reproduire puisque ce modèle ne se distingue pas particulièrement. Il est aussi important de souligner que nous avons rencontré de grandes difficultés à estimer les paramètres des modèles DD1 et DD2. Dans la plupart des applications, le modèle DD1 est équivalent au modèle à cinq paramètres avec la loi exponentielle simple pour l'intensité des cellules et le modèle DD2 donne de très mauvaises valeurs de fonctions objectif, ce qui semble montrer que cette distribution conditionnelle pour l'intensité des cellules n'est pas adaptée à la structure des données de pluie considérées. De manière générale, trois modèles semblent présenter de meilleures performances par leur adéquation aux moments statistiques utilisés dans la méthode d'estimation des paramètres : les distributions gamma et de Pareto pour l'intensité des cellules et la famille de copules Sym2 pour lier l'intensité et la durée des cellules. Le NSRPM avec la distribution de Pareto donne par exemple de bons résultats tout au long de l'année sur les données de pluie de Zürich (voir Figure 8.11), de Uccle (voir Figure 8.12), de Miami (voir Figure C.5), pour le début de l'année à Seattle (voir Figure C.6) et surpasse nettement les autres modèles sur certains mois, par exemple pour le mois de Juillet à Marseille (voir Figure C.3). Nous noterons cependant que la distribution de Pareto montre parfois ses limites (voir Figure C.7) et que ces bonnes performances sont souvent surpassées par d'autres modèles. La distribution gamma pour l'intensité des cellules présente aussi des résultats intéressants, par exemple dans le cas de l'application à Lyon (voir Figure C.2) et surpasse tous les autres modèles pour le mois de mars à Zürich (voir Figure 8.11). Concernant les modèles utilisant des familles de copules cubiques, la famille Sym2 se distingue particulièrement. Elle minimise la fonction objectif par rapport aux autres modèles de façon nette à de nombreuses reprises :

pour les mois de mars, avril, mai, juin, août, octobre, novembre et décembre à Zürich (voir Figure 8.11);

- pour les mois de janvier, février, mars, avril, juillet, septembre et novembre à Uccle (voir Figure 8.12);
- pour les mois de avril, août et juin à Val d'Or (voir Figure 8.13);
- pour les mois de septembre et novembre à Marseille (voir Figure C.3);
- pour les mois de juin, août, octobre et décembre à Seattle (voir Figure C.6);
- pour les mois de mai et de décembre à Minneapolis (voir Figure C.8).

Nous notons également que la copule Asym3 donne des résultats intéressants dans certains cas, par exemple pour le mois de janvier à Zürich (voir Figure 8.11) ou le mois de décembre à Uccle (voir Figure 8.12).

Les Tableaux 8.4 à 8.9 et en annexe les Tableaux C.1 à C.16 reportent les estimations des paramètres pour les mois de janvier et août, pour chaque station. Nous pouvons ainsi examiner les valeurs des paramètres qui minimisent la fonction objectif pour les mois de janvier et août, pour toutes les stations. Par exemple, nous pouvons constater que lorsque la modèle avec la copule Sym2 surpasse les autres modèles au regard de la fonction objectif, le tau de Kendall atteint souvent -0.4, ce qui est sa valeur minimale possible pour cette structure de dépendance (voir Tableau 4.2). Cette remarque se vérifie pour les mois de janvier et août à Zürich (voir Tableaux 8.4 et 8.5), pour le mois de janvier à Uccle (voir Tableau 8.6) ou pour le mois d'août à Val d'Or (voir Tableau 8.9). Les estimations des paramètres  $\lambda$  et  $\beta$  (relatifs au processus ponctuel proprement dit) sont très stables. Les valeurs de ces paramètres varient peu d'un modèle à l'autre. Ce constat est cohérent avec le fait que les différentes variantes du NSRPM considérées ici ne s'intéressent qu'aux propriétés des cellules (durée et intensité) et pas à la distribution des origines de cellules. Le paramètre  $\lambda$ semble être directement lié à l'intensité moyenne de pluie, puisque  $\hat{\lambda}$  est faible lorsque la station reçoit peu de pluie (en toute logique puisque les expressions de l'espérance de l'intensité de pluie sont linéaires en  $\lambda$ ). Les valeurs rencontrées pour  $\hat{\beta}$  sont en général inférieures à  $\hat{\eta}$  et se situent en-dessous de 0.5.  $\hat{\eta}$  varie classiquement dans l'intervalle [0.5,5] et  $\hat{\mu}_D$  dans l'intervalle [1,15]. Il arrive cependant quelquefois que les estimations dépassent largement les bornes supérieures de ces intervalles. Pour plusieurs applications, les valeurs obtenues pour  $\hat{\eta}$  et  $\hat{\mu}_D$  sont très élevées (voir par exemple les Tableaux C.5, C.11, C.14 et C.15). Cela pose de nombreux problèmes numériques (voir par exemple les estimations pour la famille de copules Frank cubique (Fcub) dans le Tableau C.14). Nous pouvons de plus nous questionner sur l'interprétation de ces paramètres lorsque par exemple le nombre moyen de cellules par averse dépasse la centaine, comme à Los Angeles pour le mois d'août (voir Tableau C.14). Pour ces applications, aucun modèle ne se distingue vraiment et la fonction objectif n'atteint pas des valeurs très faibles. Un début d'explication se situe dans le fait que lorsque le modèle de Neyman-Scott produit un grand nombre de cellules très courtes à chaque averse, les propriétés des distributions inhérentes à l'intensité des cellules se confondent dans l'amas de cellules (voir chapitre 5). Dans ce cas, il est aussi probable que la façon dont le modèle de Neyman-Scott distribue les origines de cellules n'est pas adapté.

Le NSRPM avec une distribution gamma pour l'intensité des cellules montre de très bons résultats lorsque son paramètre de forme  $\gamma$  est inférieur à 1. Par exemple, pour le mois d'août à Brest (voir Tableau C.2), nous obtenons  $\hat{\gamma} = 0.13$ . Cela semble indiquer que, dans certains cas, la forme des averses nécessite une distribution avec une dépendance de queue inférieure très forte, que ne permet pas de reproduire une simple distribution exponentielle. Le modèle avec une distribution gamma surpasse les autres modèles dans ces applications (par exemple pour le mois d'août à Lyon, voir Tableau C.4 ou pour le mois de janvier à Miami, voir Tableau C.9) et fait alors ressortir les limites de la distribution de Pareto ou des distributions construites à partir de copules cubiques. Au contraire, nous observons parfois les limites de la distribution gamma face aux autres modèles. Pour le mois de janvier à Brest, le modèle avec la distribution gamma est presque équivalent au modèle de base avec la distribution exponentielle (voir Tableau C.1,  $\hat{\gamma} = 0.96$ pour le modèle gamma). Rappelons que la distribution gamma est équivalente à une distribution exponentielle lorsque son paramètre de forme  $\gamma$  vaut 1. Pour cette application, c'est la copule Sym2 qui minimise la fonction objectif avec une dépendance négative ( $\tau = -0.25$ ). Dans certains cas, l'introduction de la dépendance améliore donc nettement l'adéquation du NSRPM à la structure de la pluie, le plus souvent pour des dépendances négatives (pour le mois d'août à Seattle, voir Tableau C.12 ou pour le mois de janvier à Minneapolis, voir Tableau C.15).

Le Tableau 8.10 reporte les meilleurs modèles (au sens de la minimisation de la fonction objectif) pour chaque site et chaque mois. Nous notons ainsi qu'il est difficile de discerner d'éventuels "patrons" mensuels. Ce résultat n'est pas surprenant et confirme la difficulté dans ce genre de modélisation à obtenir des conclusions systématiques.

TAB. 8.4 – Estimations des paramètres du NSRPM  $\hat{\Phi}$  à Zürich, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0105	1.30	0.068	13.10	1.27		0	8.19e-004
Ind. Gamma	0.0106	1.33	0.068	11.83	1.41	1.25	0	4.58e-004
Ind. Pareto	0.0105	1.27	0.068	13.59	19.91	27.78	0	1.16e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0106	1.12	0.067	11.83	1.17	-0.50	-0.11	4.75e-004
Fcub	0.0106	1.08	0.067	11.65	1.14	-0.23	-0.15	4.72e-004
Sym2	0.0109	0.67	0.061	7.85	0.89	-0.61	-0.40	2.34e-004
Asym3	0.0107	0.92	0.065	9.86	1.05	-0.93	-0.20	1.30e-004
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0105	1.30	0.068	13.10	0.78	0.00	-0.00	8.19e-004
DD2	0.0105	0.92	0.094	21.14	8.79	2.55	-0.34	3.11e-002

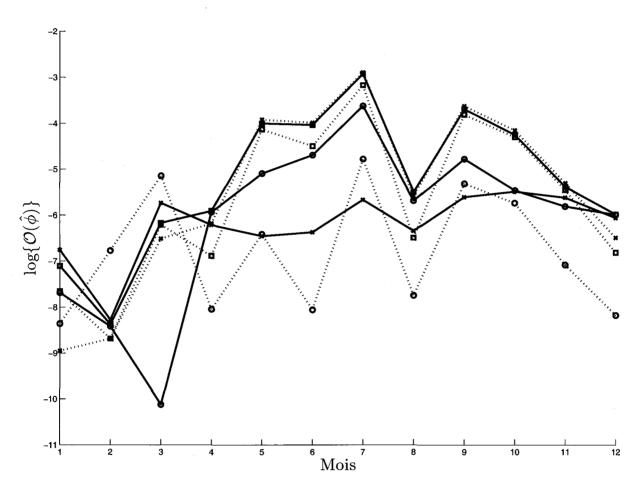


FIG. 8.11 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Zürich. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. 8.5 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Zürich, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0156	2.07	0.129	4.16	0.28		0	4.09e-003
Ind. Gamma	0.0156	2.08	0.134	4.89	0.26	0.76	0	3.38e-003
Ind. Pareto	0.0157	2.04	0.137	4.99	24.19	9.41	0	1.76e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0157	1.77	0.126	3.77	0.25	-0.59	-0.13	3.08e-003
Fcub	0.0158	1.44	0.122	3.31	0.22	-0.45	-0.30	1.51e-003
Sym2	0.0160	1.41	0.125	3.38	0.20	-0.52	-0.40	4.32e-004
Asym3	0.0157	1.78	0.124	3.64	0.25	-0.69	-0.12	3.76e-003
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0156	2.07	0.129	4.16	3.52	0.00	-0.00	4.09e-003
DD2	0.0150	0.47	0.126	3.37	21.18	1.61	-0.37	3.31e-003

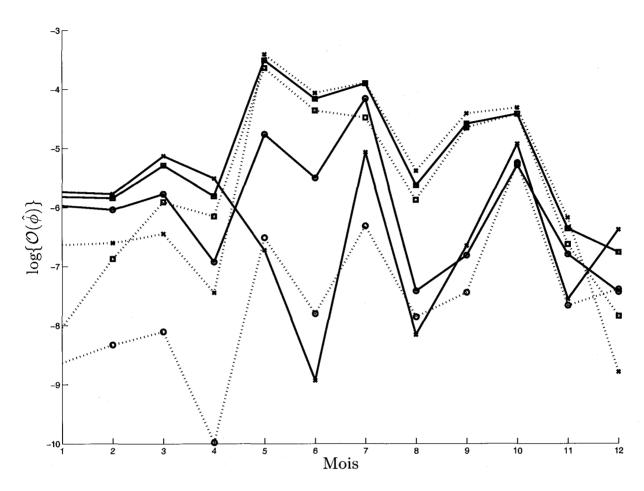


Fig. 8.12 – Logarithme de la fonction objectif pour les différents modèles testés à la station d'Uccle. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. 8.6 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Uccle, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0256	1.60	0.102	6.75	1.15		0	2.97e-003
Ind. Gamma	0.0259	1.61	0.100	6.14	1.25	1.20	0	2.55e-003
Ind. Pareto	0.0255	1.58	0.103	7.08	21.48	27.25	0	3.23e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	$\tau$	$\mathcal{O}(\hat{\phi})$
FGM	0.0262	1.18	0.097	5.31	0.97	-0.85	-0.19	1.33e-003
Fcub	0.0268	0.88	0.090	4.23	0.83	-0.54	-0.37	3.26e-004
Sym2	0.0269	0.85	0.089	4.13	0.80	-0.65	-0.40	1.78e-004
Asym3	0.0263	1.13	0.095	5.07	0.95	-0.93	-0.20	1.32e-003
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	ĉ	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0256	1.60	0.102	6.75	0.87	0.00	-0.00	2.97e-003
DD2	0.0248	1.15	0.148	11.99	19.91	4.00	-0.37	6.21e-002

TAB. 8.7 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Uccle, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0192	2.83	0.110	2.82	0.23		0	3.61e-003
Ind. Gamma	0.0192	2.89	0.128	3.81	0.19	0.62	0	6.01e-004
Ind. Pareto	0.0193	2.82	0.126	3.40	30.58	9.51	0	2.87e-004
Famille de copules	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0192	2.69	0.110	2.76	0.22	-0.29	-0.07	3.27e-003
Fcub	0.0192	2.52	0.110	2.70	0.20	-0.23	-0.15	2.82e-003
Sym2	0.0192	3.39	0.133	4.02	0.25	1.38	-0.14	3.87e-004
Asym3	0.0192	2.72	0.107	2.66	0.21	-0.42	-0.05	4.61e-003
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0192	2.83	0.110	2.82	4.37	0.00	-0.00	3.61e-003
DD2	0.0190	3.40	0.132	4.32	11.39	0.88	0.43	1.41e-003

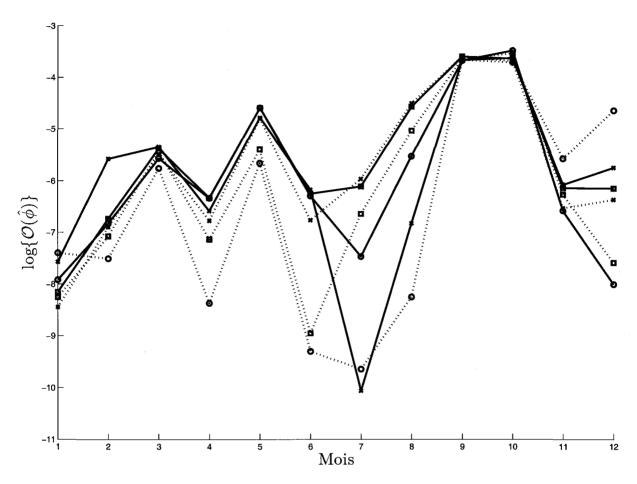


FIG. 8.13 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Val d'Or. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. 8.8 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Val d'Or, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	τ	$\mathcal{O}(\hat{\phi})$
Ind. Exponential	0.0009	0.38	0.011	4.73	1.39		0	2.87e-004
Ind. Gamma	0.0009	0.39	0.012	4.34	1.44	1.08	0	2.74e-004
Ind. Pareto	0.0010	0.38	0.012	4.53	24.53	36.62	0	5.16e-004
Famille de copules	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0009	0.37	0.011	4.66	1.37	-0.09	-0.02	2.62e-004
Fcub	0.0009	0.37	0.010	4.76	1.37	-0.04	-0.03	2.60e-004
Sym2	0.0008	0.46	0.008	7.32	1.46	1.56	-0.10	6.15e-004
Asym3	0.0009	0.37	0.011	4.55	1.32	-0.31	-0.02	2.15e-004
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0009	0.38	0.011	4.73	0.72	0.00	-0.00	2.87e-004
DD2	0.0507	0.07	0.001	1.00	10.42	1.96	-0.24	1.99e + 000

Tab. 8.9 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Val d'Or, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0198	3.25	0.213	4.15	0.20		0	1.04e-002
Ind. Gamma	0.0194	4.16	0.262	9.42	0.13	0.37	0	3.99e-003
Ind. Pareto	0.0194	5.93	0.297	10.18	14.25	4.58	0	1.09e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0198	2.99	0.212	3.97	0.18	-0.44	-0.10	9.30e-003
Fcub	0.0199	2.51	0.214	3.58	0.15	-0.45	-0.31	6.51e-003
Sym2	0.0199	4.17	0.264	6.53	0.16	0.08	-0.34	2.61e-004
Asym3	0.0198	2.97	0.206	3.78	0.18	-0.55	-0.08	1.11e-002
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0198	3.25	0.213	4.15	5.08	0.00	-0.00	1.04e-002
DD2	0.0191	7.69	0.286	14.98	15.74	0.47	0.55	2.51e-003

TAB. 8.10 – Tableau récapitulatif des meilleurs modèles lorsque le critère de minimisation de la fonction objectif  $\mathcal{O}(\dot{\phi})$  est considéré. Les 'X' indiquent les cas où les modèles sont presente tous équivalente.

ré. Les 'X' indiquent les cas où les modèles sont presque tous équivalents.	diquent l	es cas où	i les mo	dèles sor	nt presq	ue tous	équivale	nts.				
	ſ	ĹΉ	M	A	$\mathbb{Z}$	J	ſ	А	$\infty$	0	Z	D
Zürich	Asym3 Sym2	×	gam.	Sym2	Par. Sym2	Sym2	Par.	Sym2	Par.	Sym2	$\mathrm{Sym2}$	$\mathrm{Sym2}$
Uccle	Sym2	Sym2	$\mathrm{Sym2}$	Sym2	Par. Sym2	Par.	Sym2	Par. Sym2	Sym2	X	Sym2 Par.	Asym3
Val d'Or	×	Sym2	×	Sym2	×	Sym2 Fcub	Sym2 Par.	$\mathrm{Sym2}$	×	X	X	gam. Fcub
Brest	×	Sym2	×	Fcub	×	gam.	gam.	gam.	gam.	X	X	Sym2 gam.
Lyon	X	gam.	Fcub Par.	$\mathrm{Sym2}$	Sym2 Par.	gam. Par.	Par.	gam.	Sym2	Par.	Sym2	gam.
Marseille	×	Fcub	Par.	×	×	×	Par. Sym2	Sym2 Fcub	Sym2	gam.	$\mathrm{Sym2}$	gam.
Trudeau	Par.	gam.	gam.	gam.	Sym2	Par.	gam.	Par.	Par.	Sym2	Sym2 Fcub	gam.
Miami	gam.	Par.	Par.	Par.	gam. Par.	Par.	gam. Sym2	Par. gam.	Par.	Par.	Par. gam.	gam.
Seattle	Par.	Par.	X	Par.	Par.	X	X	Sym2	X	X	X	Sym2
Los Ang.	X	X	X	X	X	X	$\begin{array}{c} \text{Fcub} \\ \text{Sym2} \end{array}$	X	X	X	X	Par.
Minneap.	Sym2	X	Par.	$_{ m Sym2}$	$\mathrm{Sym2}$	Sym2 Par.	Par.	Par.	Sym2 Sym2	Par.	Par.	$_{ m Sym2}$

# 8.4 Dépendance

Dans cette section, nous nous intéressons aux degrés de dépendance que reproduisent les modèles avec copules cubiques. Dans les Figures 8.14 et 8.15, nous représentons le tau de Kendall entre la durée et l'intensité des cellules avec la copule Sym2, pour les différentes applications (en considérant deux ensembles de stations, c'est-à-dire l'Europe et l'Amérique du Nord). Il est à noter que le tau de Kendall est compris dans l'intervalle [-0.4006,0.3533] pour cette copule (voir Tableau 4.2). La borne inférieure de cet intervalle est atteinte de nombreuses fois, par exemple pour les mois de janvier, avril, août et décembre à Zürich (voir Figure 8.14) ou les mois de mai, juillet, août et novembre à Trudeau (voir Figure 8.15). Le lien entre l'intensité et la durée des cellules est très nettement négatif dans de nombreux cas, même si ce lien n'est mesuré qu'indirectement (voir chapitre 5). À Seattle et à Trudeau, le degré de dépendance est très proche de -0.4 tout au long de l'année (excepté en Février à Trudeau), suggérant que le lien pourrait être parfois plus fortement négatif s'il n'était pas limité par la structure des copules cubiques. Bien qu'il est en règle générale négatif, comme nous pouvons en avoir l'intuition (une cellule est d'autant plus courte que son intensité est élevée), le degré de dépendance est parfois nettement positif. Par exemple, pour le mois d'août à Marseille (voir Figure 8.14), le tau de Kendall s'élève à 0.3433. De plus, pour cette application et ce mois particulier, le NSRPM avec la copule Sym2 surpasse les autres modèles au niveau de la fonction objectif (voir Tableau C.6). Cela suggère que cette dépendance positive n'est pas due à un problème dans la procédure d'estimation des paramètres. Cependant, le fait que le tau de Kendall puisse varier énormément d'un mois à l'autre comme à Marseille semble peu vraisemblable en pratique. Nous pouvons alors nous demander ce qui cause cette irrégularité. Notre opinion est que cette variation est liée aux difficultés à estimer les paramètres. À Marseille, nous pouvons par exemple constater que les valeurs de fonction objectif varient beaucoup d'un mois à l'autre, tout comme le tau de Kendall. Il suffit pour s'en persuader de comparer les figures 8.14 et C.3.

Il n'est pas possible d'appliquer un test d'hypothèse afin de vérifier la significativité des valeurs obtenues pour le tau de Kendall puisqu'elles ne sont pas observées directement. Cependant, nous remarquons sur les Figures 8.14 et 8.15 que le tau de Kendall se situe très rarement autour de 0, justifiant la pertinence de la prise en compte de la dépendance entre la durée et l'intensité des cellules.

Dans l'Annexe C.2, les Figures C.9 à C.14 représentent le tau de Kendall correspondant aux copules Fcub, Asym3 et FGM pour les différentes applications. Les remarques précédentes restent valides, mais elles sont cependant moins visibles pour les copules Asym3 et FGM puisque le degré de dépendance que peuvent reproduire ces deux structures de dépendance est limité (voir Tableau 4.2).

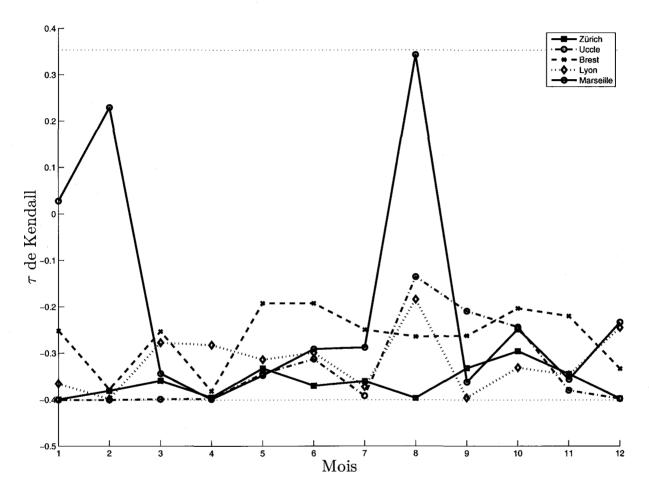


Fig. 8.14 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule Sym2 pour les stations de Zürich, Uccle, Brest, Lyon et Marseille. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

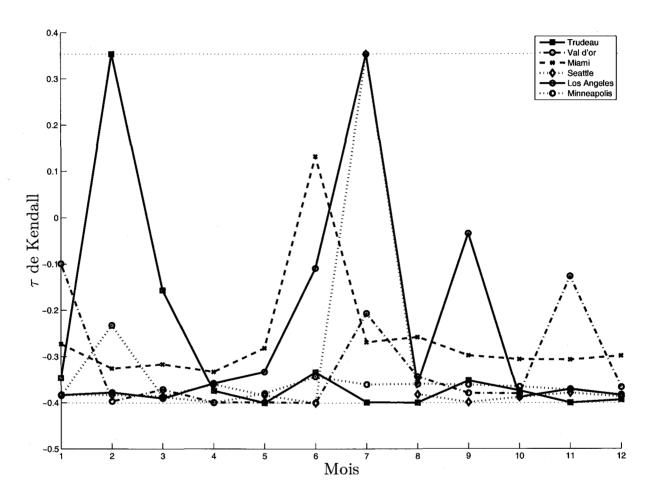


Fig. 8.15 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule Sym2 pour les stations de Trudeau, Val d'Or, Miami, Seattle, Los Angeles et Minneapolis. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

#### 8.5 Extrêmes

Comme indiqué par Onof et al. (2000), une des faiblesses majeures du NSRPM dans sa version de base est la reproduction des valeurs extrêmes. Une tendance générale présente dans la littérature de la modélisation stochastique de la pluie est la sous-estimation des valeurs extrêmes à une échelle horaire (Entekhabi et al., 1989). La capacité du modèle à reproduire les extrêmes constitue pourtant un aspect prépondérant du point de vue de la gestion des risques. Afin de visualiser les performances des différents modèles concernant ce critère, nous simulons 100 séries de pluie d'une durée de 100 ans, pour chaque modèle et chaque station. Nous calculons ensuite les maxima annuels à des échelles horaires et journalières. Qu'il s'agisse des données observées ou simulées, nous considérons les maxima sur les données horaires ou journalières brutes, c'est-à-dire que nous n'appliquons pas de fenêtres mobiles. Ces dernières sont souvent utilisées afin de construire des courbes Intensité-Durée-Fréquence (IDF) et exigent des données enregistrées à un pas de temps très inférieur à l'heure. Dans notre cas, les séries générées sont découpées en tranches d'une heure (ou 24 heures). Les cumuls de pluie sont ensuite estimés sur chacune de ces durées. Finalement, le maximum pour chaque année est retenu, à un pas de temps horaire et journalier.

Les Figures 8.18 à 8.21 (et les Figures C.15 à C.26 dans l'Annexe C.3) permettent de comparer les maxima observés et simulés pour les différents modèles considérés et pour chaque station. L'abscisse donne la période de retour en années sur une échelle de Gumbel. Nous utilisons essentiellement cette échelle afin de faciliter la représentation des maxima et nous ne prétendons pas ici que la distribution de Gumbel modélise adéquatement les valeurs extrêmes de pluie. En réalité, de récentes analyses (voir par exemple Koutsoyiannis, 2004; Neppel et al., 2006) tendent à démontrer au contraire que la distribution de Gumbel n'est en général pas appropriée pour modéliser les maxima annuels de pluie. Les quantiles empiriques des maxima sont obtenus avec la formule de Hazen  $\hat{F}[Y_{[k]}] = (k-0.5)/n$  où  $Y_{[k]}$  est le maximum annuel à une échelle horaire ou journalière correspondant au rang k. Un intervalle de confiance empirique à 95% des maxima simulés est obtenu à partir des 100 simulations de 100 ans. Nous pouvons observer l'importance du type et du degré de dépendance en examinant les intervalles de couverture empiriques.

De manière générale, nous retrouvons une sous-estimation des extrêmes annuels à une échelle horaire pour les périodes de retour de 1 à 10 ans (voir par exemple les Figures 8.19, 8.21 ou C.18). Les valeurs extrêmes à une échelle journalière sont bien restituées par les différents modèles dans la plupart des cas (voir Figures 8.16, 8.18, 8.20 et C.29). Ils sont cependant sous-estimés pour les périodes de retour de 1 à 10 ans à Val d'Or (voir Figure C.21) et à Los Angeles (voir Figure C.27) et pour les périodes de retour de 5 à 100 ans à Miami (voir Figure C.23). À Lyon, les maxima annuels journaliers observés ne semblent pas suivre une loi de Gumbel. En effet à la Figure C.17, les maxima observés ne décrivent pas une droite mais plutôt une tendance logarithmique. Or il semble que le NSRPM reproduise presque systématiquement des maxima selon une distribution de

type Gumbel. Dans ce cas particulier, les maxima journaliers observés sont ainsi surestimés pour les périodes de retour les plus fortes.

La distribution de Pareto pour l'intensité des cellules reproduit mieux les extrêmes que les autres modèles dans plusieurs cas, par exemple à Miami pour les échelles horaire et journalière (voir Figures C.23 et C.24) ou à Brest (voir Figure C.16) et Minneapolis (voir Figure C.30) à une échelle horaire. De plus, l'intervalle de confiance empirique obtenu à partir des 100 répétitions de simulations de 100 ans est plus large avec la distribution de Pareto, ce qui permet de mieux couvrir les extrêmes observés. Les modèles avec dépendance entre l'intensité et la durée des cellules donnent de bons résultats à Zürich avec les copules cubiques Fcub et Sym2 (voir les Figures 8.16 et 8.17). Le modèle DD1 livre des résultats très décevants à de nombreuses reprises, par exemple à Trudeau et Val d'Or à une échelle journalière (voir Figures 8.20 et C.21). À Seattle et à Los Angeles, les maxima simulés à partir du modèle DD1 sont complètement aberrants et surestiment très fortement les maxima observés (voir les Figures C.25 à C.28).

Nous avons également jugé pertinent de procéder à un plan de simulation de séries de pluies annuelles en sélectionnant les meilleurs modèles mois par mois. La station choisie est Lyon car, à cette station, les meilleurs modèles sont très divers d'un mois à l'autre. La construction de ce "mégamodèle" utilise ainsi les modèles indiqués au Tableau 8.10. Les Figures 8.22 et 8.23 permettent donc de comparer les maxima horaires et journaliers observés et simulés pour ce "méga-modèle" à Lyon (à comparer avec les Figures C.17 et C.18 obtenues pour chaque modèle séparément). Nous notons qu'il semble y avoir une amélioration intéressante de la reproduction des extrêmes horaires pour les périodes de retour les plus faibles, ce qui indiquerait la pertinence d'utiliser les modèles qui montrent les meilleures performances pour chaque mois, et non pas un unique modèle pour toute l'année. Généralement, nous avons vu (voir Tableau 8.10) qu'un modèle peut donner de très bons résultats certains mois, et de beaucoup moins bons pour d'autres (à titre d'exemple, le modèle sans dépendance avec distribution de Pareto à Los Angeles s'avère le plus performant en décembre mais le pire en avril et en mai, voir Figure C.7).

8.5 Extrêmes 91

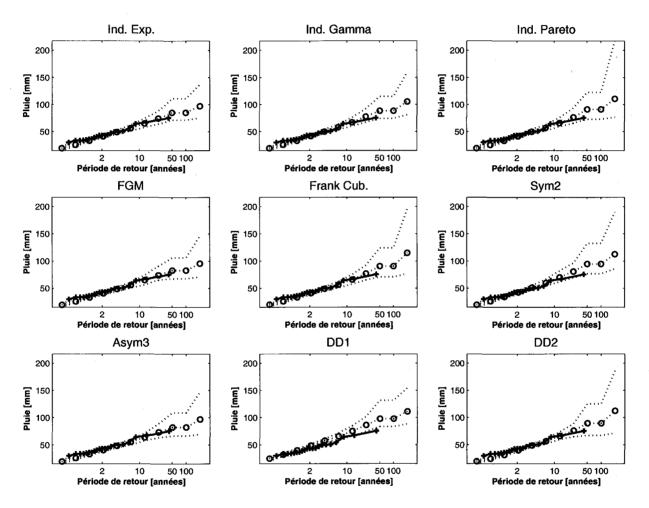


FIG. 8.16 – Maxima annuels journaliers à Zürich. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

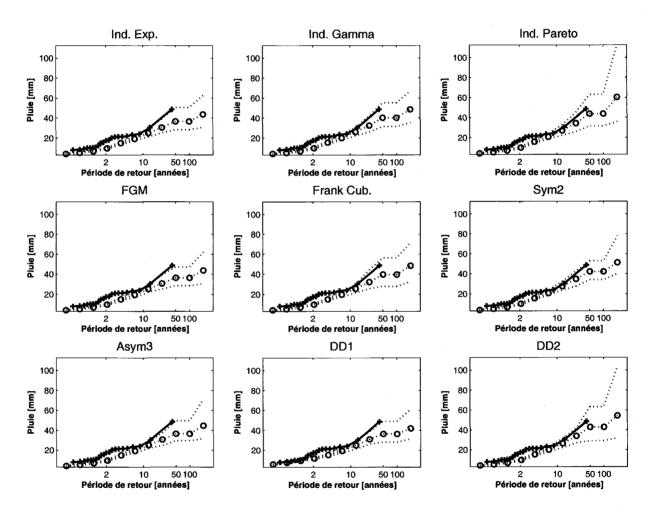


FIG. 8.17 — Maxima annuels horaires à Zürich. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

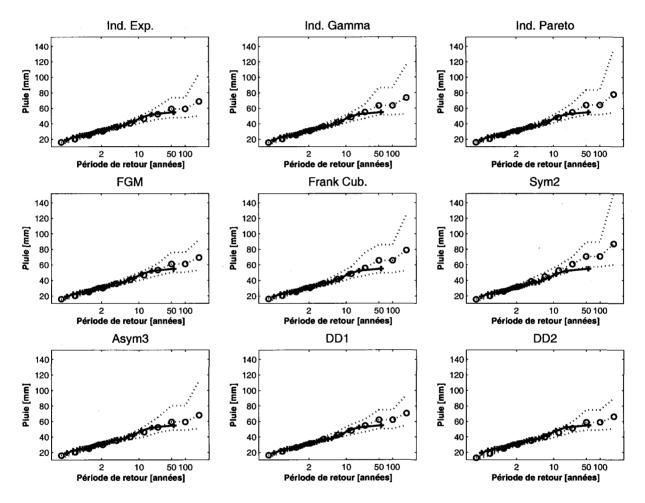


Fig. 8.18 – Maxima annuels journaliers à Uccle. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

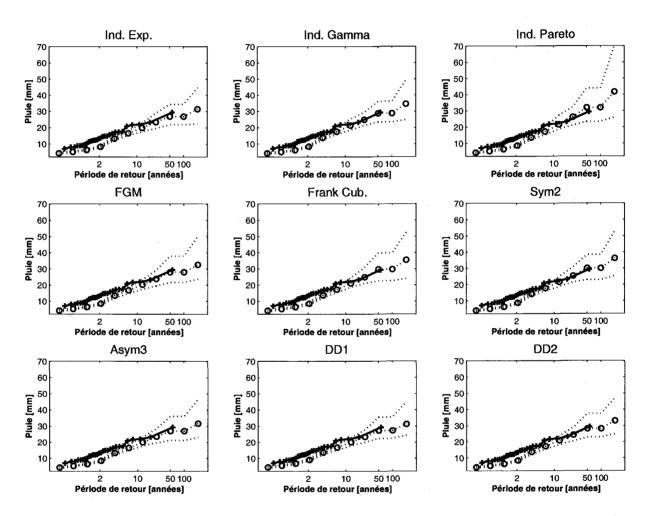


FIG. 8.19 — Maxima annuels horaires à Uccle. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

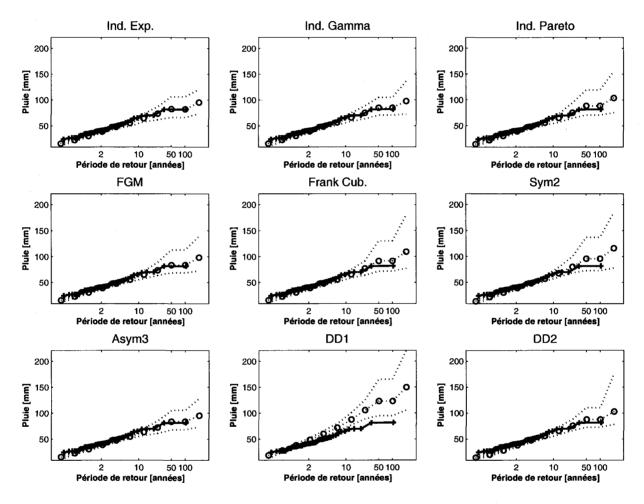


FIG. 8.20 – Maxima annuels journaliers à Trudeau. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

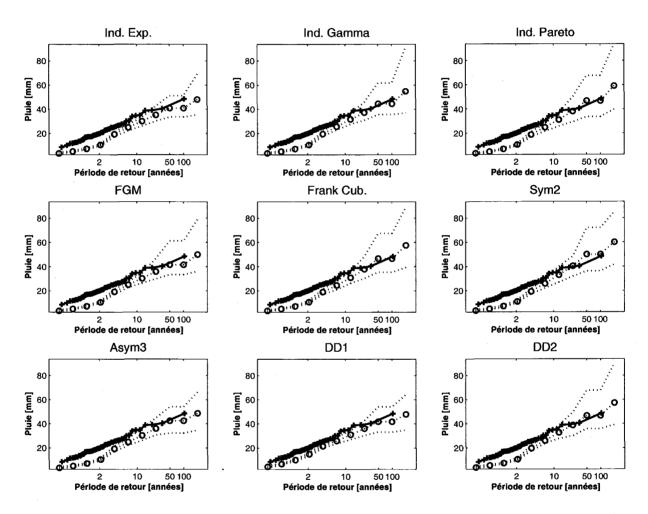


FIG. 8.21 – Maxima annuels horaires à Trudeau. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

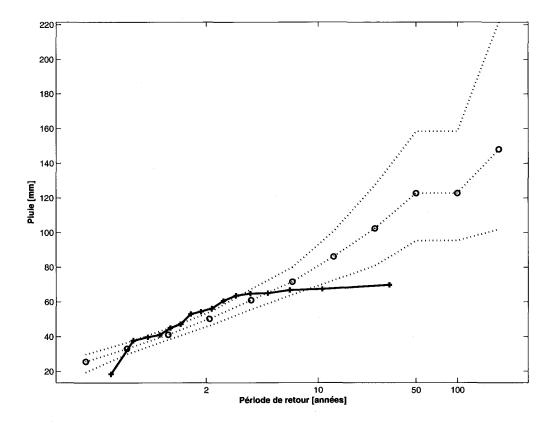


Fig. 8.22 – Maxima annuels journaliers à Lyon avec le "méga-modèle". L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

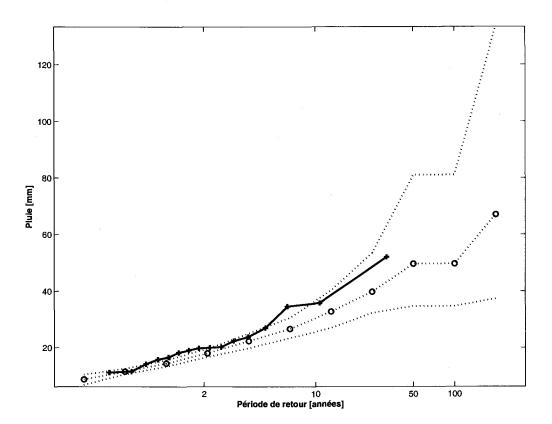


FIG. 8.23 — Maxima annuels horaires à Lyon avec le "méga-modèle". L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

8.6 Conclusion 99

#### 8.6 Conclusion

Les nombreuses applications réalisées ont permis de dresser un portrait global des performances des diverses variantes du modèle de Neyman-Scott pour différents types de climats. Le grand nombre de stations étudiées a permis de mettre en évidence certaines difficultés dans la procédure d'estimation. Lorsque l'autocorrélation journalière de décalage un est trop faible, la méthode généralisée des moments est instable. Cela nous a conduit à remplacer l'autocovariance journalière de décalage un par l'autocovariance semi-journalière (12h) de décalage un dans les propriétés statistiques prises en compte dans la méthode des moments.

Nous retiendrons de ces applications que le modèle avec la copule cubique Sym2 donne de très bons résultats sur l'adéquation du modèle aux propriétés statistiques utilisées dans l'estimation des paramètres. Pour ce critère, les modèles avec les distributions de Pareto et gamma peuvent également donner de bons résultats dans certains cas. La copule cubique semble permettre de reproduire les caractéristiques temporelles de la pluie grâce à sa capacité à modéliser des dépendances négatives. Nous nous apercevons d'ailleurs que le tau de Kendall entre l'intensité de la durée des cellules (lié à l'estimation du paramètre de la copule Sym2) est très souvent négatif.

Les valeurs extrêmes sont généralement bien reproduites par la plupart des modèles à une échelle journalière. À une échelle horaire, nous retrouvons parfois une sous-estimation des valeurs extrêmes, qui peut être importante pour les périodes de retour de 1 à 10 ans. Il semblerait alors qu'en simulant les séries annuelles à partir des meilleurs modèles pour chaque mois, les valeurs extrêmes soient mieux restituées pour ces périodes de retour.

Une hypothèse sous-jacente à la formulation du modèle avec l'intensité et la durée des cellules dépendantes était que ce lien devait permettre une meilleure reproduction des extrêmes. En effet, nous pouvions nous attendre à augmenter la capacité du modèle à produire des événements intenses puisque des cellules de courte durée accordent dans ce cas des probabilités plus fortes aux intensités élevées. D'après les résultats de ce chapitre, nous concluons qu'il n'en est pas systématiquement ainsi. Cependant, nous constatons également que la dépendance est presque systématiquement négative, et que le degré de dépendance atteint souvent la borne inférieure de l'intervalle de dépendance couvert par les copules cubiques. Nous pouvons alors supposer que des formes de dépendance atteignant des degrés de dépendance plus faibles encore auraient permis d'améliorer la performance de ces modèles, et notamment une meilleure reproduction des valeurs extrêmes en serait sans doute amélioré.

Le modèle avec la distribution de Pareto sur l'intensité des cellules permet de réduire la sousestimation des valeurs extrêmes pour les périodes de retour les plus fortes dans plusieurs cas, vraisemblablement grâce à sa queue supérieure de distribution lourde. Fort de cette constatation, une idée intéressante serait de combiner la distribution de Pareto avec une copule permettant de reproduire des degrés de dépendance très fortement négatifs. Cependant, même si la théorie des copules permet directement de construire une telle distribution (voir chapitre 4), l'application d'un tel modèle est difficile à envisager puisque le calcul des moments se heurterait alors à de grandes difficultés.

Il est important de souligner qu'il est également possible que les difficultés à reproduire des valeurs extrêmes soient dues à des propriétés extérieures à celles étudiées dans cette thèse. Par exemple, la distribution de la durée des cellules a été souvent modifiée dans la littérature et semble avoir un impact sur la reproduction des valeurs extrêmes (voir Onof & Wheater, 1994; Chandler & Onof, 2005). Comme nous l'avons vu, ce sont le plus souvent des contraintes calculatoires qui limitent les généralisations, mais nous pourrions également penser que le processus ponctuel pourrait être modifié de manière à améliorer les propriétés du modèle. Par exemple si le processus de Poisson était généralisé à un processus de renouvellement (voir chapitre 2), cela permettrait peut-être à des averses de se succéder plus fréquemment, et de donner lieu à des événements pluvieux très intenses. Nous pourrions supposer qu'une loi gamma à la place d'une loi exponentielle pour régir la distance entre les origines des cellules et celles des averses correspondantes permettrait l'occurrence de nombreuses cellules très proches tout en autorisant qu'elles soient dispersées dans d'autres cas (afin de générer des averses stratiformes). Les idées sont nombreuses mais leurs applications demeurent limitées par les expressions mathématiques à résoudre qu'elles impliquent (voir Onof, 2003).

# Chapitre 9

# Un modèle de Neyman-Scott non-stationnaire

#### 9.1 Introduction

Une des hypothèses fondamentales sur laquelle se base le modèle de Neyman-Scott concerne la stationnarité des séries de pluie générées. Cependant, la grande majorité des scientifiques intéressés par le climat semblent maintenant s'accorder sur l'existence d'un changement climatique exhibant des tendances nettes dans les séries de température et de précipitation. Ces hypothèses sont fortement appuyées par le quatrième rapport d'évaluation (AR4) du Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC, 2007). Il semble donc naturel de reconsidérer le postulat de stationnarité et d'explorer les possibilités d'introduire une non-stationnarité dans les modèles de pluie.

Les modifications des propriétés de la pluie induites par les changements climatiques sont très complexes et de nombreux efforts doivent encore être mis en oeuvre pour bien les comprendre, le phénomène pluvieux étant très variable spatialement et temporellement. Selon Trenberth et al. (2007, voir troisième chapitre du groupe de travail I du quatrième rapport d'évaluation (AR4) du GIEC¹), les observations montrent que les changements interviennent en quantité, intensité, fréquence et type de précipitations. Des tendances à long terme très prononcées de 1900 à 2005 ont été observées pour les précipitations totales dans plusieurs endroits : significativement plus humide dans le Nord-Est et le Sud de l'Amérique, l'Europe du Nord et dans l'Asie centrale et

<sup>&</sup>lt;sup>1</sup>disponible à http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter3.pdf, dernière consultation le 7 août 2008

du Nord, mais significativement plus sec dans le Sahel, le sud de l'Afrique, la Méditerranée et le sud de l'Asie. Les précipitations tombent maintenant en plus grande quantité sous forme de pluie que sous forme de neige dans les régions nordiques. Au cours des dernières années, de nombreux articles ont montré des tendances dans les séries de précipitations (voir par exemple Karl et al., 1995), principalement à un pas de temps journalier. En Amérique du Nord, nous pouvons entre autre citer les contributions de Karl & Knight (1998) aux États-Unis et Vincent & Mekis (2006) au Canada.

De nombreux articles tendent à démontrer également la présence de tendances dans les extrêmes journaliers (mensuels ou annuels). Plusieurs études ont montré que ces tendances existaient à partir de différents jeux de données dans le monde. Le lecteur pourra consulter Arnbjerg-Nielsen (2006) au Danemark, Hundecha & Bárdossy (2005) en Allemagne, Schmidli & Frei (2005) en Suisse, Iwashima & Yamamoto (1993) au Japon, Zhai et al. (2005) en Chine, New et al. (2006) en Afrique, Plummer et al. (1999) en Australie et Nouvelle-Zélande et enfin Vincent & Mekis (2006) au Canada. Si l'on considère des pas de temps inférieurs, Vaes et al. (2002) analysent une série de pluie de 100 ans (la résolution temporelle variant du jour à dix minutes) en Belgique, sans détecter de tendances significatives. Rauch & DeToffol (2006) analysent des séries jusqu'à une résolution de cinq minutes en Autriche, au Danemark et en Allemagne. Blanckaert & Willems (2006) détectent des cycles dans l'occurrence des valeurs extrêmes à partir de l'analyse d'une série de pluie à un pas de temps de dix minutes. La nature souvent chaotique des précipitations complexifie l'observation de ces changements (GIEC, 2007). Si les multiples dimensions qui influencent ce phénomène (région, saison, effet inter-annuel) sont prises en compte, des schémas clairs dans les dynamiques des précipitations sont difficiles à exhiber. La principale difficulté de l'analyse des tendances pour les extrêmes demeure toutefois les séries courtes et la grande variabilité de ces événements.

Concernant l'aspect prévisionnel pour le 21<sup>ième</sup> siècle, le résultat central des modèles climatiques est que l'augmentation croissante des concentrations de gaz à effet de serre et d'aérosols, ainsi que des changements dans l'aménagement du territoire affectent directement les processus d'interaction dans l'atmosphère et les dynamiques de pluie (voir par exemple Avissar & Liu, 1996). Avec l'augmentation de la température, il est plus probable que les précipitations tombent sous forme de pluie plutôt que de neige dans certaines régions (par exemple au Québec), spécialement en automne et au printemps au début et à la fin de la saison neigeuse, et dans certaines zones où les températures sont proches de zéro (GIEC, 2007). Cependant, les impacts appréhendés sur les précipitations vont bien au-delà de la simple augmentation de la quantité de pluie en hiver. Mailhot et al. (2007) présentent un compte-rendu de l'état de la situation au Québec en ce qui concerne l'analyse des tendances, les impacts des changements climatiques sur les précipitations et la modélisation climatique.

À notre connaissance, il y a très peu d'articles qui concernent le domaine de la modélisation non-stationnaire en temps de la pluie. Cela souligne sans doute la difficulté d'une telle approche et la complexité à valider de tels modèles. Nous notons cependant la contribution de Sansó & Guenni (2000) qui tiennent compte de la non-stationnarité à la fois dans le temps et dans l'espace. Le modèle proposé représente l'intensité de pluie observée par une variable latente tronquée qui suit un modèle linéaire dynamique multivarié. Cette structure leur permet de tenir compte de la saisonnalité, des tendances et des corrélations spatiales à travers les paramètres du modèle. La validation de ce modèle est ensuite exécutée sur des données agrégées à un pas de temps de dix jours. Cet article ne met cependant pas l'emphase sur la reproduction des tendances observées dans les propriétés de la pluie et n'exploite pas le potentiel théorique du modèle.

Nous proposons dans ce chapitre de nouveaux développements des modèles de pluie basés sur les processus ponctuels qui considèrent une fonction de non-stationnarité pour l'arrivée des averses et qui supposent que les averses seront plus fréquentes dans le futur. Cette hypothèse est en accord avec l'augmentation des précipitations annuelles moyennes. Afin de formuler correctement ce modèle de Neyman-Scott non-stationnaire (NSRPM-NS), la théorie de base de Cox & Isham (1980) est reconsidérée et les moments sont calculés jusqu'au troisième ordre dans le cas général puis avec une fonction simple pour les arrivées des averses. Cependant, ces développements mettent en lumière d'autres difficultés dans la façon dont les moments théoriques peuvent être liés aux moments empiriques. Si on considère que les propriétés statistiques varient dans le temps, le calcul des moments empiriques n'est plus direct. Dans les sections 9.4 et 9.5 de ce chapitre, une méthode d'estimation des paramètres est proposée et discutée.

# 9.2 Le processus de Poisson non-stationnaire avec amas de cellules

Les modèles de Poisson avec amas de cellules (cluster) que nous avons présentés jusqu'à présent supposent que les origines d'averses se produisent selon un processus de Poisson de taux  $\lambda$  constant. Nous désirons ici introduire une fonction  $\varphi(t)$  telle que le taux d'arrivée des averses soit dépendant du temps. Nous devons souligner ici qu'il s'agit d'une façon très particulière d'introduire une non-stationnarité, puisque nous modifions la fréquence des averses mais pas leur intensité. Une autre façon de modifier le NSRPM serait de considérer une dépendance temporelle dans la génération des intensités de cellules. Cependant, la nécessité d'obtenir des expressions analytiques (afin d'estimer les paramètres par la méthode des moments, voir section 9.5) nous contraint à retenir des hypothèses très simples pour la forme de la non-stationnarité. Nous modifions donc les relations (2.3) et (2.4) de telle façon que le processus de Poisson de taux  $\varphi(t)$  est maintenant défini de la manière suivante, pour tout t, quand  $\delta \to 0^+$ :

$$\Pr\{N(t, t+\delta) = 1 | \mathcal{H}_t\} = \varphi(t)\delta + o(\delta), \tag{9.1}$$

$$\Pr\{N(t, t+\delta) > 1 | \mathcal{H}_t\} = o(\delta), \tag{9.2}$$

où  $\varphi(t)$  est une fonction à déterminer. Nous conserverons les mêmes notations que dans le reste du document pour les autres variables aléatoires (voir chapitre 3).

#### 9.2.1 Moment du premier ordre

Comme nous l'avons mentionné au chapitre 6, une des propriétés des processus ponctuels se rapporte à l'indépendance entre la génération du processus ponctuel dN(t-u) et la distribution de l'intensité  $X_{t-u}(u)$ . D'après Cox & Isham (1980, p.75) et les équations (9.1-9.2), nous obtenons  $E\{N(t,t+\delta)\} = \mu_D \varphi(t)\delta$ . En particulier, l'espérance de l'intensité de pluie au temps t est :

$$E\{Y(t)\} = E\left\{ \int_{u=0}^{\infty} X_{t-u}(u) dN(t-u) \right\}$$

$$= \int_{u=0}^{\infty} E\{X_{t-u}(u)\} E\{dN(t-u)\}$$

$$= \mu_D \int_{u=0}^{\infty} E\{X_{t-u}(u)\} \varphi(t-u) du.$$

$$(9.3)$$

L'espérance de processus agrégé  $Y_i^h$  s'obtient à partir de l'expression suivante :

$$E(Y_i^h) = \int_{(i-1)h}^{ih} E\{Y(t)\} dt.$$
 (9.4)

#### 9.2.2 Variance et Autocovariance

L'autocovariance de décalage  $\tau$  de l'intensité de pluie instantanée Y(t) peut être exprimé (voir Rodríguez-Iturbe et al., 1987b, équation 3.2) en fonction de la covariance du processus ponctuel  $c(t,u) = \text{Cov}\left\{N(t,t+\delta_1),N(t+u,t+u+\delta_2)\right\}$  comme suit :

$$\operatorname{Cov}\{Y(t),Y(t+\tau)\} = \int_0^\infty \int_0^\infty \operatorname{E}\{X_{t-u}(u) \, X_{t+\tau-v}(v)\} c(t,\tau-v+u) \mathrm{d}u \mathrm{d}v.$$

Les propriétés du second-ordre du processus agrégé sont alors obtenues avec les relations suivantes :

$$\operatorname{Var}(Y_i^h) = \operatorname{Var}\left\{ \int_{(i-1)h}^{ih} Y(t) dt \right\}$$

$$= 2 \iint_{\substack{(i-1)h \le t \le ih \\ 0 \le \tau \le ih - t}} \operatorname{Cov}\{Y(t), Y(t+\tau)\} dt d\tau,$$

et

$$Cov(Y_i^h, Y_{i+k}^h) = Cov \left\{ \int_{(i-1)h}^{ih} Y(t) dt, \int_{(i+k-1)h}^{(i+k)h} Y(t) dt \right\}$$

$$= \iint_{\substack{(i-1)h \le t_1 \le ih \\ (i+k-1)h \le t_2 \le (i+k)h}} Cov \{Y(t_1), Y(t_2)\} dt_1 dt_2.$$

Afin de calculer les propriétés du second-ordre du processus agrégé, nous avons donc besoin d'exprimer la covariance c(t, u). Comme nous l'avons vu au chapitre 3, la façon dont les cellules sont distribuées dans le temps peut varier selon que le processus soit de Neyman-Scott ou de Bartlett-Lewis. La covariance du processus de comptage entre  $t_1$  et  $t_2$  s'écrit (Cox & Isham, 1980, p.33):

$$\operatorname{Cov}\left\{N(t_1, t_1 + \delta_1), N(t_2, t_2 + \delta_2)\right\} = \operatorname{Pr}\left\{N(t_1, t_1 + \delta_1) = N(t_2, t_2 + \delta_2) = 1\right\} - \operatorname{Pr}\left\{N(t_1, t_1 + \delta_1) = 1\right\} \operatorname{Pr}\left\{N(t_2, t_2 + \delta_2) = 1\right\} + o(\delta_1 \delta_2).$$

Suivant Cox & Isham (1980, p.77), deux points localisés en  $t_1$  et  $t_2 > t_1$  peuvent :

- 1. appartenir à différents clusters, auquel cas ils sont indépendants;
- 2. appartenir au même cluster de centre  $t_c$ .

Dans la version de base du processus de Neyman-Scott, les origines de cellules sont indépendamment séparées de l'origine des averses par des distances qui sont identiquement distribuées selon une loi exponentielle de paramètre  $\beta$ . Nous pouvons donc écrire :

$$\Pr\{N(t_1, t_1 + \delta_1) = N(t_2, t_2 + \delta_2) = 1\} = \mu_D^2 \varphi(t_1) \varphi(t_2) \delta_1 \delta_2$$

$$+ E\{D(D-1)\} \beta^2 \int_{-\infty}^{t_1} \varphi(t_c) e^{-\beta(t_1 - t_c)} e^{-\beta(t_2 - t_c)} \delta_1 \delta_2 dt_c + o(\delta_1 \delta_2).$$

Par conséquent, la covariance entre deux points en t et t + u s'exprime comme :

$$c(t,u) = \operatorname{Cov}\left\{N(t,t+\delta_1), N(t+u,t+u+\delta_2)\right\}$$

$$= \begin{cases} \operatorname{E}\{D(D-1)\}\beta^2 \int_{-\infty}^t \varphi(t_c) e^{-\beta(2t+u-2t_c)} \delta_1 \delta_2 dt_c + o(\delta_1 \delta_2) \text{ quand } u > 0 \\ \operatorname{E}\{D(D-1)\}\beta^2 \int_{-\infty}^{t+u} \varphi(t_c) e^{-\beta(2t+u-2t_c)} \delta_1 \delta_2 dt_c + o(\delta_1 \delta_2) \text{ quand } u < 0. \end{cases}$$

Le modèle de Bartlett-Lewis distribue les origines de cellules différemment, c'est-à-dire selon un processus de Poisson de taux  $\beta$ . Dans la version de base de ce modèle, la durée d'activité des averses a une durée aléatoire qui est distribuée selon une loi exponentielle de paramètre  $\varrho$ . Dans ce cas, nous obtenons :

$$c(t,u) = \begin{cases} \mu_D \beta \int_{-\infty}^t \varphi(t_c) \, \mathrm{e}^{-\varrho(t+u-t_c)} \, \delta_1 \delta_2 \mathrm{d}t_c + o(\delta_1 \delta_2) \text{ quand } u > 0 \\ \mu_D \beta \int_{-\infty}^{t+u} \varphi(t_c) \, \mathrm{e}^{-\varrho(t-t_c)} \, \delta_1 \delta_2 \mathrm{d}t_c + o(\delta_1 \delta_2) \text{ quand } u < 0. \end{cases}$$

Quand u=0, la variance du processus de comptage à l'instant t s'écrit

$$c(t,0) = \operatorname{Var}\left\{N(t,t+\delta)\right\} = \mu_D \varphi(t)\delta + o(\delta). \tag{9.5}$$

Notons que (9.5) est valide à la fois pour le processus de Neyman-Scott et pour le processus de Bartlett-Lewis.

#### 9.3 Un choix de $\varphi(t)$

 $\varphi(t)$  peut être simplement défini comme une fonction linéaire du temps t de la manière suivante :

$$\varphi(t) = \lambda(1 + \epsilon t),\tag{9.6}$$

où  $\epsilon$  correspond au taux de croissance de  $\varphi(t)$  et dépend de l'échelle considérée. Au temps t=0,  $\varphi(t)$  est équivalent à  $\lambda$ . Cependant, avec cette définition  $\varphi(t)<0$  pour  $t<-1/\epsilon$  ce qui est mathématiquement incorrect. Ainsi, nous posons :

$$\varphi(t) = \begin{cases} \lambda & \text{pour } t < 0 \\ \lambda(1 + \epsilon t) & \text{pour } t \ge 0. \end{cases}$$
(9.7)

Cette définition de  $\varphi(t)$  peut être interprétée comme un modèle de rupture, en supposant par exemple que la non-stationnarité est déclenchée par l'industrialisation au temps t=0. Le choix de l'instant pour lequel t=0 étant arbitraire, nous le faisons correspondre au début de la période d'observation.

Pour calculer l'expression de l'intensité moyenne agrégée à un pas de temps h, nous devons choisir la distribution bivariée entre l'intensité et la durée des cellules. Dans ce chapitre, nous supposons que l'intensité des cellules X et leur durée L sont des variables aléatoires indépendantes et distribuées exponentiellement. Les paramètres respectifs de ces distributions exponentielles sont  $\alpha$  et  $\eta$ . D'après (9.3) et (9.4), nous obtenons :

$$E(Y_i^h) = \frac{\lambda \mu_D \{ 2\epsilon e^{-\eta(i-1)h} (1 - e^{-\eta h}) + 2\eta^2 h + \epsilon \eta h (2h\eta i - h\eta - 2) \}}{2\alpha \eta^3}.$$

Cette fonction non-linéaire peut être adéquatement approchée par une relation linéaire en i, c'est-à-dire que l'augmentation de l'intensité moyenne agrégée, en fonction du temps, est presque constante. En fait, cette propriété est vérifiée pour chaque moment lorsque la fonction linéaire (9.7) est utilisée pour  $\varphi(t), t > 0$ . En effet, quand t est assez grand, l'effet des averses avant t = 0 devient négligeable. Dans ce cas, puisque le processus de Poisson approche un taux linéaire, l'approximation des moments par une relation linéaire est judicieuse. Cette propriété peut être utile pour l'estimation des paramètres, c'est pourquoi nous considérons une approximation du premier ordre des moments agrégées à  $t = n_h$ , où  $n_h$  dénote un nombre arbitraire d'heures. Les approximations suivantes peuvent donc être appliquées dès que  $n_h \geq 1000$ , puisqu'il est très peu vraisemblable que la durée d'une averse soit plus grande que 1000 heures (quand  $n_h$  augmente, la différence entre les fonctions (9.6) et (9.7) pour  $\varphi(t)$  disparaît, puisque l'effet des averses avant t < 0 devient négligeable). Pour l'espérance du processus agrégé, nous obtenons :

$$E(Y_i^h) \approx E_0^h + E_1^h i, \tag{9.8}$$

où la fonction  $E_0^h$  et  $E_1^h$  ne dépendent pas de i et sont donnés par :

$$E_0^h = \frac{\mu_D \lambda [\epsilon \{ 2 e^{-\eta n_h} (\eta n_h - 1)(1 + \eta n_h) - \eta h(\eta + 2) \} + 2\eta^2 h]}{2\alpha \eta^3}$$

et

$$E_1^h = rac{\mu_D \lambda \epsilon h \{ \mathrm{e}^{-\eta n_h} (1 - \mathrm{e}^{\eta h}) + h \eta \}}{\alpha \eta^2}.$$

L'expression de la variance du processus agrégé à un pas de temps h s'écrit :

$$\operatorname{Var}(Y_i^h) = \frac{\lambda[\mu_D f_v(\epsilon, \eta, \beta, h, i) + \operatorname{E}\{D(D-1)\}g_v(\epsilon, \eta, \beta, h, i)]}{4\beta^2 \eta^4 \alpha^2 (\eta + \beta)^2 (\eta + 2\beta)(\eta + 3\beta)(\beta^2 - \eta^2)},$$
(9.9)

où  $f_v(\epsilon, \eta, \beta, h, i)$  et  $g_v(\epsilon, \eta, \beta, h, i)$  sont définis dans l'Annexe D.1. L'expression de l'autocovariance de décalage k du processus agrégé à un pas de temps h est :

$$Cov(Y_i^h, Y_{i+k}^h) = \frac{\lambda[\mu_D g_c(\epsilon, \eta, \beta, h, i, k) + \mathbb{E}\{D(D-1)\}f_c(\epsilon, \eta, \beta, h, i, k)]}{4\beta^2 \eta^4 \alpha^2 (\eta + \beta)^2 (\eta + 2\beta)(\beta - \eta)},$$
(9.10)

où  $f_c(\epsilon, \eta, \beta, h, i, k)$  et  $g_c(\epsilon, \eta, \beta, h, i, k)$  sont définis dans l'Annexe D.2. Comme pour l'espérance (9.8), des approximations du premier ordre peuvent être calculées pour la variance et l'autocovariance de décalage k du processus agrégé afin d'obtenir les décompositions  $\operatorname{Var}(Y_i^h) \approx V_0^h + V_1^h i$  et  $\operatorname{Cov}\{Y_i^h, Y_{i+k}^h\} \approx C_0^h + C_1^h i$ . Cependant, les expressions sont trop fastidieuses pour être présentées dans cette thèse. Nous avons aussi calculé le moment du troisième ordre (centré) du processus agrégé. La formule obtenue s'écrit :

$$\begin{split} \xi_{i}^{h} &= \mathrm{E}[\{Y_{i}^{h} - \mathrm{E}(Y_{i}^{h})\}^{3}] = 6\mu_{D}\lambda[\epsilon(-2\eta^{3}h^{3} + 6n + 6\eta ih - 3\eta^{2}h^{2} + 6\eta^{2}ih^{2} \\ &+ 3\eta^{3}ih^{3})\,\mathrm{e}^{-\eta ih} - 6\epsilon\{\eta h(i-1) - 1\}\,\mathrm{e}^{-\eta(i-1)h} + (6\epsilon ih^{2}\eta^{2} - 18\epsilon + 12\epsilon ih\eta \\ &+ 6\eta^{2}h - 6\epsilon\eta^{2}h^{2} - 18\epsilon\eta h + 12\eta)\,\mathrm{e}^{-\eta h} - 12\eta + 18\epsilon - 12\epsilon\eta ih - 3\epsilon\eta^{2}h^{2} \\ &+ 6\epsilon\eta^{2}ih^{2} + 6h\eta^{2}]/(\eta^{5}\alpha^{3}) - 3\lambda\,\mathrm{E}\{D^{2} - D\}f_{\xi}(\eta, \beta, \epsilon, h, i)/\{8\eta^{5}\alpha^{3}(\eta^{2} - \beta^{2})^{2} \\ &\times \beta^{2}(\eta + \beta)\} + \mathrm{E}\{(D^{2} - D)(D - 2)\}\lambda g_{\xi}(\eta, \beta, \epsilon, h, i)/\{36\eta^{5}\alpha^{3}\beta^{2}(\eta^{2} - \beta^{2})^{2} \\ &\times (\eta + 2\beta)^{2}(2\eta + \beta)^{2}\}, \end{split}$$

où  $f_{\xi}(\eta, \beta, \epsilon, h, i)$  et  $g_{\xi}(\eta, \beta, \epsilon, h, i)$  sont des expressions trop longues pour être présentées dans cette thèse. Nous avons également calculé la décomposition  $\xi_i^h \approx T_0^h + T_1^h i$  pour cette expression.

Suivant Cowpertwait (1994), nous calculons la probabilité PD(h) qu'un intervalle de longueur h ne comporte pas de pluie, c'est-à-dire  $\Pr(Y_i^h=0)$ . Dans le cas d'une distribution géométrique pour le nombre de cellules, l'expression est une modification de Cowpertwait (1994, équation 2.16) et est donnée par :

$$\Pr(Y_i^h = 0) = \exp\left[-\int_0^\infty \varphi\{h(i-1) - t\}\{1 - p_h(t)\}dt - \int_0^h \varphi\{h(i-1) + t\}\{1 - p_t(0)\}dt\right],$$
(9.11)

οù

$$p_h(t) = \frac{\mathrm{e}^{-\beta(t+h)}(\eta-\beta) + \eta - \beta - \eta\,\mathrm{e}^{-\beta t} + \beta\,\mathrm{e}^{-\eta t}}{(1-\mu_D)\{\mathrm{e}^{-\beta(t+h)}(\eta-\beta) - \eta\,\mathrm{e}^{-\beta t} + \beta\,\mathrm{e}^{-\eta t}\} + \eta - \beta}$$

est la probabilité que l'intervalle (t, t+h) soit sec avec une origine d'averse à l'instant zéro. Avec  $\varphi(t)$  exprimée par (9.7), l'expression (9.11) requiert des évaluations numériques pour les deux intégrales. Puisqu'aucune formule explicite n'est disponible pour  $\Pr(Y_i^h=0)$ , des approximations linéaires ne peuvent pas être appliquées pour cette propriété statistique. La Figure 9.1 représente les valeurs théoriques de l'espérance, de la variance, de l'autocovariance de décalage un, du moment du troisième ordre et de la proportion de périodes sèches (obtenue par intégration numérique) à une échelle horaire, pour l'ensemble de paramètres suivant  $\{\lambda=0.01; \beta=0.1; \eta=1; \mu_D=5; \alpha=0.25; \epsilon=0.001\}$ . Nous noterons que les approximations linéaires sont adéquates dès que  $i\geq 6$ .

L'obtention de moments empiriques est nécessaire si l'on veut appliquer la méthode généralisée des moments. Dans notre cas, nous rencontrons un problème majeur puisque la non-stationnarité introduite dans le NSRPM implique une variation temporelle des moments. Nous ne pouvons alors plus calculer une seule estimation empirique pour chaque moment et une méthode alternative doit être développée. Nous proposons de construire une approche basée sur les approximations linéaires des moments, c'est-à-dire de segmenter le temps d'observation en intervalles réguliers, de calculer chaque statistique à partir des précipitations tombées dans ces intervalles de temps et d'estimer la pente et l'origine des relations linéaires à partir de ces statistiques.

#### 9.4 Calcul des moments empiriques

Les séries observées de pluie à une échelle h sont notées  $Y_i^h, i=1,\ldots,I$ . Considérons une fenêtre mobile qui segmente l'intervalle [1,I] en sous-intervalles de longueur  $\Omega$  qui ne se chevauchent pas. Nous calculons alors les statistiques (cela peut être la moyenne empirique, la variance, l'auto-covariance de décalage k ou le moment du troisième ordre) sur les données de pluie de chaque intervalle. Une régression robuste de ces statistiques (en fonction de i) est appliquée afin d'obtenir une estimation empirique des paramètres impliqués dans les approximations linéaires des moments théoriques. Plusieurs limitations doivent cependant être mentionnées :

- Le calcul des statistiques est appliqué sur des observations qui ne sont pas identiquement distribuées en raison de la non-stationnarité de  $Y_i^h$ ;
- une hétéroscedasticité apparaît pour chaque statistique, la variabilité augmentant avec i.

Pour ces raisons nous appliquons une méthode de carrés pondérés afin d'obtenir les paramètres de la régression. Cette régression robuste utilise un algorithme des moindres carrés itératif où les poids accordés aux observations évoluent à chaque itération en appliquant une fonction quadratique sur les résidus de chaque itération précédente (voir par exemple Rousseeuw & Leroy, 1987;

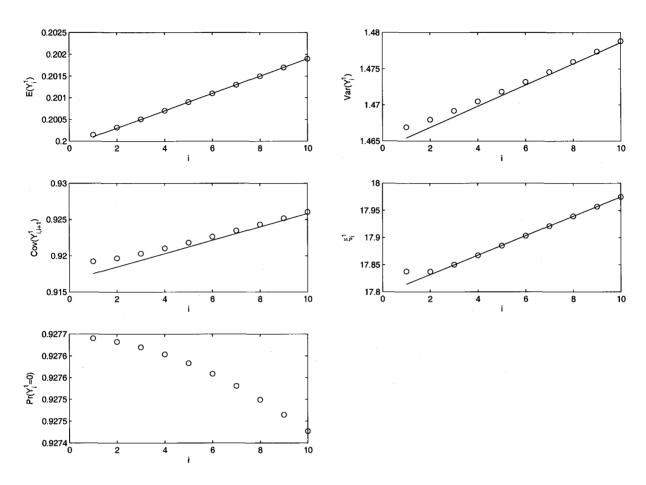


FIG. 9.1 – Approximations linéaires des moments (ligne pleine) et moments théoriques (cercles) (espérance, variance, autocovariance de décalage un, moment du troisième ordre centré et proportion de périodes sans pluie) à une échelle horaire en fonction du temps i pour l'ensemble de paramètres fixé :  $\{\lambda = 0.01; \beta = 0.1; \eta = 1; \mu_D = 5; \alpha = 0.25; \epsilon = 0.001\}$ .

Maronna et al., 2006). Pour les raisons citées plus haut, il est nécessaire d'analyser l'efficacité de la régression et d'examiner le biais possible. Nous simulons à cette fin 100 mois de pluies horaires pour 100 différents ensembles de paramètres tirés aléatoirement d'une loi uniforme entre les bornes définies ci-dessous. Une longueur de 100 mois est représentative d'une très longue série de pluies horaires, le NSRPM étant habituellement ajusté mensuellement afin de préserver la variation saisonnière. Nous contraignons les paramètres dans un espace des paramètres réaliste :  $0.002 < \lambda < 0.02 \ [h^{-1}]; \ 0.05 < \beta < 0.5 \ [h^{-1}]; \ 0.5 < \eta < 5 \ [h^{-1}]; \ 3 < \mu_C < 20; \ 0.2 < \alpha < 2; \ 0.000001 < \epsilon < 0.00001 \ [h^{-1}]$ . L'expérience montre que les paramètres évoluent généralement à l'intérieur de ces bornes. Les bornes sur  $\epsilon$  correspondent à un augmentation de la fréquence des averses de 0.9% à 8.8% par an. Il est à noter qu'avec une augmentation de 8.8%, la fréquence des averses double en 8 ans.

Une régression robuste est ensuite appliquée pour différentes valeurs de  $\Omega$  (200 heures à 20,000 heures pour l'échelle horaire et 100 jours à 800 jours pour l'échelle journalière). Cette procédure est répétée 1,000 fois afin d'observer la variabilité des estimations d'une simulation à l'autre (pour un ensemble spécifique de paramètres).

Pour chaque ensemble de paramètres, nous estimons le biais relatif des estimateurs par :

$$\frac{1}{1000} \sum_{k=1}^{1000} \frac{\hat{\kappa}_k - \kappa}{\kappa},$$

où  $\kappa$  est une pente ou une origine théorique,  $\hat{\kappa}_k$  est l'estimation empirique correspondante et k est l'indice du réplicat. Nous représentons le biais moyen relatif et un intervalle de confiance empirique à 80% ( $10^{\text{ème}}$  et  $90^{\text{ème}}$  centiles) pour chaque statistique et pour différentes valeurs de  $\Omega$  sur les Figures 9.2 à 9.5. Nous pouvons d'abord noter que, généralement, le biais diminue avec  $\Omega$  pour les échelles horaires et journalières. Sa variabilité diffère selon le moment et l'échelle temporelle considérés mais est plus faible pour les grandes valeurs de  $\Omega$ . Cependant, quand  $\Omega$  est grand, la régression s'applique sur un faible nombre de points. Pour  $\Omega = 20,000$  heures (ou  $\Omega = 800$  jours), 100 mois de pluies horaires (ou journalières) correspondent à trois intervalles de valeurs. La régression est donc appliquée avec seulement trois points lorsque  $\Omega=20,000$  heures ou  $\Omega=800$  jours. Nous reportons la valeur moyenne et les intervalles de confiance empiriques à 80% correspondants pour les différentes statistiques à une échelle horaire et journalière au Tableau 9.1 pour  $\Omega=20,000$ heures et  $\Omega = 800$  jours. Nous pouvons constater que pour chaque statistique, la pente semble plus difficile à estimer, tout spécialement pour l'autocovariance de décalage un et le moment du troisième ordre à une échelle journalière. Bien que le biais semble disparaître lorsque  $\Omega$  augmente, la variabilité est très élevée lorsque le nombre de points utilisés dans la régression est trop petit (c'est-à-dire moins que dix points). Par conséquence, un bon choix pour  $\Omega$  semble être la plus haute valeur pour laquelle au moins dix intervalles sont disponibles et dépend donc de la longueur de la série de données de pluie.

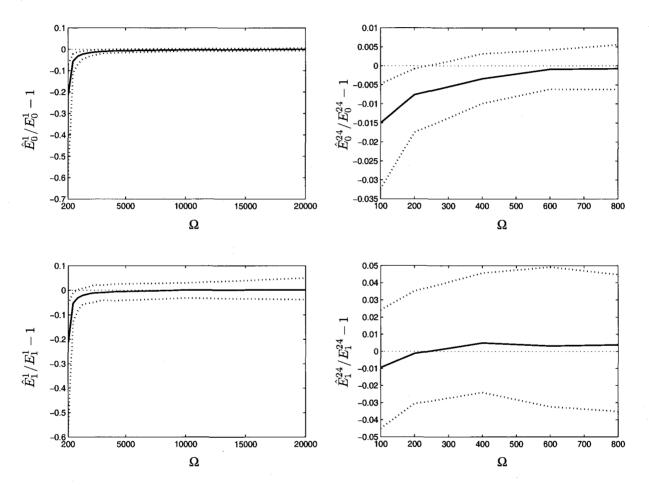


FIG. 9.2 – Biais relatif moyen et intervalle de confiance empirique à 80% ( $10^{\text{ème}}$  et  $90^{\text{ème}}$  centiles) résultant de la régression robuste pour la moyenne.

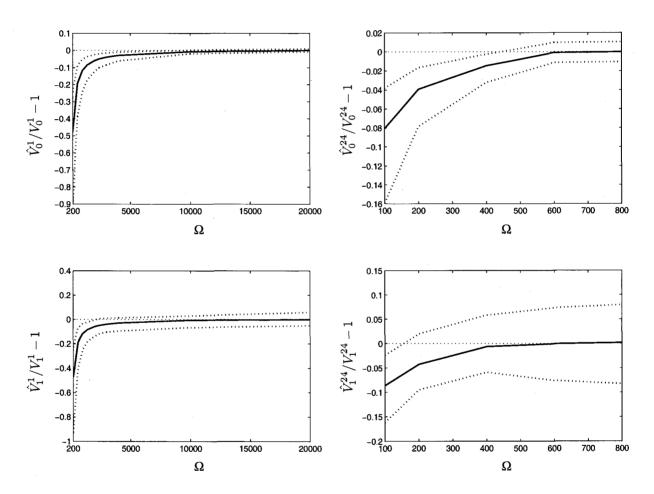


FIG. 9.3 – Biais relatif moyen et intervalle de confiance empirique à 80% ( $10^{\rm ème}$  et  $90^{\rm ème}$  centiles) résultant de la régression robuste pour la variance.

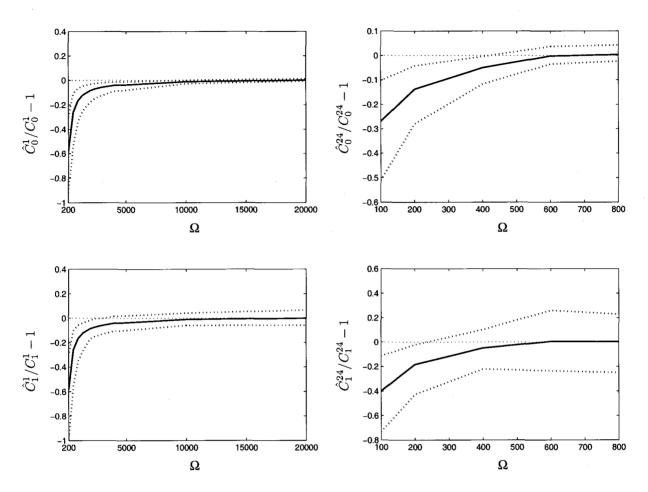


Fig. 9.4 – Biais relatif moyen et intervalle de confiance empirique à 80% ( $10^{\text{ème}}$  et  $90^{\text{ème}}$  centiles) résultant de la régression robuste pour l'autocovariance de décalage un.

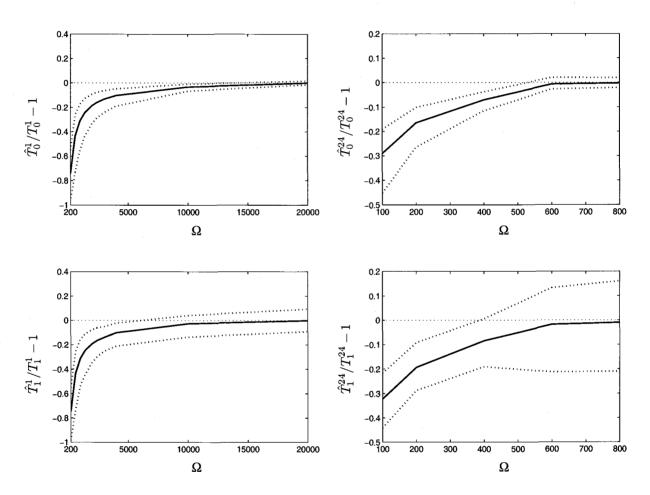


Fig. 9.5 – Biais relatif moyen et intervalle de confiance empirique à 80% ( $10^{\text{ème}}$  et  $90^{\text{ème}}$  centiles) résultant de la régression robuste pour le moment du troisième ordre (centré).

TAB. 9.1 – Biais relatif moyen et intervalle de confiance empirique à 80% ( $10^{\text{ème}}$  et  $90^{\text{ème}}$  centiles entre parenthèses) des 100 ensembles de paramètres aléatoires pour  $\Omega = 20,000$  à une échelle horaire et  $\Omega = 800$  à une échelle journalière.

		Horaire	Journalière
$\mathrm{E}(Y_i^h)$	$E_0^h$	-0.0002 (-0.0066;0.0057)	-0.0007 (-0.0062;0.0056)
	$E_1^h$	0.0011 (-0.0382;0.0496)	0.0038 (-0.0351;0.0447)
$\operatorname{Var}(Y_i^h)$	$V_0^h$	-0.0004 (-0.0082;0.0084)	-0.0002 (-0.0105;0.0103)
	$V_1^h$	-0.0010 (-0.0511;0.0591)	0.0017 (-0.0825;0.0795)
$Cov(Y_i^h, Y_{i+k}^h)$	$C_0^h$	0.0001 (-0.0083;0.0110)	0.0034 (-0.0235;0.0421)
	$C_1^h$	-0.0017 (-0.0585;0.0654)	0.0028 (-0.2504; 0.2285)
$\xi_i^h$	$T_0^h$	-0.0010 (-0.0157;0.0147)	-0.0024 (-0.0211;0.0194)
	$T_1^h$	-0.0015 (-0.0914;0.0932)	-0.0100 (-0.2100;0.1606)

#### 9.5 Méthode d'estimation des paramètres du NSRPM-NS

Dans cette section nous décrivons une méthode permettant d'estimer les paramètres impliqués dans les approximations linéaires de chaque moment. Nous proposons d'inclure ces estimations empiriques dans la méthode des moments décrite au chapitre 7. Soit  $\phi$  l'ensemble des paramètres  $(\lambda, \beta, \eta, \mu_C, \alpha, \epsilon)$ . La méthode généralisée des moments consiste à minimiser  $\mathcal{O}$  définie par :

$$\mathcal{O}(\phi) = \sum_{i=1}^{p} \omega_i \left(\frac{\hat{\kappa}_i}{\kappa_i} - 1\right)^2, \tag{9.12}$$

où  $\kappa_i$  est une origine ou une pente des approximations linéaires données à la section 9.3, c'est-à-dire  $E_0^h, E_1^h, V_0^h, V_1^h, C_0^h, C_1^h, T_0^h$  ou  $T_1^h$  pour différentes échelles de temps.  $\hat{\kappa}_i$  est l'estimation empirique correspondante et  $\omega_i$  dénote un poids positif. p représente le nombre de moments inclus dans la procédure. Le Tableau 9.2 indique nos choix des poids  $\omega_i$  qui représentent un compromis entre le souhait d'accorder une importance égale à chaque moment et la difficulté à obtenir une estimation empirique de ces moments spécifiques (voir Tableau 9.1). Précisément, nous pouvons penser qu'il sera plus facile d'obtenir des estimations empiriques pour les origines des relations linéaires que pour les pentes. Il n'est pas évident en effet que des pentes positives nettes seront observées pour tous les moments statistiques sur des données observées. C'est pourquoi nous accordons des poids plus élevés aux origines  $E_0^h, V_0^h, C_0^h$  et  $T_0^h$  qu'aux pentes  $E_1^h, V_1^h, C_1^h$  et  $T_1^h$ . Par exemple, la somme des poids du Tableau 9.2 étant égal à 1, la somme des poids associés aux origines est 0.685, alors que les poids associés aux pentes totalisent 0.315. Les poids associés à l'espérance, à la variance, à l'autocovariance de décalage un et au moment d'ordre trois totalisent respectivement des poids de 0.4, 0.3, 0.17 et 0.13. Suite aux interprétations des Figures 9.2 à 9.5, nous accordons ainsi des

		Horaire	Journalier		
$\mathrm{E}(Y_i^h)$	$E_0$	0.2	0.1		
$\mathbb{E}(I_i^{-})$	$E_1$	0.05	0.05		
$\operatorname{Var}(Y_i^h)$	$V_0$	0.1	0.1		
$\operatorname{var}(I_i)$	$V_1$	0.05	0.05		
$Cov(Y_i^h, Y_{i+k}^h)$	$C_0$	0.07	0.05		
$\bigcup_{i=1}^{n} Cov(I_i, I_{i+k})$	$C_1$	0.03	0.02		
$\xi_i^h$	$T_0$	0.05	0.015		
$ \zeta_i $	$T_1$	0.05	0.015		

TAB. 9.2 – Choix des poids dans la procédure d'estimation des paramètres du NSRPM-NS.

poids importants aux origines et pentes liées à l'espérance et à la variance et des poids beaucoup plus faibles à l'origine et à la pente du moment d'ordre trois.  $\mathcal{O}$  est minimisée numériquement à l'aide de la procédure d'optimisation suivante qui est adaptée de la méthode explicitée au chapitre 7:

- 1. Nous réalisons 1,000 optimisations contraintes dans l'espace des paramètres suivants :  $0.0001 < \lambda < 0.1 \ [h^{-1}]$ ;  $0.001 < \beta < 0.99 \ [h^{-1}]$ ;  $0.01 < \eta < 60 \ [h^{-1}]$ ;  $1 < \mu_C < 500$ ;  $0.01 < \alpha < 1000$ ;  $10^{-9} < \epsilon < 10^{-4}$ . Les paramètres sont contraints afin de rester dans un espace des paramètres physiquement vraisemblable. Les points de départ sont générés d'après une loi beta entre ses bornes. Un premier ensemble de ces paramètres est estimé en minimisant l'expression (9.12) à l'aide d'une méthode de programmation quadratique séquentielle (sequential quadratic programming, SQP, voir Fletcher, 1987). Nous obtenons ainsi 1,000 ensembles préliminaire de solutions.
- 2. Les étapes 3 à 6 sont répétées plusieurs fois (par exemple dix fois);
- 3. Seuls les 500 meilleurs ensembles de paramètres sont conservés. Le meilleur ensemble de paramètres obtenu devient le point de départ des futures optimisations.
- 4. Nous réalisons 500 optimisations contraintes. Les points de départ sont obtenus à l'aide de perturbations aléatoires autour du vecteur de paramètres précédemment obtenu. La perturbation aléatoire est tirée d'une distribution beta sur les bornes courantes (les bornes sont définies à l'étape 1 et évoluent à l'étape 6).
- 5. Nous rejetons tous les vecteurs de paramètres pour lesquels la fonction objectif dépasse la médiane des valeurs de fonctions objectives obtenues jusque-là.
- 6. Les bornes évoluent selon les meilleurs ensembles de paramètres trouvés jusque-là (c'est-à-dire ceux qui minimisent la fonction objectif). Si les meilleurs ensembles sont proches, les bornes sont restreintes. Si elles sont proches d'une borne existante, les bornes sont espacées. Ces choix sont arbitraires mais tiennent compte de l'échelle de valeurs des paramètres et sont motivés par le souhait de contraindre l'algorithme à sortir du domaine d'attraction de

9.6 Simulation

minima locaux, en tenant compte du fait que l'adaptation de ces bornes ne s'effectue qu'un faible nombre de fois (dix fois dans ce cas).

Le nombre d'optimisations contraintes réalisées à l'étape 1 est élevé afin de parcourir un espace important de paramètres, le point de départ de l'algorithme du chapitre 7 ne pouvant plus être obtenu avec l'introduction de la non-stationnarité. Cet algorithme conserve néanmoins les avantages de la précédente procédure d'estimations des paramètres, puisqu'il parcourt un important espace de paramètres, évitant ainsi que la solution obtenue soit un minimum local.

#### 9.6 Simulation

Le processus de Poisson non-stationnaire est généré à l'aide de la méthode de transformation inverse (voir par exemple Devroye, 1986; Harrod & Kelton, 2006) et s'effectue en deux étapes, c'est-à-dire :

- Générer des temps d'arrivée  $s_i$  selon un processus de Poisson stationnaire de taux 1,
- retourner le *i*-ème temps d'arrivée d'averse du processus de Poisson non-stationnaire  $t_i = \Lambda^{-1}(s_i)$ ,

où  $\Lambda(t) = \int_{-\infty}^{t} \varphi(s) ds$  est la fonction associée au taux cumulé. Dans notre cas, une borne inférieure finie doit être fixée afin d'éviter la divergence de  $\Lambda(t)$ . Par conséquent, nous devons choisir une borne inférieure suffisamment éloignée du point de départ des simulations pour que l'influence des averses qui précèdent cette borne devienne négligeable. Nous donnons un exemple de fonctions  $\varphi(t)$ ,  $\Lambda(t)$  et  $\Lambda^{-1}(s)$  avec  $\lambda=0.01$  et  $\epsilon=0.01$  à la Figure 9.6.

#### 9.7 Application

Dans cette section, nous réalisons une application du NSRPM non-stationnaire. Idéalement, nous aurions souhaité appliquer ce modèle sur des données de pluies exhibant clairement une non-stationnarité temporelle. Nous n'avons malheureusement pas trouvé de série suffisamment longue et pertinente pour notre application. Il est prévu de réaliser une application sur des données simulées à partir d'un modèle climatique développé par le Consortium Ouranos. L'extraction de ces simulations nécessite l'utilisation d'un robot qui manipule les disques de stockage contenant les données simulées pour chaque journée. Cette opération de récupération des données est donc particulièrement longue (plus de trois mois) et cette application n'a pas pu être réalisée dans le cadre de cette thèse.

Il est clair qu'il est absolument nécessaire d'analyser de façon détaillée des données observées

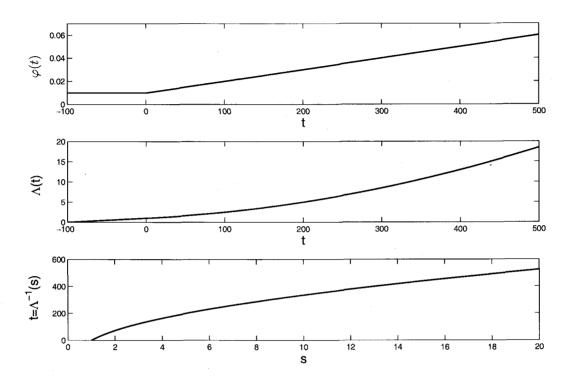


Fig. 9.6 – Exemples de fonctions  $\varphi(t)$ ,  $\Lambda(t)$  et  $\Lambda^{-1}(s)$  pour la simulation avec  $\lambda = 0.01$  et  $\epsilon = 0.01$ .

ou issues de modèles climatiques avant de procéder à des applications. De nombreuses questions doivent en effet être résolues, telles que l'application du modèle mois par mois, puisque les propriétés du modèle changent d'une année à l'autre. Il faudrait alors réfléchir à une façon de "lisser" cette non-stationnarité.

Nous sommes donc contraints d'appliquer les méthodes décrites aux sections 9.4 et 9.5 sur des pluies simulées à un pas de temps horaire à partir de l'ensemble de paramètres arbitraire suivant :  $\{\lambda=0.01,\ \eta=1.5,\ \beta=0.2,\ \mu_D=7,\ \alpha=0.25\ \text{et}\ \epsilon=0.000001\}$ . Pour cette application, 200 mois sont simulés, ce qui correspond à une série de pluies horaires très longue. Des régressions linéaires robustes sont appliquées sur les statistiques calculées à partir de la méthode décrite à la section 9.4, les résultats sont représentés aux Figures 9.7 et 9.8. Les difficultés à estimer les pentes des relations linéaires apparaissent clairement, la pente obtenue étant même négative pour l'autocovariance journalière de décalage un. Même si le biais des estimateurs proposés tend à être nul (voir Tableau 9.1), leur variabilité rend l'estimation des paramètres plus complexe. Cette variabilité semble être plus forte pour les pentes des relations linéaires, et plus encore à une échelle journalière, justifiant ainsi le choix des poids du Tableau 9.2. Les paramètres sont ensuite estimés

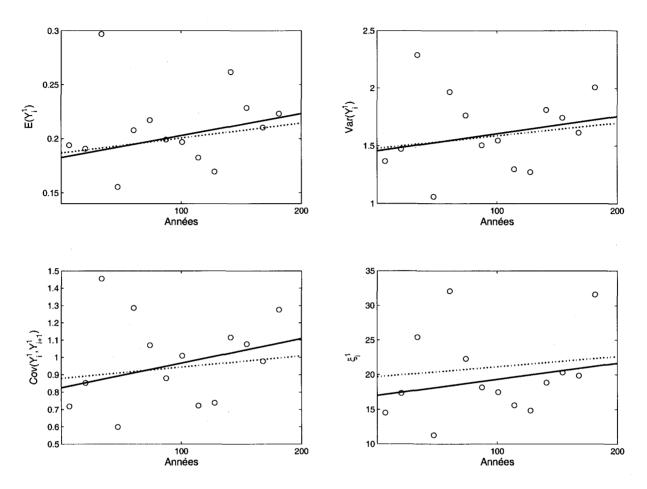


FIG. 9.7 – Moments empiriques (moyenne, variance, autocovariance de décalage un et moment d'ordre trois) et régressions linéaires correspondantes (trait plein) à une échelle horaire. Nous représentons la relation linéaire théorique en trait pointillé.

en utilisant la méthode décrite dans la section 9.5 avec  $\Omega=5000$  à une échelle horaire et  $\Omega=500$  à une échelle journalière. Les estimations obtenues sont reportées dans le Tableau 9.3. Les paramètres sont globalement estimés correctement, surtout pour  $\lambda$  et  $\epsilon$  liés au taux du processus de Poisson. L'estimation des paramètres  $\beta$  et  $\mu_D$  est plus approximative. Il est à souligner que les estimations des paramètres donnent une valeur plus petite pour la fonction objectif (0.2735) qu'avec le vrai vecteur de paramètre  $\phi$  (0.3714) ce qui indique bien que les écarts entre les estimations et les vraies valeurs de paramètres sont principalement causées par la variabilité des propriétés statistiques d'une simulation à l'autre.

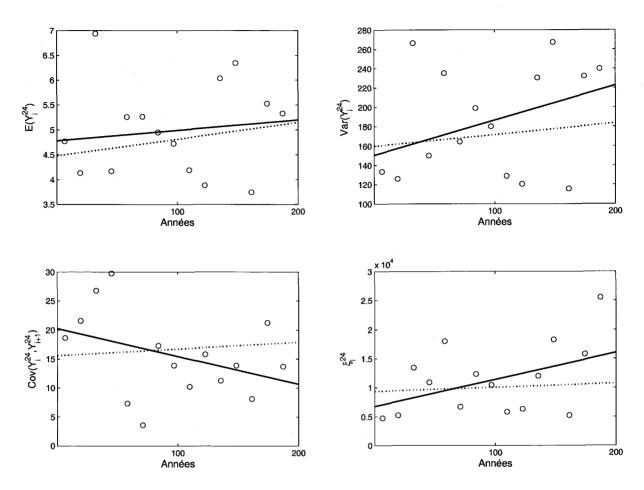


FIG. 9.8 – Moments empiriques (moyenne, variance, autocovariance de décalage un et moment d'ordre trois) et régressions linéaires correspondantes (trait plein) à une échelle journalière. Nous représentons la relation linéaire théorique en trait pointillé.

Tab. 9.3 – Estimations des paramètres du NSRPM non-stationnaire  $\hat{\phi}$ .

	λ	$\eta$	$\beta$	$\mu_{\mathbf{D}}$	$\alpha$	$\epsilon$	$\mathcal{O}(\phi)$
Vraies valeurs	0.01	1.5	0.2	7	0.25	10.e-007	0.3714
Estimations	0.012	1.159	0.116	4.291	0.231	9.0846e-007	0.2735

### 9.8 Conclusion

Dans ce chapitre, nous avons développé un modèle de Neyman-Scott non-stationnaire qui respecte les contraintes liées à l'application d'un tel modèle. Nous proposons ainsi une fonction dépendante du temps pour le taux d'arrivée des averses qui permet d'obtenir des expressions ana-

9.8 Conclusion 121

lytiques pour l'espérance, la variance, l'autocovariance et le moment d'ordre trois de l'intensité de pluie. Avec la fonction de taux d'occurrence des averses choisie, les propriétés statistiques peuvent être approchées par une relation linéaire en fonction du temps. Cette propriété est exploitée afin de développer une méthode d'estimation des paramètres. Cette approche est ensuite testée sur des données simulées.

Il est clair que l'application d'un tel modèle est limitée par la forme de l'évolution temporelle des propriétés statistiques, c'est-à-dire une croissance linéaire dans notre cas. De plus, l'approche proposée afin de modéliser les relations linéaires des propriétés statistiques souffre d'une certaine instabilité. Ce travail ne représente cependant qu'une première étape dans l'élaboration d'un NSRPM non-stationnaire. De nombreux efforts sont encore à fournir afin de décrire clairement l'évolution future des propriétés de la pluie, les prévisions n'étant pour l'instant que des réalisations possibles d'un scénario climatique. Il s'agirait notamment de déterminer la validité de nos hypothèses sur des données observées.

## Chapitre 10

### Conclusion

#### 10.1 Travail accompli

Cette thèse porte sur la modélisation des précipitations à un pas de temps court (horaire et journalier) à l'aide de processus stochastiques ponctuels. Nous proposons des généralisations au modèle dit de Neyman-Scott (Neyman-Scott Rectangular Pulses Model, NSRPM). Dans ce modèle, les averses sont simulées sous la forme d'amas de cellules pluvieuses.

# 10.1.1 NSRPM incluant un lien entre l'intensité et la durée des cellules pluvieuses

Une première généralisation permet de considérer une dépendance entre l'intensité et la durée des cellules pluvieuses, ces deux composantes étant classiquement considérées comme indépendantes. Nous présentons alors plusieurs types de distributions bivariées englobant comme cas particulier la version de base du NSRPM. Ces distributions bivariées sont construites à l'aide de copules cubiques (pour modéliser la dépendance) et de lois exponentielles (pour les lois marginales), ce qui permet d'atteindre des degrés de dépendance élevés et d'introduire une forme de dépendance souple.

Plusieurs modèles sont comparés avec des séries de pluies horaires d'origine variée (Suisse, Belgique, France, Canada et États-Unis d'Amérique). La disponibilité de longues séries de pluies horaires et le coût de calcul limitent en général l'application de ce type de modèles sur des données

124 Conclusion

de pluie avec une telle variété de climats. Un algorithme exploitant la méthode généralisée des moments est proposé afin d'estimer les paramètres des différents modèles. Les difficultés rencontrées lors des applications ont mis en lumière certaines limitations du NSRPM lorsque l'autocovariance journalière de décalage un de l'intensité pluvieuse est faible.

Concernant les forces et les faiblesses des différents modèles considérés, nous retiendrons une très bonne performance globale du modèle avec la copule cubique Sym2, particulièrement lorsque l'on considère l'adéquation du modèle aux propriétés statistiques utilisées dans la méthode d'estimation des paramètres. Pour plusieurs stations et mois particuliers, nous notons également que nous obtenons de bons résultats à l'égard du même critère avec les modèles utilisant les distributions de Pareto et gamma.

Un des objectifs de cette thèse est également de donner des informations sur la forme de la dépendance entre l'intensité et la durée des cellules, indirectement observée par le biais des applications. Principalement grâce à leur capacité à modéliser des dépendances négatives et positives dans un intervalle important, les copules cubiques employées semblent souvent améliorer la reproduction des propriétés de la pluie. Sur la plupart des applications, le tau de Kendall entre l'intensité de la durée des cellules (lié à l'estimation du paramètre de la copule cubique) est négatif. Les structures de dépendance qui donnent les meilleurs résultats sont les copules cubiques Sym2, Fcub et Asym3 dans une moindre mesure. Le fait que ce degré de dépendance soit presque systématiquement négatif (et de plus atteint souvent la borne inférieure de l'intervalle de dépendance couvert par les copules cubiques) prouve bien l'importance de tenir compte de ce lien entre l'intensité et la durée des cellules.

Le critère de reproduction des valeurs extrêmes est particulièrement important à prendre en compte lors du choix d'un modèle de pluie. Nous montrons que les maxima annuels à une échelle journalière sont généralement bien reproduits par la plupart des modèles. Cependant, nous retrouvons parfois une légère sous-estimation de ces valeurs extrêmes à une échelle horaire pour les périodes de retour les plus élevées. Pour ce problème particulier, lorsque l'intensité des cellules est distribuée selon une loi de Pareto, l'écart entre les maxima simulés et observés est réduit, vraisemblablement grâce à sa queue supérieure de distribution lourde. Bien que l'on pouvait s'attendre à une amélioration de la reproduction des valeurs extrêmes en tenant compte de la dépendance entre l'intensité et la durée des cellules, ce n'est pas systématiquement le cas. Nous pouvons alors nous demander si ce n'est pas simplement la limitation du degré de dépendance couvert par les copules cubiques qui empêche le modèle de mieux restituer les extrêmes. Autrement dit, avec d'autres formes de dépendance plus souples, nous pourrions peut-être nous attendre à obtenir un lien encore plus fort et pouvoir ainsi simuler des averses extrêmement courtes et intenses.

#### 10.1.2 NSRPM non-stationnaire

Ce mémoire de thèse propose de nouveaux développements des processus ponctuels afin d'obtenir un modèle de Neyman-Scott non-stationnaire, en gardant à l'esprit les contraintes qu'impose l'application d'un tel modèle. Dans ce NSRPM non-stationnaire, le taux d'arrivée des averses est une fonction linéaire du temps. Avec cette fonction simple, nous obtenons des expressions explicites pour l'espérance, la variance, l'autocovariance et le moment d'ordre trois de l'intensité de pluie. Ces propriétés statistiques du modèle dépendent alors du temps. Nous montrons qu'avec notre choix de fonction du taux d'occurrence des averses, ces fonctions du temps peuvent être judicieusement approchées par une relation linéaire. Nous proposons également une méthode d'estimation des paramètres qui s'appuie sur les approximations linéaires des moments pour lier les propriétés statistiques du modèle aux données observées. Une illustration de la méthode est exposée sur des données simulées. Cependant, il est évident que la validation du modèle devra également s'effectuer à travers une application sur des données observées ou issues d'un modèle climatique.

Le NSRPM non-stationnaire développé présente deux principales limitations sur lesquelles il sera nécessaire de poursuivre les efforts. En effet, les applications de ce modèle sont contraintes par la forme de l'évolution temporelle des propriétés statistiques, à savoir une relation linéaire avec le temps. La seconde problématique principale est l'instabilité de la méthode d'estimation des paramètres. Il est donc indispensable de continuer la réflexion sur la façon de lier les moments empiriques à ceux issus du modèle.

#### 10.2 Discussion et pistes des travaux à accomplir

De nombreux efforts mériteraient d'être poursuivis afin de tirer tous les avantages des modèles représentant les averses sous formes d'amas de cellules pluvieuses. Cette section propose quelques éléments de réflexion et des pistes de travaux à poursuivre.

#### 10.2.1 Estimation des paramètres

L'estimation des paramètres constitue un premier obstacle à l'application de ce modèle puisqu'il nécessite actuellement de grandes ressources de calcul lors de la minimisation de la fonction objectif. La difficulté à obtenir une fonction de vraisemblance empêche l'utilisation de méthodes bayésiennes ou le recours direct à des données de pluie. Plusieurs approches ont déjà été proposées comme alternatives à la méthode généralisée des moments mais leur application semble limitée puisqu'elles requièrent généralement une simplification du modèle (Rodríguez-Iturbe et al., 1988), à l'image de la méthode des séries de Fourier de Chandler (1995). La méthode des moments est pertinente puisqu'elle permet de lier un grand nombre de propriétés statistiques au modèle. Je pense personnellement que des travaux devraient examiner en détail le lien entre les paramètres du modèle et les propriétés statistiques de la pluie. Par exemple, la structure du modèle oppose les propriétés des averses et celles des cellules pluvieuses. Nous pourrions étudier les interactions entre le taux du processus de Poisson  $\lambda$ , la distance entre les origines de cellules et l'origine d'averse  $\beta$  et le nombre de cellules  $\mu_D$  par rapport aux propriétés statistiques des pluies journalières ( $\lambda$  semblant être lié à la moyenne). Est-ce que le paramètre  $\eta$  gérant la durée des cellules ne pourrait pas être lié assez directement à l'autocovariance horaire de décalage un? Cette problématique concerne directement l'identifiabilité des paramètres.

#### 10.2.2 Validation du modèle

De nouvelles approches concernant la validation des modèles ont été proposées dans le rapport de Chandler & Onof (2005). Ces développements permettent par exemple de tester la significativité de la différence des valeurs de fonctions objectif entre plusieurs modèles. Une méthode permettant de prendre en compte la qualité de reproduction des valeurs extrêmes est aussi présentée. L'application de ces approches à diverses séries de pluies me paraît importante puisque cela permettrait de les valider et d'illustrer leur pertinence sur différents climats.

#### 10.2.3 Vers un modèle parfait?

Les différentes applications réalisées au cours de cette thèse ont fait clairement apparaître qu'un modèle particulier ne peut pas être adapté à toutes les formes de champs de précipitation. En quelque sorte sorte tous les modèles ne sont que de pâles reflets de la réalité, comme l'indique le célèbre adage (voir Box & Draper, 1987, p.74). La question pratique est de savoir quels éléments du modèle sont les plus importants et doivent être conservés afin que ces modèles soient le plus vraisemblables possible et conservent leur utilité.

Idéalement, la distribution bivariée parfaite entre la durée et l'intensité des cellules permettrait à la fois de reproduire une densité de probabilité forte pour les faibles valeurs d'intensité de cellules (queue inférieure lourde) et un degré de dépendance important, surtout pour les valeurs négatives. Cette distribution bivariée idéale combinerait alors les forces de la distribution de Pareto ou gamma pour l'intensité des cellules et posséderait une structure de dépendance capable d'atteindre des corrélations (mesurée par le biais du tau de Kendall ou du rho de Spearman) négatives et positives fortes. La reproduction des extrêmes semble aussi nécessiter une queue supérieure lourde pour la distribution de l'intensité des cellules, telle que reproduite par la loi de Pareto. Le couplage

entre une copule cubique et une distribution de Pareto semble ainsi particulièrement intéressant. Cependant, nous avons constaté que les contraintes calculatoires limitent le choix de la distribution bivariée afin d'obtenir une expression explicite des moments théoriques issus du modèle. Il faudrait alors étudier les possibilités d'approximation des moments, à l'image des développements réalisés pour le calcul des propriétés statistiques du modèle de Bartlett-Lewis à paramètre aléatoire (Random Parameter Bartlett-Lewis Model, RPBLM), voir Onof (2003).

Un type de généralisation possible, qui pourrait permettre de s'approcher de la physique du phénomène pluvieux, agit au niveau du processus ponctuel. Nous proposons dans cette thèse la prise en compte de la non-stationnarité au niveau de l'arrivée des averses. La généralisation à un processus de renouvellement est également un aspect intéressant à approfondir. D'autres variantes des modèles de Neyman-Scott et de Bartlett-Lewis ont été proposées et testées dans la littérature, par exemple la prise en compte de plusieurs types de cellules ou l'aléation (randomization) du paramètre lié à la durée des cellules. Ces développements semblent cependant difficiles à combiner avec ceux proposés dans cette thèse. La question est alors de savoir quelles hypothèses sont les plus importantes à conserver pour chaque application.

Il me semble important de noter que les récentes applications de ce type de processus ponctuels modélisent les champs de précipitations à la fois en temps et en espace. La prise en compte de la dimension spatiale du phénomène de pluie semble maintenant concentrer tous les efforts, notamment en raison de l'accès aux données obtenues par satellite. La variabilité spatiale du phénomène pluvieux doit évidemment être considérée. Cependant, la complexité de cette tâche nécessite d'abord de résoudre les problèmes rencontrés en un point de l'espace, ou à de petites échelles spatiales.

Il est certain que ce type de modèle ne peut pas se complexifier à l'infini. Lorsque le nombre de paramètres du modèle augmente, de nombreux problèmes numériques apparaissent, principalement avec l'estimation des paramètres (lors de la minimisation de la fonction objectif). Chandler & Onof (2005) conseillent pour cette raison de privilégier les modèles comportant au plus six paramètres. De plus, les problèmes de calculs limitent le type de généralisation que l'on peut apporter (par exemple, la méthode généralisée des moments nécessite des expressions directes des moments). La généralisation à un modèle spatio-temporel et à des non-stationnarités temporelles et spatiales ne se fera alors qu'après une réflexion profonde sur les aspects les plus importants à privilégier et passera par le développement de nouvelles méthodes d'estimation des paramètres, par exemple à l'aide d'approches en plusieurs étapes (d'abord en chaque station de mesure, puis sur des régions par l'intégration de statistiques spatiales).

Conclusion

# **Bibliographie**

- ALI, M. M., MIKHAIL, N. N., & HAQ, M. S. 1978. A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8(3), 405–412.
- Anctil, F. 2008. L'eau et ses enjeux. De Boeck.
- ARNBJERG-NIELSEN, K. 2006. Significant climate change of extreme rainfall in Denmark. Water Science & Technology, 54(6-7), 1–8. doi:10.2166/wst.2006.572.
- AVISSAR, R., & LIU, Y. 1996. Three-dimensional numerical study of shallow convective clouds and precipitation induced by land surface forcing. *Journal of Geophysical Research*, **101**(D3), 7499–7518.
- BACCHI, B., BECCIU, G., & KOTTEGODA, N. T. 1994. Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology*, **155**(1-2), 225–236.
- BISSON, J., & MUTIN, G. 2000. L'eau dans le monde arabe. Carrefours de Géographie. Paris : Ellipses Édition Marketing SA.
- BLANCKAERT, W., & WILLEMS, P. 2006. Statistical analysis of clusters and cycles in the long-term historical rainfall series at Uccle. *Pages 124–128 of :* MOLNAR, P., EINFALT, T., BURLANDO, P., SEVRUK, B., & SCHMID, W. (eds), 7th extreme precipitation: Multisource data measurement and uncertainty. International Workshop on Precipitation in Urban Areas.
- BOX, G. E. P., & DRAPER, N. R. 1987. Empirical model-building and response surfaces. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons Inc.
- BROOKS, G.R., EVANS, S.G., & CLAGUE, J.J. 2001. *Flooding*. A Synthesis of Natural Geological Hazards in Canada, Geological Survey of Canada Bulletin 548. Ottawa, Ontario: G.R. Brooks (ed). Pages 101–143.
- CALENDA, G., & NAPOLITANO, F. 1999. Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation. *Journal of Hydrology*, **225**(1-2), 45–66.
- CHAMPOUX, A., & TOUTANT, C. 1988. Éléments d'hydrologie. Les éditions Le Griffon d'argile.

- CHANDLER, R. E. 1995. Parameter estimation for point process models. Technical report no. 142. Dept. of Statistical Science, University College, London.
- Chandler, R. E. 1997. A spectral method for estimating parameters in rainfall models. *Bernoulli*, 3(3), 301–322.
- CHANDLER, R. E. 2003. Moment-based inference for stochastic-mechanistic models. Internal report no. 7. DEFRA Project Improved Methods for National Spatial-Temporal Rainfall and Evaporation Modelling for BSM.
- CHANDLER, R. E., & ONOF, C. 2005. Single-site model selection and testing. Internal report no. 11. DEFRA Project Improved Methods for National Spatial-Temporal Rainfall and Evaporation Modelling for BSM.
- CHERUBINI, U., LUCIANO, E., & VECCHIATO, W. 2004. Copula methods in finance. Wiley Finance Series. Chichester: John Wiley & Sons Ltd.
- CÓRDOVA, J. R., & RODRÍGUEZ-ITURBE, I. 1985. On the probabilistic structure of storm surface runoff. Water Resources Research, 21(5), 755–763.
- COWAN, R. 1987. A bivariate exponential distribution arising in random geometry. Annals of the Institute of Statistical Mathematics, 39(1), 103–111.
- COWPERTWAIT, P. S. P. 1994. A generalized point process model for rainfall. *Proceedings of the Royal Society. London. Series A. Mathematical, Physical and Engineering Sciences*, **447**(1929), 23–37.
- COWPERTWAIT, P. S. P. 1998. A Poisson-cluster model of rainfall: high-order moments and extreme values. *Proceedings of The Royal Society of London: Series A*, **454**, 885–898.
- COWPERTWAIT, P. S. P., O'CONNELL, P. E., METCALFE, A. V., & MAWDSLEY, J. A. 1996. Stochastic point process modelling of rainfall. I. Single-site fitting and validation. *Journal of Hydrology*, 175(1-4), 17–46.
- Cox, D. R., & Isham, V. S. 1980. Point processes. London and New-York: Chapman and Hall.
- COX, D. R., & ISHAM, V. S. 1994. Statistics for the environment 2: Water related issues. Vol. 2. Chichester: Wiley. Chapter Stochastic models of precipitation, pages 3-18.
- DE MICHELE, C., & SALVADORI, G. 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *Journal of Geophysical Research*, **108**(D2). 4067, doi:10.1029/2002JD002534.
- DEVROYE, L. 1986. Nonuniform random variate generation. New York: Springer-Verlag.

BIBLIOGRAPHIE 131

DIGGLE, P. J. 1983. Statistical analysis of spatial point patterns. Mathematics in Biology. London: Academic Press Inc. [Harcourt Brace Jovanovich Publishers].

- DOWNTON, F. 1970. Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society: Series B*, **32**, 408–417.
- Duan, Q., Sorooshian, S., & Gupta, V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, **28**(4), 1015–1031.
- DUAN, Q. Y., GUPTA, V. K., & SOROOSHIAN, S. 1993. Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, **76**(3), 501–521.
- EMBRECHTS, P. 2006. Discussion of: "Copulas: tales and facts" by T. Mikosch [2006, Extremes 9(1), 3–20; MR2327842]. Extremes, 9(1), 45–47.
- EMBRECHTS, P., McNeil, A., & Straumann, D. 2002. Risk management: Value at Risk and beyond. Cambridge: Cambridge University Press.
- ENTEKHABI, D., RODRÍGUEZ-ITURBE, I., & EAGLESON, P.S. 1989. Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation. *Water Resources Research*, **25**(2), 295–302.
- Farlie, D.G.J. 1960. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47, 307–323.
- FAVRE, A.-C. 2001. Single and multi-site modelling of rainfall based on the Neyman-Scott process. Ph.D. thesis no. 2320, École Polytechnique Fédérale de Lausanne.
- FAVRE, A.-C., EL ADLOUNI, S., PERREAULT, L., THIÉMONGE, N., & BOBÉE, B. 2004a. Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, **40**. W01101, doi:10.1029/2003WR002456.
- FAVRE, A.-C., Musy, A., & Morgenthaler, S. 2004b. Unbiased parameter estimation of the Neyman-Scott model for rainfall simulation with related confidence interval. *Journal of Hydrology*, **286**(1-4), 168–178.
- FLETCHER, R. 1987. *Practical methods of optimization*. Second edn. A Wiley-Interscience Publication. Chichester: John Wiley & Sons Ltd.
- FRANK, M.J. 1979. On the simultaneous associativity of f(x,y) and x+y-f(x,y). Aequationes Mathematicae, 19, 194–226.
- FREES, E. W., & VALDEZ, E. A. 1998. Understanding relationships using copulas. *North American Actuarial Journal*, **2**(1), 1–25.

- GENEST, C., & FAVRE, A.-C. 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**(4), 347–368.
- GENEST, C., & MACKAY, R. J. 1986. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. La revue canadienne de statistique, 14(2), 145–159.
- GENEST, C., & RÉMILLARD, B. 2006. Discussion of: "Copulas: tales and facts" by T. Mikosch [2006, Extremes 9(1), 3–20; MR2327842]. Extremes, 9(1), 27–36.
- GIEC. 2007. Climate change 2007 the physical science basis: Working group I contribution to the fourth assessment report of the IPCC (climate change 2007). Cambridge University Press.
- GOEL, N. K., KUROTHE, R. S., MATHUR, B. S., & VOGEL, R. M. 2000. A derived flood frequency distribution for correlated rainfall intensity and duration. *Journal of Hydrology*, **228**(1-2), 56-67.
- GRIMALDI, S., & SERINALDI, F. 2006. Design hyetograph analysis with 3-copula function. *Hydrological Sciences Journal*, **51**(2), 223–238.
- Gumbel, E. J. 1958. Statistics of extremes. New York: Columbia University Press.
- Gumbel, E. J. 1960. Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**, 698–707.
- HALPERIN, M., & MANTEL, N. 1963. Interval estimation of non-linear parametric functions. Journal of the American Statistical Association, 58, 611–627.
- HARROD, S., & KELTON, W. D. 2006. Numerical methods for realizing nonstationary poisson processes with piecewise-constant instantaneous-rate functions. *Simulation*, **82**(3), 147–157.
- HOUGAARD, P. 1986. A class of multivariate failure time distributions. *Biometrika*, **73**(3), 671–678.
- HUNDECHA, Y., & BÁRDOSSY, A. 2005. Trends in daily precipitation and temperature extremes across western Germany in the second half of the 20th century. *International Journal of Climatology*, **25**(9), 1189–1202. doi:10.1002/joc.1182.
- HUTCHINSON, T. P., & LAI, C. D. 1990. Continuous bivariate distributions, emphasising applications. Adelaide: Rumsby Scientific Publishing.
- IWASHIMA, T., & YAMAMOTO, R. 1993. A statistical analysis of the extreme events: Long-term trend of heavy daily precipitation. *Journal of the Meteorological Society of Japan. Series II*, **71**(5), 637–640.
- JOE, H. 1997. Multivariate models and dependence concepts. Monographs on Statistics and Applied Probability, vol. 73. London: Chapman & Hall.

BIBLIOGRAPHIE 133

JOE, H. 2006. Discussion of: "Copulas: tales and facts" by T. Mikosch [2006, Extremes 9(1), 3–20; MR2327842]. Extremes, 9(1), 37–41.

- KAKOU, A. 1997. Point process based models for rainfall. Ph.D. thesis, University College, London.
- KARL, T. R., & KNIGHT, R. W. 1998. Secular trends of precipitation amount, frequency, and intensity in the United States. *Bulletin of the American Meteorological Society*, **79**(2), 231–241.
- KARL, T. R., KNIGHT, R. W., & PLUMMER, N. 1995. Trends in high-frequency climate variability in the twentieth century. *Nature*, **377**(6546), 217–220. doi:10.1038/377217a0.
- KAVVAS, M. L., & DELLEUR, J. W. 1981. A stochastic cluster model of daily rainfall sequences. Water Resources Research, 17(4), 1151–1160.
- KIM, S., & KAVVAS, M. L. 2006. Stochastic point rainfall modeling for correlated rain cell intensity and duration. *Journal of Hydrologic Engineering*, 11(1), 29–36.
- KOUTSOYIANNIS, D. 2004. Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Journal des Sciences Hydrologiques*, **49**(4), 591–610.
- KUROTHE, R. S., GOEL, N. K., & MARTHUR, B. S. 1997. Derived flood frequency distribution for negatively correlated rainfall intensity and duration. *Water Resources Research*, **33**(9), 2103–2107.
- LEE, P. A. 1981. The correlated bivariate inverted beta distribution. *Biometrical journal*, **23**(7), 693–703.
- Mailhot, A., Duchesne, S., Caya, D., & Talbot, G. 2007. Assessment of future change in intensity-duration-frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of Hydrology*, **347**(1-2), 197–210. doi:10.1016/j.jhydrol.2007.09.019.
- Mandelbrot, B. 1975. Les objets fractals. Paris : Flammarion Editeur. Forme, hasard et dimension, Nouvelle Bibliothèque Scientifique.
- MARONNA, R. A., MARTIN, R. D., & YOHAI, V. J. 2006. *Robust statistics*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd. Theory and methods.
- MASON, J. 1986. Numerical weather prediction. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 407(1832), 51–60.
- MIKOSCH, T. 2006. Copulas: tales and facts. Extremes, 9(1), 3-20.
- MITCHELL, M. 1996. An introduction to genetic algorithm. MIT press.

- MORGENSTERN, D. 1956. Einfache Beispiele Zweidimensionaler Verteilungen. Mitteilingsblatt für Mathematische Statistik, 8, 234–235.
- Nelder, J. A., & Mead, R. 1965. A simplex method for function minimization. *Computer Journal*, 7, 308–313.
- Nelsen, R. B. 2006. An introduction to copulas. Second edn. Springer Series in Statistics. New York: Springer.
- Nelsen, R. B., Quesada-Molina, J. J., & Rodríguez-Lallena, J. A. 1997. Bivariate copulas with cubic sections. *Journal of Nonparametric Statistics*, 7(3), 205–220.
- NEPPEL, L., BOUVIER, C., & NIEL, H. 2006. Quelques illustrations des sources d'incertitudes dans l'analyse de l'aléa pluvieux/ some examples of uncertainties in rainfall hazard study. *La Houille Blanche*, **6**, 22–26.
- New, M., Hewitson, B., Stephenson, D. B., Tsiga, A., Kruger, A., Manhique, A., Gomez, B., Coelho, C. A. S., Masisi, D. N., Kululanga, E., Mbambalala, E., Adesina, F., Saleh, H., Kanyanga, J., Adosi, J., Bulane, L., Fortunata, L., Mdoka, M. L., & Lajoie, R. 2006. Evidence of trends in daily climate extremes over southern and west Africa. *Journal of Geophysical Research*, 111(D14102), 147–157. doi:10.1029/2005JD006289.
- NEYMAN, J., & Scott, E. L. 1958. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B. Methodological*, **20**, 1–43.
- NORTHROP, P. 1996. Modelling and statistical analysis of spatial-temporal rainfall fields. Ph.D. thesis, University College London.
- OBEYSEKERA, J. T. B., TABIOS III, G. Q., & SALAS, J. D. 1987. On parameter estimation of temporal rainfall models. Water Resources Research, 23(10), 1837–1850.
- Onof, C. 2003. Mathematical expressions of generalised moments used in single-site rainfall models. Internal report no. 8. DEFRA Project Improved Methods for National Spatial-Temporal Rainfall and Evaporation Modelling for BSM.
- Onof, C., & Wheater, H. S. 1994. Improvements to the modelling of British rainfall using a modified Random Parameter Bartlett-Lewis Rectangular Pulses Model. *Journal of Hydrology*, **157**(1-4), 177–195.
- Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S., & Isham, V. 2000. Rainfall modelling using Poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14(6), 384–411.
- OVERNEY, O. 1997. Prédiction des crues par modélisation couplée stochastique et déterministe : méthode et analyse des incertitudes. Ph.D. thesis no. 1751, École Polytechnique Fédérale de Lausanne.

OVERNEY, O., CONSUEGRA, D., MUSY, A., LAZARO, P., BOILLAT, J-L., & SINNI-GER, R.O. 2005. Influence des changements climatiques sur le régime hydrologique et hydraulique des cours d'eau. Infoscience — Ecole Polytechnique Fédérale de Lausanne [http://de.scientificcommons.org/27998827].

- PLACKETT, R. L. 1965. A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516–522.
- PLUMMER, N., SALINGER, M.J., NICHOLLS, N., SUPPIAH, R., HENNESSY, K.J., LEIGHTON, R.M., TREWIN, B., PAGE, C.M., & LOUGH, J.M. 1999. Changes in climate extremes over the Australian region and New Zealand during the twentieth century. *Climatic Change*, **42**(1), 183–202. doi:10.1023/A:1005472418209.
- PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A., & VETTERLING, W. T. 1986. *Numerical recipes*. Cambridge: Cambridge University Press. The art of scientific computing.
- RAUCH, W., & DETOFFOL, S. 2006. Climate change induced trends in high resolution rainfall. Pages 119–123 of: MOLNAR, P., EINFALT, T., BURLANDO, P., SEVRUK, B., & SCHMID, W. (eds), 7th extreme precipitation: Multisource data measurement and uncertainty. International Workshop on Precipitation in Urban Areas.
- RENARD, B., & LANG, M. 2007. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, **30**(4), 897–912.
- RODRÍGUEZ-ITURBE, I. 1986. Scale of fluctuation of rainfall models. Water Resources Research, 22(9S), 15S-37S.
- RODRÍGUEZ-ITURBE, I., GUPTA, V. K., & WAYMIRE, E. 1984. Scale considerations in the modeling of temporal rainfall. Water Resources Research, 20(11), 1611–1619.
- RODRÍGUEZ-ITURBE, I., FEBRES DE POWER, B., & VALDÈS, J. B. 1987a. Rectangular pulses point process models for rainfall: Analysis of empirical data. *Journal of Geophysical Research*, 92(D8), 9645–9656.
- RODRÍGUEZ-ITURBE, I., COX, D. R., & ISHAM, V.S. 1987b. Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society. London. Series A. Mathematical, Physical and Engineering Sciences*, **410**(1839), 269–288.
- RODRÍGUEZ-ITURBE, I., COX, D. R., & ISHAM, V.S. 1988. A point process model for rainfall: further developments. *Proceedings of the Royal Society. London. Series A. Mathematical*, *Physical and Engineering Sciences*, **417**(1853), 283–298.
- ROUSSEEUW, P. J., & LEROY, A. M. 1987. Robust regression and outlier detection. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons Inc.

- SANSÓ, B., & GUENNI, L. 2000. A nonstationary multisite model for rainfall. *Journal of the American Statistical Association*, **95**(452), 1089–1100.
- SARMANOV, I. O. 1974. New forms of correlation relationships between positive quantities applied in hydrology. *Mathematical Models in Hydrology*, 104–109. (International Association of Hydrological Sciences, Paris).
- SCHERTZER, D., & LOVEJOY, S. 2004. Space-time complexity and multifractal predictability. *Physica A. Statistical Mechanics and its Applications*, **338**(1-2), 173–186.
- SCHERTZER, D., LOVEJOY, S., SCHMITT, F., CHIGIRINSKAYA, Y., & MARSAN, D. 1997. Multifractal cascade dynamics and turbulent intermittency. *Fractals*, 5(3), 427–471.
- SCHMIDLI, J., & Frei, C. 2005. Trends of heavy precipitation and wet and dry spells in Switzerland during the 20th century. *International Journal of Climatology*, **25**(6), 753–771. doi: 10.1002/joc.1179.
- SINGH, K., & SINGH, V. P. 1991. Derivation of bivariate probability density functions with exponential marginal. Stochastic Hydrology and Hydraulics, 5(1), 55–68.
- SINNIGER, R.O. (ed). 1997. Recherche dans le domaine des barrages crues extrêmes. Communication 5. EPFL, Laboratoire de constructions hydrauliques, Département de génie civil, Lausanne, Academic Press.
- SKLAR, A. 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris, 8, 229–231.
- SMITH, J. A., & KARR, A. F. 1985a. Parameter estimation for a model of spacetime rainfall. Water Resources Research, 26(8), 1251–1257.
- SMITH, J. A., & KARR, A. F. 1985b. Statistical inference for point process models of rainfall. Water Resources Research, 21(1), 73–79.
- STEIN, M. 1986. A simple model for spatial-temporal processes. Water Resources Research, 22(13), 2107–2110.
- STERN, R.D., & COE, R. 1984. A model fitting analysis of daily rainfall data. Proceedings of the Royal Society. London. Series A. Mathematical, Physical and Engineering Sciences, 147, 1–34.
- TRENBERTH, K.E., JONES, P.D., AMBENJE, P., BOJARIU, R., EASTERLING, D., KLEIN TANK, A., PARKER, D., RAHIMZADEH, F., RENWICK, J.A., RUSTICUCCI, M., SODEN, B., & ZHAI, P. 2007. Observations: Surface and Atmospheric Climate Change. In: Climate Change 2007: The Physical Science Basis. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Chapitre 3, pages 235–336.

BIBLIOGRAPHIE 137

VAES, G., WILLEMS, P., & BERLAMONT, J. 2002. 100 years of Belgian rainfall: are there trends? Water Science & Technology, 45(2), 55-61.

- VANMARCKE, E. 1983. Random fields. Cambridge, MA: MIT Press. Analysis and synthesis.
- Velghe, T., Troch, P. A., De Troch, F. P., & Van de Velde, J. 1994. Evaluation of cluster-based rectangular pulses point process models for rainfall. *Water Resources Research*, **30**(10), 2847–2857.
- VINCENT, L.A., & MEKIS, É. 2006. Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. *Atmosphere-Ocean*, **44**(2), 177–193.
- WAYMIRE, E., & GUPTA, V. K. 1981a. The mathematical structure of rainfall representations 1. A review of the stochastic rainfall models. Water Resources Research, 17(5), 1261–1272.
- WAYMIRE, E., & GUPTA, V. K. 1981b. The mathematical structure of rainfall representations 2. A review of the theory of point processes. *Water Resources Research*, 17(5), 1273–1285.
- WAYMIRE, E., & GUPTA, V. K. 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.
- ZHAI, P., ZHANG, X., WAN, H., & PAN, X. 2005. Trends in total precipitation and frequency of daily precipitation extremes over China. *Journal of Climate*, **18**(7), 1096–1108. doi: 10.1175/JCLI-3318.1.

## Annexe A

# Copules cubiques

Dans cet annexe, certains aspects théoriques relatifs aux copules cubiques sont exposés, en compléments du chapitre 4.

#### A.1 Conditions nécessaires

Le théorème 2 donne les conditions nécessaires pour que la fonction C(u, v), définie par la relation (4.4), soit une copule cubique.

**Théorème 2.** Soient  $\alpha, \beta : [0,1] \to \mathbb{R}$  deux fonctions satisfaisant  $\alpha(0) = \alpha(1) = \beta(0) = \beta(1) = 0$ , et soit C(u,v) une fonction définie par :

$$C(u,v) = uv + u(1-u)[\alpha(v)(1-u) + \beta(v)u], (u,v) \in [0,1]^2.$$

Alors C est une copule si et seulement si:

- $-\alpha(v)$  et  $\beta(v)$  sont absolument continues, et
- pour presque tout  $v \in [0,1]$ , le point  $\{\alpha'(v), \beta'(v)\}$  se situe dans S.

De plus, C est absolument continue.

Maintenant, limitons-nous aux copules avec des sections cubiques à la fois en u et v, c'est-à-dire les copules s'exprimant sous la forme d'un polynôme bivarié d'ordre 3. Le théorème 3 caractérise ce type de copules.

**Théorème 3.** Supposons que C ait des sections cubiques en u et en v; c'est-à-dire que pour tout  $u, v \in [0, 1]$ , C s'écrit comme suit :

$$C(u, v) = uv + u(1 - u)[\alpha(v)(1 - u) + \beta(v)u]$$

et

$$C(u,v) = uv + v(1-v)[\gamma(u)(1-v) + \delta(u)v],$$

où  $\alpha, \beta, \gamma$  et  $\delta$  sont des fonctions qui satisfont les hypothèses du théorème 2. Alors C(u, v) peut être exprimé de la même façon que dans l'équation 4.6.

Les trois sous-familles suivantes appartiennent à la classe des copules cubiques et sont citées dans Nelsen et al. (1997).

- Sarmanov Cas (i) dans Nelsen et al. (1997). La famille de Sarmanov (Sarmanov, 1974) est symétrique et inclut l'indépendance comme cas particulier.
- Frank cubique Cas (iv) dans Nelsen et al. (1997). Cette famille inclut l'indépendance pour  $\theta = 0$  et atteint les valeurs extrêmes pour  $\rho$ . Le nom que nous avons attribué à cette famille vient du fait qu'elle constitue une approximation du second ordre de la famille de Frank (voir expression 4.3.2).
- Plackett cubique Cas (v) dans Nelsen et al. (1997). Cette famille est symétrique et inclut le cas d'indépendance pour  $\theta = 0$ . Le nom que nous avons attribué à cette famille vient du fait qu'elle représente une approximation du second ordre de la famille de Plackett (voir Plackett, 1965) donnée par :

$$C(u,v) = rac{1+ heta(u+v)-\sqrt{[1+ heta(u+v)]^2-4( heta+1) heta u v}}{2 heta}$$

pour  $\theta \in [-1, 2]$ .

Ces copules cubiques ont été développées par Nelsen et~al.~(1997) (excepté pour la famille de Sarmanov).

# A.2 Minimum du tau de Kendall pour une copule avec sections cubiques en u et v

La valeur minimale du  $\tau$  de Kendall pour des copules avec des sections cubiques à la fois en u et v est la solution du problème de minimisation suivant : minimiser l'équation (4.7) où  $A_1, A_2, B_1, B_2$  remplissent les contraintes décrites dans le théorème (3). Nous notons directement que  $A_1$  et  $B_2$  sont interchangeables. Cela vaut aussi pour  $A_2$  et  $B_1$ . Cela revient donc à trouver :

$$\min_{(a,b)}(a+b)\{1/9+(a-b)/450\}$$

avec  $(a,b) \in S$ .  $(a+b)\{1/9+(a-b)/450\}$  est croissante en a et b quand  $(a,b) \in S$  de telle façon que la solution se situe sur l'ellipse  $a^2-ab+b^2-3a+3b=0$ . En posant  $b=\{a-3-\sqrt{-3a^2+6a+9}\}/2$ , a est une racine de  $-1250-1400Z+600Z^2+22Z^3+Z^4$ . Finalement, le point dans S qui minimise  $\tau$  est  $\{(-11+3\sqrt{19}-3\sqrt{-112+26\sqrt{19}})/2, (11-3\sqrt{19}-3\sqrt{-112+26\sqrt{19}})/2\}$  et nous pouvons écrire que  $|\tau| \le \sqrt{494\sqrt{19}-2053/25}$ .

# Annexe B

# Moments avec une copule cubique

Dans les sections suivantes, nous présentons les éléments complémentaires au calcul des moments du NSRPM (voir chapitre 6).

# B.1 La fonction d'autocovariance avec la durée et l'intensité des cellules dépendantes

La démonstration suivante donne le calcul de la covariance du l'intensité du processus de pluie à un décalage  $\tau$  (voir section 6.1) :

$$c_{Y}(\tau) = \text{Cov}\{Y(t), Y(t+\tau)\}\$$

$$= \text{E}\{Y(t)Y(t+\tau)\} - \text{E}\{Y(t)\} \text{E}\{Y(t+\tau)\}\$$

$$= \int_{0}^{\infty} \int_{0}^{\infty} \text{E}\{X_{t-u}(u)X_{t+\tau-v}(v)\} \text{E}\{dN(t-u)dN(t+\tau-v)\}\$$

$$- \int_{0}^{\infty} \int_{0}^{\infty} \mu_{X,L\geq u}\mu_{X,L\geq v} \text{E}\{dN(t-u)\} \text{E}\{dN(t+\tau-v)\}\$$

$$= I_{1} - I_{2}$$

Mais,

$$I_{1} = \int_{\tau}^{\infty} E\left\{X_{t+\tau-v}^{2}(v)\right\} E\left\{dN(t+\tau-v)^{2}\right\}$$

$$+ \int_{0}^{\infty} \int_{0}^{\infty} E\left\{X_{t-u}(u)X_{t+\tau-v}(v)\right\} E\left\{dN(t-u)dN(t+\tau-v)\right\}$$

$$= \int_{\tau}^{\infty} E\left\{X_{t+\tau-v}^{2}(v)\right\} E\left\{dN(t+\tau-v)\right\}$$

$$+ \int_{0}^{\infty} \int_{0}^{\infty} \mu_{X,L\geq u}\mu_{X,L\geq v} E\left\{dN(t-u)dN(t+\tau-v)\right\}$$

grâce à l'indépendance entre  $X_{t-u}(u)$  et  $X_{t+\tau-v}(v)$  quand  $\tau+u-v\neq 0$ , et  $\mathrm{d}N^2=\mathrm{d}N$ . De plus,

$$I_{2} = \int_{0}^{\infty} \int_{0}^{\infty} \delta(\tau + u - v) \mu_{X,L \geq u} \mu_{X,L \geq v} \lambda^{2} \mu_{D}^{2} du dv$$
$$+ \int_{0}^{\infty} \int_{0}^{\infty} \mu_{X,L \geq u} \mu_{X,L \geq v} \operatorname{E} \{ dN(t - u) \} \operatorname{E} \{ dN(t + \tau - v) \}$$

et  $\int_0^\infty \int_0^\infty \delta(\tau+u-v)\mu_{X,L\geq u}\mu_{X,L\geq v}\lambda^2\mu_D^2\mathrm{d}u\mathrm{d}v - 0$  en raison de la théorie de la mesure de Lebesgue. Alors

$$\begin{split} c_Y(\tau) &= \int_{\tau}^{\infty} \mathrm{E}\left\{X_{t+\tau-v}^2(v)\right\} \mathrm{E}\left\{\mathrm{d}N(t+\tau-v)^2\right\} \\ &+ \int_{0}^{\infty} \int_{0}^{\infty} \mu_{X,L\geq u} \mu_{X,L\geq v} \operatorname{Cov}\left\{\mathrm{d}N(t-u), \mathrm{d}N(t+\tau-v)\right\} \\ &= \lambda \mu_D \int_{\tau}^{\infty} \mu_{X^2,L\geq u} \mathrm{d}u + \frac{\lambda}{2} \, \mathrm{E}\left\{D(D-1)\right\} \beta \left[\int_{0}^{\infty} \int_{0}^{u+\tau} \mu_{X,L\geq u} \mu_{X,L\geq v} \, \mathrm{e}^{-\beta(u+\tau-v)} \, \mathrm{d}v \mathrm{d}u \right] \\ &+ \int_{0}^{\infty} \int_{u+\tau}^{\infty} \mu_{X,L\geq u} \mu_{X,L\geq v} \, \mathrm{e}^{-\beta(v-u-\tau)} \, \mathrm{d}v \mathrm{d}u \right] \end{split}$$

puisque  $\operatorname{Cov}\{\operatorname{d}N(t-u),\operatorname{d}N(t+\tau-v)\}$  est exprimé par (6.4) lorsque  $v< u+\tau$ . Autrement, nous avons  $\operatorname{Cov}\{\operatorname{d}N(t-u),\operatorname{d}N(t+\tau-v)\}=\operatorname{Cov}\{\operatorname{d}N(t+\tau-v),\operatorname{d}N(t-u)\}$  par symétrie de la covariance. Cela conduit à l'équation 6.5.

#### B.2 Calcul du moment du troisième ordre

Cette section fournit quelques éléments utiles à la compréhension du calcul du moment du troisième ordre du NSRPM (voir section 6.2). Puisque l'équation (6.9) n'est valide que lorsque  $t_1 < t_2 < t_3$ , différents cas doivent être considérés afin de calculer le moment du troisième ordre. Par exemple, quand  $t_1 = t_2$  et  $t_3 \neq t_1$ , le moment d'ordre trois du processus de comptage s'exprime comme :

$$E\{dN(t_1)dN(t_2)dN(t_3)\} = E\{dN(t_1)^2dN(t_3)\} = E\{dN(t_1)dN(t_3)\},$$

ce qui correspond au cas de deux cellules distinctes. Quand  $t_1$ ,  $t_2$  et  $t_3$  appartiennent à la même cellule, le moment d'ordre trois du processus de comptage s'exprime comme :

$$E\{dN(t_1)dN(t_2)dN(t_3)\} = E\{dN(t_1)^3\} = E\{dN(t_1)\} = \lambda \mu_D.$$

Nous devons considérer l'expression (6.9) en remplaçant  $t_i$  par  $t_i - u_i$  pour être en accord à l'expression (6.2). Nous intégrons  $E\{Y(t_1)Y(t_2)Y(t_3)\}$  selon l'ordre de  $t_3 - u_3$ ,  $t_2 - u_2$  et  $t_1 - u_1$ , puisque l'équation (6.9) n'est valide que lorsque  $t_1 < t_2 < t_3$ . Puisque  $t_3$  et  $t_2$  sont interchangeables dans (6.9), l'espace d'intégration peut se diviser de la manière suivante :

$$\begin{cases} t_1 - u_1 < t_2 - u_2 \\ t_1 - u_1 < t_3 - u_3 \end{cases}$$
quand 
$$\begin{cases} u_3 \in [0, t_3 - t_1 + u_1] \\ u_2 \in [0, t_2 - t_1 + u_1] \\ u_1 \in \mathbb{R}^+ \end{cases}$$

$$\begin{cases} t_2 - u_2 < t_1 - u_1 \\ t_2 - u_2 < t_3 - u_3 \end{cases} \text{ quand } \begin{cases} u_3 \in [0, t_3 - t_2 + u_2] \\ u_2 \in [t_2 - t_1 + u_1, \infty[ \\ u_1 \in \mathbb{R}^+ \end{cases}$$

$$\begin{cases} t_3 - u_3 < t_1 - u_1 \\ t_3 - u_3 < t_2 - u_2 \end{cases}$$
 quand 
$$\begin{cases} u_1 \in [0, t_1 - t_3 + u_3] \\ u_2 \in [0, t_2 - t_3 + u_3] \\ u_3 \in [t_3 - t_1, \infty[$$

Ce point est illustré dans la Figure B.1.

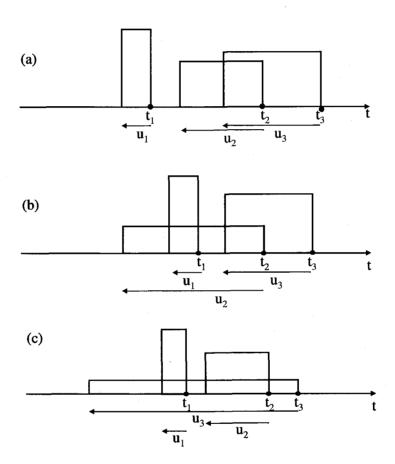


FIG. B.1 – Illustration des trois cas à distinguer lors du calcul du moment du troisième ordre du NSRPM. (a)  $t_1-u_1 < t_2-u_2$  et  $t_1-u_1 < t_3-u_3$ . (b)  $t_2-u_2 < t_1-u_1$  et  $t_2-u_2 < t_3-u_3$ . (c)  $t_3-u_3 < t_1-u_1$  et  $t_3-u_3 < t_2-u_2$ .

#### B.3 Moment du troisième ordre avec la copule FGM

Dans le moment du troisième ordre (centré) avec la copule FGM et des marges exponentielles (voir équation 6.10),  $f(\eta, \beta, \theta, h)$  et  $g(\eta, \beta, \theta, h)$  sont donnés par :

 $f(\eta,\beta,\theta,h) = -1152(\theta+4)(9\theta+16)(\beta h-1)\eta^9 + 288\theta\beta(116+21\theta)\eta^8 + 144\beta^2(180\beta\theta^2 h - 344\theta - 512 + 1040\beta\theta h + 1280\beta h - 63\theta^2)\eta^7 - 4\beta^3(8661\theta^2 + 55296 + 50552\theta)\eta^6 - 72\beta^4(1716\beta\theta h - 132\theta + 297\beta\theta^2 h + 2112\beta h - 18\theta^2 - 320)\eta^5 + 9\beta^5(4093\theta^2 + 25856 + 23520\theta)\eta^4 + 144\beta^6(260\beta\theta h - 16 + 320\beta h + 45\beta\theta^2 h - 6\theta)\eta^3 - 6\beta^7(2013\theta^2 + 12864 + 11504\theta)\eta^2 - 72\beta^9 h(\theta + 4)(9\theta + 16)\eta + \beta^9(1245\theta^2 + 8064 + 7136\theta) + 12\beta(4104\eta^4 \beta^4 \theta + 36\eta^7 \beta\theta^2 - 864\eta^6 + 2304\beta^4 \eta^4 - 864\eta^2 \beta^5 \theta - 2968\eta^6 \beta^2 \theta - 36\eta^5 \beta^3 \theta^2 + 72\eta^8 \theta^2 + 288\eta^8 \theta - 1608\beta^6 \theta \eta^2 + 54\beta^8 \theta^2 + 184\beta^8 \theta + 1206\beta^4 \eta^4 \theta^2 - 468\beta^6 \theta^2 \eta^2 - 1536\beta^2 \eta^6 + 2304\beta^4 \eta^4 - 864\eta^2 \beta^6 - 12\beta^2 \eta^7 \theta^2 h + 216\beta^4 \eta^5 \theta h - 18\beta^6 \eta^3 \theta^2 h - 144\beta^6 \eta^3 \theta h - 96\beta^2 \eta^7 \theta h + 24\beta^8 \theta h \eta + 27\beta^4 \eta^5 \theta^2 h + 3\beta^8 \theta^2 h \eta + 3072\eta^8 + 96\beta^8) \\ \times e^{-2\eta h} + 27\beta^3 \theta^2 (\beta - 2\eta)(\beta + 2\eta)(\beta - \eta)^2 (\beta + \eta)^2 e^{-4\eta h} - 32\theta\beta^3 (\beta - \eta)(\beta + 2\eta)(\beta - 2\eta)(\beta + \eta)(10\beta^2 + 6\theta\beta^2 - 28\eta^2 - 15\theta\eta^2) e^{-3\eta h} - 288\beta\eta^3 (3\theta^2 + 12\theta^2 + 1$ 

et

 $g(\eta,\beta,\theta,h) = 5(3\eta+\beta)(4\eta+\beta)(2\beta+\eta)(\beta-\eta)^2(\beta-2\eta)^2(\theta+4)(4\beta^3\theta^2\eta h - 7\beta^3\theta^2 - 56\theta\beta^3 + 32\beta^3\theta\eta h + 64\beta^3\eta h - 96\beta^3 - 21\beta^2\theta^2\eta + 12\beta^2\eta^2\theta^2 h \\ + 96\beta^2\eta^2\theta h - 168\beta^2\theta\eta - 288\beta^2\eta + 192\beta^2\eta^2 h - 18\beta\eta^2\theta^2 + 8\beta\eta^3\theta^2 h + 64\beta\eta^3\theta h - 144\beta\eta^2\theta + 128\beta\eta^3 h - 288\beta\eta^2 - 12\eta^3\theta^2 - 96\eta^3\theta - 192\eta^3) - 80 \\ \times \eta^3(\beta-\eta)(\beta-2\eta)(2\beta-4\eta-\theta\eta)(48\beta^4 + 12\beta^4\theta + 138\beta^3\theta\eta + 416\beta^3\eta + 6\theta^2\eta\beta^3 + 1168\beta^2\eta^2 + 45\theta^2\eta^2\beta^2 + 484\beta^2\theta\eta^2 + 1184\beta\eta^3 + 62\theta^2\eta^3\beta + 544 \\ \times \beta\theta\eta^3 + 24\eta^4\theta^2 + 384\eta^4 + 192\theta\eta^4) e^{-\beta h} - 4\beta^3(2\beta+\eta)(\beta-2\eta)(7\theta^3\beta^5 + 84\beta^5\theta^2 + 160\beta^5 + 240\beta^5\theta + 504\beta^4\theta^2\eta + 42\theta^3\beta^4\eta + 1440\beta^4\theta\eta + 960\beta^4 \\ \times \eta + 840\theta\beta^3\eta^2 + 288\beta^3\theta^2\eta^2 + 24\theta^3\beta^3\eta^2 + 480\eta^2\beta^3 - 4160\beta^2\eta^3 - 162\theta^3\beta^2\eta^3 - 1944\beta^2\eta^3\theta^2 - 5640\beta^2\theta\eta^3 - 5040\theta\beta\eta^4 - 1692\theta^2\beta\eta^4 - 141\theta^3\beta\eta^4 \\ - 3840\beta\eta^4 + 30\theta^3\eta^5 + 960\eta^5\theta + 360\eta^5\theta^2) e^{-2\eta h} - 240\beta^3\eta^3\theta(2\beta+\eta)(\beta-\eta)(2\beta-4\eta-\theta\eta)(2\beta+8\eta+\theta\eta) e^{-h(\beta+2\eta)} - \beta^4\theta^2(\beta+2\eta)(3\eta+\beta)(2\beta+\eta)(\beta+\eta)(\beta-\eta)^2(12+\theta) e^{-4\eta h} 8\beta^3(2+\theta)(\beta-\eta)(4\eta+\beta)(\beta+\eta)(\beta-2\eta)^2(14\theta^2\beta^2 + 140\theta\beta^2 + 320\beta^2 + 1440\beta\eta + 63\theta^2\beta\eta + 630\theta\eta\beta + 60\theta^2\eta^2 \\ + 600\theta\eta^2 + 1440\eta^2) e^{-\eta h} 960\beta^3\eta^3(2+\theta)(4\eta+\beta)(\beta-2\eta)(2\beta+6\eta+\theta\eta)(2\beta-4\eta-\theta\eta) e^{-h(\beta+\eta)} 8\beta^4\theta(\theta+10)(2+\theta)(\beta-\eta)(\beta+2\eta)(\beta-2\eta)(4\eta+\beta)(2\beta+\eta)(\beta+\eta)(\beta+\eta) e^{-3\eta h} - 40\eta^4(4\eta+\beta)(\beta+2\eta)(3\eta+\beta)(4\beta+\theta\eta+4\eta)(2\beta-4\eta-\theta\eta)^2 e^{-2\beta h} .$ 

## B.4 Formules générales des moments agrégés avec une copule cubique

Les formules générales des moments agrégés avec une copule cubique telle que définie dans le théorème (3) s'expriment de la manière suivante :

$$E(Y_i^h) = (36 + 2A_1 + A_2 + 4B_1 + 2B_2) h \lambda \mu_D / (36 \alpha \eta),$$
(B.1)

 $c_{Y}(\tau) = \lambda \mu_{D} \{ 2 e^{-3\eta \tau} (5A_{1} - 5A_{2} + 22B_{1} - 22B_{2}) - 3 e^{-2\eta \tau} (10A_{1} - 5A_{2} + 44B_{1} - 22B_{2}) + 6$   $\times e^{-\eta \tau} (36 + 5A_{1} + 22B_{1}) \} / (108\alpha^{2}\eta) + \beta \lambda E \{ D(D-1) \} \Big( e^{-3\eta \tau} (A_{1} - A_{2} + 2B_{1} - 2B_{2}) \{ A_{1} + 2(45 + A_{2} + B_{1} + 2B_{2}) \} \beta (\beta^{4} - 5\beta^{2}\eta^{2} + 4\eta^{4}) - e^{-2\eta \tau} (2A_{1} - A_{2} + 4B_{1} - 2B_{2}) (120 + 2A_{1} + 3A_{2} + 4B_{1} + 6B_{2}) \beta (\beta^{4} - 10\beta^{2}\eta^{2} + 9\eta^{4}) + 5 e^{-\eta \tau} (6 + A_{1} + 2B_{1}) \{ A_{1} + A_{2} + 2(18 + B_{1} + B_{2}) \} \beta (\beta^{4} - 13\beta^{2}\eta^{2} + 36\eta^{4}) - 30 e^{-\beta \tau} \eta [36\beta^{4} - \{468 - 24A_{1} + A_{2}^{2} - 48B_{1} + 96B_{2} + 4B_{2}^{2} + 4A_{2}(12 + B_{2}) \} \beta^{2}\eta^{2} + (36 + 2A_{1} + A_{2} + 4B_{1} + 2B_{2})^{2}\eta^{4} ] \Big) / \{ 2160\alpha^{2}\eta (\beta^{6} - 14\beta^{4}\eta^{2} + 49\beta^{2}\eta^{4} - 36\eta^{6}) \},$ 

$$Var(Y_{i}^{h}) = \lambda \mu_{D}[-7776 - 850 A_{1} - 95 A_{2} - 3740 B_{1} - 418 B_{2} + 8 e^{-3 \eta h} \{5 A_{1} - 5 A_{2} + 22 (B_{1} - B_{2})\} - 27 e^{-2 \eta h} (10 A_{1} - 5 A_{2} + 44 B_{1} - 22 B_{2}) + 216 e^{-\eta h} (36 + 5 A_{1} + 22 B_{1}) + 6 (1296 + 110 A_{1} + 25 A_{2} + 484 B_{1} + 110 B_{2}) \eta h] / \{1944 \alpha^{2} \eta^{3}\} + \lambda E\{D(D - 1)\} \Big(1080 e^{-\beta h} \eta^{3} [36 \beta^{4} - \{468 - 24 A_{1} + A_{2}^{2} - 48 B_{1} + 96 B_{2} + 4 B_{2}^{2} + 4 A_{2} (12 + B_{2})\} \beta^{2} \eta^{2} + (36 + 2 A_{1} + A_{2} + 4 \lambda_{2} + 4 \lambda_{2} (12 + B_{2})) \{2 \eta^{4}\} \{e^{\beta h} (1 - \beta h) - 1\} + 180(6 + A_{1} + 2 B_{1}) \{A_{1} + A_{2} + 2 (18 + B_{1} + B_{2})\} \beta^{2} (\beta^{5} - 13 \beta^{3} \eta^{2} + 36 \beta \eta^{4}) (e^{-\eta h} + \eta h - 1) - 9(2A_{1} - A_{2} + 4B_{1} - 2B_{2}) (120 + 2 A_{1} + 3 A_{2} + 4 B_{1} + 6 B_{2}) \beta^{2} (\beta^{5} - 10 \beta^{3} \eta^{2} + 9 \beta \eta^{4}) (e^{-2 \eta h} - 1 + 2 \eta h) + 4(A_{1} - A_{2} + 2 B_{1} - 2 B_{2}) \{A_{1} + 2 (45 + A_{2} + B_{1} + 2 B_{2})\} \beta^{2} (\beta^{5} - 5 \beta^{3} \eta^{2} + 4 \beta \eta^{4}) (e^{-3 \eta h} - 1 + 3 \eta h) / \{38880 \alpha^{2} \beta \eta^{3} (\beta^{6} - 14 \beta^{4} \eta^{2} + 49 \beta^{2} \eta^{4} - 36 \eta^{6})\},$$
(B.2)

et

$$\begin{split} &\operatorname{Cov}(Y_{i}^{h},Y_{i+k}^{h}) = \lambda \mu_{D}(\mathrm{e}^{\eta h}-1)^{2}[10A_{1}(4+8\,\mathrm{e}^{\eta h}+12\,\mathrm{e}^{2\eta h}+8\,\mathrm{e}^{3\eta h}+4\,\mathrm{e}^{4\eta h}-27\,\mathrm{e}^{\eta h(1+k)}+108\,\mathrm{e}^{2\eta h(1+k)}\\ &-54\,\mathrm{e}^{\eta h(2+k)}-27\,\mathrm{e}^{\eta h(3+k)}) - 5A_{2}\{-27\,\mathrm{e}^{\eta h(1+k)}\left(1+\mathrm{e}^{\eta h}\right)^{2}+8\left(1+\mathrm{e}^{\eta h}+\mathrm{e}^{2\eta h}\right)^{2}\} + 2\{3888\,\mathrm{e}^{2\eta h(1+k)}\\ &+297B_{2}\,\mathrm{e}^{\eta h(1+k)}\left(1+\mathrm{e}^{\eta h}\right)^{2}-88B_{2}(1+\mathrm{e}^{\eta h}+\mathrm{e}^{2\eta h})^{2}+22B_{1}(4+8\,\mathrm{e}^{\eta h}+12\,\mathrm{e}^{2\eta h}+8\,\mathrm{e}^{3\eta h}+4\,\mathrm{e}^{4\eta h}-27\\ &\times\mathrm{e}^{\eta h(1+k)}+108\,\mathrm{e}^{2\eta h(1+k)}-54\,\mathrm{e}^{\eta h(2+k)}-27\,\mathrm{e}^{\eta h(3+k)})\}]/\{3888\alpha^{2}\,\mathrm{e}^{3\eta h(1+k)}\,\eta^{3}\} +\lambda\,\mathrm{E}\{D(D-1)\}\Big(2\\ &\times\mathrm{e}^{-3\eta h(1+k)}(A_{1}-A_{2}+2B_{1}-2B_{2})\{A_{1}+2(45+A_{2}+B_{1}+2B_{2})\}\beta^{3}(\mathrm{e}^{3\eta h}-1)^{2}(\beta^{4}-5\beta^{2}\eta^{2}+4\eta^{4})\\ &-(9/2)\,\mathrm{e}^{-2\eta h(1+k)}(2A_{1}-A_{2}+4B_{1}-2B_{2})(120+2A_{1}+3A_{2}+4B_{1}+6B_{2})\beta^{3}(\mathrm{e}^{2\eta h}-1)^{2}(\beta^{4}-10\beta^{2}\\ &\times\eta^{2}+9\eta^{4})+90\,\mathrm{e}^{-\eta h(1+k)}(6+A_{1}+2B_{1})\{A_{1}+A_{2}+2(18+B_{1}+B_{2})\}\beta^{3}(\mathrm{e}^{\eta h}-1)^{2}(\beta^{4}-13\beta^{2}\eta^{2}\\ &+36\eta^{4})-540\,\mathrm{e}^{-\beta h(1+k)}\left(\mathrm{e}^{\beta h}-1\right)^{2}\eta^{3}[36\beta^{4}-\{468-24A_{1}+A_{2}^{2}-48B_{1}+96B_{2}+4B_{2}^{2}+4A_{2}(12\\ &+B_{2})\}\beta^{2}\eta^{2}+(36+2A_{1}+A_{2}+4B_{1}+2B_{2})^{2}\eta^{4}]\Big)/\{38880\alpha^{2}\beta\eta^{3}(\beta^{6}-14\beta^{4}\eta^{2}+49\beta^{2}\eta^{4}-36\eta^{6})\}. \end{split}$$

# Annexe C

# Application du NSRPM avec un lien intensité/durée des cellules

Cet annexe présente les figures permettant d'analyser la performance des NSRPM incluant une dépendance entre l'intensité et la durée des cellules (voir chapitre 8).

## C.1 Fonction objectif

Les figures suivantes présentent les valeurs de fonctions objectif minimales pour chaque modèle considéré et chaque mois, pour les applications réalisées avec les séries de pluie des stations de Brest, Lyon, Marseille, Trudeau, Miami, Seattle, Los Angeles et Minneapolis (en complément des figures de la section 8.3). Pour chaque site, nous présentons également les estimations des paramètres pour les mois de janvier et d'août.

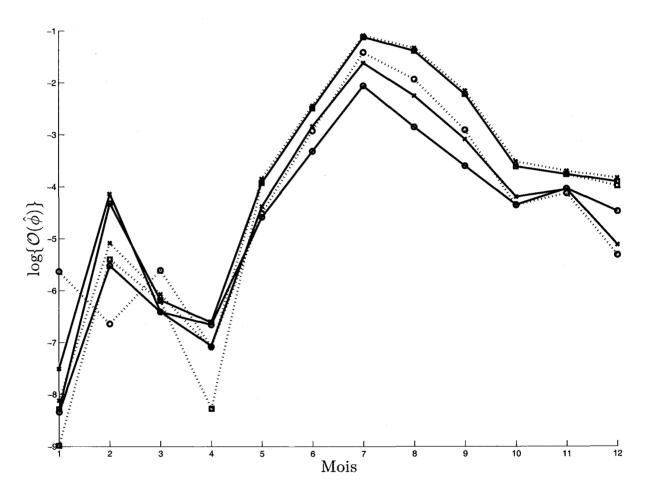


FIG. C.1 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Brest. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.1 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Brest, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponential	0.0143	1.02	0.017	12.51	0.89		0	2.52e-004
Ind. Gamma	0.0141	1.02	0.016	12.83	0.88	0.96	0	2.39e-004
Ind. Pareto	0.0140	0.99	0.016	13.20	25.55	25.40	0	5.49e-004
Famille de copules	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0142	1.10	0.017	12.82	0.93	0.38	0.09	1.44e-004
Fcub	0.0141	1.12	0.017	12.99	0.94	0.15	0.10	1.25e-004
Sym2	0.0132	1.04	0.014	17.35	0.87	0.73	-0.25	3.58e-003
Asym3	0.0144	1.08	0.017	12.17	0.88	0.00	0.06	2.98e-004
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	ĉ	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0143	1.02	0.017	12.51	1.12	0.00	-0.00	2.52e-004
DD2	0.0194	0.84	0.064	13.54	12.33	2.25	-0.34	1.31e-001

TAB. C.2 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Brest, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

all, of the de l'earboil	p) 00 105	valour.	o de la l	Olicololi	_object	$m \cup (\varphi)$	Pour I	c mois a acc
Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0171	2.72	0.034	5.96	0.34		0	2.50e-001
Ind. Gamma	0.0128	2.24	0.036	25.16	0.20	0.13	0	5.77e-002
Ind. Pareto	0.0157	1.97	0.036	9.65	3.97	3.96	0	1.05e-001
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0171	2.72	0.034	5.96	0.34	-0.01	-0.00	2.50e-001
Fcub	0.0171	2.71	0.034	5.95	0.34	-0.01	-0.00	2.50e-001
Sym2	0.0162	2.78	0.034	7.85	0.34	0.66	-0.26	1.45e-001
Asym3	0.0172	2.80	0.034	5.79	0.34	-0.07	0.04	2.64e-001
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0171	2.72	0.034	5.96	2.91	0.00	-0.00	2.50e-001
DD2	0.0137	3.87	0.036	18.07	2.86	0.00	0.58	5.28e-002

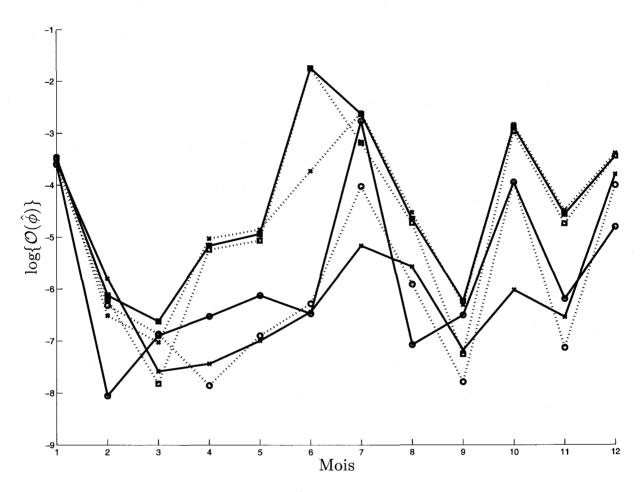


FIG. C.2 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Lyon. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.3 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Lyon, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0108	14.83	0.274	39.57	0.43		0	3.09e-002
Ind. Gamma	0.0107	16.72	0.272	70.13	0.27	0.39	0	2.76e-002
Ind. Pareto	0.0109	13.96	0.272	46.32	12.13	7.48	0	2.85e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0109	14.52	0.273	38.89	0.41	-0.21	-0.05	3.08e-002
Fcub	0.0109	9.74	0.272	33.37	0.41	-0.36	-0.25	3.10e-002
Sym2	0.0111	9.64	0.270	39.87	0.44	-0.12	-0.37	2.72e-002
Asym3	0.0109	9.77	0.272	34.87	0.55	-0.35	-0.03	3.20e-002
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0108	26.30	0.275	41.78	3.94	0.00	-0.00	3.04e-002
DD2	0.0108	19.09	0.271	71.34	24.39	2.93	0.50	2.77e-002

TAB. C.4 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Lyon, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0101	3.09	0.125	4.17	0.16		0	9.56e-003
Ind. Gamma	0.0106	3.57	0.170	10.23	0.10	0.29	0	8.47e-004
Ind. Pareto	0.0106	3.24	0.154	5.53	29.29	7.02	0	3.80e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{\alpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0101	2.98	0.126	4.11	0.15	-0.26	-0.06	9.29e-003
Fcub	0.0101	2.80	0.128	4.02	0.14	-0.22	-0.15	8.84e-003
Sym2	0.0106	3.77	0.153	6.01	0.17	1.13	-0.18	2.71e-003
Asym3	0.0100	2.99	0.123	3.96	0.14	-0.39	-0.04	1.08e-002
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0096	0.27	0.141	4.39	22.16	1.46	-0.35	9.12e-003
DD2	0.0106	4.89	0.169	9.38	14.49	0.59	0.52	8.91e-004

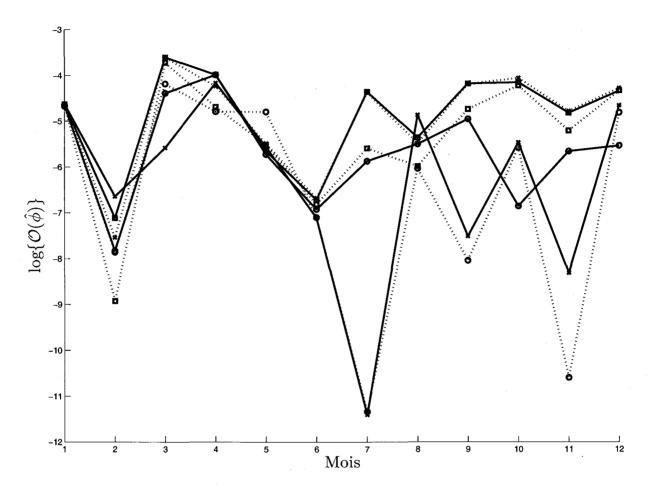


FIG. C.3 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Marseille. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.5 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Marseille, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0067	12.19	0.204	50.08	0.40		0	9.73e-003
Ind. Gamma	0.0067	10.00	0.204	54.15	0.43	0.82	0	9.54 e - 003
Ind. Pareto	0.0067	9.99	0.205	53.42	35.41	19.18	0	9.77e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0067	12.45	0.205	51.72	0.43	0.37	0.08	9.60e-003
Fcub	0.0067	9.43	0.204	48.57	0.67	0.45	0.31	9.58e-003
Sym2	0.0067	7.80	0.203	57.32	0.77	2.11	0.03	9.29e-003
Asym3	0.0067	9.77	0.204	47.22	0.49	-0.03	0.05	9.87e-003
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0067	13.35	0.205	51.77	2.70	0.03	-0.00	9.67e-003
DD2	0.0066	9.37	0.204	60.99	19.46	3.49	0.35	9.01e-003

Tab. C.6 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Marseille, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0032	2.28	0.126	3.35	0.13		0	4.68e-003
Ind. Gamma	0.0029	2.32	0.123	2.83	0.16	1.61	0	4.05e-003
Ind. Pareto	0.0032	2.23	0.131	3.69	85.10	13.83	0	7.66e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0031	2.52	0.124	3.31	0.15	1.00	0.22	3.09e-003
Fcub	0.0031	2.75	0.123	3.36	0.16	0.58	0.40	2.51e-003
Sym2	0.0032	2.80	0.124	3.54	0.16	3.00	0.34	2.39e-003
Asym3	0.0031	2.37	0.125	3.27	0.13	0.00	0.06	4.14e-003
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0030	1.96	0.124	2.97	9.20	0.18	-0.08	4.09e-003
DD2	0.0032	2.05	0.129	3.57	23.09	1.07	0.26	3.73e-003

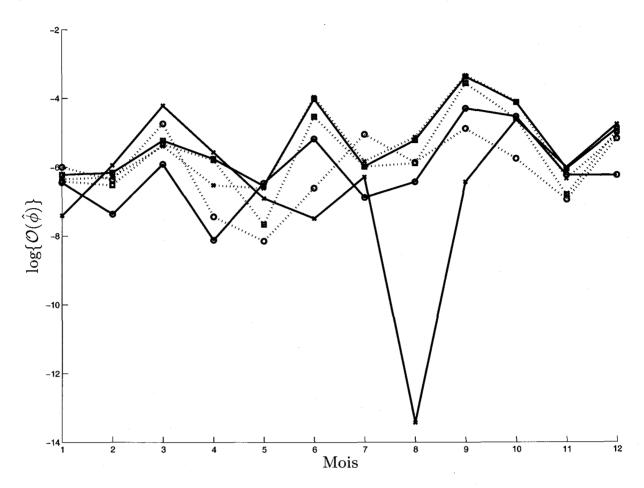


FIG. C.4 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Trudeau. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.7 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Trudeau, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0043	0.96	0.366	12.49	1.67		0	1.98e-003
Ind. Gamma	0.0044	0.86	0.363	9.04	1.50	1.09	0	1.58e-003
Ind. Pareto	0.0078	0.38	0.006	1.07	64.98	43.61	0	6.06 e-004
Famille de copules	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0043	0.99	0.359	11.35	1.65	0.44	0.10	1.82e-003
Fcub	0.0044	1.01	0.350	9.55	1.63	0.38	0.25	1.65e-003
Sym2	0.0044	0.84	0.376	12.71	1.33	0.06	-0.35	2.48e-003
Asym3	0.0043	0.92	0.366	10.51	1.52	-0.08	0.04	1.75e-003
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0043	0.96	0.366	12.49	0.60	0.00	-0.00	1.98e-003
DD2	0.0042	0.51	0.362	34.18	2.47	1.69	-0.37	6.80e-003

TAB. C.8 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Trudeau, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	$\tau$	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0161	3.48	0.205	3.11	0.13		0	5.38e-003
Ind. Gamma	0.0163	4.18	0.259	5.66	0.09	0.46	0	1.60e-003
Ind. Pareto	0.0164	3.99	0.257	4.57	35.98	6.76	0	1.48e-006
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0161	3.22	0.204	2.99	0.12	-0.44	-0.10	4.51e-003
Fcub	0.0160	2.81	0.208	2.82	0.10	-0.40	-0.27	2.76e-003
Sym2	0.0161	2.71	0.225	2.86	0.08	-0.62	-0.40	2.83e-003
Asym3	0.0160	3.20	0.197	2.86	0.11	-0.55	-0.08	5.89e-003
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0161	3.48	0.205	3.11	7.98	0.00	-0.00	5.38e-003
DD2	0.0161	6.48	0.283	8.09	25.62	0.76	0.52	1.55e-003

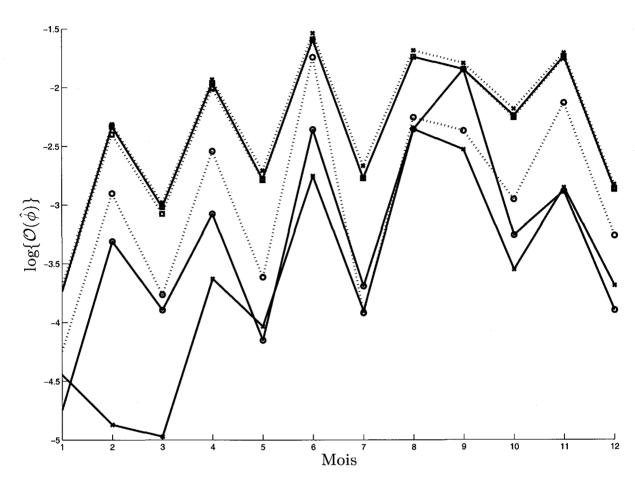


FIG. C.5 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Miami. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.9 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Miami, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0076	1.92	0.114	15.28	0.96		0	2.39e-002
Ind. Gamma	0.0074	1.64	0.114	44.57	0.63	0.19	0	8.69 e-003
Ind. Pareto	0.0078	1.59	0.116	22.06	2.00	4.41	0	1.17e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	τ	$\mathcal{O}(\hat{\phi})$
FGM	0.0076	1.89	0.114	15.13	0.95	-0.09	-0.02	2.39e-002
Fcub	0.0077	1.83	0.114	14.86	0.92	-0.08	-0.05	2.39e-002
Sym2	0.0078	1.97	0.113	19.20	0.91	0.60	-0.27	1.44e-002
Asym3	0.0076	1.95	0.115	14.79	0.92	-0.18	0.01	2.51e-002
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0072	0.18	0.116	12.78	3.22	0.74	-0.37	2.22e-002
DD2	0.0077	2.52	0.110	32.33	0.95	0.15	0.55	8.83e-003

TAB. C.10 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Miami, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0457	2.07	0.037	6.06	0.55		0	1.76e-001
Ind. Gamma	0.0375	1.72	0.037	11.14	0.38	0.33	0	9.48e-002
Ind. Pareto	0.0414	1.53	0.038	8.29	3.60	4.53	0	9.49e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0457	2.07	0.037	6.07	0.55	0.03	0.01	1.76e-001
Fcub	0.0457	2.07	0.037	6.06	0.55	0	0	1.76e-001
Sym2	0.0422	2.08	0.037	7.82	0.53	0.70	-0.26	1.05e-001
Asym3	0.0461	2.15	0.037	5.92	0.54	-0.02	0.05	1.86e-001
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0457	2.07	0.037	6.06	1.82	0.00	-0.00	1.76e-001
DD2	0.0364	2.69	0.038	13.39	1.91	0.19	0.54	8.08e-002

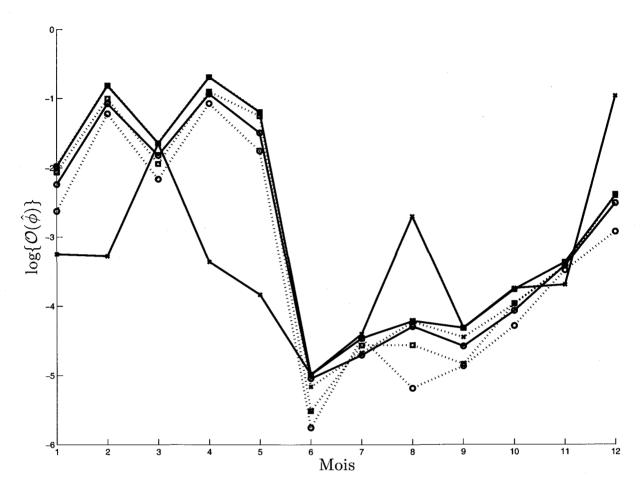


FIG. C.6 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Seattle. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.11 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Seattle, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0155	9.12	0.056	44.58	0.41		0	1.38e-001
Ind. Gamma	0.0134	5.76	0.053	86.52	0.37	0.31	0	1.06e-001
Ind. Pareto	0.0137	3.14	0.050	74.48	1.33	3.48	0	3.88e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0156	8.93	0.056	42.81	0.36	-0.44	-0.10	1.34e-001
Fcub	0.0157	8.91	0.056	41.72	0.31	-0.28	-0.19	1.27e-001
Sym2	0.0155	9.38	0.056	47.55	0.27	-0.31	-0.38	7.22e-002
Asym3	0.0157	9.09	0.057	41.75	0.36	-0.50	-0.07	1.41e-001
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0154	9.13	0.056	44.60	2.44	0.00	-0.00	1.38e-001
DD2	0.0130	5.43	0.047	82.85	2.82	0.37	0.55	9.77e-002

TAB. C.12 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Seattle, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0049	2.86	0.068	37.67	1.87		0	1.48e-002
Ind. Gamma	0.0049	2.72	0.068	49.01	1.57	0.61	0	1.36e-002
Ind. Pareto	0.0043	6.77	0.088	101.96	114.55	226.30	0	6.68e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	$\tau$	$\mathcal{O}(\hat{\phi})$
FGM	0.0050	2.54	0.068	34.45	1.66	-0.57	-0.13	1.39e-002
Fcub	0.0051	2.58	0.071	30.96	1.03	-0.58	-0.40	1.04e-002
Sym2	0.0051	2.50	0.068	37.33	1.36	-0.30	-0.38	5.59e-003
Asym3	0.0050	2.58	0.068	33.55	1.60	-0.67	-0.12	1.45e-002
modèle DD	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	ρ	$\mathcal{O}(\hat{\phi})$
DD1	0.0049	2.86	0.068	37.67	0.53	0.00	-0.00	1.48e-002
DD2	0.0048	2.83	0.066	50.28	1.16	0.78	0.42	1.51e-002

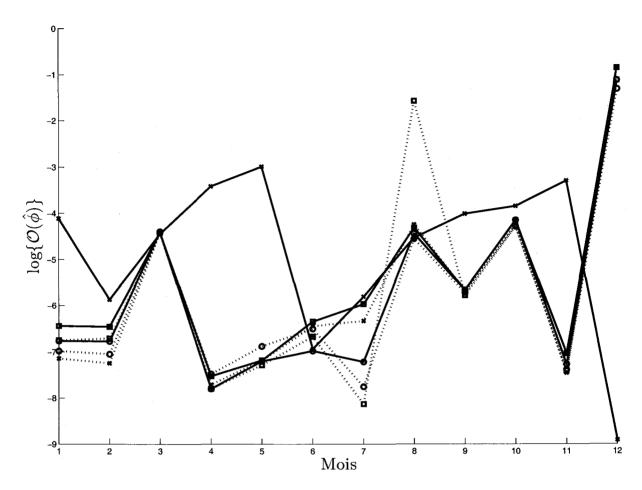


FIG. C.7 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Los Angeles. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.13 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Los Angeles, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	τ	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0043	0.59	0.053	32.10	2.08		0	1.59e-003
Ind. Gamma	0.0043	0.61	0.053	30.57	2.32	1.22	0	1.15e-003
Ind. Pareto	0.0041	1.62	0.066	188.29	33.56	146.60	0	1.63e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0044	0.50	0.053	27.96	1.87	-0.51	-0.11	1.15e-003
Fcub	0.0044	0.49	0.053	27.51	1.84	-0.20	-0.13	1.16e-003
Sym2	0.0044	0.42	0.054	26.10	1.52	-0.31	-0.38	9.28e-004
Asym3	0.0044	0.47	0.054	26.54	1.76	-0.80	-0.16	7.92e-004
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0043	0.59	0.053	32.10	0.48	0.00	-0.00	1.59e-003
DD2	0.0064	1.76	0.063	59.74	10.81	3.07	-0.22	1.24e-001

TAB. C.14 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Los Angeles, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{\alpha}$	$\hat{\gamma}$	au	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0005	20.08	0.096	64.11	0.37		0	1.33e-002
Ind. Gamma	0.0005	23.71	0.096	103.24	0.24	0.46	0	1.17e-002
Ind. Pareto	0.0006	18.29	0.096	82.23	9.31	5.94	0	1.07e-002
Famille de copules	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0005	14.17	0.095	60.45	0.46	-0.25	-0.06	1.33e-002
Feub	0.0004	0.09	0.990	105.13	65.56	-0.58	-0.40	2.07e-001
Sym2	0.0006	9.63	0.094	68.36	0.55	-0.06	-0.36	1.05e-002
Asym3	$0.0005^{\circ}$	9.77	0.094	55.92	0.65	-0.28	-0.01	1.42e-002
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0005	21.76	0.096	64.04	2.92	0.00	-0.00	1.32e-002
DD2	0.0005	15.32	0.094	102.08	16.60	2.94	0.47	1.25e-002

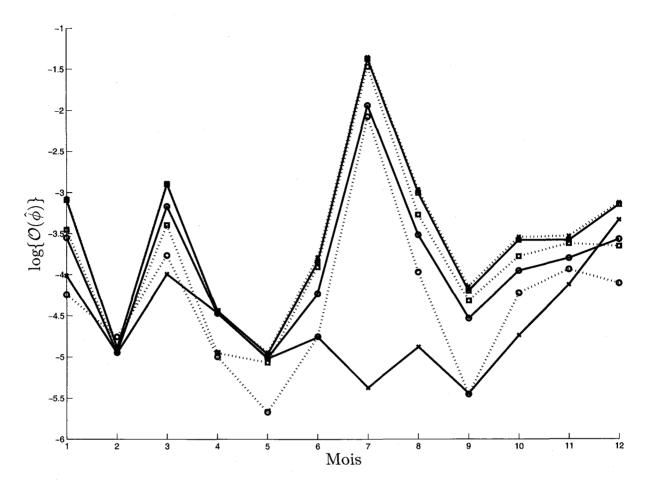


FIG. C.8 – Logarithme de la fonction objectif pour les différents modèles testés à la station de Minneapolis. Les modèles où l'intensité et la durée des cellules sont des variables aléatoires indépendantes sont indiqués en ligne pleine (avec des carrés pour la distribution exponentielle, des cercles pour la loi gamma et des croix pour la loi de Pareto). Les copules cubiques sont dessinées à l'aide d'une ligne pointillée (avec des carrés pour la copule Frank cubique, des cercles pour la copule Sym2 et des croix pour la copule Asym3).

TAB. C.15 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Minneapolis, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois de janvier.

Distribution de $X$	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	τ	$\mathcal{O}(\hat{\phi})$
Ind. Exponential	0.0071	12.90	0.107	23.32	0.51		0	4.55e-002
Ind. Gamma	0.0068	16.38	0.110	79.07	0.26	0.19	0	2.87e-002
Ind. Pareto	0.0076	9.61	0.106	32.03	3.92	4.88	0	1.81e-002
Famille de copules	$\hat{\lambda}$	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0072	13.79	0.108	22.47	0.41	-0.49	-0.11	4.31e-002
Fcub	0.0075	9.77	0.107	18.02	0.32	-0.58	-0.40	3.16e-002
Sym2	0.0075	9.77	0.106	24.34	0.46	-0.33	-0.38	1.44e-002
Asym3	0.0071	9.77	0.104	20.71	0.53	-0.60	-0.10	4.63e-002
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0071	12.90	0.107	23.32	1.96	0.00	-0.00	4.55e-002
DD2	0.0070	13.04	0.103	69.77	4.38	0.12	0.57	1.82e-002

Tab. C.16 – Estimations des paramètres du NSRPM  $\hat{\phi}$  à Minneapolis, les degrés de dépendance (tau de Kendall  $\tau$  et rho de Pearson  $\rho$ ) et les valeurs de la fonction objectif  $\mathcal{O}(\hat{\phi})$  pour le mois d'août.

Distribution de $X$	λ	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{\gamma}$	τ	$\mathcal{O}(\hat{\phi})$
Ind. Exponentiel	0.0248	2.74	0.146	8.77	0.62		0	4.95e-002
Ind. Gamma	0.0230	3.00	0.165	22.61	0.41	0.29	0	2.97e-002
Ind. Pareto	0.0241	3.59	0.189	24.44	2.08	3.67	0	7.60e-003
Famille de copules	λ	$\hat{\eta}$	$\hat{oldsymbol{eta}}$	$\hat{\mu}_D$	$\hat{lpha}$	$\hat{ heta}$	au	$\mathcal{O}(\hat{\phi})$
FGM	0.0248	2.53	0.144	8.32	0.58	-0.37	-0.08	4.82e-002
Fcub	0.0257	9.77	0.203	13.19	0.15	-0.58	-0.40	3.80e-002
Sym2	0.0250	3.02	0.162	11.51	0.49	-0.06	-0.36	1.89e-002
Asym3	0.0249	2.54	0.143	7.98	0.56	-0.49	-0.07	5.15e-002
modèle DD	$\hat{\lambda}$	$\hat{\eta}$	$\hat{eta}$	$\hat{\mu}_D$	$\hat{f}$	$\hat{c}$	$\rho$	$\mathcal{O}(\hat{\phi})$
DD1	0.0248	2.74	0.146	8.77	1.61	0.00	-0.00	4.95e-002
DD2	0.0225	6.22	0.183	39.13	2.66	0.00	0.58	1.98e-002

### C.2 Dépendance

Les figures suivantes rapportent les degrés de dépendance (mesuré par le tau de Kendall) entre la durée et l'intensité des cellules pour chaque application et pour les copules Frank cubique (Fcub), Asym3 et FGM (en complément des figures de la section 8.4). Pour chaque structure de dépendance, les résultats sont présentés en deux parties, la première étant liée aux applications des modèles en Europe, la seconde en Amérique du Nord.

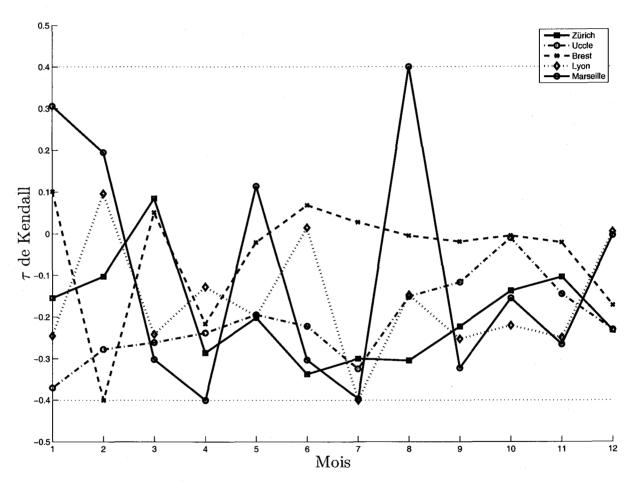


FIG. C.9 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule Fcub pour les stations de Zürich, Uccle, Brest, Lyon et Marseille. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

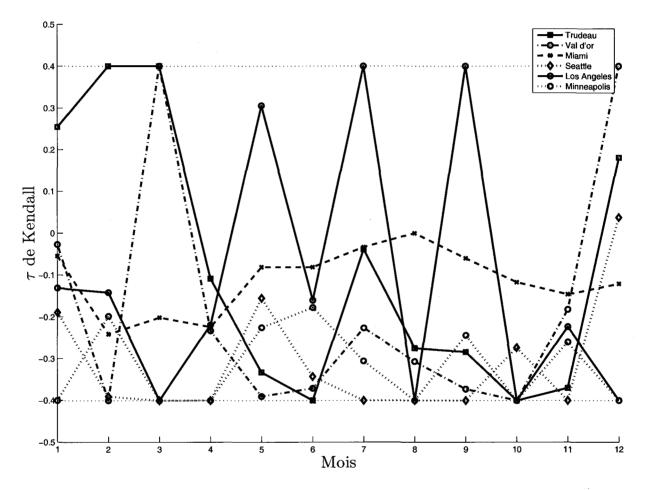


Fig.~C.10 — Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule Fcub pour les stations de Trudeau, Val d'Or, Miami, Seattle, Los Angeles et Minneapolis. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

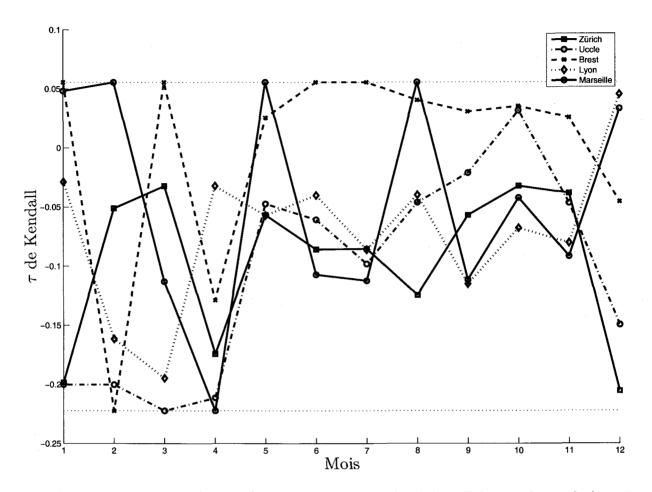


Fig. C.11 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule Asym3 pour les stations de Zürich, Uccle, Brest, Lyon et Marseille. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

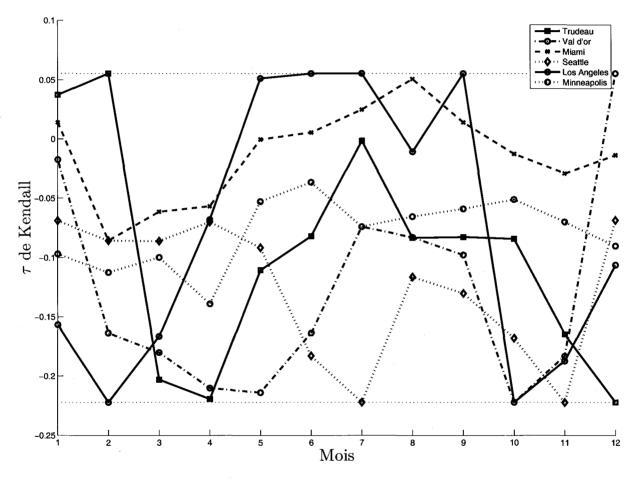


FIG. C.12 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule Asym3 pour les stations de Trudeau, Val d'Or, Miami, Seattle, Los Angeles et Minneapolis. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

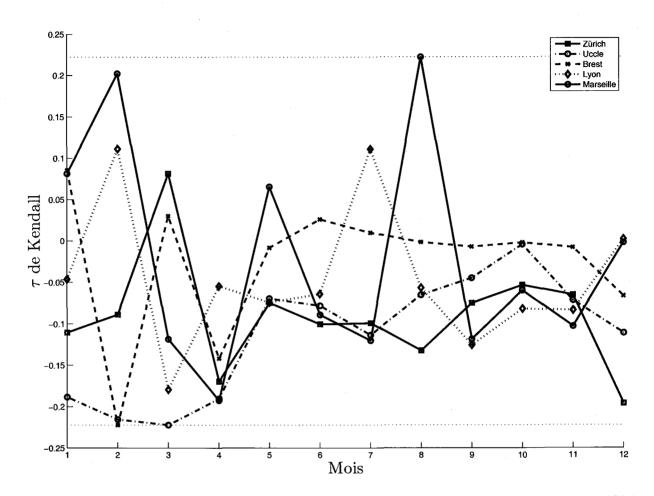


FIG. C.13 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule FGM pour les stations de Zürich, Uccle, Brest, Lyon et Marseille. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

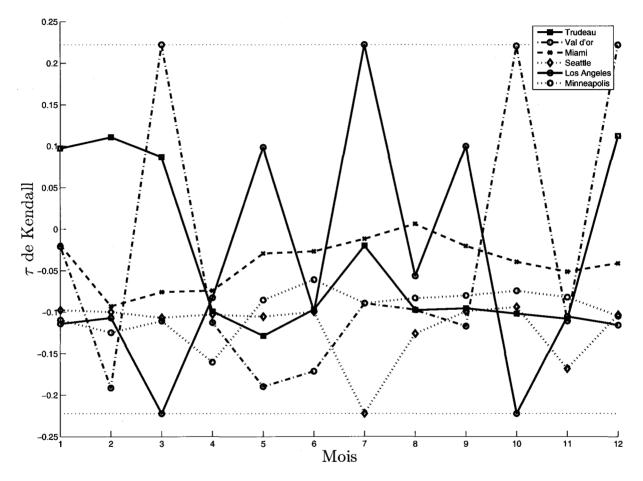


Fig. C.14 – Tau de Kendall (mensuel) entre l'intensité et la durée des cellules avec la copule FGM pour les stations de Trudeau, Val d'Or, Miami, Seattle, Los Angeles et Minneapolis. Des lignes en pointillé indiquent les bornes du tau de Kendall pour cette copule.

### C.3 Extrêmes

Les figures suivantes illustrent la reproduction des valeurs extrêmes d'intensité pluvieuse aux stations de Brest, Lyon, Marseille, Trudeau, Miami, Seattle, Los Angeles et Minneapolis (en complément des figures de la section 8.5).

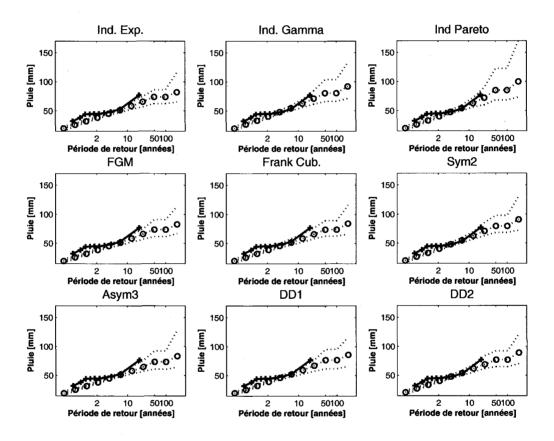


FIG. C.15 – Maxima annuels journaliers à Brest. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

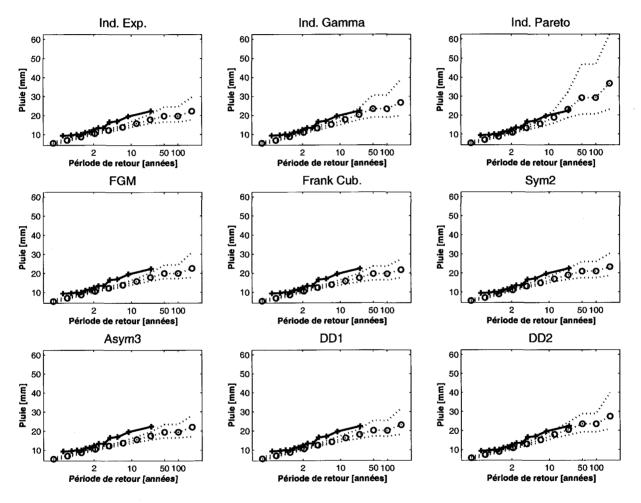


FIG. C.16 – Maxima annuels horaires à Brest. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

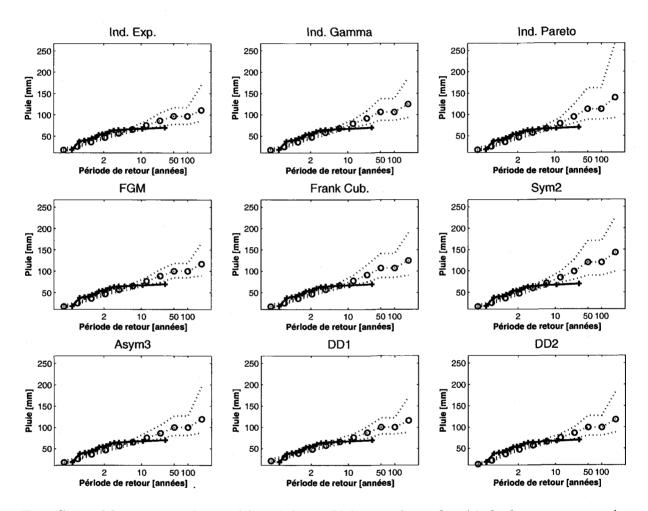


Fig. C.17 – Maxima annuels journaliers à Lyon. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

173

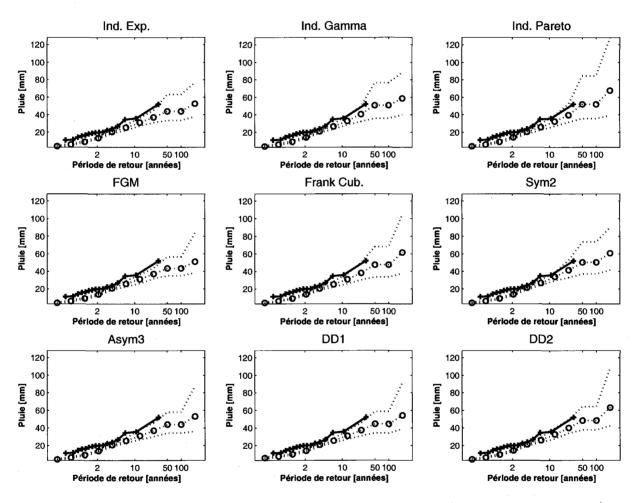


FIG. C.18 – Maxima annuels horaires à Lyon. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

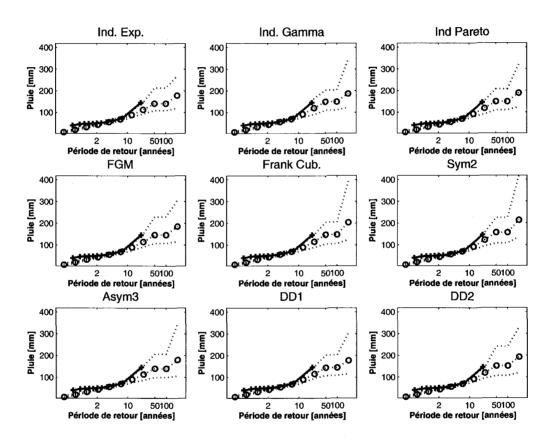


Fig. C.19 – Maxima annuels journaliers à Marseille. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

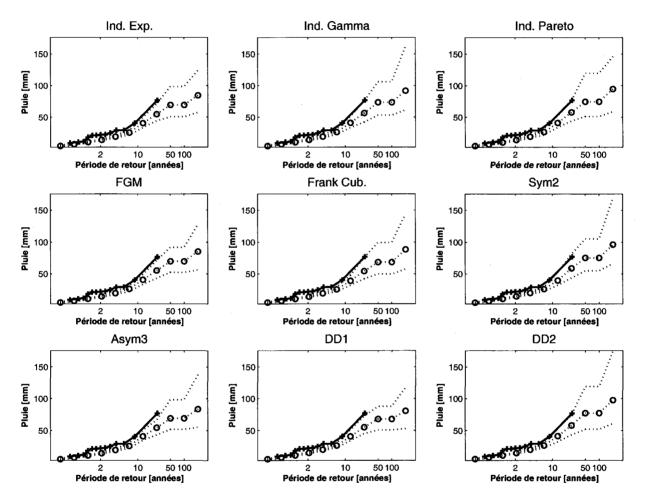


Fig. C.20 – Maxima annuels horaires à Marseille. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

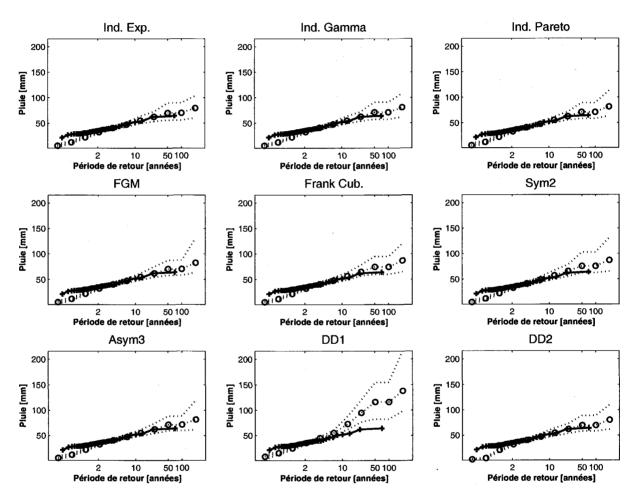


FIG. C.21 – Maxima annuels journaliers à Val d'Or. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

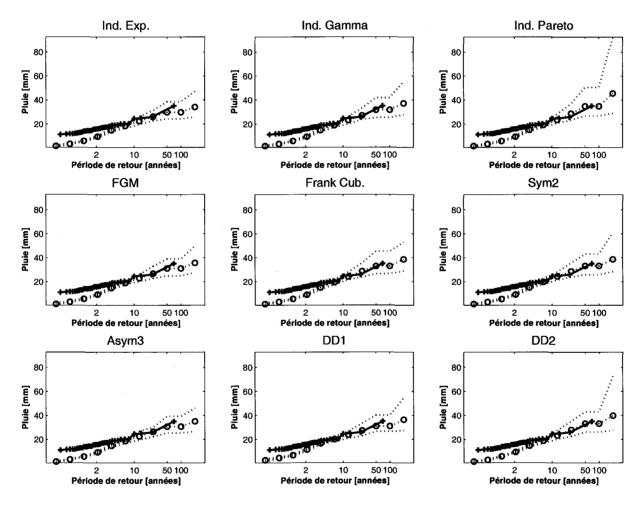


FIG. C.22 – Maxima annuels horaires à Val d'Or. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

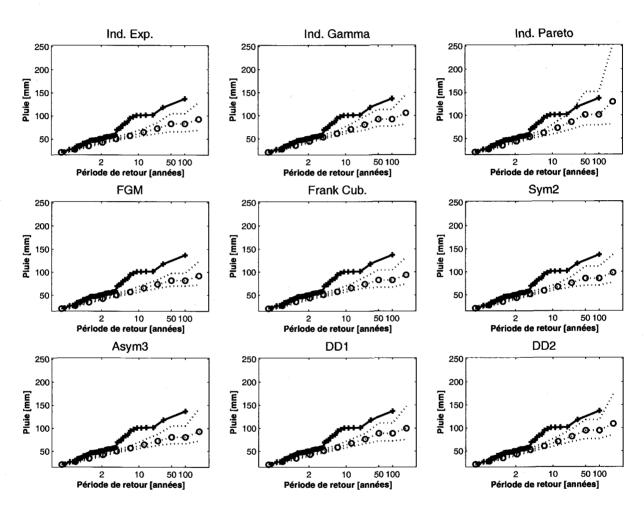


FIG. C.23 – Maxima annuels journaliers à Miami. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

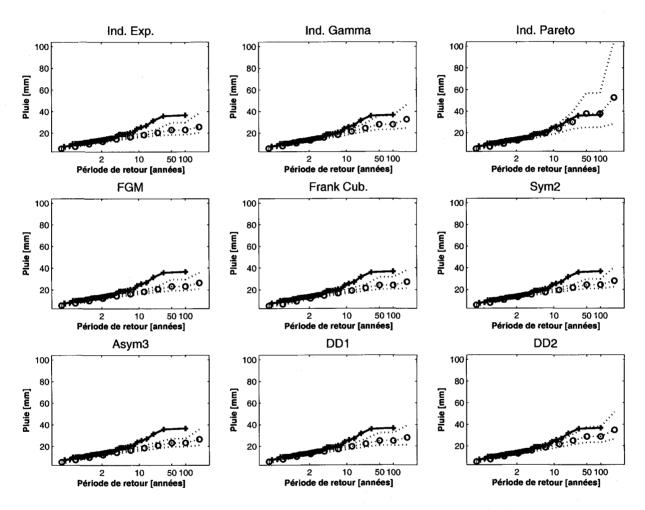


FIG. C.24 – Maxima annuels horaires à Miami. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

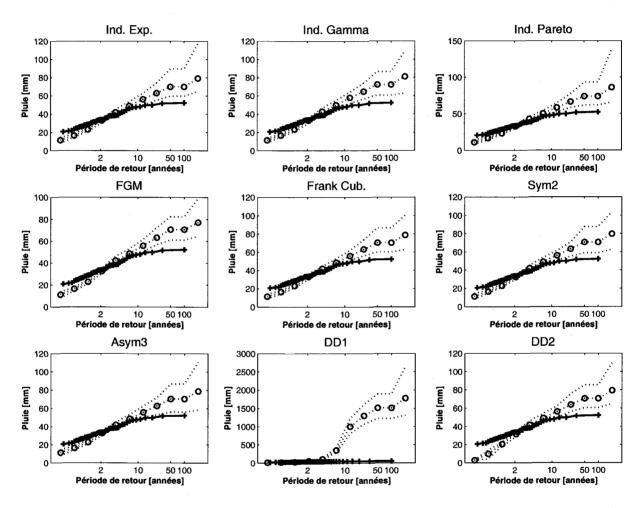


FIG. C.25 – Maxima annuels journaliers à Seattle. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

C.3 Extrêmes 181

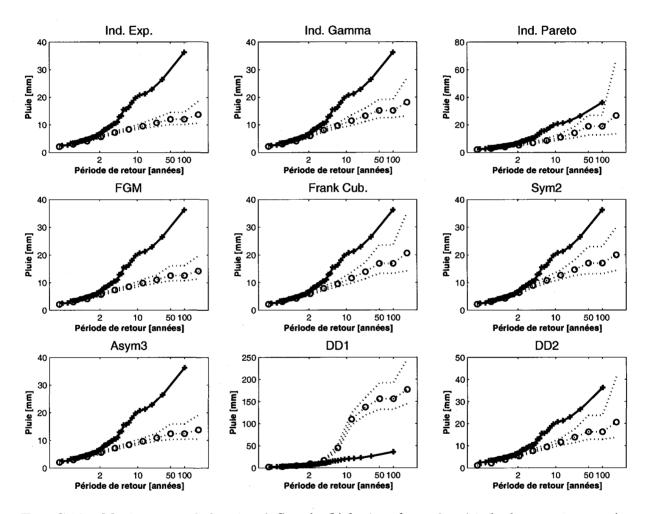


FIG. C.26 – Maxima annuels horaires à Seattle. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

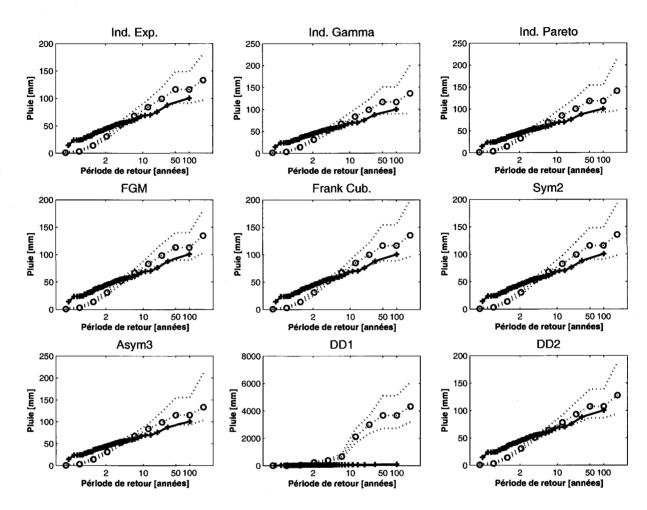


FIG. C.27 – Maxima annuels journaliers à Los Angeles. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

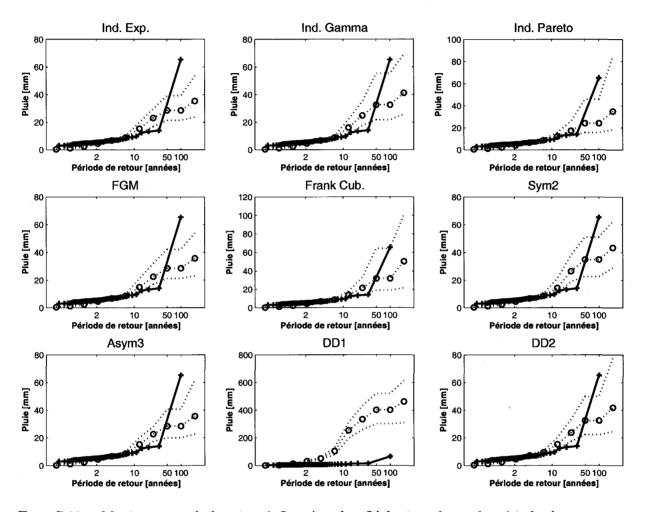


Fig. C.28 – Maxima annuels horaires à Los Angeles. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

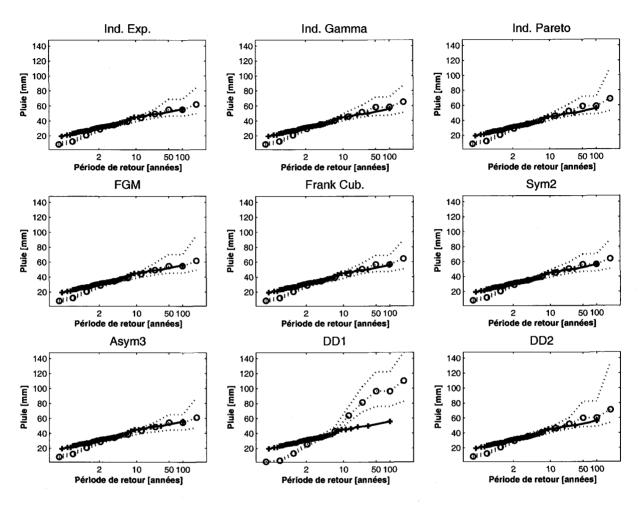


FIG. C.29 – Maxima annuels journaliers à Minneapolis. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

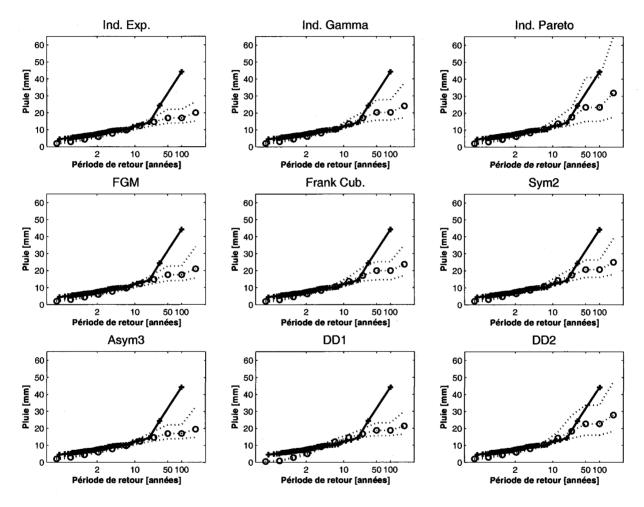


Fig. C.30 – Maxima annuels horaires à Minneapolis. L'abscisse donne la période de retour en années sur une échelle de Gumbel. La courbe en trait continu avec les symboles + représente les maxima observés. Les courbes en trait pointillé représentent la moyenne (avec o) et un intervalle de confiance empirique à 95% des maxima simulés obtenus à partir de 100 simulations de 100 ans.

## Annexe D

# Moments agrégés du processus non-stationnaire

Cet annexe donne les formules impliquées dans les expressions de la variance et de la covariance de l'intensité de pluie agrégée pour le NSRPM non-stationnaire (voir chapitre 9).

### D.1 Variance

Dans la formule (9.9) de la variance du processus non-stationnaire agrégé,  $f_v(\epsilon, \eta, \beta)$  et  $g_v(\epsilon, \eta, \beta)$  sont données par :

 $f_{v}(\epsilon,\eta,\beta,h,i) = 352\eta^{2}\epsilon\beta^{6} - 96\eta\beta^{8}\epsilon h - 320\eta^{3}\epsilon\beta^{5} - 224\eta^{5}\epsilon\beta^{3} + 544\eta\beta^{7}\epsilon - 32\eta^{6}\epsilon\beta^{2} + 176\eta^{4}\beta^{6}h + 272\eta^{3}\beta^{7}h + 96\eta^{2}$   $\times\beta^{8}h - 160\eta^{5}\beta^{5}h - 16\eta^{8}\beta^{2}h - 256\eta^{6}\beta^{4}h - 112\eta^{7}\beta^{3}h + 160\beta^{5}\eta^{4} + 16\eta^{7}\beta^{2} - 96\beta^{8}\eta + 256\eta^{5}\beta^{4} - 176\beta^{6}\eta^{3} - 272\beta^{7}$   $\times\eta^{2} + 8\eta^{8}\epsilon\beta^{2}h^{2} - 16\eta^{8}\epsilon\beta^{2}ih^{2} - 512\eta^{4}\epsilon\beta^{4} - 256\eta^{6}\epsilon\beta^{4}ih^{2} - 112\eta^{7}\epsilon\beta^{3}ih^{2} + 112\eta^{6}\epsilon\beta^{3}h + 128\eta^{6}\epsilon\beta^{4}h^{2} + 256\eta^{5}\epsilon\beta^{4}$   $\times h + 80\eta^{5}\epsilon\beta^{5}h^{2} - 160\eta^{5}\epsilon\beta^{5}ih^{2} + 160\eta^{4}\epsilon\beta^{5}h - 88\eta^{4}\epsilon\beta^{6}h^{2} + 176\eta^{4}\epsilon\beta^{6}ih^{2} + 272\eta^{3}\epsilon\beta^{7}ih^{2} - 176\eta^{3}\epsilon\beta^{6}h - 136\eta^{3}\epsilon$   $\times\beta^{7}h^{2} + 96\eta^{2}\epsilon\beta^{8}ih^{2} - 272\eta^{2}\epsilon\beta^{7}h + 16\eta^{7}\epsilon\beta^{2}h - 48\eta^{2}\epsilon\beta^{8}h^{2} + 192\epsilon\beta^{8} - 96\eta\epsilon\beta^{8}ih - 272\eta^{2}\epsilon\beta^{7}ih + 112\eta^{6}\epsilon\beta^{3}ih + 16$   $\times\eta^{7}\epsilon\beta^{2}ih + 112\eta^{6}\beta^{3} - 176\eta^{3}\epsilon\beta^{6}ih + 256\eta^{5}\epsilon\beta^{4}ih + 160\eta^{4}\epsilon\beta^{5}ih + 56\eta^{7}\epsilon\beta^{3}h^{2} + (272\beta^{7}\eta^{2} + 96\beta^{8}\eta - 192\epsilon\beta^{8} - 160$   $\times\beta^{5}\eta^{4} + 176\beta^{6}\eta^{3} - 256\beta^{4}\eta^{5} - 96\eta\epsilon\beta^{8}h - 256\eta^{5}\epsilon\beta^{4}ih + 96\eta\epsilon\beta^{8}ih - 544\eta\epsilon\beta^{7} + 176\eta^{3}\epsilon\beta^{6}ih - 160\eta^{4}\epsilon\beta^{5}ih + 256\eta^{5}$   $\times\epsilon\beta^{4}h - 352\eta^{2}\epsilon\beta^{6} + 320\eta^{3}\epsilon\beta^{5} + 272\eta^{2}\epsilon\beta^{7}ih + 160\eta^{4}\epsilon\beta^{5}h + 512\eta^{4}\epsilon\beta^{4} - 176\eta^{3}\epsilon\beta^{6}h - 272\eta^{2}\epsilon\beta^{7}h + 224\eta^{5}\beta^{3}\epsilon + 112$   $\times\eta^{6}\beta^{3}\epsilon h + 16\eta^{7}\beta^{2}\epsilon h - 16\eta^{7}\beta^{2}\epsilon ih - 112\eta^{6}\beta^{3}\epsilon ih + 32\eta^{6}\beta^{2}\epsilon - 16\beta^{2}\eta^{7} - 112\beta^{3}\eta^{6}) e^{-\eta h} + \epsilon(256\eta^{5}\beta^{4}h - 176\eta^{2}\beta^{6} + 112\eta^{5}\beta^{3} - 272\eta\beta^{7} + 160\eta^{3}\beta^{5} + 256\eta^{4}\beta^{4} - 96\beta^{8} - 272\eta^{2}\beta^{7}h + 112\eta^{6}\beta^{3}h - 96\eta\beta^{8}h + 160\eta^{4}\beta^{5}h - 176\eta^{3}\beta^{6}h + 16$   $\times\eta^{7}\beta^{2}h + 16\eta^{6}\beta^{2}) e^{-\eta ih} + \epsilon(272\eta\beta^{7} - 112\eta^{5}\beta^{3} + 176\eta^{2}\beta^{6} - 160\eta^{3}\beta^{5} - 26\eta^{4}\beta^{4} + 96\beta^{8}) e^{-\eta (i - 1)h}$ 

et

 $g_v(\epsilon,\eta,\beta,h,i) = 52\eta^3\epsilon\beta^5 - 102\eta^5\epsilon\beta^3 + 112\eta^2\epsilon\beta^6 + 92\eta\beta^7\epsilon - 42\eta^7\epsilon\beta - 102\eta^6\epsilon\beta^2 + 44\eta^4\beta^6h + 68\eta^3\beta^7h + 24\eta^2\beta^8h + 64\eta^3\beta^7h + 24\eta^3\beta^8h + 64\eta^3\beta^7h +$  $-40\eta^5\beta^5h - 4\eta^8\beta^2h - 64\eta^6\beta^4h - 28\eta^7\beta^3h - 4\beta^5\eta^4 + 28\eta^7\beta^2 - 24\beta^8\eta + 64\eta^5\beta^4 - 68\beta^6\eta^3 - 68\beta^7\eta^2 + 2\eta^8\epsilon\beta h + 2\eta^5\beta^5h - 4\eta^8\beta^2h - 64\eta^6\beta^4h - 28\eta^7\beta^3h - 4\beta^5\eta^4 + 28\eta^7\beta^2 - 24\beta^8\eta + 64\eta^5\beta^4 - 68\beta^6\eta^3 - 68\beta^7\eta^2 + 2\eta^8\epsilon\beta h + 2\eta^8\eta^2h - 2\eta^8\eta^2h$  $\times \eta^8 \epsilon \beta^2 h^2 - 4 \eta^8 \epsilon \beta^2 i h^2 - 28 \eta^4 \epsilon \beta^4 + 4 \eta^8 \beta - 64 \eta^6 \epsilon \beta^4 i h^2 - 28 \eta^7 \epsilon \beta^3 i h^2 + 34 \eta^6 \epsilon \beta^3 h + 32 \eta^6 \epsilon \beta^4 h^2 + 30 \eta^5 \epsilon \beta^4 h + 20 \eta^6 \epsilon \beta^4 h^2 + 30 \eta^6 \epsilon \beta^4 h^2 h^2 + 30 \eta^6 \epsilon \beta^4 h^2 h^2 + 30 \eta^6 \epsilon \beta^4 h^2 h^2 h^2 + 30 \eta^6 h$  $\times \beta^8 i h^2 - 24 \eta^2 \epsilon \beta^7 h - 6 \eta^8 \epsilon + 14 \eta^7 \epsilon \beta^2 h - 12 \eta^2 \epsilon \beta^8 h^2 + 24 \epsilon \beta^8 - 24 \eta \epsilon \beta^8 i h - 68 \eta^2 \epsilon \beta^7 i h + 68 \eta^6 \epsilon \beta^3 i h + 28 \eta^7 \epsilon \beta^2 i h$  $+68\eta^{6}\beta^{3}-68\eta^{3}\epsilon\beta^{6}ih+64\eta^{5}\epsilon\beta^{4}ih+4\eta^{8}\epsilon\beta ih-4\eta^{4}\epsilon\beta^{5}ih+14\eta^{7}\epsilon\beta^{3}h^{2}+(68\beta^{7}\eta^{2}+24\beta^{8}\eta-24\epsilon\beta^{8}+28\beta^{5}\eta^{4}+68\eta^{2}+6\eta^{2}+6\eta^$  $\times \beta^6 \eta^3 + 4\beta^4 \eta^5 - 24\eta\epsilon\beta^8 h + 4\eta^5 \epsilon\beta^4 ih + 24\eta\epsilon\beta^8 ih - 92\eta\epsilon\beta^7 + 68\eta^3 \epsilon\beta^6 ih + 28\eta^4 \epsilon\beta^5 ih - 4\eta^5 \epsilon\beta^4 h - 112\eta^2 \epsilon\beta^6 - 52\eta^4 \epsilon\beta^6 h + 2\eta^4 h + 2\eta^4 \epsilon\beta^6 h + 2\eta^4 h + 2\eta^4$  $\times \eta^3 \epsilon \beta^5 + 68 \eta^2 \epsilon \beta^7 ih - 28 \eta^4 \epsilon \beta^5 h - 8 \eta^4 \epsilon \beta^4 - 68 \eta^3 \epsilon \beta^6 h - 68 \eta^2 \epsilon \beta^7 h) \, \mathrm{e}^{-\eta h} + (28 \eta^7 \epsilon h - 4 \eta^8 \beta - 68 \eta^6 \beta^3 - 68 \eta^5 \beta^4 + 68 \eta^6 \beta^3 - 68 \eta^6 \beta^4 - 68 \eta^6 \beta^4$  $-24\eta^4\beta^5+6\eta^8\epsilon-28\eta^7\beta^2-4\eta^8\epsilon ih\beta+102\eta^5\epsilon\beta^3+68\eta^5\epsilon h\beta^4-68\eta^6\epsilon ih\beta^3+68\eta^6\epsilon h\beta^3-68\eta^5\epsilon ih\beta^4-24\eta^4\epsilon\beta^5 ih\beta^4-6\eta^6\epsilon h\beta^3+6\eta^6\epsilon h\beta^3+6\eta^6\epsilon h\beta^3+6\eta^6\epsilon h\beta^4+6\eta^6\epsilon h\beta^3+6\eta^6\epsilon h\beta^4+6\eta^6\epsilon h\beta^3+6\eta^6\epsilon h\beta^4+6\eta^6\epsilon h$  $-28\eta^{7}\epsilon ih\beta^{2}+24\eta^{4}\epsilon\beta^{5}h+4\eta^{8}\epsilon h\beta+36\eta^{4}\epsilon\beta^{4}+42\eta^{7}\epsilon\beta+102\eta^{6}\epsilon\beta^{2})\operatorname{e}^{-\beta h}+\epsilon(2\eta^{8}+34\eta^{6}\beta^{2}+14\eta^{7}\beta+34\eta^{5}\beta^{3}+34\eta^{6}\beta^{2}+14\eta^{7}\beta+34\eta^{6}\beta^{2}+34\eta^{6}\beta^{2$  $+12\eta^{4}\beta^{4})\,\mathrm{e}^{-\beta h(2i-1)} + \epsilon(24\beta^{8} - 24\eta\beta^{8}h + 4\eta^{5}\beta^{4}ih - 4\eta^{5}\beta^{4}h + 28\eta^{4}\beta^{5}ih - 28\eta^{4}\beta^{5}h - 68\eta^{3}\beta^{6}h + 68\eta\beta^{7} + 4\eta^{4}\beta^{6}h + 6\eta\beta^{7} + 4\eta^{6}\beta^{6}h + 6\eta\beta^{7} + 4\eta\beta^{6}h + 6\eta\beta^{7} + 4\eta\beta^{7} + 4\eta\beta^{7$  $\times \beta^4 + 68\eta^2 \beta^7 i h + 68\eta^3 \beta^6 i h + 24\eta \beta^8 i h - 68\eta^2 \beta^7 h + 28\eta^3 \beta^5 + 68\eta^2 \beta^6) \, \mathrm{e}^{\eta h (1 - 2i)} \\ - \epsilon \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 7\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 7\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^3 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^5 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^5 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 18\eta^5 \beta^6) \, \mathrm{e}^{-2\beta i h} (\eta^8 + 17\eta^5 \beta^3 + 17\eta^7 + 17\eta^$  $\times \beta + 6\eta^4 \beta^4 + 17\eta^6 \beta^2)(1 + \mathrm{e}^{2\beta h}) + \epsilon (5\eta^3 \beta^5 - 6\beta^8 + 34\eta^2 \beta^7 h + 12\eta\beta^8 h - 5\eta\beta^7 + 5\eta^2 \beta^6 - 12\eta\beta^8 ih + \eta^4 \beta^4 + 2\eta^5 \beta^4 h + 12\eta\beta^8 h - 5\eta\beta^7 + 5\eta\beta^8 h + 12\eta\beta^8 h$  $\times h - 2\eta^5\beta^4ih - 14\eta^4\beta^5ih - 34\eta^2\beta^7ih + 34\eta^3\beta^6h - 34\eta^3\beta^6ih + 14\eta^4\beta^5h) e^{-2\eta(i-1)h} - \epsilon(18\beta^8 + 34\eta^3\beta^6ih + 3$  $\times \eta^{2}\beta^{7}ih + 2\eta^{5}\beta^{4}ih + 14\eta^{4}\beta^{5}ih + 12\eta\beta^{8}ih + \eta^{4}\beta^{4} + 9\eta^{3}\beta^{5} + 39\eta\beta^{7} + 29\eta^{2}\beta^{6}) e^{-2\eta ih} + \epsilon(28\eta^{6}\beta^{3}h + 10\eta^{6}\beta^{2} - 28\eta^{6}\beta^{4}h + 12\eta\beta^{6}\beta^{2} + 2\eta\beta^{6}\beta^{6}h + 12\eta\beta^{6}\beta^{6}h +$  $\times \eta^{6}\beta^{3}ih + 68\eta^{4}\beta^{5}h + 68\eta^{5}\beta^{4}h + 4\eta^{7}\beta^{2}h - 4\eta^{7}\beta^{2}ih - 68\eta^{5}\beta^{4}ih + 6\eta^{5}\beta^{3} - 34\eta^{4}\beta^{4} + 2\eta^{7}\beta + 24\eta^{3}\beta^{6}h - 24\eta^{3}\beta^{6}ih$  $-24\eta^{2}\beta^{6} - 56\eta^{3}\beta^{5} - 68\eta^{4}\beta^{5}ih) \, \mathrm{e}^{\{-\eta(i-1) - \beta i\}h} \, - \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 14\eta^{5}\beta^{3} + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{4}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{6}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{6}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{6}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{6}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{6}\beta^{4}) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i-2)} + \epsilon(2\eta^{7}\beta + 10\eta^{6}\beta^{2} + 6\eta^{6}\beta^{2}) \, \mathrm{e}^{-\eta h(i-1) - \eta h(i-1) + \eta h(i-1) - \eta h(i-1) - \eta h(i-1) - \eta h(i-1) - \eta h($  $+\epsilon(8\eta^6\beta^2+10\eta^5\beta^3+4\eta^4\beta^4+2\eta^7\beta)\operatorname{e}^{-\eta h(i-1)-\beta h(3i-3)}-\epsilon(4\eta^4\beta^4+44\eta^2\beta^6+24\eta\beta^7+24\eta^3\beta^5)\operatorname{e}^{-\eta h(2i-1)}$  $+\epsilon(2\eta^{4}\beta^{4}+4\eta^{5}\beta^{3}+2\eta^{6}\beta^{2})\operatorname{e}^{-(\eta+3\beta)ih}+\epsilon(24\eta^{4}\beta^{4}-4\eta^{7}\beta^{2}h+24\eta^{4}\beta^{5}ih+44\eta^{5}\beta^{4}ih-44\eta^{5}\beta^{4}h+24\eta^{6}\beta^{3}ih+44\eta^{5}\beta^{4}h+24\eta^{6}\beta^{3}ih+44\eta^{5}\beta^{4}h+24\eta^{6}\beta^{3}ih+44\eta^{5}\beta^{4}h+24\eta^{6}\beta^{3}ih+44\eta^{5}\beta^{4}h+24\eta^{6}\beta^{3}ih+44\eta^{5}\beta^{4}h+24\eta^{6}\beta^{3}ih+4\eta^{5}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{5}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{5}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{5}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{5}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{6}\beta^{3}ih+4\eta^{6}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{6}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{6}\beta^{4}h+2\eta^{6}\beta^{3}ih+4\eta^{6}\beta^{4}h+2\eta^{6}\eta^{6}h+2\eta^{6}\eta^{6}h+2\eta^{6}\eta^{6}h+2\eta^{6}\eta^{6}h+2\eta^{6}\eta^{6}h+2\eta^{6$  $\times \eta^{7}\beta^{2}ih + 4\eta^{6}\beta^{2} + 20\eta^{5}\beta^{3} - 24\eta^{6}\beta^{3}h - 24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 22\eta^{5}\beta^{3} - 12\eta^{6}\beta^{2} - 12\eta^{4}\beta^{4} - 2\eta^{7}\beta) \operatorname{e}^{-(\eta + \beta)(i - 1)h} + \epsilon(-24\eta^{4}\beta^{5}h - 2\eta^{5}\beta^{5} - 12\eta^{6}\beta^{2} - 12\eta^{6}\beta^{5} - 12$  $\times \eta^{4}\beta^{4} - 4\eta^{7}\beta^{2}ih - 24\eta^{4}\beta^{5}ih - 44\eta^{5}\beta^{4}ih + 68\eta^{4}\beta^{5}ih + 68\eta^{5}\beta^{4}ih + 24\eta^{3}\beta^{6}ih + 28\eta^{6}\beta^{3}ih + 46\eta^{4}\beta^{4} + 56\eta^{3}\beta^{5}ih + 68\eta^{5}\beta^{4}ih + 68\eta^{5}\beta^{5}ih + 68\eta^{5}\beta^$  $+2\eta^6\beta^2+16\eta^5\beta^3+4\eta^7\beta^2ih+24\eta^2\beta^6-4\eta^6\beta^2-24\eta^6\beta^3ih-20\eta^5\beta^3)\,\mathrm{e}^{-(\eta+\beta)ih}.$ 

#### D.2 Covariance

Dans la formule (9.10) de la covariance du processus non-stationnaire agrégé,  $f_c(\epsilon, \eta, \beta, h, i, k)$  et  $g_c(\epsilon, \eta, \beta, h, i, k)$  sont données par :

$$\begin{split} &f_c(\epsilon,\eta,\beta,h,i,k) = \{24\eta^2\beta^5 - 24\epsilon\beta^5\eta + 8\epsilon\beta^4\eta^2 + 16\beta^6\eta + 8\beta^2\epsilon\eta^4 - 24\beta^3\eta^4 - 16\beta^6\epsilon - 8\beta^2\eta^5 + 24\beta^3\epsilon\eta^3 - 8\eta^3\beta^4 \\ &+ \epsilon(24\beta^3\eta^3 - 8\beta^4\eta^3ih + 8\beta^4\eta^2 - 24\beta^5\eta + 24\beta^5\eta^2ih + 16\beta^6\etaih - 24\beta^3\eta^4ih + 8\beta^2\eta^4 - 8\beta^2\eta^5ih - 16\beta^6)\} \\ &\times e^{-\eta h(k-1)} + (8\epsilon\beta^4\eta^2 - 24\epsilon\beta^5\eta + 24\beta^5\eta^2 + 8\beta^2\epsilon\eta^4 + 24\beta^3\epsilon\eta^3 - 16\beta^6\epsilon - 24\beta^3\eta^4 - 8\beta^2\eta^5 - 8\eta^3\beta^4 + 16\beta^6\eta \\ &+ 16\beta^6\etaih - 16\beta^6\eta h - 8\beta^4\eta^3ih + 24\beta^5\eta^2ih + 8\beta^4\eta^2 + 8\beta^4\eta^3h - 24\beta^5\eta^2h - 24\beta^5\eta + 24\beta^3\eta^3 + 8\beta^2\eta^4 - 8\beta^2\eta^5 \\ &\times ih + 24\beta^3\eta^4h + 8\beta^2\eta^5h - 24\beta^3\eta^4ih - 16\beta^6) e^{-\eta h(k+1)} + (16\epsilon\beta^6\eta h - 8\epsilon\beta^4\eta^3h + 24\epsilon\beta^5\eta^2h - 32\epsilon\beta^6\etaih - 32) \\ &\times \epsilon\beta^4\eta^2 + 96\epsilon\beta^5\eta + 16\epsilon\beta^4\eta^3ih - 48\epsilon\beta^5\eta^2ih + 16\beta^2\eta^5 + 16\beta^2\epsilon\eta^5ih + 48\beta^3\epsilon\eta^4ih - 8\beta^2\epsilon\eta^5h - 24\beta^3\epsilon\eta^4h - 32\beta^2 \\ &\times \epsilon\eta^4 - 96\beta^3\epsilon\eta^3 + 48\beta^3\eta^4 + 16\eta^3\beta^4 + 64\epsilon\beta^6 - 32\beta^6\eta - 48\beta^5\eta^2) e^{-\eta hk} + \epsilon(24\beta^3\eta^4h + 8\beta^4\eta^3h + 8\beta^2\eta^5h - 16\beta^6) \\ &\times \eta h - 24\beta^5\eta^2h) e^{-\eta(i+k)h} + \epsilon(16\beta^6\eta h - 8\beta^2\eta^5h - 24\beta^3\eta^4h - 8\beta^4\eta^3h + 24\beta^5\eta^2h) e^{-\eta(i+k-1)h} \end{split}$$

et

$$\begin{split} g_c(\epsilon,\eta,\beta,h,i,k) &= (6\eta^2\beta^5 - 4\epsilon\beta^5\eta - 2\epsilon\beta^4\eta^2 + 4\beta^6\eta + 2\eta^3\beta^4) \, \mathrm{e}^{-\eta h(k-1)} + \epsilon \{ -4\eta^3\beta^3 - 6\eta^4\beta^3h(i-1) - 2\eta^5\beta^2 \\ &\times h(i-1) + \eta^5\beta + \eta^4\beta^2 - 4\eta^3\beta^4h(i-1) - 4\eta^2\beta^4 \} \, \mathrm{e}^{-\eta h(i-1) - \beta h(i+k)} (1 - \mathrm{e}^{\beta h}) + \{ -4\eta^4\beta^3 - 2\eta^6\beta - 6\eta^5\beta^2 \\ &+ \epsilon (\eta^6 + 2\eta^4\beta^2 + 3\eta^5\beta + 4\eta^4\beta^3h - 4\eta^4\beta^3ih + 6\eta^5\beta^2h + 4\eta^4\beta^2 + 6\eta^5\beta - 2\eta^6ih\beta + 2\eta^6h\beta + 2\eta^6 - 6\eta^5\beta^2ih) \} \\ &\times \mathrm{e}^{-\beta h(k+1)} + \epsilon (4\beta^5\eta + 2\beta^4\eta^2) \, \mathrm{e}^{-\eta h(2i-2+k)} - \epsilon (3\eta^5\beta + 2\eta^4\beta^2 + \eta^6) \, \mathrm{e}^{-\beta h(2i-2+k)} + \{ \epsilon (4\beta^6\eta ih - 4\beta^6 + 2\beta^4\eta^3 ih + 6\beta^5\eta^2 ih - 2\beta^4\eta^2 - 2\beta^4\eta^3 h - 6\beta^5\eta^2 h - 6\beta^5\eta^2 + 4\beta^6 - 2\beta^4\eta^2 - 4\beta^5\eta + 2\beta^4\eta^3 ih - 2\beta^4\eta^2 - 6 \\ &\times \beta^5\eta + 6\beta^5\eta^2 ih + 4\beta^6\eta ih - 4\beta^6 + 4\beta^6\eta^2 + 6\beta^5\eta^2 + 2\eta^3\beta^4 \} \, \mathrm{e}^{-\eta h(k+1)} + (4\eta^6\beta - 2\eta^6\epsilon h\beta - 4\eta^4\epsilon\beta^3 h - 6\eta^5\epsilon \\ &\times \beta^2h - 18\eta^5\epsilon\beta + 4\eta^6\epsilon ih\beta + 8\eta^4\epsilon\beta^3 ih + 12\eta^5\epsilon\beta^2 ih - 12\eta^4\epsilon\beta^2 - 6\eta^6\epsilon + 8\eta^4\beta^3 + 12\eta^5\beta^2) \, \mathrm{e}^{-\beta hk} + \epsilon (8\beta^6 - 6 \\ &\times \beta^5\eta^2h - 2\beta^4\eta^3h - 4\beta^6\eta h + 12\beta^5\eta^2 ih + 8\beta^6\eta ih + 4\beta^4\eta^2 + 12\beta^5\eta + 4\beta^4\eta^3 ih - 4\beta^4\eta^2 - 8\beta^6\eta) \, \mathrm{e}^{-\eta h(2i+k-1)} \\ &+ \epsilon (-4\beta^6 + 6\beta^5\eta^2h + 2\beta^4\eta^3h - 4\beta^6\eta ih + 4\beta^6\eta h - 6\beta^5\eta^2 ih - 2\beta^4\eta^2 - 6\beta^5\eta - 2\beta^4\eta^3 ih) \, \mathrm{e}^{-\eta h(2i+k-2)} + (4\epsilon^2\eta^2 + 2\beta^2\eta^2) \, \mathrm{e}^{-\eta hk} - \epsilon (\eta^5\beta - \eta^4\beta^2) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-2)} + \epsilon (\eta^5\beta + \eta^4\beta^2) \, \mathrm{e}^{-\eta ih - \beta h(3i+k)} + \epsilon (\eta^4\beta^2 + \eta^5) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-3)} + \epsilon \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-2)} + \epsilon (\eta^5\beta + \eta^4\beta^2) \, \mathrm{e}^{-\eta ih - \beta h(3i+k)} + \epsilon (\eta^4\beta^2 + \eta^5) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-3)} + \epsilon \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-2)} + \epsilon (\eta^5\beta + \eta^4\beta^2) \, \mathrm{e}^{-\eta ih - \beta h(3i+k)} + \epsilon (\eta^4\beta^2 + \eta^5) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-2)} + \epsilon (\eta^5\beta + \eta^4\beta^2) \, \mathrm{e}^{-\eta ih - \beta h(3i+k)} + \epsilon (\eta^4\beta^2 + \eta^5) \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-3)} + \epsilon \, \mathrm{e}^{-\eta h(i-1) - \beta h(3i+k-2)} + \epsilon (\eta^5\beta + \eta^4\beta^2) \, \mathrm{e}^{-\eta h(2i+k)} + \epsilon \, \mathrm{e}^{-\eta h(2i+k)$$